Lecture Notes in Logistics
Series Editors: Uwe Clausen · Michael ten Hompel · Robert de Souza

Hans-Jörg Kreowski
Bernd Scholz-Reiter
Klaus-Dieter Thoben *Editors*

Dynamics in Logistics

Third International Conference, LDIC 2012 Bremen, Germany, February/March 2012 Proceedings

Springer

Lecture Notes in Logistics

For further volumes:
http://www.springer.com/series/11220

Hans-Jörg Kreowski · Bernd Scholz-Reiter
Klaus-Dieter Thoben
Editors

Dynamics in Logistics

Third International Conference, LDIC 2012 Bremen, Germany, February/March 2012 Proceedings

Springer

Editors
Prof. Dr.-Ing. Hans-Jörg Kreowski
Department of Computer Science
University of Bremen
Bremen
Germany

Prof. Dr.-Ing. Bernd Scholz-Reiter
Rector of the University of Bremen
Bremen
Germany

Prof. Dr.-Ing. Klaus-Dieter Thoben
Bremer Institut für Produktion
und Logistik (BIBA)
University of Bremen
Bremen
Germany

ISSN 2194-8917 ISSN 2194-8925 (electronic)
ISBN 978-3-642-35965-1 ISBN 978-3-642-35966-8 (eBook)
DOI 10.1007/978-3-642-35966-8
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2013934015

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

LDIC 2012 was the 3rd International Conference on Dynamics in Logistics held in Bremen (Germany) from 27th February to 1st March 2012 together with a Doctoral Workshop. The LDIC 2012 was held in conjunction with the new conference "The Impact of Virtual, Remote and Real Logistics Labs" (ImViReLL'12). Similar to its predecessors LDIC 2007 and LDIC 2009, it was organized by the Bremen Research Cluster for Dynamics in Logistics (Log*Dynamics*) of the University of Bremen in cooperation with the Bremer Institut für Produktion und Logistik (BIBA), which is a scientific research institute affiliated to the university.

The scope of the conference targets the identification, analysis, and description of the dynamics of logistic processes and networks. The spectrum of topics ranges from modeling and planning of processes and innovative methods like autonomous control and knowledge management to the new technologies provided by radio frequency identification, mobile communication, and networking. The growing dynamics in the area of logistics poses completely new challenges: logistic processes and networks have to rapidly and flexibly adapt to continuously changing conditions. LDIC 2012 provided a venue for researchers from academia and industry interested in the technical advances in dynamics in logistics. The conference addressed research in logistics from a wide range of fields, e.g., engineering, computer science, and operations research.

The LDIC 2012 proceedings consist of two invited keynotes and 49 papers selected by a strong reviewing process. The volume is organized into the following ten subject areas: "*Invited Papers*", containing the contributions of the invited speakers, "*Transport Logistics and Dynamic Routing*", "*Production Logistics and Job Shop Scheduling*", "*Modeling, Simulation, Optimization and Collaboration*", "*Identification Technologies*", "*Mathematical Modeling in Transport and Production Logistics*", "*Information, Communication, Risk and Failure*", "*Autonomous Control*", "*Global Supply Chains and Industrial Application*" and "*Considerations for a Future Internet of Things*".

We would like to thank the members of the program and organization committee and the secondary reviewers Jan Ole Berndt, Melanie Bloos, Marcus Ermler, Vaggelis Giannikas, Michael Görges, Marc-André Isenberg, Farian

Krohne, Sabine Kuske, Walter Lang, Melanie Luderer, Michael Lütjen, Jeanette Mansfeld, Afshin Mehrsai, Susanne Schukraft, Michael Teucke, Hendrik Thamer, Dirk Werthmann, and Jiani Wu for their help in the selection process. We are also grateful to Aleksandra Himstedt, Jakub Piotrowski, Ingrid Rügge, Dieter Uckelmann, and several other colleagues for their support in the local organization and for the technical assistance in running the conference system. Special thanks go to Ingrid Rügge and Marcus Seifert for organizing the doctoral workshop of LDIC 2012. We are particularly indebted to Caroline von Totth for her support in editing this volume and in careful unification of the print files of all the contributions. Moreover, we would like to acknowledge the financial support by the BIBA, the Research Cluster for Dynamics in Logistics (Log*Dynamics*), the Center for Computing and Communication Technologies (TZI), and the University of Bremen. Finally, we appreciate once again the excellent cooperation with the Springer-Verlag.

Bremen, November 2012

Hans-Jörg Kreowski
Bernd Scholz-Reiter
Klaus-Dieter Thoben

Committees

Program

Dieter Armbruster, Tempe (Arizona, USA)
Abdelaziz Bouras, Lyon (France)
Neil A. Duffie, Madison (Wisconsin, USA)
Enzo M. Frazzon, Florianópolis (Brazil)
Axel Hahn, Oldenburg (Germany)
Mark Harrison, Cambridge (UK)
Bonghee Hong, Pusan (Republic of Korea)
Alamgir Hossain, Newcastle upon Tyne (UK)
Kap Hwan Kim, Pusan (Republic of Korea)
Antonio G. N. Novaes, Florianópolis (Brazil)
Mykhailo Postan, Odessa (Ukraine)
Christian Prins, Troyes (France)
Antonio Rizzi, Parma (Italy)

Organization

Carmelita Görg
Hans-Dietrich Haasis
Aleksandra Himstedt
Michael Hülsmann
Herbert Kopfer
Hans-Jörg Kreowski (co-chair)
Michael Lawo
Rainer Malaka
Jakub Piotrowski
Robert Porzel
Ingrid Rügge

Bernd Scholz-Reiter (co-chair)
Jörn Schönberger
Marcus Seifert
Klaus-Dieter Thoben (co-chair)
Caroline von Totth
Dieter Uckelmann
Katja Windt

Contents

Part I Invited Papers

The Future of Manufacturing Utilizing Tagging Technology 3
Gisele Bennett

Fault Detection in Dynamic Vehicle Routing Operations 13
Antonio G. N. Novaes, Edson T. Bez and Paulo J. Burin

Part II Transport Logistics and Dynamic Routing

Knowledge Sharing in Intermodal Transport: A Multi-Agent Based Perspective 35
Jiani Wu and Hans-Dietrich Haasis

Emissions Minimization Vehicle Routing Problem in Dependence of Different Vehicle Classes 49
Heiko W. Kopfer and Herbert Kopfer

Product Intelligence in Intermodal Transportation: The Dynamic Routing Problem 59
Vaggelis Giannikas and Duncan McFarlane

Cargo Telematics for Operational Transport Excellence and Strategic Knowledge Management 71
Matthias Klumpp, Christof Kandel and Erik Wirsing

Part III Production Logistics and Job Shop Scheduling

Towards an Integrated Production and Outbound Distribution Planning Method . . . 85
Bernd Scholz-Reiter and Christian Meinecke

Supply Chain Integration by Human-Centred Alignment of Lean Production Systems . . . 93
Yilmaz Uygun and Natalia Straub

The Multiple Batch Processing Machine Problem with Stage Specific Incompatible Job Families . . . 113
Marc-Andrè Isenberg and Bernd Scholz-Reiter

Scheduling Multiple Batch Processing Machines with Stage Specific Incompatible Job Families . . . 125
Marc-Andrè Isenberg and Bernd Scholz-Reiter

A Comparison of Network Characteristics in Metabolic and Manufacturing Systems . . . 141
Till Becker, Mirja Meyer, Moritz E. Beber, Katja Windt and Marc-Thorsten Hütt

Part IV Modeling, Simulation, Optimization and Collaboration

Conflicting Optimization Goals in Manufacturing Networks: A Statistical Analysis Based on an Idealized Discrete-Event Model . . . 153
Reik V. Donner and Jörg Hanke

Modeling the Basic Cause-Effect Relationship Between Supply Chain Events and Performance . . . 163
Georg Heinecke, Jonathan Köber, Andreas Kunz and Steffen Lamparter

The Importance of Managing Events in a Build-to-Order Supply Chain: A Case Study at a Manufacturer of Agricultural Machinery . . . 175
Jonathan Köber and Georg Heinecke

A Dynamic Approach to Measure the Effect of Postponement Strategy on the Performance of Manufacturing and Logistics Processes . . . 187
Luiz Eduardo Simão and Mirian Buss Gonçalves

Event Management for Uncertainties in Collaborative Production Scheduling and Transportation Planning: A Review 199
Bernd Scholz-Reiter, Yi Tan, Nagham El-Berishy and José B. S. Santos Jr.

Study on the Application of DCOR and SCOR Models for the Sourcing Process in the Mineral Raw Material Industry Supply Chain 211
Raul Zuñiga, Marcus Seifert and Klaus-Dieter Thoben

Part V Identification Technologies

Service Oriented Platform for RFID-Enabled Real-Time Fleet Advanced Planning and Scheduling in Industrial Parks 223
L. Y. Pang, George Q. Huang and R. Y. Zhong

Towards Agile and Flexible Air Cargo Processes with Localization Based on RFID and Complex Event Processing 235
David Rüdiger, Moritz Roidl and Michael ten Hompel

Smart Agri-Food Logistics: Requirements for the Future Internet . . . 247
C. N. Verdouw, H. Sundmaeker, F. Meyer, J. Wolfert and J. Verhoosel

Design and Implementation of a Virtual Reader for Generating Multiple Tag Data 259
Jiwan Lee, Wooseok Ryu, Joonho Kwon and Bonghee Hong

Dynamic Tag Identity-Based Encryption Scheme for Logistic Systems 267
Jongseok Choi and Howon Kim

Queuing Buffer Method for Enhancing Reliability of Capturing Application 277
Taewoo Nam, Inhwan Lee, Byeongsam Kim and Keunhyuk Yeom

Part VI Mathematical Modeling in Transport and Production Logistics

Dynamic Optimization Model for Planning of Integrated Logistical System Functioning 291
Iryna V. Morozova, Mykhaylo Ya. Postan and Sergey Dashkovskiy

A Graph Model for the Integrated Scheduling of Intermodal Transport Operations in Global Supply Chains 301
Willian Chaves Mates, Enzo Morosini Frazzon, Jens Hartmann and Sérgio Fernando Mayerle

Exploitation of Due Date Reliability Potentials: Mathematical Investigation of the Lead Time Syndrome 313
Mathias Knollmann and Katja Windt

A Multi-Level Programming Model and Solution Algorithm for the Location of Distribution Centers with Uncertainty Demand . . . 327
Patareeya Lasunon and Raknoi Akararungruangkul

A Method to Estimate the Accumulated Delivery Time Uncertainty in Supply Networks 337
Mehdi Safaei, Safir Issa, Marcus Seifert, Klaus-Dieter Thoben and Walter Lang

An LMI Approach to Exponential Stock Level Estimation for Large-Scale Logistics Networks 349
Hamid Reza Karimi

Part VII Information, Communication, Risk, and Failure

Characterization of Thermal Flow Sensors for Air Flow Measurements in Transport Containers 361
Safir Issa and Walter Lang

Design and Implementation of Location-Based Handoff Scheme for Multimedia Data 371
Dong Hyun Kim and Jong Deok Kim

Reconfiguring Multi-Rate Wi-Fi Mesh Networks with Flow Disruption Constraints 383
Yafeng Zhou, Sang-Hwa Chung and Han-You Jeong

Maintenance Work Order Based on Event-Driven Information of Condition-Monitoring Systems—A Genetic Algorithm for Scheduling and Rescheduling of Dynamic Maintenance Orders . . . 395
Marco Lewandowski, Oliver Siebenand, Stephan Oelker and Bernd Scholz-Reiter

A Procedure for the Selection of a Supply Network Risk Mitigation Strategy in Relational Arrangements 405
José B. S. Santos Júnior, Sérgio A. Loureiro and Orlando F. Lima Júnior

Optimization of Failure Behavior of a Decentralized High-Density 2D Storage System 415
Kai Furmans, Kevin R. Gue and Zäzilia Seibold

Part VIII Autonomous Control

Distributed Reinforcement Learning for Optimizing Resource Allocation in Autonomous Logistics Processes 429
Jan Ole Berndt and Otthein Herzog

Improving Returnable Transport Equipment Management with an Autonomous Approach 441
Rebecca Aggarwal and Ming K. Lim

Stability Analysis Scheme for Autonomously Controlled Production Networks with Transportations 453
Sergey Dashkovskiy, Michael Görges and Lars Naujok

Towards the Learning Behaviour and Performance of Artificial Neural Networks in Production Control. 465
Bernd Scholz-Reiter and Florian Harjes

Part IX Global Supply Chains and Industrial Application

A Framework of the Forces Influencing the Adaptation of the SCOR Model to the Situation of the Manufacturing Industry in Developing Countries. 477
Fasika B. Georgise, Klaus-Dieter Thoben and Marcus Seifert

The Impact of Near Sourcing on Global Dynamic Supply Chains: A Case Study 489
Anna Corinna Cagliano, Alberto De Marco and Carlo Rafele

Simulation-Based Analysis of Integrated Production and Transport Logistics. 499
Enzo Morosini Frazzon, Joarez Pintarelli, Thomas Makuschewitz and Bernd Scholz-Reiter

Logistics Integrated Product Development in the German Automotive Industry: Current State, Trends and Challenges 509
Ingrid Göpfert and Matthias Schulz

Efficient and Safe Transporting of Steel Sheets with Permanent Magnets 521
Johannes Schröck and Bernd Orlik

Part X Considerations for a Future Internet of Things

The Effective Method for Distributed Information Gain in EPC Network 531
Taewoo Nam, Byeongsam Kim and Keunhyuk Yeom

Man–Machine-Interaction in the Field of Logistics: Example "Internet of Things" 545
Lars Windelband and Georg Spöttl

Sensor Reader Emulator for Integrating Sensor Networks into EPCglobal Network 557
Joo-Ho Lee, Oh-Heum Kwon and Ha-Joo Song

A Design of Distributed Discovery Service of EPCglobal. 571
Gihong Kim, Bonghee Hong and Joonho Kwon

Part I
Invited Papers

The Future of Manufacturing Utilizing Tagging Technology

Gisele Bennett

Abstract Lean, agile, just-in-time, future, green, energy efficient and robust are just some of the terms used to describe the next generation manufacturing. The focus for manufacturing is dependent on the year, the challenges, the feature technologies, the consumer demands, and business pressures. This paper discusses the potential of the future of manufacturing, which utilizes technologies that allow for real time visibility of location and condition of assets for use in decision making to anticipate the manufacturing process and meet demands.

Introduction

Manufacturing process flow dates back to 1574 when King Henry III of France watches the Venice Aresnal build complete galley ships in less than an hour using continuous flow processes (Lean Manufacturing History Timeline 2012). Manufacturing research process, supply chain logistics, energy savings, policy, regulations, and consumer demands are inputs to a decision and control algorithm that can be utilized to achieve desired outcomes in manufacturing. Even as early as 1760 French General Jean-Baptiste de Gribeauval understood the significance of standardized designs and interchangeable parts to facilitate repairs in the battlefield. At the beginning of the 20th century, automation and processes are introduced to manufacturing by Ford and Toyota leading to the introduction of just-in-time in 1938. All of the early implementations required manual oversight until robotics and automation can be introduced. The information in this chapter is intended to aid the

G. Bennett (✉)
Georgia Tech Research Institute, Georgia Tech, Atlanta, GA 30332, USA
e-mail: gisele.bennett@gtri.gatech.edu

H.-J. Kreowski et al. (eds.), *Dynamics in Logistics*, Lecture Notes in Logistics,
DOI: 10.1007/978-3-642-35966-8_1,

system engineer in how to achieve success in implementing Radio Frequency Identification (RFID) in a factory. By utilizing technology to generate data that can be transformed into information to create a revolutionary approach to manufacturing while making evolutionary changes through knowledge of location and condition of assets as inputs to optimization algorithms.

Manufacturing can be revolutionized with real time data that can be converted to information and decisions that can be implemented just in time to produce the desired outcomes. Just-in-Time (JIT) conjures up definitions in manufacturing for planned elimination of all waste and continuous improvement of productivity. Although just-in-time is commonly used in manufacturing, we use the term as part of a strategy to utilize real time data in the decision and manufacturing processes. The desired outcomes could be to improve manufacturing throughput, reduce cost, or energy consumption. The optimal outcome could be to optimize all desired outputs resulting in a multivariate optimization problem that must consider internal and external inputs. These inputs could be labor availability, health of equipment [Mean-Time-Between-Failure (MTBF)], while outside factors such as weather, traffic problems or seasonal absence due to vacation could influence workforce availability. To achieve optimization, it is important to understand the condition and location of all assets and the desired metric or goal to achieve at the manufacturing plant. Wireless manufacturing is beginning to proliferate the concept of the next generation factory (Huang et al. 2008).

To have full visibility, the manufacturer will need to have control over the data formats and supply elements. Most manufacturers, because of globalization and outsourcing, are becoming integrators. The challenge for competitiveness transforms into a management of the global network for all aspects of manufacturing from design to production and the ability to have a predictable and inexpensive logistics network. As a consequence, intelligent prediction algorithms will need to factor other elements such as human behavior, environmental, and business objectives into the optimization of the factory. The factory truly becomes a living, dynamic machine of production.

The objective of this paper is to present an overview of the challenges, opportunities, and references to research and experiments, with a focus on the use of RFID, to establish areas that need to be addressed before achieving full asset visibility. This paper will review Radio Frequency Identification (RFID) technologies for some of the asset visibility needed in manufacturing and the supply chain that supports a manufacturing environment. The use of the term assets in this paper is intended to include equipment, supplies, and people.

RFID Challenge in a Manufacturing Environment

RFID is an enabling technology that can cost-effectively identify, monitor, track, and locate assets. Assets are defined as materials, components, processes, and people. RFID systems were categorized into 5 classes during the flurry of RFID in the consumer and US Department of Defense markets in the early 21st century. The five categories begin with a basic passive tag and progress to a tag that had embedded sensors and can initiate communication. Rather than get caught up in the classes of tags and various International Organization for Standardization (ISO) standards, for the purpose of this paper restrict the reference of RFID to passive or active and to the implementation and physics of RFID systems in a manufacturing environment. Bennett and Herkert (2008) provide a detailed overview of the active RFID systems and deployment considerations. Although some of the standards have been modified, an overview of the basic physics and frequency allocations and ISO standards is provided by Bennett (2007). RFID systems are made up of a tag, air interface (tag to reader), the reader, and network to connect readers to information systems. The basic system operation is to have a reader send energy to a tag for power then the tag send Identification (ID) data back to the reader. One of the primary differences between a passive and active system is an active RFID system has tags with batteries that aid in increasing the read ranges between the reader and tag.

RFID historically has been used by Supply Chain Logistics (SCL) applications and not until the past decade have papers on the use of RFID in manufacturing started to increase in the literature. RFID has numerous challenges in a manufacturing environment because of the harsh RF conditions. The challenges, which are primarily with the basics of electromagnetics include: absorption, shielding, multi-path, diffraction, and antenna detuning to name a few. A factory has impediments to RFID systems that include racks, shelves, pallets, machinery, metal, walls, RF noise, and other assets that degrade RF communications (Remley et al. 2008; Hellstrom and Wiberg 2009; Griffin et al. 2006). However, through a combination of technologies, the right site survey and choice of RFID technology, real time asset visibility and use of RFID in a manufacturing environment is possible.

A factory or warehouse is a dynamic environment where there are constantly moving parts, moving machinery, moving tags, etc. Although it may be feasible to implement an RFID system in a static industrial environment, the dynamic aspect imposes a new set of challenges, even if the dynamic parts of the environment are completely unrelated to the system or the tags being tracked. Issues that are affected by the environment are outlined in this section as a guide for areas to consider before utilizing an RFID system in this type of environment. Some key areas to identify when choosing an RFID system to use in a manufacturing environment are:

1. Tag Orientation: Because the orientation of the tag with a reader is important, we need to understand how the changing orientation of the tag and the tagged part affect read rates and visibility. As the orientation of the tagged part is changed, the orientation of the tag will certainly be changed. This change will affect the signal strength received and therefore transmitted back to the reader. It is important to understand how this changes the performance of the specific system chosen and how can the change be mitigated. Furthermore, a change in orientation of the part might result in the part lying between the reader and the

tag. By understanding the particular manufacturing process, we would want to change the locations of the readers or the positions of the tags be chosen to minimize any consequences. Mercer et al. (2011) provide recent experiments showing the importance of tag orientation.

2. Moving Obstacles: All factories have some sort of automation, robotics, and machinery creating lots of moving, metal, unmanned pieces of equipment. Assuming that the RFID tags are chosen appropriately such that movement of the part does not greatly affect the tags' orientation relative to the reader, we would want to design the RFID deployment to mitigate read capability due to other equipment obscuring the field of view of the reader. Some experiments have been conducted on static obstructions that show the results depend on the entire area of interest, other objects and materials in the field of view, and the type of tags used (James et al. 2010).
3. Tag Distance Resolution: The ability to distinguish what order items are passing through an assembly line is referred to as sequencing and is very important if we are to have a smart and adaptable manufacturing plant that allows for custom orders. To accurately utilize sequencing, we need to be able to read one tag at a time for parts that are closely clustered. Considering that most of current manufacturing takes place on an assembly line, it is safe to assume that tags will constantly be brought into and out of range of the readers. Most readers are designed for a wide field of view leading to multiple tag reads and some tags that are missed in the scan. However, if we are trying to achieve sequencing, which requires higher resolution in distinguishing between tag 1 and tag 2, we will require a reader that has a narrower field of view. Reprogramming readers for a narrow field of view is possible giving the opportunity to actually have a sequencing feature in the asset information. Sequencing is important if we are to achieve a robust and agile manufacturing process.
4. Tag Read Rates: Current manufacturing processes can have very fast throughputs. All tags and readers are not created equal and before deployment in a high throughput section of the manufacturing processes, it will be important to determine how fast the tags can move while still allowing a high read probability. Strategic placement of readers can improve the read rates. Assuming normal operation of the manufacturing facility, it is important not to implement technology that slows the manufacturing throughput. Recent tests have been conducted to evaluate read rates as a function of objects and tag placement (Mercer et al. 2011). The results showed good read rates for semi-passive RFID tags that are battery assisted. However, when testing passive tags none achieved 100 % read rate even with increased reader exposure times (the time the reader sees the tags—180s for this experiment).

Implementation is a function of the environment, system performance, and desired outcomes. We have seen that RFID system performance changes depending on the manufacturer and advancements in technology. A manufacturer of a passive RFID system one year could yield a less than desirable performance and with advances in the next generation technology the same vendor could produce a new system with significant performance improvements. The key point is that testing never stops and should be conducted on a regular basis to determine performance. The papers referencing test results in this chapter will be obsolete in a couple of years and the experiments will need to be repeated. Especially in a manufacturing environment, site surveys and developing requirements for the application are critical to proper execution (Gaukler and Seifert 2007).

The Manufacturing Broker and Integrator

As manufacturing migrates to a supply management and a complex network integration with its suppliers and subassemblies, a network to manage the entire process becomes even more vital. Asset visibility, through RFID and other technological means, through the global supply chain will facilitate better coordination through accurate knowledge of the system state. A system as complex as a manufacturing environment can achieve inventory management using tagged and tracked data from parts for visibility. The key to success is to have agility and speed with the ability to recognize and rapidly adapt to the particular and changing demands that arise with consumer demand. Dynamic scheduling can be achieved through real-time information using intelligent tracking technologies. Although we vaguely define the manufacturing system, it is obvious that the future of most manufacturing is assembly with a complex outreach into numerous subsystems that must work together to produce the end product.

Sustainability

RFID tagged parts offer another level of future system support. RFID tagging systems lend themselves not only to the manufacturing process but the product lifecycle as well. At the end of life of the product, certain parts could easily be identified that require specific disposal methods such as those parts that contain environmentally hazardous materials. The original equipment manufacturer (OEM) through their RFID tags can identify other parts in the product, that are reusable, and information such as service history and age could be retrieved to determine the reusability. Parts that are too old to be safely reused could be immediately taken to material recycling, while other reusable modules could be reentered into supply chain.

The early adoption of RFID tags usage by Michelin Tires for tread reading and tire tracking created an added value to the manufacturer and consumer. The initial use of RFID in tires in the 2004 was to meet the US government TREAD (Transportation, Recall, Enhancement, Accountability, and Documentation) Act that mandates car makers to track tires from model year 2004 and beyond for recall problems. With the standards for tagging and the level of information provided by an EPC code, more detailed information is provided to the users for the unfortunate situations of recalls. Knowing when and where parts were manufactured, changes general recalls to targeted recalls saving significant costs in broad replacement of defective parts. The use of the RFID tag by Michelin for identification evolved into a value added feature that later included linkage to temperature and pressure that can be read through RFID readers.

Optimized Manufacturing and Smart Parts

Information about parts that are tagged and read by readers at a manufacturing station could be used to control and change manufacturing station to perform functions required for that part. Once complete, that information can be captured and entered into a central control system. The parts now are smart and part of the manufacturing assembly decision-making. The concept of smart parts has been presented in numerous configurations from warehouse management and movement of inventory to manufacturing (Zhekun et al. 2004). Automotive manufacturers have adopted the use of RFID tags to carry customer individual orders. The tags used are read/write capable and travel with the car during the production process. This tracking allows for integration into existing systems, reduces errors in the assembly of the unique car configuration, and can be reused at the end of the assembly. The use of RFID combined with GPS is presented by Brewer et al. to demonstrate the utility of multiple intelligent tracking technologies for use in manufacturing (Brewer et al. 1999).

Prognostics

Up to this point, asset visibility was focused on the use of RFID and location information. We have subsequently seen examples of RFID used for equipment control and post production information for value added and possibly product recalls. Although RFID tags are not going to provide prognostics information directly the use of RFID tags with integrated sensors could provide pieces of data that contribute to the overall health monitoring of a system. Because the factory is fully outfitted with readers, the infrastructure is in place to add smarter tags for use in monitoring temperature or relative humidity in a particular area that requires accurate control, or emissions from machinery that could be indicators of failure modes. Prognostics is a wide field and subject to extensive research but for this article we are speculating that there are enough sensing modalities that can be added to an RFID enabled tag to utilize the information for factory health monitoring. Applications of prognostics range from diagnostics and predictions of machinery failures to using prognostics in the logistics system. Lopez De La Cruz et al. (2007) define a prognostics logistics structure as data acquisition into an information system that is an input into a logistics model that yields future predictions. Chen et al. (2010) introduce the approach used in prognostics and health monitoring with RFID tags with integrated sensors such as temperature and humidity to collect data for processing. The point of introducing prognostics in this chapter is to exploit the RFID infrastructure used for asset tracking to add another data and information element to the overall manufacturing process. This information can be used for scheduling and routing of parts, e.g. rather than have scheduled maintenance for equipment use prognostics to repair on a usage basis. Having advanced knowledge of failures allows for optimal allocation of the remaining resources.

Decision Algorithms

Now that we have real time information of our assets and possibly health information on our machinery, what do we do with this information and what other information and technology do we to create the agile factory. Earlier we introduced the use of external information for manufacturing and process flow. One such information is weather—Meteorological (METAR)—data which is readily available and accurate at least 24 h or greater. Seasonal events for localized regions are also readily available. Although the future factory could have no workforce, there is a greater chance that the workforce skills will evolve and possibly be reduced. In either case, using this additional data into an optimization equation for the factory as external inputs for resource availability could change the production process. Suppose a snow storm is expected, energy consumption will increase and labor availability could be reduced. Can this information be used to change the production rate? Suppose it is spring break and the majority of the labor force follows that holiday schedule resulting in a reduction in labor. There are endless scenarios and theories on the use of external data in a factory setting and the decision algorithms to use. The focus was not on the algorithms for this paper but on how to get the data for the algorithms and to ensure accuracy and consistency in the information.

Conclusions

Real-time tracking in a manufacturing plant has challenges due to the physics behind the operation of RF systems in harsh, robust, and dynamic environments. Conducting the appropriate site survey, designing a system architecture, and constant evaluation of tagging system performance increases the probability of successful deployment of RFID or any other tagging technology. The tagging and condition information can produce benefits that include: tracking of units to gain real-time visibility; rapid location of components and supplies, which lead to minimized costs of locating inventories (reduced man-hours); aid in recalls initiated by suppliers through accurate association of unique parts thus reducing cost of bulk recalls; improved sequencing for parts installation resulting in a more agile manufacturing process; adjustment in policy to use knowledge of location and vendor availability to improve timing for replenishment orders; and automatic reconfiguration for the future manufacturing plant to adjust production flow in real time. RFID is one mode for intelligent tracking with significant benefits. For successful implementation, it is critical to evaluate the current state of the technology and develop a system architecture that can provide the necessary data in a timely manner to users or systems that need turn that data into information for decision making.

Acknowledgments The author would like to acknowledge the numerous discussions with Drs. Susan Smyth and Ninjang Huang on manufacturing and the collaborations and assistance in laboratory measurements and background research by Allison Mercer, Ryan James, James Cai, Elani Spring, and Joseph Goldberg.

References

Bennett G (2007) RFID in the USA. In: Banks J, Pachano MA, Thompson LG, Hanny D (eds) RFID Applied, Wiley, NJ

Bennett G, Herkert R (2008) Deployment considerations for active RFID systems. In: Miles S, Sarma S, Williams J (eds) RFID technology and applications, Cambridge University Press, Cambridge

Brewer A, Sloan N, Landers T (1999) Intelligent tracking in manufacturing. J Intell Manuf 10:245–250

Chen S, Tom K, Thomas L, Pecht M (2010) A wireless sensor system for prognostics and health management. IEEE Sens J 10(4):856–862

Gaukler G, Seifert R (2007) Applications of RFID in supply chains. In: Jung H, Chen FF, Jeong B (eds) Trends in supply chain design and management: technologies and methodologies, Springer, London

Griffin J et al (2006) RF tag antenna performance on various materials using radio link budgets. IEEE Antennas Wirel Propag Lett 5(1):247–250

Hellstrom D, Wiberg M (2009) Exploring an open-loop RFID implementation in the automotive industry. In Proceeding 14th IEEE conf. emerging technologies and factory automation, Mallorca, Spain, pp 1543–1546

Huang GQ, Zhang Y, Chen X, Newman ST (2008) RFID-enabled real-time wireless manufacturing for adaptive assembly planning and control. J Intell Manuf 19(6):701–713

James RK, Mercer AJ, Bennett G, McGuinness D, Smith B, Dowling J (2010) RF obstructions and European RFID systems, Poster Session IEEE-RFID Orlando, 2011

Lean Manufacturing History and Timeline (2012) Factory strategies group, LLC. http://www.superfactory.com/content/timeline.html

Lopez De La Cruz AM, Veeke HPM, Lodewijks G (2007) Applicability of RFID in the prognostics of logistics systems, IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI), Aug 2007

Mercer AJ, James RK, Bennett G, Patel P, Johnston C, Cai J (2011) Performance testing of RFID systems with RF-Harsh materials. In: Proceeding IEEE international conference on RFID-technologies and applications (RFID-TA), Sitges, Spain

Remley K et al (2008) NIST tests of the wireless environment in automobile manufacturing facilities. NIST Technical Note 1550, Oct 2008, pp 1–122

Zhekun L, Gadh R, Prabhu BS (2004) Applications of RFID technology and smart parts in manufacturing. In: Proceeding ASME 2004 Desiging Engineering Technical Conferences and Computers and Information Engineering Conference, Salt Lake City, Utah, 2004

Fault Detection in Dynamic Vehicle Routing Operations

Antonio G. N. Novaes, Edson T. Bez and Paulo J. Burin

Introduction

The explosive growth in computer, communication, and information technology in recent years, together with dramatic changes in organizations and markets, have opened new forms of operating manufacturing and transport activities in an integrated and collaborative way (Goel 2008). To optimize performance, supply-chain functions must operate in a coordinated manner. But the actual circumstances observed in these operations make it difficult to implement it in many instances. Truck breakdowns, road traffic congestions, labour absences, customer's cancel or postponement of orders, among other random events, generate deviations from the basic plans. Thus, the management of these integrated systems must be performed in a dynamic way, revising the plans and schedules whenever it becomes necessary and when system failures require corrective interventions.

An unbalanced and unstable integration of manufacturing and transport systems can impair the competiveness of supply chains. This integration is even more relevant along global supply chains due to longer transport lead-times and the network complexity of manufacturing processes. Nowadays, production and transport scheduling are still carried out sequentially, due in part to their

A. G. N. Novaes (✉)
Federal University of Santa Catarina, Rua Beija-Flor 112,
Florianopolis 88062-253 SC, Brazil
e-mail: novaes@deps.ufsc.br

E. T. Bez
Univali – Itajaí Valley University, São José, SC, Brazil
e-mail: edsonbez@gmail.com

P. J. Burin
Federal University of Santa Catarina, Florianópolis, SC, Brazil
e-mail: pburin@gmail.com

H.-J. Kreowski et al. (eds.), *Dynamics in Logistics*, Lecture Notes in Logistics,
DOI: 10.1007/978-3-642-35966-8_2,

complexity and current lack of appropriate heuristics for supporting a desirable integration at the operational level. Especially within dynamic environments, production and transport systems must be properly integrated so that efficiency, responsiveness and flexibility could be achieved and sustained. Specially, the vision of a supply chain as synchronized systems of material and information flows requires that transport capabilities, level of utilization of resources and transit lead-times be appropriately employed in order to get the most effective and sustainable production scheduling (Frazzon et al. 2010).

The increasing complexity of technological processes, the availability of advanced sensor devices, and the existence of sophisticated information processing systems, have opened the way to detect abrupt as well as latent changes in some characteristic properties of a system. Fault detection and diagnosis applied to automatic control of technical systems have been extensively investigated in the literature (Basseville and Nikiforov 1993; Isermann 1997, 2005; Simani et al. 2010), but its concepts and methods have not been extended so far to services such as freight transport and logistics. The objective of fault detection in integrated manufacturing and logistics systems is to anticipate counteractions in order to avoid malfunctions and unexpected interruptions. Many fault monitoring problems can be seen as the problem of detecting a change in the parameters of a dynamic stochastic system. Model-based fault diagnosis methods are designed to detect abnormal situations confronting real data with modelling estimates. It is assumed that a discrepancy signal is directly or indirectly linked to a fault. Care must be taken to bypass model mismatches or noise in real measurements, which can erroneously be seen as a fault, giving rise to a false alarm in detection. These considerations have led to research efforts toward robust methods, with the aim of minimizing such drawbacks (Simani et al. 2010).

Dynamic vehicle routing problems (*DVRP*) have received increasing attention among researchers (Psaraftis 1995; Larsen 2001; Ribeiro and Lorena 2005; Larsen et al. 2008; Golden et al. 2008; Novaes et al. 2011). These problems are usually related to efficiently assigning vehicles to tasks, such as picking-up components from *OEM* facilities in a row, delivering cargo, or accomplishing other services in a previously defined order, so that tasks are completed within a certain time limit and vehicle capacities are not exceeded (Figliozzi 2007, 2010). But in large and congested urban areas, particularly in developing countries, transport operators tend to assign larger numbers of visits to their vehicles in order to increase revenue. This often leads to non-performed orders at the end of the daily cycle-time, impairing the logistics service level and postponing tasks to next day, or even later. This happens because, due to the volatile traffic conditions and the great number of random variables along the route, the vehicle cycle-time usually shows great variability. But even assuming that the fleet of vehicles has been well dimensioned, there are situations in which the traffic becomes exceptionally over-congested due to severe accidents, unpredictable public transport strikes, abnormal weather conditions, etc. However, when operating in a production schedule, comprising the pick-up of components from several *OEM* facilities in a row and carrying them to an assembler company, the vehicle has to accomplish its tasks within a pre-established JIT service level.

Traffic information systems, which have been installed in some large cities of the world, tend to increase the flow of vehicles by allowing higher vehicular speeds and by offering less-congested alternative routes to drivers (Fleischmann et al. 2004). The benefits of using such traffic navigational systems in connection with vehicle routing in congested urban areas cannot be denied. But in developing countries, the required large investments to install such systems often forbid its extensive use. One of the objectives of this work is to show that simple dynamic vehicle routing procedures can dramatically improve the logistics performance of the servicing system. With an on-board computer, a fault-detection software, and simple telematics devices linking the vehicle to nearby collaborative agents (other vehicles and the central depot), it is possible to attain better performance levels. By analysing vehicle operational data at specific regeneration points along the route, it is possible to anticipate the occurrence of unperformed tasks, emitting information to other agents (vehicles, central depot), and transferring part of the tasks to other participants. With this procedure the occurrence of unperformed tasks at the end of a vehicle cycle-time can be dramatically reduced.

In a *DVRP*, not all information relevant to the planning of the routes is known by the planner when the routing process begins, and information may change after the initial routes have been constructed (Larsen et al. 2008). In the application object of this work, although the vehicle service is fully planned in advance, the possible transfer of tasks to other agents and the eventual reprogramming of visits lead to changes in the routing process, thus characterizing a dynamic behaviour (Psaraftis 1995; Larsen 2001).

The Static Routing Problem

When modelling dynamic logistics problems it is necessary to quantify a number of random parameters to be used in the main mathematical model. Larsen (2001) dedicates a full chapter of his work to the computer simulation of such data. The reason is that real-life datasets are very often not available in such detail and accuracy as to support a thorough investigation of dynamic problems. Then, randomly generated data and simulation are frequently used when designing dynamic logistics systems. We make use of such a technique to analyse important parameters related to the routing vehicle cycle.

Throughout the paper an empirical case study will be examined based on a real-life Brazilian urban scenario. Let us consider an operating district of area A containing n *OEM* suppliers. The vehicle assigned to the district leaves the depot early in the morning, goes to the assigned district, performs the collecting service visiting the *OEM* suppliers, and goes to the assembler plant when all tasks are completed, or when the maximum allowed working time per day is to be reached, whichever occurs first. This complete time sequence makes up the vehicle cycle. In some practical circumstances more than one tour per day can be assigned to the same truck. This implies extra line-haul costs, but depending on the cargo characteristics, vehicle size restrictions and other factors, multiple daily tours per vehicle might sometimes be appropriate. For the sake of simplicity, we assume that the vehicles perform just one cycle per day. The model can be easily modified to take into account multiple daily cycles. It is assumed a district of area $A = 40$ sq.km. located 7 km from the depot. The expected driving line-haul time from the depot to the district is $t_{LH1} = 14$ min and the line-haul time from the last service to the depot is admitted to be the same, i.e. $t_{LH2} = t_{LH1} = 14$ min, with a standard deviation $\sigma_{LH2} = \sigma_{LH1} = 2.8$ min, being both normally distributed. The servicing time $t_j^{(ST)}$ at a generic client location, from the instant the vehicle stops until it leaves to attend another client, is assumed to be described by a lognormal distribution, with constant $E[t_j^{(ST)}] = 11$ min, and $\sigma_j^{(ST)} = 4.5$ min. Assuming n servicing points randomly generated over the district, a combination of farthest insertion and *3-OPT* algorithms (Syslo et al. 2006) was applied in order to get the Travelling Salesman Problem (*TSP*) Euclidean route. A corrective coefficient (route factor) was then applied to the Euclidean distance to take into consideration the road network impedance (Novaes and Burin 2009).

The vehicle average speed within the district under standard traffic conditions, is $s_0 = 28$ km/h, with the velocity reduced to $s_1 = 15$ km/h during over-congested situations. The speed, both in a normal traffic condition and during an over-congested situation, is represented by lognormal distributions (section Sequential Analysis for Detecting Incipient Faults). It is assumed that over-congested situations occur with probability $p_1 = 0.20$ (hypothesis H_1), while standard conditions prevail with probability $p_0 = 0.80$ (hypothesis H_0).

Let $H = 8$ h be the maximum vehicle crew working time per day and T_C the vehicle cycle time, with $T_C \leq H$. Assuming a number n of servicing points in the district route, one is interested in estimating the expected number of visits that will be effectively performed during the daily cycle-time, with the objective of keeping it within a pre-established service level. Since n is sufficiently large in our applications, T_C can be assumed to be normally distributed according to the central limit theorem. Thus, it is necessary to estimate an upper extreme value for T_C in order to select an appropriate value for n.

Statistics of extremes have applications in many engineering domains (Gumbel 1967; Smith 2003; Haan and Ferreira 2006). Consider M samples of an i.i.d. continuous random variable X, each sample of size m taken from the same population. The asymptotic configuration, provided it exists, must be such that the largest value of any sample of size m taken from the population must have the same distribution (Gumble 1967). Let $\{X_1, X_2, \ldots, X_m\}$ represent one such sample. Let $Y_m = \max(X_1, \ldots, X_m)$ be the sample maximum. The probability that the largest value is below a generic value x is

$$Pr\{Y_m \leq x\} = F(x)^m, \tag{1}$$

$F(x)$ being the cumulative probability distribution function of x. Since a linear transformation does not change the form of the probability distribution, the probability that the largest value is below x is equal to the probability of a linear function of x (Gumbel 1967)

$$F(x)^m = F(a_m x + b_m), \tag{2}$$

the two parameters a_m and b_m being functions of m. It has been proved that, as $m \to \infty$, expression (2) tends to the cumulative probability distribution $H(x)$ of the Gumbel type in case X is described by exponential, normal, lognormal, or gamma distributions. The Gumbel distribution is (Gumbel 1967; Smith 2003)

$$\lim_{m\to\infty} F(x)^m = H(x) = exp\{-exp[-\alpha(x-u)]\}, \tag{3}$$

where α and u are coefficients obtained via calibration. Let $\hat{E}[X]$ and $\hat{\sigma}[X]$ be, respectively, the expected value and the standard variation of a sample formed by the upper extreme values extracted from M sets containing, each set, m values of a continuous variable X. Then, the estimated values of α and u are (Gumbel 1967)

$$\frac{1}{\hat{\alpha}} = \frac{\sqrt{6}}{\pi}\hat{\sigma}\,[X], \text{ and} \tag{4}$$

$$\hat{\sigma}\;\hat{u} = \frac{\pi}{C_V\sqrt{6}} - \gamma, \tag{5}$$

where $C_V = \widehat{\sigma}[X]/\widehat{E}[X]$ is the coefficient of variation and $\gamma = 0.57722$. From (4) and (5) one gets

$$\widehat{u} = \widehat{E}[X] - \frac{0.57722}{\hat{\alpha}}. \tag{6}$$

We are interested in determining the maximum value of the vehicle cycle time $T_C^{(\max)}$. Letting $x = T_C^{(\max)}$, making $H(T_C) = \vartheta$ in (3), where ϑ is a pre-assumed confidence level, and simplifying

$$T_C^{(\max)} = \widehat{u} - \frac{\ln[-\ln(\vartheta)]}{\hat{\alpha}}. \tag{7}$$

A simulated data set was generated consisting of daily cycle times forming $M = 15$ blocks, each block containing $m = 30$ simulated sample values of T_C, representing a total of 450 samples for each value of n. To perform the simulation it is necessary to assume a value for n beforehand. The objective is to get a maximum value of n such that $T_C \le H$. Five values of n were tested, as shown in Table 1. For a specific value of n, fifteen simulated blocks were produced, yielding 15 values of $T_C^{(\max)}$, one for each block. The average and standard deviation of $T_C^{(\max)}$ were obtained as shown in Table 1. Expressions (4) and (6) yielded the values of $\hat{\alpha}$ and $\hat{u}$. Assuming a $\vartheta = 0.98$ confidence level, expression (7) furnishes the overall maximum $\widehat{T}_C^{(\max)}$. We selected $n = 22$, with a value of $\widehat{T}_C^{(\max)}$ close to the permitted limit of 8 h.

Table 1 Searching for n such that $T_C \leq H$

N	$\hat{E}[T_C]$	$\hat{\sigma}[T_C]$	$\hat{\alpha}$	$\hat{u}$	$\hat{T}_C^{(\max)}$(h)
20	7.12	0.2122	6.0443	7.0285	7.58
21	7.40	0.1987	6.4547	7.3086	7.84
22	7.61	0.2100	6.1074	7.5208	8.04
23	7.96	0.1983	6.4677	7.8687	8.36
24	8.14	0.2117	6.0589	8.0480	8.53

Table 2 shows the static vehicle tour simulation framework. The line-haul travelling times from the depot to the district (outbound) and vice versa (inbound) are represented by t_{LH1} and t_{LH2} respectively in Table 2. The occurrence of hypothesis H_0 or H_1 is represented by h. Variables $d_{i-1,i}$ and $t_{i-1,i}^{(h)}$ are the distance and the travelled time, respectively, from point $i-1$ to point i in the route. The vehicle speed s is represented by log-normal distributions, which depend on the occurrence of H_0 or H_1 (section Sequential Analysis for Detecting Incipient Faults). The stopping time at point i is $t_i^{(ST)}$, and τ_i is the cumulative elapsed time up to stage i.

Table 2 Static route simulation framework

1. Assume a value for n and compute point density $\delta = n/A$;
2. $i \leftarrow 0$;
3. Generate line-haul travelling times, t_{LH1} and t_{LH2}, both normally distributed;
4. Set $\tau_0 \leftarrow t_{LH1} + t_{LH2}$;
5. Generate random number ε; if $\varepsilon \leq p_0$, $h \leftarrow 0$; if $\varepsilon > p_0$, $h \leftarrow 1$ (*);
6. $i \leftarrow i + 1$ (i is the servicing point sequencial number);
7. Generate value for $d_{i-1,i}$, Erlang distributed with parameter $\theta = 3$;
8. Generate value for the route factor k_2 (log-normal);
9. Generate stopping time $t_i^{(ST)}$ at point i, log-normally distributed;
10. If $h = 0$, then $E[s] \leftarrow s_0$, else $E[s] \leftarrow s_1$.Generate speed value s (log-normal)
11. $t_{i-1,i} \leftarrow d_{i-1,i} \times k_2/s$;
12. $\tau_i \leftarrow \tau_{i-1} + t_{i-1,i} + t_i^{(ST)}$
    (a) If $\tau_i \leq H$ and $i < n$, then $n^{(P)} \leftarrow n$ and $T_C = \tau_i$; go to (6);
    (b) If $\tau_i \leq H$ and $i = n$, then
    Begin
    $n^{(P)} \leftarrow n$; $T_C \leftarrow \tau_i$; go to (14);
    End
    (c) If $\tau_i > H$ then
    Begin
    $n^{(P)} \leftarrow [n - (i-1)]$; $T_C \leftarrow \tau_{i-1}$; go to (14);
    End
13. $n^{(U)} \leftarrow [n - n^{(P)}]$
14. Repeat the process from (2) on until the number of replications is complete

(*) $h = 0$, hypothesis H_0 and $h = 1$, hypothesis H_1;

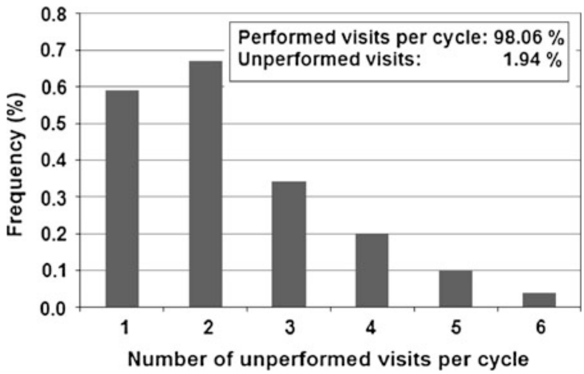


Fig. 1 Static routing: unperformed tasks per vehicle cycle

The vehicle cycle time is T_C; the number of performed tasks in the cycle is $n^{(P)}$, and the number of unperformed tasks is represented by $n^{(U)}$.

Running the simulation model for $n = 22$, with 20,000 replications and assuming over-congested traffic conditions in 20 % of the working days, it led to a 1.94 % rate of unperformed tasks during a typical working cycle. Figure 1 shows the distribution of unperformed tasks per cycle. Most frequent situations are the ones with 1 or 2 unperformed visits per cycle (1.26 % of de cases from a total of 1.94 %). On the other hand, considering only the days when over-congested traffic conditions occur, the average rate of unperformed visits raised to 10.5 %, more than five times the former value, leading to a service level much lower than the desired value.

This result deserves some considerations. First, certain traffic disruptions have severe impacts on vehicle movement as, for example, the ones caused by public transport strikes, heavy rains, etc., and they may last for some days. Those situations, when covering somewhat longer periods, may generate excessive back-logs of tasks, thus impairing the logistics operations for some time. Second, since JIT operations are quite common in global supply chains, unpredictable delays, as the mentioned situations, will lead, in the long term, to additional safety stock compensations in order to maintain the manufacturer's production line uninterrupted. With these aspects in mind, it is apparent that some improving measures are opportune.

The Dynamic Routing Problem

To avoid unexpected backlog in the pick-up process described in section The Static Routing Problem, a simple alternative is available to the logistics operator, although potentially costly: reduce the number of visiting points per vehicle by putting more trucks to perform the service. With this measure, the risk of unperformed tasks obviously will be reduced. But, as a result of the great number of random components that form the vehicle cycle time, its average value will be relatively small, meaning a low fleet usage rate. Another operating alternative is to establish a cooperative scheme, where part of the planned tasks assigned to a truck can be transferred to auxiliary vehicles whenever an excessive service load is foreseen by the on-board computer system.

Such a scheme is part of a new form of managing integrated logistic services in the supply chain, in which a set of intelligent agents, each responsible for one or more activities and interacting with other agents in planning and performing their tasks. In this new form of acting, an agent is an autonomous goal-oriented software process that operates asynchronously, communicating and coordinating with other participant agents as needed (Fox et al. 2000; Davidsson et al. 2005; Berger and Bierwirth 2010).

The smallest controlling entity in this approach (an agent) is described as anything that is able to "perceive its environment through sensors (hardware and software) and act upon that environment through actuators" (Russel and Norvig 2003). A Multi-Agent System (MAS) is a system consisting of independent intelligent control units linked to physical or functional entities such as vehicles, orders, etc. (Mes et al. 2007). Agents act autonomously by pursuing their own objectives and interact with each other using informational exchange and negotiation mechanisms (Mes et al. 2007). In this application the agents are the vehicles which perform the on-route tasks, plus the central depot which has supplementary vehicles that can be eventually assigned to the routes in case other agents do not reach agreement to exchange tasks. The objective is to eliminate or reduce as much as possible the number of unperformed tasks in the pick-up vehicle routing problem described in section The Static Routing Problem.

Fault detection and fault diagnosis methods will be employed in this work to dynamically anticipate operational counteractions in order to avoid unexpected unperformed tasks along the route. A fault-detection software is to be installed aboard, and the vehicle is assumed to be provided with a geo-referencing device and telecommunication equipment linking the vehicle to nearby collaborative agents (other vehicles and the central depot). The rationale involved will permit to infer, during the servicing process, if traffic conditions will impair the accomplishment of the planned tasks during a working day, thus transferring part of the jobs to other vehicles (agents), and leading to a collaborative scheme among them.

Fault Detection and Fault Diagnosis

The concepts and definitions of fault detection and diagnosis set forth in this section are based mostly on Isermann (1997, 2005). Other references are Basseville and Nikiforov (1993), Haan and Ferreira (2006) and Simani et al. (2010). A fault is defined as an unpermitted deviation of at least one characteristic property of a variable from an acceptable behaviour. Therefore, the fault is a state that may lead to a malfunction or failure of the system. The time dependency of faults can be classified as (a) *abrupt fault* (stepwise), (b) *incipient fault* (drift like), and (c) *intermittent* *fault*. One calls *abrupt fault* any change in the parameters of a system that occurs either instantaneously or at least very fast with respect to the sampling period of the measurements. Abrupt faults do not refer to changes in the process with large magnitude; in fact, in most applications the main question is how to detect small changes. On the other hand, incipient faults are also of interest: they are behaviour deviations that occur cumulatively over time, yet not failures, but leading to future underperformance or disruptions in the system. In our application we will deal with detection of incipient and abrupt faults.

To control technical systems, supervisory functions are installed to indicate undesired or unpermitted process states, as well as to take appropriate actions in order to maintain the operation within a pre-established service level and to avoid unexpected disruptions. Three basic function types can be distinguished (Isermann 1997): (a) *monitoring*, in which measurable variables are checked with regard to tolerances and warnings are generated for the operator; (b) *automatic protection*, normally used to control dangerous processes that cannot wait for external interventions; and (c) *supervision with fault diagnosis* which is based on the measurement of some key variables and the calculation of parameters, resulting in the identification of symptoms via change detection, followed by fault diagnosis, and leading to counteraction decisions. Type (c) function is compatible with in-depth fault diagnosis, either with *abrupt* or *incipient* time behaviour, being more comprehensive. In our application supervisory functions of type (a) and (c) will be employed.

The goal for early fault detection and diagnosis is to have enough time for counter-actions such as adding supplementary operations, reconfiguration, maintenance or repair, etc. The earlier detection can be achieved by gathering more information, especially by using the relationship among the measurable quantities in the form of mathematical models. For fault diagnosis, the knowledge of cause-effect relationships has to be used. In our analysis, two types of faults will be considered: (a) *incipient faults*, occasioned by over-congested traffic conditions, in which the measurement and analysis of commanding parameters occur cumulatively over time, and (b) *abrupt faults*, represented by unpredictable delays that may occur at *OEM* premises when transferring goods to the logistics operator, as well as vehicle breakdowns during the cycle, or another sort of exceptional random interruptions.

Incipient Faults Occasioned by Exceptional Traffic Congestion

Traffic congestion is seen as a condition of traffic delay (i.e., when vehicle flow is slowed below reasonable speeds) because the number of vehicles trying to use a road exceeds the capacity of the network to handle it (Weisbrod et al. 2003). In addition to speed reduction, congestion also introduces variability in traffic conditions, which is known as *travel time reliability* (Cambridge Systematics 2005). The resulting traffic slowdowns and travel time reliability produce negative effects on supply chain activities, including impacts on vehicle traveling costs, air quality and noise, labour efficiency, industrial and commercial productivity, customer service level, etc. The severity and pattern of congestion, as well as the effectiveness of alternative policies and interventions to address it, vary widely from place to place. That can depend on the size and layout of the urban area, its available transportation options, and the nature of traffic generators (Weisbrod et al. 2003). Congestion is usually the result of seven root causes, often interacting with one another (Cambridge Systematics 2005):

1. Physical bottlenecks, such as reduced number of lanes, narrow lane and shoulder widths, inadequate roadway grades and curves, etc., leading to reduced road capacity.
2. Traffic incidents, occasioned by events that disrupt the normal flow of traffic, such as vehicular crashes, breakdowns, etc.
3. Work zones temporally reserved for construction and repair activities on the roadway, generating lane reduction, narrower traffic spaces, lane shifts, detours, speed reduction, etc.
4. Weather conditions such as snow, flood, fog, etc., that can lead to substantial changes in driver behaviour,
5. Traffic control devices, which leads to intermittent disruption of traffic flow, such as railroad grade crossings, poorly timed light signals, traffic interferences with street cars and bus ways, etc.
6. Special events that cause severe traffic flow variations in its vicinity, being radically different from typical day-to-day patterns. For instance, a rapid transit labor strike in a city with heavy public transport patronage and generating additional flow of cars and busses.
7. Fluctuations in normal traffic which shows, on a day-to-day basis, fluctuations with days with higher traffic volumes than others. In cities with constant heavy traffic volumes, even random fluctuations can result in unreliable and over-congested traffic conditions.

Another important traffic congestion classification is due to Brownfield et al. (2003). The first type is *recurrent congestion*, which can be anticipated by road users that are acquainted with the route. The other type is *non-recurrent congestion*, which occurs at non-regular times at a site. It is unexpected and unpredictable to the driver. In our analysis, it is assumed that the logistics entity in charge of the urban transport service is aware of all programmed events, i.e. it is fully prepared to cope with recurrent congestion. Thus, from the seven factors listed above, causes (1), (3) and (5) are not considered in our application. Conversely, it is assumed that over-congested situations are originated by causes (2), (4), (6) or (7).

Although there is no existing, universally accepted, quantitative definition for traffic congestion, its analysis must rely on easy to measure elements if its impacts are to be evaluated and compared across the range of situations considered in the investigation (Brownfield et al. 2003). One frequent assumption is to assume that an urban road link is congested if its average speed is below a given upper threshold. In addition to average speed reduction to travellers, the sources of congestion also produce time variability known as *travel time reliability* (Cambridge Systematics 2005), which can be defined in terms of how travel times vary within a pre-defined period. In practical terms, it is useful to fit statistical frequency distributions to travel time, to see how much variability exists in critical sites of the road network.

Exceptionally, unpredictable and heavy traffic congestions caused by severe accidents, public transport strikes, heavy storms, etc., may occur during certain working days. In these situations the travelling speed decreases sharply. Let s_0 be the average travelling speed in a route in a generic working day, and suppose the average speed reduces to a level $s_1 \ll s_0$ when over-congested situations occur. Then, one could say that the traffic conditions are normal if $s \geq s_0$, and over-congested if $s \leq s_1$. Moreover, if $s_1 < s < s_0$, one would not decide immediately for either alternative, waiting for more information to take a decision. Of course, this is a typical statistical hypothesis testing. Nevertheless, an instantaneous travelling time increase is not, in itself, an indication of an over congested situation. In fact, many non-recurrent events have short duration, and their effects dissipate more or less rapidly. Furthermore, some recurrent events have local impact only, and their effects do not extend to other parts of the served region. Over-congested situations that are of interest in our analysis are the ones with broader geographical extension and longer duration, although in many cases they are no longer than 24 h. Thus, travel time reliability covering an expressive subset of the urban region, seems to be a good judgmental criterion to evaluate it. And, in order to measure travel time reliability it is necessary to sequentially collect and analyse traffic data. With today's on-board telematics and computing devices it is not difficult to collect and analyse real-time information on travelled distance, time and speed with satisfactory accuracy (Goel 2008). In our study, the statistical inference process to detect an over-congested condition follows a sequential analysis methodology (Wald 1947; Basseville and Nikiforov 1993; Lai 2001), which is described in the next section.

Dynamic Detection of Over-Congested Traffic Conditions

Day-to-day traffic flow variability in urban networks produces typical traffic patterns, but unexpected events cause occasional surges in traffic volumes that overwhelm the road system. Such events, of a "hectic" pattern, are generated by accidents with severe traffic interruptions, extensive public transport strikes, and long duration storms, among others. Strong changes in some characteristic properties of a system may occur occasionally in both technological and natural environments. And due to today's availability of information processing systems, complex monitoring algorithms have been developed and implemented (Basseville and Nikiforov 1993). The key difficulty in detecting a fault occurrence through the observation of some properties of a system is to separate noise from the relevant factors. In addition, some failures have a catastrophic nature, leading to an abrupt change in the control variables. But some faults occur with gradual changes in the system attributes over time. One way of tackling the latter is Sequential Analysis (Wald 1947; Basseville and Nikiforov 1993; Lai 2001).

Classical techniques of statistical inference and hypotheses testing adopt a fixed sample size. With this kind of approach one seeks to minimize the error probabilities for a given sample size. The size of the sample is defined beforehand, and following its statistical analysis one of two possible actions is taken: accept the null hypothesis H_0, or accept the alternative hypothesis H_1. The null hypothesis represents in our analysis the standard or basic situation, whereas the alternative hypothesis indicates the occurrence of an abnormal condition, leading to a fault in the system. Another way to solve hypotheses testing problems, when the sample size is not fixed a priori but depends upon the data that have already been observed, is Sequential Analysis. Now the problem is: for given error probabilities, try to minimize the sample size, or equivalently, make the decision with as few observations as possible. Contrary to the fixed sample size approach, a third possible course of action may occur in sequential analysis when the evidence is ambiguous: take more observations until the evidence strongly favours one of the two hypotheses. Thus, sequential analysis follows a dynamic sequence of observations in such a way that the decision to terminate or not the experiment depends, at each stage, on the previous test results.

Although some authors date the rudiments of sequential analysis to the works of Huyghens, Bernoulli, and Laplace, such methodology was effectively born in response to demands for more efficient testing of anti-aircraft gunnery during World War II, culminating with the development of the *Sequential Probability Ratio Test* (*SPRT*) by Wald, in 1943 (Lai 2001). A typical case of sequential estimation arises when only two unknown parameters μ and σ are required to define the distribution of the random variable x object of our analysis. Let $f(x, \mu, \sigma)$ denote the probability density function of x, when x is continuous. Conversely, if x is discrete, $f(x, \mu, \sigma)$ represents its probability. Let $x_1, x_2, \ldots, x_m$ be a set of m sequential and independent observations on x. Due to the independence of the observations, the joint probability density function is

$$f(x_1, \mu, \sigma)f(x_2, \mu, \sigma) \ldots f(x_m, \mu, \sigma). \quad (8)$$

Suppose that the distribution of the random variable x under consideration is defined by q unknown parameters (in our case, $q = 2$). A statement about the values of the q parameters is called a *simple hypothesis* if it determines uniquely the values of all q parameters. It is called a *composite hypothesis* if it is consistent with more than one value for some parameter (Wald 1947). Let us analyse the test of simple hypothesis that $\mu = \mu_0$ and $\sigma = \sigma_0$, where μ and σ are the expected value and the standard deviation of the probability distribution of x. This hypothesis is the null hypothesis denoted by H_0. The alternative hypothesis that $\mu = \mu_1$ and $\sigma = \sigma_1$ will be denoted by H_1. Thus, we shall deal with the problem of testing the simple hypothesis H_0 against the alternative simple hypothesis H_1, on the basis of a sample of m independent observations $x_1, x_2, \ldots, x_m$ on x. According to the developments of Neyman and Pearson, errors of two kinds are present when one accepts or rejects hypothesis H_0. We commit an error of first kind if we reject H_0 when it is true. On the other hand, we commit an error of the second kind if we accept H_0 when $H_1$is true. We denote the probability of an error of the first kind by α, and the probability of an error of second kind by β.

To apply the *SPRT* developed by Wald (1947) for testing $H_0 : \mu = \mu_0, \sigma = \sigma_0$ against $H_1 : \mu = \mu_1, \sigma = \sigma_1$, two positive constants A and B ($B < A$) are computed

$$A = (1 - \beta)/\alpha \text{ and } B = \beta/(1 - \alpha) \tag{9}$$

Suppose one has drawn m samples leading to the independent observations $x_1, x_2, \ldots, x_m$ on the random variable x. At this stage of the experiment the *SPRT* (Wald 1947; Basseville and Nikiforov 1993; Lai 2001) is computed as

$$\pi_m = \frac{f(x_1, \mu_1, \sigma_1) f(x_2, \mu_1, \sigma_1) \ldots f(x_m, \mu_1, \sigma_1)}{f(x_1, \mu_0, \sigma_0) f(x_2, \mu_0, \sigma_0) \ldots f(x_m, \mu_0, \sigma_0)}. \tag{10}$$

Three situations may occur:

1. If $B < \pi_m < A$, the experiment continues by taking an additional observation;
2. If $\pi_m \geq A$, the experiment terminates with the rejection of H_0;
3. If $\pi_m \leq B$, the experiment terminates with the acceptance of H_0.

For purposes of mathematical simplification, it is more convenient to compute the logarithm of the ratio π_m. Let

$$z_i = ln\left(\frac{f(x_i, \mu_1, \sigma_1)}{f(x_i, \mu_0, \sigma_0)}\right). \tag{11}$$

Define

$$\pi_m^* = \ln(\pi_m) = z_1 + z_2 + \cdots + z_m. \tag{12}$$

The test is addictive now. The experiment continues if $\ln B < \pi_m^* < lnA$ by taking an additional observation; the process terminates with the rejection of H_0 *if* $\pi_m * \geq lnA$; and it terminates with the acceptance of H_0 if $\pi_m * \leq lnB$.

In practical cases, composite hypothesis may occur. One way to solve sequential analysis problems with composite hypothesis is the method of a weighting function associated with the generalized likelihood ratio algorithm (Basseville and Nikiforov 1993). To do this, two weighting probability distributions, with density functions $g(H_0)$ and $g(H_1)$, depending on H_0 and H_1 respectively, are introduced into the model. The *SPRT* is now transformed into a weighted likelihood ratio test (Basseville and Nikiforov 1993). But, in order to do this, it is necessary to fit distributions $g(H_0)$ and $g(H_1)$ to the data, which depends on detailed information not commonly available in real settings. In the application considered in this paper, a more tractable composite hypothesis test is adopted. This composite hypothesis testing is represented by $H_0^{'} : \mu \leq \mu_0, \sigma \leq \sigma_0$ versus $H_1^{'} : \mu \geq \mu_1, \sigma \geq \sigma_1$, such that $\mu_1 > \mu_0$ and $\sigma_1 > \sigma_0$. This model is usually sufficient for practical purposes (Lai 2001). Assuming that the probabilities of the errors of first and second kind also do not exceed α and β, one can use the *SPRT* of the simple hypothesis $H_0 : \mu = \mu_0, \sigma = \sigma_0$ versus $H_1 : \mu = \mu_1, \sigma = \sigma_1$, with the same error probabilities α and β. However, while this *SPRT* has minimum expected sample size at $\mu = \mu_0$, $\sigma = \sigma_0$ and at $\mu = \mu_1, \sigma = \sigma_1$, its maximum expected sample size over μ and σ can be larger than the optimal fixed sample size (Lai 2001). This means that sometimes the sequential test will not be sufficient to detect hypothesis H_1 during the daily tour, generating unperformed tasks at the end of the working day. But unperformed tasks will be eliminated or drastically reduced when compared with the static alternative, as it will be shown in section Sequential Analysis for Detecting Incipient Faults, a fact that justifies the adoption of the dynamic setting in our model.

Sequential Analysis for Detecting Incipient Faults

In this application, the variable that commands the decision whether to seek help from another agent or to proceed along the planned routing process is the vehicle speed s. In fact, since link lengths vary along the route, and consequently the resulting displacement times also vary, speed is a more appropriate variable to measure traffic variations. The renewal epoch (stochastic regenerating point) of the sequential decision process is defined as the instant when the vehicle crew has just terminated a task at an OEM location and is ready to depart for the next visit. At such an instant, the on-board computer evaluates the displacement time $t_{i-1,i}$ over the traveled segment linking the last visiting stop $i - 1$ to the present one i. The corresponding speed s is simply obtained by dividing the travelled segment extension by its respective displacement time, both elements assumed to be available on the on-board system. For the district under analysis it is assumed that there are enough historical data on speed values covering the standard traffic condition and the over-congested scenario. In particular, the average speed $s \geq s_0$ is related to standard traffic conditions and $s \leq s_1$ represents the over-congested scenario. This information, together with the series of data collected up to that point, will serve as the basis for inferring whether the traffic is normally behaved or is over-congested, thus leading to the appropriate operational decision.

As discussed in section Dynamic Detection of Over-Congested Traffic Conditions, it is necessary to define a probability distribution $f(x,\ \mu,\ \sigma)$ to represent the random variable that commands the decision process. A sample was gathered in a representative route located in the urban area under analysis, involving 40 vehicle travel speeds during typical working days. A log-normal distribution was fitted to the data (Fig. 1):

$$f(s) = \frac{1}{\sigma\sqrt{2\pi}} exp\left\{ -\frac{1}{2}\left[\frac{\ln(s) - \mu}{\sigma}\right]^2 \right\},\ s > 0, \tag{13}$$

where s is the speed in km/h, and μ and σ are the parameters of the log-normal distribution given by

$$\mu = ln\left\{\frac{E[s]^2}{\sqrt{var[s] + E[s]^2}}\right\} \text{ and } \sigma = \sqrt{ln\left\{\frac{var[s]}{E[s]^2} + 1\right\}}, \tag{14}$$

where $E[s]$ and $var[s]$ are the expected value and the variance of s respectively. For the mentioned sample of travelling times, assumed to represent the normal traffic conditions in our application, one has $E[s] = 28$ km/h and $var[s] = 44.84$, leading to $\mu_0 = 3.3044$ and $\sigma_0 = 0.2354$.

For the over-congested traffic conditions one has $E[s_1] = 15$ km/h. It was assumed that, for this situation, the speed is also represented by a log-normal distribution. It was assumed further that the coefficient of variation is the same for hypotheses H_1 and H_0, i.e. $C_V = \sqrt{44.84}/28.0 = 0.239$. Depending on real data, this assumption may be changed, fitting a value of $var[s_1]$ directly over the real data. Thus, for the over-congested traffic condition in our application one has $var[s_1] = C_V^2 \times E[s_1]^2 = 12.852$, leading to $\mu_1 = 2.680$ and $\sigma_1 = \sigma_0 = 0.2354$. Substituting (13) into (11) and (12), and making the necessary simplifications, one gets the *SPRT* parameter,

$$\pi_m^* = m\, ln\left(\frac{\sigma_0}{\sigma_1}\right) + \frac{1}{2}\sum_{i=1}^{m}\left[\frac{\ln\left(s^{(i)}\right) - \mu_0}{\sigma_0}\right]^2 - \frac{1}{2}\sum_{i=1}^{m}\left[\frac{\ln\left(s^{(i)}\right) - \mu_1}{\sigma_1}\right]^2, \tag{15}$$

where m is the sequential number of the test and $s^{(i)}$ is the vehicle speed measured at stage i along the route within the district (Fig. 2).

At each regeneration point (stage) the *SPRT* value π_m^* (15) is computed. Depending on the *SPRT* value, three scenarios are defined:

(a) scenario $sc = 0$, when $\pi_m^* \leq B$;
(b) scenario $sc = 1$, when $\pi_m^* \geq A$;
(c) scenario $sc = 2$,when $B < \pi_m^* < A$.

Countermeasures are taken if scenario $sc = 1$ occurs; otherwise, the routing process continues unchanged until the next stage. Figure 3 shows a schematic representation of the vehicle routing sequence and the decision stage where the *SPRT* is performed. Assume that the vehicle agent AG_A left the depot with the assignment of n visits. Suppose the sequential test indicates the occurrence of hypothesis H_1 at stage 3, as shown in Fig. 3. At that point, the on-board computer checks how many of the remaining visits should be transferred to another vehicle agent. Let k be the number of visits to be transferred. Upon negotiation, agent AG_B agrees to perform the k tasks. Of course, depending on the number of visits to be transferred, more than one agent can be involved in the transference.

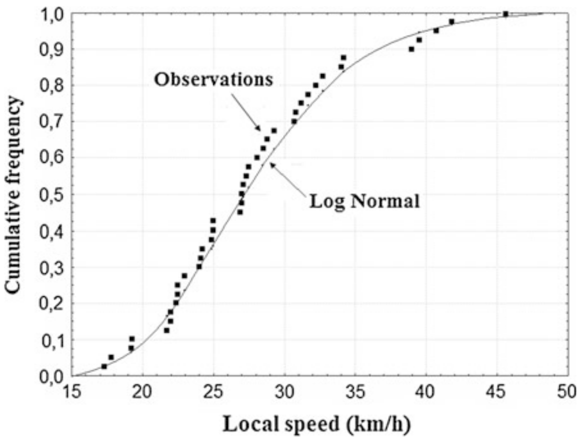


Fig. 2 Fitting a log-normal distribution to the local speed

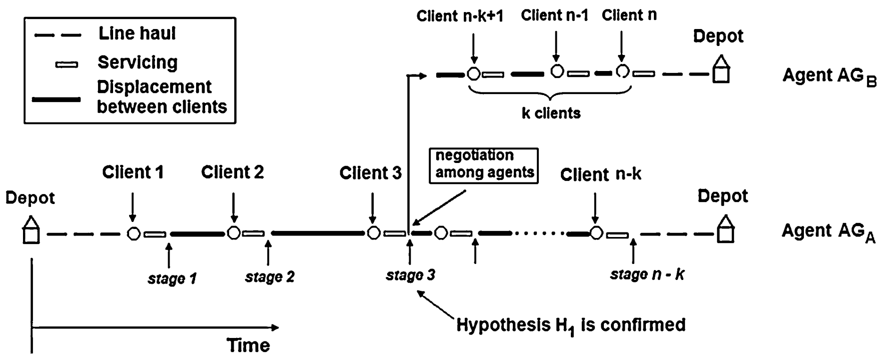


Fig. 3 The vehicle routing sequence and the decision stage

Transference of Tasks

Let i be the actual stage of the routing process. Recall that stage i corresponds to the instant when the ith pick-up service has just terminated, and the vehicle is ready to depart to the next stop. Let us analyse first the occurrence *of incipient faults*. Define $\tau_i^{(h)}$ as the cumulative vehicle time along the route, measured from the departure from the depot, up to stage i, given hypothesis h (either $h = 0$, or $h = 1$) occurs. It is given by

$$\tau_i^{(h)} = t_{LH1} + \sum\nolimits_{j=2}^{i} t_{j-1,j}^{(h)} + \sum\nolimits_{j=1}^{i} t_j^{(ST)}. \qquad (16)$$

On the other hand, let $\omega_{i,j}^{(h)}$ be the elapsed time from the actual stage i to stage j ($j > i$), assuming hypothesis h occurs, and also assuming that the vehicle returns to the depot just after visit j

$$\omega_{i,j}^{(h)} = \sum\nolimits_{m=i+1}^{j} t_{m-1,m}^{(h)} + \sum\nolimits_{m=i}^{j} t_{m}^{(ST)} + t_{LH2} \tag{17}$$

We have assumed in the application that all random variables are independent. Due to the central limit theorem and for i sufficiently large, variable $\omega_{i,j}^{(h)}$ can be approximately represented by a normal distribution. Thus, for a 98 % significance level, the maximum expected value of $\omega_{i,j}^{(h)}$ is

$$\overline{\omega}_{i,j}^{(h)} = \max \omega_{i,j}^{(h)} \cong E[\omega_{i,j}^{(h)}] + 2.06\sqrt{var[\omega_{i,j}^{(h)}}\tag{18}$$

At the beginning of the routing process is not yet known which traffic condition is prevailing, thus hypothesis H_0 (which shows higher probability) is assumed, i.e. $h = 0$. Further, at every stage i the SPRT is performed. Suppose scenario $h = 0$ occurs. This indicates that the routing process should proceed unchanged, with hypothesis H_0 prevailing. On the other hand, suppose scenario $h = 1$ occurs, meaning one should accept hypothesis H_1. Then, in order to define the possible number of visits to be performed in the route, one has to seek for the largest value of j such that the expected total time to accomplish the tasks is not greater than H

$$\tau_i^{(h)} + \bar{\omega}_{i,j}^{(h)} \leq H, \text{ with } h = 1 \text{ and } j > i. \tag{19}$$

Thus, the total number of visits to be performed in the tour is $n^{'} = i + j$, and the number of visits to be transferred to other agents is $n^T = n - (i + j)$, which represents the expected number of *incipient faults* in the application. Of course, since n^T tasks are transferred, there will be less tasks remaining to be considered at the next stages of the process, i.e. $n \leftarrow n'$, with $n^{'} < n$.

After handling incipient faults, the model investigates the occurrence of *abrupt faults*. Here, an abrupt fault refers to the occurrence of unperformed tasks at the end of a daily vehicle cycle occasioned by exceptional unpredictable delays at *OEM* premises when picking-up manufactured orders, lorry breakdowns when travelling along the route, etc. At each decision stage the on-board system estimates the maximum cycle time to perform all visits, considering the effective elapsed time so far, plus the eventual observed delays and the remaining visits to be done. If the daily cycle time limit is surpassed, the on-board system estimates how many visits are to be transferred to other agents.

Suppose an exceptional and unpredictable manufacturing delay Δ_i occurs at stage i. Care must be taken not to consider as exceptional delays situations already contemplated in historical data variability. An exceptional delay ∇_m may also occur at a transport link m. Relation (19) is now modified as follows

$$\tau_i^{(h)} + \overline{\omega}_{i,j}^{(h)} + \Delta_i + \nabla_m \leq H, \text{ with } h = 0 \text{ or } 1. \quad (20)$$

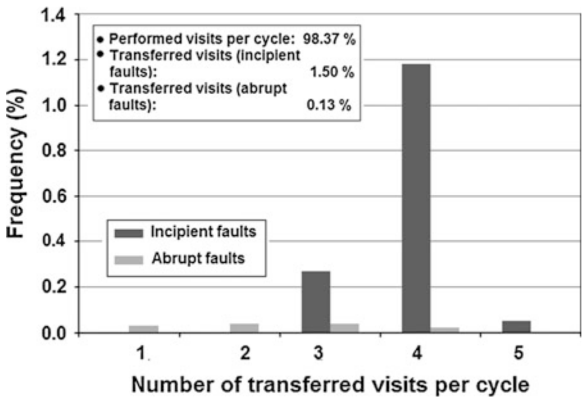
Again, one looks for the largest value of j such that the cycle time constraint is respected, and estimating the number of remaining visits that will be performed and the tasks that have to be transferred to other agents.

Simulation Results

The simulation of the dynamic model for $n = 22$, with 20,000 replications and assuming over-congested traffic conditions in 20 % of the working days, resulted in a 1.63 % rate of transferred tasks during a typical working cycle. From that, 1.50 % of the transferences were generated by incipient faults, and 0.13 % by abrupt faults. The transference total of 1.63 % is less than the 1.94 % level of unperformed tasks observed in the static case. This happens because, as the sequential test is performed, followed by abrupt fault detection, the number of remaining tasks to be done decreases, thus reducing the possibility of occurring other additional faults. Figure 4 shows that the occurrence of four transferred tasks per cycle is the most frequent situation in the dynamic case. In fact, since it takes some time until the *SPRT* can detect hypothesis H_1, the number of prospective tasks to be transferred tends to increase, and are likely to happen all at the same moment. Abrupt tasks also occur, but at a significantly reduced frequency.

On the other hand, considering only the days when hypothesis H_1 occurred, the rate of transferred tasks was 7.93 %, from which 7.50 % was generated by incipient faults and 0.43 % by abrupt faults. In those two situations it was admitted $\Delta_i = 0$ and $\nabla_m = 0$ in (20), meaning that abrupt faults were not generated by exceptional delays, but were occasioned by intrinsic variations in the random variables that form the cycle time. Of course, if those elements were not nil, the results would reflect their presence.

Fig. 4 Dynamic routing: transferred tasks per vehicle cycle



Conclusions

Although potentially suitable for applications, a number of points are still open for further research. First, it remains to be investigated the criteria to decide which tasks should be transferred to other agents, considering the vehicle routing configuration and the corresponding schemes of other prospective agents. Another important point is that, depending on the *OEM* plant locations and the time of the day, it might be impossible to transfer tasks, requiring other anticipating measures from the central depot. A third question refers to the routing optimization process based on the *TSP* criterion, which is the prevalent case in most dynamic routing problems reported in the literature, where one searches for the route that minimizes travelled distance or time. If the components or products of the diverse *OEM* manufacturing plants show different added values, the ones with the highest values should not be located at the end of the picking up process due to the higher inventory costs occasioned by unperformed tasks. Further research is planned with the objective of developing new heuristics to solve this kind of vehicle routing problem incorporating product value considerations in the optimization criteria.

Acknowledgments This research has been supported by the Brazilian Capes Foundation and by DFG — German Research Foundation, Bragecrim Project n° 2009-2.

References

Basseville M, Nikiforov I (1993) Detection of abrupt changes: theory and application. Prentice-Hall, New Jersey

Berger S, Bierwirth C (2010) Solutions to the request reassignment problem in collaborative carrier networks. Transp Res Part E 46:627–638

Brownfield J, Graham A, Eveleigh H, Maunsell F, Ward H, Robertson S, Allsop R (2003) Congestion and accident risk. Road safety research report No 44, Department of Transport, UK

Cambridge systematics (2005) Traffic congestion and reliability. Federal highway administration, Washington, DC

Davidsson P, Henesey L, Ramstedt L, Törnquist J, Wernstedt F (2005) An analysis of agent-approaches to transport logistics. Transp Res Part C 13:255–271

Figliozzi MA (2007) Analysis of the efficiency of urban commercial vehicle tours: data collection, methodology, and policy implications. Transp Res Part B 41:1014–1032

Figliozzi MA (2010) The impacts of congestion on commercial vehicle tour characteristics and costs. Transp Res Part E 46:496–506

Fleischmann B, Gnutzmann S, Sandvo*β* E (2004) Dynamic vehicle routing based on on-line traffic information. Transp Sci 38(4):420–433

Fox MS, Barbuceanu M, Teigenm R (2000) Agent-oriented supply chain management. Int J Flex Manuf Syst 12:165–188

Frazzon EM, Makuschewits T, Scholz-Reiter B, Novaes AG (2010) Assessing the integrated scheduling of manufacturing and transportation systems along global supply chains. In: 12th World conference on transport research, paper no 1238, Lisbon, July 11–15

Golden B, Raghavan S, Wasil E (eds) (2008) The vehicle routing problem: latest advances and new challenges. Springer, New York

Goel A (2008) Fleet telematics. Springer, New York
Gumbel EJ (1967) Statistics of extremes. Columbia University Press, New York
Haan L, Ferreira A (2006) Extreme value theory: an introduction. Springer, New York
Isermann R (1997) Supervision, fault-detection and fault-diagnosis methods—an introduction. Control Eng Practice 5(5):639–652
Isermann R (2005) Model-based fault detection and diagnosis—status and applications. Annu Rev Control 29:71–85
Lai TL (2001) Sequential analysis: some classical problems and new challenges. Statistica Sinica 11:303–408
Larsen A (2001) The dynamic routing problem. PhD Dissertation, Technical University of Denmark, Lyngby
Larsen A, Madsen OBG, Solomon MM (2008) Recent developments in dynamic vehicle routing. In: Golden B, Raghavan S, Wasil E (eds) The vehicle routing problem: latest advances and new challenges. Springer, New York, pp 199–218
Mes M, van der Heijden M, van Harten A (2007) Comparison of agent-based scheduling to look-ahead heuristics for real-time transportation problems. Eur J Oper Res 181:59–75
Novaes AG, Burin PJ (2009) A dynamic vehicle routing problem (in Portuguese). In: Proceedings XXIII ANPET—Brazilian association of transport research and teaching, 9–13 Nov, Vitória
Novaes AG, Frazzon EM, Burin PJ (2011) Dynamic vehicle routing in over congested urban areas. In: Kreowski H-J, Scholz-Reiter B, Thoben K-D (eds) Dynamics in logistics. Springer, Berlin Heidelberg, pp 49–58
Psaraftis HN (1995) Dynamic vehicle routing: status and prospects. Ann Oper Res 61:143–164
Ribeiro GM, Lorena L (2005) Dynamic vehicle routing using genetic algorithms (in Portuguese). In: Proceedings of XVI conference on transport research and education —ANPET, Recife
Russel S, Norvig P (2003) Artificial intelligence: a modern approach. Prentice Hall, Englewood Cliffs
Simani S, Fantuzzi C, Patton RJ (2010) Model-based fault diagnosis in dynamic systems using identification techniques. Springer, London
Smith RL (2003) Statistics of extremes, with applications in environment, insurance and finance. Department of Statistics, University of North Carolina, Chapel Hill. (http:www.stat.unc.edu/postscript/rs/semstatrls.ps)
Syslo MM, Deo N, Kowalik JS. (2006) Discrete optimization algorithms with pascal programs. Dover Books on Mathematics, Mineola
Wald A (1947) Sequential analysis. Dover Publications, New York
Weisbrod G, Vary D, Treyz G (2003) Measuring the economic costs of urban traffic congestion to business. Transportation research record no 1839, Transportation Research Board, Washington, DC

Part II
Transport Logistics and Dynamic Routing

Knowledge Sharing in Intermodal Transport: A Multi-Agent Based Perspective

Jiani Wu and Hans-Dietrich Haasis

Abstract Due to the problems of road congestion, environmental concerns and traffic safety, intermodal transport is a significant factor in the movement of freight. Furthermore, knowledge sharing is crucial for efficient coordination in intermodal transport in reducing transit time and cost. However, the discussion of knowledge sharing tools in intermodal transport is few in academia. Therefore, the research for appropriate knowledge sharing tools becomes imperative. In view of the distributed nature of involved actors and knowledge resources in intermodal transport, this paper proposes a multi-agent based perspective for this filed. It designs a multi-agent based knowledge sharing platform (MA-KSP), which identifies four types of agents according to their functions on three layers. In addition, on the basis of activities of intermodal transport, it specifies the knowledge content which is potentially involved in practical applications. Finally, one of the scenarios in intermodal transport illustrates how the MA-KSP works in the real world.

Keywords Knowledge sharing · Intermodal transport · Multi-agent system

J. Wu (✉) · H.-D. Haasis
University of Bremen, International Graduate School for Dynamics in Logistics,
Institute of Shipping Economics and Logistics, 28359 Bremen, Germany
e-mail: wu@isl.org

H.-D. Haasis
e-mail: haasis@isl.org

H.-J. Kreowski et al. (eds.), *Dynamics in Logistics*, Lecture Notes in Logistics,
DOI: 10.1007/978-3-642-35966-8_3,

Introduction

Intermodal transport is typically characterized by a multiplicity of actors. It covers a wide range of transport activities supported by operators of local road haulage, rail, sea or inland waterway, and other independent organizations. Integrating multiple transport modes requires a process or system approach for execution and “a higher degree of skills and broader knowledge of the transportation/supply chain processes equipments, and infrastructures” (Muller 1999). Due to this distributed nature of actors and information resources involved in multi-organizations, knowledge sharing among actors becomes highly necessary. As it moves from a focus on infrastructure components to a holistic focus on processes or systems, knowledge is the key to sustain competitive processes and systems in today’s global economy. The success of intermodal transport depends largely on the knowledge sharing between multiple players. Knowledge sharing can generate their knowledge, skills, core competences and resources, so as to enhance competitive capabilities and respond better to business opportunities (Liu et al. 2011).

The motivation of establishing a communication platform for intermodal transport stems from the lack of commonly accepted tools to share quantifiable and live information. Examples of the information are position of freight and vehicles, expected arrival and waiting times as well as alternative routes and transport modes (Bernaer et al. 2006). The increasing use of multi-agent system (MAS) provides a significant opportunity for the transport industry. For example, a MAS is capable to make route planning and vehicle allocation in real time considering the challenges in a city logistics. Changing customer order data and vehicle conditions can be incorporated dynamically (Haasis et al. 2009). It also has the potential to improve the quality of intermodal transport. The MAS allows qualified information to be more smoothly shared among the actors, which could bring about collaborative efficiency. The collaborative efficiency supports the fulfillment in terms of production, infrastructure, technology, quality and policies making of intermodal transport.

This paper designs a knowledge sharing platform for intermodal transport from a multi-agent based perspective. The remainder of this paper is organized as follows. Section Background analyses the strategic importance of knowledge sharing for improving the intermodal transport performances; as well as proposes hat a multi-agent based solution is suitable for this field. In Sect. Multi-Agent Based Knowledge Sharing Platform, after introducing the structure of a basic agent of MA-KSP, it provides a checklist of involved knowledge content in practical application, and the architecture of MA-KSP is depicted. It details the constituent agents and their interactive behaviors. Section Example offers an example taking a scenario in real-world of intermodal transport to show how the MA-KSP works.

Background

Intermodal transport means the movement of goods in one and the same loading unit or road vehicle, which uses successively two or more modes of transport (rail, sea, air, or inland waterway), without moving the goods themselves in changing modes (ECMT 1993). It offers manufacturers a full range of transportation modes and routing options, allowing them to coordinate supply, production, storage, finance, and distribution functions to achieve efficient relationships (Rondinelli and Berry 2000). The operation of intermodal transport is supported by independent actors. These may be: the demand side (shippers, forwarders, ocean shipping lines, logistics service providers), the supply side (infrastructure providers, transport operators, intermodal transport operator), the policy side (authorities, regulators, associations) and the surrounding side (quarantine and inspection service, customs, insurance and bank) (Vrenken et al. 2005).

Figure 1 is a diagram briefly tracing the evolutions of intermodal transport. It shows some typical characteristics at different time stage (past-current-future).

Three representative aspects are indicated:

- Solutions such as collaboration solution, stakeholder relationship, communication mode
  - Past: single-way processing mode, short of flexibility, easily fail
  - Current: networking for multi-channel applications, risk reduction
  - Future: multi-level networking, adequately utilizing resource, a better inter modality performance and efficiency of social cost
- Visibility relates to the transparency level of information sharing, process tracing and accessing
  - Past: unknowable process, information asymmetry, high-risk
  - Current: more communication technology, improved information acquisition
  - Future: transparent information exchanging, reduction of the administration and planning errors

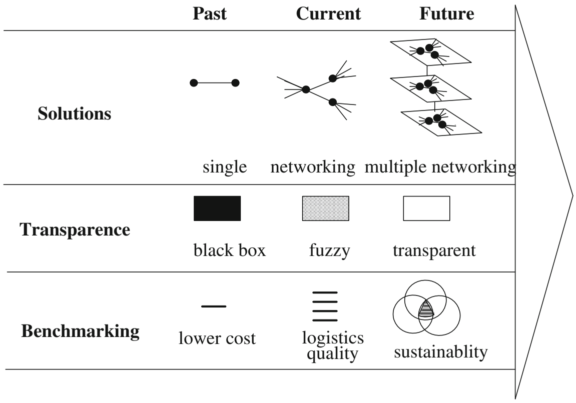


Fig. 1 Evolution of intermodal transport

- Benchmarking shows the assessment criteria for the intermodal transport performance
  - Past: orienting lower cost, one-sided view, much hidden risks, undesirable consequences
  - Current: combined considerations about logistics qualities (time, cost, reliability, flexibility and legitimization, etc.)
  - Future: sustainable development—balance of economic benefit, ecological protection and social welfare.

Nowadays, intermodal transport tends to be collaborative/cooperative, information/knowledge-based, and sustainability-oriented. One of the primary challenges in intermodal transport management is to coordinate both several inter-dependent activities and the communication between the multiple actors (Davidsson et al. 2005). Therefore, the tendency leads to the concept of "Knowledge Sharing". Knowledge sharing is the act of not only contributing one's own knowledge but also seeking and receiving others' knowledge within the system. It allows for retrieving the existing organizational knowledge, exchanging the existing personal knowledge and generating new knowledge. The knowledge can be represented with the usage of specific data structures in established computerized knowledge base. Consequently this reduces errors, enhances the quality of cooperation and makes it possible to monitor the chain of intermodal transport.

Agent technology is an appropriate approach to design and develop a distributed system for knowledge sharing (Van Elst et al. 2007). An agent is used to denote hardware or (more usually) software based computer system that owns one or more of the following properties: autonomy, social ability, reactivity, pro-activeness, cooperation, learning, mobility and communication (Wooldridge and Jennings 1995). Due to the limitations of system resources quantity and the skills of the agent itself, the tasks performed by a single agent are very limited. In contrast, the MAS is a group of agents which works as a single system to integrate their functions and to perform large, complex tasks. The MAS offers a new dimension for cooperation and coordination in distributed collaborative environment, and provide an effective approach for coordination and cooperation to help the members of a team to share their knowledge (Zhang et al. 2008). Consequently, intermodal transport is characterized by the distributed collaborative environment, and this paper investigates knowledge sharing from a multi-agent based perspective.

Multi-Agent Based Knowledge Sharing Platform

The aim of knowledge sharing platform (KSP) is to establish a strong network for intermodal transport players, thereby obtaining more efficiencies and sustainability. A multi-agent based knowledge sharing platform called MA-KSP is a collection of agents co-operating with each other, which enables to fulfill the missions of knowledge sharing—knowledge retrieval, knowledge exchange and knowledge creation.

Starting with an introduction of the basic agent within MA-KSP, and then it provides a checklist of knowledge content for practical using; subsequently a multi-agent based architecture for KSP is depicted in the context of intermodal transport, with the details about the constituent agents and their interactive behaviors.

Basic Agent

Knowledge sharing capability seems reach less, since nowadays lower hierarchical levels have no information access to the same or a higher level. In essence, agent is an entity that performs autonomous actions based on information; it may rely on other agents to acquire or share information to achieve its goals. In principle, an agent on a level is able to access all information from other agents (Lefebvre 2003).

The abstract structure of a basic agent in MA-KSP is depicted in Fig. 2. In this paper, each agent has its own aims when sharing knowledge. Moreover, the term "agent" will be used to modules that are autonomous, interactive, learning, and have the ability to communicate over a network (Wooldridge and Jennings 1995):

- Autonomous: operate without direct intervention of humans or others, and have some kind of control over their actions and internal states;
- Interactive: the capability of interacting with other agents and possible humans via an agent-communication language;
- Learning: the ability to learn while acting and reacting in its environment;
- Communication: the ability to exchange information between disparate entities. This information exchange could be as simple as a trip signal, or could be complex such as detailed information about a remote location.

In order to make knowledge available to others and meanwhile absorbing suitable knowledge, agents of MA-KSP perceive their environment through sensors and act upon their environment through actuators. Their environment consists of other agents and various knowledge flows. In intermodal transport, sensors

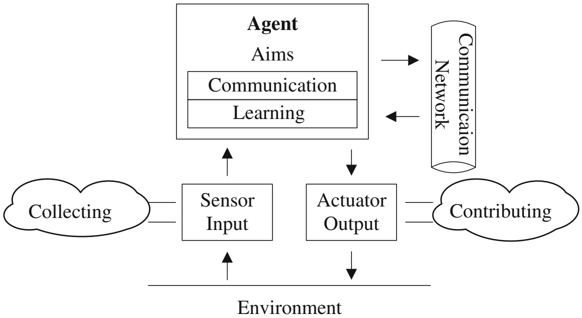


Fig. 2 Abstract structure of agents in MA-KSP

inputs might include impact factors (such as cost, transit time, reliability and services offered), online operating data, transportation monitoring. Examples of actuators outputs are strategic and tactical planning, sound decisions, selection of transport modes, operational control.

Knowledge Content

A basic understanding of knowledge content in intermodal transport enables the design of and the use of MA-KSP. Rooting in the professional activities and their objectives in intermodal transport, Table 1 summarizes a list of knowledge content involved in MA-KSP. In line with the requirements of logistics service and the characteristics of intermodal transport, the involved knowledge orientates in such

Table 1 Knowledge content in intermodal transport

Orientations		Knowledge items
Time		Technical speeds of different traffic modes
		Limitations of infrastructure capacity, and legal capacity to these speeds
		Influences of route distances on the timeliness of different traffic modes
		Extra handling time (transfer of loading units between modes, services or players)
		Scheduled time table
Cost		Budget of total cost
		Price of different transport modes and infrastructures
		Indemnity agreement of loss or damage insurance
		Intangible cost (disturbance of ecosystem, obtrusive to urban and social groups)
		Invoicing issues
Reliability	Matched capacity	Analysis and assessment of opportunities
		Route choices at the planning phase
		Simulation and prognosis of the transport capacities needed
		Adaptability to the operation environment
	Punctuality	Percentage of shipments arriving late
		Degree of delay
	Security and safety	Awareness on the subject of safety issues
		Safety quality of candidates
		Solutions for loss and damage control
		Causes of damages
Flexibility		Requirements and conditions on dangerous goods shipping
		Unforeseen changing circumstance and disturbances (natural disasters, war, strike etc.)
		List of circumstances which would require using alternate traffic modes
Legitimization		Border procedures (customs inspections, official quarantines)
		Current situation and tendency of related laws and regulations

dimensions as time, cost, reliability, flexibility and legitimization. Corresponding items are listed following these dimensions, and they represent the sources of the knowledge flows in intermodal transport.

Architecture of MA-KSP

The architecture of MA-KSP is designed as Fig. 3. It presents different agent types and their functions in the KSP. These agents are distinguished by their own special tasks: resource store, communicating facilitator and functional work.

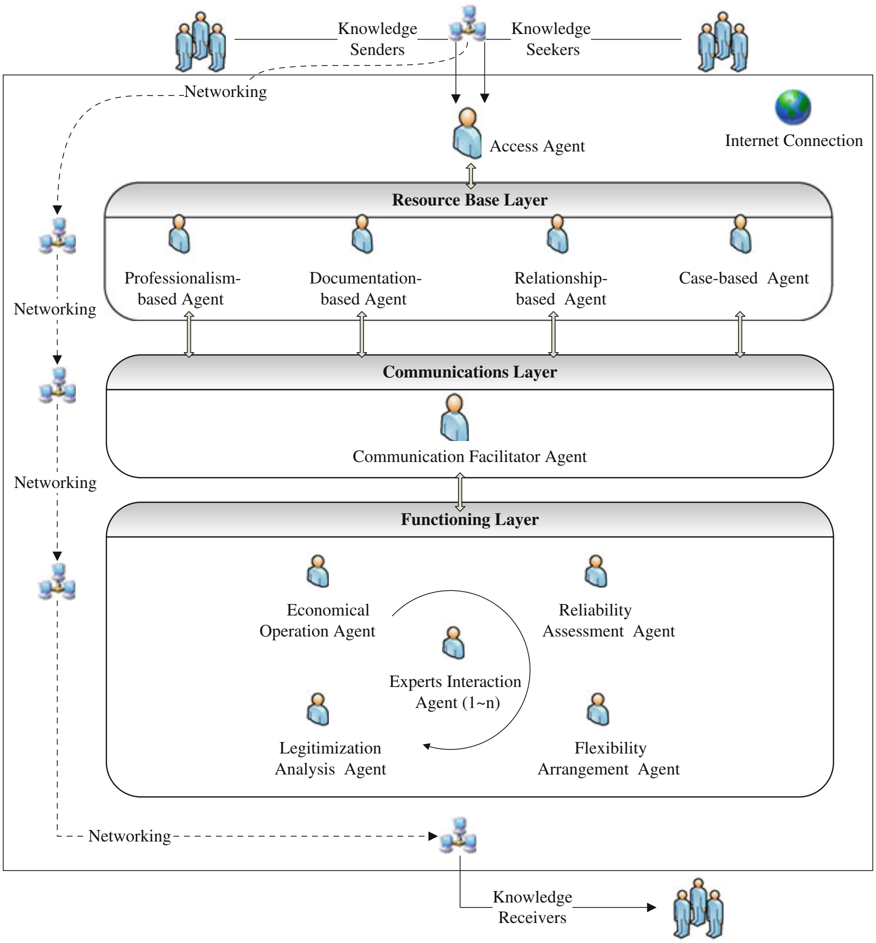


Fig. 3 Architecture of MA-KSP for intermodal transport

These distributed agents communicate and finish the tasks of knowledge sharing by a computer networking. Each agent holds its owner "background knowledge base", "data base", "model base" and "case base". In detail, background knowledge is the key for the intelligent decisions, and it can be acquired from experts' experiences. Moreover, models and algorithms offer the sources for quantitative analysis, which rely on the improvement of computing technologies in the field of transport controlling. In the process, the learning module of each agent conducts continual experience review, information collection, data mining, and knowledge discovery. Consequently the new knowledge is added into the knowledge base, meanwhile the mode base and case base are both updated, which result in more improved abilities to adapt a dynamic environment.

Any actor in intermodal transport could be a promoter driving activities of knowledge sharing. In general, they can be classified into three roles: knowledge senders, seekers and receivers.

This architecture contains four types of agents at three layers. Their responsibilities and interactive behaviors are detailed as following:

1. Access agent identifies the entered knowledge

The access agent ensures the security of the entered knowledge through browsing and identification via networking. The access agent can do in two ways. First, several pages of information could be browsed ahead in the background. Second, possible interesting pages from those browsed would be identified then presented to other agents (Kerckhoffs and Vangheluwe 1996). Access agent mainly observes the information from knowledge senders and knowledge seekers. On one hand, since an individual sender's type is unobservable to the receiver, normally there is no mechanism for a sender to credibly communicate the value of his knowledge to the receiver. Hence, before reusing or creating knowledge, access agent should identify the knowledge provided by senders are that are not-repeated or updated. Then, the qualified knowledge is generalized for storing or processing in next steps. On the other hand, facing various requests from knowledge seekers, the assess agent should ensure that the requests are potentially contributing or can be addressed currently.

2. Resource-based agents provide and assimilate knowledge

Then the permitted knowledge (knowledge sent or requests) flows towards the resource-based layer. Agents at resource-based layer can not only provide existing knowledge resource for seekers' requests, but also selectively assimilate the knowledge provided by the senders. Four agents are identified including professionalism-based agent, documentation-based agent, relationship-based agent, and case-based agent.

- Professionalism-based agent administrates the tacit knowledge attributed to a professional field, such as enterprise spirit, professions code, professional ethics, and creativity etc.

- Documentation-based agent holds the knowledge concerning address lists, documented manuals, project reports, knowledge maps.
- Relationship-based agent manages mechanisms of stakeholder relationships, which is in the form of knowledge roles definition and actors' relationships.
- Cased-based agent relates to cases storage and it offers corresponding solutions to different problems. Such solutions can be either set in advance, or obtained from the problem solving process, and only agreed excellent solutions can be added in the case base.

3. Communication facilitator agents connect and coordinate work

A special group of agents called communication facilitator agent (CFA), which are broker agents functioning in two folds: (a) they constitute a communication platform linking the resource base layer and the functioning layer. They are responsible for resource allocation in the MA-KSP; (b) CFA coordinates interactions of agents. They seek and obtain knowledge from other agents or existing knowledge base. Examples include the "Terminal Checker Agent" and the "Transport Checker Agent". One agent can contact other agents to acquire knowledge about terminal suppliers or transport suppliers. Behind the scenes the CFA will comb for the requested knowledge to make a uniform outcome.

4. Functioning agents process unusable knowledge

Unusable knowledge at previous layers will be processed in the functioning layer. The four basic functioning agents (Economical Operation Agent, Reliability Assessment Agent, Flexibility Arrangement Agent, and Legitimization Analysis Agent) are designed considering the quality of intermodal transport. Each of them is responsible for special issues, and proposes opinions from its own view.

- Economical Operation Agent (EOA) processes information relating to systematic economical efficiency, loss or damage, arriving time etc., as well as offer suggestions and solutions on how to reduce loss, enhance transport efficiency. For this, EOA uses corresponding models, algorithm (e.g. modes connecting planning, vehicle load forecast, energy consumption planning, storage turnover planning) and strategies of economical concerns.
- Reliability Assessment Agent (RAA) offers real-time assessment on reliability of intermodal transport. Similarly, RAA uses corresponding models, algorithm (e.g. transport capacities simulation, delay degree prognosis) and related knowledge. In actual working, they put forward "suffering forecast" and always notify the bottlenecks of intermodal transport.
- Flexibility arrangement agent (FAA) deals with the emergency situations. Once it receives the request of solving emergency, FAA endeavors to find out the location and cause of the emergent problems, and offers solutions.

Legitimization analysis agent (LAA) is responsible for the counseling knowledge about laws, policies and regulations. LAA contributes to give suggestions on legal feasibility, also keeps tracking on legal status and issues regarding to intermodal transport.

Expert interaction agents (EIAs) (1 – n) generate solutions in integrated manner-negotiation on the opinions of basic functioning agents. Usually they not only guide the work of other functioning agents, but also offer a final answer to the knowledge receivers through negotiations.

Example

In order to explain how the MA-KSP works in practical application, this section offers an example using MA-KSP in intermodal transport. Figure 4 outlines a real-world scenario in intermodal transport using MA-KSP. This example relates an "Intermodal Operator" in the planning phase of intermodal transport activities.

The MA-KSP serves the knowledge seeker (intermodal operator) for reusing and creating new knowledge, thereby making an optimal plan. In general sense, an intermodal operator is anyone who undertakes to arrange for a transport of goods using more than one mode of transport and who issues one transport document for the entire cargo journey. Often this refers to door to door transport (Roemer 2007). They integrate all activities in this transport chain, procure transport and trans-shipment services and take over commercial risks.

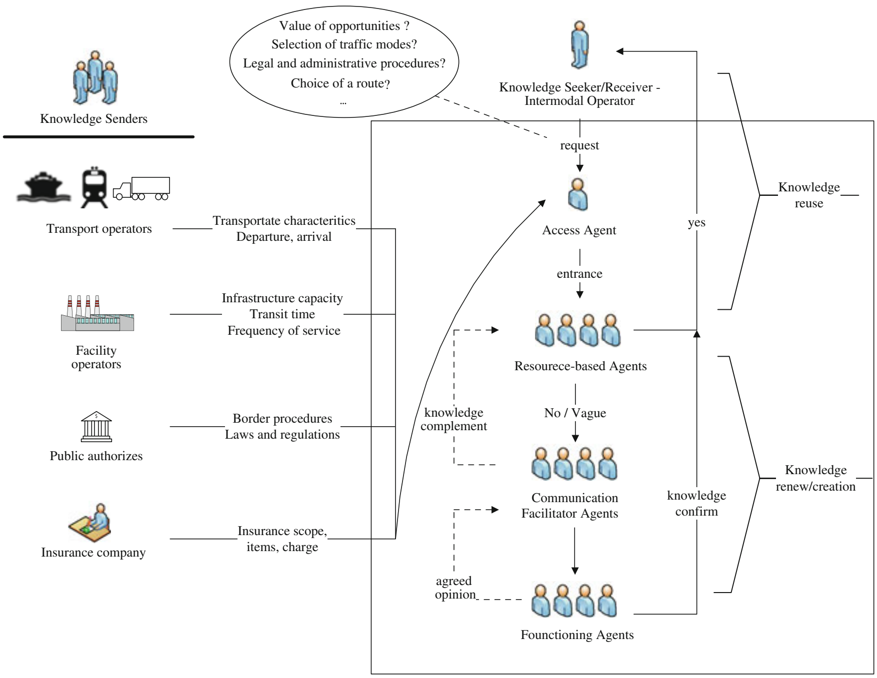


Fig. 4 A scenario of using MA-KSP in intermodal transport

Besides, “knowledge senders” (such as “transport operators”, “facility operators”, “public authorities”, and “insurance company”) send their knowledge in various forms into the KSP. They not only supply for the resource-based agents, but also participate in the negotiation phrase. Knowledge can be directly used from the resource-based layer to avoid unnecessary exchanges of related messages. Otherwise the KSP make them clear or systematic through functioning agents’ negotiation.

At the beginning, facing an opportunity, the knowledge seeker comes up with a serious of questions, e.g. whether it is worth taking the intermodal transport offer? How to select traffic modes? What are the legal and administrative procedures? What is the best route? In order to obtain the answers, the knowledge seeker requests the access agent for an entry. After the verification of identities, resource-based agents retrieve in their knowledge base, try to abstract matched knowledge for answers. If the initial questions can be answered (“yes”) in this phase, that will mean “one-way” approach for knowledge sharing, which mainly reflects the process of knowledge reusing. In detail, the intermodal operator obtains clear knowledge from these resource-based agents in such ways: similar or same cases stored tell the intermodal operator the worth of taking this task opportunity or the reason of giving away; existing documentation offers the selection of traffic modes or a complete transport route; professionalism-based and relationship-based agent can provide knowledge about corresponding legal and administrative procedures, as well as responsibilities of a intermodal operator in this task. These compose a definite answer to the request. During the knowledge abstracting and reusing process, experience learning is viewed as a leading role.

If the resource-based agents cannot offer matched knowledge or the answer is vague, the other phase (knowledge renew/creation) of knowledge sharing will occur. It is an interactive negotiation involving multiple functioning agents. Functioning agents make analysis of knowledge backgrounds of involved knowledge senders, and also do predictive parsing and analog simulation. The whole relies on the information offered by authorized knowledge senders, and the ability of functioning agents themselves, as well as their communication and negotiation skills. Usually authorized knowledge senders in this scenario are potential cooperative partners for the freight forwarder, thus their knowledge sent to the KSP are helpful. The knowledge scope ranges from transport operators conditions (such as transport characteristics, departure, and arrival) to facility ability (infrastructure capacity, transit time, frequency of service). In addition, it covers border procedures of potential routes, related laws and regulations, insurance scope, items, charge of possible insurance companies. Functioning agents respectively work out their own optimal solutions, and show attitudes to other agents. Considering the suggested optimal solutions and synthesize assessment opinions, expert interaction agents give the confirmed answer (acceptance/rejection of the potential intermodal transport task—choice of transport modes—arrangement of transmit nodes—plan of the freight route—items of preparation for the administrative affairs—kind of insurance buying) through their negotiation. Meanwhile, the agreed opinion can be used by resource-based agents to amend their knowledge bases.

A great many scenarios in intermodal transport need the MA-KSP to smooth knowledge sharing channels, e.g. settlement of a freight transport, dealing with contract problems, addressing the delay of arrival time etc. In addition, a most important factor is the active contribution from knowledge senders (they can provide quantifiable and live information). In the interaction environment, the roles of knowledge seeker, sender and receiver are always changing. Even in one scenario, one actor often plays more than one roles in the process of knowledge sharing.

Conclusions

Intermodal transport involves multiple actors working together. It accomplishes transport tasks through coordination, joint adventure, sharing profits. Nowadays it tends to be collaborative/cooperative, information/knowledge-based, and sustainability-oriented. Knowledge sharing is the act of making knowledge available to others, which can consist with this trend for better performances.

The MAS has been applied successfully for a variety of industrial problems. For the specific capabilities of MAS turn it into a promising candidate in providing a knowledge sharing solution, this paper proposes a solution to investigate knowledge sharing for intermodal transport from a multi-agent perspective. The main task of this paper is the design of the architecture of MA-KSP. It is constituted of four types of agents identified by their active roles: access agent, resources-based agent, communication facilitator agent, and functioning agent. These agents can interact and collaborate to help the users retrieving, exchanging and creating knowledge. Besides, a basic understanding of knowledge content is provided such as the knowledge relate to time, cost, reliability, flexibility and legitimization. In order to show how the MA-KSP works in the real-world, one of the scenarios in intermodal transport is illustrated. Results of this paper are expected to contribute effects to actors in intermodal transport through MA-KSP.

References

Bernaer S, Causmaecker PD, Berghe GV, Vermeulen T (2006) Agent models to control complexity in multi modal transport paper presented at the European conference on the use of modern information and communication technologies (ECUMICT)

Davidsson P, Henesey L, Ramstedt L, Törnquist J, Wernstedt F (2005) An analysis of agent-based approaches to transport logistics. Transp Res Part C 13:255–271

ECMT (1993) Terminology on combined transport (English-French-German-Russian), (Complete edition), European conference of ministers of transport (ECMT)

Haasis HD, Wildebrand H, Plöger M, Zimmermann F (2009) AMATRAK-traffic avoidance through intelligent control of freight service on the basis of autonomous multi-agent transport coordination. Paper presented at the XIII euro working group on transportation

Kerckhoffs EJH, Vangheluwe HLM (1996) A multi-agent architecture for sharing knowledge and experimental data about waste water treatment plants through the internet. Paper presented at the EUROMEDIA

Lefebvre S (2003) Dynamic output compensation between selected channels in power systems autonomous systems and intelligent agents in power system control and operation (pp 155–178). Springer, Berlin Heidelberg

Liu P, Raahemi B, Benyoucef M (2011) Knowledge sharing in dynamic virtual enterprises: a socio-technological perspective. Knowl-Based Syst 24:427–443

Muller G (1999) Intermodal freight transportation (3rd ed), Eno Transportation Foundation, United States

Roemer J (2007) Multimodal transport operators. Background paper on multimodal transport, from http://www.sla.org.sg/mto

Rondinelli D, Berry M (2000) Multimodal transportation, logistics, and the environment managing interactions in a global economy. Eur Manage J 18(4):398–410

Van Elst L, Choi I, Song M (2007) An integration architecture for knowledge management systems and business process management systems. Comput Ind 58:21–24

Vrenken H, Macharis C, Wolters P (eds) (2005) Intermodal transport in Europe. European Intermodal Association (EIA), Brussels

Wooldridge MJ, Jennings NR (1995) Agent theories, architectures, and languages: a survey. In: Wooldridge MJ, Jennings NR (eds) Intelligent agents: ECAI-94 workshop on agent theories, architectures, and languages. Springer, Berlin, pp 1–39

Zhang C, Tang D, Yanxia Liu, You J (2008) A multi-agent architecture for knowledge management system. Paper presented at the Fifth international conference on fuzzy systems and knowledge discovery

Emissions Minimization Vehicle Routing Problem in Dependence of Different Vehicle Classes

Heiko W. Kopfer and Herbert Kopfer

Abstract This contribution presents a specific objective function referring to sustainability within vehicle routing and it proposes an extension of "green" vehicle routing problems by introducing an inhomogeneous fleet. For the basic problem it is assumed that the amount of CO_2 emission depends on the distances to be travelled and on the degree to which the used vehicles are loaded. The goal of minimizing the total travel distance is contrasted with the goal of minimizing the total amount of emission. The basic problem is extended by the availability of various types of vehicles with different capacities and specific fuel consumption. This work includes a mathematical formulation of the extended problem and a small instance that illustrates the problem. Computational experiments for the basic and the extended problem formulation are performed by using CPLEX.

Introduction

The extremely high amount of emission caused by road transport gives reason for intensifying the research on ecologically oriented vehicle routing. The amount of CO_2 emission is proportional to the amount of fuel consumed for the fulfillment of

H. W. Kopfer (✉)
Department Maschinenbau, Universität Siegen,
Paul-Bonatz-Straße 9-11,
57076 Siegen, Germany
e-mail: heiko.kopfer@uni-siegen.de

H. Kopfer
Universität Bremen, Lehrstuhl für Logistik,
Wilhelm-Herbst-Straße 5,
28359 Bremen, Germany
e-mail: kopfer@uni-bremen.de

H.-J. Kreowski et al. (eds.), *Dynamics in Logistics*, Lecture Notes in Logistics,
DOI: 10.1007/978-3-642-35966-8_4,

tours (ICF 2006). An example for green transportation planning on the operational level is given by Kara et al. (2007). In this approach the fuel consumption related to a transportation plan depends on the flow of transported quantities. But fuel consumption and consequently CO_2 emission are a function of the actual weight of the used vehicles (Figliozzi 2010a; Ubeda et al. 2010) including their dead weight.

The approach introduced in this paper presents a new green version of the traditional Vehicle Routing Problem (VRP) (Dantzig and Ramser 1959). This new version aims at reducing CO_2 emissions by applying an objective function based on realistic values for the fuel consumption of modern trucks. This introduced version of an ecological VRP is called EVPR throughout this paper. Furthermore, the EVRP is extended by considering different types of trucks with specific CO_2 emissions. The extended problem introduced here is called the Emission Minimization Vehicle Routing Problem with Vehicle Categories (EVRP-VC).

Ecological Objectives and the Operational Environment

There are only few approaches for ecological vehicle routing and scheduling in literature. They mostly aim at minimizing the emitted CO_2 by considering e.g. the average speed (Figliozzi 2010a), average speed combined with acceleration rates (Figliozzi 2010b), topology (Scott et al. 2010; Ubeda et al. 2010) or the payload (Jaramillo 2010; Scott et al. 2010). Scott et al. (2010), Ubeda et al. (2010) analyze the carbon dioxide emission of fleets with various vehicle types and utilization rates. None of the known publications considers individual fuel consumption functions for accurately defined vehicle classes.

The model presented in this paper is an extension of the approach of Jaramillo (2010) with respect to both, the introduction of various vehicle categories and the minimization of the estimated fuel consumption instead of ton-kilometers. To the best of the knowledge of the authors, there is no approach presented in literature which considers and investigates these two extensions.

The vehicle categories (see Table 1) proposed in this paper are defined in accordance with the current regulations in the EC. The relevant regulations refer to license categories, acceptance tests and toll charges. The suggested categories represent a sensible graduation in compliance with current traffic laws. This has also motivated the manufacturers to establish these vehicle categories. The proposed vehicle categories differ with respect to the specific fuel consumption and

Table 1 Classification of vehicle categories

Vehicle category VC_{VWG} (–)		Gross vehicle weight GVW (to)	Load capacity Q_k (to)	Fuel consumption F_K (l/100 km)
VC_{40}		40	25	$26 + (0.36/to) \cdot q$
VC_{12}		12	5.5	$20 + (0.76/to) \cdot q$
$VC_{7.5}$		7.5	3.25	$15 + (1.54/to) \cdot q$
$VC_{3.5}$		3.5	1.5	$8 + (3.31/to) \cdot q$

thus CO_2 emissions. Table 1 shows the entries for fuel consumption which are empirical values based on information of carriers and of test reports for vehicles (see e.g. Spritmonitor 2012, Eurotransport 2012). The variable q in column 4 of Table 1 denotes the weight of the cargo carried by a vehicle.

Figure 1 shows a first simple example. For a given scenario, the difference between possible optimal solutions of the VRP on the left side and the EVRP-VC on the right side is demonstrated. In case of the VRP, one large vehicle serves all locations in a single tour providing the minimum distance solution. In order to minimize the CO_2 emission, the EVRP-VC generates a solution with three smaller vehicles out of several vehicle categories.

The plan on the left side and the plan on the right side do not only differ with respect to the clustering but also with respect to the sequencing of the customers. In the left plan, customers are served in the sequence q_8, q_{10}, q_9 before the vehicle returns to the depot. On the right side, an ecological oriented objective function is assumed. That is why it is possible that the solution deviates from the shortest path in order to obtain a solution with minimal emissions. In the case shown here, the weight of the customer request q_9 is very high. As a consequence, the emission is reduced by serving this customer q_9 first before travelling to q_{10}.

The flexibility of sequencing the customers is increased when several sub-tours are considered instead of an entire big tour. For Example, the orientation of the tour q_1, q_2, q_3 can be determined independently of the orientation of the other tours on the right side. In contrast, changing the orientation of the sub-tour q_1, q_2, q_3 on the left side will affect the integration of this sub-tour in the total tour.

The comparison of the VRP, the EVRP and the EVRP-VC is illustrated using a small example. The example consists of one depot at location [0] and 7 customer locations [1,…,7]. Table 2 shows the distance between each pair of locations.

The customer demands (given in brackets) are as follows: customer 1 (0.5 t), customer 2 (1 t), customer 3 (2 t), customer 4 (4 t), customer 5 (3.5 t), customer 6 (8 t) and customer 7 (4 t). The sum of all customer demands is 21 t, i.e., all customers can be served by a single vehicle out of category VC_{40}. The optimal solutions for the above example have been generated by our approach which will

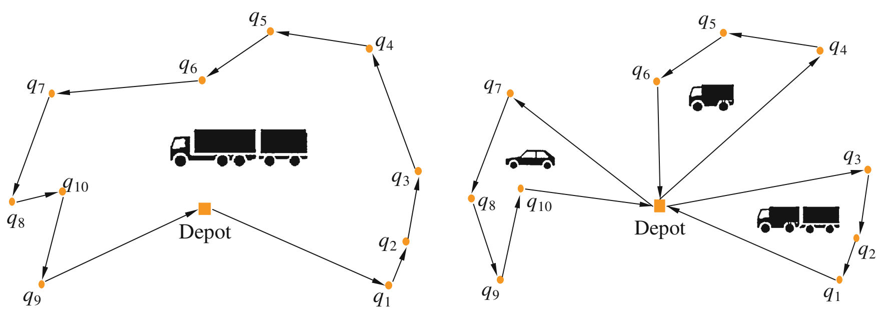


Fig. 1 Clustering and sequencing of the vehicles by different objective functions

Table 2 Distance matrix D (one depot [0] and seven customers [1,…,7]); distance d_{ij} in (km)

	0	1	2	3	4	5	6	7
0	0	85	31	73	50	77	47	92
1	85	0	60	151	59	161	122	78
2	31	60	0	97	30	109	66	75
3	73	151	97	0	123	34	86	160
4	50	59	30	123	0	85	63	53
5	77	161	109	34	85	0	123	191
6	47	122	66	86	63	123	0	70
7	92	78	75	160	53	191	70	0

Table 3 Solutions for the different models

		VRP	EVRP	EVRP-VC
Route	Vehicle VC_{40}	{0-5-3-6-7-1-4-2-0}	{0-6-3-5-4-7-1-2-0}	{0-6-7-4-0}
	Vehicle VC_{12}	–	–	{0-5-3-0}
	Vehicle $VC_{7.5}$	–	–	–
	Vehicle $VC_{3.5}$	–	–	{0-2-1-0}
Fuel consumption [l]		140.18	137.96	119.83
Total traveled distance (km)		465	474	580

Fig. 2 Vehicle routing by using the EVRP-VC

be presented in section Computational Experiments. The results are shown in Table 3. The solution for the EVRP-VC of this example is illustrated in Fig. 2.

Because of the triangle inequality, the VRP and the EVRP will always use only one vehicle if there is a vehicle which is big enough to serve all customers in a single tour. That is why, and for reasons of a fair comparison, the VRP and the EVRP are using within this example only the vehicle out of category VC_{40}, although the EVRP could possibly save fuel by using the smallest vehicle which is big enough to serve all customers. The optimal routes generated for the VRP and EVRP differ. Compared with the VRP, the EVRP-solution reduces the fuel consumption slightly by 1.5 % and increases the traveled distance by a small amount of 1.9 %. The optimal solution of the EVRP-VC shows that using trucks of different categories can reduce the amount of CO_2 emission tremendously (by 13 % compared to the EVRP and by 14.5 % compared to the VRP). On the other hand, the number of used trucks increases considerably. As a consequence, the sum of the lengths of the routes is also increasing significantly (by 22.3 and 24.7 % for the VRP and EVRP, respectively).

Model Formulation

The mathematical formulation for the EVRP-VC is built by extending a traditional VRP-formulation (Dantzig and Ramser 1959). The main extensions are:

1. the vehicle specific values for the load dependent fuel consumption must be considered in the objective function (if these vehicle specific values are equal for all available vehicles, the EVRP-VC turns out to be an EVRP),
2. the weight q_{ijk} of the goods transported by vehicle k from location i to location j must be known for all i, j, k and must be considered in the objective function.

Indices:

i, j Locations: $i, j \in I$ where 0 represents the depot, $I = \{0, 1, \ldots, n\}$.
k Vehicles: $k \in K$ where k describes the vehicle parameters, $K = \{1, \ldots, m\}$.

Parameters:

d_{ij} Distance between locations i and j.
q_j Demand of customer j for $j = 1, \ldots, n$.

Constants:

a_k Fuel consumption of the empty vehicle k per kilometer.
b_k Fuel consumption for the load of vehicle k per ton and kilometer.
Q_k Maximum load capacity of vehicle k.

Variables:

q_{ijk} Cargo of vehicle k traveling between locations i and j

x_{ijk} 1 if vehicle k serves location j immediately after serving location i, 0 otherwise.

y_{jk} 1 if location j is served by vehicle k,0 otherwise.

u_i Arbitrary real variable.

Objective Function:

$$\min \sum_{i=0}^{n} \sum_{j=0}^{n} \sum_{k=1}^{m} d_{ij} \cdot \left[a_k \cdot x_{ijk} + b_k \cdot q_{ijk} \right] \tag{1}$$

Subject to:

$$\sum_{j=1}^{n} q_j \cdot y_{jk} \leq Q_k \quad \forall k \in K \tag{2}$$

$$\sum_{i=0}^{n} x_{ijk} = \sum_{i=0}^{n} x_{jik} \quad \forall k \in K; \forall j \in I \tag{3}$$

$$\sum_{i=0}^{n} x_{ijk} = y_{jk} \quad \forall k \in K; \forall j \in I \tag{4}$$

$$\sum_{k=1}^{m} y_{jk} = 1 \quad \forall j \in I \backslash \{0\} \tag{5}$$

$$\sum_{j=1}^{n} x_{0jk} \leq 1 \quad \forall k \in K \tag{6}$$

$$u_i - u_j + n \sum_{k=1}^{m} x_{ijk} \leq n - 1 \quad \forall i \in I; \forall j \in I \backslash \{0\} \tag{7}$$

$$\sum_{i=0}^{n} q_{ijk} - \sum_{i=1}^{n} q_{jik} = q_j \cdot y_{jk} \quad \forall k \in K; \forall j \in I \backslash \{0\} \tag{8}$$

$$q_{ijk} - Q_k \cdot x_{ijk} \leq 0 \quad \forall k \in K; \forall i \in I; \forall j \in I \tag{9}$$

$$q_{ijk} \geq 0 \quad \forall k \in K; \forall i \in I; \forall j \in I \tag{10}$$

$$u_i \geq 0 \quad \forall i \in I \tag{11}$$

$$x_{ijk} \in \{0, 1\} \tag{12}$$

$$y_{jk} \in \{0, 1\} \tag{13}$$

The objective function (1) minimizes the fuel consumption. Constraints (2–7) and (11–13) are the usual restrictions of the VRP with an MTZ formulation using the variables u_i in (Eq. 7) for subtour elimination (see e.g. Toth and Vigo 2002, p 13). The usual VRP formulation is enlarged by constraint sets (8, 9 and 10). Constraints (8) are responsible for balancing the flow of goods. These equations allow the determination of the amount of the freight flow on each edge. Constraints (9) inhibit any transportation on unused edges. Otherwise it would be possible that the demanded quantities take paths differing from those of the vehicles. Finally, the transport of negative payload is excluded by constraints (10).

Computational Experiments

Using the plain VRP as a basis for our investigations has the advantage that CPLEX can be used on a personal computer with 4 GB active store and a 2.6 GHz processor for solving problem instances up to a size of ten customers, four truck categories and twelve trucks to optimality within a few seconds. Additionally, focusing on the VRP has the benefit that the analysis and comparison of the models can show the pure effects of the ecological objective function and the net effect of introducing truck categories. The introduction of time windows would have made the situation and its analysis much more complicated.

For the computational experiments we assume that a heterogeneous fleet with an unlimited number of vehicles for each category of Table 1 is available. We consider 10 problem instances with one depot at location [0] and 10 customer locations [1,…,10]. The locations are equal for all instances. They have been generated randomly with the depot in the center of the distribution area. Table 4 shows the common distance matrix for all instances.

The customer demands for the considered problem instances are given in Table 5. They have been arranged in a systematical and balanced way. For instances 1–4, all customers of an instance demand identical quantities, whereas

Table 4 Distance matrix (one depot [0] and ten customers [1,…, 10]); distance d_{ij} in (km)

	0	1	2	3	4	5	6	7	8	9	10
0	0	104	31	14	50	86	98	92	59	81	90
1	104	0	127	114	154	221	186	191	157	178	34
2	31	127	0	21	30	62	81	75	35	68	115
3	14	114	21	0	41	77	90	84	49	72	100
4	50	154	30	41	0	42	59	53	33	47	134
5	86	221	62	77	42	0	68	45	36	62	181
6	98	186	81	90	59	68	0	28	120	17	202
7	92	191	75	84	53	45	28	0	76	30	178
8	59	157	35	49	33	36	120	76	0	93	144
9	81	178	68	72	47	62	17	30	93	0	171
10	90	34	115	100	134	181	202	178	144	171	0

Table 5 Demand matrix [demand q in (to)]

Inst.	q_1	q_2	q_3	q_4	q_5	q_6	q_7	q_8	q_9	q_{10}	Σ
1	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	2.5
2	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	5
3	1	1	1	1	1	1	1	1	1	1	10
4	2.75	2.75	2.75	2.75	2.75	2.75	2.75	2.75	2.75	2.75	27.5
5	0.25	0.5	1	2	3	3	2	1	0.5	0.25	13.5
6	3	2	1	0.5	0.25	0.25	0.5	1	2	3	13.5
7	1	2	3	4	5	5	4	3	2	1	30
8	5	4	3	2	1	1	2	3	4	5	30
9	0.5	1	2	4	8	8	4	2	1	0.5	31
10	8	4	2	1	0.5	0.5	1	2	4	8	31

the quantities are increasing from instances 1 to 4. For instances 5–10, the demands of different customers vary within a single instance. For these instances, the total demand of all customers is increasing starting with 13.5 t and ending with 31 t. The total demands and the individual demands are equal for instances out of the pairs (5,6), (7,8) and (9,10). The instances within a single pair only differ with respect to the assignment of demands to the customers.

The results of the experiments are shown in Table 6. Compared to the VRP, the average fuel consumption related to the 10 solutions of the EVRP can be reduced by 4.26 % from 152.46 l to 145.96 l, whereas the traveled distances on average increase by only 0.3 % from 527.2 km for the VRP to 528.9 km for the EVRP. This comparison clearly shows that, at least for the 10 instances considered in this paper, the EVRP produces solutions with slightly increased tour lengths but with a considerable decrease with respect to fuel consumption. In general, the experiments demonstrate that it is worth to investigate the characteristics and the advantages of the EVRP in more detail and to analyze the trade-off between the solutions of the VRP and the EVRP. For six instances of Table 6, the length of the routes generated for the VRP and the length of those generated for the EVRP are totally equal. The VRP-solutions and the EVRP-solutions for these six instances differ only with respect to the orientation of the generated routes. The VRP is indifferent to reversing the orientation of routes, i.e., a given solution and the solution with the reverse order for serving the customers are considered to be equal. For the EVRP this is not true because the fuel consumption depends on the actual weight of the vehicle on the legs of the entire route. That is why the EVRP-solution always chooses the orientation with less fuel consumption. This orientation is referred to as "light" orientation whereas the other one is referred to as the "heavy" one. Considering all instances of Table 6 with equal tour length, the average difference between the fuel consumption (*fc*) for the "heavy" orientation (shown in Table 6 as *fc* for the VRP) and for the "light" orientation (shown as *fc* for the EVRP) amounts to 5.46 l. This means that in Table 6 most of the fuel reduction (84 %) that has been achieved by switching from the VRP to the EVRP is realized by choosing the right orientation of the shortest route.

Table 6 Fuel consumption (*fc*) and distances (*d*) in dependence of the chosen objective function

Inst.	VRP		EVRP		EVRP-VC	
	fc in [l]	*d* in (km)	*fc* in [l]	*d* in (km)	*fc* in [l]	*d* in (km)
1	136.53	514	135.38	514	54.98	576
2	139.41	514	137.10	515	69.79	699
3	145.18	514	139.81	521	113.95	684
4	151.22	521	147.42	521	136.09	521
5	155.55	514	140.71	514	102.79	522
6	146.57	514	140.70	521	129.57	704
7	157.25	541	154.62	541	130.31	705
8	152.82	521	147.41	521	147.41	521
9	187.63	598	169.01	600	126.80	644
10	152.39	521	147.48	521	147.48	521

The results of Table 6 which have been attained for the EVRP-VC show that a huge potential for fuel saving can be reached by introducing an inhomogeneous fleet of vehicles.

Applying the EVRP-VC, the average fuel consumption for the instances of Table 6 decreases to 115.9 l, which means a reduction of 24 % compared to the VRP. It is worth to mention that the amount of reduction varies a lot over the considered instances from 3 % (instance 10) to 60 % (instance 1). Even very similar instances differ a lot with respect to the potential of the EVRP-VC to reduce fuel consumption and to increase travel distances. See e.g. instances 9 and 10 having identical customer demands which have been assigned to the customers in a different way. The numbers of used vehicles as well as the traveled distances are increasing drastically by applying the EVRP-VC instead of the VRP. The travel distances increase on average by 16 % and the number of used vehicles is increased for seven of ten instances. For the instance 3, there are even five vehicles used instead of one vehicle for the VRP. In order to reduce the problems arising from a heterogeneous and enlarged own fleet, the remedy of subcontracting suitable requests to specialized carriers with small and cheap vehicles is recommended. This can keep the increase of costs within a limit since the regulations for smaller vehicles are much more lax to many respects.

Conclusions and Future Research

The results of the computational experiments show that it is worth to solve the EVRP additionally to the VRP and to contrast these two approaches and the solutions generated by them. Introducing heterogeneous fleets of vehicles with different capacities and fuel consumption opens up a tremendous potential for saving CO_2 emissions. Solutions of the EVRP-VC show that this potential can reach a reduction of 20 % and even more. The price to be paid for this reduction is that the total travel distances as well as the number of used vehicles increase drastically. For balancing the aforementioned benefits and drawbacks, an extended model with an integrated cost function considering fuel consumption, travel distances and fixed costs of the vehicles will be developed.

For a continued analysis of the relations among the VRP, EVRP and EVRP-VC, it is necessary to expand the above experiments. On the one hand, several characteristic scenarios represented by sensibly configured large sets of problem instances have to be defined and analyzed. On the other hand, powerful heuristics are needed to enable the solution of the relevant optimization models for typical medium sized and large sized problem instances. Finally, in successive research, conclusions for the configuration of vehicle fleets can be derived.

References

Dantzig GB, Ramser RH (1959) The truck dispatching problem. Manage Sci 6:80–91

Eurotransport (2012) www.eurotransport.de/test/Vergleichstest-mercedes-axor-volvo-fm-scania-g-man-tgs-431402.html, read out at Jan 4 2012

Figliozzi M (2010a) Emissions minimization vehicle routing problem. In: 89th annual meeting of the transportation research board, USA, Jan 2010

Figliozzi M (2010b) Vehicle routing problem for emissions minimization. Transp Res Rec: J Transp Res Board 2197:1–8

ICF (2006) Assessment of greenhouse gas analysis techniques for transportation projects. 6 NCHRP Project 25-25 Task 17, ICF Consulting, Fairfax, VA

Jaramillo JR (2010) The green single vehicle routing problem. In: SE INFORMS annual meeting, Myrtle Beach, SC Oct 2010

Kara I, Kara BY, Yetis MK (2007) Energy minimizing vehicle routing problem. In: Dress A, Xu Y, Zhu B (eds) Combinatorial optimization and applications. Lecture notes in computer science, vol 4614. Springer, Berlin, pp 62–71

Scott C, Urquhart N, Hart E (2010) Influence of topology and payload on CO_2 optimized vehicle routing. In: Chio C (ed) Applications of evolutionary computation. Springer, Berlin

Spritmonitor (2012) www.spritmonitor.de, read out at 4 Jan 2012

Toth P, Vigo D (2002) The vehicle routing problem. SIAM, Philadelphia

Ubeda S, Arcelus FJ, Faulin J (2010) Green logistics at Eroski: a case study. Int J Prod Econ 131(1):1–8

Product Intelligence in Intermodal Transportation: The Dynamic Routing Problem

Vaggelis Giannikas and Duncan McFarlane

Abstract Intermodalism has been recognised as a promising way to efficiently reduce logistics costs and travel times of shipments in today's freight industry. Further, it has the potential to add resilience to the transportation network by providing options in the case of disruption. However, intermodalism still faces some operational challenges such as the routing of shipments in the intermodal network and the control of these shipments during their transit. In this paper we focus on the dynamic version of the routing problem in intermodal transportation that is the control of a shipment while it moves in the transportation network. We present an approach (based on the so-called "intelligent product" notion) for efficient dynamic routing and discuss its applicability and differences with current practice.

Introduction

Intermodal transportation can be defined as "[…] the transportation of a person or load from its origin to its destination by a sequence of at least two transportation modes" (Crainic and Kim 2006). The underlying idea of intermodalism is that the cooperation of different transport modes can lead to substantial advantages resulting from the combination of their individual strengths in order to build an efficient and independent transport system (Buchholz et al. 1998). As an example

V. Giannikas (✉) · D. McFarlane
Institute for Manufacturing, University of Cambridge, 17 Charles Babbage Road, Cambridge, CB3 0FS, UK
e-mail: eg366@eng.cam.ac.uk

D. McFarlane
e-mail: dcm@eng.cam.ac.uk

H.-J. Kreowski et al. (eds.), *Dynamics in Logistics*, Lecture Notes in Logistics,
DOI: 10.1007/978-3-642-35966-8_5,

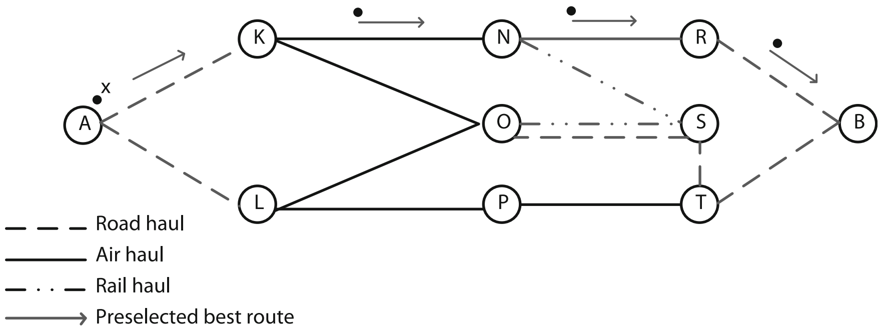


Fig. 1 An intermodal transportation network

of an intermodal network, Fig. 1 depicts a scenario where an order x has to move from place A to place B through terminals $K - T$ having a set of available routes and modes that it can use. We discuss this specific scenario later.

One of the main operational problems that intermodal operators face is the so-called intermodal routing problem (Chang 2008), that is the selection of the best route for the transportation of a shipment given a set of options that the transportation network can offer. Furthermore, due to disruption factors in the network, an intermodal operator may need to control shipments during transportation and make decisions about alternative routes in case the initial route is no longer available. The management of *intermodal dynamic routing* involves these aforementioned decisions and is characterised by a number of special requirements which may affect its performance.

In this paper we are particularly interested in addressing intermodal dynamic routing problems using the so-called "intelligent product" paradigm. The intelligent product notion involves control architectures where industrial operations are managed and controlled more via their moving parts (e.g. inventory, finished products) and less by centralised decision systems (McFarlane et al. 2003; Kärkkäinen et al. 2003). Such an approach has been shown to bring special benefits in terms of disruptions management and resilience in logistics environments of manufacturing processes (Pannequin et al. 2009). Moreover, it is argued that it can improve logistics and transportation operations in supply chain management again by increasing the robustness of the system (Meyer 2011). For these reasons, we recognize potential synergies between the notion of product intelligence and intermodal dynamic routing and we discuss the potential benefits.

Intermodal Routing Problem

In this section we begin with describing the so-called "a priori intermodal routing problem" which is addressed before the shipment of an order. Then we focus on the dynamic version of this problem which deals with the handling of the routes of shipments while they are in transit.

A-Priori Intermodal Routing

Intermodal routing deals with route and service choices in existing intermodal networks. Intermodal operators who are responsible for this process act as customers/users of the intermodal infrastructure and services trying to find the optimal route for a certain shipment. They buy the services offered by different modes and manage the whole process, from the collection of the product until its delivery (Macharis and Bontekoning 2004). As an example, in the scenario depicted in Fig. 1, let us assume that the pre-selected best route for order x from A to B is $A \rightarrow K \rightarrow N \rightarrow R \rightarrow B$. The nature of the a priori intermodal routing problem makes its solution a very complex task for several reasons. One of the main factors that affect intermodal routing is time. An intermodal operator may have a different set of available options for the routing of his shipment during a period of time as scheduled lines can vary a lot. Moreover, since by definition this particular transportation problem can potentially use the whole intermodal network, different people and organisations have to coordinate even for the distribution of a single shipment. Ziliaskopoulos and Wardell (2000) and Chang (2008) summarise the important characteristics that increase the complexity of the intermodal routing problem in the multi-objective nature of the problem as different stakeholders may have different concerns, the discontinuities of the fixed schedule lines that many modes often use, the delays in switching points and the disruptions while in transit as well as the violation of the first-in-first-out policy on certain links.

Dynamic Intermodal Routing

The a priori version of the routing problem described in the previous section deals with the process of finding the best path in an intermodal network before the dispatch of a shipment and there is a number of studies that try to solve it (Macharis and Bontekoning 2004; Caris et al. 2008; Chang 2008). However, in real practice, it is often the case that while an order is in transit, its a priori defined route is blocked or disrupted and a revised route would ideally be assigned (Azadian et al. 2012). We call this the *intermodal dynamic routing problem* which is defined as the problem of identifying the best new route for an order while it is in transit and part or all of its pre-selected route is blocked or disrupted. The problem also involves all the decisions that need to be made in order for this new route to be agreed and used. From a mathematical perspective, the problem is closest to the stochastic time-dependent shortest path problems. The reader is referred to (Azadian et al. 2012) for a very thorough review on this type of problems.

Back to the example illustrated in Fig. 1, let as assume that when the order x is in transit and has arrived at Terminal K, the route $K \rightarrow N$ is no longer available due to a strike in the airport located in N making the $K \rightarrow N \rightarrow R$ part of the initial route inaccessible. The alternative decisions that an intermodal operator can choose from in case of such a disruption are summarised in Table 1.

Table 1 Intermodal dynamic routing decisions

Decision	Description	Example
Do nothing—Wait for next available movement	Shipments will wait in a terminal until the next similar way of transportation is available	Order *x* can wait until the strike in terminal *N* is over and aircraft can use it again
Select a different path in the transportation network	A shipment is re-planned to use a different set of terminals and modes for its transportation	Order *x* can move through *O*, *S*, and *T* or *O*, *L*, *P* and *T* in order to reach *B* from *K*
Schedule truck movement	Arrange the transportation of a shipment using a truck. Road transportation is normally more flexible as trucks can be scheduled at any time	Scheduling of a truck movement from *K* to *N* creating an alternative route for order *x* that did not exist before
Schedule additional aircraft/ship/train movements	Similar to scheduling truck movements. However, extra air, sea and rail transportation is normally much more difficult and costly to be scheduled than road transportation	As before

In the case of dynamic intermodal routing the identification of a new route for the order becomes even more complex than in the a priori case, having a number of special requirements:

1. *Order-level information and decisions*: Since disruptions may have different effects on each shipment, different decisions may need to be made for each one based on their individual special characteristics.
2. *Lifecycle information*: Information spanning different phases of the transportation process is required in order to make the best decision about the next steps of each order. Here, the lifecycle of a shipment starts with a request for transit and finishes with its delivery.
3. *Distributed decision making*: Often, there is not a single organisation controlling the whole transportation network. Thus, it is not possible for any organisation to make all the decisions concerning all the stages of transportation of a shipment. For this reason, planning and controlling decisions have to be made and executed in a distributed environment, among different companies and geographical places.
4. *Multi-objective nature of decisions*: As discussed above, this is one of the main characteristics of routing problems. Since there is no one person/organisation planning and managing the whole transportation of a order, he cannot manage services provided by different modes based on current demand. A larger set of criteria have to be taken into account when decisions about the routing of an order are being made.

5. *Time-critical decisions*: The time that a decision is being made may affect the options that the decision maker may have, normally limiting rather than increasing their number.
6. *Time-consuming problem solving*: The high complexity of finding a new best path does not allow re-computation of solutions in real time.

Product Intelligence in Logistics

Many of the requirements identified in the previous section can be addressed within an intelligent product approach. For this reason, in this section we introduce the intelligent product notion, its potential benefits as well as its similarities with other concepts.

The intelligent product approach was first proposed as a building block for control architectures that could offer benefits to manufacturing, supply chain management and logistics in terms of more customised management of individual products and orders (Wong et al. 2002). Prior to this, many influences can be seen in the multi-agent based industrial control field (e.g. Bussmann and Schild 2000) and holonic manufacturing (e.g. Van Brussel et al. 1998). These control architectures, such as the *Auto-ID driven* control architecture (McFarlane 2002), the *Distributed, Intelligent Product Driven* control architecture (McFarlane et al. 2003), the *Inside-out* control architecture (Kärkkäinen et al. 2003) or the *Product-Instance-Centric* control architecture (Hribernik et al. 2006), suggested that industrial operations could be managed and controlled more via their moving parts and less by using centralised decision making systems oriented around the organisation and its resources. In this context, an *intelligent product* was recognised as a physical and information based representation of a product which has its own identification, can collect, retain and exchange information with its environment as well as participate in the decision making process about its next steps (Wong et al. 2002) (See Table 2 for complete definition).

Although these initial developments are well accepted in the literature, the characteristics of an intelligent product and the fundamental ideas behind it can also be found in other emerging technological topics such as *smart objects*, objects in *autonomous logistics* and the *Internet of Things*. For example, Table 2 shows the connection between the concept of product intelligence and objects autonomy comparing their definitions as well as the different levels of intelligence/autonomy of objects in these approaches showing that there might be significant synergies among the different concepts (Uckelmann et al. 2010; McFarlane 2011). A good example of how the principles of software agents and autonomous logistics can be used for supply network management problems is analysed in (Schuldt 2011).

The adoption of an intelligent product approach in industrial operations has been argued to bring special benefits to its users such as increased robustness in the

Table 2 Characteristics of intelligent products and autonomous logistics

	Intelligent products (Wong et al. 2002; Kärkkäinen et al. 2003)	Autonomous objects (Hülsmann and Windt 2007; Scholz-Reiter et al. 2008)
Definitions	Possesses a unique identity	Self identification and detection system
	Communicates effectively with its environment	Communication ability ICT
	Retains or stores data about itself	
	Deploys a language to display its features etc.	Communication ability ICT
	Is capable of participating in or making decisions relevant to its own destiny	Information processing ability to identify alternatives evaluation system execution
Levels of intelligence/ autonomy	Information handling	Identify and store static data
		Gather and store dynamic data
	Decision making	Process data to create new ones and solve tasks
		Exchange data with other objects and interact with them
		Intelligent information based material handling is used to initiate actions

face of change and effective management of disruptions. Apart from some qualitative statements of this argument (McFarlane et al. 2003; Morales-Kluge et al. 2011), it has also been quantitatively shown that autonomy and intelligence in products can create more robust and flexible systems. As an example, the adoption of an autonomous product-driven control architecture in a production cell is shown to improve work in process levels and lead times especially in cases where disturbance factors like perturbations take place (Pannequin et al. 2009). In another study, it is shown that dynamic routing algorithms can lead to increased levels of robustness and adaptability in production logistics (Sallez et al. 2009). In the area of vehicle routing and transportation, Meyer (2011) argues that an intelligent product approach could have significant impact in problems of supply networks based on interviews with users of a prototype that uses such an approach.

Towards an Intelligent Product Approach to Intermodal Dynamic Routing

In this section we now make the connection between the intelligent product approach and the intermodal dynamic routing problem and its potential implementation is discussed.

Linking Intelligent Products with Intermodal Routing Problems

Since one the main benefits of an intelligent product approach seem to be the improved management of disruptions as we saw in the previous section, we conjecture that in the case of intermodal transportation and logistics, an intelligent product approach can have a very positive impact in the planning and managing decisions in the dynamic routing problem. Table 3 illustrates how the intermodal routing characteristics developed in a previous section map onto characteristics of an intelligent product approach. From the table we can see that an intelligent product approach might be appropriate for the solution of the intermodal dynamic routing problem:

At the information-handling level (characteristics 1–3), an intelligent product approach can support important serial-level information necessary for the support of fast and distributed decision making. First of all, by communicating with its environment a product or order or shipment will be capable of gathering information about its status during its whole lifecycle as well as collecting goals and objectives of different stakeholders while it is in transit. Utilising the above information can help operators identify and compare available alternatives which they could not normally find before. For example, this information can be used for the identification of alternative routes for an order ("different path in a transportation network" option), something that is not very usual in current practice.

Table 3 Linking intermodal dynamic routing and intelligent products

Intermodal dynamic routing requirements	Intelligent product characteristics				
	1. Unique identification	2. Comms with environment	3. Data gathering and storage	4. Language to display features	5. Decision making
Order-level information	*	*	*		
Lifecycle information	*	*	*		
Distributed decision making			*	*	*
Time-critical decisions			*	*	
Multi-objective decisions	*				*
Time consuming problem solving	*				*
Order-level decisions	*			*	*

Secondly, this information (e.g. order's current location, delay) can be accessed by organisations in different locations in order to optimise their decisions about the movement of an order. Negotiations can be more effective when all the important information is available. Moreover, even in cases when negotiation is not necessary, a single organisation can optimise its decisions when it holds information about the whole lifecycle of a shipment rather than its next step only. Finally, since each order will be uniquely identified and linked with individual data, the time required for the process of identifying the optimal route can be also reduced as all the required information will be easily accessible.

Including the decision making level as well (characteristics 4–5), an intelligent product approach can also support the decision making processes in the intermodal dynamic routing in two ways: Firstly by support communication between several operators of specialised information and secondly by enabling serial-level decisions to be made for different orders.

In the first case, customised views of appropriate information regarding a shipment's individual history and future needs can improve the final decisions about their next steps. For example, a terminal operator could reroute an order more effectively if he knew its final destination instead of just having access to the order's next station. This could reduce the time needed for the transportation of the order compared with the case where an order is put in the next available mode that operates the specific route. In the second case, orders make their own decisions about their next steps. Here, the intermodal dynamic routing problem could benefit in terms of facilitation of distributed and multi-objective decision making as the product could take into consideration several goals from multiple stakeholders and then choose the optimal solution. Orders, represented by software agents could negotiate with several stakeholders (who in turn will be represented by other software agents) and reach to an agreement without the participation of people, thus saving important working time.

Issues in Implementing an Intelligent Product Approach

As implied in an earlier section, the dynamic routing of orders in an intermodal network is not a futuristic scenario. As intermodal operators already face disruptions and other similar problems in their networks, they have found ways to overcome these difficulties. Current practice follows a rather *organisation oriented* approach where, when a new decision about a shipment's next steps has to be made, the organisation in charge of the shipment makes this decision (Christopher 2011, Chap. 7). For example, a carrier may decide to delay a shipment and expedite another one if this will maximise its benefit, irrespective of the end customer's and the logistics provider's needs. Such an approach often satisfies the needs of one organisation without taking into account the other stakeholders. Moreover, the tools that are currently in use mainly facilitate the adoption of the "next available movement" option (see Table 1) leaving aside other alternatives that could be more beneficial for the organisation in charge as well as its customers.

On the other hand, an intelligent product solution would provide a more *shipment oriented* or even *customer oriented* approach where, when a new decision about the next steps of a shipment has to be made, the shipment itself will seek for the best option. In this case, a shipment, representing a customer's/owner's needs, would negotiate with the different organisations participating in the planning and management of the shipment and reach to a decision regarding its next steps. Human travelling is a good analogy for this case. As people do, a shipment will decide which path is best for it given a set of parameters like the cost of transportation, the lead time etc. When a certain disruption happens, a shipment will seek for the available alternatives and decide what will maximise its benefit.

Since the transition from an organisation oriented approach to one which purely uses intelligent products would require fundamental changes to a number of elements of current practice (such as operational processes, data availability, technology used etc.), an intermediate approach might be used as the transitional phase between the two. In such an approach, when a new decision about a shipment's next steps has to be made, the different stakeholders will negotiate in order to come to an agreement. For example, in case a carrier needs to delay a shipment, he will negotiate about the alternative solutions for this shipment with the logistics provider and the customer before making the final decision. This *shipment oriented (but organisation based)* approach is already used by some organisations today. However, the role of the customer is not very significant and the tools in use for these negotiations can be very time consuming for all actors (Schuldt 2011).

Conclusions

In this paper we have examined the way in which an intelligent product approach can be used for the improvement of the control of dynamic routing in intermodal networks. We did that by presenting a number of synergies that exist between the particular characteristics of this approach and the operational requirements of the intermodal dynamic routing problem. Among them, the identification and comparison of alternative routes as well as the negotiation between stakeholders for the decision of the final route seem to be the most important benefits that an intelligent product approach can bring to this problem.

We also discussed the different issues associated with the deployment of the intelligent product one starting from current practice and moving through a transitional phase. We summarise the main differences between these approaches:

- *Data access*: A shift from organisation-based to open access data models where information is distributed among and transferred between organisations.
- *Action space*: Broadened from considering only organisation-centric actions to a set of all available actions in the whole intermodal network.

- *Business values/priorities*: A shift from organisation-centric priorities to more customer-centric priorities where customer satisfaction plays a crucial role.

References

Azadian F, Murat AE, Chinnam RB (2012) Dynamic routing of time-sensitive air cargo using real-time information. Transp Res Part E: Logist Transp Rev 48(1):355–372

Buchholz J, Clausen U, Vastag A (1998) Handbuch der Verkehrslogistik. Springer, Berlin

Bussmann S, Schild K (2000) Self-organizing manufacturing control: an industrial application of agent technology. In: 4th international conference on multi-agent systems, pp 87–94

Caris A, Macharis C, Janssens G (2008) Planning problems in intermodal freight transport: accomplishments and prospects. Transp Plan Technol 31(3):277–302

Chang T-S (2008) Best routes selection in international intermodal networks. Comput Oper Res 35(9):2877–2891

Christopher M (2011) Logistics and supply chain management. Pearson education, Harlow

Crainic T, Kim K (2006) Intermodal transportation. In: Barnhart Cynthia, Laporte Gilbert (eds) Handbooks in operations research and management science: transportation. North-Holland, Amsterdam, pp 467–537

Hribernik K, Rabe L, Thoben K, Schumacher J (2006) The product avatar as a product-instance-centric information management concept. Int J Prod Lifecycle Manage 1(4):367–379

Hülsmann M, Windt K (2007) Understanding autonomous cooperation and control in logistics: the impact of autonomy on management, information communication and material flow. Springer, Berlin

Kärkkäinen M, Holmström J, Främling K, Artto K (2003) Intelligent products–a step towards a more effective project delivery chain. Comput Ind 50(2):141–151

Macharis C, Bontekoning YM (2004) Opportunities for OR in intermodal freight transport research: a review. Eur J Oper Res 153(2):400–416

McFarlane D (2002) Auto ID based control systems-an overview. In: 2002 IEEE international conference on systems. Man and Cybernetics

McFarlane D, Sarma S, Chirn JL, Wong CY, Ashton K (2003) Auto ID systems and intelligent manufacturing control. Eng Appl Artif Intell 16(4):365–376

McFarlane DC (2011) Distributed intelligence in manufacturing & service environments. In: 1st international SOHOMA workshop, Paris

Morales-Kluge E, Farideh G, Scholz-Reiter B (2011) Intelligent products—towards autonomous logistics processes—a work in progress paper. In: International conference on product lifecycle management

Meyer GG (2011) Effective monitoring and control with intelligent products. PhD thesis, University of Groningen

Pannequin R, Morel G, Thomas A (2009) The performance of product-driven manufacturing control: an emulation-based benchmarking study. Comput Ind 60(3):195–203

Sallez Y, Berger T, Trentesaux D (2009) A stigmergic approach for dynamic routing of active products in FMS. Comput Ind 60(3):204–216

Scholz-Reiter B, Uckelmann D, Gorldt C, Hinrichs U, Tervo J (2008) Moving from RFID to autonomous cooperating logistic processes. In: Miles B et al. (eds) RFID technology and applications, Cambridge University Press, Cambridge, pp 183–197

Schuldt A (2011) Multiagent coordination enabling autonomous logistics. Springer, Berlin

Uckelmann D, Isenberg M-A, Teucke M, Halfar H, Scholz-Reiter B (2010) Autonomous control and the internet of things: increasing robustness, scalability and agility in logistic networks. In: Ranasinghe DCC, Sheng QZZ, Zeadally S (eds) Unique radio innovation for the 21st century, Springer Berlin, pp 163–181

Van Brussel H, Wyns J, Valckenaers P, Bongaerts L, Peeters P (1998) Reference architecture for holonic manufacturing systems: PROSA. Comput Ind 37(3):255–274

Wong C, McFarlane D, Zaharudin A, Agarwal V (2002) The intelligent product driven supply chain. In: IEEE international conference on systems, man and cybernetics, p 6

Ziliaskopoulos A, Wardell W (2000) An intermodal optimum path algorithm for multimodal networks with dynamic arc travel times and switching delays. Eur J Oper Res 125(3):486–502

Cargo Telematics for Operational Transport Excellence and Strategic Knowledge Management

Matthias Klumpp, Christof Kandel and Erik Wirsing

Abstract The paper at hand deals with two major trends in logistics: On the one hand, increasing importance of information and communication technology for logistics processes and on the other hand, development towards lifelong-learning and professional education in logistics to improve companies' knowledge management. The authors discuss future research directions of these two developments by means of cargo telematics, especially for logistics service providers. Therefore operational process benefits for creating customer value, performance measurement and future planning challenges are connected to logistics learning capabilities and corporate knowledge management and result in a conceptual framework to implement technical developments in learning environments.

Introduction

Flexibility and variability of supply chains increase because of globalization influences and growing market competition. Industry underlies strong fluctuations which strike the logistics sector in a multiple way. That is why dynamics in logistics are of particular importance for the transportation sector and why the transport

M. Klumpp (✉) · C. Kandel
FOM University of Applied Sciences, Institute of Logistics, Essen, Germany
e-mail: matthias.klumpp@fom-ild.de

C. Kandel
e-mail: christof.kandel@fom-ild.de

E. Wirsing
Schenker Deutschland AG, Kelsterbach, Germany
e-mail: erik.wirsing@dbschenker.com

H.-J. Kreowski et al. (eds.), *Dynamics in Logistics*, Lecture Notes in Logistics,
DOI: 10.1007/978-3-642-35966-8_6,

processes are often bottlenecks within production networks (Meers et al. 2010). For this reason logistics service providers are influenced by this development. They have to tackle these challenges by using information and telecommunication support to react on nowadays market requirements, like e.g. increasing transport volume, short reaction times, high security levels and process stability which are demanded by customers (Vahrenkamp 2007). Modern cargo telematics or state-of-the-art track and trace solutions can tackle these tasks by providing several added services for the users. They offer the possibility to localize shipped goods at any time in real-time and to display travelled routes exactly. This benefits future optimization and planning challenges and lead to higher process performance at the transportation execution level (Schenk and Richter 2007). On top these solutions support company's learning capabilities by transforming tacit into explicit knowledge (Abidi et al. 2011; Smith 2001), for example by supporting non-local truck drivers with customer-related information or logistics managers with style sheets for transport calculations (Klumpp et al. 2011). First, the paper at hand gives an overview about current solutions and developments in cargo telematics and defines differences between discrete and continuous track and trace systems. Second it is discussed how logistics service providers can be supported on an operational level with benefits by using continuous track and trace systems, before the relevance of these benefits for company's knowledge management is outlined.

State-of-the-Art in Track and Trace

Track and trace in the courier, express and parcel sector is established. Service providers offer this standard service to their customers. With an identification number they get the possibility to check the current status of their shipments. For private purposes this event monitoring is an additional benefit but for industry issues accuracy of localization and for example estimated arrival times are of higher interest. However, only few solutions offer these added values to customers at present. Therefore continuous tracking solutions have to be realized based on satellite navigation (Brewer et al. 1999).

Barcoding or Radio Frequency Identification (RFID) are solutions which only provide the identification of discrete events (Schöch and Hillbrand 2006). In fact, GPS or in future GALILEO provides the possibility to realize a continuous track and trace of shipments but demands active power supply. Then, localization is possible without an external reader device (Kärkkäinen et al. 2004). Boundaries of satellite tracking are for example warehouses, production plants or other closed buildings wherein localization is not possible. Important for the real-time access of data is a combination of such a GPS tracking module with an active communication device so that the tracking data can be transmitted. Therefore, for example mobile internet communication based on GPRS (General Packet Radio Service) can be used, but other communication devices like UMTS as well. During air cargo or on seaway transports real-time position transmission is still not possible with this technology. Therefore other communication devices are necessary: For example satellite communication for seaway transports or standardized interface connections to the carriers' telematics. Beside satellite navigation a continuous tracking is also possible based on the cell location information of GSM: By calculating runtimes to at least three cell towers a position can also be estimated at any time. However, accuracy of this position is poor; it varies from 100 m up to 35 km (Schöch and Hillbrand 2006). Figure 1 shows an overview of different solutions for track and trace.

The most common solution in practice to realize identification and localization of shipments anytime is a *quasi-continuous* solution which is also displayed in Fig. 2: If the vehicle is fitted with telematics a GPS localization of the vehicle is possible. If in addition the telematics unit is able to administrate shipments or transfer orders based on barcode or RFID information, a mapping of shipments to the GPS data of the vehicle is possible and a real-time positioning can be done. This leads to a high computable complexity, especially if different logistics service providers work together within one network or if many sub-contractors are applied (Carlino et al. 2009; He et al. 2009; Blümel et al. 2008; Wang and Potter 2008; Brewer et al. 1999).

The major problem of using a tracking system within a supply chain or a distribution network is the holistic integration of one system, especially if actors within the supply chain vary from time to time or if the logistics service provider executes jobs in different networks (Meers et al. 2010). Therefore a battery powered solution could solve this difficulty (Kärkkäinen et al. 2004). Existing solutions to track shipments during transportation in real-time are based on container or swap trailer tracking. For instance, DB Schenker have fitted their 4,100 swap trailers with a GPS positioning system and on request individual shipments can be provided with additional sensors to detect damage or forbidden access and to monitor temperature or other important transport restrictions. Goal of these solutions is to offer a better service to customers by allowing direct access to the transport chains of their shipments (Wirsing 2011).

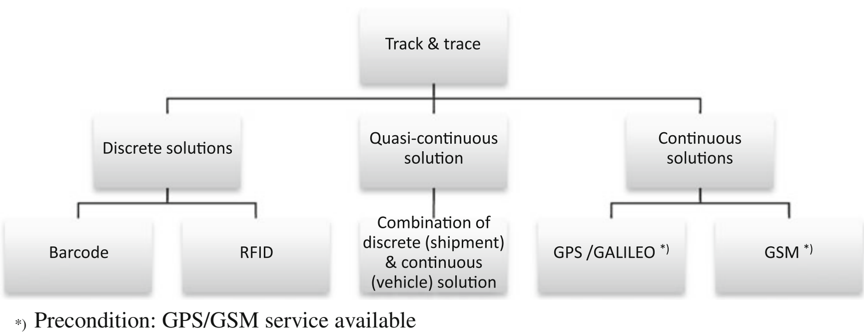


Fig. 1 Methods for track and trace

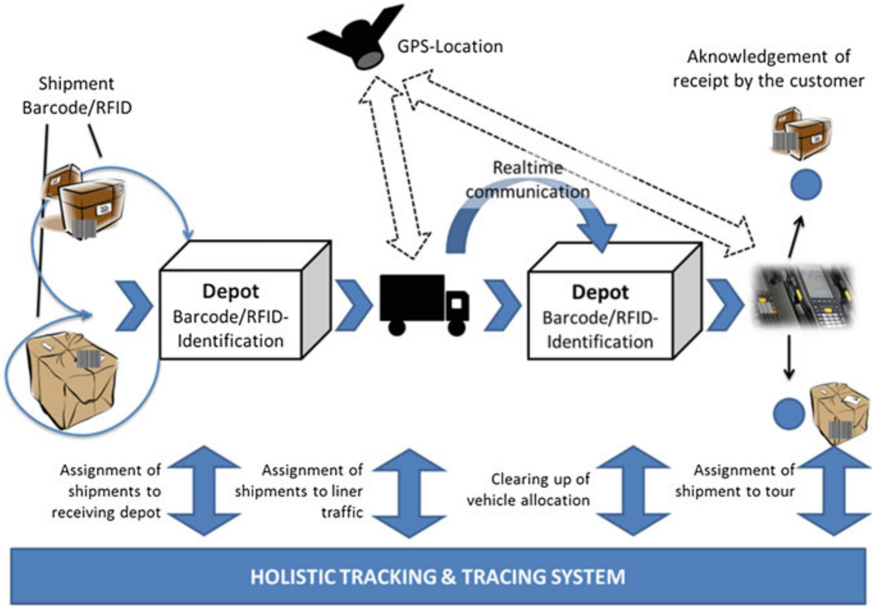


Fig. 2 Holistic quasi-continuous track and trace system

Operational Process Benefits of Continuous Track and Trace

Development of Additional Services

Because logistics products—in fact services—are characterized by a high degree of homogeneity, logistics service providers have to distinguish themselves from market competitors to secure their existence (Hülsmann et al. 2011). Therefore new services have to be generated and offered to the customers, e.g. based on a continuous and holistic track and trace system.

Customers demand safer and more reliable transport chains. Especially for high value or dangerous goods, tracing is important to safeguard the general objective of logistics: The seven Rights (Batarlienė 2007). This supply chain security is a major request by customers because costs of lost goods are only a small part. Further costs appear because of substitute orders, production down-time or penalties for late deliveries (Urciuoli 2010). Nowadays logistics (tracking) solutions can only identify actual status of the shipment. But with an implementation of a continuous track and trace solution the customer could get much more influence on the transport process. Figure 2 shows a possibility to realize a quasi-continuous solution within a groupage or parcel network.

In pre-carriage and on last-mile GPS handhelds with barcode scanners or RFID readers can be used and within the main run the truck telematics must provide an interface for transmitting shipment information (Blümel et al. 2008). For customers, several advantages could be realized because of two main attributes of such a solution: On the one hand real-time information allows customers to influence the transportation process directly; on the other hand information can be used automatically within supply chain or production control software to react autonomously after critical events as well as for performance measurement issues. This mainly results in more process transparency for logistics service providers. Additionally, confidence of loaders increases because they are able to follow the transport process and to safeguard delivery. So the customer business value increases and maybe results in follow-up orders. On top, some critical shipments can be fitted with additional sensors to control for example horizontal position, temperature, vibration during transportation or forbidden access (Wirsing 2011).

Improvement of Performance Measurement

Another major advantage is the possibility of automatic generation of key performance indicators for performance measurement for controlling issues. Nowadays, a continuous performance measurement based on shipments is not common for logistics industry. But studies show that a continuous visibility of shipments within the supply chain increases the delivery performance and lowers logistics costs (Aberdeen Group 2006; Goel 2010). So the potential of controlling, especially for process optimization, is not sufficiently used in logistics (Krupp et al. 2009). One example is the evaluation of drivers and target agreements of driving performance in connection with incentives (Ogle et al. 2002). A post cost calculation, an automatic generation of service-levels or the evaluation of sub-contractors are possibilities to increase performance measurement by data generation based on continuous tracking information of individual shipments or cargo telematics in general. The sales department can use tracking data and transport performance for negotiations as well.

Also for green logistics indicators, especially greenhouse gas (GHG) emission data, continuous track and trace data can be used for calculation. The release of the draft version of the European standard dealing with *methodology for calculation and declaration on energy consumptions and GHG emissions in transport services* strengthen the pressure on logistics service providers to calculate and publish their GHG emissions (DIN EN 16258 2011). Modern telematics systems measure all data which is necessary to calculate truck based emissions (Mikulski and Kwasny 2010; Hague 2008). This is comprehensible because the amount of GHG only depends on fuel consumption (Kranke 2009). By identifying transport sections of specific vehicles operations systems, truck based emissions can be allocated automatically on shipments using computer aided calculation when pick-up and delivery is known and the position saved (Wick et al. 2011; Wick and Klumpp 2010; Noland et al. 2004). This automatic measurement is the first step to reduce emissions and to install a strategic green sector in the performance measurement as well as in the whole company (Piotrowicz 2011). This allows logistics service providers to present themselves as first-movers in the strategic field of sustainable logistics and to realize a competitive edge in the logistics industry (Léonardi and Baumgartner 2004).

Supporting Planning and Scheduling

Because of data generation by the operating system for a continuous visibility of transport routes, transport or production planners get a better and more reliable support for their daily decisions and planning challenges. For example, arising disruptions within transportation processes could be identified instantly, based on a supply chain event management, so that date of receiving this information is earlier. This causes a longer time window to respond to this disruption with a counteraction. This is possible because of a more fluent and faster flow of information (Hülsmann et al. 2011; Goel 2010; Reclus and Drouard 2009; Schmid and Brockmann 2006). On top, the reasons for disruptions could be analyzed in detail which allows process changes in general, e.g. avoiding specific routes during defined time windows. Another advantage is the early possibility of calculating estimated times of arrival of different shipments. By this, scheduling of the subsequent process task after the transport process, which could be for example an allocation of shipments on different tours or a batch of jobs on a machine, can be executed during the transportation simultaneously and dynamically. The overall scheduling result is therefore more improved because there is a longer planning time available (Kandel et al. 2011).

Many existing models of scheduling research assume a static environment and do not consider real-time information as well as the cooperation and interaction of production and transportation (Cowling and Johansson 2002). In fact production planning and control develops from a historical intra-company to a nowadays complex inter-company challenge (Schuh et al. 2008). Therefore, logistics has to be considered holistically within planning and optimization processes. Because of the growth of smart loading equipment and intelligent cargo it can be assumed that continuous tracking information of nearly all transport processes are available in the future (Urciuoli 2010; Blümel et al. 2008). Based on this, the idea of supply chain scheduling arises. The aim of supply chain scheduling is the combination of production planning and scheduling with the operational execution of transportation (Scholz-Reiter et al. 2011; Herrmann 2010; Chen and Vairaktarakis 2005; Hall and Potts 2003). A future research goal must be the development of a cooperative production and transport planning approach within a supply chain based on real-time tracking information to reduce communication problems und to increase efficiency, especially at supply chain interfaces.

In addition, data basis which is generated by daily application of a continuous tracking system can be used for future planning tasks for network design or calculation of warehouses because the transportation data is not only saved in quantitative data, also visualization of transport streams, network bottlenecks or other useful presentation and analysis can be done based on the tracking data (Schmid and Brockmann 2006).

Learning Capabilities and Corporate Knowledge Management

In general the question of learning and knowledge management is an increasing interest in business research as well as management in logistics (Warden et al. 2011; Whitchurch and Gordon 2011; King 2009). The following general trends in education and learning management also apply to the transport sector and are the background for this research contribution about the new prospective 'marriage' between new information and transparency concepts (RFID, GPS, telematics) and knowledge management in logistics:

- Basically the overall trend towards a knowledge-based society and economy drives many changes as for example the need for more and life-long learning among the adult workforce (Mark et al. 2006; Bleiklie 2005).
- This also implies that knowledge not anymore is only created in dedicated institutions such as universities and research institutions and afterwards transferred into business practice ('mode 1') but today also is created in different contexts such as business practice and project networks ('mode 2', Gibbons et al. 1994)—which brings about many changes towards learning and knowledge management in companies and other institutions outside the formal education sector (for example Klumpp and Zelewski 2009; Teichler 1998).
- Furthermore the trend in education management towards learning outcome orientation also brings about a new accountability for knowledge management endeavors—indicating e.g. that specified operative advantages of such concepts have to be proved (Klumpp 2009; European Commission 2008; Landoni and Verganti 2006).
- Furthermore the concept of intrinsic motivation (Abouserie 1995) as well as explicit versus tacit knowledge has been developed in order to explain and manage practical organizational learning environments (Smith 2001).

All these trends also have implications for logistics processes, especially as they also are increasingly knowledge-intensive—i.e. through new technologies as RFID or GPS (Münchow et al. 2011). First drafts for the transfer of these general knowledge management trends into the logistics sector exist, for example for the learning structure of qualifications frameworks (Bioly et al. 2010), the implementation of competence evaluations according to Berufswertigkeit concept in logistics or even a draft for a logistics industry qualifications framework to support continuing education and learning especially in small and medium-sized businesses in the logistics sector (Klumpp et al. 2010).

For example especially for logistic service providers who execute pre-carriage and last-mile transport processes the use of modern mobile GPS telematics modules as described above offers a large potential for personal learning and corporate knowledge management. For example drivers could get more flexibility according their staff assignment on different tours: The tracing of past routes offers the possibility to navigate drivers on routes which were already successful used by drivers in the past who have local knowledge of the region. In addition actual telematics modules could be rigged with a digital camera so that pictures of the shipping address could be saved and displayed on the navigation screen. With this visualization drivers save much time when delivering shipments at unknown addresses because they would not have to search for loading ramps or stock receipts. The individual—tacit—knowledge of each driver could be accessed through this integration of data by any other employee (Ahrens 2008; Hennig 2008). So the clustering of delivery or pick-up tours gains in flexibility.

These applications would require that a strategic knowledge environment is established in a company as previous research showed that different levels of tacit and explicit knowledge exist in relation to the overall learning motivation and setup of different logistics companies (Abidi et al. 2011). For example the following context of the interaction between organizational learning environment and individual learning motivation could be applied to this question in the wake of new telematics concepts (Fig. 3).

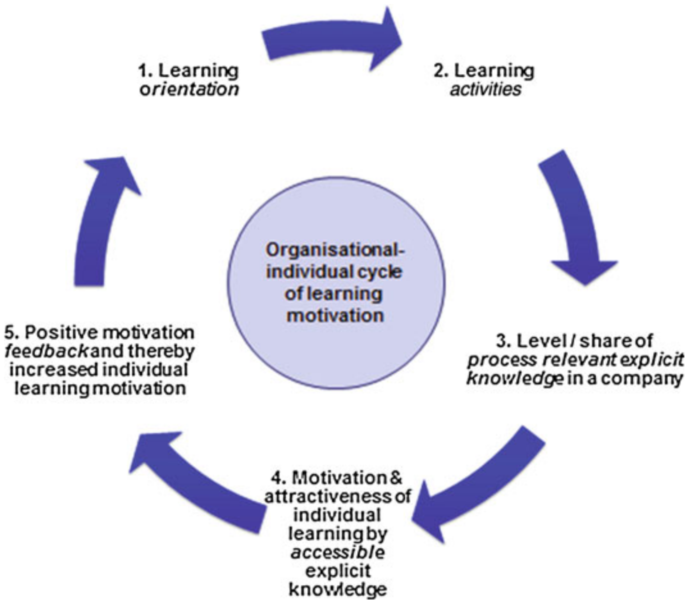


Fig. 3 Organizational and individual interaction in motivation for knowledge management

	(a) Value added services	**(b) Carbon performance evaluation**	**(c) Dynamic scheduling**
(1) Learning orientation	GPS-based information system with referenced customer information	System with support functions for learning about transport carbon implications	GPS-based information system with automated re-scheduling options
(2) Learning activities	Group-based feedback sessions for delivery tours to document information	Workshops and presentations for customers and truck drivers	Testing of re-scheduling options in practice and learning about implications
(3) Level of explicit knowledge	GPS-based routing system using data	Web-based carbon information system	Documentation of suggestions & results
(4) Motivation for individual learning	Easy-to-use handheld providing individualized customer info	Responses to sustainability requests e.g. from customers	Using a case registry for scheduling decisions
(5) Motivation feedback	Customer satisfaction (“thanks”) for service	Innovation image by knowledge transfer	Increased production efficiency, flexibility

Fig. 4 Exemplified telematics application knowledge management environments

In order to outline this strategic draft the described steps in the knowledge motivation and acquisition circle are retraced for the three described application areas for telematics in logistics according to the Fig. 4. Therein three specific examples for a generalized knowledge management system with the five drafted steps is outlined (theoretical) in order to flesh out the basic principles of the dynamic interaction of telematics and knowledge management and their synergies.

Conclusion

This research contribution showed the emerging trend towards telematics applications in the logistics sector and the expected changes derived from that in the example fields of value added services, performance measurement i.e. carbon accounting and production scheduling. The innovative approach of the paper consists out of the transfer and matching of these telematics trends towards education and knowledge management trends. The example described concepts of outcome orientation as well as tacit and explicit knowledge learning environments by addressing the three described telematics applications and provided a lively picture of future strategic knowledge management concepts in logistics—this model could be used as a blueprint for further strategic learning concepts for new technologies.

Further research would possibly have to address the questions of efficiency and acceptance of such concepts as well as the connection to traditional technology and application seminars e.g. provided by software companies. Furthermore suggested research questions could be the intercultural dimension and feasibility of such learning concepts or the global integration in company networks.

Acknowledgments This research is connected to the national excellence research cluster *LogistikRuhr* (www.effizienzcluster.de), funded by the German Federal Ministry of Education and Research (BMBF) in the project funding line 01|C10L19D.

References

Abidi H, Klumpp M, Keuschen T (2011) Industry qualifications framework logistics: explicit and tacit knowledge and qualifications in the logistics industry. In: Blecker T, Jahn C, Kersten W (eds) Maritime logistics in the global economy, current trends and approaches. Eul, Lohmar, pp 327–341

Aberdeen Group (2006) The supply chain visibility roadmap—Moving from vision to true business value. Aberdeen Group, Boston

Abouserie R (1995) Self-Esteem and achievement motivation as determinants of students' approaches to studying. Stud High Educ 20(1):19–26

Ahrens D (2008) Jenseits des Mythos vom "gläsernen Fahrer": Die Rolle der Telematik im Transportprozess. In: Funken C, Schulz-Schaeffer I (eds) Digitalisierung der Arbeitswelt, Zur Neuordnung formaler und informeller Prozesse in Unternehmen. VS, Wiesbaden, pp 69–91

Batarlienė N (2007) Implementation of advanced technologies and other means in dangerous freight transportation. Transport 22:290–295

Bioly S, Keuschen T, Klumpp M (2010) Radio frequency identification (RFID) qualification in Logistics. In: Grubbström RW, Hinterhuber HH (eds) 16th International working seminar on production economics, conference proceedings, Innsbruck, vol 3, pp 25–36, 1–5 March 2010

Bleiklie I (2005) Organizing higher education in a knowledge society. High Educ 49(1):31–59

Blümel E, Boevé W, Recagno V, Schilk G (2008) Ship, port and supply chain security concepts interlinking maritime with hinterland transport chains. WMU J Marit Aff 7:205–225

Brewer A, Sloan N, Landers T (1999) Intelligent tracking in manufacturing. J Intell Manuf 10:245–250

Carlino M, Confessore G, Liotta G (2009) Potentials of satellite-based traceability services in logistics. In: Proceedings of the European navigation conference—global navigation satellite systems 2009, Naples, Italy, pp 1–6, 3–6 May 2009

Chen Z-L, Vairaktarakis GL (2005) Integrated scheduling of production and distribution operations. Manage Sci 51:614–628

Cowling P, Johansson M (2002) Using real time information for effective dynamic scheduling. Eur J Oper Res 139:230–244

DIN EN 16258 (2011) Methodology for calculation and declaration on energy consumptions and GHG emissions in transport services (good and passenger transport). Beuth, Berlin

European Commission (2008) Empfehlung des Europäischen Parlaments und des Rates vom 23. April 2008 zur Errichtung eines Europäischen Qualifikationsrahmens für Lebenslanges Lernen. European Commission, Brüssel

Gibbons M, Limoges C, Nowotny H, Schwartzmann S, Scott P, Trow M (1994) The new production of knowledge: the dynamics of science and research in contemporary societies. Sage, London

Goel A (2010) The value of in-transit visibility for supply chains with multiple modes of transport. Int J Logist Res Appl 13:475–492

Hague M (2008) Taking steps to reduce your carbon footprint. Logist Transp Focus 10(9):18–20

Hall NG, Potts CN (2003) Supply chain scheduling: batching and delivery. Oper Res 51:566–584

He W, Tan EL, Lee EW, Li TY (2009) A solution for integrated track and trace in supply chain based on RFID & GPS. In: Proceedings of the IEEE conference on emerging technologies & factory automation, Mallorca, Spain, pp 1–6, 22–25 Sept 2009

Hennig U (2008) Satellitengestützte Sendungsverfolgung, Vorteile der GPS-Technologie für Kurier–, Paket und Expressdienste. FM Das Logistik Magazin 12/2008, pp 50–51

Herrmann J (2010) Supply chain scheduling. Gabler, Wiesbaden

Hülsmann M, Schwientek A, Korsmeier B, Austerschulte L (2011) Creating customer value in logistics: contributions and limitations of autonomous cooperation-based technologies. In: Hülsmann M, Scholz-Reiter B, Windt K (eds) Autonomous cooperation and control in logistics. Springer, Berlin, pp 15–25

Kärkkäinen M, Ala-Risku T, Främling K (2004) Efficient tracking for short-term multi-company networks. Int J Phys Distrib Logist Manage 34:545–564

Kandel C, Klumpp M, Keusgen T (2011) GPS based track and trace for transparent and sustainable global supply chains. In: Proceedings of the 17th international conference on concurrent enterprising, Aachen, Germany, pp 252–259, 20–22 June 2011

King RW (2009) Knowledge management and organizational learning. Springer, Dordrecht

Klumpp M (2009) Berufswertigkeit als Konzept zur Evaluation informellen Lernens und Wissens. In: Helmstädter HG, Tippe U (eds) Durchlässigkeit und Anrechnung im Hochschulalltag, Dem lebenslangen Lernen Türen öffnen, Schriftenreihe zu Fernstudium und Weiterbildung, Band 4. Fachhochschule Brandenburg, Brandenburg, pp 45–62

Klumpp M, Kowalski M, Zelewski S (2011) Beurteilung der Wirtschaftlichkeit von Gütertransporten mit der Eisenbahn. In: Zelewski S, Jene S (eds) Kooperationen zwischen Eisenbahnverkehrsunternehmen, Grundlagen—Konzepte—Praxisanwendungen. Logos, Berlin, pp 541–586

Klumpp M, Peisert R, Keuschen T (2010) Sectoral qualifications framework logistics: what can we learn from Berufswertigkeit and Graduate surveys? In: Blecker T, Kersten W, Lüthje C (eds) Pioneering solutions in supply chain management—a comprehensive insight into current management approaches. Schmidt, Berlin, pp 279–292

Klumpp M, Zelewski S (2009) Moderne Wissensbasierte Kompetenzanforderungen an Führungskräfte nach dem Konzept Berufswertigkeit und resultierenden Eigenschaften wissenschaftlicher Weiterbildung. In: Beyersdorf M, Christmann B (eds) DGWF Jahrestagung 2008: Strukturwandel der Arbeit—Zukunft der wissenschaftlichen Weiterbildung. DGWF, Hamburg, pp 85–97

Kranke A (2009) CO2-Berechnung—Wie Sie die CO2-Emissionen bei LKW-Transporten konkret berechnen können. LOGISTIK inside 05/2009, pp 47–49

Krupp T, Bohlmann B, Stockinger D, Bauer C (2009) Controlling bei Logistikdienstleistern—Aktueller Stand und zukünftige Herausforderungen. Controlling 21:464–471

Landoni P, Verganti R (2006) Fostering knowledge and technology transfer through evaluation systems at a regional level. Int J Technol Transf Commercialisation 5(4):355–372

Léonardi J, Baumgartner M (2004) CO_2 efficiency in road freight transportation: status quo, measures and potential. Transp Res Part D 9:451–464

Mark R, Pouget M, Thomas E (eds) (2006) Adults in higher education—learning from experience in the New Europe. Lang, Bern

Meers S, Hennes G, Nyhuis P (2010) Logistische Herausforderungen in Produktionsnetzen. Zeitschrift für wirtschaftlichen Fabrikbetrieb 105:949–952

Mikulski J, Kwasny A (2010) Role of telematics in reducing the negativ environmental impact of transport. In: Mikulski J (ed) Transport system telematics. communications in computer and information science. Springer, New York, pp 11–29

Münchow A, Kowalski M, Zelewski S (2011) Modellierung RFID-gestützter wissensintensiver Prozesse. Prod Manage 16(4):26–29

Noland RB, Ochieng WY, Quddus MA, North RJ, Polak JW (2004) The vehicle emissions and performance monitoring system: analysis of tailpipe emissions and vehicle performance. Transp Plan Technol 27:431–447

Ogle J, Guensler R, Bachman W, Koutsak M, Wolf J (2002) Accuracy of global positioning system for determining driver performance parameters. Transp Res Rec 1818:12–24

Piotrowicz W (2011) Monitoring performance. In: Cetinkaya B, Cuthbertson R, Ewer G, Klaas-Wissing T, Piotrowicz W, Tyssen C (eds) Sustainable supply chain management. Springer, Berlin, pp 57–80

Reclus F, Drouard K (2009) Geofencing for Fleet & Freight management. In: Proceedings of the 9th international conference on intelligent transport systems telecommunications, Lille, France, pp 353–356, 20–22 Oct 2009

Schenk M, Richter K (2007) Telematik und RFID—Elektronische Beobachter gestalten die sichere Warenkette. In: Bullinger H-J, ten Hompel M (eds) Internet der Dinge. Springer, Berlin, pp 77–89

Schmid S, Brockmann C (2006) Marktübersicht: Tracking & Tracing Lösungen In Der Logistik. PPS Managament 11(2):52–61

Schöch R, Hillbrand C (2006) Ein integrierter Ansatz für diskrete und stetige Sendungsverfolgung auf Stückgutebene mittels RFID und GSM. In: Mannfeld DC, Suhl L (eds) Informationssysteme in Transport und Verkehr. DS&OR Lab, Paderborn, pp 89–102

Scholz-Reiter B, Novaes AGN, Makuschewitz T, Frazzon EM (2011) Dynamic scheduling of production and inter-facilities logistic system. In: Haasis H-D, Kreowski H-J, Scholz-Reiter B (eds) Dynamics in logistics. Springer, Berlin, pp 443–453

Schuh G, Stich V, Schmidt C (2008) Produktionsplanung und -steuerung in Logistiknetzwerken. In: Nyhuis P (ed) Beiträge zu einer Theorie der Logistik. Springer, Berlin, pp 249–273

Smith EA (2001) The role of tacit and explicit knowledge in the workplace. J Knowl Manage 5(4):311–321

Teichler U (1998) The changing roles of the university and the non-university sectors. Eur Rev 6(4):475–487

Urciuoli L (2010) Supply chain security—mitigation measures and a logistics multi-layered framework. J Transp Secur 3:1–28

Vahrenkamp R (2007) Logistik, Management und Strategien. 6., überarbeitete und erweiterte Auflage. Oldenbourg, München

Wang Y, Potter A (2008) The application of real time tracking technologies in freight transport. In: Proceedings of the 3rd international IEEE conference on signal-image. Shanghai, China, pp 298–304

Warden T, Porzel R, Gehrke JD, Langer H, Herzog O, Malaka R (2011) Knowledge management for agent-based control under temporal bounds. In: Hülsmann M, Scholz-Reiter B, Windt K (eds) Autonomous cooperation and control in logistics. Springer, Berlin, pp 229–246

Whitchurch C, Gordon G (2011) Some implications of a diversifying workforce for governance and management. Tert Educ Manag 17(1):65–77

Wick C, Klumpp M (2010) Logistics carbon footprinting in practice. In: Blecker T, Kersten W, Lüthje C (eds) Innovative process optimization methods in logistics—emerging trends, Concepts and Technologies. ESV, Berlin, pp 41–62

Wick C, Klumpp M, Kandel C (2011) Shapley value simulation for allocating GHG emission savings due to logistics pooling within ECR cooperations. In: Proceedings of the 2011 European simulation and modelling conference, Guimaraes, Portugal, pp 240–244, 24–26 Oct 2011

Wirsing E (2011) Sichere Lieferketten durch Real-Time-Monitoring bei Lagerung, Transport und Umschlag. DB Schenker. Presentation at the 5th FOM Forum Logistik, 26 Oct 2011. Duisburg, Germany. http://www.fom.de/download/519-Wirsing_-_Schenker.pdf. Accessed 15 Nov 2011

Part III
Production Logistics and Job Shop Scheduling

Towards an Integrated Production and Outbound Distribution Planning Method

Bernd Scholz-Reiter and Christian Meinecke

Abstract The competition on global markets has changed towards a competition between supply chains instead of single competing companies. Thus, optimization approaches like integrated planning of production and outbound distribution comprise multiple companies within a supply chain. This paper provides requirements towards an integrated planning method and highlights the differents with existing integrated planning approaches. Additionally, this contribution presents a concept for a new integrated planning and outbound distribution planning method. This planning method provides an improved integration of the subproblems production and transportation planning. The paper also presents the operating principle of the new integrated planning method and identifies relevant research questions for further development.

Introduction

Companies are facing many challenges in today's global markets like frequently changing products as well as a decreasing delivery times and an ongoing individualization of products (Wiendahl and Lutz 2002). As a consequence, companies focus on their core competencies and outsourcing other areas towards other companies and service providers. Thus, the supply chains length increases and relationships with new companies have to be established. These relationships and interfaces cause an additional coordination and collaboration effort (Arnold et al. 2008). Nevertheless, companies have to collaborate closely in order to stay

B. Scholz-Reiter · C. Meinecke (✉)
BIBA—Bremer Institut für Produktion und Logistik at the University of Bremen, Hochschulring 20, 28359 Bremen, Germany
e-mail: mei@biba.uni-bremen.de

H.-J. Kreowski et al. (eds.), *Dynamics in Logistics*, Lecture Notes in Logistics,
DOI: 10.1007/978-3-642-35966-8_7,

competitive or to improve the competitiveness of the entire supply chain. As a result there will be competition between supply chains instead of a competition between single companies.

The Supply Chain Management (SCM) concept provides numerous tools to tackle these challenges. The goal of SCM is a holistic design, planning and scheduling of the flow of goods and information in the entire supply chain (Arnold et al. 2008). Traditionally, process optimization focuses on improvements of single planning and scheduling problems like production and transport. In contrast to this, SCM requires an integrated approach across involved companies (Baumgarten and Darkow 2004). All planning activities have to concentrate on the entire supply chain. Thus, single planning problems of supply chain partners have to be combined in one integrated planning problem. Thereby, requirements and restrictions of different companies are an input in one optimization problem and can be considered simultaneously (Sarmiento and Nagi 1999; Chen 2010). By applying the integrated planning approach production and logistics processes can be linked seamlessly with one another in the entire supply chain. This represents the goal of integrated planning (Tan 2001).

A supply chain comprises legally independent companies with different responsibilities and conflictive goals. In practice each single company represents a profit centre and carries out its own process and order planning and scheduling. The results of this separated planning activities are in accordance with individual targets of the corresponding company. It is obvious, that this kind of separated and sequential planning is not optimal for the performance of the entire supply chain (Chandra and Fisher 1994; Chen and Vairaktarakis 2005; Pundoor and Chen 2005). An integrated planning approach provides an increased supply chain performance and competitiveness by coordination of the single planning problems. An intensive exchange of relevant information between involved supply chain partners is essential for a coordinated planning approach (Stadtler 2005).

In this paper we focus on development of a concept for a planning method to solve the integrated production and outbound distribution planning problem. Therefore this contribution is organized as follows. Section Literature Review provides an overview of existing approaches in the field of integrated production and outbound distribution planning methods and presents the motivation for the development of an integrated planning concept. This concept will be described in Section Concept for a New Integrated Planning Method as well as relevant research questions identified. We finish with a conclusion.

Literature Review

Integrated production and outbound distribution planning is relevant in many industrial applications (Geismar et al. 2008). In the literature there is a number of terms like e.g. integrated production and outbound distribution scheduling problem (IPODS) or production distribution problem (PDP) dealing with the integration of production and distribution planning (Chen 2010). The scope of consideration and the level of detail differs in all existing approaches but covers the field of integrated production and distribution planning.

Sarmiento and Nagi reviewed existing integrated planning approaches on strategic and tactical level and give a categorization regarding the scope of consideration and its combinations (e.g. production/distribution/inventory) (Sarmiento and Nagi 1999). Furthermore, there is an additional categorization regarding the planning horizon (one period vs. multi period). The authors state, that dynamic effects like a variation in demand or machine failure are not considered in existing approaches because of the high complexity. Furthermore the authors summarize, that many existing approaches were developed for specific scenarios or are on a high abstraction level. This prevents a direct comparison of existing approaches and a general application in practice.

Chen reviews integrated planning approaches only on the operational level (Chen 2004). Existing approaches considered had been categorized regarding the scope of consideration as well (e.g. in-house transportation/production/outbound transportation). Furthermore, there is a categorization considering planning horizon, delivery type and variation in demand as well as economic indicators like transport and inventory costs (Chen 2010). The author states, that dynamic effects, transport capacities and transport times have been neglected in existing planning approaches on the operational level so far. The author figures out, that existing approaches considered focus on specific scenarios or are on a very generic level. Thus, the development of fast and robust algorithms for practical applications is a future research field.

Both production and outbound distribution planning are two NP-hard problems. Due to complexity of the integrated planning problem many approaches use a decomposition method first in order to separate the integrated problem in two subproblems. After this production planning and transport planning are considered separately and sequentially using a push or a pull principle. This means that the planning process is production or transport oriented (Hermes et al. 2009). Afterwards, they combine these two results of the sub-problems in an integration step. Therefore, sorting algorithms (e.g. Chen and Vairaktarakis 2005) as well as priority rules and searching algorithms (e.g. Chandra and Fisher 1994) are used. The goal of this integration step is to put similar orders together in order to reduce transportation costs or setup times. In fact, it is not sure if there a possibility to combine similar orders. In the worst case the planning result after integration step is equal to the planning result after finishing sequential planning. Nevertheless, integrated planning methods using this sequential approach provide better results regarding defined indicators then traditional sequential planning approaches (Chandra and Fisher 1994; Chen and Vairaktarakis 2005; Pundoor and Chen 2005).

There is another group of integrated planning methods beside this sequential approaches to solve the integrated problem. These recently developed planning methods are metaheuristics applying an improvement procedure on a complete solution for a simultaneous modification of production and transportation plan in order to improve the total costs. For example, Boudia and Prins present a memetic algorithm with a dynamic population management using a genetic algorithm and a local search method to improve an initial solution by allowing elementary changes (moves) (Boudia and Prins 2009). Each single solution comprises a tuple of lists, matrices and vectors representing e. g. delivery quantities and trips. The authors emphasise, that the algorithm developed works on complete solutions instead of considering the single sub-problems. Archetti et al. modify the distribution plan using a remove and reinsert method in the improvement procedure and determine the corresponding production plan in order to minimize the total costs (Archetti et al. 2011).

Nevertheless, the achievement of a higher integration could result in an increased acceptance of the integrated planning result by the involved supply chain partners in comparison with traditional sequential planning approaches. Furthermore, an increased integration could lead to an improved integrated planning result regarding defined indicators. Thus, this research is motivated to fill the gap identified by development of an integrated planning approach for the operational level considering capacity constraints in production and distribution as well as dynamics and economic aspects. This research is not about production planning and scheduling or transportation planning methods but on the integration of this separate planning problems. A concept of this integrated planning method is presented in the following paper.

Concept for a New Integrated Planning Method

An increased integration level is an important goal for the development of a new integrated planning method. The aim is to ensure an improved coordination of the single planning problems to achieve an increased supply chain performance as well as an increased competitiveness and an improved acceptance of the integrated planning result within the supply chain. In order to improve the integration we first cluster already known orders using a backward scheduling and a selection criteria like the delivery date. For the backward scheduling the already known production time and the transport time of the orders are used. By using e.g. a Bayes classificator we receive a number of clusters as a first result. Each cluster contains orders with a defined range of remaining time until delivery date and thus with a different priority. This range depends on the cluster criteria, which has to be defined previously.

After this clustering step we start a planning step with the cluster having the highest priority and containing all urgent orders with the lowest remaining time until delivery date. The planning step comprises a separate planning of production and transport using already known planning methods from the respective fields. For production planning dispatching rules or shifting-bottleneck-heuristics are applicable. The sweep-heuristics or nearest-neighbour-rule for example could be used for generating transport plans. At the end of the planning step there is a separated and initial production plan and a transportation plan for all orders of the first cluster and we go on with an optimization step.

In the optimization step we combine the two initial planning results. Therefore, sorting and evolutionary algorithms as well as permutation based rules and local search algorithms could be used. The aim is to put similar orders together and to reduce transportation costs or setup times. The optimization step is equal to the integration step of existing integrated planning methods. At the end of the optimization step a so called cluster plan is available.

After the optimization step of the first cluster an initial integrated production and transport plan is available. It can be fixed for this cluster because of the highest order priority. The further successive proceeding of the planning and the optimization step for the cluster with a lower priority is the same as described for the first cluster. After finishing the optimization step of each cluster the integration of planning results into the integrated plan is necessary. In the integrated plan, which contains fixed slots form clusters with higher priorities, unused production and transportation slots are available. These slots can be used for insertation of the cluster plan and completion of the integrated plan. The procedure ends after insertation of cluster plan n in the integrated production and transport plan. By this way the integrated plan will be completed step by step based on the results of previous cluster and integration plans. The operating principle of the described integrated production and outbound distribution planning method is depicted in Fig. 1.

In each cluster we combine production and transportation issues and restrictions and use these result as a basis for the integration of the subsequent cluster. Thus, the integration level of this conceptual integrated method is on a distinct higher level than the integration level of existing planning methods.

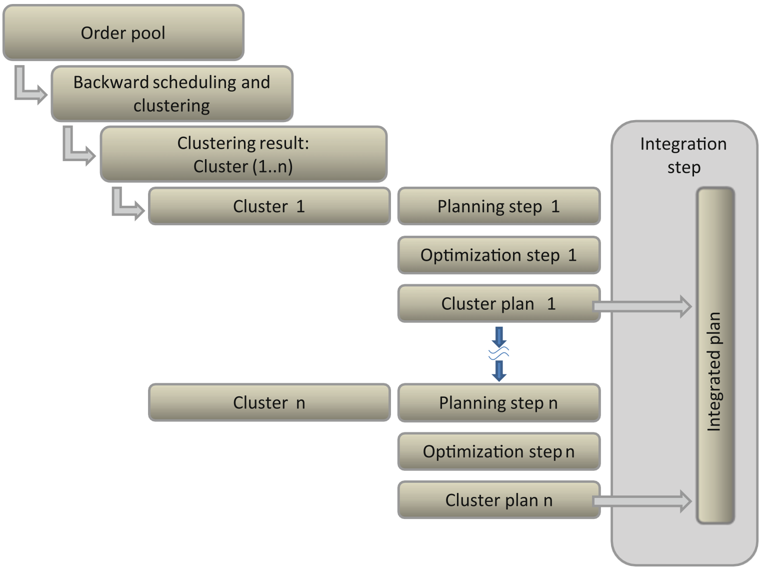


Fig. 1 Operating principle of the integrated planning method

In order to transfer this concept into a detailed integrated planning method which is applicable in practical problems we identified a number of relevant research questions. The first focus of our future investigations is on clustering. The goal is to identify appropriate clustering methods and to define cluster boundaries depending on the order structure and additional production and transportation restrictions. A large number of cluster reduces the problem complexity within one cluster and thus the CPU time but there is also a reduction of the optimization potential.

The formulation of an objective function for the optimization step is the second goal of our future research. This objective function is necessary to decide if a feasible improvement in the optimization step will have a positive effect for the cluster plan and for the target indicators of the integrated plan as well. The design of the objective function represents planning assumptions in the supply chain. Therefore, weighting factors as well as economic and process indicators have to be defined and included. The most important focus of future investigations is the development respectively the selection of methods which will be used for the planning and optimization steps.

For evaluation the developed integrated method will be implemented in existing models from the literature and additionally in a model of a use case deduced from a practical use case. The model from the literature will be used to check the functionality. Furthermore, there will be a comparison with existing integrated planning methods and with production or transport oriented sequential planning processes. Especially transport oriented planning methods as presented in Scholz-Reiter et al. (2008, 2009, 2010) are relevant for the comparison of process indicators because of their orientation themselves on fulfillment of customer requirements and indicators.

Conclusion

In this paper we presented a concept for an innovative integrated planning method for the production and outbound distribution problem. As a result of a literature review it can be stated that existing integrated planning methods focus on specific scenarios or are on a very generic level with a low integration of respective sub-problems. This was the motivation for the development of the provided concept for an integrated planning method with a distinct higher integration. We presented the operating principles as a concept for an integrated planning method, which will be detailed in the future. A number of relevant research questions were identified and described. This approach is promising because of the high integration and its probably higher acceptance at the planning departments of the supply chain partners.

References

Archetti C, Bertazzi L, Paletta G, Speranza M (2011) Analysis of the maximum level policy in a production-distribution system. Comput Oper Res 38:1731–1746

Arnold D, Isermann H, Kuhn A, Tempelmeier G, Furmans K (2008) Handbuch Logistik. Springer, Berlin

Baumgarten H, Darkow L (2004) Konzepte im Supply Chain Management. In: Busch A, Dangelmaier W (eds) Integriertes Supply Chain Management—Theorie und Praxis effektiver unternehmensübergreifender Geschäftsprozesse, 2nd edn. Gabler, Wiesbaden

Boudia M, Prins C (2009) A memetic algorithm with dynamic population management for an integrated production-distribution problem. Eur J Oper Res 195:703–715

Chandra P, Fisher M (1994) Coordination of production and distribution planning. Eur J Oper Res 72:503–517

Chen Z (2004) Integrated production and distribution operations: taxonomy, models and review. In: Simchi-Levi D, Wu S, Shen Z (eds) Handbook of quantitative supply chain analysis: modelling in the e-business era. Kluwer Academic Publishers, Dordrecht

Chen Z (2010) Integrated production and outbound distribution scheduling: review and extensions. Oper Res 58:130–148

Chen Z, Vairaktarakis G (2005) Integrated scheduling of production and distribution operations. Manage Sci 51:614–628

Geismar H, Laporte G, Lei L, Sriskandarajah C (2008) The integrated production and transportation scheduling problem for a product with a short lifespan. Informs J Comput 20:21–33

Hermes A, Preuß M, Wagenitz A, Hellingrath B (2009) Integrierte Produktions- und Transportplanung in der Automobilindustrie zur Steigerung der ökologischen Effizienz. http://www.in-ter-trans.eu/12.html. Accessed 21 Dec 2010

Pundoor G, Chen Z (2005) Scheduling a production-distribution system to optimize the tradeoff between delivery tardiness and distribution cost. Naval Res Logist 52:571–589

Sarmiento A, Nagi R (1999) A review of integrated analysis of production-distribution systems. IIE Trans Sched Logist Spec Issue Manuf Logist 31:1061–1074

Scholz-Reiter B, Böse F, Dizdar C, Windt K (2008) Transportorientierte Reihenfolgeplanung. Industrie Management—Zeitschrift für industrielle Geschäftsprozesse 4:15–17

Scholz-Reiter B, Meinecke C, Ruthenbeck C (2009) Development of a cross company planning and scheduling method for the supply chain management of RoRo and ConRo-Ports. In: Laptaned U, Banomyong R (eds) Proceedings of the 1st international conference on logistics and transport. UP Organizer and Publication Co Ltd, Bangkok

Scholz-Reiter B, Meinecke C, Ruthenbeck C (2010) Planungsmethode für die unternehmensübergreifende Auftragsabwicklung. Industrie Management—Zeitschrift für industrielle Geschäftsprozesse 5:39–42

Stadtler H (2005) Supply chain management and advanced planning—basics, overview and challenges. Eur J Oper Res 163:575–588

Tan K (2001) A framework of supply chain management literature. Eur J Purchasing Supply Chain Manag 7:39–41

Wiendahl H, Lutz S (2002) Production in networks. Ann CIRP 51(2):573–586

Supply Chain Integration by Human-Centred Alignment of Lean Production Systems

Yilmaz Uygun and Natalia Straub

Abstract In spite of acceptable achievements of application of lean production in and between many companies, the durable success still falls short of expectations. Even 20 years after the first lean wave, companies could not achieve the same efficiency like Toyota. Regarding the application of lean methods and tools between companies in supply chains the status and quality of application is even worse. This is mainly caused by two aspects. On the one hand the shop floor and especially the effects of employees on the application of lean production are not considered acceptably. On the other hand there is a common dissatisfaction regarding mere technical implementation concepts. Focussing on these aspects facilitates the integration and collaboration in supply chains. According to this, a new model for the holistic implementation of lean production focusing on soft factors is presented in this chapter which enables the effective and sustainable application of lean production in supply chains.

Introduction

Supply chains of industrial companies focus more and more on efficient organisation of their interlinked production processes. Within this context the application of lean production along the value chain gains increasingly in importance. To achieve an optimal effectiveness in the application of lean production systems, it is

Y. Uygun (✉) · N. Straub
TU Dortmund University, Leonhard-Euler-Straße 5, 44227 Dortmund, Germany
e-mail: uygun@LFO.tu-dortmund.de

N. Straub
e-mail: natalia.straub@tu-dortmund.de

H.-J. Kreowski et al. (eds.), *Dynamics in Logistics*, Lecture Notes in Logistics,
DOI: 10.1007/978-3-642-35966-8_8,

on the one hand necessary to connect all organisational measures along the value chain, and on the other hand to build up a common understanding among the companies which are involved in the production process (RIF 2007; Rivera et al. 2007; Faust 2009a). Guided by both the idea of creation of an integrated and inter-organisational lean production system (LPS) and the existence of an improvement possibility in the inbound area, the common and cooperative process optimisation within the inter-organisational implementation of LPS offers considerable optimisation potentials (RIF 2007; Faust 2009b; Hofbauer et al. 2009). But current studies show that existent cooperation of companies with their suppliers within the inter-organisational implementation of LPS are evaluated in the most cases as not very successful mainly because the expected positive developments and enhancements of processes were disappointing (Uygun et al. 2009; RIF 2007).

On this, many possible reasons are discussed in literature, where currently more and more an unsuitable problem solving culture is cited. According to this, the competitive advantage of a company lies not exclusively in the ready-made methods and tools, but in fact in the ability of the employees to perceive a situation and create suitable and smart solutions for this (Rother 2009). It becomes evident that the action fields within the sustainable implementation of LPS in supply chains address rather the soft and human-centred success factors (Uygun et al. 2009).

In this context Nell-Breuning speaks of the rediscovery of the human together with his assumption that the success of an enterprise primarily depends on the ability to create conditions in which the employees can reach personal development by adequate activities. He also forecasts a "Copernican Revolution" for business management (Nell-Breuning 1950) if it desists from looking at the employees merely in an instrumental manner and considers the interaction between the organisation members and the company instead (Drumm 2008). While the enterprise helps people achieve the satisfaction of their demand for self-realization, the employees, for their part, do their utmost to attain the companies' objectives. In this manner unused potentials, which lead an organisation and supply chains as a working team of humans to the optimum success, can be tapped (Hesch 1997).

In the same way, Brödner speaks of a "renaissance of the human work" (Brödner 1993) and Pfeiffer and Weiß speak of a "change of perspectives from tangible asset to human asset", the latter having a paradigmatic nature (Pfeiffer and Weiß 1992). Similarly, Rapaport declares that "management today recognises that to have a winning organisation, it has to be more knowledgeable and competent in dealing with and developing people. That is the most fundamental change" (Rapaport 1993). Therefore, the human becomes a crucial success factor of the lean production system of an enterprise and supply chains, whose biggest challenge is to activate and promote the innovation and creativity potential as well as to activate productivity and employee commitment (Brödner 1993).

Literature Review

In literature there exist numerous concepts which describe the implementation and inter-organisational application of LPS in supply chains and networks, but a model for a holistic and human-based implementation in supply chains does not yet exist. In the followings relevant concepts will be divided according to their way of approach in three main groups: lean production concepts, lean supply chain concepts and supplier development concepts. In this connection the organisational and human-centred aspects of the implementation will be focused.

Lean Production Concepts

The implementation of lean production is described in literature by numerous concepts and roadmaps (Oeltjenbruns 2000; Baumgärtner 2006; Dombrowski et al. 2008). These concepts explain the tasks and the timeline of the company-wide implementation of LPS and provide important knowledge for the project management and the sequence of the planning and introduction activities which are partially also relevant for the inter-organisational implementation. Although in some concepts (Dombrowski et al. 2008) the last stage of the implementation contains elements like supplier development and expansion of the LPS in the network, organisational and substantial hints for the concrete execution are not included.

Regarding the human-centred aspects it becomes evident that the realisation of preparatory qualification measures for lean methods and tools are regarded as necessary in order to bring employees close to the implementation of LPS, but a concretisation of these aspects does not exist there.

Lean Supply Chain Concepts

After reviewing the relevant literature regarding lean supply chain (LSC) it is statable that the implementation concepts of LSC have different perspectives of consideration and concretisation. This is also shown by Anand and Kodali who illustrate a comprehensive literature review regarding the status quo of LSC concepts. This research contents "that the concepts and theory of LSC are not yet fully developed […] and only very few papers that addressed the concept of applying Lean Manufacturing principles to the whole supply chain are available" (Anand and Kodali 2008).

Based on this analysis relevant concepts for the inter-organisational lean implementation will be depicted in the following, like RIF (2007), Kuhn and

Uygun (2008), König (2009), Dolcemascolo (2006), Phelps et al. (2003), Anand and Kodali (2008), Rivera et al. (2007).

In the research project "Lean Production Systems along the Value Chain" by RIF a concept for both a company-wide and an inter-organisational implementation and application of LPS was developed. The focus of the inter-organisational implementation lies in the identification of the network-wide interface-affine lean methods and their effective linkage in the whole network. This concept provides important advice for the structural organisation of the LPS implementation in networks (RIF 2007). Based on these findings Kuhn and Uygun developed a concept for the management of the networked application of LPS elements. Besides the phases which run recursively, the tasks of the network authority is given (Kuhn and Uygun 2008). The detailed procedure of the implementation and important aspects like selection of implementation partners are in both concepts not regarded in detail. The human-centred aspects of the LPS implementation are considered in terms of preparatory qualification measures but the implementation and cross-linkage of lean methods come to the fore.

König developed a concept for the constitution of inter-organisational networks (ION) for the cooperative optimisation of LPS which describes a systematic procedure which range from network initiation to the subsequent network operation. The main idea of the ION is the identification of the strengths and weaknesses of the co-opted network partners in order to enable an exchange for the continuous improvement by well-directed learn arrangements. In this respect the ION is a non-profit organisation without any contractual basis. So, the cooperation is based upon voluntariness and engagement. This concept provides a methodology for the configuration of a knowledge network and describes both the conceptual and organisational design field of a collective improvement of principles and methods of LPS (König 2009). Although human-centred and cultural aspects are not explicitly regarded, this idea of developing an inter-organisational problem solving culture with respect to the LPS optimisation converges partially with the human-centred implementation perspective.

The concept by Dolcemascolo "Improving the Extended Value Stream" illustrates a methodology for the step-by-step expansion of the lean manufacturing concept along the value chain. Through selection of appropriate partners and a systematic implementation of lean principles, improvement of target achievement with respect to efficiency, quality, and time is aimed in the whole network. The methodology contains a systematic approach for the configuration and transformation of existing supply chains in lean supply chains. The human-centred aspects are not regarded explicitly. However, Dolcemascolo defines an eighth waste "underutilisation of employees' mind and ideas". In respect of the whole supply chain this means the suboptimal utilisation of ideas and the creativity of suppliers (Dolcemoscolo 2006). This concept is method-oriented but the organisational aspects of implementation are relatively well explained and provide an important input.

The concept by Anand and Kodali "LEADER"—Conceptual Framework for Lean Supply Chain and Its Implementation" is also a methodology for the step-by-step transformation of an existing supply chain into a lean supply chain. This concept is effected by five lean principles which have to be implemented connectedly in the whole supply chain. The adjustments during the inter-organisational implementation of the lean principles are shown based on the illustration of the differences between lean supply chain and lean manufacturing concepts. This concept is characterised by a systematic presentation of the implementation steps and the integration of necessary tools in the specific phases. But organisational aspects of network-wide implementation are not considered. So, the focus of this concept is a conceptual frame for lean supply chains. The cultural and human-centred aspects of implementation such as "leadership, commitment and involvement of all stakeholders and cultural change across the supply chain" are emphasised, but the authors give no concrete proposals how to realise these changes (Anand and Kodali 2008).

Phelps et al. focus on "Developing Lean Supply Chains—A Guidebook", which describes how lean principles in the whole supply chain are to be implemented through intensified cooperation with the suppliers. It is the aim to identify and eliminate systematically waste at both suppliers and inter-organisational interfaces. Through the extensive illustration of tasks in the particular phases this concept is relatively detailed. The employee-centred aspects are mentioned and the cultural change in the whole supply chain is seen as a necessary support of the implementation. However, the human-centred aspects are not at the centre of consideration and the training measures are method-oriented (Phelps et al. 2003).

Rivera et al. (2007) provide a conceptual frame for lean supply chains which is based upon the concept by Phelps et al. (2003) and extends the latter with one further phase. However, this concept possesses a low level of detail compared to other concepts.

Supplier Development Concepts

The relevant supplier development literature makes contribution to the process of supplier development which is in reality deployed so as to enable the suppliers to implement lean methods self-dependently (RIF 2007). Here, the concepts by Wagner and Boutellier (2003), Handfield et al. (2006), and Delphi according to Liker and Meier (2005) can be stated. In summary, the supplier development concepts give important information regarding configuration and coordination of inter-organisational teamwork between manufacturer and supplier. The concepts are in general derived from the praxis and are based on empirical studies and possess a relatively low level of detail.

Resume

The existing concepts divide the implementation process into different phases which vary in name, content, and duration. Furthermore, the concepts have different levels of detail. The number of phases is affected strongly by the starting and end point of implementation. Regarding the end point of implementation the concepts differ in the extension to other in-company departments and other network partners. Based on the analysis of the phase division, four phases of the network-wide implementation of LPS can be classified in general in which all mentioned phases can be integrated: initialisation, planning, implementation, and evaluation and expansion.

Summing up, the existing implementation concepts aim at the transfer of lean principles and methods to network partners. Here, the concepts do not describe the organisational integration of the network-wide implementation, but are rather conceptual approaches. Moreover, cultural aspects of LPS are mentioned under the term "soft factors", but are neither in praxis nor in theory considered adequately. So, there is a need for action in developing a concept for human-centred LPS implementation in networks which considers both human-centred aspects and network-wide organisational operationalisation.

Human as Key Factor

Focusing on the effects employees have on the application of lean production, central importance is attributed to the company's underlying image of human (Korge 2003a). Hence, the image of human as a rational-economic being, which has been dominant for almost one century, still shapes the actions and behaviour patterns of a huge number of companies. This image of human is contrasted in the following with the idea of human of lean production systems (Füser 2001; Hesch 1997).

The scientific management of Taylorism regards the company as an economic-technical system (Hansmann 1992; Stürzel 1992). The aim of the company is to reduce requirements for human labor at the operation level to a minimum through predetermining courses of actions and breaking them down into partial tasks (Reichwald 1993) in order to improve the quantification and controlling of work (Hesch 1997). Taylorism aims primarily at a comprehensive and universal planning and controlling of production processes. That means that efficient solutions can be found through reduction of room of action, which in turn causes reduction of possible influence on operating sequences by single company members (Bayerl 1991; Pfeiffer and Weiß 1992). By contrast, lean production systems are designed to use the achievement potential as well as the practical knowledge of employees in a most comprehensive manner through reduction of direct work guidelines and an increase in influence (Bayerl 1991; Kopp 1993). These contrary ways of thinking reveal that whilst arranging the production system, both organisational concepts take different skills and achievement potentials as well as the wishes and needs of staff as a basis. Therefore they have a divergent image of human (Korge 2003b).

Implied Image of Human in Lean Production Systems

Due to the topicality of the subject matter, no image of human within lean production systems exists at the moment. Therefore, an implied integrated image of human in this production philosophy is being deduced and theoretically founded according to the findings of the so-called Dimensional Ontology by Frankl.

Frankl describes the human as a three-dimensional being, which unites within itself the physiological/biological, psychological/sociological and intellectual/noetic dimension, i.e. the physical, emotional, and intellectual reality characterises the human as a whole (Hesch 1997). On the bottom dimension, special meaning is attributed to the sensuous perception ability of the human as a physical being in lean production systems. The second dimension establishes the basis for multiple qualifications and lasting learning by characterising the human as an emotional-social being and considering the cooperation of thinking, feeling, and willing of employees in particular. In his intellectual dimension, the human strives for a permanent advancement of his being, above all for the need for authentic raison-d'être. While lean production systems particularly require creativity, confidence, autonomy, and emotions, as well as co-operative and integrating behaviour, companies characterised by the rational-economic image of human stress qualities such as obedience, egoism, business competition, dependence, control, and rationality. The rational-economic image is being replaced by the implied image of the holistic human. Consequently a paradigm change can be noted in the image of human of business administration (Hesch 1997). At the same time the dichotomy between employees and managers increasingly disappears and the human becomes increasingly regarded as an "enterpriser in the enterprise" (Hesch 1997). It becomes evident, that the employee-company relationship is no longer entirely based on an economic exchange, but depends on an inner bond.

For the conception of lean production systems this means that pure implementation and cross linking of best practice methods and best practice tools to establish a regulatory framework, and thereby the visible positive effects of a lean production (e.g., Kanban, Single Minute Exchange of Dies or Poka-Yoke), are not of prime concern (Dannecker 1993). The latter are merely success indicators, i.e. the symptom level (Dannecker 1993), and consequently represent impressive features of a lean and thereby exceedingly efficient organisation (Reiß 1992a). The significance rather lies within invisible courses of events, processes, and ideologies, which provide an essential basis for continuous development and improvement of success indicators. Therefore, they are marked as success determinants. Being a desirable form of thought and behavioural patterns of all employees, which should be a reaction to a certain situation or problem (Rother 2009), the continuous improvement process or Kaizen takes a prominent position as a hybrid component (Reiß 1992b), i.e. as a connector between the human-centred management and the operative components.

The Kaizen philosophy serves an enterprise as a mind-set, starting from the actual state and being based on continuous evolution and improvement. It guides employees through unknown and unpredictable terrain and moves them towards the unknown visionary target state by making employees study, understand the current situation on-the-spot, and react to it accordingly to their intellectual and behavioural routines (Rother 2009). Therefore, the quintessence of this production philosophy is the creation of a "community of scientists" at the Gemba level (Spear and Bowen 1999a, b). In an atmosphere of continuous striving for perfection, as well as an environment which promotes a lasting change (Liker 2004), this community aims at perfection and thereby fills the lean production system with life (Liker and Hoseus 2007), according to the motto "You can make it better—find out".

New Demands on Employees

Based on this view of employees, there is a need for multifunctional, qualified employees, who have basic technological knowledge of different technical teams as well as methodical and social competence, who are able to operate problem analysis and cause study as well as to work out proposals for solutions in the working area and help with its implementation (Becker 1992). Therefore, there is a requirement for an employee expertise, extended or different from the one of the traditional fordist-tayloristic system structures. This means that there is a need for a holistic competence to act, which is formed by the synergetic interplay of professional, methodical, and social competence (Regber and Zimmermann 2007; Faix et al. 1994).

Hence, the responsibilities and functions of managers and employees are no longer primarily defined by formal structure. Distinctions between individual organisation members and groups such as problem definition, responsibilities, personal interest, and self-conception, and thereby traits of psychological nature, which in a certain way run through the "heads of the managers and employees" gain more importance (Hesch 1997; Schmidt 1993). As a consequence, particularly attitudes or thinking structures and principles have to be changed in addition to the basic behaviour patterns during the introduction of lean production systems. Kieser, like Reichwald who refers to the "taylorism in the heads of the reengineers" (Reichwald 1995) sees the main obstacle for the introduction of new forms of organisation in the mental pressure of Taylorism (Kieser 1993; Hesch 1997).

In addition to thinking structures which define modes of cooperation and distribution of tasks, traditional mechanic-analytical, plainly functional-oriented thinking (Schultheiß 1995; Reinhardt 1995) in lean production systems must be overcome in favor of proactive, sensitive, holistic, potential-oriented, and economic thinking principles. These must be anchored in the minds of employees (Bösenberg and Metzen 1993; Keidel 1995). Within the scope of a proactive thinking principle, planning with foresight and design of future actions by means of continuous development of individual strengths and motives plays a major role. The sensitive thinking principle has to be regarded as a willingness to change. It is based on interior and outward information openness and therefore considers facts, individual moods, and emotions in decision taking. In the account of the holistic way of thinking, an intellectual interlude between part and wholeness, the fitting of partial findings in overall concepts as well as mutual thinking on different abstraction levels (Bleicher 1990) shall take place. This means that problems and their correlations are to be captured and approached appropriately from some sort of "bird's eye view". The core of the potential-oriented thinking principle is the development and usage of all resources. The economic thinking principle aims at an employee-driven elimination of the three types of waste *Muda* (waste), *Muri* (overburden), and *Mura* (unevenness) and must therefore be taken as a basis of all forms of thinking and all forms of behaviour of organisation members (Oeltjenbruns 2000; Ohno 1988; Keßler 2008). It becomes evident that, in addition to new organisation structures, there is first and foremost a need for new thinking structures and thinking principles as well as fundamentally new forms of behaviour (Hartmann 1992), i.e. the successful and sustainable implementation of lean production systems requires a revolution in thought and action of all employees of a company in particular (Baumgärtner 2006).

Human-Centred LPS Implementation in Supply Chains

According to the findings in the literature review of lean supply chain and supplier development concepts it is obvious that in most cases the company-wide successful implemented methods should be transferred to suppliers and other network partners. Here, the mere copying of methods and techniques of other companies do not lead to the desired results (Rother 2009). This is due to the fact that the production system and the methods are a result of a creative act of individual companies based on the thinking and behavioural routines which influences the daily working life of employees in companies (Rother 2009; Röhrle 2009).

The competitive advantage of a successful enterprise has not so much been with the concrete solutions like lean methods and tools but in the ability to perceive a situation, detect weaknesses and obstacles, and find suitable and smart solutions (Rother 2009; Korge and Lentes 2009). This ability can be understood as corporate culture of continuous improvement Liker and Hoseus 2007) and is an internalised mode of thinking, behaviour, and action of employees to question every situation, look for its improvement and realise the associated measures (Kämpf 2007). The employees deal with the problems and develop creative solutions. The sources of error should be avoided by their identification and elimination (Dannecker 1993; Müller 1995). At this, the existing processes and methods are always scrutinised, exhausted and trimmed to a high performance level in order to achieve an effective and efficient state of the whole supply chain (Spear and Bowen 1999a, b).

It is up to managers to develop the ability of their employees on the one hand, whereby this can lead to the evolvement of potentials for problem solving, process optimisation, and creativity (Rother 2009). On the other hand, the willingness of employees should be awakened to be aware of the advantage of engaging himself extensively: to improve own working conditions, to increase company profit, and to secure own workplace (Korge 2005).

Human-Centred Model for Implementation of Lean Production Systems

So there is a need for a new model for implementation of a lean production system which puts the employee at the centre. For this, the human-centred model for implementation of lean production systems by Uygun and Wagner (2011) is regarded, which is based on the few relevant concepts which exist in this context.

This model stresses the political-behaviour oriented and value-cultural dimension and focuses on the value stream of employees (see Fig. 1). As a consequence, the model aims at the internal core of transformation process in order to create a "fertile soil" for the company's specific implementation and development of operative methods and tools. It is based on a sustainable change of behaviour and attitude of all employees at the gemba level.

This model consists of an input, a nuclear transformation process, subsystems (instruments for conveying spirit and purpose and human-centred infrastructure) and an output. The output consists of five elements, which represent in its entirety an interdependent, self-contained system. The underlying image of human as well as the staff available to an enterprise function as input, in addition to design principles, philosophy, and the required new ways of thinking and behaviour

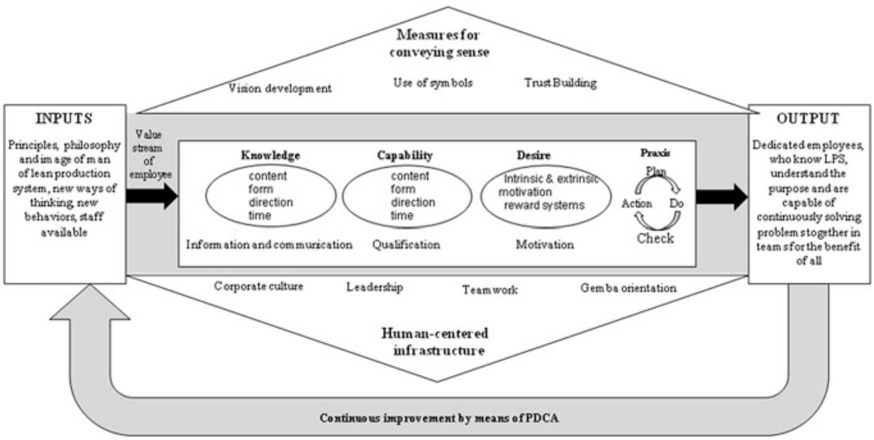


Fig. 1 Schematic illustration of the human-centred model (Uygun and Wagner 2011)

patterns. The core is formed by a change in behaviour of employees, which can be realised by suitable information and communication instruments, qualification instruments, and motivation instruments. Measures for conveying spirit and purpose must be used in this context to foster the middle to long-term change of consciousness of the employees based on a behaviour change. The human-centred infrastructure is amongst other things a basic requirement to meet the aspired evolutionary competences of a learning organisation. The output or the aim of this concept is the construction of acceptance for lean production systems, i.e. the creation of dedicated employees who know the production system and understand its purpose and are able to strive for continuous improvement on the basis of the Kaizen philosophy (Uygun and Wagner 2011).

Concept of Human-Centred LPS Implementation in Supply Chains

Based on this new model, the developed concept for employee-centred implementation of LPS in supply chains is divided into four phases, which build up on each other: planning, implementation, evaluation and expansion (see Fig. 2).

For a universal and clear implementation of this concept significant tasks of respective phases are subdivided into supporting steps. These were extracted from consisting approaches and were subjected to content enlargement and modification

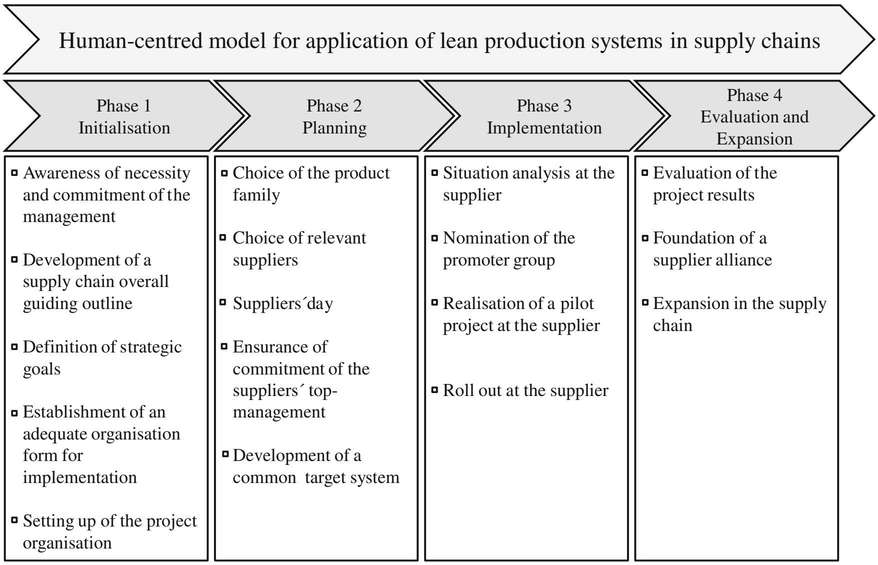


Fig. 2 Human-centred model for application of lean production systems in supply chains

according to corresponding requirements. The tools and methods used thereby support the phase-specific actions.

As particular aspects are sufficiently considered in literature, the focus of following statements lies on the remarks on employee-centred aspects of the particular phases.

Initialisation

The initialisation phase represents the initial position of implementation on the basis of the awareness of necessity and commitment of the management. Along with the development of the supply chain overall guiding outline and the definition of the superordinate strategic objectives the choice of an adequate organisation form as well as the founding of a project organisation are also included. Thereby the implementation strategy as well as the coordination form are appointed and anchored organisationally.

Therefore this phase begins with a positive decision of the management, which can first of all refer to an examination of the effectiveness of the network-wide LPS implementation or already to a complete development of the implementation (Dombrowski et al. 2008; Baumgärtner 2006; Hofbauer et al. 2009). Hereafter, a long-term vision is to be developed and to be specified in form of a supply chain overall guiding outline (Rapaport 1993). With the description of the leading idea and the direction of future inter-organisational operations the aimed long-term vision statement of the network or rather the network culture is to be presented. The basic principles enable the target group seeing the implementation of the human-centred LPS clearer and more relevant (Rother 2009; Rudnitzki 2002). The network partners as well as own employees should thereby get an identification opportunity with the plan and should be encouraged to do active participation (Hungenberg 2008; Becker 2007; Daniel 2001).

The following definition of superordinate strategic objectives is necessary to answer the question on what should be aimed at and achieved within an implementation of employee-centred LPS in the supply chain. By precise definition of goals from the outset, a target-oriented approach is possible for elaboration of the implementation and identification of the action demand (Hofbauer et al. 2009).

On the basis of defined strategic objectives the next significant task is the choice and establishment of an adequate organisation form for implementation and coordination of the plan, which has been sufficiently discussed in literature (Baungärtner 2006; RIF 2007).

Within the scope of arrangement and setting up of the organisation form necessary organisational project committees must be established for implementation and control of the implementation process. The establishment of an interdisciplinary project organisation ensures to a great extent that all cross-functional registers and experiences have been captured and processed. Moreover, organisational interfaces are overcome (Hartmann 2005; RIF 2007). The establishment of a steering committee, a coordination team, a support team and LPS experts teams are suggested in literature (RIF 2007; Barthel and Korge 2002; Korge 2003c).

On the basis of the employee-centred point of view the support teams and LPS experts teams must be mentioned. The latter consist of trained experts which posses a high competence concerning human-centred LPS. They are responsible for development and continuous improvement of standardised learning approaches which is used for transfer of employee-centred LPS within the network. Moreover, the team offers content support in all further planning functions (Korge 2003c). The support teams are responsible for operational realisation of the project and consist of mentors. Their task is to train the suppliers locally in implementation and to provide advice during the implementation of the human-centred management. Through the specific coaching of the improvement routine, which as a routine of thinking and acting mobilises the human abilities for improving and solving of problems, the development of the problem solving culture at the supplier is promoted. Moreover, with the help of a systematic approach an orientation along precise problems is offered (Rother 2009; RIF 2007).

Planning

Significant milestones of the planning phase are the choice of the product family and choice of an adequate network partner for realisation of the implementation project.

Thereby it is first of all useful to start with the supply chain of a product family which has a big leverage concerning improvement effects (Anand and Kodali 2008; Dolcemascolo 2006; Phelps et al. 2003) and to include the 1-tier-suppliers in the implementation. With the help of direct customer-supplier relation, motivation of the supplier and a successful realisation can be guaranteed (Weidemann 2005). The choice of adequate 1-tier-suppliers can be carried out with the help of a rating. Thereby, in contrast to the conventional supplier assessments within the supplier relationship management, the focus of the assessment within the scope of this concept lies not on the assessment of the performance of the supplier in the past, but on its development potential (Pauli et al. 2010; König 2009). Besides, the development potential describes the supplier‘s suitability for intensive inter-organisational cooperation within the scope of the LPS implementation. In this context evaluation measures could be: motivation of the supplier, lean experience of the supplier, mutual trust, availability of resources, readiness to communicate openly, lasting cooperative relationship, commitment of the top management, relation skills with own suppliers and strategic significance of the supplier (Dolcemoscolo 2006; Phelps et al. 2003; Krause and Elram 1997; Wagner and Boutellier 2003).

Within the scope of suppliers’ day, intention and plans of the concept as well as ideas of the human-centred LPS implementation can be communicated to the representatives of the supplier and feedback or suggestions referring to this can be collected. The event serves primarily as a possibility to give the supplier an understanding of advantages of a human-centred LPS implementation and to awaken his willingness (Handfield et al. 2006; Korge 2003c).

The ensurance of commitment, continuous support and cooperation of the top management of the suppliers offer a necessary condition for the success of the human-centred LPS implementation and for the final decision in favour of cooperation with the supplier within the scope of the LPS implementation (Oeltjenbruns 2000; Vitasek et al. 2005).

Finally, a common target system and measurement system should be compiled. The human-centred implementation of LPS can also be integrated as part of target agreements and thus contribute to supplier assessment (Hartmann 2005).

Implementation

The implementation phase starts with the situation analysis. In addition to traditional instruments for illustration of the actual state (e.g., value stream analysis, collection of key performance indicators and audits), the main focus within the scope of the presented concept is the assessment of soft factors in the company. Important for the duration and strategy of the implementation is the cultural level of an organisation (Scharfenberg 1997). To achieve a holistic and objective image it is necessary to break down the corporate culture in such characteristics which influence the development of problem consciousness directly. For this purpose the assessment according to Cottril and Gary (2006) represents an adequate methodology which assesses the corporate culture of the supplier in assessment categories and degrees of maturity steps.

After the situation analysis the nomination of the promoter group at the supplier is carried out. The promoter group must include different departments and different levels to guarantee cross-functional competence (MacDuffie and Helper 1997). Starting with this promoter group the human-centred implementation begins at the supplier. The team should learn the principle and the methodology of the improvement routine. Thereby, according to Rother the promoter group represents no staff position or lean department which will later be responsible for complete mentoring, training of the improvement routine, or improvements at the process level. The team members rather have the task of monitoring, adjustment and advancement of the learning approach of the organisation. In addition to the development of promoters first activities of the team can be used within the scope of the pilot project for the training of internal trainers. Therefore it is suitable to extend the team with employees who are suitable for this role (Rother 2009).

Thus, it makes sense to begin the implementation of the human-centred LPS with a clear pilot project. In particular the pilot project is considered as a place of qualification for the first internal coordinators and trainers of the improvement routine (Rother 2009). Starting with the promoter group in a defined area the mentors begin with the development of the internal improvement competence with the supplier (Liker and Meier 2007) following a pattern such as: Discussion and evaluation of the results of the situation analysis, implementation of a seminar to awaken awareness, development of the improvement routine and the coaching competence. During this phase the mentors are involved actively, teach and support the supplier with the development of his internal coordinators and “teachers” for improvement routines.

As the new approach cannot be introduced and implemented at once “overnight”, this changeover or the rolling out is carried out “step by step”. The duration of the roll-out depends on specific factors (maturity of partners, scope of the supply chain, etc.) Thereby, first experiences which are won by means of feedback loops can contribute to the optimisation of measures of the improvement culture (Hofbauer et al. 2009). After first experiences within the pilot project the roll out procedure can be continued independently by members of the promoter group. Thereby, they take over the role of mentors and begin the coaching of the roll out teams. The aim is the training of additional internal trainers of the improvement routine at all hierarchy levels of the company. These trainers on their part should develop thinking patterns in terms of the improvement routine within their team members (Rother 2009).

Evaluation and Expansion

Within the last phase of Evaluation and Expansion status and results of the implementation of the supplier are evaluated in order to reflect the target fulfillment in an assessing way. After evaluation of the results continuous improvement of network activities is carried out. On this occasion, the expansion of implementation within the supply chain is stressed, again beginning with the first phase. Thereby, the concept is carried out in a continuous cycle. Nevertheless, due to a cyclic enlargement of the users within the supply chain and network these cycles does not represent a ring-shaped, but a spiral development. With the concept developed here an ideal and generic procedure is suggested which is based on the examined theoretical and practical approaches. Within the realisation of the transfer of the human-centred LPS in the network this sequence serves as an assistance and orientation and does not claim to be exhaustive. The sequence of particular steps of the respective phases must therefore not be kept compulsory in the presented form and order. An adaptation according to specific company needs can be carried out within the practical implementation.

Evaluation and monitoring of the implementation project represent in that way an inalienable element of the successful inter-organisational cooperation (Howaldt and Ellerkmann 2007; König 2009) and contain thereby following aspects: in-process monitoring, evaluation of project results, evaluation and if necessary adaptation of the concept for the human-centred LPS implementation.

In addition to this, a supplier alliance can be constituted. This alliance is a union of suppliers involved in the implementation who establish a network for exchange of expert knowledge and experiences on the basis of regular meetings. Moreover, this represents thereby a permanent organisational structure, which exists during the implementation as well as during the following application and improvement stage (Dolcemoscolo 2006; Hines et al. 2000). By this means the organisation of common strategies and control systems, improvement initiatives and continuous learning is intended to develop sustainably a culture of network-wide striving for improvement. Furthermore, a common communication basis and information basis between customer and supplier as well as among suppliers is developed (Liker and Meier 2005).

The final expansion is carried out in two parallel processes. On the one hand, the OEM starting with the first phase triggers a new implementation project with another supply chain partner or another supply chain. On the other hand the supplier as multiplier also starts with its own implementation project with selected downstream suppliers and thus integrates further steps of the supply chain into the implementation. These again carry out the same approach with their own suppliers, etc. (Phelps et al. 2003).

Conclusion

The developed human-centred model for application of lean production systems in supply chains describes the structural and procedural roll out of the human-centred LPS implementation in the supply chain. It consists of four phases which have to be transacted consecutively. The human-based implementation of LPS supports the inter-organisational collaboration and integration along the supply chain by aligning the specific production process of the supply chain partners. The holistic view with an emphasis on employee-centred aspects will guarantee a more effective and sustainable application of LPS in supply chains because an LPS in its original meaning is all about people who are the enabler of the continuous improvement process.

Dealing with the human-centred implementation of lean production systems in supply chain it is obvious that the employees could expand their experience and competence horizon by means of other unconventional patterns so as to achieve their full potential. New performance levels should be reached by use of effective working methods and forms of cooperation based on mobilisation of creativity, problem awareness, and effectiveness (Rother 2009; König 2009). Besides, the implementation can be realised within the scope of an active supplier development with strategic suppliers.

References

Anand G, Kodali R (2008) A conceptual framework for lean supply chain and its implementation. Int J Value Chain Manage 3:313–357

Barthel J, Korge A (2002) Implementierung Ganzheitlicher Produktionssysteme als Aufgabe des Managements—Ergebnisse einer Studie in Brownfield-Werken der Automobilindustrie In: Institut für angewandte Arbeitswissenschaft e.V. (ed) Ganzheitliche Produktionssysteme—Gestaltungsprinzipien und deren Verknüpfung. Bachem, Köln

Baumgärtner G (2006) Reifegradorientierte Gestaltung von Produktionssystemen. Theoretische und empirische Analyse eines Gestaltungsmodells. TCW, München

Bayerl A (1991) Japanische Unternehmen. Ein Gegenentwurf zum abendländischen Modell moderner Organisation. Universität Erlangen-Nürnberg

Becker K (1992) Mit der Lean Production steigt wieder der Inhalt der Arbeit. Arbeitgeber 13(14):497–499

Becker T (2007) Leitbildentwicklung in Kooperationen. In: Becker T (ed) Netzwerkmanagement. Mit Kooperation zum Unternehmenserfolg, 2nd edn. Berlin, Heidelberg

Bleicher K (1990) Zukunftsperspektiven organisatorischer Entwicklung. Von strukturellen zu human-zentrierten Ansätzen. Zeitschrift für Führung + Organisation 3:152–161

Bösenberg D, Metzen H (1993) Lean Management. Vorsprung durch schlanke Konzepte, 2nd edn. Moderne Industrie, Landsberg

Brödner P (1993) Die Rückkehr der Arbeit in die Fabrik. In: Doleschal R (ed) Lean Production. Schlanke Produktion. Neues Produktionskonzept humanerer Arbeit? IAT Proceedings, Düsseldorf

Cottrill K, Gary L (2006) How soft metrics can make hard choices easier. Harvard business review supply chain strategy. Dezember 2005 to Januar 2006, pp 7–9

Daniel A (2001) Implementierungsmanagement. Ein Anwendungsorientierter Gestaltungsansatz. Dissertation, Universität Hohenheim, Gabler, Wiesbaden

Dannecker B (1993) Lean Management—Japanisierung oder eigener Weg?—Modifikation und Implementierungsfähigkeit einer angepassten deutschen Form dieser Managementkonzeption. o. V., Konstanz

Dolcemascolo D (2006) Improving the extended value stream. Lean for the entire supply chain. Productivity Press, New York

Dombrowski U, Zahn T, Grollmann T (2008) Roadmap für die Implementierung Ganzheitlicher Produktionssysteme. Z W F 12:1120–1125

Drumm H-J (2008) Personalwirtschaft. Springer, Berlin, Heidelberg

Faix W-G, Buchwald C, Wetzler R (1994) Der Weg zum schlanken Unternehmen. Moderne Industrie, Landsberg

Faust P (2009a) Zweite Lean-Welle—die sieben Thesen. Z W F 3:157–163

Faust P (2009b) Lean suply chains. Der Kunde als Taktgeber. Z W F 4:235–241

Füser K (2001) Modernes management. Lean management, business reengineering, benchmarking und viele andere Methoden. Deutscher Taschenbuch, München

Handfield RB et al (2006) Avoid the pitfalls in supplier development. In: Rhodes E, Warren J-P, Carter R (eds) Supply chains and total product systems a reader. Blackwell Publishing, Oxford

Hansmann K-W (1992) Industrielles management. Oldenbourg, München Wien

Hartmann D (1992) Lean Management. Beziehung Kunde/Hersteller. Toyotas Philosophie. In: RKW e. V. (ed) Lean production. Tragweite und Grenzen eines Modells. Tagungsband zum Workshop im Mai 1992, o. V., Düsseldorf

Hartmann H (2005) Lieferantenbewertung—Lieferantenentwicklung. Kernaufgaben eines gezielten Lieferantenmanagements. WEKA Median, Kissing

Hesch G (1997) Das Menschenbild neuer Organisationsformen. Mitarbeiter und Manager im Unternehmen der Zukunft. Gabler, Wiesbaden

Hines P et al (2000) Value stream management. Strategy and excellence in the supply chain. Financial Times, Prentice Hall, Harlow

Hofbauer G, Mashhour T, Fischer M (2009) Lieferantenmanagement. Die Wertorientierte Gestaltung der Lieferbeziehung. Oldenbourg, München

Howaldt J, Ellerkmann F (2007) Entwicklungsphasen von Netzwerken und Unternehmenskooperationen. In: Becker T et al (ed) Netzwerkmanagement. Mit Kooperation zum Unternehmenserfolg, 2nd edn. Springer, Berlin Heidelberg

Hungenberg H (2008) Strategisches management. Ziele-Prozesse-Verfahren, 5th edn. Gabler, Wiesbaden

Kämpf R (2007) Prozesse. Kontinuierlicher Verbesserungsprozess (KVP). In: Gienke H, Kämpf R (eds) Handbuch der Produktion. Carl Hanser, München

Keidel S (1995) Das Lean Management Konzept. Ursprünge—Bestandteile—Auswirkungen. Eine kritische Beurteilung unter normativen Aspekten. Dissertation, Hochschule St. Gallen

Keßler S (2008) Entwicklung eines Gestaltungsrahmens für Ganzheitliche Produktionssysteme bei Logistikdienstleistern. Dissertation, Technische Universität Dortmund, Praxiswissen, Dortmund

Kieser A (1993) Managementlehre und Taylorismus. In: Kieser A (ed) Organisationstheorien. W. Kohlhammer, Stuttgart

König C (2009) Interorganisationale Netzwerke zur kooperativen Optimierung Ganzheitlicher Produktionssysteme. Dissertation, TU Carolo-Wilhelmina, Braunschweig

Kopp R (1993) Lean control—Die Kontrollform der schlanken Produktion. In: Howaldt J, Minssen H (eds) Lean, leaner…? Die Veränderung des Arbeitsmanagements zwischen Humanisierung und Rationalisierung. Kontania Druck—und Verlagsgesellschaft, Dortmund

Korge A (2003a) Die unternehmensspezifische Ausgestaltung. Von Anforderungen zur Lösung. Log_X, Stuttgart 2003

Korge A (2003b) Nichts ist so praktisch wie eine gute Theorie. Hintergrundwissen für den Manager. Log_X, Stuttgart

Korge A (2003c) Beginnen Sie, ehe Sie es müssen. Ein Produktionssystem implementieren. In: Spath D (ed) Ganzheitlich produzieren-Innovative Organisation und Führung. Log_X, Stuttgart

Korge A (2005) Die Arbeitskultur für Lean Production gestalten. KVP-Impuls: Ein bewährter Methodenbaukasten zur Messung und Weiterentwicklung der Arbeitskultur. wt Werkstattstechnik online 2:11–18

Korge A, Lentes HP (2009) Ganzheitliche Produktionssysteme. Konzepte, Methoden, Erfolgsfaktoren In: Bullinger HJ (ed) Handbuch Unternehmensorganisation. Strategien, Planung, Umsetzung, 3rd edn. Springer, Berlin, Heidelberg

Krause DR, Elram LM (1997) Success factors in supplier development. Int J Phys Distrib Logistics Manage 1:39–58

Kuhn A, Uygun Y (2008) Robuste Netzwerke durch kollaborative Anwendung Ganzheitlicher Produktionssysteme. In: Pfohl H-C, Wimmer T (eds) Wissenschaft und Praxis im dialog. Robuste und sichere Logistiksysteme. Deutscher Verkehr, Hamburg

Liker JK, Meier D (2005) The Toyota way fieldbook: a practical guide for implementing Toyota's 4Ps. McGraw-Hill Professional, New York

Liker JK, Meier D (2007) Toyota talent: developing your people the Toyota way. McGraw-Hill Professional, New York

Liker J-K (2004) The Toyota way. 14 management principles from the World's greatest manufacturer. McGraw-Hill. New York

Liker J-K, Hoseus M (2007) Toyota culture. The heart and soul of the Toyota way. McGraw-Hill, New York

MacDuffie JP, Helper S (1997) Creating lean suppliers: Diffusing lean production through the supply chain. IMVP Working Papers (W 0165a)

Müller HJ (1995) Schlüsselqualifikation—Die evolutionäre Tiefenstruktur beruflicher Qualifikationen. In: Geißler H, Behrmann D, Peterson J (eds) Lean management und Personalentwicklung. Peter, Frankfurt

Nell-Breuning O (1950) Der Mensch im Betrieb. Zeitschrift für Betriebswirtschaft 5:257–266
Oeltjenbruns H (2000) Organisation der Produktion nach dem Vorbild Toyotas. Analyse, Vorteile und detaillierte Voraussetzungen sowie die Vorgehensweise zur erfolgreichen Einführung am Beispiel eines globalen Automobilkonzerns. Dissertation, Technische Universität Clausthal, Innovationen der Fabrikplanung und -organisation, 3rd edn. Shaker, Aachen
Ohno T (1988) Toyota production system—beyond large-scale production. Productivity Press, New York
Pauli M et al (2010) Logistische Lieferantenentwicklung in der Automobilindustrie. Ein heuristischer Ansatz zur Auswahl von logistischen Entwicklungskandidaten. In: Bogaschewsky R et al (ed) Supply management research. Aktuelle Forschungsergebnisse. Gabler, Wiesbaden
Pfeiffer W, Weiß E (1992) Lean Management. Grundlagen der Führung und Organisation industrieller Unternehmen, Erich Schmidt, Berlin
Phelps T, Hoenes T, Smith M (2003) Developing Lean Supply Chains. A Guidebook. Altarum Institute, The Boeing Company and Messier Institute. Discussion Paper. http://www.newvectors.net/publications/pdfs/esd_LeanSupplyChainGuideBook.pdf. Accessed 17 Mai 2010
Rapaport R (1993) To build a winning team: an interview with coach bill walsh. Harvard Bus Rev 71:111–120
Regber H, Zimmermann K (2007) Change-Management in der Produktion. Prozesse effizient verbessern im Team. Mi-Fachverlag, Landsberg
Reichwald R (1993) Der Mensch als Mittelpunkt einer ganzheitlichen Produktion. Innovative Organisationskonzepte aus betriebswirtschaftlicher Perspektive. Gabler, München
Reichwald R (1995) Der Taylorismus in den Köpfen der Reengineerer. Office Manage 5:12–13
Reinhardt R (1995) Lean management lernen. In: Geißler H, Behrmann D, Peterson J (eds) Lean Management und Personalentwicklung. Peter, Frankfurt
Reiß M (1992a) Schlanke Produktion: Primär Personalführung ist gefordert. Personalführung 6:456–461
Reiß M (1992b) "Lean management" ist "heavy management". Office Manage 5:38–41
RIF (2007) Dortmunder Initiative zur rechnerintegrierten Fertigung e.V.: Ganzheitliche Produktionssysteme entlang der Wertschöpfungskette. Abschlussbericht, Dortmund
Rivera L et al (2007) Beyond partnerships: the power of lean supply chains. In: Jung H, Chen F, Jeong B (ed) Trends in supply chain design and management. Technologies and methodologies. Springer, London
Röhrle J (2009) Der Mensch im ganzheitlichen Produktionssystem. In: Bullinger HJ (ed) Handbuch Unternehmensorganisation. Strategien, Planung, Umsetzung, 3rd edn. Springer, Berlin, Heidelberg
Rother M (2009) Toyota Kata. Managing people for continuous improvement and superior results. McGraw-Hill Education, New York
Rudnitzki J (2002) TANDEM—Die Lieferantenkooperation von DaimlerChrysler. In: Hahn D, Kaufmann L (ed) Handbuch Industrielles Beschaffungsmanagement. Internationale Konzepte—Innovative Instrumente—Aktuelle Praxisbeispiele, 2nd edn. Gabler, Wiesbaden
Scharfenberg H (1997) Implementierungsmanagement—effektiv und effizient. In: Nippa M, Scharfenberg H (ed) Implementierungsmanagement. Über die Kunst, Reengineeringkonzepte erfolgreich umzusetzen. Gabler, Wiesbaden
Schmidt J (1993) Die sanfte organisations-revolution. Von der Hierarchie zu selbststeuernden Systemen. Campus, Frankfurt New York
Schultheiß W (1995) Lean management. Strukturwandel im Industriebetrieb durch Umsetzung des Management-Ansatzes. Expert, Renningen-Malmsheim
Spear S, Bowen H-K (1999a) Decoding the DNA of the Toyota production system. Harvard Bus Rev 9(10):97–106
Spear S, Bowen K (1999b) Decoding the DNA of the Toyota production system. Harvard Bus Rev September/October 1999:96–106

Stürzel W (1992) Lean production in der Praxis. Spitzenleistung durch Gruppenarbeit. Jufermann, Paderborn

Uygun Y, Keßler S, Stausberg J (2009) Verbreitung Ganzheitlicher Produktionssysteme. Erkenntnisse aus einer deutschlandweiten Studie. wt Werkstattstechnik online 3:136–140

Uygun Y, Wagner SU (2011) Guidelines for human-based implementation of lean production. In: Proceedings of the 44th CIRP international conference on manufacturing systems

Vitasek K, Manrodt KB, Abbott J (2005) What makes a lean supply chain? Supply Chain Manage Rev 7:39–45

Wagner SM, Boutellier R (2003) Lieferantenentwicklung in der supply chain—Eine Bestandsaufnahme. Logistik Manage 4:50–70

Weidemann M (2005) Auswahl von Gestaltungsmöglichkeiten logistischer Schnittstellen in Produktionsnetzwerken mittelgroßer Stückgutfertiger unter besonderer Berücksichtigung der Zulieferindustrie. Shaker, Aachen

The Multiple Batch Processing Machine Problem with Stage Specific Incompatible Job Families

Marc-Andrè Isenberg and Bernd Scholz-Reiter

Abstract This contribution introduces the problem of batching and scheduling n jobs in non-permutation flow shops with $m \geq 2$ batch processing machines in sequence and stage specific incompatible job families. Batches are build at each stage according to family-specific parameters. This creates a stage-interdependent batching and scheduling problem. We formulate the model by considering different options in terms of release dates, due dates, objective functions as well as setup times and define a corresponding integer linear program.

Introduction

In this contribution we define the multiple batch processing machine problem with stage specific incompatible job families ($MBPM_{SIF}$) and test its complexity factor. This problem consist of batching and scheduling multiple products in flow shop environments with multiple stages and intermediate storages in between. In general we understand batching as the task of selecting jobs and building batches out of jobs. Scheduling generates machine specific batch sequences by determining starting and ending times of batch productions. Each stage has one batch processing machine (BPM) with limited capacity. The product orders are represented by jobs during the production phase. All jobs have to be processed on all machines in the same technological order. However, the job sequence is allowed to change between the stages. This is called a *non-permutation schedule*. For changing the job sequence and for temporary storage we take intermediate storages with infinite capacities into our

M.-A. Isenberg (✉) · B. Scholz-Reiter
Planning and Control of Production Systems, University of Bremen,
Hochschulring 20 D 28359 Bremen, Germany
e-mail: ise@biba.uni-bremen.de

H.-J. Kreowski et al. (eds.), *Dynamics in Logistics*, Lecture Notes in Logistics,
DOI: 10.1007/978-3-642-35966-8_9, © Springer-Verlag Berlin Heidelberg 2013

model consideration. Additionally, we include transport times from machine to storage and vice versa. We assume the options of common or distinct due dates. Furthermore, we introduce stage specific release dates which are the maximum of the *availability* and *realisability* of jobs. A job is available in a stage if it has completed all previous stages. A job is realisable if all materials which are required for stage specific job processing are available. The realisability options contain the material availability at *time zero* as well as *randomly-distributed* which describes the usage of an arbitrary distribution. There is full knowledge about the production demand and the resulting jobs for a specified time period. The time period can be divided into time slots. The jobs belong to *stage specific job families* with family-specific processing times. They are incompatible to each other. Such incompatibility results from the production of multiple products and considers realistic production scenarios where products have several product variants. Since product variants have partially the same components or colours jobs for different products can build common batches within a certain stage. This causes a reduction of the variety of different products to a stage specific number of different job families. The feature combination of multiple BPM in sequence and stage specific job families lead to *inconsistent batches* thus the job constellation of batches only persist for single stages. This is contrary to other batching and scheduling problems where the batch is only build once at the first stage and then endures. Furthermore, the batch building of jobs follows the *batch availability model* with *anticipatory setups*, *batch sequence-dependent setup-times* and a *non-preemptive scheduling* procedure [compare to (Allahverdi 2008; Potts and Kovalyov 2000)]. Following the batch availability model the jobs have the same starting and end time and become available for subsequent stages at once. Opposing to this in the job availability model jobs of a batch are independent from each other in terms of their completion times. Anticipatory setups specify the knowledge about the next batch in queue for a machine to arrange the setup before its arrival. Batch sequence-dependent setup-times appear if the previous batch on a machine belongs to a different family than the subsequent one. Non-preemptive scheduling exclude the possibility of setting manufacturing priorities for individual jobs. Since the $MBPM_{SIF}$ induces stage-interdependencies in batching and scheduling, reasonable objective functions are the minimisation of total flow time, total completion time or makespan. Following the notations of Graham et al. (1979), Potts and Kovalyov (2000), we denote our model as Fm/s_{ifg}, b_i, $d_j/\sum f_j$ (Graham et al. 1979; Potts and Kovalyov 2000). For more detailed descriptions about notations and characteristics we recommend (Allahverdi 2008; Potts and Kovalyov 2000).

The research is inspired by a real word scenario. Automakers partially ships their export cars in the so-called Completely-Knocked-Down procedure. Completely-Knocked-Down describes the shipment of cars as assembly sets to save import taxes in the country of destination. Within the preparation phase of the shipment the car components are picked and packed into boxes. To gain economies of scale this process is done by building batches out of boxes with the same content. Following, the boxes come into an intermediate storage. Subsequent the boxes are stuffed into containers with different destinations. The transfer to the above model works as follows: the *picking and packing process* is the first BPM, the *stuffing process* the second one. The creation of batches derives from clustering boxes with identical content and stuffing boxes with equal destinations into the same container. Due to this exigences both BPM have their own job families. Also they are limited in space and volume thus they have a maximum batch size.

The objective of this contribution is the formulation of the $MBPM_{SIF}$ achieving a comprehensive understanding of the problem and thus enabling the development of solution approaches. The article is structures as follows: section Related Work presents a literature review concerning related work. section Problem Statement illustrates the problem and formulates the mathematical model and integer linear program respectively. A summary of the contribution as well as an outlook are presented in section Summary and Outlook.

Related Work

There exist a number of reviews and surveys which precis the state of the art for research concerning batching and scheduling (Allahverdi 2008; Mathirajan and Sivakumar 2006; Potts and Kovalyov 2000; Webster and Baker 1995). We started our literature research on the basis of their categorizations and went over to more specified criteria of our problem such as *multiple batch processing machines in sequence* and *stage specific incompatible job families*. These characteristics are the most important ones since they cause discontinues processing and disjunctive sets of jobs for batch formation thus increasing the complexity of batching and scheduling. Furthermore, we only considered batch availability models and offline-algorithm. The focus on offline-algorithm derives from our model which underlies full knowledge of the production demand of a future period. That induces deterministic behaviour.

The combination of multiple BPM in sequence and stage specific job families automatically leads to the appearance of *inconsistent batches*. There is a very limited number of papers which explicitly consider this characteristic. Ng and Kovalyov (2007) study flow shops with several BPM in sequence and machine-dependent setup-times. They prove for several problems that a permutation schedule is the optimal one (Ng and Kovalyov 2007). However, their demonstration only includes compatible job families. The characteristic of stage specific incompatible job families would impede the feasibility of a permutation schedule. Buscher and Shen (2009) develop a mixed integer programming formulation implying inconsistent batches for a flow shop scheduling problem involving several BPM. Their program is only able to solve small instances and does not consider incompatible job families. Additionally, there are some more approaches on flow shops with multiple BPM in sequence composing consistent batches. For instance Kumar Manjeshwar et al. (2009) as well as Lei and Guo (2011) apply meta-heuristics such as simulated annealing and variable neighbourhood search (Kumar Manjeshwar et al. 2009; Lei and Guo 2011).

Multiple stages and incompatible job families are considered by Fu et al. (2011), Kim et al. (2001), Tajan et al. (2006). Fu et al. (2011) analyse the problem of a two-stage flow shop containing a single BPM following by a discrete processor and a limited buffer in between Fu et al. (2011). Kim et al. (2001) as well as Tajan et al. (2006) study the reverse order which consists of a serial processor feeding a BPM (Yd et al. 2001; Tajan et al. 2006). In all models the incompatible job families are only valid for the batch processing stage. Besides the flow shop focus there are several papers which concentrate on batching and scheduling of single batch processing machines regarding incompatible job families ($SBPM_{IF}$). Yao et al. (2012) consider dynamically arriving jobs and analyse different dominance properties. They deduce several lower bounds and propose a basic branch-and-bound algorithm, which they improve to a decomposed branch-and-bound algorithm regarding the dynamic job arrival. The objective is the minimisation of the total completion time (Yao et al. 2012). Koh et al. (2005) study the $SBPM_{IF}$ in a multi-layer ceramic capacitor production line and propose a number of heuristics as well as hybrid genetic algorithm to solve problems with an extensive number of jobs in reasonable computation time (Koh et al. 2005). A potential transfer of the $SBPM_{IF}$ approaches to our model is in adapting them onto each stage of our model. However, the stages would be planned separately and without any consideration of interdependencies out of the stage specific scheduling.

Additionally, many researchers analyse the batching and scheduling problems of single stages including identical or non-identical parallel batch processing machines with incompatible job families ($PBPM_{IF}$).Their research is mostly motivated by semiconductor wafer fabrication. Jula and Leachman (2010) propose exact and heuristic algorithms for the $PBPM_{IF}$. Their model also includes feeding and removing apparatus (Jula and Leachman 2010). Chiang and Cheng (2010) develop a memetic algorithm for minimizing total weighted tardiness in $PBPM_{IF}$ regarding dynamic job arrival and incompatible job families. Their algorithm plans the batch formation and batch sequence simultaneously (Chiang and Cheng 2010). Since the papers of this category mainly focus on the usage of common resources and do not regard scheduling dependencies across stages we do not consider more approaches and refer to Mathirajan and Sivakumar (2006).

Summarising the results of our literature research we notice that no analysed model or approach combines the characteristics of *multiple batch processing machines in sequence* and *stage specific incompatible job families*. However, the illustrated research presents related and interesting work and deals with certain elements of our problem. Hence, they may support the following problem formulation.

Problem Statement

In general, batching and scheduling problems consist of two fundamental decision for each BPM. The first one concerns the batch formation, the second one contains the batch sequencing. Regarding our model we need to be aware that both problems are related to each other due to the characteristics such as incompatible job families and BPM capacities. Naturally the batch formation precedes. However, it influences the completion time of a batch by arranging its batch size and thus determining the availability of the containing jobs for subsequent BPM. Since the following BPM has another job families than its predecessor the batch cannot automatically proceed as a whole, it could contain incompatible families for the BPM. This effects a necessary re-formation of the batches between the stages. Smaller batch sizes could reduce this problem, but they are causing more setup-times by an increased changing rate of batch families and decrease the efficiency of the BPM. For permuting the job sequence and composing new batch formations there are storages with unlimited capacities located between the BPM. This ensures that BPM never block. Due to the option of distinct due dates and the re-formation of batches we notice that jobs of the same batch may vary in their stage specific release dates and deadlines. In addition, the planning problem could be more complex by introducing production time slots. They have limited time capacities and batches are mostly not allowed to split across them. They need to be completed within one time slot. However, the option of different time slots are not considered in the following notation and problem formulation.

Mathematical Problem Formulation

Following, we present the notation and underlying assumptions of the integer linear program of the $MBPM_{SIF}$. We use the following notation:

Subscripts

$k = 1,\ldots,m$	Machine index
$j, i, c = 1,\ldots,n$	Job index
$s = 1,\ldots,x$	Storage index
$a_k = 1,\ldots,g_k$	Family index on machine k

Parameters

$P_{k,j}$	Processing time of job j on machine k
Z_j	Due date of job j
$u_{k,j}$	Setup time for job j on machine k (equal for all family members)
$SMIN_k$, $SMAX_k$	Minimum and maximum batch size at machine k
$T_{k,s}$, $T'_{s,k}$	Transport time for batches from k to s and vice versa
$EoS_{k,j}$	Economies of scale for job j if its batch has more than one job
$bigM$	Big integer number

Binary variables

$\delta_{k,j,i}$	1 if job j is scheduled before job i on machine k
$\theta_{k,j,i}$	1 if job j is direct predecessor of job i on machine k
$\gamma_{k,j,i}$	1 if job j and i belongs to the same family on machine k
$\pi_{k,j}$	1 if job j has a directly preceding job on machine k which belongs to the same family
$B_{k,j,i}$	1 if job j and job i belongs to the same batch on machine k
$\phi_{k,j,i,c}$	1 if job c is scheduled between j and i on machine k

Integer variables

C_j	Completion time of job j
f_j	Flow time of job j
C_{max}	Makespan
$r_{k,j}$, $r'_{k,j}$	Start and completion time of job j on machine k
$v_{k,j}$	Shortfall below $SMIN_k$ of the batch containing job j
$D_{k,j}$	Deadline of job j on machine k
$A_{k,j}$	Realisability date of job j on machine k
$R_{k,j}$	Release date of job j on machine k

There are several assumptions. Besides the intermediate storages additional storages are located before the first and after the last batch processor. Hence, the flow shop has $x = m + 1$ storages. All jobs have to pass all storages. Transport times are from machines to storages and vice versa and count for the whole batch. The production planning is far ahead of the end of the production period thus we assume an infinite time horizon. The following integer linear program focuses mainly on the minimisation of the total flow time. Thereby it uses batch sequence-dependent setups. Both can be easily substituted by other objectives and setup approaches.

Problem Formulation

Objective function

$$min\ v_{k,j} + \begin{cases} \sum_{j=1}^{n} f_j & \text{if objective is total flow time} \\ \sum_{j=1}^{n} C_j & \text{if objective is total completion time} \\ C_{max} & \text{if objective is makespan} \end{cases} \tag{1}$$

Environmental constraints and definition

$$f \geq \sum_{k=1}^{m} \sum_{j=1}^{n} v_{k,j} + \sum_{j=1}^{n} (Z_j - r_{1,j}) \tag{2}$$

$$r'_{k,j} \geq r_{k,j} + (1 - \pi_{k,j}) \cdot u_{k,j} + P_{k,j} \cdot \sum_{i=1}^{\mathrm{n}} B_{k,j,i} - EoS_{k,j} \cdot (-1 + \sum_{i=1}^{n} B_{k,j,i}) \quad \forall k,j \tag{3}$$

$$\sum_{j=1}^{n} \theta_{k,j,i} \cdot \gamma'_{k,j,i} - \pi_{k,i} \cdot bigM \leq 0 \quad \forall k, i \tag{4}$$

$$\sum_{j=1}^{n} \theta_{k,j,i} \cdot \gamma'_{k,j,i} + (1 - \pi_{k,i}) \cdot bigM \geq 1 \quad \forall i \tag{5}$$

$$r_{k,j} \geq r'_{k-1,j} + T_{k-1,s} + T'_{s,k} \quad \forall k > 1, j, s \tag{6}$$

$$r'_{m,j} + T_{m,x} \leq Z_j \quad \forall j \tag{7}$$

Machine constraints

$$\sum_{j=1}^{n} B_{k,j,i} \leq \ SMAX_k \quad \forall k, i \tag{8}$$

$$v_{k,j} + \sum_{j=1}^{n} B_{k,j,.i} \geq \ SMIN_k \quad \forall k, i, v_{k,j} \leq 1 \tag{9}$$

$$r_{k,j} \geq r'_{k,i} - bigM \cdot (\delta_{k,j,i} + B_{k,j,i}) \quad \forall k, j, i, j \neq 1 \tag{10}$$

$$r_{k,i} \geq r_{k,i} - bigM \cdot (1 - \delta_{k,j,i} + B_{k,j,i}) \quad \forall k, j, i, j \neq 1 \tag{11}$$

$$\delta_{k,j,i} + \delta_{k,i,j} + B_{k,j,i} \leq 1 \quad \forall k, j, i \tag{12}$$

$$\delta_{k,j,i} + \delta_{k,j,c} + \delta_{k,c,i} - \phi_{k,j,i,c} \cdot bigM \leq 2 \quad \forall k, j, i, c \tag{13}$$

$$\delta_{k,j,i} + \delta_{k,j,c} + \delta_{k,c,i} + (1 - \phi_{k,j,i,c}) \cdot bigM \geq 3 \quad \forall k, j, i, c \tag{14}$$

$$\sum_{c=1}^{n} \phi_{k,j,i,c} - (1 - \delta_{k,j,i} + \theta_{k,j,i}) \cdot bigM \geq 1 \quad \forall k, j, i, j \neq i \tag{15}$$

$$\sum_{c=1}^{n} \phi_{k,j,i,c} - (2 - \delta_{k,j,i} - \theta_{k,j,i}) \cdot bigM \leq 0 \quad \forall k, j, i, j \neq i \tag{16}$$

$$\delta_{k,j,i} \geq \theta_{k,j,i} \quad \forall k, j, i \tag{17}$$

Batch constraints

$$r_{k,j} \geq r_{k,i} + 1 - (1 - \delta_{k,j,i}) - bigM \quad \forall k, j, i \tag{18}$$

$$r_{k,j} \geq r_{k,j} - \delta_{k,i,j} \cdot bigM \quad \forall k, j, i \tag{19}$$

$$B_{k,j,i} \leq \gamma'_{k,j,i} \quad \forall k, j, i \tag{20}$$

$$B_{k,j,j} \geq 1 \quad \forall k, j \tag{21}$$

$$\delta_{k,j,i} + \delta_{k,i,j} + B_{k,j,i} \cdot bigM \geq 1 \quad \forall k, j, i \tag{22}$$

The objective function (1) includes the options of minimising the total flow time, the total completion time or the makespan. Additionally, there are also the conceivable options of minimising the mean flow time $n^{-1}\sum f_j$ or mean completion time $n^{-1}\sum C_j$. Since they are equal to total flow time and total completion time we do not formulate them explicitly. The choice of the objective criterion depends on the motivation for analysing the MBPM_{SIF}. While the usage of the total flow time criterion allows a more intensive consideration of storages, the total completion time criterion focuses on the BPM. As the Eq. (2) describes, the above formulation focuses on the total flow time. However, the exclusively pursuance of this objectives would partially in contrast to the fundamental idea of batch processing which is driven by economic efficiency. Supposing the abandonment or insignificance of setup-times decision-makers would prefer batches of size one since they do not cause waiting times in storages until other jobs of the same job family are available. Hence, we introduce a weak constraint (9) for the minimal batch size of a BPM. This weak constraint is regarded in the objective function by the term of $v_{k,j}$ which defines the shortfall of the minimal batch sizes. It is restricted to allow a shortfall of at most 1. In contrary to Eq. (9) constraint (8) refers to the maximum number of jobs within a batch. Due to the physical bounded space it is constructed as a hard restriction which needs to be regarded strictly. The batch sequence-dependent setup events are determined by the constraints (4) and (5). These constraints work as an *either-or-option*. They analyse if one of the directly preceding jobs of j belongs to the same family as j does. The term (3) describes the completion time of a job depending on its production start on the machine, a potentially setup as well as the duration of the corresponding batch. The batch duration derives from the sum of the jobs processing times on machine k as well as on the economies of scale. This effect allegorises the savings of time which occur by the aggregation of material provisions, identical working procedures, learning effects etc. We advert that this formulation of setups only specifies *when* a setup precedes a batch, it does not quantify their duration. These could be constant for all families, family-dependent, machine-dependent or both. Constraint (6) defines the technological order of machines. All jobs needs to pass all machines and all storages. Equation (7) ensures that all jobs are in the last storage until their due dates. The constraints (10) and (11) identify the order of jobs and avoid overlapping of batches. They are also formulated as a *either-or-option*. In the case that job j is scheduled at any time before job i $\delta_{k,j,i}$ is *1*. If the jobs build a common batch or if j is scheduled after i $\delta_{k,j,i}$ is 0. Equation (12) avoids bidirectional hierarchies and ensures that two jobs either have one hierarchy proportion or that they constitute a common batch. The complementary constraint to this is (22). For the case that there is no hierarchy proportion it reasons that the jobs build a common batch. The constraints (13 and 14) determine whether a job c is between job j and job i. This serves as a basis for the Eqs. (15 and 16). They count the number of jobs between job j and job i. If the sum of the intermediate jobs is 0 job j is a direct predecessor of i. In this case the sequence variable $\theta_{k,j,i}$ is set to 1. Constraint (17) defines that the sequence variable $\theta_{k,j,i}$ only can be set to 1 if the order variable $\delta_{k,j,i}$ also is 1. That means that job j can only be a predecessor of job i if they have an equal hierarchical order. The Eqs. (18 and 19) defines the same starting time for jobs of the same batch. The constraint (20) makes sure that jobs only can constitute a batch if they belong to the same job family on machine k. Equation (21) completes the matrix of $B_{k,j,i}$ and ensures that the belongings of two jobs to the same batch are bidirectional.

Analysing the problem size and complexity of the integer linear program we applied the pre-solver of GUROBI (Bixby et al. 2012) The scenario consist of $m = 2$ BPM, random job family processing times $P_{k,j}$ between 20 and 60 min, EoS is 15 % of the job′s process duration, the due dates Z_j out of an interval of 100 min, family-specific setup times $u_{k,j}$ of 50 % of a full sized batch of the specific family and transport times $T_{k,s}$ as well as $T'_{s,k}$ of 5 min. The other parameters $SMIN_k$, $SMAX_k$, g_k as well as the numbers of integer variables, binary variables and nonzeros after pre-solving are described for exemplary cases in Table 1.

Extending the above described model we further present the mathematical formulations of stage specific release dates and deadlines. They are necessary for the development of heuristically solution approaches solving the $MBPM_{SIF}$.

$$R_{k,j} = \begin{cases} \max\left\{A_{k,j}, r'_{k-1,j} + T_{k-1,s} + T'_{s,k}\right\} & \text{if } k > 1 \\ A_{j,k} & \text{else} \end{cases} \quad \forall k, s = k \tag{23}$$

$$D_{k,j} = \begin{cases} r_{k+1,j} + T'_{s,k+1} - T_{s,k} & \text{if } k < m \\ Z_j - T_{k,s} & \text{else} \end{cases} \quad \forall k, s = k \tag{24}$$

According to Eq. (23) the release date of job j on machine $k + 1$ results out of the maximum of its realisability date and its availability date. The realisability date corresponds to the provision of required material in the specific stage. The job availability date results from the completion of its batch on the previous machine $k - 1$ and the transport times for the batch from machine $k - 1$ to storage $s = k$ to

Table 1 Exemplary problem sizes

n	g_1	g_2	$SMIN_1$	$SMAX_1$	$SMIN_2$	$SMAX_2$	Integers	Binaries	Non-zeros
10	3	4	3	4	2	3	1,765	1,721	15,924
15	3	4	3	4	2	3	6,348	6,273	61,761
20	5	4	3	4	3	5	15,166	15,046	147,992
30	5	4	3	4	3	5	52,089	51,909	513,344
50	9	6	3	4	3	5	25,3847	243,547	2,427,108
60	9	6	3	4	3	5	42,3063	422,703	4,223,142

machine k. If $k = 1$ the availability of the job only depends on its realisability. The deadline $D_{j,k}$ of a job j for machine k results by calculating the reverse order in (24). If $k = m$ the job′s deadline derives from its due date instead of its starting time on the following machine. This calculations are equal if the $\mathrm{MBPM_{SIF}}$ contains common or distinct due dates or if the realisability dates are chosen as time zero or distributed. If the focus of analysing the $\mathrm{MBPM_{SIF}}$ excludes storages the transport times can be set to zero.

Problem Complexity

Since we allow a high variety of combinations in terms of due dates, deadlines, objective function etc. we only address exemplary cases for providing different perspectives on the complexity of $\mathrm{MBPM_{SIF}}$. First we want to consider the case that we have different time slots with equal time capacity. The batches have non-identical durations. If we pursuit the minimisation of makespan $\mathrm{C_{max}}$ it is advantageous to realise a maximum usage of the time slots, i. e. no idle times of machines. This problem layout is equal to the one dimensional bin packing problem where we have bins of limited capacity and boxes of non-identical sizes. The bin packing problem is known as NP-hard (Garey et al. 1976). However, this reduction only regards the scheduling problem and does not work if we have only one time period with infinite time capacity. For this case Glass et al. (2001) proof for a flow shop of $m = 2$, $g_k = 1$ for all k, inconsistent batches and the objective of minimising the makespan that it is NP-hard in the strong sense. But they do not explicitly regard incompatible families. This is done by Webster and Baker (1995) for $\mathrm{SBPM_{IF}}$ and different objectives. According to Uzsoy (1994) and for the objective of minimising total flow time they show that the problem is NP-hard (Webster and Baker 1995). The case of distinct release dates is considered by (Yuan et al. 2006). They proof that the $\mathrm{SBPM_{IF}}$ is strongly NP-hard even for two distinct release dates and also if the processing times of the jobs are equal and the job families have identical setup-times. The $\mathrm{MBPM_{SIF}}$ has distinct release dates in each stage of the flow shop. Garey and Johnson (1990) proved that finding the minimum mean flow time schedule in a flowshop of $m \geq 2$ machines is NP-complete (Garey et al. 1976). If we take the job constellation for batches as a given, the challenge of $\mathrm{MBPM_{SIF}}$ only consist of finding the best schedule. Due to the presented results we assume that the $\mathrm{MBPM_{SIF}}$ is also NP-complete. This degree of complexity is not effected by regarding or omitting distinct due dates or varying realisability dates. Different release dates in a stage results from the feature combination of multiple stages and stage specific incompatible job families. Therefore, all presented variants of the $\mathrm{MBPM_{SIF}}$ are NP-hard.

Summary and Outlook

This contribution introduces and formulates the $MBPM_{SIF}$ as a logical extension of $SBPM_{IF}$ and $PBPM_{IF}$. The mathematical formulation describes a general model which allows flexibility in terms of the objective function, release dates, due dates, setup-times and storage consideration. Using a solver for model consolidation we illustrate the problem complexity for small scale test cases. According to the results of other researches we classify the $MBPM_{SIF}$ as NP-hard.

Further research should be done to find feasible solution approaches for the $MBPM_{SIF}$. Due to the hardness of the problem exact algorithms or MILP solver only can solve small instances. They could contribute to the understanding of the problem by determining the significance of single elements for the problem complexity. However, feasible solutions for more extensive instances only can be provided by heuristics. The development of heuristic needs to be made for specified characteristics of the $MBPM_{SIF}$ since this ensures a higher quality of the computed solution. Further one, such a solution could be serve as a basis method for improvement by developing and applying meta-heuristics [compare to Kumar Manjeshwar et al. (2009); Lei and Guo (2011)]. The $MBPM_{SIF}$ does not need restricted to the described options. Many more modifications and extensions are thinkable as long they do not change the characteristics of multiple stages and stage specific incompatible job families.

References

Allahverdi A (2008) A survey of scheduling problems with setup times or costs. Eur J Oper Res 187:985–1032

Bixby R, Gu Z, Rothberg E (2012) Gurobi. URL http://www.gurobi.com. Accessed 12/03/16

Buscher U, Shen L (2009) MIP-based approaches for solving scheduling problems with batch processing machines. Oper Res 132–139

Chiang T, Cheng H (2010) A memetic algorithm for minimizing total weighted tardiness on parallel batch machines with incompatible job families and dynamic job arrival. Comput Oper Res 37(12):2257–2269

Fu Q, Sivakumar A, Li K (2011) Optimisation of flow-shop scheduling with batch processor and limited buffer. Int J Prod Res p Forthcoming article the multiple batch processing machine problem 11

Garey MR, Johnson DS (1990) Computers and intractability; a guide to the theory of NP-completeness. W. H. Freeman & Co., New York

Garey MR, Johnson DS, Sethi R (1976) The complexity of flowshop and jobshop scheduling. Math Oper Res 1(2):117–129

Glass CA, Potts CN, Strusevich VA (2001) Scheduling batches with sequential job processing for two-machine flow and open shops. INFORMS J Comput 13(2):120–137

Graham RL, Lawler EL, Lenstra JK, Kan Rinnooy A (1979) Optimization and approximation in deterministic sequencing and scheduling: a survey. Ann Discrete Math 5:287–326

Jula P, Leachman R (2010) Coordinated multistage scheduling of parallel batch-processing machines under multiresource constraints. Oper Res 58(4-Part-1):933–947

Kim Yd, Kim Jg, Choi B, Kim Hu (2001) Production scheduling in a semiconductor wafer fabrication facility producing multiple product types with distinct due dates. IEEE Trans Robot 17(5):589–598
Koh S, Koo P, Kim D, Hur W (2005) Scheduling a single batch processing machine with arbitrary job sizes and incompatible job families. Int J Prod Econ 98(1):81–96
Kumar Manjeshwar P, Damodaran P, Srihari K (2009) Minimizing makespan in a flow shop with two batch-processing machines using simulated annealing. Robot Comput Integrated Manuf 25(3):667–679
Lei D, Guo XP (2011) Variable neighbourhood search for minimising tardiness objectives on flow shop with batch processing machines. Int J Prod Res 49(2):519–529
Mathirajan M, Sivakumar A (2006) A literature review, classification and simple meta-analysis on scheduling of batch processors in semiconductor. Int J Adv Manuf Tech 29(9–10):990–1001
Ng CT, Kovalyov MY (2007) Batching and scheduling in a multi-machine flow shop. J Sched 10(6):353–364
Potts C, Kovalyov M (2000) Scheduling with batching: a review. Eur J Oper Res 120(2):228–249
Tajan JBC, Sivakumar AI, Gershwin SB (2006) Performance of a serial-batch processor system with incompatible job families under simple control policies. Mechanical Engineering (65), URL http://hdl.handle.net/1721.1/29835
Uzsoy R (1994) Scheduling a single batch processing machine with nonidentical job sizes. Int J Prod Res 32(7):1615–1635
Webster S, Baker KR (1995) Scheduling groups of jobs on a single machine. Oper Res 43(4):692–703
Yao S, Jiang Z, Li N (2012) A branch and bound algorithm for minimizing total completion time on a single batch machine with incompatible job families and dynamic arrivals. Comput Oper Res 39(5):939–951
Yuan JJ, Liu ZH, Ng CT, Cheng TCE (2006) Single machine batch scheduling problem with family setup times and release dates to minimize makespan. J Sched 9(6):499–551

Scheduling Multiple Batch Processing Machines with Stage Specific Incompatible Job Families

Marc-Andrè Isenberg and Bernd Scholz-Reiter

Introduction

We develop and evaluate several heuristics for generating a valid production plan for a multi-product multi-stage flow shop according to the $MBPM_{SIF}$. The heuristics are simple and fast thus they can serve as basis methods for meta-heuristics such as genetic algorithms. Since we assume full knowledge about the production demand for a certain period our general approach follows the idea of backward planning. Based on the job's due dates the machine scheduling is made in retrograde order. This procedure enables a demand-oriented planning and reduces the usage of the flow shop storages, but it assumes the timely and quantitative sufficiently material provision in each stage. Furthermore, we consider distinct due dates, machine and family-dependent setup-times, intermediate storages as well as transport times to storages and machines. The setup-times occur batch sequence-dependent. We analyse their influence on the developed heuristics by varying their durations. The processing approach follows the batch availability model thus the jobs of a batch have the same starting and end time, and are available at once for subsequent steps. Required material are available in all stages at time zero thus the stage specific release dates only depend on the completion of the jobs on the previous machines and necessary transport times. Due to the combination of multiple batch processing stages and stage specific job families *inconsistent batches* occur. That means that the job constellation of batches only persist for single stages, i.e. batches are build for each machine. For a broader overview of flow shop characteristics we recommend (Potts and Kovalyov 2000).

M.-A. Isenberg (✉) · B. Scholz-Reiter
Planning and Control of Production Systems, University of Bremen,
Hochschulring 20, D 28359 Bremen, Germany
e-mail: ise@biba.uni-bremen.de

H.-J. Kreowski et al. (eds.), *Dynamics in Logistics*, Lecture Notes in Logistics,
DOI: 10.1007/978-3-642-35966-8_10,

The description of the $MBPM_{SIF}$ follows in section Introduction. Section Problem Statement describes the underlying principles and pseudo-codes of our heuristics. The evaluation of the algorithms is presented in section Experiments. A summary and outlook are part of section Summary and Outlook.

Problem Statement

The $MBPM_{SIF}$ describes non-permutation flow shops with multiple products, multiple stages and stage specific incompatible job families. Each stage has one batch processing machine (BPM). The BPM $k = 1,\ldots, m$ differ in their maximum job capacity $SMAX_k$ and are weakly restricted to a minimum batch size $SMIN_k$ for economic reason. Intermediate storages $s = 1,\ldots, x$ are located between as well as before the first and after the last BPM ($k = 1,\ k = m$). This induces $x = m+1$ storages. There is a high demand for products $p = 1,\ldots,w$. All jobs $j = 1,\ldots,n$ of the set J have to pass all x storages. The transport times from machine k to storage s takes $T_{k,s}$ and $T'_{s,k}$ vice versa. Transport times count for the whole batch. A batch $B_{b_{k,o}}$ out of $b_{k,o} = 1,\ldots,h_{\mathrm{k,o}}$ is located on machine k in time slot o. The corresponding duration $P_{b_{k,o}}$ is the sum of the processing times of the containing jobs. The jobs are partitioned in machine specific incompatible job families $a_k = 1,\ldots,g_k$. The number of families g_k is a stage specific consolidation number of the number of products. This assumption follows by considering real production scenarios where product variants have partially the same components or colours. Hence, jobs for different products can build common batches within certain stages. This reduces the number of different products to a stage specific number of different job families. The production time of a single job j on machine k starts at $\mathrm{r_{j,k}}$. Additionally, the jobs have a due date $\mathrm{Z_j}$ as well as a machine specific release date $R_{j,k}$ and deadline $D_{j,k}$. A batch on machine k precedes a setup-time $u_{a,k}$ if it is the first batch in the time slot o or if it belongs to another job family a_k as the previous batch. The time slots $o = 1,\ldots,q$ represent production periods such as working days. Their time capacity L_o are equal. We assume an infinite time horizon. The objective criterion is the minimisation of the mean flow time $n^{-1}\sum f_j$ which is equal to the total flow time $\sum f_j$. We define the flow time of a job j as the difference of its due date and its starting time on the first machine $f_j = Z_jj - r_{j,1}$. The starting point for calculations $\mathrm{Z_j}$ arises by the approach of backward planning. The later a job starts in its production, the shorter its flow time gets; thus enabling a reduction of the number of supplies by aggregating purchase order quantity. According to the notation of Graham et al. (1979) and Potts and Kovalyov (2000) we denote our model as $\mathrm{Fm/s_{ifg},b_i,d_j/\sum f_j}$. According to Garey et al. (1976) finding the minimum mean flow time schedule in a flowshop of $m \geq 2$ machines is NP-complete. If we take the job constellation for batches as a given, the challenge of $MBPM_{SIF}$ only consist of finding the best schedule. Hence, we assume that the $MBPM_{SIF}$ is also NP-complete. Since exact algorithms would only be capable to solve small instances we pursuit the development of heuristics. To our knowledge no approach exist for solving the $MBPM_{SIF}$. There are several approaches for single stages with incompatible job families [compare to Koh et al. (2005) and Yao et al. (2012)]. However, applying these approaches to the $MBPM_{SIF}$ would necessitate their implementation in each single stage. The stages would be planned separately without any consideration of stage across interdependencies.

Solution Approaches

In general, solution approaches have to comply two tasks: the batch formation and the batch sequencing. The batch formation selects the jobs for building a common batch. Batch sequencing schedules the batches on the machines considering their capacities. Since we apply backward planning the stage specific deadlines $D_{j,k}$ are the initial points for all methods generating batch formations and sequences. The calculation of $D_{j,k}$ is done in reverse order and works as follows:

$$\mathrm{D_{j,k}} = \begin{cases} r_{j,k+1} - T'_{s,k+1} - T_{k,s} & \text{if} \quad k < m \\ \mathrm{Z_j} - \mathrm{T_{k,s}} & \text{else} \end{cases} \quad \forall k, s = k \tag{1}$$

$$V_{b_{k,o}} = \min_{j \in B_{b_{k,o}}} \left\{ D_{j,k} \right\} \tag{2}$$

Equation (1) shows that the deadline of job j on machine k results by subtracting all transport times from machine $k + 1$. Since $x = m+1$ defines the number of storages in proportion to the number of machines, the relevant transport times causes from storage $s = k+1$ to machine $k = k+1$ and from machine k to storage $s = k+1$. Equation (2) describes the deadline of a batch $V_{b_{k,o}}$ which depends on the earliest deadline of the containing jobs. We develop two groups of heuristic methods. The first group consists of two methods which work in a sequence-oriented manner for batch formation. The second group also consists of two methods. However, those use a clustering technique.

All methods apply the same basic procedure for finding free production intervals on the machines. Each BPM k has a time table organising its sequencing. The procedure searches on the basis of this time table by considering the batch duration and a preferred date for its completion date on machine k. The search for an adequate free interval is done in a retrograde manner. In dependence of the surrounding batches of the free interval the methods checks if a setup is required. If the free interval does not allow an additional setup the method continues its search. According to the backward planning approach all methods start with $k = m$ and proceed to the previous machines. They terminate when they have processed $k = 1$.

Sequence-Based Approaches

The sequence-based approaches are divided into the Next-Fit-Decreasing (NFD) and the Proportional-Fit-Decreasing (PFD) algorithms. Both approaches sort the jobs by their stage specific deadlines in decreasing order. The batch formation process of the algorithms realise batch sizes of a given *formation-threshold*. For instance the formation-threshold can be $SMIN_k$ or $SMAX_k$.

Next-Fit-Decreasing

The simple and intuitive algorithm of NFD is illustrated in Table 1. According to Eq. (1) NFD firstly sorts the jobs by their deadlines on machine k in descending order. Subsequent it partitions the jobs into subsets by their families on k. Since the jobs are still in decreasing order the next step of batch formation (row 4–7 in Table 1) groups jobs which are timely close to each other. Finally, NFD sorts the batches first by their job families and last by their deadlines in decreasing order. The criterion combination induces that batches of the same deadline and family are next to each other. It reduces the mean flow time and decreases the number of setups.

Proportional-Fit-Decreasing

The idea of PFD bases on a proportional consideration of the job family occurrences in the deadline descended job sequence. The pseudo-code of the PFD is depicted in Table 2. After sorting the jobs by their deadlines the partitioning starts. In contrary to NFD, PFD does not divide the tasks of partitioning and batch formation; it integrates them into each other. At first, PFD allocates the job to the

Table 1 Pseudo-code of the Next-fit-decreasing algorithm

01	Start NFD for stage k with J, create empty batch list τ
02	Sorting$J$ according to deadline $D_{j,k}$ in descending order
03	Partitioning J according to job families in g_k subsets
04	For all subsets of $a_k = 1,\ldots,g_k$ do
05	Group jobs of subset a_k to batch $B_{b_{k,o}}$ according to *formation-threshold*
06	Remove grouped jobs from subset a_k
07	Add $B_{b_{k,o}}$ to τ
08	Descending sorting of τ by job family a_k and deadline $V_{b_{k,o}}$
09	For all batches $b_{k,o} = 1,\ldots,h_{k,o}$ of τ do
10	Reverse scheduling of $b_{k,o}$, preferred completion time is $V_{b_{k,o}}$
11	If $k > 1$ start NFD($k - 1$, J), else terminate

Table 2 Pseudo-code of the proportional-fit-decreasing algorithm

Start NFD for stage k with J, create set of occurrence counter Ω
Sorting J according to deadline $D_{j,k}$ in descending order
Start partitioning J according to job families in g_k subsets
  For all jobs $j = 1,\ldots,n$ do
    Allocate job j to the subset of its job family a_k
    Add position value $n - j$ to occurrence counter of a_k
    If subset of a_k achieves formation-threshold
      Calculate Q.75-threshold
      If a_k in $\Omega \geq Q$.75-threshold
        Group jobs of subset a_k to batch $B_{b_{k,o}}$ according to threshold
        Remove grouped jobs from subset a_k;
        Adjust occurrence counter of a_k;
        Reverse scheduling of $B_{b_{k,o}}$, preferred completion time is $V_{b_{k,o}}$
Batching + scheduling residual job of $a_k = 1$; g_k by descending order of Ω
If $k > 1$ start PFD($k - 1$, J), else terminate

subset of its family a_k. Following, it calculates the position value of the job by $n - j$. That means: the later the deadline of a job, the higher its position value. The position value is added to an occurrence counter related to the family a_k. If the number of the jobs in the subset of a_k achieves the formation-threshold PFD determines the $Q_{.75}$-*threshold* which describes the minimal occurrence value of the best quartile of all subsets. If the occurrence counter of a_k exceeds $Q_{.75}$-*threshold* a batch $B_{b_{k,o}}$ is formed and directly scheduled. Subsequent, the occurrence counter of a_k is adjusted to the size of the residual jobs in subset a_k. The simultaneously partitioning, batching and scheduling of PFD forms a permanent online-ranking with thresholds. It calculates in an ongoing process the significance of the job families in proportion to their occurrences thus enabling the consideration of changing priorities in the job sequence.

Cluster-Based Approaches

The cluster-based approaches are divided into the Next-Fit-Clustering algorithm (NFC) and the Best-Fit-Clustering algorithm (BFC). They work with the same hierarchical clustering technique of *complete-linkage clustering* [compare to Xu and Wunsch (2009)]. This is an agglomerative approach which starts with n clusters containing a single job. In each iteration the algorithm determines all distances between all clusters. Our measure of distance is the time difference of the deadlines of different jobs. The distance between the clusters X and Y is defined as

$$\delta\{X|Y\} = \max_{\substack{j_x \in X \\ j_y \in Y}} \left\{ \left| D_{j_x,k} - D_{j_y,k} \right| \right\}$$

That means that the distance of two clusters X and Y results by comparing all job deadlines of cluster X to all job deadlines of cluster Y and taking the maximum. The two clusters with the closest distance to each other are joined and a new iteration starts. The clustering procedure terminates if no clusters can be joined unless the new cluster would exceed $SMAX_k$. The advantage of applying this clustering technique is that batches finally contain the job combinations with the closest average distance to each other, i.e. smallest variance concerning the jobs deadlines. Additionally, we implement the option of a *trade-off function* for considering the jobs families in the previous process. The trade-off function is only regarded by initialising the distance matrix and reduces the distance of the clusters X and Y by 25 % if their jobs have the same job family at stage $k-1$. This leads to groups of jobs whose deadlines on machine k are close to each other *and* whose jobs are able to form a batch on $k-1$. This feature extends the clustering technique to a stage family-anticipating procedure and may help to reduce the mean flow time.

Next-Fit-Clustering

The NFC is a simple algorithm too. Its pseudo-code is depicted in Table 3. It starts with partitioning all jobs into family-specific subsets. All jobs of each single subset are clustered according to the complete-linkage technique described above. Subsequently, the algorithm transfers the clusters to batches, makes a descending sorting by their job families and batch deadlines and schedules them one by one.

Best-Fit-Clustering

The BFC meets the awareness of planning interdependencies between batch formation and batch sequencing [compare to Potts and Kovalyov (2000)]. Naturally,

Table 3 Pseudo-code of the next-fit-clustering algorithm

Start NFC for k with J, create empty batch list τ and empty cluster list ϕ
Partitioning J according to job families in g_k subsets
For all subsets $a_k = 1,\ldots, g_k$ do
    Build clusters for subset a_k by using complete-linkage clustering
    Add all clusters of a_k to ϕ
For all clusters $c = 1,\ldots, v$ of ϕ
    Create batch $B_{b_{k,o}}$, out of cluster c
    Add batch $B_{b_{k,o}}$ to τ
Descending sorting of τ by job family ak and deadline $V_{b_{k,o}}$
For all batches $b_{k,o} = 1,\ldots,h_{k,o}$ of τ do
    Reverse scheduling of $b_{k,o}$, preferred completion time is $V_{b_{k,o}}$
If $k > 1$ start NFC($k-1$, J), else terminate

the batch formation precedes the batch sequencing. However, by grouping jobs the batch formation only regards family-dependencies. But there is no consideration of the duration of free production intervals. Hence, batches might produced much closer to their deadlines if they would contain less jobs. Therefore, BFC adjust batch sizes to the duration of free production intervals, still regarding $SMIN_k$. The algorithm of BFC is illustrated as pseudo-code in Table 4. As all other algorithms do, BFC starts by partitioning the jobs into subsets according to their job family. A while loop follows which terminates if all subsets are empty. Within the loop the first steps are always the clustering, the transfer of clusters into batches and the descending sorting of batches by their deadlines. Subsequently, an iteration goes over the batches. The jobs of each batch $b_{k,o}$ are sorted by their deadlines in decreasing order. This helps in the following steps. The algorithm builds a minimal batch b_{min} according to $SMIN_k$ and determines its required production interval ρ

Table 4 Pseudo-code of the best-fit-clustering algorithm

Start BFC for k with J, create empty batch list τ + empty cluster list ϕ
Partitioning J according to job families in g_k subsets
While all subsets $\neq$ empty do
  Clear ϕ
  For all subsets $a_k = 1,\ldots,g_k$ do
    Build clusters for subset a_k by using complete-linkage clustering
    Add all clusters of a_k to ϕ
  Clear τ
  For all clusters $c = 1,\ldots, v$ of ϕ
    Create batch $B_{b_{k,o}}$ out of cluster c
    Add batch $B_{b_{k,o}}$ to τ
  Descending sorting of τ by job family a_k and deadline $V_{b_{k,o}}$
  Set *Reduction* = false
  For all batches $B_{b_{k,o}} = 1,\ldots,h_{k,o}$ of τ do
    Descending sorting of $b_{k,o}$ by deadline
    Determine minimal batch b_{min} of $b_{k,o}$ according to $SMIN_k$
    Calculate required production time ρ of b_{min}
    Determine deadline D_{min} of b_{min}
    Start anticipating schedule
      While no feasible free production interval
      Request for interval $\geq \rho$ and $\leq D_{min}$
      If interval $\geq \rho$ + *possible setup* terminate search
    If $P_{b_{k,o}}$ + *possible setup* $\geq$ *interval start reduction*
      Remove jobs from $b_{k,o}$ with $D_{j,k} \leq$ interval end or who do not fit
      Set *Reduction* = true
    Reverse scheduling of modified $b_{k,o}$, preferred completion time is $V_{b_{k,o}}$
      Reduction of corresponding subset a_k
      If *Reduction* = true break
If $k > 1$ start BFC($k - 1$, J), else terminate

(without a setup) as well as its deadline. Due to the sorting b_{min} consists of the latest jobs in term of their deadlines. Afterwards an anticipating step follows. Within this step BFC makes a reverse search for a free production interval in the time table of machine k which is longer than ρ. Additionally, the algorithm checks if the surrounding batches of the free production interval are of the same family or if a setup is required. If the free production interval covers the production of b_{min} and a potential setup it further checks if the original batch fits into the interval. If not, the algorithm reduces the original batch $b_{k,o}$ as less as necessary and sets the Boolean *Reduction* to true. The anticipating step ends, $b_{k,o}$ is scheduled and the subset of the job's family a_k is reduced by deducting the scheduled jobs. If a reduction was necessary BFC repeats the clustering step, otherwise it continues with the next batch.

Experiments

We implemented the algorithm of NFD, PFD, NFC and BFC in Java. For evaluating the algorithms we developed a test scenario framework for creating and testing the problem instances. All random numbers are generated by using the fast variant of the Mersenne Twister number generator [compare to Matsumoto and Nishimura (1998)].

Setup

The experimental setup for the generation of problem instances defines fixed values as well as potential intervals for certain attributes. In general, the flow shop consist of $m = 5$ stages, each containing a single BPM. We created scenarios of $p = 5, 10, 20,$ 30 and 40 products. Each scenario consist of $n = 1{,}000$ jobs, thus the product to job ratio differs between the product levels. The jobs belong to 50 different due dates which are randomly distributed within a range of 14 days. Since our backward planning occurs sufficiently in advance we want to mention that the present date is not relevant for our planning. The due dates always are on the end of the working day thus implying that the corresponding products will ship the next morning. The number of stage specific families g_k derives by multiplying the product level with a stage specific consolidation factor demonstrating the production of different product variants with partially the same components. The consolidation factor is randomly chosen within the interval 0.3–0.6. Two jobs of the same product type have the same families on all machines. However, there are several product types which belong to the same family a_k due to the consolidation of the number of product types to the number of stage specific families. The job processing time $P_{j,k}$ of all families among all machines are random numbers of the interval 5–20 min. The transport times $T_{k,s}$ and $T'_{s,k}$ from storages to machines and vice versa are all constant and take 5 min for the whole batch. Each machine has a random capacity $SMAX_k$ of the interval 8–12 jobs and a weak minimum bound $SMIN_k$ of 3–8 jobs. A time slot o has a capacity L_o of 16 h. If all these attributes are specified the problem instance is analysed by applying various levels of setup-times. As we defined, setups occur batch sequence-dependent. The duration of a setup depends on its family a_k. For calculating the family-specific setup-times we multiply the setup levels ua_k of 0, 5, 10, 20, 40 and 80 % with the time which the production of a batch of full size $SMAX_k$ of family a_k takes on machine k. That means that a batch of a family whose full sized batch would need the production time of 10 min would precede a setup-time of 0, 0.5, 1, 2, 4 or 8 min according to the specific setup level. Summarising the explanations each scenario consist of a certain product level, specified attributes and a certain setup level. For each product level we created 250 random instances with specified attributes. In a second step we vary each of these instances by the level of setup-time. That results in problem instances with equal attributes but different setup-times. Due to 5 product levels, 250 random instances and a variation of 6 setup levels we receive 7,500 different problem constellations. Table 5 summarises the experimental setup.

Basing on all problem instances we analysed all algorithms. The sequence-based approaches are tested by using the two different formation-thresholds of $SMIN_k$ and $SMAX_k$. Therefore, we denote the variants as NFD_{MIN}, NFD_{MAX}, PFD_{MIN} and PFD_{MAX}. The cluster-based approaches are analysed by regarding and omitting the option of the trade-off function. They are denoted as NFC, NFC_{TF}, BFC and BFC_{TF}.

Results and Evaluation

Table 6 presents the average performance measures of the mean flow time $n^{-1}\sum f_j$ (denoted as F) in days, the coefficient of variation CV_F of the mean flow time which is defined as the standard deviation σ_f divided by $Ø_F$, the makespan C_{max} in

Table 5 Setup of test scenarios

Meta data			
Number of instances (each comibination)	250	p	{5, 10, 20, 30, 40}
		u_{ak} (%)	{0, 5, 10, 20, 40, 80}
Constant attributes			
Number of due dates	50	m	5
$T_{k,s}$, $T'_{S,k}$ in minutes	5	n	1,000
L_O in hours	16		
Interval attributes			
$SMAX_k$ in jobs	$[8, 12] = \{SMAX_k \in \mathbb{N} \vert 8 < SMAX_k < 12\}$		
$SMIN_k$ in jobs	$[3, 8] = \{SMIN_k \in \mathbb{N} \vert 3 < SMIN_k < 8\}$		
g_k	$[0.3 * p, 0.6 * p] = \{g_k \in \mathbb{N} \vert 0.3 * p < g_k < 0.6 * p\}$		
$P_{j,k}$ of a_k in minutes	$[5,20] = \{P_{j,k} \in \mathbb{N} \vert 5 < P_{j,k} < 20\}$		
Z_j in days	$[0,14] = \{Z_j \in \mathbb{N} \vert 0 < Z_j < 14\}$		

days, the average usage of $SMAX_k$ (denoted as U) in percent and the average processing time in storages S_p waiting for the next machine for all algorithms variants and product level. S_p is determined by a sample rate of 10 min for each intermediate storage. Additionally, the results of the product level $p = 20$ of BFC_{TF} are exemplary mentioned. A superior performance of one algorithm to another is demonstrated by a reduction of F, C_{max} and S_p or by an increase of U.

Comparison of Sequence-Based Approaches

The comparison of the PFD_{MIN} algorithm to the NFD_{MIN} algorithm lead to an ambiguous result. While PFD_{MIN} performs better mean flow times for the product level of 10 (−0.36 %), 20 (−0.37 %) and 30 (−0.27 %), NFD_{MIN} dominates the product level of 5 (0.23 %) and 40 (0.10 %). However, all values differ in a small range. It is noticeable that PFD_{MIN} continuously perform worse in C_{max} and better in U than NFD_{MIN}. The inferiority of C_{max} increases with an advanced level of products (5P = 0.24 %, 10P = 0.40 %, 20P = 0.47 %, 30P = 0.81 %, 40P = 1,20 %). The superiority of U also increased with an advanced level of products (5P = 0.24 %, 10P = 0.40 %, 20P = 0.47 %, 30P = 0.81 %, 40P = 1.20 %). This may be related to the jobs per product ratio and provides an indication of a product level specific suitability of algorithms.

The comparing analyses of PFD_{MAX} and NFD_{MAX} shows similar results for C_{max} (5P = 0.14 %, 10P = 0.58 %, 20P = 0.80 %, 30P = 1.35 %, 40P = 1.64 %) and U (5P = 0.20 %, 10P = 0.86 %, 20P = 0.83 %, 30P = 0.97 %, 40P = 1.13 %). An interesting criteria is the coefficient of variation CV_F which is continuously better for PFD_{MAX} and decreases with an advanced product level (5P = −0.70 %, 10P = −2.86 %, 20P = −8.21 %, 30P = −11.88 %, 40P = −14.23 %). This significant improvement may influenced by the increased U which generally is on a higher level thus working as a lever. However, it is partially paradox that in spite of a decreased CV_F the makespan of PFD_{MAX} still underperform those of NFD_{MAX}.

Contrasting the effects of the threshold values of $SMIN_k$ and $SMAX_k$ shows considerable differences. Comparing NFD_{MAX} to NFD_{MIN} we determine in average a longer F of 18.39 % but an increased U of 47.19 %. This is a logical consequence of the higher threshold: bigger batch sizes cause longer mean flow times since we apply inconsistent batches and the batches contain jobs with a higher variation of stage specific release dates. Thus, jobs waiting longer in the intermediate storages until all jobs of their batches are available. An interesting effect is the relative improvement of NFD_{MAX} to NFD_{MIN} in terms of F with an increasing setup level (exemplary for $p = 20$: 0S = 64 %, 5S = 52.97 %, 10S = 42.08 %, 20S = 23.94 %, 40S = 3.21 %, 80S = −12.67 %). This is an observable trend for all product levels and has a linear-proportional slope. However, the trend has product level specific heights. It leads to a predominance of NFD_{MAX} in F for all product levels with the setup level of 80 %. Since this trend is also observable for the comparison of PFD_{MAX} to PFD_{MIN} (exemplary for $p = 20$:

Table 6 Average performance measures of solution approaches

$\varnothing_P$	NFD_{MIN}					PFD_{MIN}					NFD_{MAX}					
	F	CV_F	C_{max}	U	S_P	F	CV_F	C_{max}	U	S_p	F	CV_F	C_{max}	U	S_p	u_{ak}
$\varnothing_5$	4.30	0.48	21.23	50.36	278	4.31	0.48	21.28	50.55	278	5.03	0.43	21.92	99.38	319	
$\varnothing_{10}$	5.14	0.45	22.34	49.90	322	5.12	0.45	22.42	50.70	320	5.88	0.37	22.30	98.44	395	
$\varnothing_{20}$	6.38	0.42	23.98	49.39	405	6.36	0.41	24.09	50.29	403	7.33	0.35	23.34	96.83	546	
$\varnothing_{30}$	7.29	0.41	24.79	49.47	442	7.27	0.39	24.99	50.44	438	8.71	0.34	24.25	95.30	663	
$\varnothing_{4O}$	8.14	0.41	25.84	48.63	488	8.15	0.38	26.15	49.72	487	10.11	0.35	25.56	93.73	787	
$\varnothing$ all	6.25	0.43	23.63	49.55	387	6.24	0.42	23.79	50.34	385	7.41	0.37	23.47	96.74	542	
	PED_{MAX}					NFC					BFC					
$\varnothing_5$	5.02	0.42	21.95	99.58	318	4.60	0.55	21.27	85.11	311	4.05	0.52	20.59	81.17	275	
$\varnothing_{10}$	5.89	0.36	22.43	99.30	395	5.12	0.59	21.67	80.84	335	4.44	0.54	21.01	76.48	280	
$\varnothing_{20}$	7.34	0.32	23.52	97.66	545	5.86	0.58	22.57	77.82	388	5.19	0.53	21.89	73.10	339	
$\varnothing_{30}$	8.75	0.30	24.57	96.27	661	6.54	0.54	23.04	76.24	419	5.89	0.51	22.42	70.61	376	
$\varnothing_{4O}$	10.21	0.30	25.98	94.86	790	7.14	0.53	23.75	75.47	466	6.47	0.51	23.08	68.86	425	
$\varnothing$ all	7.44	0.34	23.69	97.53	542	5.85	0.56	22.46	79.10	384	5.21	0.52	21.80	74.04	339	
	NFC_{TF}					BFC_{TF}					$BFC_{TF,}p = 20$					
$\varnothing_5$	4.59	0.55	21.27	85.13	310	4.05	0.53	20.58	81.18	275	2.86	0.57	17.55	71.43	231	0
$\varnothing_{10}$	5.12	0.59	21.66	80.92	336	4.43	0.54	20.99	76.59	279	3.48	0.55	18.50	72.56	293	5
$\varnothing_{20}$	5.84	0.58	22.54	77.96	388	5.18	0.54	21.87	73.24	338	3.84	0.54	19.24	72.96	310	10
$\varnothing_{30}$	6.51	0.54	23.03	76.34	418	5.89	0.51	22.40	70.66	375	4.63	0.53	20.79	73.55	344	20
$\varnothing_{4O}$	7.12	0.53	23.70	75.61	464	6.46	0.51	23.05	68.97	424	6.33	0.51	24.12	74.21	394	40
$\varnothing$ all	5.84	0.56	22.44	79.19	383	5.20	0.52	21.78	74.13	338	9.94	0.50	31.00	74.72	454	80

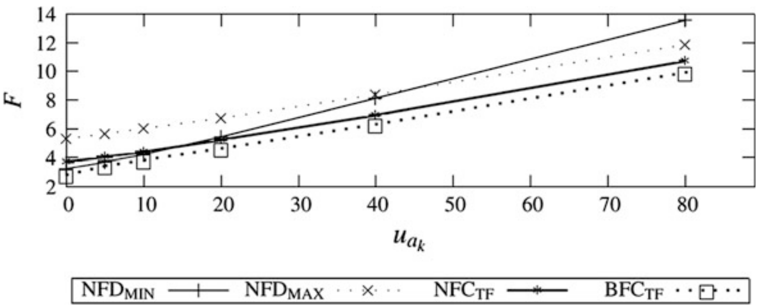


Fig. 1 Impact of the setup level on the mean flow time (p = 20)

0S = 63.80 %, 5S = 52.42 %, 10S = 42.08 %, 20S = 24.91 %, 40S = 3.85 %, 80S = −11.88 %) we reason that a bigger batch size compensates its disadvantage of longer waiting times in storages and longer mean flow times by reducing the average number of batches thus decreasing the absolute number of necessary setups. The breakeven point of setup-time is close to 40 %, little varying on the product level. Figure 1 illustrates the breakeven point by contrasting NFD_{MIN} to NFD_{MAX}, NFC_{TF} and BFC_{TF} in terms of F.

Comparison of Cluster-Based Approaches

The comparison of BFC to NFC shows a strong dominance of BFC. Except U all criteria are better for all products and setups to those of NFC. The inferiority of U results by adjusting the batch sizes to the durations of the free production intervals. In the average case BFC underlies NFC in U by −6.39 % and outperforms NFC in F (−10.95 %), CV_F (−6.11 %), C_{max} (−2.96 %) and S_p (−11.65 %).

The comparison of BFC_{TF} to NFC_{TF} results equal dominance properties and similar values. BFC_{TF} underlies NFC_{TF} in U by −6.39 % and outperforms NFC_{TF} in F (−10.88 %), CV_F (−5.94 %), C_{max} (−2.96 %) and S_p (−11.77 %).

Analysing the influence of the trade-off function we compare NFC_{TF} to NFC and BFC_{TF} to BFC. In average NFC_{TF} performs slightly better than NFC for the criteria F (−0.26 %), C_{max} (−0.09 %), U (0.1 %) and S_p (−0.17 %) and slightly worse for CV_F (0.09 %). However, there are divergent dominance properties within several product and setup levels thus this predominance is not valid for all cases. BFC_{TF} and BFC have similar proportions. Except for CV_F (0.27 %) all criteria F (−0.19 %), C_{max} (−0.09 %), U (0.08 %) and S_p (−0.30 %) are slightly better for BFC_{TF}.

Overall Comparison

Since BFC_{TF} seems to dominate the other cluster-based algorithms we only compare it to the algorithm variants of NFD and PFD. The comparison of the average mean flow time F shows that BFC_{TF} clearly outperforms NFD_{MAX} (−29.83 %), PFD_{MAX} (−30.12 %), NFD_{MIN} (−16.82 %) and PFD_{MIN} (−16.7 %). Since the average U of BFC_{TF} is below those of NFD_{MAX} (−22.61 %) and PFD_{MAX} (−23.40 %) this results confirm the general hypothesis of the impact of the batch size onto the mean flow time. However, BFC_{TF} still better perform in F than NFD_{MIN} and PFD_{MIN} which have considerable less U than BFC_{TF} (compare to Fig. 2). That shows that there is no mono-causality between the batch size and the mean flow time and that there are another important impact factors such as the procedure of batch formation. Since the approaches mainly differ in their way of batch formation these results illustrate the significance of the selection technique and emphasise the quality of the clustering approach. For almost all other criteria (total average) BFC_{TF} also dominates NFD_{MAX} ($C_{max} = -7.21$ %, $S_p = -37.62$ %), PFD_{MAX} ($C_{max} = -8.07$ %, $S_p = -37.60$ %), NFD_{MIN} ($C_{max} = -7.86$ %, $S_p = -12.64$ %) and PFD_{MIN} ($C_{max} = -8.45$ %, $S_p = -12.29$ %). However, BFC_{TF} has a considerable higher CV_F ($NFD_{MAX} = 42.92$ %, $PFD_{MAX} = 54$ %, $NFD_{MIN} = 20.71$ %, $PFD_{MIN} = 24.3$ %). Comparing all methods to NFD_{MIN} Fig. 2 depicts the inferior and superior criterias. This comparison is made by using the total averages over all problem instances.

There are some more general observations. The variation of the setup-time has no influence on U for all methods except for BFC_{TF}. An advanced setup level leads to a higher U applying BFC_{TF}. This effect is described in Table 6. Furthermore, increasing the number of products cause a reduction of U and CV_F and an increase of F, C_{max} and S_p for all approaches, independent from the level of setup. This may be related to the more extensive interdependencies arising from an advanced product level. A higher product level entails a higher number of different job families thus inducing smaller batch sizes, additional setups and finally the increase of F, C_{max} and S_p. Additionally, we observe a higher CV_F for the the

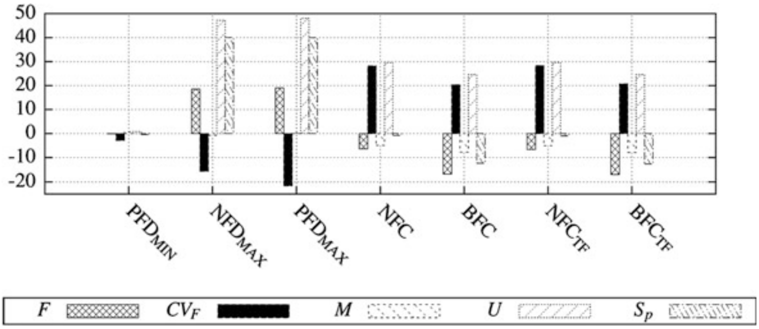


Fig. 2 Comparing the performance of all algorithms to NFDMIN (in %)

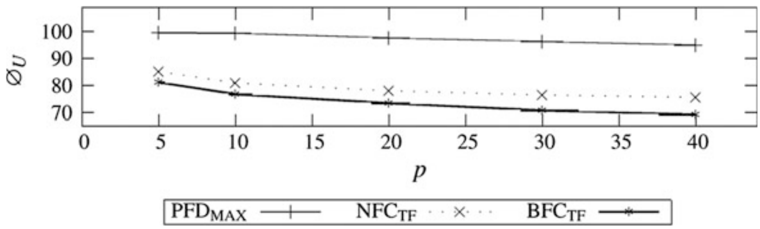


Fig. 3 Impact of the job/product-ratio on the average usage $SMAX_k$

cluster-based approaches. The effect of a decreasing U with an advanced product level is depicted in Fig. 3 for the algorithms PFD_{MAX}, NFC_{TF} and BFC_{TF}. It bases on the average U for all setup levels for each specific product level.

As Fig. 2 depicts there is no approach which is superior for all criteria for all product levels and for all setup levels. Following our objective of minimising the mean flow time, BFC_{TF} outperforms all other approaches. However, considering U we mention a relative inferiority of BFC_{TF} to other approaches. Although BFC_{TF} fulfils the requirement of $SMIN_k$ for all problem instances a further improvement is desirable. In addition, the evaluation shows that the approaches have different suitabilities for different setup levels (compare to Fig. 1). Further research should intensify the analyses for the corresponding reasons of the setup level dependent superiorities for identifying key factors. Moreover, future research should develop methods which do not base on greedy procedures as the developed ones do. Greedy algorithms enable adequate results with low computation effort; but potentially their results are only local optima.

Summary and Outlook

This contribution introduces and evaluates four heuristics each with two variants for the $MBPM_{SIF}$. According to the objective of minimising the mean flow time BFC_{TF} outperforms all other heuristic variants. Considering the average batch size BFC_{TF} partially underperforms but still fulfils the weak restriction of $SMIN_k$. In addition, we analyse that the batch formation procedure itself has a significant impact on the mean flow time. The increase of the product level also extends the mean flow time by increasing the stage interdependencies. Varying the setup levels shows that break even points exists where bigger batch sizes do not prolong flow times. Due to their simple architecture and low computation effort the developed heuristics are able to serve as basis methods within meta-heuristics such as Genetic Algorithms or Simulated Annealing. Hence, further research may be effected on the development of such meta-heuristics as well as their comparison with a lower bound, calculated by a linear solver. Furthermore, we only considered job availability at time zero. Applying uncertainty regarding the material provision at the stages necessitates the development of reacting methods.

References

Garey MR, Johnson DS, Sethi R (1976) The complexity of flowshop and job-shop scheduling. Math Oper Res 1(2):117–129

Graham RL, Lawler EL, Lenstra JK, Kan Rinnooy A (1979) Optimization and Approximation in Deterministic Sequencing and Scheduling: a Survey. Ann Discrete Math, vol 5. Elsevier, pp 287–326

Koh S, Koo P, Kim D, Hur W (2005) Scheduling a single batch processing machine with arbitrary job sizes and incompatible job families. Int J Prod Econ 98(1):81–96

Matsumoto M, Nishimura T (1998) Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. ACM Trans Model Comput Simulat 8(1):3–30

Potts C, Kovalyov M (2000) Scheduling with batching: a review. Eur J Oper Res 120(2):228–249

Xu R, Wunsch DC (2009) Clustering. IEEE series on computational intelligence, Wiley

Yao S, Jiang Z, Li N (2012) Abranch and bound algorithm for minimizing total completion time on a single batch machine with incompatible job families and dynamic arrivals. Comput Oper Res 39(5):939–951

A Comparison of Network Characteristics in Metabolic and Manufacturing Systems

Till Becker, Mirja Meyer, Moritz E. Beber, Katja Windt and Marc-Thorsten Hütt

Abstract Both metabolic and manufacturing systems face fluctuating environmental influences and thus share the common challenge to maintain a high level of efficiency for a variety of different conditions. Therefore, transferring methods used for analyzing one of the systems can lead to gaining new insights in the other. Following-up on previous findings on analogies in metabolic and manufacturing systems, our approach now is to analyze and compare complex network measures such as centrality or flow activity in both systems to identify quantified relations. The results show that both systems also display distinct statistical differences in addition to their various structural similarities.

Introduction

The metabolism of a cell and the material flow network of a manufacturing system are both faced with highly variable environmental influences, such as fluctuating input factors or disturbances within the system. Thus, they share comparable challenges: they have to efficiently and sustainably cope with uncertain and varying system conditions while also displaying a high performance under normal circumstances. Therefore, exploring parallels between metabolic and manufacturing systems seems promising, as analysis or control methods that are applicable in one of the two systems can be suitable for the other one as well.

Some analogies of metabolic and manufacturing systems have already been described in the past, such as parallels between certain objects in the systems (e.g.,

T. Becker (✉) · M. Meyer · M. E. Beber · K. Windt · M.-T. Hütt
Jacobs University, Campus Ring 1, 28759 Bremen, Germany
e-mail: t.becker@jacobs-university.de

H.-J. Kreowski et al. (eds.), *Dynamics in Logistics*, Lecture Notes in Logistics,
DOI: 10.1007/978-3-642-35966-8_11,

enzymes in metabolism and machines in manufacturing) (Tharumarajah et al. 1998) or the process of evolutionary optimization in both systems (Armbruster et al. 2005). In a previous approach, we have enhanced those ideas by pointing out further similarities, such as in network topology, system dynamics and flow control (Becker et al. 2011). Although the analogies between the two systems seem obvious, approaches so far have stayed on a rather descriptive level, and no quantifiable relation in system behavior of metabolic and manufacturing systems has been revealed yet.

Therefore, in this paper we seek to give a quantitative statement on parallels between metabolic and manufacturing systems by applying and comparing complex network measures to a dataset of a metabolic and a manufacturing network. This allows us, on the one hand, to verify the validity of the existing qualitative analogies between the two system types. On the other hand, this analysis can help finding improved methods for the analysis of networks in manufacturing systems.

The paper is structured as follows. The second section will thus give a description of the analogies between metabolic and manufacturing systems as yet proposed. In the third section, different complex network measures and their application in metabolic and manufacturing systems will be presented. The fourth section illustrates and investigates the results of our comparison of complex network measures in metabolic and manufacturing systems. A discussion and conclusion is then given in the fifth section.

Analogies in Metabolic and Manufacturing Systems

Metabolic and manufacturing systems share obvious commonalities regarding their structure and their functions. Metabolism can be seen as a system dealing with transportation, decomposition, and production of compounds. Helbing et al. (2009) claim that logistics as the organization, coordination, and optimization of material flows is an omnipresent characteristic of biological systems. This similarity has often been the trigger for the development of bio-inspired approaches in logistics and manufacturing. In logistics, methods for transport network design using fungal networks (Bebber et al. 2007; Tero et al. 2010) or optimization methods for logistics processes such as vehicle routing based on ant algorithms (Bell and McMullen 2004) have been proposed. In spite of the high number of bio-inspired approaches in logistics, there have been few investigations so far which attempt to unravel the fundamental mechanisms that enable manufacturing systems to benefit from biological structures. First works point out the structural similarities between biological and manufacturing systems and their components (e.g., cells and production units) (Tharumarajah et al. 1998) as well as similarities in control between the two systems (Ueda et al. 1997).

Our previous work includes a qualitative synopsis of matching elements between metabolic and manufacturing systems on the levels of network topology, network elements, flow organization, and system dynamics (Becker et al. 2011).

The *network topology* of material flow networks in manufacturing systems can be modeled by a directed graph, which consists of *network elements*, in particular of nodes representing machines or assembly stations and (weighted) links representing the corresponding material flow between the machines. Similarly, metabolism as a complete set of biochemical reactions within an organism (e.g., a cell) can be seen as a sequence of transformations of substrates into products. The interactions between substrates, reactions, and products can then represented as a network. The substrates and products correspond to the raw material and finished products in a manufacturing system. Most reactions in a metabolism are enabled through the catalytic activity of enzymes, which correspond to machines in manufacturing.

Flow control in both network types depends on the structure and plasticity of the network. Although the control mechanisms themselves are rather different, there are similar layers of control: on the one hand, there are global 'strategies', represented by a production plan in manufacturing systems and by external signals that control metabolic functions. On the other hand, there are local feedback mechanisms, e.g., dispatching rules in manufacturing and concentration-triggered suppression or stimulation of reactions.

The *system dynamics* layer depicts the actual routing functions within a material flow or metabolic network. This routing happens at each node in the system and determines the subsequent path of the flow elements through the network. In manufacturing systems, the system dynamics are influenced by lot sizes, setup times, and technological restrictions. Metabolic structures are more complex, as metabolic regulation is controlled by gene regulation, covalent enzyme modifications, and regulation of the enzyme through non-covalent binding with other molecules.

Complex Network Theory and Measures

Research on complex systems using approaches from graph theory or statistical mechanics (Albert and Barabási 2002) is a well-established scientific field. Different types of networks, such as random graphs (Erdös and Renyi 1959), small-world (Watts and Strogatz 1998), or scale-free networks (Barabási and Albert 1999) have been identified and analyzed. Therefore, a large variety of network measures regarding topology and structure of networks (Costa et al. 2007; Ziv et al. 2005) has been defined.

As a considerable number of real world networks, such as social or biological networks, implicate the characteristics of complex systems, a wide range of scientific disciplines focus on the application of findings from complex network theory to real networks in order to understand and predict system behavior. So far, many different system types, such as communication (Albert et al. 1999), biological (Barabási and Oltvai 2004), or social networks (Newman and Park 2003), have been analyzed using complex network measures.

The study of metabolic networks from a graph-theoretical point of view has been a search for specific characteristics that can be attributed to principles of evolution. In this context, degree distribution (Jeong et al. 2000), modularity (Hartwell et al. 1999), hierarchy (Ravasz et al. 2002), and architectural robustness (Giaever et al. 2002; Papin and Palsson 2004) have been investigated and a very characteristic non-randomness of metabolic systems has been established. Despite the success of topological analyses the *true* features of metabolic networks are a matter of ongoing debate (Montañez et al. 2010), and a connection to evolutionary design principles is difficult (Papp et al. 2009).

In manufacturing, complexity poses an increasing challenge: manufacturing systems tend to become larger (i.e., increase the amount of elements) and highly connected. Therefore research in manufacturing focuses on finding ways to describe, measure, and manage complexity, in order to competitively deal with it (Papakostas et al. 2009; Vrabič and Butala 2011). First applications of complexity measures, such as degree distributions or clustering coefficients, exist for material flow networks in logistics (Hammel et al. 2008; Peters et al. 2008). Thus, complex networks measures seem as an ideal common analysis method for the comparison of metabolic and manufacturing systems on a structural level.

Comparison of Network Measures in Metabolic and Manufacturing Systems

In this section, we analyze and compare a metabolic and a manufacturing network using different complex network measures. Here, we use a network representation of all reactions occurring in *Escherichia coli* K-12 metabolism. Currency metabolites (energy carrier molecules) were manually removed to reveal the underlying disparate structure (Ma and Zeng 2003). The resulting bipartite network (bipartite, as it contains both metabolite and reaction nodes) was then projected onto the reactions nodes. The resulting network contains only reactions nodes and they are linked if they share a common metabolite. The metabolic fluxes were computed using a flux balance approach (Varma and Palsson 1994) on the latest model *E. coli* metabolism (Feist et al. 2007).

The manufacturing network is based on a job shop production system of a tool manufacturer. The product range comprises around 5,000 different variants and the system consists of around 300 workstations. To depict this system as a graph, the nodes represent the workstations, whereas the links represent the material flow between the work stations. The amount of material flow was derived from a data set of feedback data from one year of production in the job shop.

Firstly, we compare the in-degree distributions in both networks. The degree of a node indicates its connectivity within the network. The out-degree and the total degree of the nodes are similarly distributed for both networks, so we only show the in-degree, which is most relevant for the dynamic data. Figure 1 illustrates the

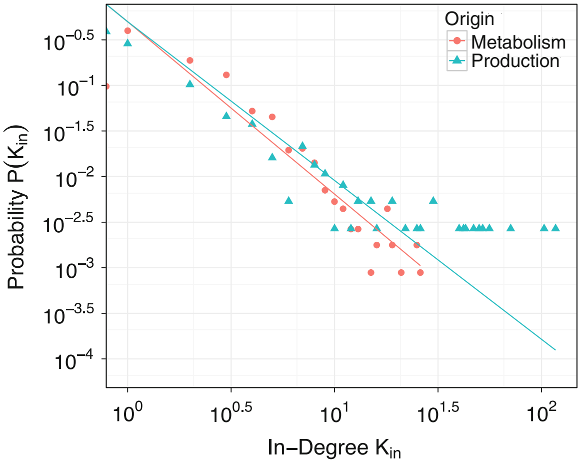


Fig. 1 In-degree distribution of the metabolic and the manufacturing network. The data was fit with a power-law using maximum likelihood estimation. The power-law coefficient alpha was determined to be 1.742 with a 95 % confidence interval of 1.644–1.848 for the production network and 1.891 with an interval of 1.835–1.951 for the metabolic network

in-degree distribution for the metabolic and the manufacturing network. The distributions clearly show that both networks are characterized by a small number of nodes with high degrees, whereas the majority of nodes have only a few links connected to them. This points to the existence of "hub" nodes in the network having a central task or a gateway function.

The second measure regarding network topology is the betweenness centrality distribution. The betweenness centrality B_k, of a node k, is defined as the sum over the number of shortest paths between all pairs of nodes i and j that pass through node k divided by the total number of shortest paths between i and j (Freeman 1977). It similarly quantifies the importance of the respective node in the system. Usually, there is a significant correlation between node degree and betweenness centrality in a network, which is the case for the networks considered here. As one would expect, the betweenness centrality values of the metabolic and the manufacturing networks indicate that there is a limited number of central nodes, while the majority of nodes is of moderate centrality (see Fig. 2). Therefore, we conclude that, although both networks originate from two particular different domains, their topology in terms of a connectivity pattern is to a high extent alike. In order to question this likeness, we produced random ensembles for both networks with an edge-switch algorithm (Milo et al. 2002). The deviation from randomness is apparent but cannot be adequately systematized.

Thirdly, we compare the distribution of active nodes, in order to determine how the activity of flows is distributed among the nodes in the networks in contrast to the topological key figures presented above. The metabolic system has been simulated in a variety of environments. For each node we have counted the number of environments in which this particular node has been active. As the manufacturing system does not operate in different environments, we have alternatively separated the manufacturing data in periods of time of equal length. These periods can serve as surrogates for environments, because each period comprises distinct demand represented by an individual set of production orders. After analyzing

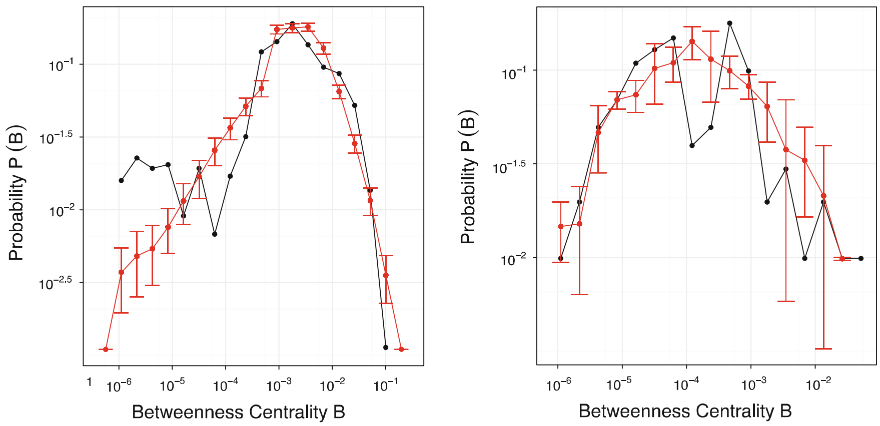
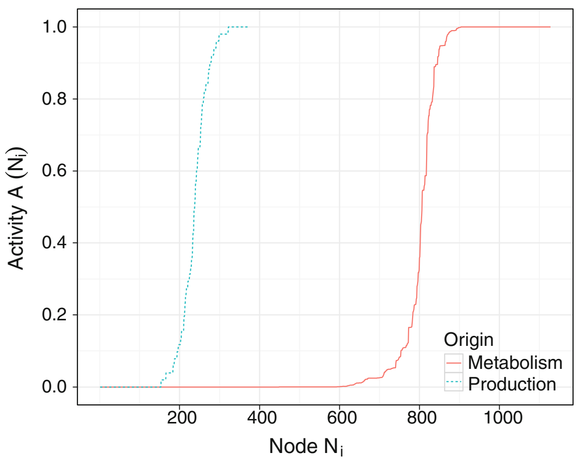


Fig. 2 Betweenness centrality distribution in the metabolic (*left*) and the manufacturing network (*right*). Comparison of the network in *black* with an ensemble of randomized networks in *red*

different period lengths, we decided to use a length of seven days, resulting in 51 complete periods. All alternative period lengths (except extreme values like one day or 365 days) basically show a similar pattern of distribution. We picked one week as an appropriate period, because it is able to represent a common, yet individual set of processing orders in this specific manufacturing company.

Figure 3 illustrates the activity distribution of the metabolic and the manufacturing network. The general shapes of the two curves are fairly similar and of sigmoidal shape. The curve of the metabolic system shows a long plateau of highly active nodes and a rather long tail of seldom-active nodes, which means that there are a high number of standard reactions as well as extremely specialized reactions, whereas only few reactions show intermediate activity. Compared to the metabolic

Fig. 3 Distribution of active nodes in the metabolic network (for different media) and in the manufacturing network (for different time intervals)

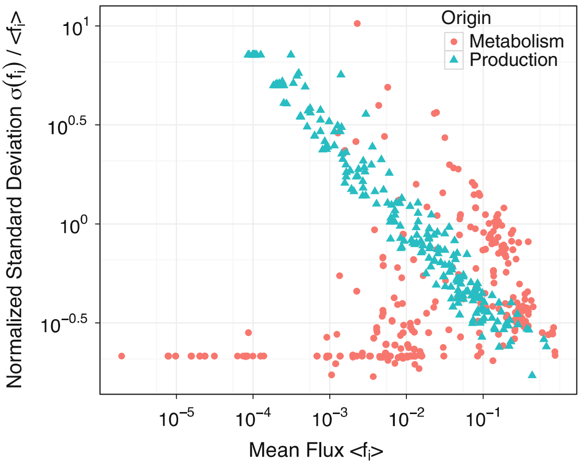


system, the manufacturing system seems slightly less specialized. The tail indicates that a considerable amount of machines is not active most of the time, meaning that a higher amount of nodes is active for the majority of environments (i.e., for most of the production orders carried out). However, as mentioned above, the shape of the curve depends to a certain extent on the sampling of environmental conditions and, in the case of the industrial network, on the time discretization.

The fourth measure to be compared is the mean throughput per node in relation to its standard deviation, normalized by the mean throughput. We chose a logarithmic scale for analyzing the flows in the networks as we are focusing on smaller flows and activity (see Fig. 4). For the metabolic system, there is a tendency that flows have a higher relative variation in general, whereas the variations of the production system seem to follow a more distinct pattern: the majority of nodes have a small flux yet a high variation. We assume that there is a focus on the stability of high fluxes while smaller fluxes show stronger fluctuations.

Finally, we want to analyze the interaction of flow with network topology. Therefore, Fig. 5 depicts the betweenness centrality of a node depending on the mean flux. This enables us to check whether there is a relation between the importance of a node in terms of throughput and its topological position in the network. Although we saw in the beginning that on a topological level, hub nodes exist, it now looks as if these hubs are not that clearly visible in terms of flow intensity. Seemingly, there is no correlation between flux and centrality in the metabolic network. However, the manufacturing network shows a slight relation between centrality and flux for higher centrality values.

Fig. 4 Mean throughput per node in relation to the standard deviation normalized by the mean throughput in the metabolic and the manufacturing network. The distribution of the fluxes in the production network follows a tight power-law (not shown)



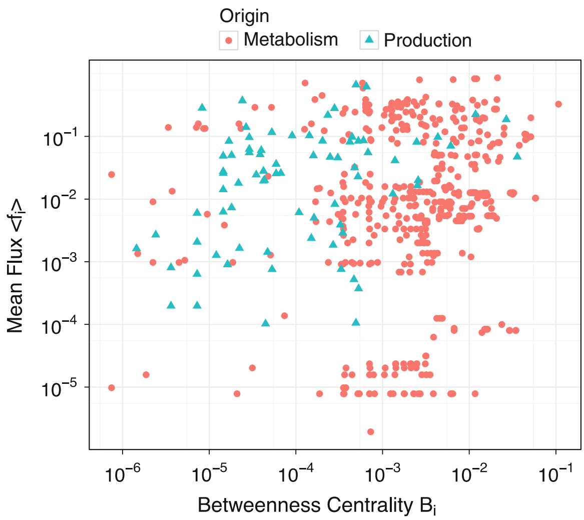


Fig. 5 Mean throughput per node in relation to its betweenness centrality

Conclusion and Outlook

In this paper, we have analyzed and compared different network measures for data from a metabolic and a manufacturing network. Following-up on other approaches that compare biological and engineered systems on a network level (6), we use networks as a common language that allows for the comparison of abstract measures between systems with fundamentally different purpose. We show for the first time that technical and biological 'production networks', in addition to their many structural similarities also display distinct statistical differences. Strikingly, the differences identified here concern the dynamics of material flow in these networks. In particular, the variation of intermediate-sized and very large fluxes is systematically suppressed (i.e., evolutionarily controlled) in the metabolic network, compared to the manufacturing network. Also, in the manufacturing network the mean flow through a node is coupled to the node's betweenness centrality, which is not observed in the metabolic network. We see further potential in the analysis of such dynamic aspects, yet future approaches will also have to focus on a larger-scale analysis based on datasets of several metabolic and manufacturing networks.

Acknowledgments We thank Nikolaus Sonnenschein (UC San Diego) for the simulation data of metabolic fluxes and for providing a curated metabolic network structure. The research of Katja Windt is supported by the Alfried Krupp Prize for Young University Teachers of the Krupp Foundation. Marc Hütt acknowledges support from Deutsche Forschungsgemeinschaft (grant HU-937/6).

References

Albert R, Barabási A-L (2002) Statistical mechanics of complex networks. Rev Mod Phys 74:47–97

Albert R, Jeong H, Barabási A-L (1999) Diameter of the world-wide web. Nature 401:130–131

Armbruster D, Degond P, Ringhofer C (2005) Continuum models for interacting machines. In: Armbruster D, Kaneko K, Mikhailov A (eds) Networks of interacting machines. World Scientific Publishing, Singapore

Barabási A-L, Albert R (1999) Emergence of scaling in random networks. Science 286:509–512

Barabási A-L, Oltvai ZN (2004) Network biology: understanding the cell's functional organization. Nat Rev Genet 5:101–113

Bebber DP, Hynes J, Darrah PR, Boddy L, Fricker M (2007) Biological solutions to transport network design. Proc R Soc B 274:2307–2315

Becker T, Beber ME, Windt K, Hütt M-T, Helbing D (2011) Flow control by periodic devices: a unifying language for the description of traffic, production, and metabolic systems. J Stat Mech: Theory Exp 2011:P05004

Bell JE, McMullen PR (2004) Ant colony optimization techniques for the vehicle routing problem. Adv Eng Inform 8:41–48

Costa LDF, Rodrigues FA, Travieso G, Boas PRV (2007) Characterization of complex networks: a survey of measurements. Adv Phys 56:167–242

Erdös P, Renyi A (1959) On random graphs, Publicationes Mathematicae (Debrecen). 6:290—297

Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, Karp PD, Broadbelt LJ, Hatzimanikatis V, Palsson BØ (2007) A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. Molecular Systems Biology 3 (June 2007)

Freeman LC (1977) A set of measures of centrality based on betweenness. Sociometry 40:35–41

Giaever G, Chu AM, Ni L, Connelly C, Riles L, Veronneau S, Dow S, Lucau-Danila A, Anderson K, Andre B et al (2002) Functional profiling of the Saccharomyces cerevisiae genome. Nature 418:387–391

Hammel C, Flemming A, Schulze F, Peters K (2008) Anwendung von Methoden aus der Theorie Komplexer Netzwerke für die Optimierung der Layouts von MFS, Technische Universität Chemitz (Hrsg.): 4. Fachkolloquium der WGTL, pp. 81–91

Hartwell LH, Hopfield JJ, Leibler S, Murray AW (1999) From molecular to modular cell biology. Nature 402:C47–C52

Helbing D, Deutsch A, Diez S, Peters K, Kalaidzidis Y, Padberg-Gehle K, Lämmer S, Johansson A, Breier G, Schulze F et al (2009) Biologistics and the struggle for efficiency: concepts and perspectives. Adv Complex Syst 12:533–548

Jeong H, Tombor B, Albert R, Oltvai ZN, Barabasi A-L (2000) The large-scale organization of metabolic networks. Nature 407(6804):651–654 October 5, 2000

Ma H, Zeng A-P (2003) Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms. Bioinformatics 19(2):270–277 January 22, 2003

Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U (2002) Network motifs: simple building blocks of complex networks. Science 298(5594):824–827 October 2002

Montañez R, Medina MA, Solé RV, Rodríguez-Caso C (2010) When metabolism meets topology: reconciling metabolite and reaction networks. BioEssays: News Rev Mol, Cell Dev Biol 32(3):246–256 March 2010

Newman MEJ, Park J (2003) Why social networks are different from other types of networks. Phys Rev E 68:036122-1–036122-8

Papakostas N, Efthymiou K, Mourtzis D, Chryssolouris G (2009) Modelling the complexity of manufacturing systems using nonlinear dynamics approaches. CIRP Ann Manufact Technol 58:437–440

Papin JA, Palsson BO (2004) Topological analysis of mass-balanced signaling networks: a framework to obtain network properties including crosstalk. J Theor Biol 227:283–297

Papp B, Teusink B, Notebaart RA (2009) A critical view of metabolic network adaptations. HFSP J 3(1):24

Peters K, Seidel T, Lämmer S, Helbing D (2008) Logistics networks: coping with nonlinearity and complexity. In: Helbing D (ed.): Managing complexity: insights, concepts, applications, Springer

Ravasz E, Somera AL, Monaru DA, Oltvai ZN, Barabási A-L (2002) Hierarchical organization of modularity in metabolic networks. Science 297(5586):1551–1555 August 2002

Tero A, Takagi S, Saigusa T, Ito K, Bebber DP, Fricker MD, Yumiki K, Kobayashi R, Nakagaki T (2010) Rules for biologically inspired network design. Sciene 327:439–442

Tharumarajah A, Wells AJ, Nemes L (1998) Comparison of emerging manufacturing concepts. In: SMC'98 Conference. Proceedings of 1998 IEEE international conference on systems, man, and cybernetics, pp. 325–331

Ueda K, Vaario J, Ohkura K (1997) Modelling of biological manufacturing systems for dynamic reconfiguration. CIRP Ann Manufact Technol 46:343–346

Varma A, Palsson BØ (1994) Metabolic flux balancing: basic concepts, scientific and practical use. Nat Biotech 12(10):994–998 October 1994

Vrabič R, Butala P (2011) Computational mechanics approach to managing complexity in manufacturing systems. CIRP Ann Manufact Technol 60:503–506

Watts DJ, Strogatz SH (1998) Collective dynamics of 'small-world' networks. Nature 393:440–442

Ziv E, Koytcheff R, Middendorf M, Wiggins C (2005) Systematic identification of statistically significant network measures. Phys Rev 71:1–8

Part IV
Modeling, Simulation, Optimization and Collaboration

Conflicting Optimization Goals in Manufacturing Networks: A Statistical Analysis Based on an Idealized Discrete-Event Model

Reik V. Donner and Jörg Hanke

Abstract Performance optimization is a crucial issue in present-day manufacturing systems. Here, we reconsider the problem of conflicting optimization goals, with a focus on low inventory levels and short throughput times. Based on an idealized discrete-event model of a complex small-scale manufacturing network, the impact of different production strategies as well as order policies on both quantities is carefully examined and systematically compared. Qualitative similarities of different scenarios regarding inventory levels and throughput times are investigated in detail by means of cluster analysis. Our results provide new insights into the influence of different key parameters on the complex material flow dynamics in manufacturing networks, and their reflection in different optimization goals.

Introduction

The economic success of present-day manufacturers is largely determined by their ability to meet omnipresent changes of market conditions resulting from the successive shortening of innovation and product life cycles, the development of new technologies, changing demands for specific products and services and, last but not least, the globalization of economy resulting in the emergence of new competitors. As a result, in the last decades manufacturers have often reacted to accelerating changes of market conditions by a successive specialization of goods and services,

R. V. Donner (✉)
Research Domain IV—Transdisciplinary Concepts and Methods, Potsdam Institute for Climate Impact Research, Telegrafenberg A31, 14473 Potsdam, Germany
e-mail: reik.donner@pik-potsdam.de

R. V. Donner · J. Hanke
Institute for Transport and Economics, Dresden University of Technology, Würzburger Str. 35, 01187 Dresden, Germany

H.-J. Kreowski et al. (eds.), *Dynamics in Logistics*, Lecture Notes in Logistics,
DOI: 10.1007/978-3-642-35966-8_12,

leading to outsourcing of secondary tasks and, consequently, the emergence of networks with considerable structural complexity (Kuzgunkaya and ElMaraghy 2006). Besides these observations for present-day production and logistics networks, the mathematical theory of complex networks has provided many examples that in general, the structure of networks constraints the resulting dynamics (i.e., complex structures are typically accompanied by complex dynamics) (Boccaletti et al. 2006). Specifically, like comparable material flow systems in biology (Gross and Blasius 2008; Gross and Sayama 2009), production and logistics are characterized by a co-evolution of structure and dynamics which is adaptive with respect to changing "environmental" conditions (Pathak et al. 2007).

The multiplicity of factors relevant for the successful operation of manufacturing systems implies the need for flexible and cost-minimizing strategies for organizing logistics processes in production and distribution. So far, these different aspects have typically been optimized separately by applying tools from operations research. However, traditional approaches from this field often lack a sufficiently holistic view on the mutual interdependences of logistics processes at different levels, and classical optimization tools have often reached their limits due to the complexity of both structure and resulting dynamics. One possibility to meet the latter problem is applying local instead of global strategies for production planning and control, giving the individual production units a certain degree of autonomy for their actions (Hülsmann and Windt 2010; Hülsmann et al. 2011). However, one general problem persists: there is a persistent lack of knowledge on how possible strategies influence the different characteristics of logistics performance to be considered in the optimization of the whole manufacturing process.

This contribution systematically addresses the problem of potentially conflicting goals of performance optimization in production logistics by comparing the influence of changing conditions on both inventory levels and throughput times as two particularly relevant quantities. For this purpose, an idealized discrete-event model of a closely interwoven manufacturing network is used for simulating the effect of different production strategies as well as order policies. In section Optimization Goals in Production Logistics, a general discussion of conflicting optimization goals is provided, followed by a brief description of our simulation model given in section Description of the Model. The results of our investigations are statistically analyzed by means of cluster analysis. The most relevant findings of this analysis are described in section Results. Finally, some conclusions regarding the optimization of real-world manufacturing processes are summarized.

Optimization Goals in Production Logistics

The general purpose of logistics systems is to make a given quantity of a certain product available at a predetermined time and place with a given quality. In order to achieve this goal for a multiplicity of production processes at the same time, a careful operation of the material handling and storage systems as well as transportation and production processes is necessary. Associated with this general statement, logistics operations involve various different types of costs, e.g., for storage, machining, and transportation, the sum of which has to be minimized for an economically successful operation. The variety of different economic factors is represented by a set of production-related key figures (Nyhuis and Wiendahl 1999) including inventories, throughput times, adherence to due dates, and machine usage (Fig. 1).

In general, the aforementioned key figures are not mutually independent. For example, minimization of throughput times requires low machines usage rates and, hence, vanishing queues in front of production units. As a result, the manufacturing process becomes transparent (e.g., individual jobs can be easily located in and, hence, directed through the production network) and flexible, storage costs are negligible. However, a low machine usage implies that the available infrastructure is not optimally used (i.e., the available production capacity is considerably higher than the actually used one), which implies high fixed costs per unit. Thus, exclusively focusing on low throughput times can become a disadvantage since it may reduce the manufacturer's economic fitness compared with possible competitors. On the other hand, with a higher capacity usage, queues emerge (i.e., the inventory and, hence, the amount of stored capital is rising) leading to higher throughput times. Even though buffered stocks of material or semi-finished products imply delays and additional costs, they can in fact be important for balancing intermittent variations in demand and supply that would be even more costly otherwise. These exemplary considerations demonstrate that optimizing production processes is an extremely complex task with a multiplicity of conflicting goals to be taken into account.

There is a variety of methods for obtaining a proper balance between the different key figures. For example, a careful analysis of different types of material involved in the production process allows giving individual priorities to each type, which can be considered in the ordering and storage processes. In general, one can distinguish two general approaches for ordering resources at a proper time in a proper quantity: demand-driven orders (based on the production plan for a given future period of time) and usage-driven orders (based on the actual usage of resources in a previous period of

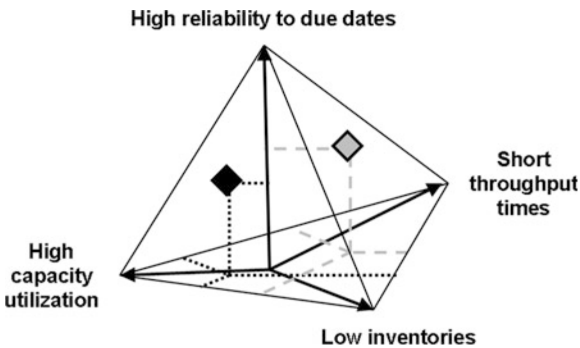


Fig. 1 Schematic representation of the four main conflicting optimization goals in production logistics as axes of a 3-simplex (*tetrahedron*). The diamonds represent two example choices of relative priorities of the different goals (represented by distances from the respective opposite faces) with a focus on high capacity utilization (*black*) and short throughput times with low inventories (*gray*)

time). The latter approach includes different types of order policies distinguishable by whether or not the order periods and volumes are fixed. Specifically,

1. order point policies continuously compare the actual stocks with a critical minimum value necessary to maintain the production process (i.e., the order intervals are variable)—a special case are provision policies where this critical value is zero;
2. periodic order policies are characterized by fixed order intervals, but possibly variable order volumes;
3. periodic control policies combine the main ideas behind order point and periodic order policies.

For a detailed discussion of the properties of these different strategies, we refer to (Daganzo 2005) or other related textbooks.

Description of the Model

The previous general considerations call for further scientific investigations in order to improve our present-day knowledge on the effect of different control strategies on the multiplicity of potentially conflicting optimization goals in production logistics. For this purpose, in the following we will study a generic setting inspired by a real-world production and logistics network from a European manufacturing company (Fig. 2a) with factories in different countries which maintain a continuous material flow between each other (Windt 2002). Specifically, different factories are specialized on fabricating distinct goods that are needed by others for finalizing their products. This setting can be considered as an example for complex manufacturing networks without explicit hierarchy, which are typical for sub-networks of factories belonging to the same company.

As an idealization of such a non-hierarchical manufacturing network, we consider a symmetric network of $N = 4$ factories displaying comparable properties as the aforementioned real-world system (Fig. 2b). This is, every production unit has the same substructure consisting of four production lines for different products fabricated for distinct recipients, and four sort-pure buffers for different resources provided by different suppliers. Every unit delivers exactly one type of product to every other factory, and one final product to some external market. The implementation of this system in the discrete-event simulation environment Plant Simulation has been described in detail in previous work (Donner et al. 2008a, b). In this work, we focus on the impact of three different order policies (order point, provision, and periodic order policy) on inventories and throughput times as two selected key figures of logistics. Performing extensive simulations with a variety of different settings of relevant parameters (such as machining and transportation times, initial inventories, and order volumes) for both static and periodic demand of the external market demand yields a large data set that can be considered representative for real-world manufacturing processes at least on short to intermediate time horizons. Based on this data base, in

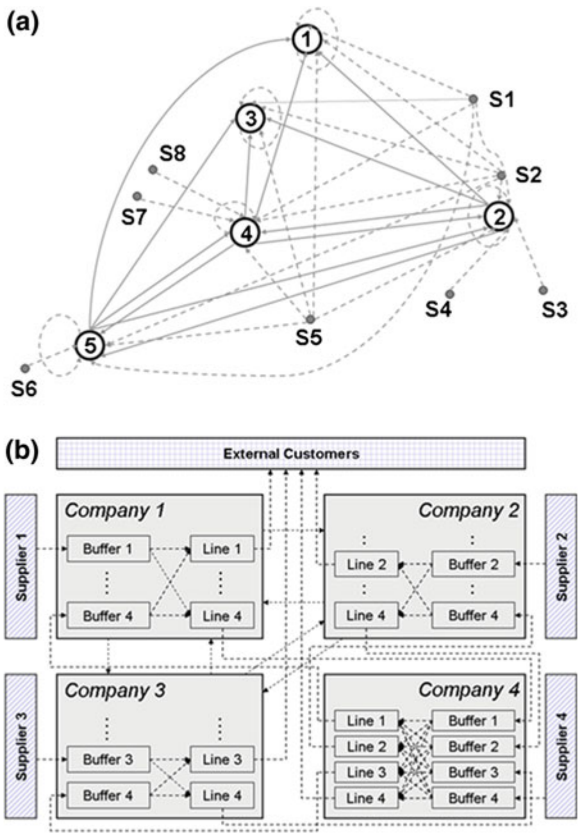


Fig. 2 **a** Schematic representation of a real-world manufacturing network (Windt 2002) consisting of five specialized factories (1–5) and 8 external suppliers (S1–S8). Material flows between different factories are represented by solid lines, such from suppliers to individual factories by dashed lines. **b** Mathematical idealization of a non-hierarchical network with $N = 4$ production units each consisting of four production lines and sort-pure buffers

the following, we will statistically classify scenarios leading to similar characteristics with respect to inventory levels or throughput times, respectively, by means of hierarchical cluster analysis. Subsequently, we compare the dendrograms obtained for both logistics observables and discuss their similarity or dissimilarity for different order strategies applied.

Results

Statistical Methodology

The specific system considered in this work has already been intensively studied for identifying and classifying different sources of instability in small-scale manufacturing systems (Donner et al. 2008c) as well as quantitatively characterizing the logistics performance by means of nonlinear methods of time series analysis (Donner et al. 2011). All previous analyses have, however, been restricted to the dynamics of inventory levels recorded for the individual sort-pure buffers of semi-finished material. In this work, we integrate for the first time information on inventory levels and throughput times into a common analysis. Since unlike inventories, throughput times do not exhibit any prominent temporal variation, it is reasonable to restrict this first analysis to very simple statistical characteristics of the corresponding probability distribution functions. For a better comparability, we will treat the inventory time series in exactly the same way. In order to get initial insights into general leading-order interdependences between inventories and throughput times, all further analyses will focus exclusively on the mean values of both observables for each production line (throughput times) and buffer (inventories), respectively. We note that it is straightforward to extend the basic analytical framework presented in the following by including further statistical characteristics, or even the complete shape of the probability distribution functions.

In order to detect similarities between different scenarios (characterized by different parameter settings and/or order strategies), we quantify them in terms of the mean-squared distance between the 16 respective mean values for each scenario,

$$d_{ij} = \sum\nolimits_{k=1}^{16} \left(\overline{X}_i^{(k)} - \overline{X}_j^{(k)} \right)^2,$$

where $\bar{X}_j^{(k)}$ represents the mean inventory (throughput time) of the k-th buffer (production line) in scenario i. A large distance d_{ij} implies that the scenarios i and j do strongly differ with respect to the mean inventories (throughput times), whereas the mean values for all buffers (production lines) approach equal values in the case of $d_{ij} = 0$. Based on the thus obtained mutual similarities, we perform an agglomerative cluster analysis of the set of all considered scenarios using the single linkage (minimum-distance) method for successively combining disjoint clusters in a hierarchical cluster tree (dendrogram).

The thus obtained dendrograms based on inventories and throughput times, respectively, are then systematically compared in order to detect similarities in the respective grouping of the considered scenarios. We do not particularly address here the problem of identifying the optimum number of groups, for which there is a variety of approaches available in the statistical literature. In order to quantify the similarities of a clustering into a fixed number K of disjoint groups, we propose utilizing the so-called Hamming distance which characterizes the statistical similarity between general binary structures (Hamming 1950). Specifically, a clustering of objects can be described by a binary matrix with entries $A_{ij} = 1$ if the elements i and j belong to the same cluster, and $A_{ij} = 0$ otherwise. Then, the Hamming distance is defined as

$$H = \frac{2}{N(N-1)} \sum_{i=1}^{N} \sum_{j=i+1}^{N} \left| A_{ij}^{(I)} - A_{ij}^{(TT)} \right|,$$

where the superscripts *(I)* and *(TT)* indicate inventories and throughput times, respectively. A Hamming distance $H = 0$ thus indicates that at the considered number of groups K, the groupings based on inventories and throughput times are the same, whereas high values of H (the maximally possible values depend on K) indicate that the respective groupings do strongly differ from each other.

Influence of Model Parameters on Mean Inventories and Throughput Times

For an order point policy, we find that moderate variations of the transportation times have only very minor influence on the resulting mean inventories as well as average throughput times, which is reflected by two marked clusters (for moderate and large maximum values, respectively) including almost all corresponding settings when the remaining model parameters are kept constant. This behavior can be easily understood since large transportation times imply that the next processing step at the production unit to be delivered may be delayed, which leads to the formation of queues and, hence, larger throughput times. However, if there are already queues in the network, a moderate increase of transportation times will not further postpone the production processes significantly. In a similar way, we observe only minor influences of machining times and initial inventories on both mean inventory levels and throughput times. Qualitatively similar findings are obtained for the provision policy, which is expected since this policy is in fact a special case of the order point policy with an order point of zero. In both cases, we can thus conclude already from a rough inspection of the resulting dendrograms that inventories and throughput times react in a qualitatively similar way to changes of the relevant production parameters of our model.

In comparison with the order point policy, periodic order policies are typically characterized by a much more complex dynamics (Donner et al. 2008a, b). This general observation is also reflected in the cluster analysis. Specifically, small variations of transportation and machining times as well as initial inventories can lead to rather similar as well as completely different mean inventories and throughput times.

We emphasize that the general findings reported above can be made for both static and periodic demand patterns of the external market. Hence, we conclude that the observed clustering is a robust feature resulting from the conceptually different types of order policies, which determine the variety of possible dynamics much more strongly than the specific choices of individual production parameters (given that they are only varied within reasonable ranges).

Regarding the dendrograms obtained for both considered logistics observables, we have carefully examined the resulting Hamming distances as a function of the number of groups K (Fig. 3). As one could expect from theoretical considerations, due to the associated average cluster size the Hamming distance has in general a maximum at a very low number of clusters and decays towards larger K. In fact,

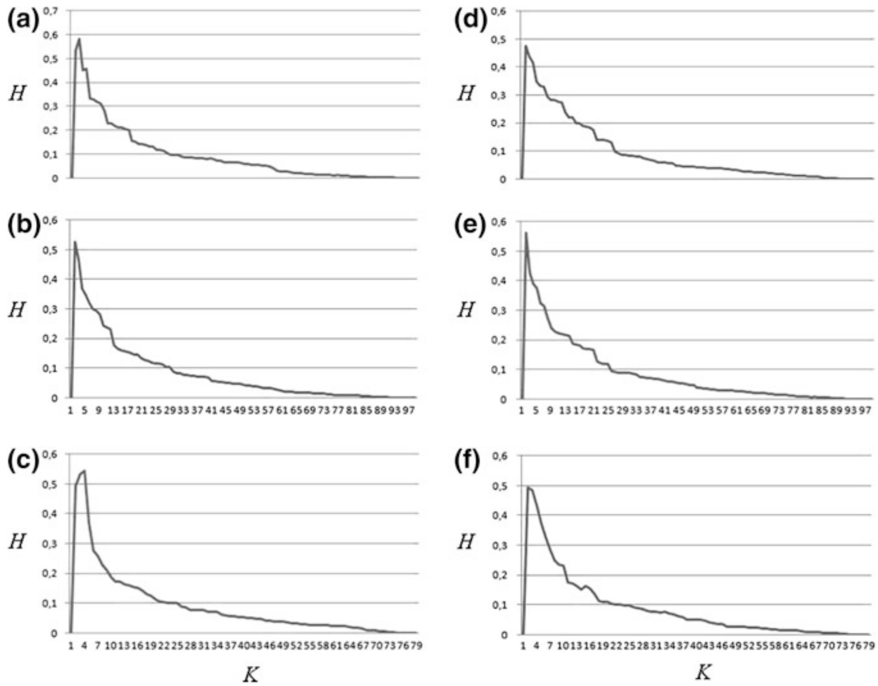


Fig. 3 Dependence of the Hamming distance *H* between dendrograms obtained based on mean inventories and throughput times on the number of groups *K* for order point, provision, and periodic order policy (from *top* to *bottom*) with static (*left panels*) and periodic (*right panels*) market demands

we do not observe any considerable differences between the curves obtained for the different order policies with static or dynamic demand patterns. When considering a reasonable number of clusters, say, *K* being of the order of 10 for in total 97 simulation runs for order point and provision policies, and 79 runs for periodic order policy (with one parameter less—order volume—than for the other two policies) considered in each setting, we find that the Hamming distance is in general rather low, indicating a certain degree of coincidence between the clusterings obtained based on inventories and throughput times. We can thus conclude that both properties are mutually dependent and should be considered together when discussing potential optimization strategies in production logistics.

Conclusions

Based on the statistical analysis of a large set of discrete-event simulations of an idealized small-scale manufacturing network model, we have provided quantitative evidence for significant interdependences between the quantitative behaviors of inventory levels and throughput times under variations of some relevant production parameters. Specifically, our investigations revealed that for order point and provision policies, transportation and machining times as well as initial inventories have only minor influence on the performance of the production system with respect to these two key figures of production logistics. Even for periodic order policies, which are commonly characterized by a much more irregular dynamics of material flows (Donner et al. 2008a, b), this general result remains qualitatively correct.

While this contribution has attempted to establish a widely applicable statistical framework for investigating the effects of order policies and production parameters on conflicting goals of logistics optimization, we have to emphasize that the restriction to mean values made here leaves a lot of practically relevant information unconsidered. Therefore, we outline a generalization of the used methodology to the consideration of higher-order statistical characteristics of inventories and throughput times, their complete probability distribution functions, or even their variations with time. The only restriction we have to make here is the availability of a suitable distance measure between the considered characteristics obtained for individual runs, which is present in all aforementioned cases. Additional research is necessary to further support our general findings reported in this work, e.g., by considering a model of the network architecture displayed in Fig. 2 validated based on real-world production data.

In summary, our results illustrate the necessity of considering the mutual interdependences between different key figures of production logistics (which are reflected in conflicts between optimization goals such as low inventories and short throughput times) in the performance optimization of manufacturing networks. In this respect, the general findings reported here have potentially important implications for the economically successful and efficient operation of arbitrary real-world production systems. Specifically, they provide fundamental information for the future development of improved planning and control strategies for networked manufacturing systems which need to take these interdependences better into account than existing approaches.

Acknowledgments This work has been partially supported by the Leibniz association (project ECONS—Evolving Complex Networks). Simulation data used in this study have been kindly provided by the BIBA—Bremer Institut für Produktion und Logistik GmbH at the University of Bremen. Corresponding discussions with Uwe Hinrichs are gratefully acknowledged.

References

Boccaletti S, Latora V, Moreno Y et al (2006) Complex networks: structure and dynamics. Phys Rep 424:175–308. doi:10.1016/j.physrep.2005.10.009

Daganzo C (2005) A theory of supply chains. Springer, Berlin

Donner R, Scholz-Reiter B, Hinrichs U (2008a) Nonlinear characterization of the performance of production and logistics networks. J Manufact Syst 27:84–99. doi:10.1016/j.jmsy.2008.10.001

Donner R, Hinrichs U, Scholz-Reiter B (2008b) Symbolic recurrence plots: a new quantitative framework for performance analysis of manufacturing networks. Eur Phys J ST 164:85–104. doi:10.1140/epjst/e2008-00836-2

Donner R, Hinrichs U, Scholz-Reiter B (2008c) Mechanisms of instability in small-scale manufacturing networks. In: Haasis HD, Kreowski HJ, Scholz-Reiter B (eds) Dynamics in logistics—first international conference, LDIC 2007—Bremen, Germany, August 2007—Proceedings, Springer, Berlin, pp 161–168 doi:10.1007/978-3-540-76862-3_15

Donner R, Hinrichs U, Schicht C, Scholz-Reiter B (2011) Complexity-based evaluation of production strategies using discrete-event simulation. In: Kreowski HJ, Scholz-Reiter B, Thoben KD (eds) Dynamics in Logistics—second international conference, LDIC 2009—Bremen, Germany, August 2009—Proceedings, Springer, Berlin, pp 423–432 doi:10.1007/978-3-642-11996-5_38

Gross T, Blasius B (2008) Adaptive co-evolutionary networks: a review. J R Soc Interface 5:259–271. doi:10.1098/rsif.2007.1229

Gross T, Sayama H (eds) (2009) Adaptive networks: theory, models and applications. Springer, Heidelberg

Hamming RW (1950) Error detecting and error correcting codes. Bell Syst Techn J 29:147–160

Hülsmann M, Windt K (eds) (2010) Understanding autonomous cooperation and control in logistics. Springer, Berlin

Hülsmann M, Scholz-Reiter B, Windt K (eds) (2011) Autonomous cooperation and control in logistics. Springer, Berlin

Kuzgunkaya O, ElMaraghy HA (2006) Assessing the structural complexity of manufacturing systems configurations. Int J Flex Manuf Syst 18:145–171. doi:10.1007/s10696-006-9012-2

Nyhuis P, Wiendahl HP (1999) Logistische kennlinien. Springer, Berlin

Pathak SD, Day JM, Nair A et al (2007) Complexity and adaptivity in supply networks: building supply network theory using a complex adaptive systems perspective. Decis Sci 38:547–580. doi:10.1111/j.1540-5915.2007.00170.x

Windt K (2002) Optimierung von Lager- und Distributionsstrukturen in Logistiknetzen am Beispiel eines weltweit agierenden Maschinenbauers. In: Tagungsband zum Wissenschafts-symposium Logistik der BVL, Huss-Verlag, Munich, pp 235–251

Modeling the Basic Cause-Effect Relationship Between Supply Chain Events and Performance

Georg Heinecke, Jonathan Köber, Andreas Kunz and Steffen Lamparter

Abstract The increased integration of global supply networks with a reduction in stock and time buffers has raised their vulnerability to disturbances (i.e., events) that erode efficiency and competitiveness. The complex interrelationships in these networks lead to a cascade of knock-on effects that affect company performance in varying degrees. Hence, causality analysis is the key for early identification of critical events. To this end, a dynamic model is developed in this contribution. It describes the basic cause-effect relationship between events, their knock-on effects and company-internal performance indicators. To make the complexity of this task manageable, an effects-based classification of events is delineated from the EPCIS format. The utility and functionality of the model is illustrated on two examples. It is proposed that it serves as a basic outline for the logic foundation of the discovery component of supply chain event management (SCEM) systems.

G. Heinecke (✉) · A. Kunz
Institute of Machine Tools and Manufacturing,
Swiss Federal Institute of Technology Zurich,
Tannenstrasse 3 8092 Zurich, Switzerland
e-mail: georghe@ethz.ch

A. Kunz
e-mail: kunz@iwf.mavt.ethz.ch

G. Heinecke · S. Lamparter
Siemens AG—Corporate Technology,
Otto-Hahn-Ring 6 81739 Munich, Germany
e-mail: steffen.lamparter@siemens.com

J. Köber
CLAAS Selbstfahrende Erntemaschinen GmbH,
Muensterstraße 33 33428 Harsewinkel, Germany
e-mail: jonathan.koeber@claas.com

H.-J. Kreowski et al. (eds.), *Dynamics in Logistics*, Lecture Notes in Logistics,
DOI: 10.1007/978-3-642-35966-8_13,

Introduction

Over the past decades, lean management has brought a revolution to manufacturing that has reshaped supply chains and restructured factories (Holweg and Pil 2005). Loose bonds between companies have made way for strong partnerships of a few key-players with equal clout that conduct (joint) research and development of sophisticated modules. Ever more intertwined interdependencies of the material and information flows between these partners caused a shift from linear supply chains to complex, globalized supply networks (Wagner and Bode 2006). At the same time, factories have become capable of producing countless product variants with little time and stock buffers according to the build-to-order philosophy.

However, while lean management with prevalent just in time (JIT) deliveries requires smooth material flows, the reality of collapsed supply chains with idle productions illustrates that stochastic influences on a global scale not only cause uncertainty but eventually materialize in costly disruptions (Blackhurst et al. 2005; Radjou et al. 2002). Between 2000 and 2009, an annual average of 387 natural disasters were recorded worldwide (Guha-Sapir et al. 2011). This equates to more than one disaster a day with potential adverse effects on supply networks. In 2010, Asia, the manufacturing center of the world, was struck by 134 disasters alone, reflecting a global share of 34.8 % (Guha-Sapir et al. 2011). Once labor disputes, demand volatility, supplier failures, transport delays, and others are added to this count, it comes as no surprise that today's complex supply networks are increasingly experiencing reliability problems. A company's performance and its customers are affected by unexpected disturbances that produce deviations of different magnitude between target and actual states (Otto 2003).

Supply chain event management (SCEM) addresses this problem. It is the practice of observing, prioritizing and reacting to events that occur during the operation of a supply chain (Tribowski et al. 2009). Events in this context can be defined as disturbances that impede, in varying degrees, the execution of supply chain processes as they were originally scheduled. However, although SCEM addresses a fundamental problem, i.e., that intra- and inter-organizational processes are rarely executed as planned, it has received less attention from the academic community than supply chain risk management (SCRM) (Otto 2003). This is surprising given that supply glitches amplify along supply networks (Radjou et al. 2002). Thus, if they are caught early when their effects are still insignificant, the worst effects up- or downstream could be avoided.

Liu et al. (2007) state, however, that this is the crux of the matter: Early identification of events with significant impacts on operational performance is difficult because its criticality to downstream processes depends on several dynamic factors. These factors change over time and continually alter the current state of the supply network and with that the criticality of an event. Thus, the same event at different points in time could have varying implications. For instance, the criticality of a two hour delivery delay depends as much on stock levels at the affected company as on the time buffer between scheduled delivery and start of production.

Liu et al. (2007) pointed out that precisely these dynamics in integrated, complex supply chains may cause a storm of events and that “causality analysis is the key to controlling such a storm”. In this contribution it is therefore argued—contrary to the common practice of defining static threshold values for event identification—that available event information (e.g., two hour delay) should be used to gauge the knock-on effects on a company’s performance given the current state of the network. This evaluation can then be repeated in an online fashion when the event information is updated (e.g., three hour delay). To this end, a novel approach to event classification is presented and incorporated into the modeling of the cause-effect relationship between supply chain events, their knock-on effects, and company performance. It is proposed that the developed model serves as a basic outline for the logic foundation of event identifications through SCEM systems.

Literature Review

The origin of SCEM as a research area can be traced back to the beginning of the new millennium (Knickle 2001). Since then the concept has languished in the international academic community, rarely going beyond high-level, conceptual publications (e.g., Otto 2003). Meanwhile, in German-speaking countries, it has received more attention, albeit still in an insufficient depth (Straube et al. 2007; Tribowski et al. 2009). The lack of attractiveness is partly attributable to the popularity of a more mature, neighboring discipline: Supply chain risk management (SCRM). Although both essentially deal with events along the supply chain, however, they need to be clearly distinguished from each other. Roughly speaking, while SCRM aims at mitigating the *potential* of events (Tang 2006), SCEM focuses on mitigating the knock-on *effects* of critical events after they are observed (Bearzotti et al. 2012; Otto 2003). To put it simply, SCEM comes into play when SCRM has failed. On the one hand SCEM is primarily concerned with the early identification of events and mitigation of their effects with short-term operational measures (Bearzotti et al. 2012; Otto 2003). SCRM on the other hand deals with medium- and long-term strategic decisions that could mitigate the risk of e.g., a supplier failure by implementing a dual sourcing strategy (Tang 2006).

Unsurprisingly, given the state of the research area, SCEM specific mechanisms for early identification of critical events on the basis of the dynamic relationships between events and supply chain factors are largely missing. One exception is Liu et al. (2007) where Petri nets were used to model event relationships in a supply chain to analyze the cause-effect relationships between consecutive events. It was shown that the variation of event parameters (e.g., probability of successful alternative sourcing) affects supply chain performance (fill rates, lead times, etc.,). This led to the conclusion that by managing events it is possible to manage supply chain performance. However, investigations into the relationship between events and company performance were not part of the research.

Apart from cause-effect relationships, tools from statistical control can be used for event identification. It focuses on the differences between target and actual states. Since processes are stochastic, different realizations of certain (local) process parameters can be obtained. The resulting frequency curve allows for the definition of threshold values below or above which a parameter is considered to be out of control and an alarm (to correct the process) is raised. This approach is of limited use in SCEM. First, an alarm suggests that effects have already considerably aggravated and containment options are limited or not existent anymore. Second, the analysis neglects dynamic supply chain factors like time and stocks. Breaching a threshold indicates a problem but it lacks information about the state of the following processes where e.g., enough security stock could be available.

Another impediment to event identification is a widespread lack of real-time status data (Günthner et al. 2010). Supply chain transparency is a critical success factor and real-time monitoring systems are a prerequisite for effective SCEM systems. For various reasons these monitoring systems are rarely realized within companies and even less among partners (Curtin et al. 2007; Günthner et al. 2010). If they are present, reservations against data exchange are common, although the same companies trade large quantities of material (Lee and Whang 2000).

Elements of the Cause-Effect Relationship

Due to the dynamics of supply networks, it is argued that event identification based on threshold values of process parameters gives insufficient insight into the criticality of an event. Cause-effect relationships, however, facilitate decision making through the estimation of fluctuations in performance indicators, given a certain event and supply network state. The magnitude of change in these indicators, which depends heavily on dynamic factors (e.g., stocks, order situation, production schedule), enable decision makers to better evaluate the observed situation and, if necessary, devise counter-measures. Before these estimations are possible, crucial elements that constitute the cause-effect relationship are defined in the following subsections.

Definition and Classification of Events and Their Basic Effects

To enable an SCEM system to identify critical events from monitoring systems' data streams, the possible occurrences have to be defined first. This task can prove difficult for two main reasons. First, it is problematic to pinpoint the precise origin of an event, since frequently only the cascade of knock-on effects is noticed. Thus, the original source of an event may stay hidden. For instance, the bullwhip effect starts with small fluctuations in downstream demand and leads to much more severe fluctuations further upstream (Lee et al. 1997). Second, it is unlikely that all possible interactions between the elements of a complex system can be foreseen, making it difficult to define all events that could affect it up front.

To facilitate the event definition process, risk classifications of SCRM that place specific occurrences into general classes can be transferred to SCEM. Chopra and Sodhi (2004) propose nine classes (disruptions, delays, systems, forecast, etc.,). Jüttner (2005) identify three classes: Supply-side, demand-side and environmental risks. Gaonkar and Viswanadham (2004) propose an event classification based on the severity of their effects (deviations, disruptions, catastrophes). All classifications, however, leave it to the decision maker to specify concrete events. This approach seems arbitrary because in complex systems events can be overlooked and allocation is bound to be ambiguous at times.

To avoid these drawbacks and to reduce the complexity of the event definition process, it is proposed to employ a novel approach that is based on the *immediate, object-related effects* of an event on a company. They can be derived from the presence or absence of object-related observations at RFID-based monitoring points along the supply chain. We propose to describe the resulting primitive, real-time data streams of these observations with the EPCIS format (EPCIS 2007). It contains information that essentially answer the question of what object is when, where and why—the latter refers to the business context of the observation and is described by a business step ID. Hence, this information can be leveraged to notice cases when the supply chain in general and preceding logistics processes in particular fail to fulfill their purpose: To bring the right product, at the right time, to the right place, in the right quantity and the right quality (Fig. 1). These five event classes can be easily identified by a comparison of reality (i.e., observation or its absence at a monitoring point) with plan information from a central database. For instance, suppose a traffic jam delays a truck delivery for two hours. The monitoring system first notices that the truck has not passed the monitoring point at the scheduled (i.e., planned) time. Once it passes the monitoring point, the delay can be quantified and expected effects on performance indicators gauged. Once the material arrives, however, the bulk-reading of the cargo, in conjunction with information from the database, reveals that too little was delivered. Thus, a delay becomes more severe when only part of the required material arrives.

The example illustrates one advantage of using an event classification (Fig. 1) that builds on the information provided through the EPCIS format—namely, once the resulting effects of each of these five classes on the processes of an affected company are examined more closely, it is apparent that they result in only two distinct cases of cause-effect relationships. First, *material delivery delay* constitutes an event where the observation at a specific point of the monitoring system was delayed. Second, *material delivery failure* constitutes an event where all or part of a shipment is useless (i.e., wrong product and/or quality) or unfulfilled (wrong quantity and/or place). These event classes also cover demand side risks

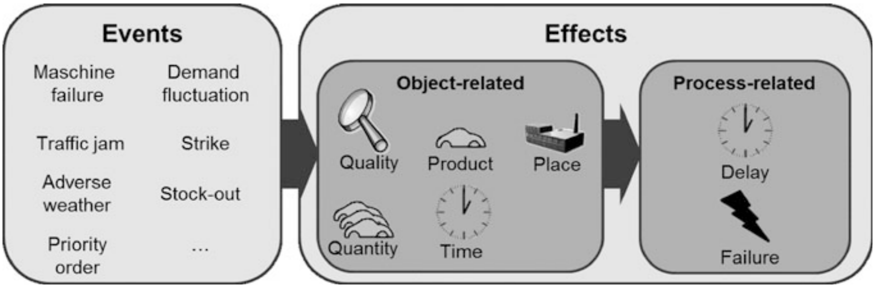


Fig. 1 Effects-based event classification

such as priority orders or sudden demand fluctuations (e.g., wrong quantity delivered). For instance, a short-term increase in the order quantity by the customer results in a delivery failure when the supplier has insufficient material stocked for handling the increase, or the material is already on its way to the manufacturer.

It is apparent that an *effects-based classification* of events not only reduces (external) events to two distinct cases but, more importantly, relates them to time as the sole measuring unit. In short, the classification illustrates that all events, based on their effects, can be placed in a temporal context. Many of the external events that can occur at the supplier, however, can also happen at the company under consideration. These internal events (e.g., strike, machine failure, etc.) affect the shipment of the product and with this also the company's performance. Therefore, they have to be included when the cause-effect relationship of events on company performance is examined. Thus, two analog process-related event classes are proposed: *material production delay* and *material production failure*.

Events and Vulnerability Drivers

Wagner and Neshat (2010) state that "supply chain vulnerability is a function of certain supply chain characteristics and that the loss a firm incurs is a result of its supply chain vulnerability to a given supply chain disruption". Wagner and Bode (2006) identify five external drivers that increase vulnerability: single and global sourcing, customer and supplier dependence and supplier concentration.

Besides these external vulnerability drivers, internal ones also threaten a company's performance. For instance, small delivery delays quickly paralyze modern production systems due to their reliance on JIT deliveries that need little time and stock buffers. Furthermore, mass customization has led to an explosion of product variants, which has made the timely provisioning of specific material critical. At the same time, production systems lack flexibility that would allow a short-term adaption of the production program in case some material cannot be provided.

Although diverse, internal and external vulnerability drivers influence the event classes comparably by increasing the *likelihood* of occurrence and the *severity* of their effects. For instance, single sourcing of a specific part considerably raises the likelihood of a material delivery failure. Similarly, global sourcing of a part raises the likelihood of a material delivery delay due to larger distances. In contrast to SCRM, which aims at mitigating these drivers, SCEM incorporates their implications but otherwise regards them as fixed. Thus, an intelligent SCEM system preserves the advances of lean management (e.g., little time and stock buffers) while ensuring high performance in an event-prone environment.

Event Criticality and Performance Indicators

To establish the cause-effect relationship, events have to be linked to meaningful indicators. Performance can be measured in various ways and literature proposes many different indicators (e.g., Gunasekaran et al. 2001, 2004). This contribution aims at relating an event and its knock-on effects to the success of a company. These effects are first noticed in production processes when material is delayed or missing. Therefore, rather than approaching performance measurement from a strategic, long-term perspective (e.g., customer satisfaction), the operational indicators of Nyhuis and Wiendahl (2006) are employed. These are affected first and allow a quick and concise estimation of the severity of an event.

Given a JIT delivery strategy with little buffers and no intervention, the most immediate effect of an event is a reduction of the work in process *inventory* as queues in front of production assets deplete. This reduction leads to (1) shorter *cycle times* since waiting times are cut and (2) lower asset *utilization* due to increased idle times. Hence, also the *production rate* and overall output are cut. The customer notices the effects of an event when, through a lack of material, his outstanding orders cannot be fulfilled as scheduled and *delivery delays* accumulate.

Modeling the Cause-Effect Relationship

Systems dynamics is well-suited for the modeling and visualization of supply networks because the coherent notation can relate its many variables and their general interdependencies. According to Özbayrak et al. (2007) "the ability of understanding the network as a whole, analyzing the interactions between the various components of the integrated system […] make systems dynamics an ideal methodology for modeling supply chain networks".

The following subsections integrate the previous considerations into a simplified stock and flow diagram. First, the basic JIT manufacturing system with its interdependencies will be presented. It serves as baseline system. In the next subsection, the model will be enhanced through the integration of the event classes that affect a company. The final model helps in gauging the criticality of an event in a dynamic environment by quantifying their effects on performance indicators.

Basic Manufacturing System with Performance Indicators

In very generic terms, a supply chain can be represented by several, consecutive accumulations (i.e., stocks) that are each increased by an inflow and decreased by an outflow (Sterman 2000). Operating according to the build to order principle, the supply chain of Fig. 2 is initiated through product orders by the customer. These increase the backlog and equate to the material order rate of the manufacturer. This material order rate is the inflow for the stock of material that is on order but not yet delivered. The outflow is regulated by the delivery rate. It would normally increase the material inventory but is superfluous in a JIT system. Thus, the delivery rate directly equates to the production start rate that is responsible for the inflow into the work in process (WIP) inventory. It is drained by the production rate. The resulting product inventory is then drained by the ship rate. It equates to the fulfillment rate that decreases the backlog again. Thus, in a deterministic setting with constant flow rates the entire system quickly reaches equilibrium in which only a fixed time passes between the placement of the order by the customer and the fulfillment of the order by the manufacturer.

The performance indicators of the system are stocks, cycle time, utilization, and delivery delays. Stocks are given by the WIP and production inventories. The cycle time is calculated using Little's Law, while utilization is derived from the production rate (Hopp and Spearman 2007). Lastly, the delivery delay is calculated from the backlog and the fulfillment rate.

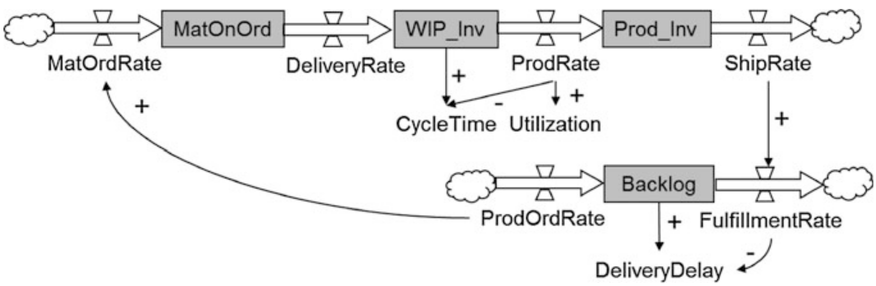


Fig. 2 JIT manufacturing system

Event-Prone JIT Manufacturing System

Figure 3 incorporates the event classes and vulnerability drivers into the model. They directly influence the flow levels that are not constant anymore but fluctuate depending on the severity of the event. For instance, a tsunami will have a more profound but less sudden impact on the delivery rate than a traffic jam. The delivery rate is influenced, in the opposite direction, by the probabilities of a material delivery delay of different length and material delivery failure of different size: A larger probability leads to a lower average delivery rate as they occur more frequently over the simulation time. The production rate is similarly influenced by the internal event classes. All classes are affected by vulnerability drivers. For instance, global sourcing increases the distance between manufacturer and supplier, making a delivery delay or failure more likely and severe.

In Fig. 3 the knock-on effects of events are now noticeable in a consecutive, time-dependent *reduction* first of the stocks and then of the flows of the system. Once the effects of an event have rippled through the manufacturing system, the shipping rate is reduced. Due to constant demand levels, several customer orders cannot be satisfied in this situation. The reduction in the fulfillment rate leads to an increase in the backlog. The manufacturing system is now clogged by orders that should already have been shipped. The orders that are supposed to be processed have to increasingly wait their turn, and delivery delays accumulate.

Scenario Description and Performance Analysis

To briefly verify and evaluate the usefulness of the presented concept, a supply "hiccup" (i.e., material delivery delay) of different length in a JIT environment is simulated. In the first scenario half of the required material for time $t + 6$ is

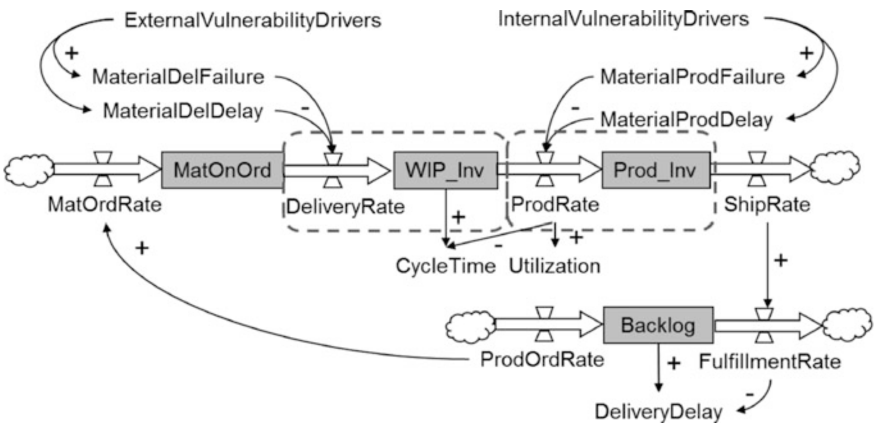


Fig. 3 Event-prone JIT manufacturing system

delayed and arrives at $t + 7$. The second scenario increases this delay to two days. Thus, half the material is delayed in $t + 6$ and $t + 7$ and only arrives in $t + 8$ and $t + 9$ respectively. Both scenarios assume constant customer orders and production capacities of 10 units per day. The simulations were carried out using Vensim PLE.

Figure 4 illustrates the development of the WIP inventory. It is apparent that both systems reach equilibrium on day 3 where 10 units arrive and 10 units are processed. The one-(left diagram) and two-day (right diagram) delays then influence the system to the same effect, albeit in different magnitude: The average long-term WIP inventory increases. This circumstance is due to (1) a lack of supplies inventory in JIT systems and (2) no flexibility in production capacity. The first results in a permanent loss of unused production capacity while the second prevents a new alignment between customer demand and WIP inventory.

Figure 5 shows the development of the delivery delay. In both scenarios it also reaches equilibrium on day 3 when the average delay between order placement and order shipment is 3 days—i.e., the order to delivery time. In case of the one-day material delay (left) the delivery delay briefly doubles and then stays at 3.5 days. In the second scenario (right) it increases to 7 days before leveling out at 4 days. This analysis illustrates the fragility of a JIT system with no flexibility in production capacity. A brief delay permanently increases the WIP inventory and overall lead time as older orders are clogging the manufacturing system.

The examples illustrate that the basic logic of the model enables the *isolation* of the undistorted, performance-related effects of an event. The simulated performance analysis helps a decision maker, who faces the prospect of a one- or two-day material delivery delay, to make a more informed decision of whether the particular event is deemed critical to the operations and the success of the company. Obviously, this decision can be automated by defining thresholds for each indicator. If these are breached, an event is considered critical and planers can prepare counter-measures well before the effects of it reach the company. Besides costly emergency logistics schemes, measures then include, but are not limited to, economic alternatives such as a rescheduling of the order sequence by utilizing various flexibility potentials of the production. Hence, early event identification is

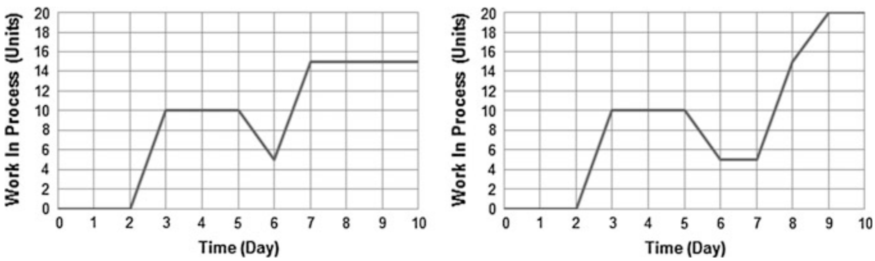


Fig. 4 Work in process (WIP) inventory of a one-(*left*) and two-day (*right*) delay

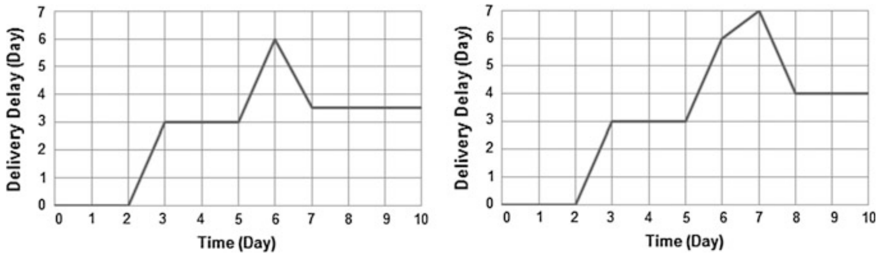


Fig. 5 Delivery delay of a one-(*left*) and two-day (*right*) delay

critical because once the production sequence is fixed (i.e., frozen zone), rescheduling becomes an impossible task and affected products have to be reworked at high costs.

Concluding Remarks

The increasing appearances of different supply chain events present a serious threat to the efficiency and competitiveness of global supply networks. This necessitates concepts and systems that can identify events early and react to their effects. To this end, a dynamic model is developed that describes the basic cause-effect relationship between events, their knock-on effects and performance indicators. To facilitate this task an effects-based classification of events on the basis of the EPCIS data format is proposed in this contribution. It is demonstrated that the model serves as a basic outline for the logic foundation of the event discovery component of supply chain event management (SCEM) systems. The simulation of two different delay events showed that the quantification of their performance-related effects can help a decision maker to estimate their criticality in advance and more accurately. In future the proposed model is applied to an industrial case study to further investigate the validity of the proposed cause-effect relationship. Also, additional research will be done in regard to (1) the combined effect an events when conflicting optimization goals are pursued, (2) the decision about the criticality of an event and (3) adequate response policies (e.g., rescheduling).

References

Bearzotti LA, Salomone E, Chiotti OJ (2012) An autonomous multi-agent approach to supply chain event management. Int J Prod Econ 135(1):468–478

Blackhurst J, Craighead CW, Elkins D, Handfield RB (2005) An empirically derived agenda of critical research issues for managing supply-chain disruptions. Int J Prod Res 43(19):4067–4081

Chopra S, Sodhi MS (2004) Managing risk to avoid supply-chain breakdown. MIT Sloan Manage Rev 46(1):53–62

Curtin J, Kauffman RJ, Riggins FJ (2007) Making the 'most' out of RFID technology: a research agenda for the study of the adoption, usage and impact of RFID. Inf Technol Manage 8(2):87–110

EPC Global (2007) EPC Information Services (EPCIS). Version 1.0.1 Specification. September, 2007

Gaonkar R, Viswanadham N (2004) A conceptual and analytical framework for the management of risk in supply chains. In: Proceedings—IEEE international conference on robotics and automation pp 2699–2704

Günthner W, Meißner S, Conze M, Fischer R (2010) Stand und Entwicklung des RFID-Einsatzes in der Automobillogisitk: Ergebnisse einer empirischen Studie. RFID Anwenderzentrum, München

Guha-Sapir D, Vos F, Below R, Ponserre S (2011) Annual disaster statistical review 2010. Centre for Research on the Epidemiology of Disasters. Université Catholique de Louvain, Brussels

Gunasekaran A, Patel C, Tirtiroglu E (2001) Performance measures and metrics in a supply chain environment. Int J Oper Prod Manage 21:71–87

Gunasekaran A, Patel C, McGaughey RE (2004) A framework for supply chain performance measurement. Int J Prod Econ 87(3):333–347

Holweg M, Pil FK (2005) The second century: reconnecting customer and value chain through build-to-order moving beyond mass and lean in the auto industry, vol 1. The MIT Press, Cambridge

Hopp W, Spearman M (2007) Factory physics, 3rd edn. Irwin McGraw-Hill, NY

Jüttner U (2005) Supply chain risk management: Understanding the business requirements from a practitioner perspective. Int J Logistics Manage 16(1):120–141

Knickle K (2001) Supply chain event management—the next best thing to supply chain perfection. AMR Research Outlook, Boston

Lee HL, Whang S (2000) Information sharing in a supply chain. Int J Technol Manage 20:373–387

Lee HL, Padmanabhan V, Whang S (1997) Information distortion in a supply chain: the bullwhip effect. Manage Sci 43(4):546–558

Liu R, Kumar A, van der Aalst W (2007) A formal modeling approach for supply chain event management. Decis Support Syst 43(3):761–778

Nyhuis P, Wiendahl HP (2006) Logistic production operating curves—basic model of the theory of logistic operating curves. CIRP Ann Manuf Technol 55(1):441–444

Otto A (2003) Supply chain event management: three perspectives. Int J Logistics Manage 14(2):1–13

Özbayrak M, Papadopoulou TC, Akgun M (2007) Systems dynamics modelling of a manufacturing supply chain system. Simul Model Pract Theory 15(10):1338–1355

Radjou N, Orlov LM, Nakashima T (2002) Adapting to supply network change. Tech. rep., Forrester Research

Sterman JD (2000) Business dynamics: systems thinking and modeling for a complex world, vol 53. Irwin McGraw-Hill, NY

Straube F, Vogeler S, Bensel P (2007) RFID-based supply chain event management. In: RFID Eurasia, 2007 1st Annual, pp 1–5

Tang CS (2006) Perspectives in supply chain risk management. Int J Prod Econ 103(2):451–488

Tribowski C, Goebel C, Guenther O (2009) EPCIS-based supply chain event management—a quantitative comparison of candidate system architectures. In: International Conference on complex, intelligent and software intensive systems, pp 494–499

Wagner SM, Bode C (2006) An empirical investigation into supply chain vulnerability. J Purchasing Supply Manage 12(6):301–312

Wagner SM, Neshat N (2010) Assessing the vulnerability of supply chains using graph theory. Int J Prod Econ 126(1):121–129

The Importance of Managing Events in a Build-to-Order Supply Chain: A Case Study at a Manufacturer of Agricultural Machinery

Jonathan Köber and Georg Heinecke

Abstract This application-oriented research focuses on customer-orientation in the context of build-to-order supply chain management (BOSC) and supply chain event management (SCEM). The last decade was characterized by increasing volatility and complexity in the supply network. As a consequence, companies with traditional supply chain management were confronted with more exogenous and endogenous disturbances. As a result, customer-oriented supply chains show poor performances in regard to the service rate, delivery time, inventory level and capacity utilization. Finally a "Lose-Lose-Situation" occurs, because customers do not get what they want and enterprises lose profit. This paper presents a case study of a manufacturer of agricultural machinery whose production system is following a BOSC strategy. Recently two supply chain disturbances, a strike in the production and a supplier shortage of a just-in-sequence part, significantly affected the service rate. The performance-related effects of these two disturbances are demonstrated in an elaborated system dynamics model, policies for their mitigation (e.g. capacity flexibility) proposed and their effectiveness evaluated.

J. Köber (✉)
CLAAS Selbstfahrende Erntemaschinen GmbH,
Münsterstraße 33, 33428 Harsewinkel, Germany
e-mail: jonathan.koeber@claas.com

G. Heinecke
Swiss Federal Institute of Technology Zurich,
Tannenstrasse 3, 8092 Zurich, Switzerland
e-mail: georghe@ethz.ch

G. Heinecke
Siemens AG Corporate Technology, Gleiwitzerstraße 555,
90475 Nuremberg, Germany
e-mail: georg.heinecke.ext@siemens.com

H.-J. Kreowski et al. (eds.), *Dynamics in Logistics*, Lecture Notes in Logistics,
DOI: 10.1007/978-3-642-35966-8_14,

Introduction

In the field of supply chain management, firms find their suppliers to be unreliable because the required delivery rates are often missed. In turn, suppliers find the ordering patterns of their customers to be volatile and unpredictable. Generally, it creates a vicious cycle between sales, production and suppliers (Christopher and Holweg 2011; Holweg and Pil 2001). Inside each firm, managers find their forecasts of incoming orders are rarely correct and always changing. On the one hand, the sales forecasts lose all credibility with the production people. On the other hand, the marketing and sales organizations complain that unreliable production makes forecasting and selling difficult. The endogenous instability caused by exogenous events and the structure of a supply chain undermine trust within and between supply chain partners. The conflict creates supply chain instability and is usually aggravated by a lack of transparency. The latter is strongly needed as a prerequisite for a successful, early identification of supply chain events from real-time status data (Heinecke et al. 2011).

As a result the approach of build-to-order (BTO) production has become a popular operations paradigm after the successful implementations at Dell Computers. It refers to a demand-driven production approach, where a product is scheduled and built in response to a confirmed customer order (Holweg and Pil 2001). Hence, firms like the one in the following case study follow the approach of BOSC management to (re)act to the fluctuations in demand and to handle the increasing complexity of the value network (Gunasekaran and Ngai 2005; Gunasekaran 2007). The BOSC model is now being actively pursued in several different industries such as computers, automotive and manufactures of construction and agricultural machinery (Holweg and Pil 2004; Parry and Graves 2004; Salvador et al. 2004). BOSC can be defined as ‘‘the value chain that manufactures quality products or services based on the requirements of an individual customer […] at competitive prices, within a short span of time by leveraging the core competencies of partnering firms or suppliers and information technologies […] to integrate such a value chain" (Gunasekaran and Ngai 2005).

However, supply chains are continuously subjected to disturbances e.g. in demand, production rates and material deliveries. These disturbances or events constantly knock BOSC out of equilibrium. In the literature, different approaches are discussed to define the right product delivery strategy (Christopher and Towill 2001; Olhager 2010). Furthermore, research has strongly focused on the topic of preventive measures to minimize the potential for events. Approaches like supply chain risk management and BOSC that aim at achieving a robust and economical supply chain are state of the art in theory and practice. The case study will show, however, that supply chain events occur nevertheless and knock event-prone BOSC out of their equilibrium. The authors have thus conceived a generic system dynamics model to evaluate the performance-related effect of supply chain events.

According to Huang et al. (2007) events can be classified into three categories: deviations, disruptions and disasters. Deviation events refer to changes in parameters from their expected or mean value. Disruption events are occurrences that are so significant and far reaching that normal operation is disturbed considerably. For example, if product mixes or demand have changed significantly, it will take a span of time for the system to recover from deteriorating performance. During the recovery, performance becomes unpredictable and remedial actions must be taken in order to bring the system back to a more stable state. Disastrous or catastrophic failures lead to a temporary irrecoverable shut-down of a supply chain network (Gaonkar and Viswanadham 2004).

The authors present a case study of a supply chain whose operations are affected by unpredictable events. When these occur, firms must take immediate actions to assess potential impacts and, if necessary, activate contingency plans for mitigation of their worst effects. The case study is based on the supply chain of a manufacturer of agricultural machinery who follows the BTO principle to achieve high performance in an increasingly complex environment. Simultaneously unpredictable events, however, erode the stability of the production program and endanger overall performance. Thus, as this paper's case study will illustrate, BTO production systems do not guarantee high performance when they lack flexibility. In a nutshell, the paper focuses specifically on the following key points:

1. Building a generic system dynamics model based on the manufacturing supply chain of the case study.
2. Reconstructing the two supply chain events (strike and delivery failure of a supplier) in the model that led to a lower service rate. Verification of the assumption that variations or disturbances knock BOSC out of equilibrium.
3. Identification and evaluation of possible policies (e.g. capacity flexibility) that can be applied to ensure high performance of a BOSC with respect to market and operational targets.

Case Study

Manufacturing Supply Chain

The present case study is based on data from the supply chain of a medium-sized manufacturer of agricultural machinery that produces combine harvesters, forage harvesters, balers, forage harvesting machinery and tractors. These markets are characterized by low volumes, seasonal demand, series production, increasing product variety and globalization of operations. The production system is following a BTO principle where every production job is triggered by a customer order. Hence, disturbances and volatile demand have both a significant impact on the behaviour of the production system. Due to the competitive landscape, the agricultural machinery market has high requirements on quality, service, price, delivery reliability and short delivery times to generate customer satisfaction.

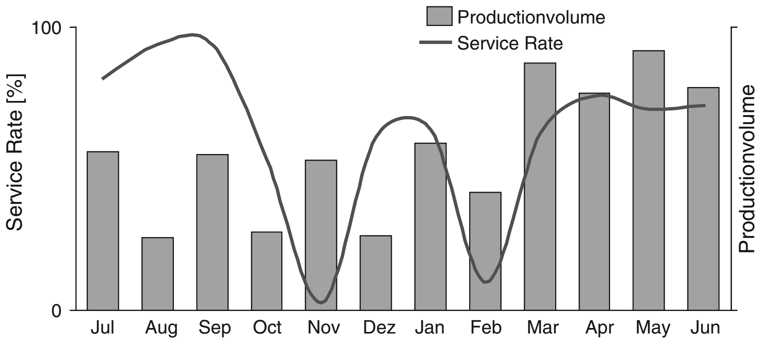


Fig. 1 Service rate and production volume at the manufacturer of the case study

Furthermore, the BOSC is strongly influence by unplanned endogenous and exogenous factors like supplier issues, sales campaigns, strikes and demand and price changes.

Figure 1 shows the service rate (line) and production volume (bars) over a period of 12 months. In general, the production volume in August and December is lower than the other months because of the summer and winter break. The other two significant falls in production volume in November and February, however, are due to two major disruptions. Both had a delayed effect on the service rate, which is defined as the percentage of punctually fulfilled orders compared to the total amount of orders during a certain time span.

In October a strike affected a plant of the manufacturer for over one week. As a result of this event, most of the scheduled machines were only produced weeks later. Now, if an event leads to a delay in production then all following, sequenced orders are not punctual regarding to their confirmed delivery date. Thus, many of the confirmed delivery dates were missed and the service rate declined significantly in November. As a main result, the monthly service rate fell dramatically. A second supply chain event happened in late January and early February. A specific module that is sourced just-in-sequence from a certain supplier could not be delivered in the required quantity for weeks. Hence, the OEM was unable to manufacture certain product variants and the service rate significantly declined again in February.

Key Performance Indicators

Key performance indicators (KPIs) of a production system according to Wiendahl are (1) delivery time, (2) service rate, (3) capacity utilization and (4) inventory level (Wiendahl 2010). From the market perspective a very short delivery time and adherence to delivery dates, which lead to a high service rate, are crucial. From the perspective of the operational level, the goals is to have a high and stable

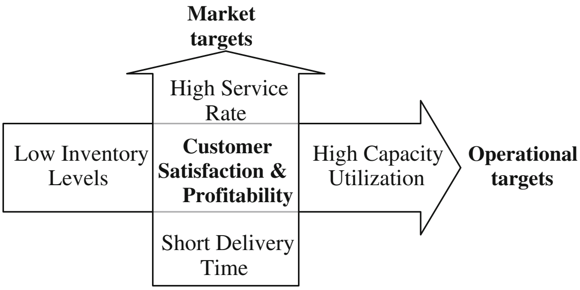


Fig. 2 Key performance indicators and the polyemma of production controlling

utilization of capacity and low inventory levels of raw material, work in process and final stock. The four performance measures are shown in Fig. 2. The four targets have a huge influence on customer satisfaction and on enterprise profitability. On the one hand they create an internal conflict between having high service rates and short delivery times. On the other hand to have low inventory levels and high capacity utilization. In the literature this contrary target set is called the polylemma of production controlling (Wiendahl 2010), which means it has at least more than two conflicting objectives (options) compared to a dilemma situation. Regarding to the presented case study, where the manufacturer operates in a saturated buyer market, the operational targets are arguably less important than the market targets. The latter ensure lasting customer loyalty for services and future product versions even in the face of severe competitive pressure. Nevertheless, since operational targets affect enterprise profitability, all four KPIs will be embedded in the simulation model. This approach enables a comprehensive evaluation of the manufacturing supply chain performance.

Description of the System Dynamics Model

The described system of the manufacturing supply chain is transformed into a continuous simulation model to evaluate the cause-effect structure and the impact of supply chain events on the KPIs. Furthermore, different supply chain reactions are simulated and evaluated in order to determine appropriate policy decisions that can mitigate the effects of these events.

System dynamics is a specific approach of system theory and is adequate to analyse complex system behaviour. The benefits of system dynamics are a unified notation of continuous flow modelling and a graphical interface. The user-friendliness, aggregation of system parameters and understanding of the system behaviour are considered as another important advantage. In addition to the possibilities of continuous flow modelling based on differential equations, it also offers the possibility of implementing discrete elements like random and trigger functions, which facilitate an intuitive modelling (Scholz-Reiter et al. 2008; Morecroft 2007). In short, "system dynamics is well suited for supply chain modelling and policy design. It is a method to enhance learning in complex systems and […] to help us learn about dynamic complexity, understand the sources of policy resistance, and design more effective policies" (Stearman 2000). Potentially, other simulation tools (e.g. Plant Simulation, OTD-NET Simulator) may be also adequate to demonstrate this purpose. However, system dynamics fits perfectly to build a generic model, which illustrates the system behaviour of a supply chain on an abstracted level. This generic model makes it possible to verify and evaluate the performance-related effects of supply chain events (variations and disturbances) in manufacturing supply chains.

The generic model in Fig. 3 is based on the manufacturing model in (Stearman 2000). Additionally to the basic manufacturing supply chain model, the KPIs and supply chain events are added. The simulation was created with the software Vensim® PLE. Figure 3 illustrates the model with its three main areas: The supply chain structure (a), a basic market structure (b) and key performance indicators (c). The supply chain structure in block a consists of the order, material and production flow from the supplier, via the manufacturer, to the customer. These supply flows are limited by the maximum available capacity (e.g. in production or material supply). Delays and adjustment times (e.g. material delivery time and production time) are embedded in the simulation model. The material and production rates are determined with a time delay. The market structure (block b) contains a customer order rate per time unit, backlog of orders, order fulfilment rate, target of delivery delay and desired shipment rate. The backlog shows the total amount of orders per time unit. The desired shipment rate defines by the quotient between the backlog and target delivery delay. The order fulfilment rate is determined by the shipment rate. The delivery delay is equal to the backlog divided by the order fulfilment rate. If the final stock is greater than or equal to the desired shipment rate, the service will be 100 %. Hence, the service rate measures the number of punctually fulfilled

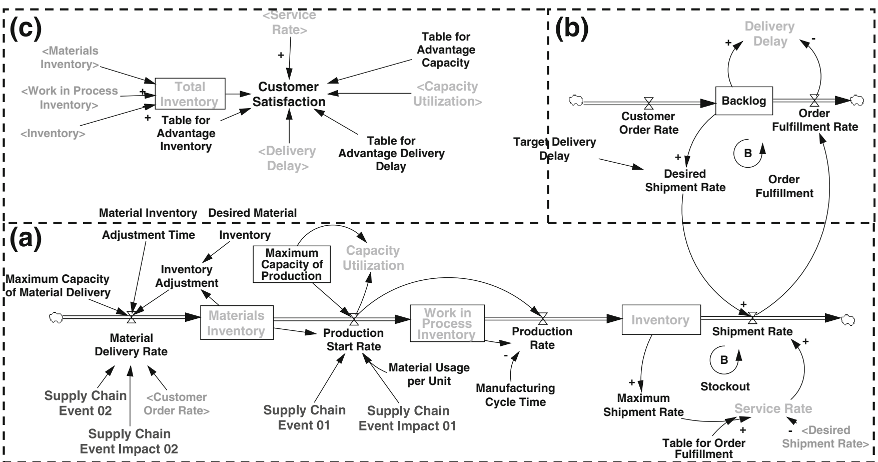


Fig. 3 System dynamics model of a manufacturing supply chain with the KPIs

orders. Besides the service rate, the three other KPIs (block c) of production control are embedded. Those indicators have a direct and important link to the customer satisfaction and ultimately also to enterprise profitability.

Analysis of the Base Case Scenario

This subsection compares performance of the model (1) when operations are stable and (2) when they are disrupted by two events to verify the performance-related effects of supply chain events on BOSC. Table 1 shows the employed parameters of the base case scenario.

The dashed line in Fig. 4 shows that a stable supply chain leads to the levelling of the market targets, service rate as well as the delivery delay. In this base scenario the manufacturing supply chain achieves a high service rate of about 90 % and a delivery delay of about nine weeks. The effect of the first weeks is neglected because the simulation needs several weeks to calibrate the behaviour of the system.

The first supply chain event (e.g. strike at the OEM) occurs from week 15 to 17. The service rate falls sharply from 90 to 48 % and needs more than 10 weeks to recover to 65 %, where it levels off. The delivery delay is directly linked to this performance. Its value increases from 9 to 17 weeks and, after several weeks, balances itself at 12 weeks. This effect happens because the production constantly operates at its maximum capacity limit of 600 units. Customer orders still arrive at the same rate, however, and therefore the production is continually booked with old orders and the new arrivals have to wait longer than they usually would. Furthermore, the unfulfilled orders increase the order backlog and desired

Table 1 Base case parameter for the manufacturing supply chain model

Parameter	Base case value
Simulation time	52 weeks
Customer order rate	600 units/week
Target delivery delay	8 weeks
Manufacturing cycle time	4 weeks
Maximum capacity	600 units
Desired material inventory	600 units
Material inventory adjustment time	4 weeks
Maximum capacity of material delivery	600 units
Material usage per unit	1
Supply chain event time 1 (strike at OEM)	week 15
Supply chain event impact 1	−600 units
Supply chain event time 2 (supplier delivery issue)	week 32
Supply chain event impact 2	−600 units

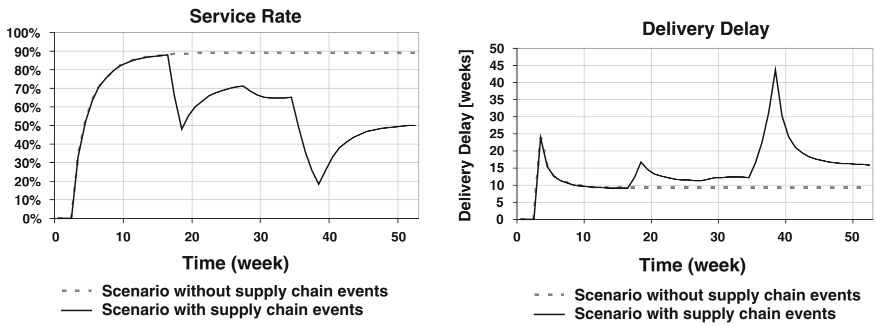


Fig. 4 Service rate and delivery delay in the base case scenario

shipment rate. Hence, if an event like a production strike happens, the delivery delay increases permanently as lost production capacity cannot be made up. The second event, e.g. a shortage of 600 units of a just-in-sequence module, happens from week 35 to 39. The effect is similar to the first event of a production disturbance. Again, the service rate falls sharply - this time from 65 % to under 20 % - and the delivery delay increases from 12 to over 40 weeks, later balancing itself at 16 weeks. It can be concluded that a disturbance of the supply chain temporarily destabilizes the system and leads to generally lower performance levels for the considered KPIs.

Analysis of the Scenario with Capacity Flexibility

A lot of different approaches of counter-measures exist to avoid performance degradation. However, an often discussed solution to avoid performance degradation is to increase flexibility in capacity (Salvador et al. 2007; Howard 2002; Fredriksson and Gadde 2005; Ahlert et al. 2009) to absorb deviations. With the developed model is it possible to show the effectiveness of an increase in flexibility regarding to production capacity and material supply rate.

In the simulation model with production flexibility the maximum production rate, maximum capacity of material delivery and desired material inventory are increased from 600 to 700 units and then in another simulation from 600 to 800 units. Figure 5 illustrates the effects of an increased capacity to 700 units on the KPIs. The average service rate is 15 % higher in the scenario with capacity flexibility. Furthermore, the capacity flexibility also has a positive effect on the delivery delay. The average delivery time is reduced from 14 to 11 weeks. On the other side, the operational targets are worse than before. An increase of the maximum capacity leads to unused resources and consequently the average capacity utilization falls by 6 %. The inventory level shows no noteworthy effects. In essence, the capacity flexibility is only required in case of disturbances in the supply chain. If supply chain events happen, capacity flexibility has a positive effect on the two market targets, service rate and delivery delay. This overall positive effect is illustrates in the customer satisfaction and profitability diagram (bottom left diagram in Fig. 5).

In another simulation the maximum capacity is increased to 800 units. This setup leads to a very high performance regarding the market targets but to a much lower level for the operational ones. Hence, customer satisfaction and profitability achieve an overall lower level because of an over-sized production capacity. The simulations show that flexibility is costly. The operational costs are important factors that need to be considered. Idle capacity and/or high inventory levels do not

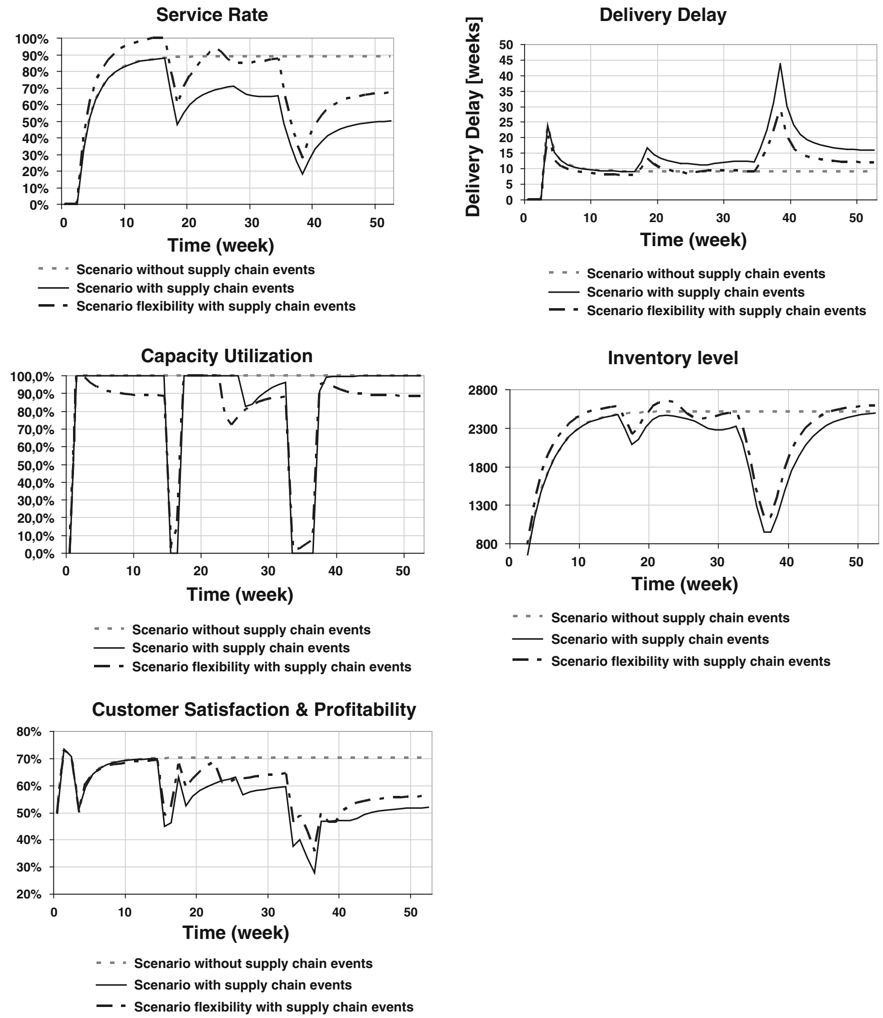


Fig. 5 Performance indicators of market and operational targets

add value to the product itself but still have a direct influence on product prices as these costs have to be carried by the customers.

In summary, the simulation model illustrates the influences of supply chain events and enables their evaluation. Beyond that, different policies can be validated and verified to realize an effective BOSC, which considers the influences of supply chain events. This study showed that increased capacity flexibility helps to stabilize BOSC when disturbances influence manufacturing supply chain performance. Furthermore, the cost of flexibility has to be compared with the loss of operational performance.

Conclusion and Outlook

Although BTO has done much regarding the reduction of costs and, hence, increase customer satisfaction and profitability of the OEM and, it is questionable whether the increased vulnerability of production systems and performance do not necessitate another paradigm shift. It is important to have a holistic view on the supply chain behaviour and the balance of operational and market targets. The developed system dynamics model is a suitable tool to understand the behaviour of a manufacturing supply chain and to simulate different supply chain setups and mechanism of counter-measures to absorb the vulnerability through supply chain events. The results of this contribution can help to design and optimize a BOSC and to achieve a “Win - Win-Situation” between customers and enterprises in future. This paradigm shift can be envisioned as a hybrid production system that utilizes build-to-stock and build-to-order principles combined with an optimized degree of mix and volume flexibility in the supply chain, and consequently harness their respective advantages. A comparison with similar industrial cases will be relevant to validate the methodology, generic system dynamic model and the presented results. Additionally, fundamental research has to focus on the fields of vulnerability of BOSC and the intelligent mitigation of effects of supply chain events through SCEM systems. In further research the elaborated simulation model will be extended to evaluate the cost-benefit rate of a production system that considers also the influence of supply chain events. Furthermore, focusing on the organizational and strategic factors that drive supply chain management decisions with respect to the implementation of a BTO strategy represents a relevant future direction for research.

References

Ahlert KH et al (2009) Capacity management in order-driven production networks—A flexibility-oriented approach to determine the size of a network capacity pool. Int J Prod Econ 118:430–441. doi:10.1016/j.ijpe.2008.11.013

Christopher M, Holweg M (2011) “Supply Chain 2.0”: managing supply chains in the era of turbulence. Int J Phys Distrib Logist Manage 41(1):63–82. doi:10.1108/09600031111101439

Christopher M, Towill D (2001) An integrated model for the design of agile supply chains. Int J Phys Distrib Logist Manage 31(4):235–246. doi:10.1108/09600030110394914

Fredriksson P, Gadde LE (2005) Flexibility and rigidity in customization and build-to-order production. Ind Mark Manage 34:695–705. doi:10.1016/j.indmarman.200505.010

Gaonkar R, Viswanadham N (2004) A conceptual and analytical framework for the management of risk in supply chains. In: Proceedings of the IEEE international conference on robotics and automation, pp 2699–2704

Gunasekaran A (2007) Build-to-order supply chain management. Int J Oper Prod Manage 27(11) guest editorial

Gunasekaran A, Ngai EWT (2005) Build-to-order supply chain management: a literature review and framework for development. J Oper Manage 23:423–451. doi:10.1016/j.jom.2004.10.005

Heinecke G, Lamparter S, Kunz A, Wegener K, Lepratti R (2011) Process transparency: effects of a structured read point selection. In: Proceedings of the 21st international conference on production research

Holweg M, Pil FK (2001) Successful build-to-order strategies start with the customer. MIT Sloan Manage Rev 43(1):73–84

Holweg M, Pil FK (2004) The second century: reconnecting customer and value chain through build-to-order moving beyond mass and lean in the auto industry. The MIT Press, Cambridge

Howard M (2002) 3DayCar production flexibility: solving the barriers to build-to-order with production and process technology. Executive Summary, University of Bath http://www.3daycar.com/mainframe/publications/library/prodflex.pdf. Accessed 30 Dec 2011

Huang HY et al (2007) A dynamic system model for proactive control of dynamic events in full-load states of manufacturing chains. Int J Prod Res 47(9):2485–2506. doi:10.1080/00207540701484913

Morecroft J (2007) Strategic modelling and business dynamics: a feedback system approach. Wiley, Chichester

Olhager J (2010) The role of the customer order decoupling point in production and supply chain management. Comput Ind 61:863–868. doi:10.1016/j.compind.2010.07.011

Parry G, Graves A (2004) Build to order: the road to the 5-day car. Springer, London

Salvador F et al (2004) Build-to-Order is not that easy: adding volume flexibility to mass customization. Working paper, University of Madrid http://ideas.repec.org/p/emp/wpaper/wp04-16.html. Accessed 30 Dec 2011

Salvador F et al (2007) Mix flexibility and volume flexibility in a build-to-order environment: Synergies and trade-offs. Int J Oper Prod Manage 27(11):1173–1191. doi:10.1108/01443570710830584

Scholz-Reiter B et al (2008) Dynamik logistischer Systeme. In: Nyhuis P (ed) Beiträge zu einer Theorie der Logistik. Springer, Berlin

Stearman JD (2000) Business dynamics: thinking and modeling for a complex world. Irwin McGraw-Hill, London

Wiendahl HP (2010) Betriebsorganisation für Ingenieure, 7th edn. Carl Hanser Verlag, Munich

A Dynamic Approach to Measure the Effect of Postponement Strategy on the Performance of Manufacturing and Logistics Processes

Luiz Eduardo Simão and Mirian Buss Gonçalves

Abstract The current dynamic business environment dominated by competition between supply chains, factors such as the increase in demand variability, the proliferation of product variety, the process and supply uncertainty and the increased complexity of manufacturing and logistics networks are crucial in the increasingly globalized competitive arena. One way to meet this challenge is to use of postponement as a strategy to delay the point of product differentiation on space and time along the supply chain. This strategy has been increasingly used as an important tool for supply chain management. Thus, this article aims to present a dynamic method and a set of attributes and metrics to measure and evaluate the effect of postponement on the performance of manufacturing and logistics processes. Subsequently, the method has been applied in a case study on a Brazilian industrial company, considering two different scenarios, represented by total anticipated demand in space and time (decentralized inventory/make-to-stock) and the total demand postponement in space and time (centralized inventory/make-to-order).

L. E. Simão (✉)
CAPES Foundation Ministry of Education of Brazil, Brasília, DF, Brazil
e-mail: luizes2011@hotmail.com

L. E. Simão · M. B. Gonçalves
UFSC—Federal University of Santa Catarina, Florianópolis, SC, Brazil
e-mail: mirianbuss@deps.ufsc.br

L. E. Simão · M. B. Gonçalves
Campus Universitário, Caixa Postal 476, Trindade, Florianópolis, SC CEP 88040-900, Brazil

H.-J. Kreowski et al. (eds.), *Dynamics in Logistics*, Lecture Notes in Logistics,
DOI: 10.1007/978-3-642-35966-8_15,

Introduction

In these times of economic uncertainty, it is increasingly difficult to forecast customer demand. The major problem in most supply chains is their limited visibility of real demand (Christopher 2000). In this new environment, some companies have developed a new business model based on response time, supported by the great advance in information and communication technologies (Bowersox and Closs 1996). In contrast to the traditional business model, based on anticipation, the business model based on time, enable the creation of organizational networks (Van Hoek 1998), also known as supply chains. Thus, in this business environment dominated by supply chains, the increase in demand variability, the proliferation of product variety, the process and supply uncertainty and the increased complexity of the manufacturing and logistics networks are crucial in the increasingly globalized competitive arena. This requires an agile supply chain where the manufacturing and logistics systems to be able to responding a fast and flexible way to costumer's requests. The aim of the agile supply chain should be to carry inventory in a generic form, that is, standard semi-finished products awaiting final assembly or localization. This is the concept of 'postponement', a vital element in any agile strategy (Christopher 2000). Therefore, this article aims to present a dynamic method and a set of attributes and metrics to measure and evaluate the effect of postponement strategy on the performance of manufacturing and logistics processes across the supply chain.

Related Works

Postponement means delaying activities in the supply chain until customer orders are received with the intention of customizing products, as opposed to performing those activities in anticipation of future orders (Van Hoek 2001). The first related work that addressed the principle of postponement dates back to the 1920s (Knight 1921). The theoretical and practical basis was developed by Alderson (1950) and later expanded by Bucklin (1965) with regard to demand by the concept of speculation and also by Starr (1965) with regard to supply by the concept of product modularization. Basically, these authors argue that the costs related to risks and uncertainties are intrinsically related to the intensity of differentiation in space–time that occurs during the manufacturing and logistics processes. These authors also asserted that the greater final configuration of the products is postponed along the supply chain, the smaller the losses are and the higher is the level of service to the customer. At that time these authors recognized that the growing demands for customization would not be satisfied by the mass production system and it would be necessary to involve the areas of manufacturing and logistics in order to maximize the possibility of differentiation. However, the works of Alderson (1950), Bucklin (1965) and Starr (1965) did not trigger the deserved attention of executives due to the predominant mentality of mass production of that time. Only in the late 1980s, when Zinn and Bowersox (1988) published an article and resumed the study of the subject, it was recognized that postponement was an effective strategy to improve distribution systems. In the 1990s, Christopher (1992) broadened the scope of the concept, highlighting postponement as an effective and feasible way to reorganize global supply chains. Therefore, the author recommends a careful examination of the entire value chain, in order to look for opportunities to postpone the final configuration of the product, i.e., encourage the order of components and intermediate packaging in global supply networks for final assembly in the countries of destination. Subsequently, several works have been published (Lee and Billington 1994; Van Hoek 1997; Lee and Tang 1997; Feitzinger and Lee 1997; Pagh and Cooper 1998; Battezzati and Magnanimous 2000; Van Hoek 2001; Lee and Swaminathan 2003; Yang and Burns 2003; Yang et al. 2004; Boone et al. 2007; Kisperska-Moron and Swierczek 2010), and so on. However, despite the increased attention of professionals and academics to use the strategy of postponement in supply chains, their applications are still not used as expected (Van Hoek 2001; Yang and Burns 2003; Boone et al. 2007).

According to Lee and Billington (1994), it is possible to postpone the point in the supply chain at which the customization of the final product can be configured, through appropriate design of the product structure, manufacturing and logistics processes. Bowersox and Closs (1996) stated that there are two types of postponement that are critical for business: (1) production postponement (form) and (2) logistics postponement (time and place). According to these authors, these two types of postponement are able to reduce risk, but in different ways. The production (form) postponement focuses on the shape of the product, moving unfinished items and customization along the logistics network before delivery to the costumers. In other words, form postponement means delaying the final differentiation of the product until the customer orders are received (Skipworth 2003).The logistics postponement (time and place) has two focuses: place—which proposes the storage of finished products in centralized facilities; and time—which refers to performing the movement of products at the last possible moment in time, in order to respond quickly to orders received from the customer.

Moreover, in recent years, some studies showed that postponement has been used successfully to:

- **Manage the proliferation of product variety** (Randall and Ulrich 2001; Ramdas 2003; Blecker and Abdelkafi 2006; Stäblein et al. 2011);
- **Reduce the complexity of the supply chain** (Van Hoek 1998; Znou and Mill 2002; Blecker and Abdelkafi 2006; Christopher and Holweg 2011);
- **Reduce demand variability** (Van Hoek et al. 1999; Christopher 2000; Van Hoek 2001; Yang et al. 2004; Graman and Magazine 2006);
- **Reduce process and supply uncertainty** (Van Hoek 2001; Brown et al. 2000; Yang and Burns 2003; Yang et al. 2004).

In order to achieve these objectives, it is also necessary an aligned supply chain strategy to facilitate manufacturing, logistics and outsourcing decisions along the supply chain (Mikkola and Skjøtt-Larsen 2004). Yang and Burns (2003) claimed that due to the different possible combinations between the various types of postponement and supply chain configurations for an industrial company, taking into account the different aspects of the operation, product and demand, postponement can be applied fully or partially through three operational strategies known as:

- **ETO (engineer-to-order)**;
- **MTO (make-to-order)**;
- **ATO (assemble-to-order).**

A 2 × 2 matrix with four alternatives for the choice of supply chain strategy was developed by Pagh and Cooper (1998). These alternatives are derived from the combination of the possible choices of manufacturing and logistics strategies in relation to demand in space and time. In general, this reflects a combination:

1. **Strategy of anticipated total demand in space and time**
2. **Anticipation strategy in space and postponement strategy in time**
3. **Postponement strategy in space and anticipation strategy in time**
4. **Strategy of postponement of total demand in space and time.**

Yet, according to Pagh and Cooper (1998), each of these four approaches has advantages and disadvantages in terms of cost and operational flexibility. The strategy of early total demand in space and time (decentralized inventories/make-to-stock) is the most widely adopted by companies. Based on sales forecasts, all manufacturing operations are carried out by anticipating the shipment of products to distribution centers. In turn, shipment is made prior to order placement by end customers. Economies of scale in production and distribution are the main advantages of this strategy, for the processing of materials/products always occurs in large batches. However, investment in inventory is high, and the risk of obsolescence and the cost of transfers between facilities are higher.

In the anticipation strategy in space and postponement strategy in time (decentralized inventories/make-to-order) the early stages of production are centralized, aiming at achieving economies of scale. The decentralization is operationalized by sending stocks of semi-finished products to various distribution centers close to the end customer, and it is intended to provide reduced delivery times. The main advantage of this strategy is to reduce the number of SKU levels of safety stock of finished products in various distribution centers. However, there is an increase in the cost of the final stages of production due to the decentralization of operations, for example, packaging, assembly or mixing. There is also an increase in costs of processing applications due to the need to coordinate the physical distribution with the final stages of production.

With regard to the strategy of postponement strategy in space and anticipation strategy in time (centralized inventories/make-to-stock), manufacturing operations are centered and directed to the formation of stocks based on sales forecasts, before the physical distribution. The advantages of this strategy are related to the reduction of inventory levels of finished goods due to the centralization. The disadvantages refer to higher delivery costs, as a result of the higher frequency of shipments and hiring ad-hoc express transportation. In this case, economies of scale in production are preserved.

Finally, in the strategy of postponement of the total demand in space and time (centralized inventories/make-to-order) production operations are fully centralized and initiated on-demand. The main advantage is the reduction of inventory levels of raw materials, semi-finished and finished products at all stages of production. In this case, economies of scale in production are virtually eliminated.

A Method and Metrics for Measuring and Evaluating the Impact of Postponement Strategy on the Performance of Manufacturing and Logistics Processes

One way to measure and evaluate of the impact of postponement in manufacturing and logistics processes is by checking the magnitude of the gap in performance indicators before and after the implementation of this strategy. Thus, a radar chart can be made with a set of indicators with their respective values for different scenarios; subsequently the gaps between the indicators may be evaluated in these two situations (Frazelle 2002) along a time interval (day, week or month).

Moreover, despite the increasing importance and interest in implementing the strategy of postponement, few models for their measurement and evaluation were developed specifically for this purpose. Some works presented an evaluation of cost-benefit analysis and results obtained using postponement, and pointed out some metrics and approaches. However, all works have focused on specific supply chain processes or specific aspects for their measurement and evaluation (Zinn and Bowersox 1988; Lee and Billington 1994; Lee and Tang 1997; Van Hoek 1999). However, there are gaps in the literature, specifically in analyzing which of these measures and approaches would be more appropriate to assess the impacts on the use of postponement on the performance of the supply chain. Nevertheless, Zhang and Tan (2001) developed a specific model for joining metric measurement and evaluation of postponement. Based on the prospects of the business approach they classified the performance measurement of the postponement in two dimensions: (1) internal and (2) external. The internal dimension was sub-divided into four types of performance attributes: (1) total cost (2) cost, (3) customer service, and (4) asset management. The external dimension includes only one attribute of performance: (1) environmental cost.

This study proposes a new method for measuring the performance of postponement in order to enable quantifying the efficiency and effectiveness of adopting this strategy on supply chain. Thus, the impact of postponement over the manufacturing and logistics processes should be evaluated through six performance attributes: (1) Inventory, (2) Use of assets, (3) Time, (4) Customer Value, (5) Financial and (6) Environment. The Fig. 1 demonstrates this methods impact.

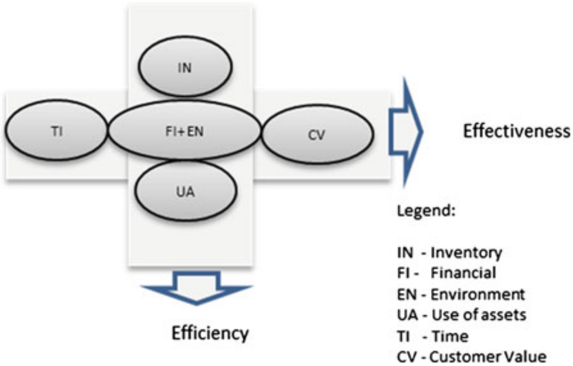


Fig. 1 Performance measurement attributes for postponement strategy

On the performance measurement attributes for postponement in the Fig. 1, the attributes related to inventory (IN) and use of assets (UA) are strongly related to the internal perspective (efficiency) of the supply chain performance in terms of costs, while the attributes related to time (TI) and customer value (CV) are related to the external perspective (effectiveness) of the supply chain performance in terms of customer service needs. With regard to the two central attributes, financial (FI) and environment (EN) impact, it is important to achieve the maximum profit with minimal use of resources in supply chain, which related to both perspectives (efficiency and effectiveness). These last attributes are able to assess the performance results of the whole supply chain. Each attribute has its meaning and metrics, as described below:

- **Inventory (IN)** this attribute evaluates performance in terms of inventory of raw material, work-in-process and finished goods. This measure includes: inventory level, inventory turns, inventory cost and capital invested.
- **Use of Assets (UA)** this attribute focuses on the use of capital investment in plants, vehicles and equipment. This measure includes: productivity, machine utilization and vehicle utilization.
- **Time (TI)** this attribute encompasses the set of performance measures related to time from the customer point of view. This measure include: total lead time, transport lead time and manufacturing lead time;
- **Customer Value (CV)** this attribute includes the set of performance measures used to assess customer satisfaction. This measure includes: fulfillment rate, product variety, on-time delivery and quality rate.
- **Financial (FI)** this attribute evaluates the performance impact in terms of costs and value from the stakeholder's point of view. This measure includes: profitability, total cost, cost as percentage of sales, product cost, manufacturing cost and transport cost.
- **Environment (EN)** this attribute evaluates the performance impact in terms of their effect on the environment and resource use. This measure includes: energy consumption, water consumption, CO_2 emission and solid waste production.

Once that the method and a set of attributes and measures has been identified, it is possible to measure and evaluate the impact of postponement in the performance of manufacturing and logistics processes.

A Case Study

In order to analyze how the measurement model can be applied to evaluate the impact of postponement on the performance of manufacturing and logistics processes, one case study has been conducted in Brazilian industrial company. The manufacturer produces more than 150 different models of garden equipment and delivers it for over one hundred retailers in Brazil and South America. It has more than 800 employees who produce over 4,000 equipment units a day. The manufacturing system employs a make-to stock strategy and use forecasting techniques to forecast customer demand. The production starts before customers' orders are received and the company operates in batches model. The usual lead time for delivery products to customers is between 20 and 30 days. The order is fulfilled when it is completely manufactured and delivered from warehousing to shipping. The manufacturing and logistics network is described in Fig. 2.

Figure 2 represents an example of a manufacturing and logistics network without the application of postponement in manufacturing and logistics processes. This strategy anticipates total demand in space and time (decentralization/make-to-stock). In this approach, based on sales forecast, all manufacturing operations are carried out in anticipation to product shipment to the industry distribution center before to costumers.

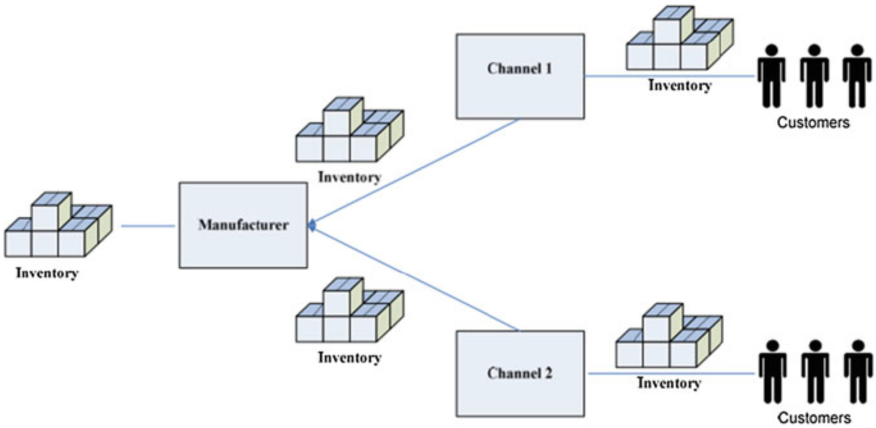


Fig. 2 Process flow chart of the garden equipment company with manufacturing and logistics processes without postponement

After adopting the postponement strategy the company employed a make-to-order strategy to fulfill customers‘ orders. Figure 3 illustrates the changes on the manufacturing and logistics network with the implementation of postponement.

The Fig. 3 describes the strategy of postponement total demand in space and time (centralization/make-to-order), in which all tasks are delayed until receiving the customers’ orders. First, raw materials are transformed to standardized parts (generic products). When the actual orders are received, these parts are assembled together with other customized parts (e.g., engines, knifes, electrical system, plastic parts, user‘s manual and boxes) to specifics finished goods in an acceptable time. The main advantages of this strategy are the reduction of lead time, inventory levels of raw materials, semi-finished and finished products at all stages of the manufacturing and logistics network and while maintenance a high level of product variety. This approach is able to improve order fulfillment and economies of scale in production and logistics may be reduced.

The proposed model for performance gap analysis of postponement may be used for measuring and analyzing the performance of manufacturing and logistics processes. For this specific case, there were used 10 metrics: total cost, raw material inventory, transport cost, inventory turns, costs as percentage of sales, manufacturing cost, total lead time, finished goods inventory and work-in-process inventory. Based on the performance indicators listed, a radar chart was made composed of the set of metrics with their respective values for the two different scenarios described in Figs. 2 (without postponement) and 3 (with postponement).The results of the performance gap analysis of postponement for the Brazilian industrial company case study are presented in Fig. 4.

Figure 4 illustrates that the postponement strategy yielded positive impacts for nine out of 10 metrics proposed measurements. Therefore, postponement was able to reduce the demand variability in production schedule and consequently reduce the process and the supply uncertainty and the complexity of the manufacturing

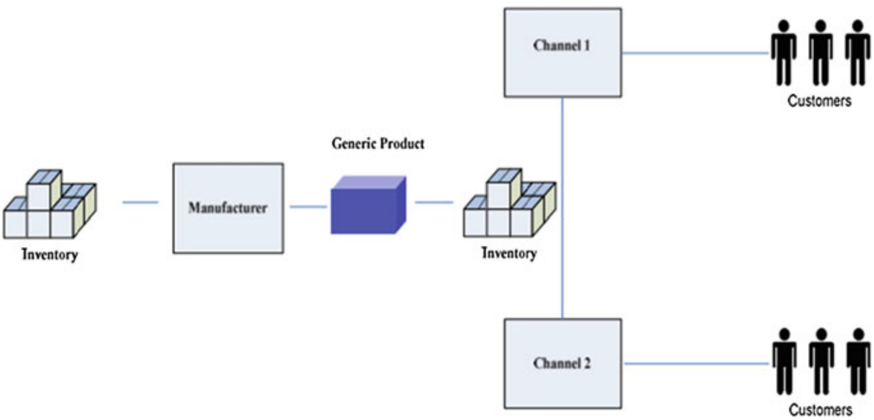


Fig. 3 Process flow chart of the garden equipment company with manufacturing and logistics processes with postponement

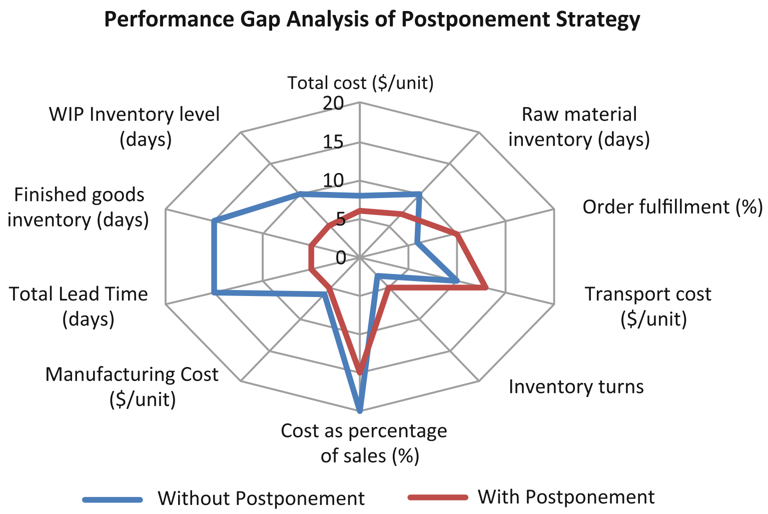


Fig. 4 Performance gap analysis of postponement strategy of the company case study

and logistics network. At the same time this strategy allowed maintaining a high degree of product variety for fulfilling delivery customers' needs upon delivery. Finally, this method allowed assessing the dynamic effect of postponement strategy on the manufacturing and logistics processes.

Conclusions

The highly competitive and dynamical changing environment has required companies' right decisions to plan carefully the implementation of manufacturing and logistics systems along the supply chain. The method here developed aims at assessing performance gap analyses of postponement and its effect on the supply chain performance. Thus, a case study has been conducted in Brazilian industrial company in order to demonstrate its applicability, considering a set of ten measures to quantify the efficiency and effectiveness of adopting that strategy on the supply chain. Two different scenarios have been measured and evaluated—without postponement and with postponement. The method allowed the visualization of the dynamic effect of postponement on the manufacturing and logistics processes. The case study yield positive impact for nine out of 10 metrics used to measure and evaluate the impact of postponement strategy had positive impact. Moreover, the study demonstrated that postponement was able to reduce the demand variability in production scheduling, the process and supply uncertain and the complexity of the manufacturing and logistics network. At the same time the postponement strategy maintained a high degree of product variety for fulfilling costumer needs.

References

Alderson W (1950) Marketing efficiency and the principle of postponement. Cost Profit Outlook 3:15–18

Battezzati L, Magnanimous R (2000) Supply chains for FMCG and industrial products in Italy: practices and advantages of the postponement. Int J Phys Distrib Logistics Manage 30(5):413–424

Blecker T, Abdelkafi N (2006) Complexity and variety in mass customization systems: analysis and recommendations. Manag Decis 44(7):908–929

Boone CA, Craighead CW, Hanna JB (2007) Postponement: an evolving supply chain concept. Int J Phys Distrib Logistics Manage 37(8):594–611

Bowersox DJ, Closs DJ (1996) Logistical management: the integrated supply chain process. McGraw-Hill, New York

Brown AO, Lee HL, Petrakian R (2000) Xilinx improves its semiconductor supply chain using product and process postponement. Interfaces 30(4):65–80

Bucklin LP (1965) Postponement, speculation and the structure of distribution channels. J Mark Res 2:26–31

Christopher M (1992) Logistics and supply chain management: strategies for reducing cost and improving services. Financial Times Pitman Publishing, London

Christopher M (2000) The agile supply chain: competing in volatile markets. Ind Mark Manage 29(1):37–44

Christopher, M, Holweg M (2011) Supply chain 2.0: managing supply chains in the era of turbulence. Int J Phys Distrib Logistics Manage 41(1):63–82

Frazelle E (2002) Supply chain strategy: the logistics of supply chain management. McGraw-Hill, NY

Kisperska-Moron D, Swierczek A (2010) The selected determinants of manufacturing postponement within supply chain context: An international study. Int J Prod Econ 133(1):192–200

Feitzinger E, Lee HL (1997) Mass customization at Hewlett-Packard: the power of postponement. Harvard Bus Rev 75(1):116–121

Graman GA, Magazine MJ (2006) Implementation issues influencing the decision to adopt postponement. Int J Oper Prod Manage 26(10):1068–1083

Knight FH (1921) Risk, Uncertainty and profit. Hart, Schaffner and Marx, Houghton Mifflin Company, Boston

Lee HL, Billington C (1994) Designing products and processes for postponement. Management of design: engineering and management perspectives. Kluwer Academic, Boston, pp 105–122

Lee HL, Tang CS (1997) Modelling the costs and benefits of delayed product differentiation. Manage Sci 43(1):40–53

Mikkola JH, Skjøtt-Larsen T (2004) Supply-chain integration: implications for mass customization, modularization, and postponement strategies. Prod Plan Control 15(4):352–361

Pagh JD, Cooper MC (1998) Supply chain postponement and speculation strategies: how to choose the right strategy. J Bus Logistics 19(2):13–33

Ramdas K (2003) Managing product variety: an integrative review and research directions. Prod Oper Manage 12(1):79–101

Randall T, Ulrich K (2001) Product variety, supply chain structure, and firm performance: an empirical examination of the U.S. bicycle industry. Manage Sci 47(12):1588–1604 INFORMS

Skipworth H (2003) The application of form postponement in manufacturing. PhD Thesis, Cranfield University, England

Stäblein T, Holweg M, Miemczyk J (2011) Theoretical versus actual product Variety: how much customization of the customers really demand? Int J Oper Prod Manage 31(3):350–370

Starr MK (1965), "Modular production, a new concept", Harvard Bus Rev 53(6):131–142.

Swaminathan JM Lee HL (2003) Design for Postponement. In: de Kok A.G., Graves S.C.(eds.): Handbooks in Operations Research and Management Science: Supply Chain Management:Design, Coordination and Operation. Elsevier Publishers, pp 199–228

Van Hoek RI (1997) Postponed manufacturing: a case study in the food supply chain Suppl Chain Manage, , 2(2): 63–75.

Van Hoek RI (1998) Reconfiguring logistic systems through postponement strategies. J Bus Logistics 19:22

Van Hoek, R.I., (1999) Postponement and the reconfiguration challenge for food supply chains,Supp Chain Manage, 4(1): 18–34.

Van Hoek RI, Peelen PDE, Commandeur PDHR (1999) Achieving mass customization through postponement: a study of international changes. J Market-Focused Manage 3(3):353–368

Van Hoek RI (2001) The rediscovery of postponement: a literature review and directions for research. J Oper Manage 19(2):161–184

Yang B, Burns ND (2003) Implications of postponement for the supply chain. Int J Prod Res 41(9):2075–2090

Yang B, Burns ND, Backhouse CJ (2004) Postponement: review and an integrated framework. Int J Oper Prod Manage 24(5):268–487

Zhang C, Tan GW (2001) Classification of postponement strategies and performance metrics framework. In: Proceedings of the 4th PACIS

Zinn W, Bowersox DJ (1988) Planning physical distribution with the principle of postponement. J Bus Logistics 9(2):117–136

Znou D, Mill L (2002) An empirical study of the role of postponement application in reducing supply chain complexity. In: IEEE international engineering management conference, vol 1, pp 448–453

Event Management for Uncertainties in Collaborative Production Scheduling and Transportation Planning: A Review

Bernd Scholz-Reiter, Yi Tan, Nagham El-Berishy and José B. S. Santos Jr.

Abstract This paper presents a review of using event management to deal with the uncertainties in production scheduling and transportation planning processes at the operational level. Moreover, it argues the importance of considering uncertainties and the application of event management in a collaborative production and transportation planning process at the operational level.

Introduction

Supply chain management (SCM) requires a precise coordination of flows of raw materials, finished goods, information and financial resources. With the fast development of information technology and the global market, collaboration between different functional units in a supply chain (SC) has become one of the key success factors for companies involved in SCs.

B. Scholz-Reiter · Y. Tan (✉) · N. El-Berishy
Planning and Control of Production Systems (PSPS), University of Bremen, Bremen, Germany
e-mail: tan@biba.uni-bremen.de

B. Scholz-Reiter
e-mail: bsr@biba.uni-bremen.de

N. El-Berishy
e-mail: elb@biba.uni-bremen.de

J. B. S. Santos Jr.
Learning Lab in Logistics and Transport, Civil Engineering Architecture and Urban Design School at the University of Campinas,
Campinas, São Paulo, Brazil
e-mail: jbened@fec.unicamp.br

H.-J. Kreowski et al. (eds.), *Dynamics in Logistics*, Lecture Notes in Logistics,
DOI: 10.1007/978-3-642-35966-8_16,

However, to achieve this collaboration companies are facing new challenges. On the one hand, this collaboration requires companies to respond to changes of customer needs quickly. On the other hand, due to shorter product life cycles as well as an increased number of product variants and the increasing dependence of supply chain functional units, the entire supply chain has become more dynamic than ever before (Stank et al. 2011).

In this dynamic environment, uncertainties of the SC activities, represented by the gap between the planned and the actual system status, occur frequently and therefore must be considered in the SC planning processes. This gap shows a new challenge for the management of complex and dynamics SC processes. The entire organizations and the SC partners are more susceptible to unexpected events or situations not covered in the SC planning stages (Christopher et al. 2011). Processes, such as the coordination between transportation planning stages for dealing with long lead times, the ability to quick response to relevant changes in demand or supply, the synchronization between production scheduling and transportation planning, the outsourcing of products and services, the reduction of inventory levels through just in time (JIT), the collaboration with suppliers, etc., emphasize this important management issue.

This paper focus on two SC processes: production and transportation. The aim of this paper is to review the state of the art of using event management to deal with the uncertainties in these two processes. Indeed, this work argues the importance of considering the uncertainties in a collaborative production and transportation planning at the operational level. The paper is organized as followed: section Supply Chain Event Management and Supply Chain Risk Management presents an overview of Supply Chain Event Management (SCEM) and Supply Chain Risk Management (SCRM). Section Event Management for Uncertainties in Production Scheduling shows a review of the literature regarding event management and rescheduling techniques in the production environment to deal with uncertainties. Section Event Management for Uncertainties in Transportation Planning summarizes the important issues and methods in transportation planning linked with uncertainties. Section Event Management in Collaborative Production and Transportation Planning discusses the collaborative production and transportation planning considering uncertainties. The paper closes with a conclusion and description of future research.

Supply Chain Event Management and Supply Chain Risk Management

In order to properly react to these unexpected deviations, a high degree of transparency and visibility in the supply chain processes are necessary. The concept of Supply Chain Event Management covers the visibility and transparency requirements focusing on two main goals. First, identifying possible unexpected deviations and reducing their impacts, to assure the customer satisfaction and the operation efficiency; second, creating the supply chain visibility (Otto 2003).

In order to fulfill these goals SCEM identifies deviations in SC processes between the planned and the actual status. Afterwards, it starts actions to minimize the gaps for the whole SC, according to some predefined rules. The main role of applying the SCEM is to eliminate or to reduce the delay time between the moments when an unexpected event occurs and when the decision maker addresses and solves the problem caused by this event.

Supply Chain Risk Management consists of the identification and evaluation of risks as well as the consequent losses. In addition, this approach mitigates these losses and ensures the supply chain outcomes through the implementation of coordinated strategies among the SC partners (Manuj and Mentzer 2008).

In general, the SCRM considers four steps: risks identification, analysis, evaluation and monitoring. In the first step all risks for the supply chain are determined. At the analysis phase, a deep understanding of the risk's identification must be done. The purpose of the evaluation step is to define the most appropriate management response for each risk or combination of risks. Finally, a monitoring and control procedure has to be implemented to manage the risks (Ritchie and Brindley 2007a, b; Khan and Burnes 2007). This step encompasses the major part of the SCEM process. This fact shows the strongest link between SCEM and SCRM.

Basically, potential sources of risks can be identified in accordance with: environmental characteristics, industrial characteristics, network configuration, network partners, organization's strategy, problem specific variables, and decision making. The risks associated with SC management are classified as: risks related to the operation (variability in demand, disruption in the supply process, etc.,); risks related to natural disasters (hurricanes, tsunamis, earthquakes, etc..); and, risks caused by direct action of human (war, financial bubbles, etc.,) (Tang 2006). This paper focuses on the operation risks. Table 1 presents the most relevant uncertainties at the operational level focusing on these two SC processes.

The risk-taking decisions influence the selection of risk management strategy for operations in a supply network. For example, the postponement[1] strategy in a manufacturing environment could increase the product development costs and the investments to the higher flexibility needed for the assembly lines.

However, planning process needs to react for these exceptions to give a solution for the system's current status. In an effort to extend the SCEM research, this paper proposes an approach considering the uncertainties in the integrated production and transportation planning systems at the operational level simultaneously.

[1] The decision to delay some manufacturing activities like assembly, labeling or packaging.

Event Management for Uncertainties in Production Scheduling

In order to deal with unexpected events occurring during a production process (see the rescheduling triggers for production planning listed in Table 1), rescheduling has been investigated in the last decade. Rescheduling focuses on the operational level of the production planning.

When unexpected events take place and invalidate the planned production schedule, rescheduling is applied. It updates the current production schedule or generates a completely new schedule according to the current state of the production system. Rescheduling as a reaction to events enhances the reliability, robustness and performance of a production planning and control system. Thus maintains the desired manufacturing objectives, such as on-time completion of customer orders, factory throughput or production cycle time.

Compared with the conventional scheduling, rescheduling has to fulfill two new requirements:

- *Efficiency*: In contrast to the offline scheduling approaches that run a long time (from several hours to several days) to generate a near optimal schedule prior to a production process starts, rescheduling is applied during a production process. Hence, the production has to stop and wait for the new schedule. Any delay of the production process caused by the computational time of the rescheduling approach lengthens flow time of the production process and affects the delivery dates of the customer orders. Therefore, rescheduling is time critical and typically has to be a real-time application.
- *Flexibility*: The initial state of a rescheduling problem is normally more complex than that of a scheduling problem. Rescheduling has to take more aspects of the production systems state into account, e.g., release times of machines.

Furthermore, a necessary information system (sensors, communication networks, hardware and software) for monitoring unexpected events as well as the manufacturing system state (jobs and machines) is a precondition to apply the rescheduling.

Vieira et al. (2003) summarizes works in production rescheduling and presents a framework (see Table 2) for understanding and classifying rescheduling research. The framework includes rescheduling environments, strategies, policies, and methods. "The rescheduling environment identifies the set of jobs that need to be scheduled. A rescheduling strategy describes whether or not production schedules are generated. A rescheduling policy specifies when rescheduling should occur. Rescheduling methods describe how schedules are generated and updated".

For event management in production we focus on the predictive-reactive rescheduling strategy and its event-driven rescheduling policy, in which an initial schedule is generated before a production process start and this schedule will be updated if an event occurs during the production process. For this strategy and policy Vieira et al. (2003) introduces three categories of rescheduling methods to

Table 1 Uncertainties related with production and transportation planning processes at the operational level

Type of uncertainty	Triggers related with production planning	Triggers related with transportation planning
Supply	Raw material delay	Single sourcing under JIT strategies
	Quality problem	Lack of carrier flexibility (time, fleet options and capacity)
Production	Machines breakdown	Lack of integration with production
	Job cancelled (dropped/destroyed)	Production system delay
	Urgent order (rush or "hot") arrival	Production order cancelled
	Over/under estimated processing time	
	Job goes on hold	
	Number of jobs exceed threshold	
	Schedule not followed by personal	
	Unexpected maintenance	
	Rework or quality problems	
Transportation	Transportation system delay	Shipment time variability
		Loss or damage of materials in the shipping operation
		Fleet breakdown or lack of drivers
		Unexpected repairs for preferred transport routes
Information	Poor stock auditing and poor quality control system	Poor stock auditing and poor quality control system
	Database inaccuracy	Database inaccuracy
Demand	Urgent (rush or "hot") job arrival	Customer order cancelled (dropped/destroyed)
People management	Personnel delay	Lack of drivers

Table 2 Rescheduling framework of Vieira et al. (2003)

Rescheduling environments				
Static (finite set of jobs)		Dynamic (infinite set of jobs)		
Deterministic (all information given)	Stochastic (some information uncertain)	No arrival variability (cyclic production)	Arrival variability (flow shop)	Process flow variability (job shop)
Rescheduling strategies				
Dynamic (no schedule)		Predictive-reactive (generate and update)		
Dispatching rules	Control-theoretic	Rescheduling policies		
		Periodic	Event-driven	Hybrid
Rescheduling methods				
Schedule generation		Schedule repair		
Nominal schedules	Robust schedules	Right-shift rescheduling	Partial rescheduling	Complete regeneration

repair the initial schedule. They are right-shift rescheduling, partial rescheduling and complete regeneration.

Pfund et al. (2006) processed a survey in 14 semiconductor companies and shows that the state of the practice for rescheduling is to use dispatching rules. The framework of Vieira et al. (2003) defines dispatching rules as a dynamic rescheduling strategy without generating a new schedule. Nevertheless, with light extensions [e.g., by means of the active schedule generator of Giffler and Tompson (1960)[2]] dispatching rules can also generate schedules and be applied as the predictive-reactive rescheduling strategy. Dispatching rules can repair the current schedule partially or generate a new schedule completely.

In the practice, dispatching rules are simple to be implemented and applied in production processes. In addition, they fulfill the efficiency and flexibility requirements of rescheduling. Priore et al. (2001) as well as Chen and Yih (1996) indicate that the performance of dispatching rules depends on the state of the manufacturing system at each moment, and no single rule exists that is better than the rest in all the possible system states.

Tan and Aufenanger (2011) show a trend and the potential of using artificial intelligence in the event management for uncertainties in the production scheduling. They introduce a machine learning process to acquire the knowledge about the relationship between system states and dispatching rules. With this acquired knowledge they develop a knowledge-based rescheduling approach, which can dynamically select the most appropriate dispatching rule depending on the current production system state.

Figure 1 illustrates the learning process and the rescheduling approach of Tan and Aufenanger (2011). The approach consists of offline and online two phases. In the offline learning phase it uses scheduling approaches to generate schedules for the previous scheduling problems of the production system. Then, each pair of problem and its schedule is analyzed by an analysis heuristic. Training data in the form of production system state, then dispatching rule are generated that denotes for one given state of a previous problem which dispatching rule is selected by its schedule. Afterwards, a machine learning process is applied on these training data to acquire knowledge about the relationship between system states and dispatching rules in the considered production system.

In the online phase, when this production system has an unexpected event (e.g., delay of the arrival of raw materials due to a transportation delay) and the current production schedule has to be updated, the rescheduling approach considers the current system state and dynamically selects the most appropriate dispatching rule for this state, based on the acquired knowledge. The approach iteratively updates the system state and selects the best dispatching rule for each state, until the whole new schedule is completed.

[2] The set of active schedules is a subset of feasible schedules for a scheduling problem. Giffler and Tompson (1960)proved that at least one optimal schedule is active schedule. Their work also presents a heuristics, which can generate all possible active schedules. Dispatching rules are usually used to lead the search directions in this heuristics to generate active schedules.

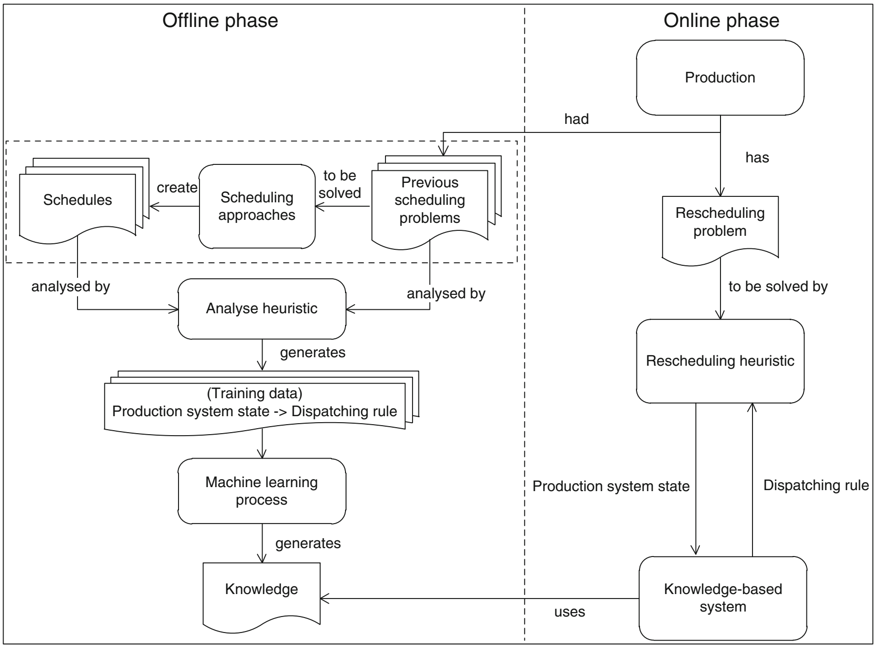


Fig. 1 The knowledge-based rescheduling concept for dynamically selecting dispatching rules of Tan and Aufenanger (2011)

Tan and Aufenanger (2011) reported this knowledge-based rescheduling approach outperforms all single dispatching rules that they studied. In addition, the offline knowledge acquisition phase saves the computational time of the online phase and ensures the efficiency (near real-time) of the rescheduling.

Event Management for Uncertainties in Transportation Planning

A broad view of transportation management's role in an integrated SC is a must nowadays because it is the key to improving the SC performance. Traditionally, the focus of managing uncertainty has been on production operations, with little attention paid to the causes and consequences of uncertainty within freight transport operations. Consequently, there has been little integration of transport in supply chains (Rodrigue et al. 2007).

Transport planning is influenced by the uncertainties. The transport represents an important part of the supply chain due to the dynamics of the current business environment, which is characterized by short products life cycles and large varieties of the products. Performance of transport can impact the wider supply chain (Larsen et al. 1999). Therefore, effective and efficient logistic management has become a requirement rather than a competitive advantage under these complex and dynamic environment. Accordingly, increased flexibility in transport services to meet a variety of customer demands, and the involvement of the shipper, carrier and customer as part of a logistics trade are highly required.

The design of transport network has to consider a number of requirements:

1. *Sustainability*: It is increasingly seen as essential to delivering long-term profitability. The targeted benefits include shorter product life cycles, faster product development cycles, globalization and customization of product offerings, and higher overall quality (Fortes 2007). Construct automated decision support system helps the decision maker in determining the optimal distribution schedules and the optimal distribution sequences (Bonfill et al. 2008).
2. *Flexibility*: The increase of structural and dynamic complexity of production and logistics systems could be observed. Companies should be able to respond to diversity or change of environment.
3. *Visibility*: It is about measuring time and accuracy of information transfer. The transfer of information between operational levels could be delayed or inaccurate, therefore affect the efficiency of the whole network.
4. *Reliability*: It affects customers trust through different levels of the supply chain. The importance is the supplier's performance, how consistent suppliers deliver raw materials on time in good condition.

Uncertain events can affect the ability of transport operations to satisfy customers' requirements. The customer satisfaction could associate with the service before or after purchasing products or service as well as with service elements directly involved in the physical products distribution (Li and Schulze 2011). Thus, there is a need to identify the sources of transport uncertainty as a mean of improving the effectiveness of management (Sanchez-Rodrigues et al. 2010a). From transport perspective, various sources of uncertainty can exist within logistics networks. Uncertainty can be initiated from one source and can possibly affect many of the logistics processes (Sanchez-Rodrigues et al. 2008). The uncertainty sources are mainly categorized into five types which are (Sanchez-Rodrigues et al. 2010a):

1. *Shipper*: any uncertainty originating from the sender of products in the logistics trade, which directly impacts transport performance. These may relate to raw material sourcing, the production processes or the activities involved in the dispatching process such as: shipment time variability, loss or damage of materials in the shipping operation and lack of integration with production.
2. *Customer*: any uncertainty that is produced by the receiver of products. Examples include forecasting and ordering products or any delivery restrictions. Uncertainties could be caused by variations in customer demand for transport, rigid delivery window, equipment breakdowns, or lack of integration within SCs.

3. *Carrier*: any anomalies that can be originated from the carrier and directly affect the delivery process, such as transport delays due to internal reasons like: vehicle failure or a lack of drivers, insufficient fleet capacity, lack of carrier flexibility in terms of time, lack of integration between transport modes and providers or lack of flexibility of shipment and transport schedule.
4. *Control systems*: any problems caused by inadequate and fragmented information and communications technology systems within the logistics trade such as lack of integration when SC companies assess, transport and inventory systems are not properly integrated, lack of visibility of information regarding inventories or work in process, capacity, order status within the SC, or the lack of physical monitoring systems such as poor auditing or quality control systems.
5. *External uncertainty*: any disruption caused by uncontrollable transportation factors, including congestion, labor shortages, unexpected repairs for preferred transport routes and volatility of fuel prices or even problems that cannot be predicted in any way, such as political and natural disasters.

SCEM has tools that can reduce or eliminate uncertainties when they applied. The most important tools that are: strategic optimization (such as network modeling software), which ensures that the distribution networks are robust to disruptions, operational optimization (such as vehicle scheduling), which allows businesses to respond to uncertainty as quickly as possible while minimizing the overall impact, quality management (like total quality management), which allows to address the causes of uncertainty, to reduce their frequency and to impact in the longer term, and demand forecasting, which are designed to improve accuracy and, therefore, reducing uncertainty (Sanchez-Rodrigues et al. 2010b).

Event Management in Collaborative Production and Transportation Planning

In a collaborative environment of SC the requirement for synchronization of production and transport planning is characterized by the breadth of effects of unexpected events of one process in the other. For example, an urgent customer order (rush order) leads changes in the production planning, and consequently generates a delay in the loading operation of the vehicle, which causes another event affecting the transportation planning.

This mutual influence of unexpected events in production and transport system requires planning makers to consider the production and transport planning in an integrated way, where the visibility of the two simultaneous processes is provided. Scholz-Reiter et al. (2011) presents a mathematical model to describe the integrated production and transportation scheduling problem. Their model considers one production facility and its associated transportation along the SC.

Based on an integrated production and transportation planning model at the operational level, strategies and methods of event management for uncertainties in production planning can be extended to transportation planning, and reversely. Hence, an integrative event management for uncertainties in production and transportation planning can be achieved.

Conclusions and Future Work

This paper introduced the event management for uncertainties for collaborative production and transportation planning at the operational level. That adds value to SCs in a dynamic environment. Considering uncertainties of SC processes simultaneously leads to maximize the benefit of reducing or eliminating risks.

For future research, more review related to integration between production and transportation is required. In addition risk based planning and scheduling approach is one of the possible research directions. This approach combines the features from a simulation model that generates a detailed resource-constrained deterministic schedule and a probability-based risk analysis considering the deviations from SC process. Furthermore, global supply network aspects such as culture aspects and specific regulations should be taken into account.

Sustainability in the supply chain is increasingly seen as essential to delivering long-term profitability. Achieving a level of integration between production and distribution decisions will yield these benefits. These benefits include shorter product life cycles, faster product development cycles, globalization and customization of product offerings, and higher overall quality (Fortes 2007).

Acknowledgments This research was supported by the International Graduate School for Dynamics in Logistics (IGS) at the University of Bremen, by Capes as part of the Brazilian-German Collaborative Research Initiative on Manufacturing Technology (BRAGECRIM) and by Deutscher Akademischer Austausch Dienst (DAAD) and the Egyptian Government under Grant GERLS 2010.

References

Bonfill A, Espuna A, Puigjaner L (2008) Decision support framework for coordinated production and transport scheduling in SCM. Comput Chem Eng 32:1206–1224
Chen CC, Yih Y (1996) Identifying attributes for knowledge-based development in dynamic scheduling environments. Int J Prod Res 34:1739–1755
Christopher M, Mena C, Khan O, Yurt O (2011) Approaches to managing global sourcing risk. Supply Chain Manag: Int J 16(2):67–81
Fortes J (2007) Green supply chain management: a literature review. Otago Manag Graduate Rev 7:51–62
Giffler B, Tompson GL (1960) Algorithms for solving production-scheduling problems. Oper Res 8:487–503
Khan O, Burnes B (2007) Risk and supply chain management: creating a research agenda. Int J Logistic Manag 18(2):197–216

Larsen ER, Morecroft JDW, Thomsen JS (1999) Theory and methodology: complex behavior in a production-distribution model. E J Oper Res 119:61–74

Li L, Schulze L (2011) Uncertainty in logistics network design: a review. In: Proceedings of the international multi conference of engineers and computer scientist II, IMECS 2011

Manuj I, Mentzer JT (2008) Global supply chain risk management strategies. Int J Phys Distrib Logistics Manag 38(3):192–223

Otto A (2003) Supply chain event management: three perspectives. Int J Logistics Manag 14(2): 1–13

Pfund ME, Mason SJ, Fowler JW (2006) Semiconductor manufacturing scheduling and dispatching. In: Herrmann JW (ed) Handbook of production scheduling. Springer, New York

Priore P, De La Fuente D, Gomez A, Puente J (2001) A review of machine learning in dynamic scheduling of flexible manufacturing systems. Artif Intell Eng Des Anal Manuf 15:251–263

Ritchie B, Brindley C (2007a) Supply chain risk management and performance: a guiding framework for future development. Int J Oper Prod Manag 27(3):303–322

Ritchie B, Brindley C (2007b) An emergent framework for supply chain risk management and performance measurement. J Oper Res Soc 58(11):1398–1412

Rodrigue VS, Stantche D, Potte A, Naim M, Whitein A (2007) Establishing a transport operation focused uncertainty model for the supply chain. In: 14th international annual Euroma conference

Sanchez-Rodrigues V, Potter A, Naim M (2010a) Evaluating the causes of uncertainty in logistics operations. Int J Logistics Manag 21(1):45–64

Sanchez-Rodrigues V, Potter A, Naim M (2010b) The impact of logistics uncertainty on sustainable transport operations. Int J Phys Distrib Logistics Manag 40(1):61–83

Sanchez-Rodrigues V, Stantchev D, Potter A, Naim M, Whiteing A (2008) Establishing a transport focused uncertainty model for the supply chain. Int J Phys Distrib Logistics Manag 38(5):388–411

Stank TP, Dittmann JP, Autry CW (2011) The new supply chain agenda: a synopsis and directions for future research. Int J Phys Distrib Logistics Manag 41(10):1–25

Scholz-Reiter B, Novaes AGN, Makuschewitz T, Frazzon EM (2011) Dynamic scheduling of production and inter-facilities logistic systems. Dynamics in logistics. Springer, Berlin

Tan Y, Aufenanger M (2011) A real-time rescheduling heuristic using decentralized knowledge-based decisions for flexible flow shops with unrelated parallel machines. In: Proceedings of the 9th IEEE international conference on industrial informatics. pp. 431–436

Tang CS (2006) Perspectives in supply chain risk management. Int J Prod Econ 103:451–488

Vieira GE, Herrmann JW, Lin E (2003) Rescheduling manufacturing systems: a framework of strategies, policies, and methods. J Sched 6:39–62

Study on the Application of DCOR and SCOR Models for the Sourcing Process in the Mineral Raw Material Industry Supply Chain

Raul Zuñiga, Marcus Seifert and Klaus-Dieter Thoben

Abstract One of the main avenues to improve the whole supply chain performance in today's business environment is a better understanding about the behavior of the processes in the supply chain. This is more critical if these processes belong to the early part of the supply chain in the mineral raw materials industry. This paper examines how the sourcing process in this industry is represented in today's supply chains. Most of the concepts, variables, and factors related to supply chain and supply chain management, mineral raw materials industry, the exploration, development and extraction processes are extracted from literature. All of them are presented by using the existing models, Design Chain Operations Reference (DCOR) model and Supply Chain Operations Reference (SCOR) model. This work finds that DCOR and SCOR models can be extended to model the sourcing process. Furthermore, it demonstrates the possibility to extend the SCOR model to the rest of the early supply chain processes of this industry.

R. Zuñiga (✉) · M. Seifert · K.-D. Thoben
Bremer Institut für Produktion und Logistik GmbH an der,
Universität Bremen, Hochschulring 20, 28359 Bremen, Germany
e-mail: Zun@biba.uni-bremen.de

M. Seifert
e-mail: Sf@biba.uni-bremen.de

K.-D. Thoben
e-mail: tho@biba.uni-bremen.de

R. Zuñiga
Engineering Department, Arturo Prat University,
Arturo Prat 2120 Av, Iquique, Chile

R. Zuñiga
International Graduate School for Dynamics in Logistics,
University of Bremen, Hochschulring 20,
28359 Bremen, Germany

H.-J. Kreowski et al. (eds.), *Dynamics in Logistics*, Lecture Notes in Logistics,
DOI: 10.1007/978-3-642-35966-8_17,

Keywords Mineral raw material industry · Supply chain · Supply chain management · Sourcing process · DCOR and SCOR models

Introduction

Supply chain models have been implemented nowadays in several industries. However, as the supply chain model was originally designed to fit mostly in a manufacturing environment, few related research has been focused on the mineral raw material industry (Tardelli et al. 2004). This industry and its sourcing process in particular—present many atypical characteristics, which are difficult to comply with existing Supply Chain models, and very limited research has been conducted covering this early part of the supply chain. A better understanding about the complexity and characteristics of the processes in the mineral raw materials industry can contribute to improve the early supply chain processes. Thereby improve the entire supply chain performance. For this, it is relevant to examine how the sourcing process works in this industry. The problem is, that the current modelling approaches to supply chains consider raw material industry as an infinite source, a black-box with unlimited resources. Therefore the supply of mineral raw materials has been more or less excluded from existing models so far (Vial 2004). This research examines how the sourcing process of this industry is currently represented in supply chains. In addition, it identifies how this process fits into the existing DCOR and SCOR models.

The remainder of this paper is a follows. Section Mineral Raw Materials Industry provides a review of some of the relevant literature on the mineral raw materials industry characteristics, mine life cycle and mine project development. Section Supply Chain, Supply Chain Management Sourcing Process, DCOR and SCOR Models gives some definitions on supply chain management applied to this industry and introduces the reference models, SCOR and DCOR. Finally, Sect. Sourcing Process by Using DCOR and SCOR Models defines the selected model to be used in this research to model the sourcing process of this industry.

Mineral Raw Materials Industry

A unique characteristic of mineral raw materials industry, intrinsic of its own nature, is the fact that the main raw material—the ore—is originated from an internal source, the mineral deposit (Tardelli et al. 2004). For that, the main characteristic in this industry in comparison to other industries is determined by the geological environment around the mineral deposit. Also, the mine life cycle determines the projects developments that influence the processes of this industry.

Features and Constraints

The special features of mineral resources deposits, initially unknown, fixed in physical size and location, and variable in quality, result in characteristics that create both problems and opportunities (Maksaev 2004). For this, the following characterize the mineral resource deposit.

For one, mining activities involve high-risk investments (Vial 2004). Mineral deposits have to be located and delineated, entailing large exploration costs, before considering normal industrial development and production decisions. Consequently mineral exploration is an integral part of the mineral raw materials industry; in fact successful exploration is essential for mining companies to survive over time. Also, mineral deposits once they are discovered are fixed size determined by nature, and therefore they are subject to depletion during the normal course of mining production. Consequently, it is necessary to engage in continuous and successful exploration efforts just to maintain existing production levels.

The inherent variability in quality and other geological parameters can also result in the variability of mine site revenue, and the operating and capital costs that ultimately affect the returns on investment. For this, one of the most important competitive advantages of a mining company is the high quality of its mineral deposits. To avoid the cost of transporting non-mineral waste rock with the product, companies choose to process ore adjacent to the mineral deposit. For that, mineral deposits cannot be moved to a convenient location.

There is always a level of uncertainty and an estimation error in geologic modelling and geostatistics systems, estimated ore quality parameters, such as mineralogy, metallurgical and chemical grades, granulometry. These parameters will only be known after the exploitation, already in the production process. Also, the uncertainties are in the production process and therefore it is required to work with stockpiles of intermediate and final products. Mining companies normally produce raw materials to stock, not to order (Tardelli et al. 2004).

Mine Life Cycle

The main phases of the mining cycle are: mineral exploration, mine development, mine operation, and mine closure (Natural Resources Canada 2006). By understanding the activities and behaviour of processes over their complete mine life cycle, it can potentially improve the processes in the early part of the supply chain. Thereby, enhance the supply of raw materials in the marketplace.

Mineral exploration. Every new mine starts as an exploration project. However, most exploration projects will not advance to become mines. EMMMV (2010) indicates that the types of mineral exploration are: Greenfield, Brownfield, and on-mine-site. It will take from 7 to 10 years before the start of a new mine if the exploration programs get good results. The most relevant mineral exploration activities are: prospecting, mining claim staking, detailed exploration, sampling and drilling, environmental baseline work, and preliminary deposit evaluation.

Mine development. The objective of this phase is to learn about the potential value of a mineral deposit. If initial exploration gets good results, the project moves from exploration to the deposit evaluation and mine planning stage. It is in this stage where the mining company will increase its activities and investment to determine if the mineral deposit is worthwhile and if a mine can be developed. Also the company will prepare the design of the mine. The construction can start once evaluation and planning are completed and a decision on building a mine has been taken.

Mine operation. This phase involves the hiring, training, commissioning, production, and mine expansion. There are two types of mines: underground and open pit. In the production process is the extraction of ore, separation of minerals, disposal of waste and shipment of ore/minerals. Additional sampling, drilling, planning and mapping is required if a mine expands its life.

Mine closure. The mineral deposits have a finite life. A mining company closes for different reasons, but the two most common are: running out of the ore resource; and low commodity or metal prices, which make the mine uneconomic to operate.

New Projects Development and Projects for Preparation

Mining companies reduce the size of their mineral resources each day and eventually need to replace the mineral deposit. However, finding, evaluating, and developing a new mineral deposit is technically and economically complex, expensive, and long-term. Thus, competition in this industry plays itself out predominantly in competition of new mine projects, rather than in a customer-focused approach to winning market share.

The mining company will try to increase production when the market conditions improve. When this happens, the only way to achieve permanent increases is to expand the mine. Mine expansion can include: enlarging the existing mine; opening up more mine areas; buying more equipment and hiring more people; expanding the processing plant to process more ore; changing the processing plant to process faster; and doing more exploration work to try to find more ore. For this expansion is required to perform new project development (Greenfield or Brownfield).

Greenfield projects are for new minerals deposits and Brownfield projects are for existing operations. Creating a new mine expansion requires project involving, research in geosciences (exploration), product (ore) research, design and constructs the mine. Upon completion of these projects, the production of many units of those mineral raw materials begins.

The project development for preparation is performed in current operations. In the production process, waste still needs to be excavated in order to keep uncovering more ore. This is now called “development” in the underground mining and "stripping" in open-pit mining. Development and/or stripping go hand-in-hand with production and are critical to maintain production. Also, exploration on-site-mine is required, and study, design and construct activities.

Supply Chain, Supply Chain Management, Sourcing Process, DCOR and SCOR Models

The physical existence of mineral deposits in nature and mineral raw materials demand in the domestic or global economy are the basic incentives for mineral supply in the supply chain. The way how to achieve economical production of mineral raw materials consists of a sequence of processes through which minerals are transformed from an unknown geological resource to a marketable material (see Fig. 1). A better understanding about the behaviour of the processes in the early part of the supply chain in this industry increases the chances to improve the whole supply chain performance in today's business environment. For this, it is relevant to examine how the sourcing process works in this industry. This means to include the first supplier processes of raw materials to this industry, and to describe the existing DCOR and SCOR models for modelling these processes.

The Early Part of the Supply Chain and the Sourcing Process

Lambert et al. (1998) define Supply Chain Management as "the integration of business processes from end user through original suppliers that provides products, services, and information that add value for customers". According to the above definition, the role of mineral raw materials industry in the supply chain is to find, delineate and develop mineral deposits and then to extract, process and sell the raw materials derived from deposits. Consequently, the mineral deposits are a central point of the mineral supply process to the raw materials industry.

Guoqing et al. (2003) indicates that the difference between the mineral raw materials industry and the manufacturing industry is the way to obtain raw materials. But compared under the supply chain management view, some common

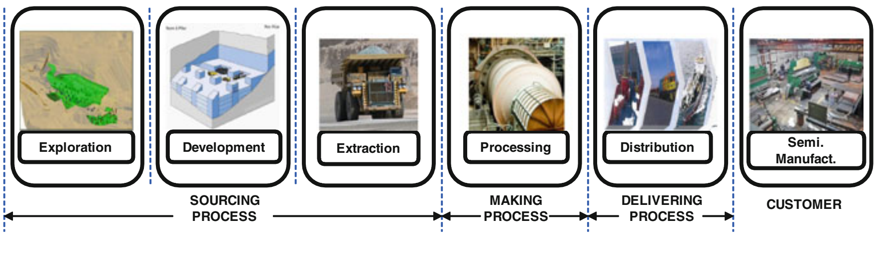


Fig. 1 Processes of the mineral raw materials industry

Fig. 2 The extraction in manufacturing industry (*left*) and the extraction in mineral raw material industry (*right*)

points can be found. The cost of exploring and mining in the mineral raw materials industry can be regarded as the purchase cost in manufacturing industry. However, this comparison is valid when the sourcing process is considered as a black-box in the supply chain modelling. For understanding the complexity and behaviour of this process in the early part of the supply chain, the exploration, development and extraction processes must be included in the sourcing process modelling.

Tardelli et al. (2004) emphasize that another important restraining factor for the planning and scheduling function is the sequencing of the mine exploitation. In order to be properly extracted, one specific ore block must be at the face of the open pit or underground mine, so mining equipment can have access to it (Fig. 2 right). A manufacturing plant does not have this kind of restraining factor, as any component in a production material warehouse can be easily accessed at once and fed into the assembly line (Fig. 2 left). Most of the traditional frameworks to supply chain modelling are focused on manufacturing industry and they do not include the sourcing process for the raw material industry. Thus, it is important to study how the existing DCOR and SCOR models can be used to model the sourcing process of this industry.

The DCOR and SCOR Models

The Supply Chain Council, based on the successful supply chain operations reference model, proposed a DCOR model to assist firms constructing their framework. The model consists of five primary processes: Plan, Research, Design, Integrate and Amend (SCC 2006). The DCOR concept provides considerable reference value for firms. However, DCOR version 1.0 only offers preliminary definitions, such as those for major design procedures, and lacks a best practice example. Furthermore, few studies have attempted this issue (Wu et al. 2007). This research considers the concepts of design chain, exploration process and mine design process in proposing a model based on DCOR, as a part of the sourcing process in the mineral raw material industry.

The framework receiving the most attention is the SCOR model. SCOR is a model that evolved from a company's effort to introduce efficiency into the Supply Chain. SCOR describes a Supply Chain as consisting of five primary processes; Plan, Source, Make, Deliver, and Return (SCC 2010). These processes are Level 1 processes within the SCOR hierarchy. Level 2 processes describe three process types; Planning, Execution and Enablement. Level 3 are the standardized operations of the Level 2 processes. Level 4 enhances each of the Level 3 processes specific to the organizations needs.

Sourcing Process by Using DCOR and SCOR Models

Mineral raw materials companies focus on the mine design to produce a product (ore), by using the exploration results created by the exploration and feasibility processes. The tighter the collaboration between these processes, the better the design of the entire mine. These processes can be represented by using the DCOR model. Then, the construction and extraction processes are modeled by using the SCOR model.

The Exploration and Feasibility and Design with DCOR Model

Figure 3 depicts the DCOR model with an example as follows. The company has a dedicated R and D team focused on exploration and feasibility studies. The work of this team is driven by the overall product (ore) roadmap. The Plan processes a link to the DCOR processes. Research and Design are driven by this plan. The design team works off the same product (ore) roadmap. But, in this case the key element is the mine design. The research for the mine design encompasses the linkage to the work of the exploration and feasibility team. This team works with the design team to "integrate" the exploration and feasibility results into the mine design. Thus, Integrate processes link to Research processes.

The Construction and Extraction Processes with SCOR Model

Construction industry is different to the typical manufacturing industry for the following cases: Customers should come to the place and use the product because

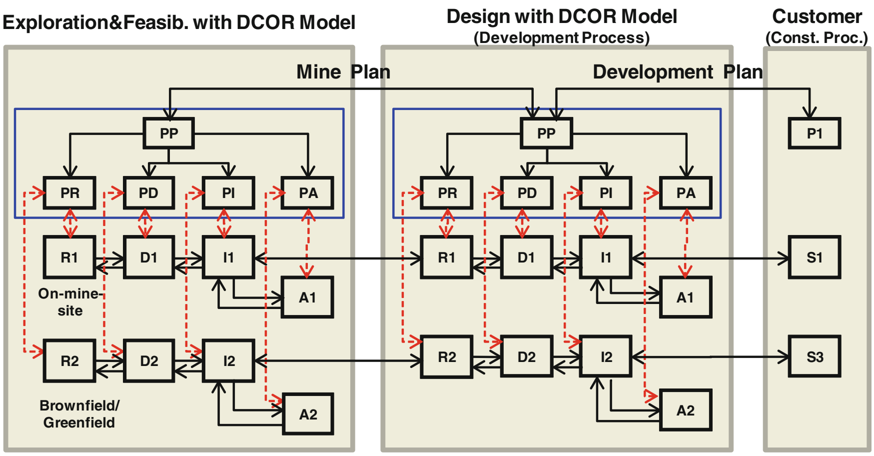


Fig. 3 The exploration and feasibility and design processes with DCOR model

there is no possibility to transport the final product to another place. There are no rapid changes in technology, for example in comparison with the computer industry. There are no changes in customer demands and sometimes there is no chance of any change for customers after defining the agreed goal. Hue et al. (2007) depicts in Fig. 4 (left) that construction is a multi-organization process, which involves client/owner, designer, contractor, subcontractors, specialist contractors, and suppliers.

Morris and Pinto (2007) indicate that the SCOR model can be adapted to describe a project supply chain. They explain that in a project supply chain context, the planning process encompasses all aspects of planning, including the integration of the individual plans of all supply chain members, into an overall project supply chain plan. For example, in the sourcing process the focus is on all processes related to procurement, such as identifying, selecting and qualifying suppliers, contract negotiation, and inventory management, and so on.

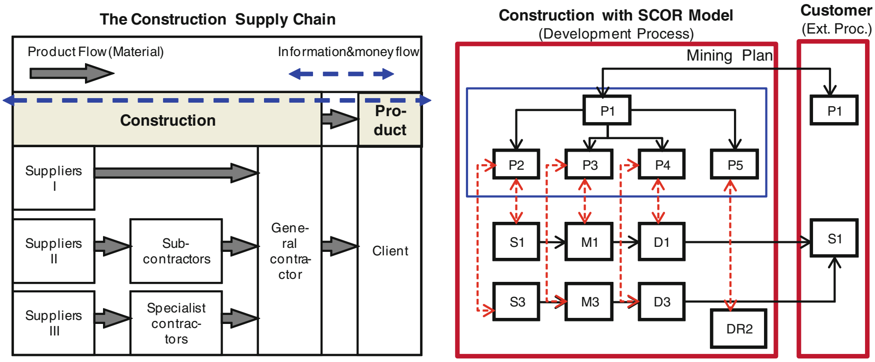


Fig. 4 The construction process supply chain (*left*) and its SCOR model (*right*)

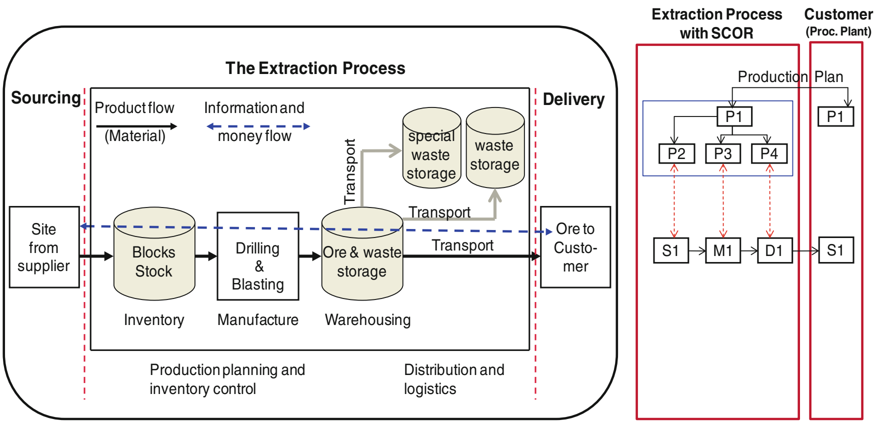


Fig. 5 The extraction process supply chain (*left*) and its SCOR model (*right*)

Figure 5 illustrates the extraction process modelling. This process can be explained as follows. The processing plant (customer) requests a specific quantity of ore from the extraction process. Information starts to flow from the customer to the extraction process. This process then searches the warehouse (ore stock), which would be an ore storage warehouse, to establish whether the ore is available or not. If available, the ore is sourced from the warehouse and delivered to the customer via a transport system. The delivery mechanism is the logistics of how to distribute the product (ore), and infrastructure is part of the distribution planning. When the ore is not available in the warehouse, it has to be manufactured by extraction process. The extraction process then plans the ore and sends information to interrogate the inventory for ore blocks stock. Some of the ore blocks stock needed for the ore are held in inventory. The sourcing request is sent out to the supplier, the development process, which includes the design and construction processes. Then, the supplier delivers an available site with the ore blocks requested. The extraction process keeps blocks in inventory until they are needed to extract the ore. This is the production planning and inventory control. Once the ore has been completed, the extraction process delivers ore to the customer via the warehouse as described above.

In addition, it has also shown the flow of information and money between the different processes. Thus, a product (ore) produced by the extraction process can be managed using supply chain management. Also, Fig. 5 (right) depicts the extraction process by using the SCOR Model.

Conclusions

This paper showcases how SCOR and DCOR models may be extended in order to model the sourcing process of the mineral raw materials industry. Based on the results shown in this paper the following conclusions can be obtained.

1. The SCOR model, which is a quasi-standard in manufacturing industry, cannot be fully extended to model this process.
2. This sourcing process has an important difference in comparison to the sourcing process of manufacturing industry.
3. The DCOR model allows modelling the earliest processes of the sourcing process and then, it can be integrated with the SCOR model.
4. It requires the analysis of Level 3 processes to identify the Key Performance Indicators and Best Practices, which are more suitable to this sourcing process.
5. There exist a promising potential of SCOR model to be extended to the rest of the processes in the early part of the supply chain in this industry.

References

Exploration Mining Metals and Minerals Vertical EMMMV (2010) The exploration and mining business reference model. The open group https://www.opengroup.org/emmmv/uploads/40/22706/Getting_started_with_the_EM_Business_Model_v_01.00.pdf. Accessed 2 Nov 2011

Guoqing L, Nailian H, Xuchun H (2003) Study on the application of supply chain management (SCM) for jiaojia gold mine. Applications of computers and operations research in the minerals industries. South African Institute of Mining and Metallurgy

Hue X, Wang Y, Shen Q et al (2007) Coordination mechanisms for construction supply chain management in the internet environment. Int J Project Manage 25:150–157

Lambert DM, Stock JR, Ellram LM (1998) Fundamentals of logistics management. Irwin/McGraw-Hill, Boston

Maksaev V (2004) Introducción a la economía minera: definición y aplicación [online]. Available from http://www.cec.uchile.cl/∼vmaksaev/ECONOMIA%20MINERA.pdf. Accessed 4 Oct 2010

Morris PW, Pinto JK (2007) The wiley guide to project technology, supply chain and procurement management. Wiley, Inc. Hoboken

Natural Resources Canada (2006) Mining information kit for aboriginal communities: exploration, development, operation and closure. Available from http://www.pdac.ca/pdac/advocacy/aboriginal-affairs/2006-mining-toolkit-eng.pdf. Accessed 10 June 2011

Supply Chain Council SCC (2010) Supply chain operations reference model SCOR version 10.0. The supply chain council, Inc. SCOR: the supply chain reference ISBN 0-615-20259-4 (binder)

Supply chain council SCC (2006) Design chain operations reference model (DCOR) version 1.0. The supply chain council, Inc.

Tardelli P, Barbin F, Cesare de Tomi G (2004) ERP in mining industry: studying the software functionality and the value chain. In: second world conference on POM and 15th annual POM conference, Cancun México, 30 April–3 May 2004

Vial J (2004) Modeling commodity markets in the global economy: familiar findings and new strategies. The earth institute at Columbia University, working papers series center on globalization and sustainable development CGSD (18)

Wu Y, Yeh S, Fang L (2007) The development of a collaborative design chain reference model for the motorcycle industry. Int J Adv Manufact Technol 35:211–225

Part V
Identification Technologies

Service Oriented Platform for RFID-Enabled Real-Time Fleet Advanced Planning and Scheduling in Industrial Parks

L. Y. Pang, George Q. Huang and R. Y. Zhong

Abstract Industrial parks have been playing an important role in stimulating the economic and industrial development. Within industrial parks, materials flow with both input and output is heavy. The fleet management includes methods to maximize the utilization of fleet resources and efficiency of transportation plan becomes challenging. This paper proposes to design and develop a service-oriented system to assists stakeholders in an industrial park for collaborative transportation decision-makings and operations with considering real-time fleet resources and operational constraints. RFID technologies are used in industrial parks to create smart work environments for enhancing real-time information sharing. Some important standards such as ISA-95 and business-to-manufacturing markup language (B2MML) are adopted for sharing information among diversified enterprises.

Introduction

Industrial parks are a tract of land developed and planned for the specific industrial purposes (HEPD/IPPS 1997). Within an industrial park, companies are clustered together and designed to share common utilities and infrastructures. Centralized enterprises within a planned area favour the integration of supply chain that allows

L. Y. Pang · G. Q. Huang (✉) · R. Y. Zhong (✉)
Department of Industrial and Manufacturing Systems Engineering,
The University of Hong Kong, Hong Kong, People's Republic of China
e-mail: gqhuang@hku.hk

R. Y. Zhong
e-mail: zhongzry@hku.hk

L. Y. Pang
e-mail: h0664249@hku.hk

H.-J. Kreowski et al. (eds.), *Dynamics in Logistics*, Lecture Notes in Logistics,
DOI: 10.1007/978-3-642-35966-8_18,

for more holistic and coordinated planning and implementation strategies (Tudor et al. 2007). Transportation between manufacturers is managed by third-party logistics (3PLs) operators. The materials flow with both input and output are immense (Geng and Hengxin 2009). The seamless supply of materials to industries is a prerequisite for manufacturers' production planning and scheduling operations. Vehicle fleets play the important role in transportation. However, the fleet management becomes a challenge due to the stochastic arrival of delivery orders, dynamic status of vehicles, drivers etc.

This paper is motivated by some real-life problems in industrial parks. The first one is resource visibility. Resources such as vehicles and drivers are prone to be stochastic, non-predictable and difficult to acquire. Planners do not have actual information on hand for precise decision making resulting in inefficient planning and scheduling. The second problem is resources utilization. The current decision processes in company are based on workers' experience. There is no standard procedure or guideline on decision-making. Hence, the utilization of resources and efficiency of plan are fluctuated. The third problem is collaboration problem between different stakeholders. They lack a collaborative mode and suitable information system to help them cooperative decision-making and operations.

To address the above problems, many technologies and methods have been introduced and applied in fleet management environment. RFID technologies have been widely applied in warehousing and fleet management to collect real-time information (Chow et al. 2006; Böse et al. 2009). Many algorithms have been proposed to solve vehicle routing and scheduling problem such as tabu search heuristic, genetic algorithm and ant system (Gendreau et al. 2008; Montemanni et al. 2005; Baker and Ayechew 2003; Yizhong et al. 2008). Advanced planning and scheduling (APS) integrates decision procedures with considering optimization methods and operation data such as loads, vehicles and operators information (Meyr et al. 2005). Also, APS provides a standard decision procedure to guide decision maker in transportation decision-making (Fleischmann 2005).

The recent movements of APS mainly focus on its problem scope, framework, mathematical models and relevant algorithms. The real-life cases of APS design and implementation are rarely studied. Some practical limitations challenge the implementation of APS. First, the feasibility and practicality of APS heavily rely on operational data, but there are no methods on how to synchronize APS data and real-time operation data. Second, APS usually utilizes specific algorithms to address specific planning and scheduling problems, so it is hardly generalized to diversified enterprises which may own different business processes. Third, APS is difficult to change system parameters such as user interfaces and data formats to achieve simple-to-use and easy-to-implement customization. Gebauer and Schober (2005) have been highlighted that flexibility to change the decision support system is needed in order to cope with the occurrence of less foreseeable events.

In this paper, a service oriented system is proposed and designed to support real-time information sharing, optimize fleet planning and scheduling processes, and enhance collaborative decision among decision makers. The objectives of this paper are summarized as follows. Firstly, to develop a hierarchical fleet advanced planning and scheduling (FAPS) decision procedure for industrial parks. Secondly, to deploy RFID enabled smart objects and gateways to create smart operation environments and obtain real-time operating information. Thirdly, to develop a service oriented platform to support the FAPS decision procedure and collaborative decision-making between different users as well as increase the system flexibility, and scalability.

Overview of Fleet Advanced Planning and Scheduling

Figure 1 shows the overall real-time fleet management system infrastructure. The purposes of this proposed infrastructure are to first, apply RFID/Auto-ID technologies in fleet operational environment. Second, to develop a framework for fleet advanced planning and scheduling processes with taking account of real-time operational information. Third, to develop data exchange adaptors that applying international standard schema to exchange information among FAPS system modules, heterogeneous enterprise information systems (EISs), and RFID/Auto-ID devices.

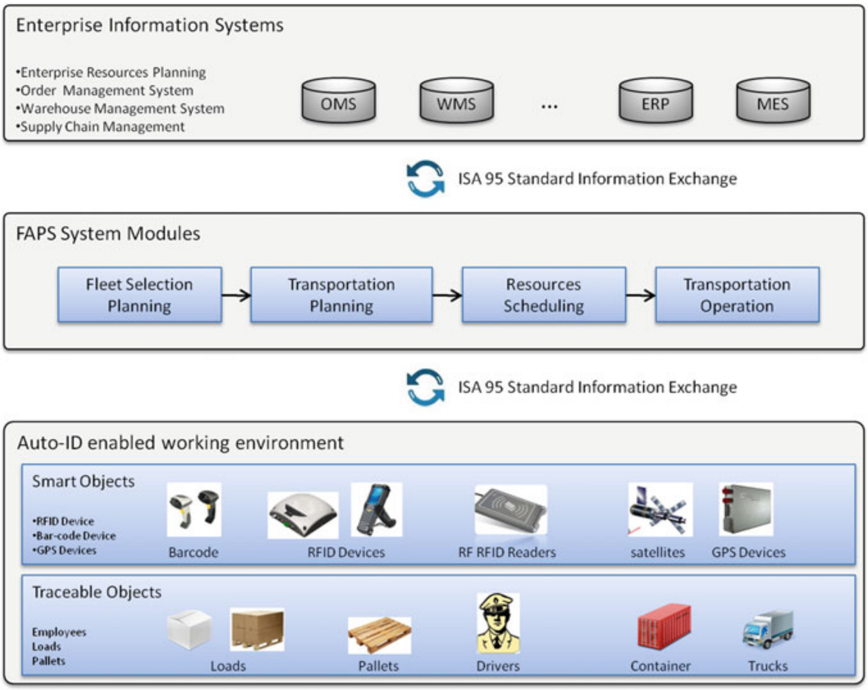


Fig. 1 Real-time fleet management system infrastructure

The real-time fleet management system infrastructure can be mainly divided into three levels. The higher level includes different EISs such as WMS, TMS, and ERP from different enterprises. The middle level is the FAPS system modules. These modules are developed for major fleet management decision makers and operators in an enterprise. The lower level is Auto-ID enabled working environment by using smart gateway and smart objects. Information adaptors that are compliant ISA-95 standard will be used for information exchange between FAPS system modules and external data sources such as EISs and RFID/Auto-ID devices.

Auto-ID Enabled Working Environment

The purpose of Auto-ID enabled working environment is to allow users to share real-time operation information. The development of this environment is composed of two steps, first step is creation of Smart Objects (SOs) for data collection and second step is deployment of Smart Gateways for data processing. The detailed steps will be introduced as follow:

The first step of creating RFID enabled working environment is to convert conventional objects into SOs and to equip the information value-adding operations points with suitable RFID readers for data collection. Pallets, locations, and loads are equipped with barcodes and RFID tags to make them identifiable. Their information will be collected when they pass through the information value-adding point which refers to a location or procedure that will update the transportation operations. For example, loads are moved out the warehouse representing customer required materials are in transited.

The second step is to deploy Smart Gateways for integrating and managing all the SOs deployed. The Smart Gateway has a hardware hub and a suite of management software to host all the SOs within a certain working area. There are 3 types of Smart Gateway, namely stationary, mobile and portable Smart Gateways. Stationary Smart Gateway is placed at a fixed location, such as the gate of a warehouse. A mobile Smart Gateway is installed to a moveable resource, such as a forklifts and trucks. A portable Smart Gateway is a handheld device responsible for distributed item identification within a certain area or along a certain process, e.g., with a driver in a logistics process.

FAPS System Modules

FAPS system modules are at the centre of the infrastructure. The system collects real-time information from different SOs through the associated gateways for adaptive transportation planning and scheduling decision and processes control. The FAPS system includes four main modules;

- **Fleet Selection Planning** module is responsible for deciding which fleet handles the manufacturers' transportation requests.
- **Transportation Planning** module is responsible for generating material transportation plans which includes what type and quantity of materials from the source location to destination within a specific time with an assigned route.
- **Resources Scheduling** module is responsible for assigning transportation resources to transportation plans which are created from the transportation planning module, based on real-time resources information.
- **Materials Transportation Operations** module is responsible for executing the scheduled transportation plans and updating the transportation progress. It includes two types of operation, external transportation operations which ship materials between different companies, and internal warehousing operations such as materials picking, moving, and storing within the company.

Fleet Advanced Planning and Scheduling Process

As shown in Fig. 2, the overall process of Fleet Advanced Planning and Scheduling (FAPS) contains 7 steps. From the perspective of process, the FAPS process can mainly divide into two levels, decision level and operational level. Generally, the aim of decision level modules is to help decision makers collaboratively making planning and scheduling decision in an adaptive approach with considering and integrating real-time operational information in the industrial park. The aim of operational level modules is to help operators executing scheduled transportation plan and real-time updating the operation progress.

Fleet Selection Planning

The key user of the module is the warehouse order manager. First, manufacturers' material transportation requests in the transportation request pool will be loaded into the fleet selection planning module. The transportation request pool is to collect and store materials transportation requests from manufacturers. The materials transportation requests contain required material type, quantity, and expected transportation time. Second, conditions of the requests, materials physical properties, geographical properties, logical properties, and handling requirements are obtained for evaluating the requests. They are shortlisted and released for optimization. Third, a proper fleet selection model will be selected and associated parameters will be defined before triggering the fleet selection planning service. The transportation request may be split into several sub transportation tasks for different fleets to collaborate execution. For example, a transportation task may split into two sub tasks, one sub task for forklift fleet prepares materials and another sub task for truck fleet transports prepared materials to destination.

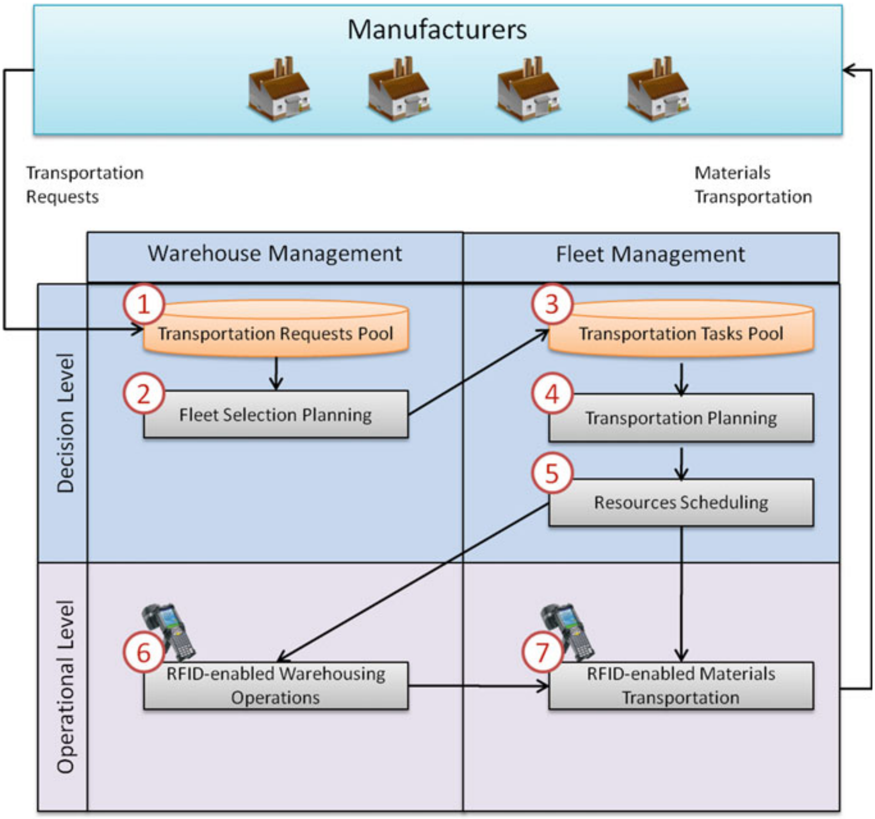


Fig. 2 Overall FAPS process

Then, after invoking the planning service, the results will be listed for reviewing and manually adjusting if necessary. Finally, the results will be released to corresponding fleet transportation tasks pool.

Transportation Planning

The key user of this module is the fleet planner. There are two types of fleets, external truck fleets for shipping materials between different companies, and internal forklift fleets for moving materials inside a company. Each fleet has their own transportation tasks pool. First, transportation tasks input from transportation tasks pool to transportation planning module. Second, related information such as conditions of the tasks, vehicle capacity, and other information will be loaded into the module. Third, a proper vehicle routing model will be selected and associated parameters will be defined. Forth, after invoke vehicle routing services, operators can review and fine-tune the result if necessary. Finally, the vehicle routing plan will be released to next module.

Resources Scheduling

The key user of this module is the fleet dispatcher. The objective of the dispatcher is to decide who is to do the transportation plan with what type of facility. The scheduling module involves several steps. First, transportation tasks are imported from transportation planning module. Second, transportation tasks are re-prioritized based on user's criteria. Third, scheduling services are invoked to assign specific tasks to specific operators. Forth, the assignment results are listed for scheduler to review and adjust. Finally, scheduled tasks are confirmed and released to individual operators.

Auto-ID Enabled Warehousing and Transportation Operation

The key user of this module is the forklift or truck drivers. Warehousing operations include order picking, loading and unloading, and put away. Transportation operation is to ship materials between different companies. All of the operations are similar and share a typical operation cycle. First, the operator obtains a scheduled transportation tasks from his/her own task pool. Second, the operator locates and takes suitable equipment (e.g., forklift, truck) for the chosen task. Finally, the operator follows the task (specification) sheet to pick up and move the materials to the designated location. Mobile SOs are used in all operations to help operators record and update the operation progress.

Design Considerations

FAPS system is designed for different users, including manufacturers, the public warehouse, and 3PL companies, in industrial park for collaborative transportation decision-making and operations. FAPS system and its main components are shown in Fig. 3.

FAPS system is designed with high usability by less trained users in operation. In order to meet these requirements, several key innovations have been adopted in designing, implementing and deploying FAPS system. They are summarized as follows:

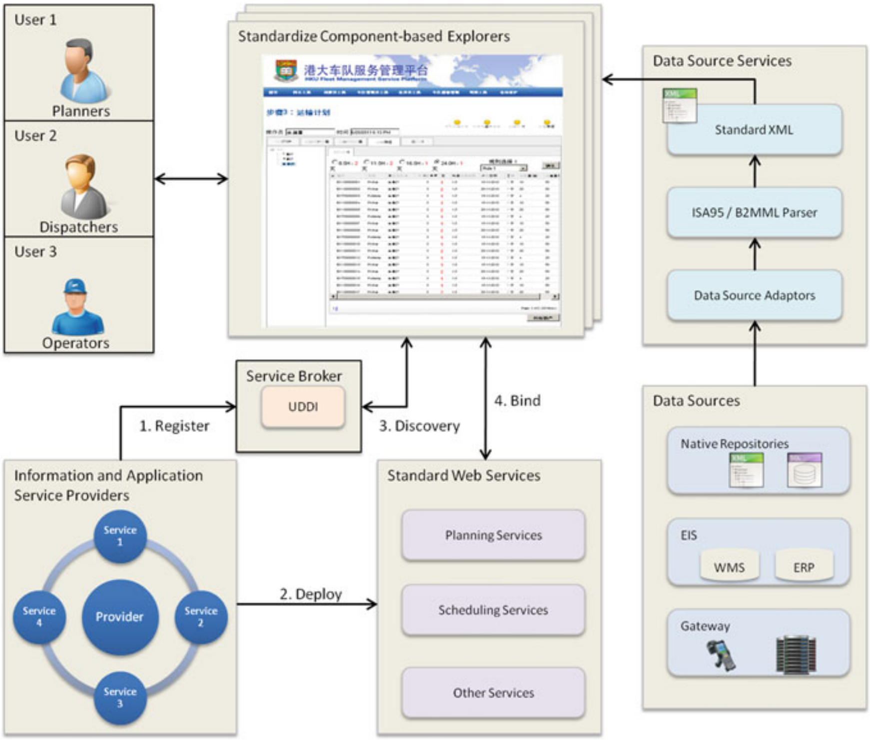


Fig. 3 Service oriented based FAPS system architecture

- Innovative use of Auto-ID and RFID technologies for creating smart operation environment.
- Innovative use of Service-Oriented Architecture (SOA) for integrating different services, both internal and third party, in decision and operation process.
- Reconfigurable web-based user interfaces driven by XML data models.
- Centralized data source service for standardize data model from heterogeneous enterprise information systems (EISs).

FAPS system is a web-based system designed and developed following SOA. The FAPS system architecture as shown in Fig. 3 can be divided into three categories. The first category includes data source services to enable information communication between the explorers and heterogeneous data sources. Industrial standards such as ISA95 for manufacturing will be applied to facilitate information exchange. The second category includes standard web services by which planning and scheduling optimization algorithms are developed and deployed. It also includes services which are provided by third party information and application service providers. The third category includes standard component-based explorers. The user interface is component-based and the development of component is based on standard data models in the domain of manufacturing and logistics. Standard components enhance the reusability of user interface and enable developers to customize solutions without high costs and long development cycles.

A complete SOA process involves four main phases: Register, Deploy, Discovery and Bind as shown at the bottom of Fig. 3. Service developers/providers deploy their web services at the designated server sites and publish the details of these services including location, capability as well as interfacial description. Service consumers search and select appropriate web services from the published database. Values must be specified and provided before services are invoked during a specific application process.

The implementation of SOA is web services technology (Carey 2008). Web services make functional building-blocks accessible over standard Internet protocols independent of platforms and programming languages. The web service architecture involves three categories, namely universal description, discovery and integration (UDDI), web services description language (WSDL) and simple object access protocol (SOAP). UDDI is a platform-independent, XML-based registry for distributed services to list themselves on the internet, while WSDL standard provides a uniform way of describing the abstract interface and protocol bindings of these services. In simple terms, a WSDL describe what a web service can do, where it resides, and how to invoke it. Simple object access protocol (SOAP) is a platform-independent protocol for invoking those distributed web services through exchanging XML-based messages.

Conclusions

It has been argued that supply chain should be integrated and different stages should be considered as a whole in order to achieve the overall optimal performance throughout the entire supply chain processes. Industrial parks gather different industries within a planned area. Various types of industries such as manufacturing, warehousing and transportation mimic a small supply chain. This environment favours the integration of supply chain to achieve global optimization in transportation planning and scheduling. It also favours sharing expensive RFID equipments between different members. This paper presents a service-oriented platform for fleet advanced planning and scheduling which enables essential information sharing and significant improving transportation planning and scheduling in industrial parks. Hierarchical transportation planning and scheduling has been adopted in FAPS system to guide the decision making. The use of the FAPS system leads to improve the fleet management, increase decision and operation efficiency as well as fleet resources utilization.

The FAPS system demonstrates several key contributions. Firstly, RFID devices are used in working environment, and smart gateways are adapted and developed to create smart operation environments where changes and distances are tracked and traced on a real-time basis. Secondly, RFID-enabled real-time information traceability and visibility plays essential roles in coordinating decisions and operations of different parties involved in transportation process. Thirdly, adaptive optimization models and solution algorithms are developed and deployed as web services to solve integrated transportation planning and scheduling problems.

The FAPS system is highly extensible and usable due to the innovative use of cutting-edge technologies for industrial informatics. Firstly, the platform follows a standard service-oriented architecture with key modules implemented as web services. This maximizes system integrity and scalability. Secondly, user interfaces are web-based and developed with web components supported by standard data models. This allows developers easily customize the user interface for different type of users. Finally, a centralized data-source web service is used to manage heterogeneous data sources. The third-party customizability and usability are achieved with minimum efforts without substantial re-coding efforts.

Although R&D efforts have been carried out in close consultation with real-life cases, the FAPS system is yet to be tested in a real-life industrial environment. The library of optimization services must be consolidated for different application scenarios. The work has been done with a good intention is to adopt domain standards for enterprise information management. However, standards do not seem to be comprehensive enough for both manufacturing and logistics processes while the transportation seems to be related to both processes. Therefore, more efforts are necessary to better understand these standards for better exploitation.

References

Baker BM, Ayechew MA (2003) A genetic algorithm for the vehicle routing problem. Comput Oper Res 30(5):787–800

Böse F, Piotrowski J, Scholz-Reiter B (2009) Autonomously controlled storage management in vehicle logistics—applications of RFID and mobile computing systems. Int J RF Technol: Res Appl 1(1):57–76

Carey MJ (2008) SOA What? Computer 41(3):92–94

Chow HKH, Choy KL, Lee WB, Lau KC (2006) Design of a RFID case-based resource management system for warehouse operations. Expert Syst Appl 30(4):561–576

Fleischmann B (2005) Distribution and transport planning supply chain management and advanced planning. In: Stadtler H, Kilger C (eds). Springer, Berlin Heidelberg, pp 229–244

Gebauer J, Schober F (2005) Information system flexibility and the performance of business processes. University of Illinois at Urbana-Champaign, College of Business, Urbana

Gendreau M, Iori M, Laporte G, Martello S (2008) A Tabu search heuristic for the vehicle routing problem with two-dimensional loading constraints. Networks 51(1):4–18

Geng Y, Hengxin Z (2009) Industrial park management in the Chinese environment. J Cleaner Prod 17(14):1289–1294

HEPD/IPPS (1997) Industrial estates: principles and practice. UNIDO Publication, Austria

Meyr H, Rohde J, Wagner M, Wetterauer U (2005) Architecture of selected APS supply chain management and advanced planning. In: Stadtler H, Kilger C (eds). Springer, Berlin Heidelberg, pp 341–353

Montemanni R, Gambardella L, Rizzoli A, Donati A (2005) Ant colony system for a dynamic vehicle routing problem. J Comb Optim 10(4):327–343

Tudor T, Adam E, Bates M (2007) Drivers and limitations for the successful development and functioning of EIPs (eco-industrial parks): a literature review. Ecol Econ 61(2–3):199–207

Yizhong W, Ho OKW, Huang GQ, Da L, Huafang H (2008) Study on RFID-enabled real-time vehicle management system in logistics. In: IEEE international conference on automation and logistic (ICAL) 1–3 September 2008. pp 2234–2238

Towards Agile and Flexible Air Cargo Processes with Localization Based on RFID and Complex Event Processing

David Rüdiger, Moritz Roidl and Michael ten Hompel

Abstract How can one transfer the concept of a real-time enterprise to the air cargo business? What are the challenges from a technological and information technology point of view? In this article the processes of an airfreight hub are analyzed under the aspects of flexibility, quality and transparency, in order to demonstrate the economic potential of a real-time locating system (RTLS) implementation at a cargo hub. Following this, concrete industrial applications are developed by using the idea of an event-driven architecture (EDA) and complex event processing (CEP) methods which enable immediate reactions to critical business transactions. This includes the development of CEP logics for airfreight processes using the CEP language Esper. In addition, first test results are presented by applying data from an RTLS of an industrial partner.

Keywords RFID · Real-time localization (RTLS) · Event-driven architecture (EDA) · Complex event processing (CEP) · Esper · Air cargo logistics · Airfreight business

D. Rüdiger (✉) · M. ten Hompel
Fraunhofer Institute for Material Flow and Logistics, Joseph-von-Fraunhofer-Straße 2–4, 44227 Dortmund, Germany
e-mail: david.ruediger@iml.fraunhofer.de

M. ten Hompel
e-mail: tenhompel@iml.fraunhofer.de

M. Roidl
Chair of Materials Handling and Warehousing, TU Dortmund, Emil-Figge-Str. 73 44227 Dortmund, Germany
e-mail: moritz.roidl@flw.mb.tu-dortmund.de

H.-J. Kreowski et al. (eds.), *Dynamics in Logistics*, Lecture Notes in Logistics,
DOI: 10.1007/978-3-642-35966-8_19,

General Introduction

In air cargo operations transshipment activities are affected by several disturbances which may rapidly occur and to which reactions have to be initiated. For instance a canceled or a delayed flight as well as internal process delays might interfere with the general planning and scheduling of operations. Highly sophisticated hardware, such as RFID and RTLS, and software, such as CEP and EDA enable air cargo operators to automate processes and, as a result, reduce complexity of operations management.

Analysis of the Airfreight Business

This paper contains a brief analysis of the airfreight business with a special focus on air cargo hub operations. As air transportation used to have a sovereign function, even to this day, countries demand or maintain a national air cargo sector. Nonetheless, private ownerships are in the majority, since deregulation activities in key markets. Nowadays, the central industry trade group of airlines International Air Transport Association (IATA) holds about 230 members; a steady growth since their foundation in 1919, August, 25th, starting with six member organizations (Allaz 2005). Still the growth has not reached its peak (IATA 2011). The causes are seen in the on-going trend toward global markets in terms of sourcing and sales and the increase of dynamic, globalized value-creation in terms of production (De Meyer 1998). After all, the general tendency of shorter product life-cycles encourages short-term planning of global supply chains. At present, to be able to meet market demands and for reasons of efficiency, effectiveness and regulatory issues, air cargo carriers (e.g., Lufthansa, United Airlines) offer worldwide transport services in strategic alliances (e.g., Star Alliance) (Delfmann et al. 2005). In this context co-operation, communication and information exchange between supply-chain participants play a decisive role.

Complexity and Scheduling Challenges in Airfreight Hubs

Air transportation offers the advantage of speed and is usually used for long-distance transport. Compared to other means of transport, air cargo transport is the most expensive one. Typically, carried goods are high value goods (e.g., electronic products, chemicals) or goods, which demand a quick transfer, such as perishables or animals. Ensuring rapid processes the air cargo business defines and maintains its competitiveness (Arnold et al. 2008).

Each year, millions of airfreight shipments are transferred at international airfreight hubs. For instance, in 2010 approx. 4.1 million tons of air cargo was shipped via Germany (Destatis 2011). Facing infrastructural aspects, the transfer of shipments usually occurs in hub-and-spoke-systems. Hub-and-spoke systems enable airfreight carriers to organize a nationwide route network among economic aspects. Instead of connecting one station directly with another one, it is possible to reach a high utilization of means of transport when shipments are combined to batches. The disadvantage of a multi-level freight transport is the fact that longer transport routes have to be accepted (Arnold et al. 2008).

The concentration of large numbers of transports/shipments at one hub leads to high complexity in terms of general hub management and process/production planning. In addition, due to given flight schedules of airports, time windows for internal cargo handling are generally low (especially in cargo transit processes). As a result approaches of reducing complexity are of the utmost significance.

The main focus of approaches in complexity reduction should be put on the synchronization of inbound and outbound processes, which will be discussed in the next paragraph. The handling of cargo in these processes, especially concerning the break-down and build-up of Unit Load Device (ULDs—standardized air cargo pallet) is very time-consuming and often regarded as a central bottleneck of airfreight processes. However, the optimization of performance is not as easy as it may seem. Transshipment activities are affected by several disturbances which may rapidly occur and to which reactions have to be initiated (e.g., changing weather conditions).

These days air cargo carriers are competing on a monthly base in maximizing the air cargo key performance indicator FAP (shipments flown-as-planned) (IATA 2009). For this reason, approaches of forming agile and flexible air cargo hub processes meet with positive market response.

Airfreight Hub Processes

Based on a hub-and-spoke system, air transport processes can be divided into export (transfer from landside to airside), transit (airside–airside) and import (airside–landside). Four different functional areas can be distinguished at air cargo stations worldwide: export handling area, outbound handling area, inbound handling area and import handling area. Figure 1 describes the main tasks in the different areas that influence production planning and scheduling (PPS), which builds the base of the analyzed RFID scenario.

Following the idea of assembly line production, shipments are routed from one area to another. Thus, the management of crossing material flows can be seen as a central challenge of hub management. In such a context, the subjects of having detailed process information, exchanging information between different processes and autonomous processing of information play a crucial role.

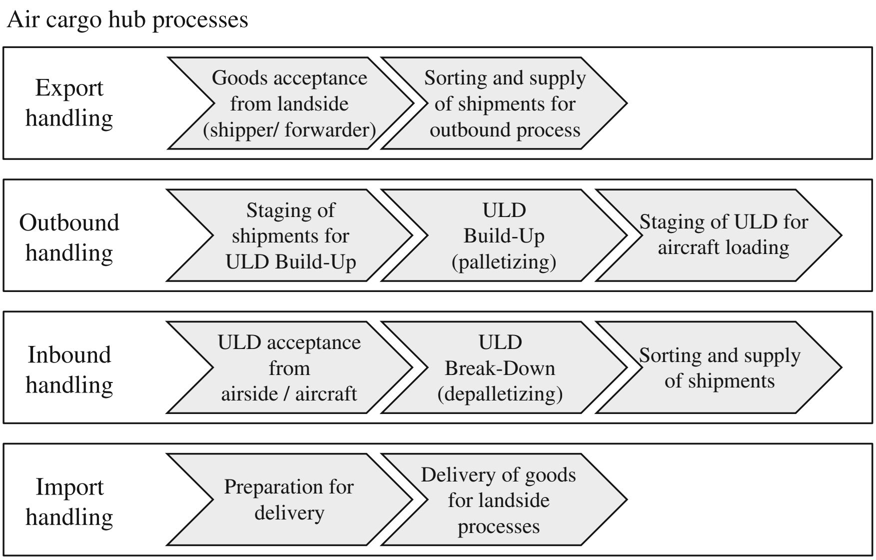


Fig. 1 Functional areas and processes in air cargo hubs

Introduction to Central Aspects of an RFID-Triggered CEP Solution

The following section gives a short introduction to RFID in the sector of consumer goods logistics and to Real-Time Locating System (RTLS) using passive RF communication. Further Complex Event Processing (CEP) is presented in the context of Event-Driven Architecture (EDA) and the idea of a real-time enterprise.

Passive, open loop RFID solutions are developed in line with requirements of the sector of consumer goods and retail. “In open loop environments, tagged objects are usually permanently associated and identified by RFID tags, items are not limited to a predetermined set of partners, and seamless data exchange, enabled by open standards among participants, is required” (Al-Kassab and Rumsch 2008). EPCglobal defines itself as the required standard. Being mapped with already existing standards like GTIN, SSCC, and GRAI the standard assures a high level of flexibility for user-specific functionality (GS1 2011).

Real-Time Localization Based on RFID

Several approaches for the localization of active and passive RFID tags have been proposed in the past (Weichert et al. 2010). Nowadays, RTLS solutions based on active RFID components (e.g., Identic Solutions, Ubisense) are widely used in industrial applications. In contrast, passive RTLS solutions represent a niche market. Recently, two U.S. American solutions attract specialists' interest: Mojix Inc. and RF Controls. In 2004 Mojix released their innovative RTLS solution based on passive RF communication (UHF), with reading capabilities "beyond previous RFID reader offerings" (Mojix 2008). As stated in (Wessel 2009 and Roberti 2008) Mojix provides a practical and market-feasible solution.

Exchanging data passively is highly relevant regarding the restrictions and requirements of modern airfreight transport business. International security standards prohibit the unaccompanied transport of active sensor elements during flight operations. The reason of the ban is the danger of influencing aircraft instruments. Because of this, airline passengers are generally informed about switching-off their cellular phones as the aircraft is ready for departure.

The Mojix system offers a real-time track and trace solution for objects being tagged with passive (UHF) RFID transponders. Figure 2 illustrates the basic functionality of the system. The operating mode of the Mojix RFID RTLS technology benefits from the capability of a central receiver to extract extreme low-level backscattering data signals (approx. 120 dBm). A standard set-up requires a single central receiver and several (up to 512) excitation points, which are installed systematically on the ceiling of a production facility. Following the no-line-of-sight principle the centralized reading unit is able to read tags in a distance up to 200 m, although the strength of the EM field does not exceed 2 W ERP. As stated

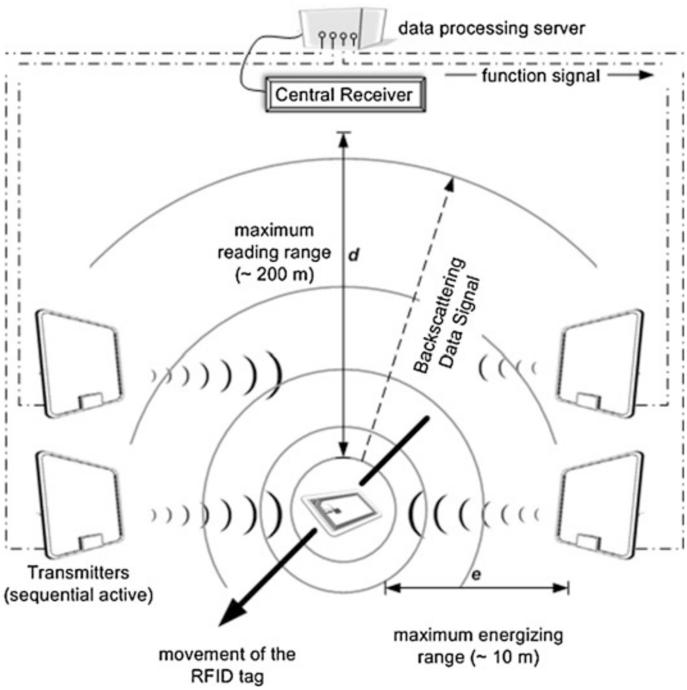


Fig. 2 Real-time locating system of Mojix Inc

in the Mojix patent (US 7873326) the tracking algorithm uses a combination of "Received Signal Strength" (RSS) and ''Time Difference of Arrival'' (TDOA). In addition, only one transmitter is active simultaneously, which allows combining several signals in form of triangulation. The standard output signal of Mojix comprises EPC, x-position, y-position, z-position and a system time stamp.

Event-Driven Architecture and Complex Event Processing

Improving business agility in terms of enterprise software architecture specifies the goal of an event-driven architecture (EDA). Within an EDA events are transmitted among loosely coupled software components and services (Mani Chandy and Schulte 2008). In doing so, EDA does not compete with the concept of service-oriented architecture (SOA), but rather extends its functionality with the ability to process events in real-time (Bruns and Dunkel 2010).

Complex event processing (CEP) and EDA are closely connected. CEP—extensively researched by Luckham (2002)—can be defined as a "software infrastructure that can detect patterns of events (and expected events that didn't occur) by filtering, correlating, contextualizing, and analyzing data captured from disparate live data sources to respond as defined using the platform's development tools" (Gualtieri and Rymer 2009). In opposition to traditional database management, which follows the concept of "store-and-analyze", a CEP engine analyzes permanently pre-defined rules and queries to data streams. As a consequence a high level of reactivity is provided.

The CEP market landscape contains large vendors (Tibco, IBM, Progress Software) as well as specialists (Agent Logic, EsperTech) (Gualtieri and Rymer 2009). This article uses the open-source CEP software Esper, which is designed for high-volume correlation of millions of events and which can fully be embedded in existing Java based architectures (Esper 2011). The programming language of Esper (a tailored Event Processing Language) shows a certain similarity to standard SQL statements. In contrast, it allows expressing rich event conditions respectively event correlations and spanning time windows for event pattern matching (Esper 2011). EDA in combination with CEP can provide a high level of reactivity and therefore qualifies companies for the concept of real-time enterprise and its central idea of permanent information availability about business procedures (Taylor et al. 2009).

Improving Agility and Flexibility of Air Cargo Processes

Accurate status information on process developments is required for efficient and agile process management. These days, information is rather generated selectively within air cargo hubs. It is standard practice to use the Barcode technology, which impedes applying a full-scale information system on shipment-level. In contrast, the combination of a real-time localization system and a framework of continuous data processing allow to increase the accuracy of status information and to ensure a high level of compliance between planned processes and time schedules. The reasons will be explained hereinafter.

Some research projects are already investigating the organizational and technical possibilities of increasing the accuracy of status information in air cargo processes: Integrated Air Cargo Hub (IACH 2011), DyCoNet (Dyconet 2011) and IATA e-Freight initiative (eFreight 2011). Nonetheless, applying agile and smart IT architectures, such as CEP and EDA, to air cargo processes form an entirely new approach.

Advantages of Real-Time Monitoring in an Air Cargo Context

In air cargo context, the total production quality of a hub is strongly dependent on the time-critical transit process. By providing localization information in real-time, process irregularities can be detected automatically and already in the phase of initiation. Consequently, reactions to problems can be launched directly without wasting time and letting a mistake develop. Not only managerial staff can take advantage of instantaneous status control but also the functioning of the RTLS systems is monitored permanently. In the case when no location information is provided after a certain period of time, one could infer a broken or hidden RFID tag.

Permanent real-time status control (on shipment-level) allows quick decisions in general hub management. The following example focuses on real-time control of shipment movements through an air cargo hub. Based upon an underlying time schedule the movement of a shipment is permanently opposed to a given time schedule (Fig. 3). Advantages of the new approach are as follows: real-time reactions to process delays in terms of resource prioritization and near-time total process re-organization for enhancing system performance in general.

Complex Event Processing in an Air Cargo Context

Even small-sized RFID installations generate a high volume of rapidly changing data for central IT systems. Facing the large areas to be covered by an RTLS in an air cargo hub and the large number of RFID readers, it is not possible to connect the readers directly to central IT systems. Figure 4 illustrates the required multi-layer IT architecture. At integration layer gathered localization events are filtered, contextualized and correlated to meaningful events (complex events). Those events, which are the result of an automated event pattern matching, activate finally subscribed software components at application level.

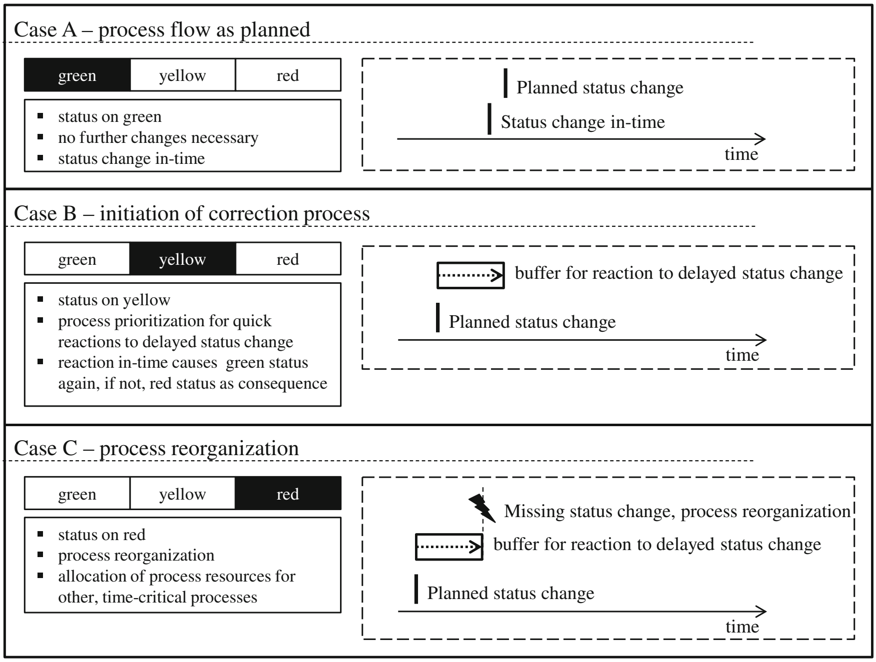


Fig. 3 Instantaneous status control of processes

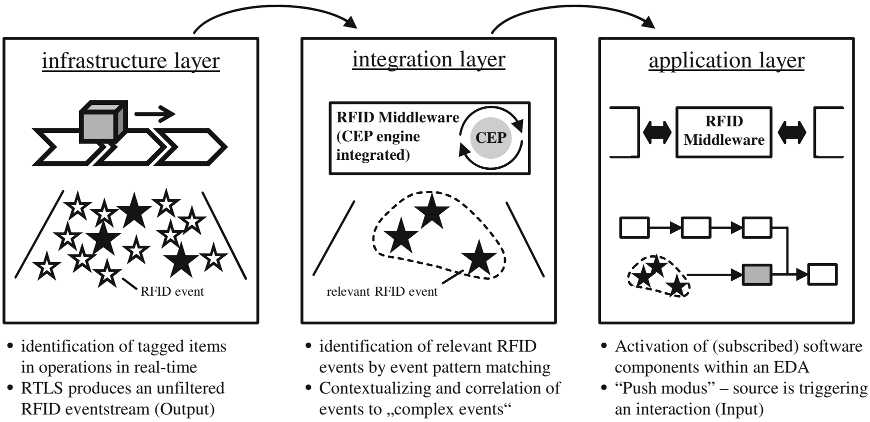


Fig. 4 System architecture

Implementing a solution for the illustrated air cargo use case in Esper requires in a first step the combination of three event streams: movement event stream (output of the RTLS), plan data event stream (operational temporal information) and topology event stream (spatial information on system topology). These

Plandata Events	Movement Events*	Topology Events
• unique ID of shipment (EPC) - *in Esper: EPC* • unique ID der area / process - *in Esper: locationId* • planned start time of process - *in Esper: requiredStartTime* • planned end time of process - *in Esper: requiredExitTime* • timestamp - *in Esper: timestamp*	• unique ID of shipment (EPC) - *in Esper: EPC* • x-Position - *in Esper: xPosition* • timestamp - *in Esper: timestamp* * Output RFID/RTLS	• unique ID of area / process - *in Esper: locationId* • start of area in x-direction - *in Esper: xStart* • End of area in x-direction - *in Esper: xEnd* • timestamp - *in Esper: timestamp*

Fig. 5 Types of event streams in the given example

different sources of information, which establish a basis for event pattern matching in terms of process-monitoring, are illustrated in the following Fig. 5.

Defining all data information as dynamic event streams characterizes the innovative approach of Complex Event Processing. Even topology information is considered dynamically and variable in time. Consequently dynamics in logistics can be mapped with maximum flexibility.

The relating Esper statements for the given example can be grouped into four parts: First the combination of location information with spatial information (Fig. 6), second the comparison of actual process progress and planed process progress (Fig. 7), third the extraction of all those events that caused a warning

```
INSERT INTO  EventBuffer1
SELECT       mov.EPC as EPC,mov.timestamp as timestamp, top.locationId as locationId
FROM         MovementEvent.win:keepall() as mov, TopologyEvent.win:keepall() as top
WHERE        mov.xPosition >= top.xStart AND mov.xPosition < top.xEnd
ORDER BY     mov.timestamp
```

Fig. 6 Combining location information with spatial information in Esper

```
INSERT INTO  EventBuffer2
SELECT       Buff1.EPC as EPC, plan.locationId as locationId, plan.timestamp as timestamp,
             Buff1.timestamp as timestamp2, Buff1.locationId as locationId2,
             (case when (Buff1.locationId – plan.locationId) < 0 then "WARNING"
             else "OK" end) as status,
FROM         EventBuffer1.win:length(1) as Buff1, PlandataEvent.win:keepall() as plan
WHERE        Buff1.timestamp <= plan.timestamp
ORDER BY     Buff1.timestamp
```

Fig. 7 Comparison of actual and planed process progress

```
INSERT INTO  EventBuffer3
SELECT       Buff1.EPC as EPC, Buff1.timestamp as timestamp,
             Buff1.locationId as locationId
FROM         EventBuffer1.win:keepall() as Buff1
WHERE        prev(Buff1.locationId) != Buff1.locationId
ORDER BY     Buff1.timestamp
```

Fig. 8 Extraction of critical events

```
SELECT       Buff2.locationId, Buff2.timestamp, Buff3.locationId, Buff3.locationId
FROM         EventBuffer2.win:length(1) as Buff2, EventBuffer3.win:length(1) as Buff3
WHERE        Buff2.locationId = Buff3.locationId AND Buff3.timestamp > Buff2.timestamp
ORDER BY     Buff3.timestamp
```

Fig. 9 Monitoring of reactions

status (Fig. 8), and fourth the extraction of all those events that describe a reaction to a warning message (Fig. 9).

Based on these Esper statements a visual output with concrete information can be displayed to an employee. This may include information about the object that causes a warning status (e.g., last recorded object position, planned process progress, remaining time for process correction) or in a more sophisticated approach concrete action recommendations how to solve the identified problem. This may be realized in combining the CEP engine with tools of production planning and scheduling.

First Test Results from an Air Cargo Facility

For testing we used a Mojix RTLS installation within the facilities of a global-operating air cargo carrier. Following the prerequisites of the given example, shipments were equipped with passive UHF RFID tags (UPM Shortdipole Monza 3). The main focus was laid on the monitoring and analysis of the build-up process of air cargo pallets in terms of time and quality.

The movement of shipments was recognized by the Mojix RTLS and location data was provided to the CEP engine (Fig. 10). The developed Esper statements made it possible to derive status changes, such as “shipment on staging area” and “shipment on ULD”, automatically and in real-time. Altogether, the test was terminated successfully. Nevertheless, due to some imprecise location information—partly an inaccuracy of 1 meter was observed—discrepancies of actual occurring processes and the interpretation of the CEP engine were detected. The uncontrolled industrial test environment prevents from determining the exact sources of errors.

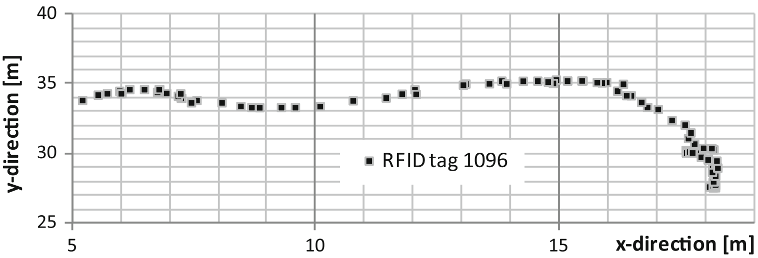


Fig. 10 Example localization data of Mojix RTLS. The system was configured to record tag positions each second. This figure shows the slow movement of a tag that was attached to a ULD

Summarizing Conclusion

The development and transfer of air cargo's specific rules into logics for efficient processing of RFID localization events gives evidence that complex event processing (CEP) is well-suited for monitoring and controlling the processes of air cargo hub processes. Moreover, seamless process transparency, given by an appropriate RTLS, facilitates large-scale process improvements, such as the elimination of manual scan processes and immediate quality checks. Nonetheless, the fundamental challenge is to implement a comprehensive RTLS that has extensive coverage and reliable localization results. In case of localization systems a general inaccuracy has to be accepted for physical reasons. In order to identify an appropriate RTLS product, companies need to define a maximum inaccuracy. This raises one main question: Can inaccuracies be handled with adequate calibrations of the logistic systems? Assuming that, the combination of RTLS and CEP enables implementing a smart and agile information management system, which is in line with the concept of real-time enterprise and its central idea of permanent information about on-going business. Thus, it sharpens a company's ability to react to internal/external irregularities and to communicate differentiated with costumers (e.g., detailed and differentiated forecasts on delivery).

As deducted in the context of air cargo operations, logistic processes and networks face continuously changing conditions. Combining market-feasible track and trace solutions with new approaches of computer science allows it to cope rapidly and flexibly with changing conditions in logistic systems.

References

Al-Kassab J, Rumsch W (2008) Challenges for RFID cross-Industry standardization in the light of diverging industry requirements. Syst J IEEE 2:170–177

Allaz C (2005) History of air cargo and airmail from the 18th century. Christopher Foyle Publishing Limited, London

Arnold D et al (2008) Handbuch logistik, 3rd edn. Springer, Heidelberg

Bruns R, Dunkel J (2010) Event-driven architecture: softwarearchitektur für ereignisgesteuerte geschäftsprozesse. Springer, Heidelberg

De Meyer A (1998) Manufacturing operations in Europe: where do we go next? Eur Manage J 16(3):262–271

Delfmann W et al (2005) Strategic management in the aviation industry. Ashgate Publishing Limited, Hampshire

Destatis (2011) Luftfracht 2010: erstmals mehr als vier millionen tonnen befördert. http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Presse/pm/2011/02/PD11__063__464,templateId=renderPrint.psml. Accessed 31 Oct 2011

Dyconet (2011) Dynamisch, autonomes, energieautarkes container netzwerk in der Luftfrachtindustrie. http://www.dyconet.de/. Accessed 31 Oct 2011

Esper (2011) Event processing with Esper and NEsper. http://esper.codehaus.org/. Accessed 31 Oct 2011

GS1 (2011) GS1 system landscape. http://www.gs1.org/docs/gsmp/architecture/GS1_System_Landscape.pdf. Accessed 31 Oct 2011

Gualtieri M, Rymer J (2009) The Forrester wave^{TM}: complex event processing (CEP) platforms, Q3 2009. http://www.waterstechnology.com/digital_assets/1261/progress_apama_forrester_report.pdf. Accessed 31 Oct 2011

IATA (2009) Quality and industry standard for the global air cargo industry. http://www.iata.org/whatwedo/Documents/cargo-2000/C2K-Standard-Presentation-V.2.24-2009-08-28.pdf. Accessed 31 Oct 2011

IATA (2011) Vision 2050 report. http://www.iata.org/about/Documents/vision-2050.pdf. Accessed 31 Oct 2011

eFreight IATA initiative (2011).http://www.iata.org/whatwedo/cargo/efreight/Pages/index.aspx. Accessed 31 Oct 2011

IACH—Integrated Air Cargo Hub (2011). http://www.iml.fraunhofer.de/en/fields_of_activity/project_center_aviation_logistics/projekte/effizienz_luster.html. Accessed 31 Oct 2011

Luckham D (2002) The power of events: an introduction to complex event processing in distributed enterprise systems. Addison-Wesley, Boston

Mani Chandy K, Roy Schulte R (2008) Event processing: designing IT systems for agile companies. McGraw-Hill, New York

Mojix (2008) Brochure on Mojix star system. http://www.mojix.com/pdf/Mojix_STAR_System.pdf. Accessed 31 Oct 2011

Roberti M (2008) Mojix takes passive UHF RFID to a new level. http://www.rfidjournal.com/article/view/4019. Accessed 31 Oct 2011

Taylor H et al (2009) Event-driven architecture: how SOA enables the real-time enterprise. Addison-Wesley, Boston

Weichert F et al. (2010) Marker-based tracking in support of RFID controlled material flow systems. Logistics Res 2:13–21 (Springer, Heidelberg)

Wessel R (2009) Rewe deploying long-range real-time location RFID system. http://www.rfidjournal.com/article/view/5187. Accessed 31 Oct 2011

Smart Agri-Food Logistics: Requirements for the Future Internet

C. N. Verdouw, H. Sundmaeker, F. Meyer, J. Wolfert and J. Verhoosel

Abstract The food and agribusiness is an important sector in European logistics with a share in the EU road transport of about 20 %. One of the main logistic challenges in this sector is to deal with the high dynamics and uncertainty in supply and demand. This paper defines requirements on Future Internet (FI) technologies that have to be met to accomplish the specific challenges of agri-food logistics. It identifies a set of generic technical enablers as input for the realisation of a FI core platform. This technology foundation is to be developed and tested in a Future Internet public–private partnership (FI-PPP) environment of over 150 organisations.

Introduction

Agri-food is an important sector in European logistics. The agri-food logistics covers 19 % of the transports within the EU and 25 % of the international EU transports (Verweij and van der Moolen 2009). One of the main logistic challenges for the agri-food industry is to deal with high dynamics and uncertainty in supply and demand (Verdouw et al. 2010a; Trienekens et al. 2011). There is great uncertainty regarding fresh product quality as well as available volumes in time on

C. N. Verdouw (✉) · J. Wolfert
LEI Wageningen UR, The Hague, The Netherlands
e-mail: Cor.Verdouw@wur.nl

H. Sundmaeker · F. Meyer
ATB Institut für Angewandte Systemtechnik Bremen GmbH, Bremen, Germany

J. Verhoosel
TNO, Delft, The Netherlands

H.-J. Kreowski et al. (eds.), *Dynamics in Logistics*, Lecture Notes in Logistics,
DOI: 10.1007/978-3-642-35966-8_20,

a specific place. The sector is characterised by last-minute changes and rush-orders. As a consequence, the required prediction and planning concept and accompanying logistics system needs to be very flexible, enabling last minutes changes and reallocations, but also provide a robust planning (Verdouw et al. 2010a; Trienekens et al. 2011; van Wezel et al. 2006; Verdouw et al. 2010b).

In this paper it will be argued that current internet limitations are important bottlenecks to accomplish the specific challenges of agri-food logistics. Important limitations that are addressed by the EC's FI Architecture Group include (1) a lack of data integrity, reliability, provenance and trust; (2) a lack of data integration and federated storage solutions; (3) lack of flexibility and adaptive control; and (4) segmentation of data and control (Zahariadis and Papadimitriou 2011).

The intended development of Future Internet (FI) technologies is promising to overcome these limitations. The FI is a general term that labels the emergence of a new era in the evolution of the Internet. It combines several trends in internet development into an integrated approach. These trends include (FI-WARE 2011b):

- the on-going industrialization of IT in the form of cloud computing and open service delivery platforms;
- new wireless networking technologies and the deployment of fibre that are paving the way for new (real-time) applications;
- the breakthrough of the Internet of Things, with the vision of ubiquitously connecting intelligent devices and sensors.

The objective of the research presented in this paper is to define requirements on Future Internet (FI) technologies of the food and agribusiness domain. Therefore, after an introduction of the method, the remainder of this paper first describes the logistic challenges in the agri-food sector and the envisaged FI-based solution. Next, an analysis of user requirements for FI logistics is presented.

Method

Future Internet PPP and the Smart Agri-Food Project

The research presented in this paper is carried out as a part of the Smart Agri-Food project ("Future Internet for Safe and Healthy Food from Farm to Fork") within the ICT theme of the FP7 (www.smartagrifood.eu). Smart Agri-Food is one of eight currently being funded users case studies of the "Future Internet Public Private Partnership" programme (FI-PPP www.fi-ppp.eu).

The FI-PPP programme aims to advance Europe's competitiveness in FI technologies and systems and to support the emergence of FI-enhanced applications of public and social relevance. This will be achieved by (1) development of a smart, Europe-wide internet-powered infrastructure, (2) increasing the suitability for processing larger amounts of data, (3) ensuring a more secure and reliable use of Internet, allowing real-time data processing for real-time applications.

In technical terms the FI-PPP targets a versatile and open network and service platform, supported by reusable, standardised and commonly shared technology enablers serving a multiplicity of use cases in different sectors (European Commission 2011).

The FI-PPP programme currently (2012) includes over 150 partners. It follows a user-driven approach, in which 8 use case projects define the requirements for the realisation of a FI core platform. Subsequently, this technical foundation of the FI is to be developed and tested in its public–private partnership environment.

The Smart Agri-Food project addresses the food and agribusiness as a use case for the FI. It aims to boost the application and use of FI ICTs in the agri-food sector by identifying and describing/developing:

- technical, functional and non-functional FI-specifications for experimentation in the food and agribusiness;
- smart agri-food-specific capabilities and conceptual prototypes, demonstrating critical technological solutions including the feasibility to further develop them in large scale experimentation and validation;
- existing experimentation structures and start user community building, resulting in an implementation plan for the next phase of the FI-PPP programme.

Methodology

The Smart Agri-Food project is focussing on three sub systems of the sector: Smart Farming, Smart Agri-Food Logistics and Smart Food Awareness (consumer). This paper presents the results of the definition for experimentation of the second sub system: Smart Agri-Food Logistics. Therefore, seven specific application scenarios have been identified and a template for the investigation of the application was developed. Experts from academia, non-profit research institutes, ICT industry and food industry were analysing the scenarios to identify ICT related aspects and to derive specific user related functional requirements. The requirements were clustered into functional blocks based on (1) a logical association of the different functional requirements that address similar tasks, and (2) the so-called *generic enablers* that shall be provided by the FI Core Platform (FI-WARE 2011a, b).

Smart Logistics in the Agri-Food Sector

Specific Logistic Challenges in Agri-Food Supply Chains

The sector-specific characteristics of the food and agribusiness heavily impact logistics in this sector (Verdouw et al. 2010a; Trienekens et al. 2011; van Wezel et al. 2006; Verdouw et al. 2010b):

- High supply uncertainty due to natural production: unpredictable variations in quality and quantity of supply, which demands for flexibility in logistic processes and planning and early warning and pro-active control mechanisms;
- High perishability of fresh food products, which demands for temperature-conditioned transportation and storage (cold chains) and very short order-to-delivery lead-times;
- Seasonable growing, which demands for global sourcing to ensure year-round availability;
- High demands on food safety, quality and (environmental) legislation, which demands for the ability to trace production information of products in transit;
- High flow complexities, due to a combination of continuous and discrete product flows, diverging and converging processes and by-products; this demands for advanced tracking and tracing and logistic planning capabilities;
- Important role of import/export, including additional phytosanitary and veterinary inspections;
- Complex network structure where small and medium enterprises trade with huge multinationals in the input and retail sector; this demands for proper collection and regional orchestration in logistic mainports and proper allocation mechanisms to connect aggregated demand with fragmented supply.

Due to these characteristics, the application of existing ICT solutions is not always obvious and less straightforward than it might be in other industries. For example: there is a high need for interoperability but this should be accompanied with flexibility to deal with the high dynamics in agri-food logistics. Furthermore, the need for sector-specific ICT solutions is high, but at the same time the possibilities for investments are low due to the large number of SMEs. Consequently, the current state of the art of ICT in the agri-food logistics is characterized by large amounts of available data, but there is a poor level of integration and the support for intelligent use of these data is insufficient. The complexity of current solutions is too high and jeopardizes the development and operation of affordable solutions. As a result, the adoption of internet for basic information services is high, but the use for more advanced functionalities is limited.

These complexities contrast with the high and constantly increasing need for better information, control and transparency in the agri-food sector. Consequently, there is a mismatch between the state of information technology in agri-food and the high and increasing need for intelligent solutions that combine interoperability with flexibility and that are both sector-specific and suitable for SMEs. The next section introduces a vision on how to deal with these challenges.

Smart Agri-Food Logistics Enabled by the Future Internet

The envisaged solutions focus on the enhancement of new types of efficient and responsive logistics networks with flexible chain-encompassing tracking and tracing

Fig. 1 Smart agri-food logistics

systems and decision support based on that information. These systems effectively virtualise the logistics flows from farm to fork, support a timely and error-free exchange of logistics information and provide functionality for intelligent analysis and reporting of exchanged data to enable early warning and advanced forecasting. Three critical features of these systems are distinguished: real-time virtualization, logistics connectivity and logistics intelligence (see Fig. 1).

As argued in the previous section, current solutions do not sufficiently provide these features. The FI offers a realistic scenario to solve this because it aims at a provision of an integrated architecture that overcomes the basic limitations of current information systems for agri-food logistics.

The envisaged Smart Agri-Food Logistics solutions will build on the Core Platform that is being developed by the FI-PPP programme (FI-WARE 2011b; European Commission 2011; FI-WARE 2011a). This platform aims at the provision of an innovative infrastructure for cost-effective creation and delivery of versatile digital services, providing high Quality of Service and security guarantees. The basic underlying approach is that the Core Platform will offer reusable and commonly shared capabilities and functionalities (Generic Enablers) which can be flexibly customized, used and combined for many different Usage Areas (pick, plug and play). The FI Core Platform will contain Generic Enablers for Cloud Hosting, Data/Context Management, Applications/Services Ecosystem and Delivery Framework, Internet of Things (IoT) Services Enablement, Interface to Networks and Devices (I2ND) and Security (FI-WARE 2011a, b).

Products implementing these Generic Enablers can be picked and plugged together with complementary products (Specific Enablers) in order to build domain-specific instances (FI-WARE 2011b). Smart Agri-Food Logistics is intended to be one of those instances (as part of the Smart Agri-Food project). The next section elaborates on the specific requirements that should be met for that purpose.

Requirements on the Future Internet

Basic Future Internet Challenges in Agri-Food Logistics

Smart Agri-Food Logistics imposes specific demands on the development of FI technologies. The identified basic challenges are :

- *Internet of Things (IoT)*: Various types of sensors (e.g., temperature loggers) are used in the agri-food logistics sector. These applications are mainly limited in scope and focus on specific subparts of the logistics chain. There are communication breaks, especially when transport units or ownership is changed. Communication aspects of the IoT are already fairly well covered by existing technologies, such as networked RFID, wireless sensor networks and near-field communication. A FI is expected to enable a capturing at an integrated global sensor information level in which combinations of sensor information can be used to constantly monitor the product status and enable the identification of exceptions or deviations from expected conditions from a chain perspective.
- *Telematics Systems*: To promptly react to changes of ambient parameters (e.g., temperature, light, ethylene concentration, etc.), the sensors are already applied in some parts of agri-food logistics. When merging this with further information about location, speed and context a new dimension of forecasting and reactions will be possible. However, dynamic change of interrelationships of food chain partners is imposing a key challenge for integrating different systems and communicating information. There is a need for an advanced decentralised management of authentication, authorisation and revocation of access rights for enabling advanced governance models while assuring privacy of data.
- *Tracking and Tracing*: Well-known services are in place, but often limited to an internal level of traceability or to an external approach by service providers. If we increase the level of traceability and focus on the whole supply-chain, the system needs to be able to trace across and through the companies in the chain. However, there are some holistic approaches available but generally need to apply a central entity that is compiling all the information. At the same time the food chain actors are reluctant to provide all data to such a central entity.
- *Autonomous Systems*: Most available logistics systems implement a strict centralized approach and manage logistic processes in a top-down way. Opposite to centralised systems, decentralised systems are trying to translate the current predominant hierarchy approach into a heterarchy model. To achieve this, each entity inside the logistics processes is able to process information, take decisions autonomously and communicate with other entities. Similar to the IoT topic, these capabilities implicate the existence of a virtual identity representing the physical one, but encapsulating the intelligence and ensuring the integrity of content and certifying "mobile software components" to be used on different systems operated by the actors in the food chain.

- *Business Intelligence*: To support strategic and functional decisions, BI systems analyse company-, competitors- and market-relevant electronic accessible data and process models by acquisition, processing and dissemination. This domain can be considered as a beneficiary of the technologies discussed before. Hence, as soon as the exploitation of the other ICTs will advance, tremendous improvements of BI seem possible.

In addition to these technological challenges, it became also evident that the realisation of a FI cannot be accomplished purely on a technology-driven basis. Organisational conditions are crucial success factors, in particular: a proper business case for all involved participants, agreement on the distribution of benefits, trust in the privacy of data and a business process orientation.

Specific Application Scenarios

Specific application scenarios are elaborated as kind of test beds to prepare a larger scale experimentation with respect to FI potentials for the agri-food logistics. Those application scenarios are defined based on the type of logistics information. Distinction is made between identification, static and dynamic state attributes, including condition parameters, product specifications and compliance information. Based on this classification the following Table 1 is identifying the product categories and user groups that are mainly involved.

Expert-Based Definition of User Requirements

The application scenarios are used to define in total 46 FI requirements for smart agri-food logistics. The requirements are grouped into 14 functional blocks. Some 30 % of the requirements are addressing blocks related to data analysis, collection and management, 20 % are addressing the configuration and communication functionalities, while some 15 % are addressing service based architecture and loose coupling principles. Also requirements related to e.g., the IoT, interoperability and immediate user notification are identified.

Subsequently, the requirements are analysed from a more technical perspective and are additionally grouped with respect to envisaged enabling technologies. These functionalities are associating agri-food related aspects, but are serving as trigger to identify and elaborate the envisaged generic enablers that shall be provided by the FI core platform:

- *Peer to peer Services*: Enabling to map real-life communication inside a supply-chain it should be possible to easily create, join and leave P2P networks. This includes centralised, hybrid and pure P2P networks. Additionally problems like routing of messages through the chain, not requiring a central authority that

Table 1 Main involved users of the application scenarios

Application scenario	Product category	Main involved user groups
1. Intelligent supply chain event management systems for the future food supply chain	Fruit (mango)	Grower, forwarder, packer/repacker
2. Exception notification based on fruits/vegetables chain	Fruit and vegetables	Farmer, laboratory, trader, transporter, distribution centre, retailer, consumer, service provider
3. Real-time and trusted information regarding product specifications and compliance	Meat	Primary producer, slaughtering/deboning, processing, transporter, storing, wholesale, retail
4. Legal compliance and quality control	Vegetables (tomato)	Producer (or product owner), transporter, packer, wholesaler/trader, certification body, consumer
5. Quality controlled logistics in the flower chain	Flowers	Grower, auction, trader, retailer
6. Intelligent retail store replenishment of fresh products	Fresh produce (fruits/vegetables)	Retailer (warehouse), retail outlet (supermarkets, hypermarkets, etc.)
7. RFID on pallets from warehouse to retail store	Fresh products (such as meat)	Retailer (warehouse), retail outlet (supermarkets, hypermarkets, etc.)

tracks the latest IP should be addressed, while taking care for routing, prioritisation, message handling and storage.

- *Decoupled/Asynchronous transport*: To allow working in rural areas and to comply to insufficient network coverage, the overall system should compensate disconnects from the internet on mobile devices by allowing asynchronous communication, enabling to send data without waiting for immediate response.
- *Mobile Services/Agents*: As indicated before, rural areas not necessarily allow a permanent connection to the internet/cloud. A FI shall enable a transparent user support. This could possibly be supported by migrating services/agents to user centric devices, since assuring both quality of services and a real time interaction from a user point of view.
- *Online profiles*: Applications need to fetch information about a given object, like user or machinery profiles. Beyond central or device individual storage, a FI service cloud could also enable the decentralised and distributed storage for a more efficient exploitation of scarce technical resources.
- *Updateable profile for objects*: Enabling the timely update and/or synchronisation of especially logistics information (e.g., data for scheduling, notification) that is stored decentralised and distributed by diverse objects in the food chain.
- *Entity authentication/authorisation*: Mechanism for all types of entities (e.g., actors, objects, etc.) shall be supported, also enabling a decentralised security approach for issuing certificates without a central root. This shall also enable the collaboration in local networks (e.g., devices, machines or ICT supported actors in the field) that are decoupled from the Internet.

- *Automatically add/revoke access rights*: Automatically adding and/or revoking access rights, based on the changes in the physical as well as in the virtual world. This is especially relevant for the physical product flow and related change of e.g., ownership and location along the agri-food chain.
- *Identify aggregations of objects*: Agri-food logistic objects must be unique identifiable on different levels of aggregation (e.g., cases, pallets, containers). Services for mapping or integration of different unique identification schemes (e.g., GTIN, SSCC, GLN, GRAI) need to be enabled.
- *Virtual Identity*: Supporting the generation of virtual identities, while aggregating the identities for objects that are grouped in the physical world, avoiding the need to use active digital objects for each virtual entity (e.g., batch, location, owner related), but also enabling the generation of individual virtual identities.
- *Decentralised Trust*: Beyond technically trusted relationships, agri-food logistics security systems should allow for human related trusted interventions. This is specifically related to the integration of business process related certification schemes (e.g. Orgainvent). The logistic objects should facilitate the decentralised generation of such human-trusted relationships.
- *Mapping of Interfaces*: The Internet assures the standardised information exchange based on e.g., http, ftp. There are also divers data exchange standards. However, different legacy systems are offering semantically similar but technical incompatible interfaces. At the same time, they need to import and export information from/to a product related digital/virtual object, while the interfaces of the different systems remain quite stable. A FI shall support the mapping of those technically different interfaces that can be executed by an experienced user, generally familiar with the semantics.

Conclusions and Discussion

This paper has argued that the FI can be an important enabler of the dynamic logistic systems of the future, as required in the agri-food sector, if it provides specific capabilities. The paper has identified these capabilities. Its objective was to define the requirements on FI technologies that have to be met to accomplish the specific challenges of agri-food logistics.

The paper has first addressed the basic demands for FI logistics concerning Internet of Things (IoT), Telematics Systems, Tracking and Tracing, Autonomous Systems and Business Intelligence. Next, 7 specific application scenarios are used to define 46 specific requirements, grouped into 14 functional blocks. Last, a set of the required technical enablers is identified, i.e., peer to peer services; decoupled/asynchronous transport; mobile services/agents; online profiles; updateable profile for objects; entity authentication/authorisation; automatically add/revoke access rights; identify aggregations of objects; virtual identity; decentralised trust; mapping of interfaces. These generic enablers are provided to the FI-PPP initiative as input for the realisation of a FI core platform. This technology foundation of the FI is to be developed and tested in a public–private partnership environment of over 150 organisations.

The main theoretical contribution is that the Smart Agri-Logistics approach utilizes a generic and standardized internet platform to instantiate specific solutions for logistics information systems in the agri-food sector. As a result, it overcomes current bottlenecks and enables the development and operation of affordable solutions that are independent from geographic locations and independent from specific implementation choices. This potentially will boost the application of intelligent information systems for logistics management in agri-food supply chains.

The foremost remaining challenge from a scientific perspective is how to configure dedicated logistic information systems, that provide the right functionality in the specific context of agri-food supply chains, based on a generic internet platform. A complicating factor behind this challenge is that supply chains in the food and agribusiness are characterised by a high variety and variability of process configurations. As a consequence, logistic information systems should enable to take part in different configurations concurrently and to switch quickly to new or adjusted configurations.

The ICT related functionalities, as presented in this paper, will be further considered when specifying the envisaged conceptual prototypes that shall serve for the verification of a large scale implementation. They need to be further coordinated with the envisaged FI core platform in an interactive approach to clearly separate the generic enablers from the envisaged food domain specific enablers.

Acknowledgments The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007–2013) under grant agreement n 285 326. The authors greatly acknowledge the involved individuals and companies for their support. In particular our gratitude goes to Michael van Bekkum (TNO), Adrie Beulens (Wageningen University), Yann Cassing (SGS), Marta Fontseré (Bon Preu), Elena Mansilla (SGS), Eloi Montcada (Bon Preu), Joan Sabartés (Bon Preu), Angela Schillings-Schmitz (GS1), Ralph Troeger (GS1), and Nikola Vucic (Huawei).

References

European commission; information and communication technologies, work programme 2011–12; EC; July 2010; http://cordis.europa.eu/fp7/ict/

Verweij CA, van der Moolen BJ (2009) De agrologistieke kracht van Nederland. NDL/HIDC, Platform Agrologistiek, TNO p 12

FI-WARE (2011a) High-level description (product vision). Version 1.1 Accessed 31 Aug 2011

FI-WARE (2011b) Overall FI-WARE vision. Project Wiki. URL: http://forge.fi-ware.eu/plugins/mediawiki/wiki/fiware/index.php/Overall_FI-WARE_Vision. Accessed 13 Jan 2011

Trienekens JH, Wognum PM, Beulens AJM, van der Vorst JGAJ (2011) Transparency in complex dynamic food supply chains. Adv Eng Info 26(1):55–65

van Wezel W, Van Donk DP, Gaalman G (2006) The planning flexibility bottleneck in food processing industries. J Oper Manage 24(3):287

Verdouw CN, Beulens AJM, Trienekens JH, Verwaart T (2010a) Mastering demand and supply uncertainty with combined product and process configuration. Int J Comp Int Man 23(6):515–528

Verdouw CN, Beulens AJM, Trienekens JH, Wolfert J (2010b) Process modelling in demand-driven supply chains: a reference model for the fruit industry. Comp Electr Agr 73(2):174–187

Zahariadis T, Papadimitriou D et al (2011) Towards a future internet architecture. In: Domingue J (ed) Future internet assembly 2011, Springer, LNCS, 6656, pp 7–18

Design and Implementation of a Virtual Reader for Generating Multiple Tag Data

Jiwan Lee, Wooseok Ryu, Joonho Kwon and Bonghee Hong

Abstract Recently, RFID technology has become essential to a wide area of applications such as logistic, manufacture and pharmacy. Testing of RFID middleware is a crucial process to increase stability of RFID system. However, it required to install large numbers of RFID readers and RFID tags, which costs us lots of time and expense. Due to this limitation, it is better to use RFID reader emulations for testing RFID middleware. In this paper, we propose a mechanism for generating RFID tags data which closely emulates data generated by the real RFID reader. To do this, we define a probability model for one tag and generate multiple tag data. The experimental results show that the error ratio of our approach is below 9.2 % and feasibility of RFID reader emulation for testing RFID middleware.

J. Lee · W. Ryu · B. Hong
Department of Computer Engineering, Pusan National University,
Busan, Republic of Korea
e-mail: wldhks85@pusan.ac.kr

W. Ryu
e-mail: wsryu@pusan.ac.kr

B. Hong
e-mail: bhhong@pusan.ac.kr

J. Kwon (✉)
Institute of Logistics Information Technology, Pusan National University,
Busan, Republic of Korea
e-mail: jhkwon@pusan.ac.kr

H.-J. Kreowski et al. (eds.), *Dynamics in Logistics*, Lecture Notes in Logistics,
DOI: 10.1007/978-3-642-35966-8_21,

Introduction

Radio Frequency Identification (RFID) which automatically identifies product has become the key technology in ubiquitous computing environment. EPCglobal Inc. establishes hardware and software standards (2008; 2006; 2005) of RFID infrastructure for wide uses of RFID systems in industry. Generally, RFID system consists of RFID tag which stored electronic code for identifying product, RFID reader which reads tag information from RFID tag, and RFID middleware (Mohd Hashim et al. 2008) which collects RFID data and responses results of user queries. Numerous RFID readers transmit a huge stream of tag identifications to the middleware. Performance of RFID middleware is critical because it should correctly process the stream in real time. For example, A large retailer annually handles 60 billion items (Bornhoevd et al. 2004). It means that about 1,900 items should be processed per second. Therefore, the RFID middleware needs to be tested in such harsh environments to ensure performance of the middleware.

Generation of the data using real readers causes a lot of money as to test RFID middleware. Therefore, it is common approach to emulate RFID reader for the testing. However, previous researches for emulating RFID reader do not consider the constraint of reader such as probability of reading tags in real environment. There were only functional emulation (Palazzi et al. 2009; Frischbier et al. 2006) which emulate reader protocols to verify functional correctness. In order to provide the proper emulation of the RFID reader, the emulation should include pattern which generates tags data according to the number of tags. RFID reader cannot always guarantee to read tags. It means that RFID reader can read tags successfully or fail to read tags.

In this paper, we propose a mechanism for generating multiple RFID tags. Our approach is to generate RFID tags data when multiple tags move into RFID reader sight. To do this, we define probability model for one tag, and suggest mechanism for generating multiple RFID tags based on probability model. We also perform an experimental study to evaluate our approach.

The rest of the paper is organized. In section Related Works, we discuss the related work on RFID reader. In section Generation Model, we propose a mechanism for generating RFID tags data. In section Implementation and Evaluation, we show the experiment results and suggest design and implementation. Summary and future works are presented in section Conclusions and Future Work.

Related Works

RFID reader needs to be emulated for testing of RFID middleware. Previous researches of reader emulation (Palazzi et al. 2009; Frischbier et al. 2006) have proposed simply to transfer Electronic Product Code (EPC) to the RFID middleware. Rifidi Emulator is a program that emulates reader protocol (Pramari LLC 2007).

There is related work to builds a virtual store by using Rifidi Emulator. This study introduces a characteristic of reader for emulating reader. It is as follow: every RFID reader has an IP address and a specific port. The reader works like a server. It puts on the socket the information about a tag, which will be read by the application client. Another previous research for emulating RFID reader is handling the functional factor of physical reader. It is not suitable to test RFID middleware. RFID reader may occur an error when reads tags because of the surrounding environment.

However, previous research is also not suitable for emulating RFID reader. Because it is generating tag data in the RP (RP: Reader Protocol) level. The data generator which generates tag data by RP is not similar compared with RFID reader. In this paper, we propose a model for generating RFID tags data based on probability model.

Generation Model

In this section, we propose a model for generating data of RFID tag base on probability. The proposed method is to similarly emulate data generation of RFID reader. We assume that RFID reader collects tags data by 100 % in section Assumption. We present our probability model for one tag in section Probability Model and mechanism for generating data of multiple RFID tags in section Mechanism of Generation.

Assumption

We assume that if the tags are read within RFID Readers sight, it has 100% probability of being a valid read. Figure 1 show our assumption and notation. The notation t_s presents start time and t_e presents end time when RFID reader collects tags information.

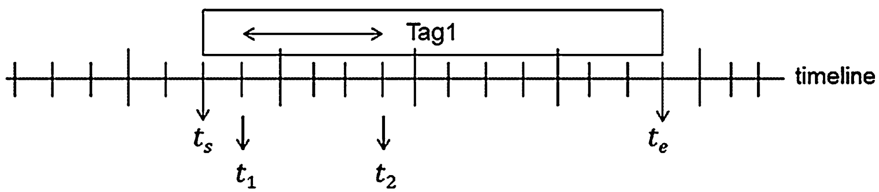


Fig. 1 Case1: fully covered reading time of tag

Probability Model

To emulate generating data of RFID tag, we need to make a probability model. Because RFID readers have characteristic of failed reading when readers collect tags information. Therefore, to emulate generating data of RFID tags, a probability model is needed. We define probability model for one tag at a random interval which is period when RFID reader collects tags information. The probability model $P_s(t_1, t_2)$ is presented as below:

$$P_s(t_1, t_2) = \begin{cases} 100\,\% & t_s < t_1 \; and \; t_2 < t_e, \\ \frac{t_2 - t_s}{t_2 - t_1} \times 100\,\% & t_1 < t_s < t_2 < t_e, \\ \frac{t_e - t_1}{t_2 - t_1} \times 100\,\% & t_s < t_1 < t_e < t_2, \\ 0\% & otherwise \end{cases} \quad (1-4)$$

The probability model can be divided into four intervals between t_s and t_e. Each interval consists of t_1 and t_2 exist between t_s and t_e, t_s exists between t_1 and t_2, t_e exists between t_1 and t_2 and t_1 and t_2 does not exist between t_s and t_e.

Case 1. fully covered reading time of tag: The probability of reading tags is 100 % according to our assumption in section Assumption. Figure 1 show case1.

Case 2. partially covered reading time of tag: The probability of reading tags is decided by included area between t_1 and t_2 as shown in Fig. 2. It is calculated by dividing interval between t_1 and t_s , or t_e to interval between t_1 and t_2. The expression presents in Eq. (2–3).

Case 3. fully uncovered reading time of tag: The probability of reading tags is 0 % according to our assumption in section Assumption. Figure 3 shows case 3.

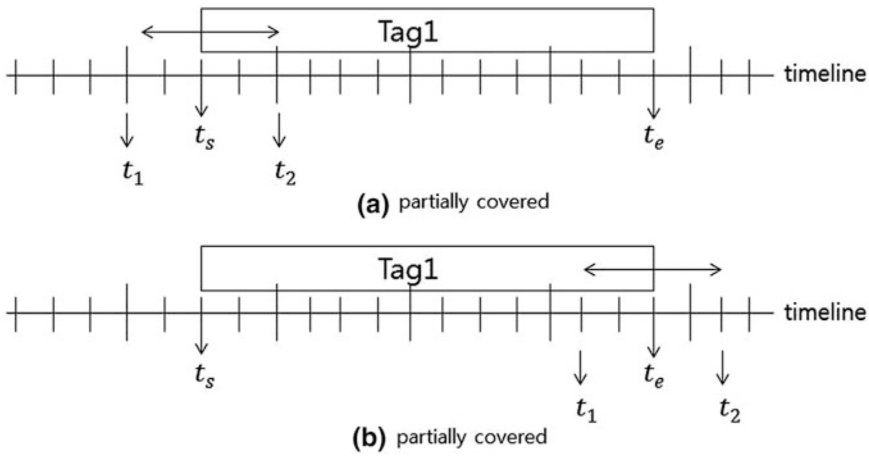


Fig. 2 Case2: partially covered reading time of tag

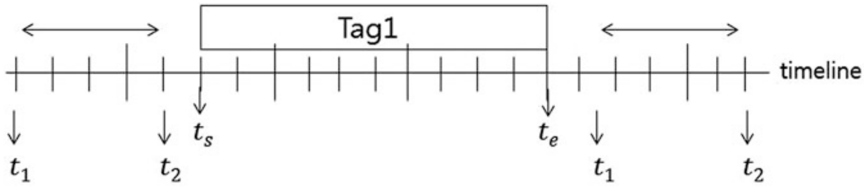


Fig. 3 Case3: fully uncovered reading time of tag

Mechanism of Generation

We propose a model for generating data of multiple RFID tags based on probability model. There are three patterns of data when RFID reader collects tags information. First, RFID reader generates empty report of tags in the RFID reader sight without tags. Second, RFID reader generates report of tag when one of tag moves into RFID reader sight. Third, RFID reader generates report of tags when multiple tags move into RFID reader sight. Therefore, we define tag count function in Eq. (5). LT is the list of tags. $|LT|$ is the number of tags. $\lfloor \vdash \rceil$ is round-off and $LT_i.t_s$ is start time of collecting tags. Each tags have a different reading probability interval t_1 and t_2. Because the enter time of each tag into RFID reader sight is different. To do this, it is to define different probability.

$$F_{tc}(\mathrm{LT}, t_1, t_2) = \begin{cases} 0 & |LT| = 0, \\ \left\lfloor \sum_{i=0}^{|LT|} P_s(t_1 - LT_i.t_s, t_2 - LT_i.t_s) \right\rceil & |LT| > 0 \end{cases} \quad (5)$$

Implementation and Evaluation

Design and Implementation

Based on our probability model, we have design a virtual reader to emulate data generation of RFID reader. The virtual reader system consists of tag data generation, tag count generation and probability model. Figure 4 shows our propose architecture of virtual reader. And we implement the virtual reader system using Java language JAVA (jdk1.6.0.2). We executed the program on a intel core2 quad 2.4 Ghz machine with 2Gb RAM running WindowsXP Pro.

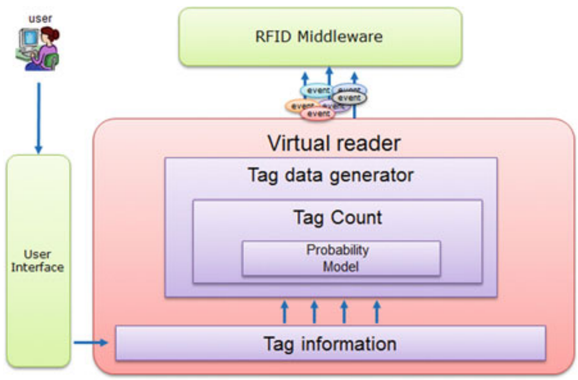


Fig. 4 Architecture of virtual reader for generating tags data

Evaluation

We performed experiments to evaluate proposed mechanism comparing with alien 9,800 RFID reader. In the experiment, we fix the tag's moving speed, the number of tags, interval between each tag and the number of antennas. The initial setting is given in Table 1.

In the first experiment, we compared proposed mechanism with alien 9800 RFID reader. The proposed mechanism which assumes that RFID reader collects data of RFID tag by 100 % is based on probability model. In the result, alien 9800 RFID reader generates 164 data, and our approach generates 227 data. The error ratio is 38 % which compares alien reader with proposed mechanism as Fig. 5 shows.

In the second experiment, we performed to reduce error ratio in Fig. 6. So we applied probability that collects one tag with alien reader. Figure 6 shows ideal reading probability by assumption and empirical reading probability by alien9800, respectively. We applied ideal reading probability to approach1 and empirical reading probability to approach2. After applying the probability of reading, alien 9800 RFID reader generates 164 data, our approach 1 generates 227 data and our approach 2 generates 184 data. The error ratio is 9.8 % comparing alien reader with our apporach 2. From Fig. 7, we can see that our approach 2 reduces error ratio more than approach 1.

Table 1 Settings for experiments

Property	Value
Number of tags	10
Interval between each tag	0.3 sec
Tag's moving speed	40 m/min
Number of antennas	1

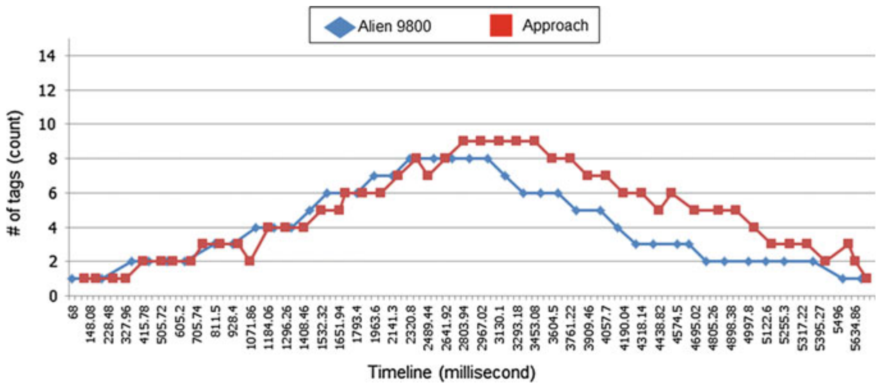


Fig. 5 Comparing our approach with alien 9800 RFID reader

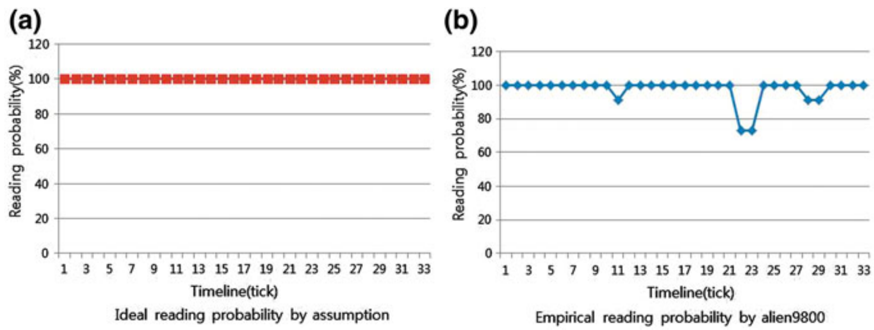


Fig. 6 Alien reader's probability of reading

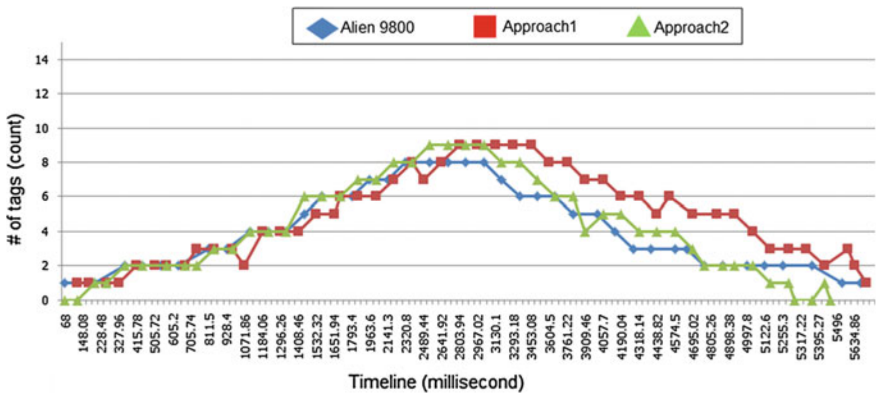


Fig. 7 Comparing our approaches with alien 9800 RFID reader

Conclusions and Future Work

Emulating RFID reader is essential to reduce cost for testing RFID middleware. However, previous work which generates tag data in Reader Protocol level cannot express real RFID readers. These studies cannot provide testing of RFID middleware in environments that move multiple RFID tags into RFID reader sight.

In this paper, we proposed a function for generating multiple RFID tags data based on probability model. To do this, we defined probability model for one tag, and proposed mechanism which generated multiple RFID tags data. We also performed experiments for evaluating the proposed mechanism. The results shows that proposed mechanism was similar compared to alien 9800 RFID reader with 90.2 %. In the future work, we need to design methods for testing RFID middleware by using our proposed virtual readers.

Acknowledgments This research was supported by Technology Development Program for Agriculture and Forestry, Ministry for Food, Agriculture, Forestry and Fisheries, Republic of Korea.

References

Bornhoevd C, Lin T, Haller S, Schaper J (2004) Integrating automatic data acquisition with business processes—experiences with SAP's auto-ID infrastructure. In VLDB, pp 1182–1188

EPCglobal Inc. (2008) The application level events (ALE) specification version 1.1 ratified standard. Part I: core specification, http://www.gs1.org/gsmp/kc/epcglobal/ale

EPCglobal Inc. (2006) The reader protocol version 1.1 ratified standard, http://www.gs1.org/gsmp/kc/epcglobal/llrp

EPCglobal Inc. (2005) The EPCglobal architecture framework, http://www.gs1.org/gsmp/kc/epcglobal/architecture

Frischbier S, Sachs K, Buchmann A (2006) Evaluating RFID infrastructures. Workshop RFID Intelligente Funketiketten—Chancen und Herausforderungen, Erlangen, Germany

Mohd Hashim SZ, Mardiyono Anuar N, Wan Kadir WMN (2008) Comparative analysis on adaptive features for RFID middleware. International conference on computer and communication engineering (ICCCE), pp 989–993

Palazzi CE, Ceriali A, Dal Monte M (2009) RFID emulation in rifidi environment. Dipartimento di Matematica Purae Applicata Universita degli Studi di Padova, pp 4–13

Pramari LLC (2007) Rifidi, Software Defined RFID, http://www.rifidi.org.

Dynamic Tag Identity-Based Encryption Scheme for Logistic Systems

Jongseok Choi and Howon Kim

Abstract In recent years, there is an increasing interest in harbor security and safety due to growing threats by international terrorism. There have been many efforts to cope with international terrorism on the harbor and transport. One of these efforts, department of homeland security of USA has been tried to make a CSD (conveyance security device) security standard such as ISO 18185 to achieve secure transport and logistics. However, since the CSD security standard is based on the symmetric key cryptosystem, which requires the centralized key management system, the CSD security is not widely used for logistic security. It is known that current CSD security standard has difficulty in mutual authentication and signature generation/verification. If we use the PKI (Public Key Infrastructure) based public key cryptosystem for CSD security, we can use the authentication and key management capabilities more easily in eSeal and CSD logistics applications. However, the PKI requires high computational cost, communication overhead, and high storage cost. In this reason, we propose a novel public key cryptosystem for logistics security, which is called dynamic tag ID-based encryption scheme. The proposed scheme requires a trusted agency (TA) to reduce the storage and communication overhead. It is more efficient than conventional PKI cryptosystem from the viewpoint of communication cost because proposed scheme communicates only between tags and TA. Also since proposed scheme does not require storage for a certificate, it is more efficient than conventional PKI based cryptosystem from the point of storage usage.

J. Choi · H. Kim (✉)
Information Security and System LSI Lab, Pusan National University,
Busandaehak-ro 63beon-gil, Geumjeong-gu, Busan, Korea
e-mail: howonkim@pusan.ac.kr

J. Choi
e-mail: jschoi85@pusan.ac.kr

H.-J. Kreowski et al. (eds.), *Dynamics in Logistics*, Lecture Notes in Logistics,
DOI: 10.1007/978-3-642-35966-8_22,

Introduction

The current logistics system only uses symmetric key cryptography that has been limited to authenticate entities because of it needs to share keys among tags. In order to provide secure authentication the system has to adopt public key cryptography (PKC) that requires public key infrastructure (PKI). Unfortunately, it is hard to structure PKI framework for the logistics environment since PKI framework requires very heavy communication and computation cost. Therefore we needed to design the public key cryptography that drive out PKI framework by IBE schemes. Unfortunately an adversary can duplicate the tag by their IDs by clone or Sybil attack. For the issue standards on CSD and E-seal for the logistics system also needs to include an anonymous protocol.

In 2000, Joux (2004) devised three-party two-round key agreement protocol. Boneh and Franklin proposed ID-based encryption (IBE) scheme (Boneh and Franklin 2001) in 2001. The short signature scheme of Boneh et al. (2001) was proposed in 2001. The three schemes were of create interest to cryptographers who started studying on the applications of pairing-based cryptography.

Recently, users and agencies have considered a protocol for user anonymous since Das et al. (2004) had proposed a dynamic ID-based scheme for remote user in 2004. Unfortunately, Chien and Chen (2005) revealed that the dynamic ID-based scheme proposed by Das et al. could not protect anonymous of the users. However, the scheme proposed by Chien and Chen also had a problem of impersonation attack. Researchers had started investigating an authentication scheme for anonymous and an impersonation attack. There has been some dynamic ID-based schemes (Liao and Wang 2009; Wong et al. 2006; I-En Liao et al. 2005), which did not fully satisfy.

Currently, the schemes based on pairing cannot provide user anonymous. Usually, public key cryptography (PKC) should inform the users for public key infrastructure (PKI) which authenticate the legitimacy of a user's public key. Therefore public key cryptography including the scheme based on pairing commonly cannot provide the user's anonymity.

In this paper, we proposed pairing-based scheme which provides the anonymity and the encryption without PKI among two tags and trusted agency. Our approach is to share public key between two tags without revealing their identities using a third party, trusted agency. Two tags generate their public key for encryption every session. After that, trusted agency can authenticate legitimacy of a tag's public key. Finally, two tags can recognize the legitimacy of public key of another tag and securely communicate using the public key authenticated by trusted agency.

The rest of the paper is organized as follows: identity-based encryption of Boneh and Franklin is discussed in section Preliminaries. We describe bilinear maps with properties in section Bilinear Map and give our proposal in section Our Scheme. Our scheme is evaluated in section Analyses. Section Conclusions finally gives conclusions.

Preliminaries

This section describes Boneh and Franklin (2001) scheme. Since Shamir (1984) proposed the ID-based scheme in 1984, which has a problem on key revocation, there have been several proposal for IBE Scheme (Desmedt and Quisquater 1986; Huhnlein et al. 2000; Maurer and Yacobi 1991; Tsuji and Itoh 1989; Tanaka 1987). However, they were not fully satisfactory. Boneh and Franklin firstly proposed the IBE from pairing which is important in this field. The scheme of Boneh and Franklin is described by three algorithms; keygen, encrypt and decrypt.

Keygen

Given a security parameter $k \in \mathbb{Z}^+$, the algorithm performs as follows:

Step1 Run **g** on input **k** to generate two prime order groups $\mathbb{G}_1$, $\mathbb{G}_2$ and a bilinear map $\hat{\boldsymbol{e}}$: $\mathbb{G}_1 \times \mathbb{G}_1 \rightarrow \mathbb{G}_2$. Let **q** be the order of $\mathbb{G}_1$, $\mathbb{G}_2$. Choose a random generator $\mathbf{P} \in \mathbb{G}_1$.

Step2 Pick a random $\boldsymbol{s} \in \mathbb{Z}_{\boldsymbol{q}}^*$ and set $\boldsymbol{P_{pub}} = \boldsymbol{sP}$. Pick a random $\mathcal{Q}_{\boldsymbol{ID}} \in \mathbb{G}_1^*$

Step3 Choose a cryptographic hash function $\boldsymbol{H}_2 : \mathbb{G}_2 \rightarrow \{0,1\}^n$ for some **n**.

Step4 The public key is $\langle q, \mathbb{G}_1, \mathbb{G}_2, \widehat{e}, P, P_{pub}, \mathcal{Q}_{ID}, H_2 \rangle$. The private key is $\langle q, \mathbb{G}_1, \mathbb{G}_2, e, P, P_{pub}, \mathcal{Q}_{ID}, H_2 \rangle$.

Encrypt

To encrypt $M \in \{0,1\}^n$ choose a random $r \in \mathbb{Z}_q^*$ and set the ciphertext to be:

$$C = \langle rP, M \oplus H_2(g^r) \quad \textit{where } g = \hat{e}(\mathcal{Q}_{ID}, P_{pub}) \in \mathbb{G}_2^* \rangle \tag{1}$$

Decrypt

Let $C = \langle U, V \rangle$ be a ciphertext created using the public key P_{pub}.

$$P_{pub} = \langle q, \mathbb{G}_1, \mathbb{G}_2, \hat{e}, n, P, P_{pub}, \mathcal{Q}_{ID}, H_2 \rangle \tag{2}$$

To decrypt C using the private key $d_{ID} \in \mathbb{G}_1^*$ compute:

$$V \oplus H_2(\hat{e}(d_{ID}, U)) = M \tag{3}$$

Our Approach

Recently, researchers, who study on the private security, have been interest in the dynamic ID-based scheme for the logistics system. Various schemes after the scheme of Das et al. did not perfectly overcome the problem on the anonymity. Therefore we propose the dynamic ID-based encryption scheme based on pairing, which can be fully satisfactory without PKI, in this section. Theretofore, we describe the properties of bilinear map used to our scheme. Thereafter, we explain the scheme we proposed in this section.

Bilinear Map

There are two groups $\mathbb{G}_1$ and $\mathbb{G}_2$ of prime order q. For clarity, let $\mathbb{G}_1$ be an additive group and $\mathbb{G}_2$ be a multiplicative group. Sometimes $\mathbb{G}_1$ can be also written multiplicatively.

We assume that P is a generator in $\mathbb{G}_1$, and we can write as follows.

$$aP = \overbrace{P + P + \cdots + P}^{atimes} \tag{4}$$

The useful bilinear map used to our scheme has to fully satisfy following properties.

$$\hat{e} : \mathbb{G}_1 \times \mathbb{G}_1 \rightarrow \mathbb{G}_2 \tag{5}$$

Bilinearity It is satisfactory as following equation.

$$\forall P, Q \in \mathbb{G}_1, \forall a, b \in \mathbb{Z}_q^*, \hat{e}(aP, bQ) = \hat{e}(P, Q)^{ab} \tag{6}$$

Non-Degeneracy If a point is not the identity in $\boldsymbol{G}_1$, the maps result non-identity in G_2.

$$\forall P \in \mathbb{G}_1, P \neq 0 \Rightarrow \hat{e}(P, P) \neq 1 \tag{7}$$

Computability $\hat{e}$ has to be computable.

Our Scheme

We describe the dynamic identity-based scheme we proposed. The scheme uses the ID-based encryption of Boneh and Franklin to exchange some information between a user and a trusted agency (TA). This scheme is very similar in the scheme of Boneh and Franklin except using dynamic ID instead of static ID. Our basic idea is to change the tags' ID whenever a tag wants to securely transport messages with another tag. Therefore, the scheme needs to exchange some information generating a public key of a tag every session. However, this protocol can be more efficient than public key infrastructure (PKI) authenticating legitimacy of a public key. There are three parties, trusted agency and two users, in our scheme. The scheme is composed installation, key exchange, encrypt and decrypt.

Installation

This phase is invoked when a user register in the trusted agency. The agency has two groups $\mathbb{G}_1, \mathbb{G}_2$ of prime order q, a generator P of $\mathbb{G}_2$, two hash function $\mathrm{H}_1 : \{0,1\}^* \to \mathbb{G}_1^*, H_2 : \mathbb{G}_2 \to \{0,1\}^n$, a bilinear map $\hat{e} : \mathbb{G}_1 \times \mathbb{G}_1 \to \mathbb{G}_2$, a public key $TA_{pub} = \alpha P$ and the secret key α. Note that a user U_i stores the public key of the agency TA in this phase. Therefore our scheme cannot use PKI to authenticate the legitimacy of TA's public key.

Step1 $\boldsymbol{TA}$ sends the following parameters to $\boldsymbol{U_i}$ through secure channel.

$$params = \left\langle q, \mathbb{G}_1, \mathbb{G}_2, \hat{e}, n, P, P_{pub}, H_1, H_2 \right\rangle \tag{8}$$

Step2 $\boldsymbol{U_i}$ stores ***params*** in the storage.

Key Exchange

This phase is very important for user's anonymity in our scheme. We have already described that our basic idea is to change ID whenever a user wants to securely communicate with another user. Therefore, two users need to legitimately exchange the dynamic public key through trusted agency every session. Note that the agency can authenticate the legitimacy of user's public key. Consequently, this scheme cannot depend on the PKI, which is very inefficient. We assume that a user U_1 wants to transport messages with a user U_2. Our scheme works as following:

Step1 $\boldsymbol{U}_1$ generates two random numbers $\boldsymbol{x}, \boldsymbol{c_1} \in \mathbb{Z}_{\boldsymbol{q}}^*$.

Step2 $\boldsymbol{U}_1$ sends the following message $\boldsymbol{M}_1$ to $\boldsymbol{TA}$.

$$M_1 = \langle \hat{e}(xP, c_1 H_1(ID_{U1})) \rangle \tag{9}$$

Step3 Upon receiving the message, $\boldsymbol{TA}$ computes $\hat{\boldsymbol{e}}(\boldsymbol{xP}, \boldsymbol{c_1 H_1}(\boldsymbol{ID_{U1}}))^{\alpha}$ and sends the message $\boldsymbol{M}_2$ to $\boldsymbol{U}_1$.

$$M_2 = \langle \hat{e}(xP, c_1 H_1(ID_{U1}))^{\alpha} \rangle \tag{10}$$

Step4 $\boldsymbol{U}_1$ compares as follows.

$$M_2 = ?\langle \hat{e}(xP, c_1 H_1(ID_{U1}))^{\alpha} \rangle \tag{11}$$

If it holds, $\boldsymbol{U}_1$ proceeds next step. Otherwise, $\boldsymbol{U}_1$ terminates the session.

$$U1_{pub} = \hat{e}(\alpha x P, c_1 H_1(ID_{U1})) \tag{12}$$

Step5 $\boldsymbol{U}_1$ generates a random number $\boldsymbol{r}_1$ and computes as following.

$$V_1 = H_2\left(g_{TA}^{r1}\right) \oplus c_1 H_1(ID_{U1})||xP \tag{13}$$

Step6 $\boldsymbol{U}_1$ sends the message $\boldsymbol{M}_3$ to $\boldsymbol{U}_2$.

$$M_3 = \left\langle r_1 P, xP, V_1, U1_{pub} \right\rangle \tag{14}$$

Step7 $\boldsymbol{U}_2$ generates two random numbers $\boldsymbol{y}, \boldsymbol{c}_2 \in \mathbb{Z}_{\boldsymbol{q}}^*$.
Step8 $\boldsymbol{U}_2$ sends the message $\boldsymbol{M}_4$ to $\boldsymbol{TA}$.

$$M_4 = \langle \hat{e}(yP, c_2 H_1(ID_{U2})), M_3 \rangle \tag{15}$$

Step9 $\boldsymbol{TA}$ computes as following.

$$\hat{e}(\alpha ID_{TA}, r_1 P) = \hat{e}(ID_{TA}, P)^{\alpha r_1} = \hat{e}(ID_{TA}, \alpha P)^{r} = g_{TA}^{r_1} \tag{16}$$

Thereafter, $\boldsymbol{TA}$ computes as follows.

$$H_2\left(g_{TA}^{r_1}\right) \oplus V_1 = c_1 H_1(ID_{U1})||xP \tag{17}$$

Consequently, $\boldsymbol{TA}$ can authenticate the public key of $\boldsymbol{U}_1$ as following.

$$\hat{e}(xP, c_1 H_1(ID_{U1}))^{\alpha} = ?U1_{pub} \tag{18}$$

Finally, $\boldsymbol{TA}$ accepts the U_1's public key.

Step10 Like Step3, $\boldsymbol{TA}$ helps $\boldsymbol{U}_2$ generate the public key of $\boldsymbol{U}_2$ by computing.

$$U2_{pub} = \hat{e}(yP, c_2 H_1(ID_{U2}))^{\alpha} \tag{19}$$

and then $\boldsymbol{TA}$ generates a random number $\boldsymbol{r_{TA}}$ and lastly computes as follows.

$$\mathrm{V}_2 = H_2\left(U2_{pub}^{r_{TA}}\right) \oplus \mathrm{H}_2\left(U1_{pub}\right) \tag{20}$$

$\boldsymbol{TA}$ finally sends the following message $\boldsymbol{M}_5$ to $\boldsymbol{U}_2$.

$$M_5 = \left\langle r_{TA}\alpha P, U2_{pub}, V_2 \right\rangle \tag{21}$$

Step11 Like Step4, $\boldsymbol{U}_2$ compares and computes as follows.

$$U2_{pub} = ?\hat{e}\left(TA_{pub}, c_2 H_1(ID_{U2})\right)^{y} \tag{22}$$

$$\hat{e}(r_{TA}\alpha P, yc_2 H_1(ID_{U2})) = \hat{e}(\alpha yP, c_2 H_1(ID_{U2}))^{r_{TA}} = U2_{pub}^{r_{TA}} \tag{23}$$

Thereafter compares as following:

$$V_2 \oplus H_2\left(U2_{pub}^{r_{TA}}\right) = ?H_2\left(U1_{pub}\right) \tag{24}$$

If it holds all of equations in the Step11, $\boldsymbol{U}_2$ proceeds the next step.

Step12 $\boldsymbol{U}_2$ generates a random number $\boldsymbol{r}_2$ and computes as following.

$$V_3 = \left(U1_{pub}^{r_2}\right) \oplus H_2\left(U2_{pub}\right) \tag{25}$$

$\boldsymbol{U}_2$ finally sends the message $\boldsymbol{M}_6$ to $\boldsymbol{U}_1$.

$$M_6 = \left\langle r_2 \alpha P, V_3, U2_{pub} \right\rangle \tag{26}$$

Step13 Like the Step11, U_1 computes as follows.

$$\hat{e}(r_2 \alpha P, x c_1 H_1(ID_{U1})) = \hat{e}(\alpha x P, c_1 H_1(ID_{U1}))^{r_2} = U1_{pub}^{r_2} \tag{27}$$

$\boldsymbol{U}_1$ can verify $\boldsymbol{U2_{pub}}$ by following computation.

$$V_3 \oplus H_2\left(U1_{pub}^{r_2}\right) = ?H_2\left(U2_{pub}\right) \tag{28}$$

Eventually, two users have successfully exchanged the public key of each user.

Encrypt

Two users have exactly exchanged the public key of each user. Therefore, two users can encrypt the data similar to IBE of Boneh and Franklin. However, note that our scheme does not need to reveal the ID of users. And the U_1's public key, $U1_{pub} = \hat{e}(x\alpha P, c_1 H_1(ID_{U1}))$, and U_2's public key, $U2_{pub} = \hat{e}(y\alpha P, c_2 H_1(ID_{U2}))$, are used to encrypt the messages. In this phase, we assume that a public key for encryption is $P_{pub} = \hat{e}(\beta\alpha P, \gamma H_1(ID))$ where β and γ are random numbers and αP is the public key of the *TA* stored in each user's storage. We suppose that a user *A* wants to send a message to a user*B*.

Step1 ***A*** generates a random number ***r*** and computes as follows.

$$V = H_2\left(P_{pub}^{r}\right) \oplus M \tag{29}$$

Step2 Consequently, ***A*** sends the message to ***B***.

$$r\alpha P, V \tag{30}$$

Decrypt

Two users generated the secret keys, U_1's x and U_2's y, corresponding to the public key in key exchange. We assumed that the public key for encryption is $P_{pub} = \hat{e}(\beta\alpha P, \gamma H_1(ID))$. We can easily recognize that the secret key is β in decrypt. Upon receiving the message, $\langle r\alpha P, V \rangle$, the user *B* performs as following.

Step1 ***B*** computes

$$\hat{e}(r, \alpha P, \beta\gamma H_1(ID)) = \hat{e}(\alpha P, \gamma H_1(ID)^{r\beta} = \hat{e}(\beta\alpha P, \gamma H_1(ID))^{r} \tag{31}$$

Step2 Thereafter, $\boldsymbol{B}$ obtains the plain text.

$$V \oplus H_2(\hat{e}(\beta\alpha P, \gamma H_1(ID))^r) = M \tag{32}$$

Analyses

Our scheme partially depends on the security of the scheme of Boneh and Franklin (2001) against some attacks. Consequently, this section focuses on the man-in-the-middle (MiTM) attack and impersonation attack. Note that the users exactly know the public key of the trusted agency.

Man-in-the-Middle Attack

This scheme is composed installation, key exchange, encrypt and decrypt. MiTM attack is that the attacker E can try to concurrently establish two sessions, user-attacker session and attacker-trusted agency session, at the Step2 and the Step3 as following:

Step1 $\boldsymbol{E}$ can intercept the message $\boldsymbol{M}_1$.

Step2 $\boldsymbol{E}$ generates $\boldsymbol{z}$ and $\boldsymbol{c_E}$, sends the following message $\boldsymbol{M}_1^{'}$ to $\boldsymbol{TA}$.

$$M_1^{'} = \langle \hat{e}(zP, c_E H_1(ID_E)) \rangle \tag{33}$$

Step3 $\boldsymbol{TA}$ computes $\hat{\boldsymbol{e}}(\boldsymbol{zP}, \boldsymbol{c_E H_1}(\boldsymbol{ID_E}))^{\alpha}$ and sends the message,

$$M_2^{'} = \langle \hat{e}(zP, c_E H_1(ID_E))^{\alpha} \rangle \tag{34}$$

to $\boldsymbol{E}$. $\boldsymbol{E}$ can perform similar to the original Step4 and establish session with $\boldsymbol{TA}$.

Step4 $\boldsymbol{E}$ has to send the message corresponding to $\boldsymbol{M}_1$. Therefore $\boldsymbol{E}$ cannot help sending $\boldsymbol{M}_2^{'}$ or following messages,

$$M_2^{''} = \langle \hat{e}(xP, c_1 H_1(ID_{U1}))^{z} \rangle \tag{35}$$

to $\boldsymbol{U}_1$ because of $\boldsymbol{E}$ cannot find out the secret key α of $\boldsymbol{TA}$.

Step5 Upon receiving the message from $\boldsymbol{E}$, $\boldsymbol{U_1}$ performs the original Step4. If $\boldsymbol{E}$ sent $\boldsymbol{M}_2^{'}$ to $\boldsymbol{U}_1$, $\boldsymbol{U}_1$ sees the equation,

$$\hat{e}(zP, c_E H_1(ID_E))^{\alpha} = ?\hat{e}(\alpha P, c_1 H_1(ID_{U1}))^{x} \tag{36}$$

Likewise, if $\boldsymbol{E}$ sent $\boldsymbol{M}_2^{''}$ to $\boldsymbol{U}_1$, the equation is following:

$$\hat{e}(xP, c_1 H_1(ID_{U1}))^{z} = ?\hat{e}(\alpha P, c_1 H_1(ID_{U1}))^{x} \tag{37}$$

However, E sends $M_2^{'}$ or $M_2^{''}$, the message will be dropped as above.

Impersonation Attack

Impersonation attack is that the attacker establishes a session disguised as the trusted agency. However, our scheme can overcome the attack by how the scheme detects MiTM attack. Likewise, this attack can be tried to do in key exchange phase. The adversary can try to perform the attack as following:

Step1 $\boldsymbol{E}$ can intercept the message $\boldsymbol{M}_1$.

Step2 $\boldsymbol{E}$ would try to impersonate $\boldsymbol{TA}$ in this step. Accordingly, E computes as following: and then sends $\boldsymbol{M}_2'$ to $\boldsymbol{U}_1$.

$$M_2' = \hat{e}(xP, c_1H_1(ID_{U1}))^{\alpha'} \tag{38}$$

Step3 In this step, $\boldsymbol{U}_1$ verifies the received message as following:

$$\hat{e}(xP, c_1H_1(ID_{U1}))^{\alpha'} = ?\hat{e}(\alpha P, c_1H_1(ID_{U1}))^{x} \tag{39}$$

Note that the adversary cannot discover or compute the secret key α of TA. Therefore, the message M_2' cannot be passed to the next step because of the above equation cannot be held with the attack.

Conclusions

We proposed a pairing based dynamic ID-based scheme, which provides a complete anonymity for the RFID tag. Since our proposed scheme is based on the non-PKI, tag needs to generate key pairs and a dynamic ID for each session. Also, we can easily prove the security of our proposed scheme because our scheme can decrypt/encrypt data with similar way to the IBE scheme by Boneh and Franklin.

In our proposed scheme, the trusted agency provides the legitimacy of each tag's public key without participating the communication between two tags. Of course, the trusted agency can't decrypt the encrypted data, which are encrypted with tag's private keys. These mean that our proposed scheme provides complete anonymity of tags and confidential communications between tags.

The scheme we proposed in this paper can be applicable to logistics system, which is hard to apply a complex PKI system.

Acknowledgments This work was supported by the Grant of the Korean Ministry of Education, Science and Technology (The Regional Core Research Program/Institute of Logistics Information Technology).

References

Boneh D, Franklin M (2001) Identity-based encryption from the Weil pairing. CRYPTO 2001:213–229

Boneh D, Lynn B, Shacham H (2001) Short signatures from the Weil pairing. ASIACRYPT 2001:514–532

Chien HY, Chen CH (2005) A remote authentication scheme preserving user. IEEE AINA'05 245–248

Das Manik Lal, Sacena Ashutosh, Gulati Ved P (2004) A dynamic id-based remote user authentication scheme. IEEE Trans Consum Electron 50(2):629–631

Desmedt Y, Quisquater J (1986) Public-key systems based on the difficulty of tampering. CRYPTO '86 111–117

Huhnlein D, Jacobson M, Weber D (2000) Towards practical non-interactive public key cryptosystems using non-maximal imaginary quadratic orders. Sel Areas Crypt 275–287

Joux Antoine (2004) A one round protocol for tripartite Diffie–Hellman. J Cryptol 17(4):263–276

Liao Y-P, Wang S-S (2009) A secure dynamic ID based remote user authentication scheme for multi-server environment. Comput Stand Interfaces 31(1):24–29

Liao I-E, Lee C-C, Hwang M-S (2005) Security enhancement for a dynamic id-based remote user authentication scheme. International conference on next generation web services practices (NWeSP'05), pp 437–440

Maurer U, Yacobi Y (1991) Non-interactive public-key cryptography. CRYPTO '91 498–507

Shamir A (1984) Identity-based cryptosystems and signature schemes. CRYPTO '84 47–53

Tanaka H (1987) A realization scheme for the identity-based cryptosystem. CRYPTO '87 341–349

Tsuji S, Itoh T (1989) An ID-based cryptosystem based on the discrete logarithm problem. IEEE J Sel Areas Commun 7(4):467–473

Wong KHM, Zheng Y, Cao J, Wang S (2006) A dynamic user authentication scheme for wireless sensor networks. IEEE international conference on sensor networks, Ubiquitous, and Trustworthy Computing (SUTC'06), pp 244–251

Queuing Buffer Method for Enhancing Reliability of Capturing Application

Taewoo Nam, Inhwan Lee, Byeongsam Kim and Keunhyuk Yeom

Abstract The RFID systems used in industries have several problems. When a massive number of RFID events are generated, bottlenecks cause data losses, and duplicated event data cause a decrease in reliability. In this paper, we propose a Queuing Buffer Method that can generate efficient and reliable Event Data to solve these problems. This Queuing Buffer Method temporarily stores the EPC data generated in real time to prevent data losses by bottlenecks and remove unnecessary duplicated Event Data. This method can be applied to various middleware and application systems in RFID environments to obtain reliable Event Data.

Keywords RFID · EPC network · Capturing application · Queuing buffer

T. Nam · I. Lee · K. Yeom (✉)
Department of Computer Science and Engineering, Pusan National University,
30 Jangjeon dong, Geumjeong Gu, Busan 609-735, Republic of Korea
e-mail: yeom@pusan.ac.kr

T. Nam
e-mail: kaluas@pusan.ac.kr

I. Lee
e-mail: ihlee@pusan.ac.kr

B. Kim
Food Safety and Distribution Research Group, Korea Food Research Institute,
Anyangpangyo-ro, Bundang Gu, Seongnam, Gyeonggido 1201-62, Republic of Korea
e-mail: bskim@kfri.re.kr

H.-J. Kreowski et al. (eds.), *Dynamics in Logistics*, Lecture Notes in Logistics,
DOI: 10.1007/978-3-642-35966-8_23,

Introduction

Radio Frequency Identification (RFID) is a general term for technology that uses a radio frequency to recognize and identify people or objects (Juels 2006; Lewis 2004). Such RFID technology is able to recognize multiple tags simultaneously. Furthermore, it can recognize more information by storing data in internal memory. This technology is being extended to use throughout industry and can be employed in various fields, including logistical, supply chain, and ubiquitous applications. For this reason, EPCglobal, a nonprofit organization for the public standardization of RFID technology, has suggested an international standard called EPC Network (Juels 2006; Lewis 2004; Walmart 2005).

EPC Network provides an architecture for collecting and filtering Electric Product Code (EPC), acquiring product information, and tracking products. Effective logistical, supply chain and ubiquitous application systems can be developed using RFID by applying EPC Network (H.K. launches RFID supply Chain Project 2005).

A Capturing Application is an element of EPC Network used to process the EPCs provided by ALE Middleware for EPCIS Event data and stored to EPCIS. Therefore, a Capturing Application is needed when an RFID Event occurs to store it to EPCIS. However, a Capturing Application based on the EPC Network standard has some problems. In particular, at a product line, when attaching an RFID Tag to a product and generating a Produce Event at high speed, there are some problems with processing the Tag as an EPCIS Event, like data losses. For example, the ALE Middleware periodically filters the duplicated Raw Data and sends it to the Capturing Application, but cannot remove it in the receiving interval. In this situation, the Capturing Application sometimes generates duplicated Event data. Therefore, further study is needed on the Capturing Application for reliable EPC. Moreover, when processing the EPC, duplicated EPCIS Events are often generated, which is an important problem that needs to be solved.

In this paper, we propose a Queuing Buffer Method that can be applied to an international standard Capturing Application to solve these problems in an RFID environment. In Fault Detection in Dynamic Vehicle Routing Operations, we provide EPC Network explanations and problems and introduce the Queuing Buffer Method in Dynamic Optimization Model for Planning of Integrated Logistical System Functioning. In Emissions Minimization Vehicle Routing Problem in Dependence of Different Vehicle Classes, we discuss an experiment and analysis results for a Capturing Application that applied the proposed method. In Product Intelligence in Intermodal Transportation: The Dynamic Routing Problem, we explain the contribution of our results and give some conclusions and information about further work.

Related Works

EPC Network Architecture

The current RFID technology was suggested by ISO/IEC and EPCglobal, a standard organization for RFID in the industrial field, in late 1990. In particular, EPC Network is the de facto standard in industry. EPC Network provides EPC technology, which is an identification system that gives a unique code to each object for identification. In addition, it provides a storage/management function for the EPC written in the RFID Tag and recognized by a Reader. In this, EPC Network is an effective base architecture used to construct logistical and supply management systems based on RFID (Retailer RFID Spending Projected To Research 2005; EPCglobal 2007a).

Figure 1 shows an EPC Network Architecture suggested by EPCglobal. This figure illustrates how an RFID Tag is read by a Reader, processed as an EPCIS Event through an ALE and Capturing Application, and stored to EPCIS.

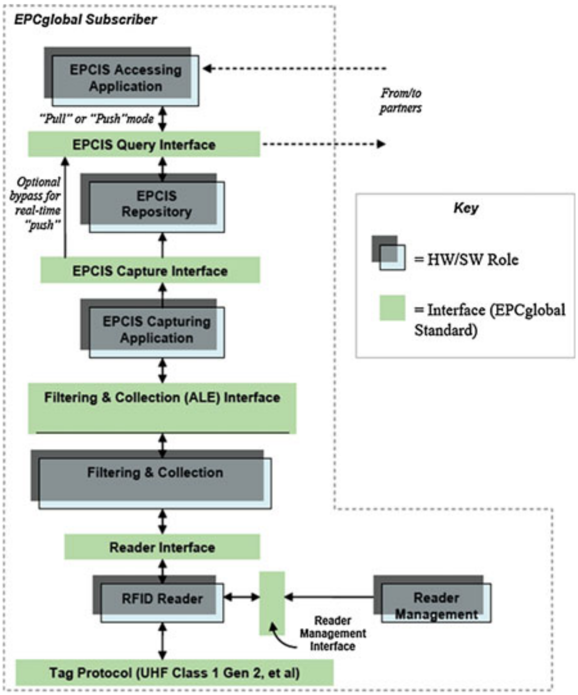


Fig. 1 EPC network architecture

Application Level Event

In Fig. 1, an Application Level Event (ALE) is performed according to the definition of the Event Cycle Specification (ECSpec). ECSpec is registered through a Capturing Application or RFID application's ALE Interface, and generates an ECReport according to the defined specification. This figure illustrates how an RFID Tag is read by a Reader, processed as an EPCIS Event through the ALE and Capturing Application, and stored to EPCIS. The main function of ECSpec is collecting and filtering physical data (RFID Tag) during a certain time (Duration), processing it into the EPC and ECReport formats during a certain interval, and then generating these (EPCglobal 2009b; EPCglobal 2007b; Park et al. 2010). For this reason, ALE is an important middleware used for collecting physical data in an RFID environment.

Capturing Application

The Capturing Application receives an ECReport from ALE and combines it with information related to the EPC. This combined information is processed into the EPCIS Event format before being stored to EPCIS. The number of generated EPCIS Events depends on the number of EPCs in the received ECReport. Therefore, all of the EPCs must be processed into the ECPIS Event format. An EPCIS Event in EPCIS explicitly expresses an RFID Tag's movement. The various flows are used to construct the event data for a particular RFID Tag. Hence, the Capturing Application's role is important in an application system based on RFID. A generated EPCIS Event must be guaranteed to be correct and reliable.

EPC Information Services

EPC Information Services (EPCIS) is the storage system for EPC information. The goal of EPCIS is to enable disparate applications to leverage EPC data via EPC-related data sharing, both within and across enterprises (Park et al. 2010). EPCIS stores the event and master data. High level applications can get event or master data from EPCIS and then add product movement and detail information. The Capturing Application generates an EPCIS Event and sends it to EPCIS like a query. EPCIS receives and processes the query. When receiving an EPCIS Event, the system stores the information and sends it to EPC Discovery Services. EPC Discovery services are at a higher level than EPCIS. Thus, clients can access this high level and get information through EPC Discovery Services and EPCIS.

Therefore, the reliability of an EPCIS Event is very important to EPCIS. If EPCIS's information is incorrect, the reliability of higher level applications may be questionable.

Problems with RFID Middleware

Figure 2 shows the ECSpec flow at ALE. We know that a generated ECReport is delivered to a client application making a request. In Fig. 2, ECReports 1 and 2 in Client1 have duplicated EPCs 3, 4, and 5, and the ECReport in Client 2 has duplicated EPCs 3 and 5.

If the Capturing Application receives an ECReport that includes duplicated EPCs, it will generate unnecessary EPCIS Events. These unnecessary EPCIS Events will cause a problem with the event data's reliability. In addition, if ECReport 2 is received while processing ECReport 1, it will cause a bottleneck because the received EPC data is going to be put in a standby condition.

If the RFID application system only uses master data for immovable products, a duplicated EPC is not a big problem, but when products are being moved, like in

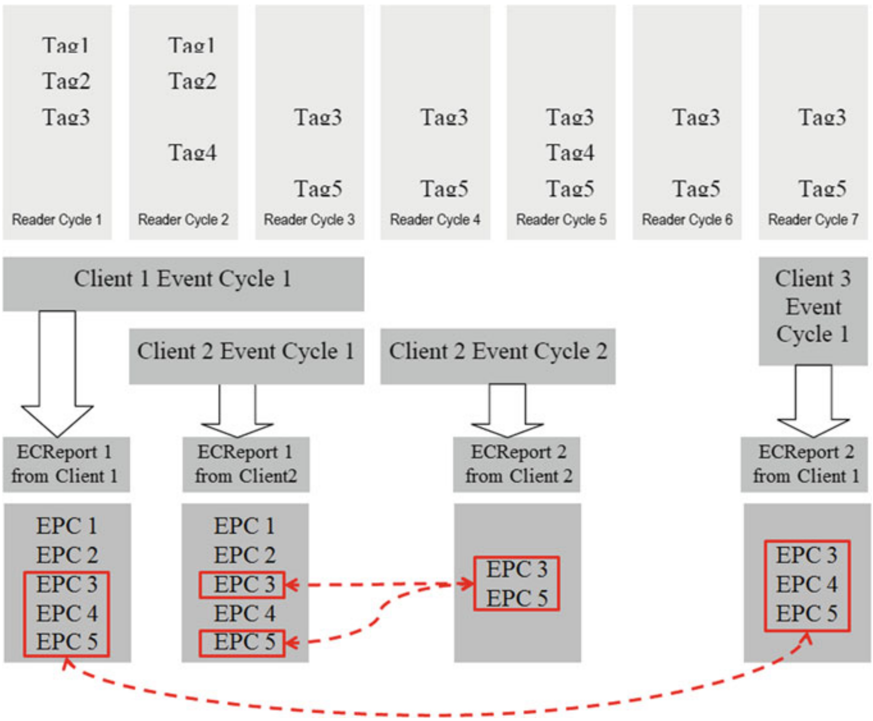


Fig. 2 ECReport generation flow of ALE

a factory or logistical center's release process or a retail center's storage process, there will be a critical problem with the reliability of the RFID information. These problems cause the RFID application system to send unreliable event data.

Queuing Buffer Method for Enhancing Reliability

In this paper, we suggest a solution for the problems described in Fault Detection in Dynamic Vehicle Routing Operations. This solution uses the Queue data structure.

The Queuing Buffer is a memory space that has a First In First Out (FIFO) property like a queue. This buffer is used for the temporary storage of an ECReport received from the ALE. The Buffer stores ECReports in time order and generates EPCIS Events sequentially. It also has a function that removes duplicated EPCs from an ECReport.

This Queuing Buffer Method has two major functions. First, it prevents data losses from bottlenecks. Second, it prevents the generation of duplicated EPCIS Events.

Prevents Data Losses from Bottleneck

In The Capturing Application collects the ECReports from many ALEs and generates one EPCIS Event. Therefore, if one of the ALE's Event Cycles is too

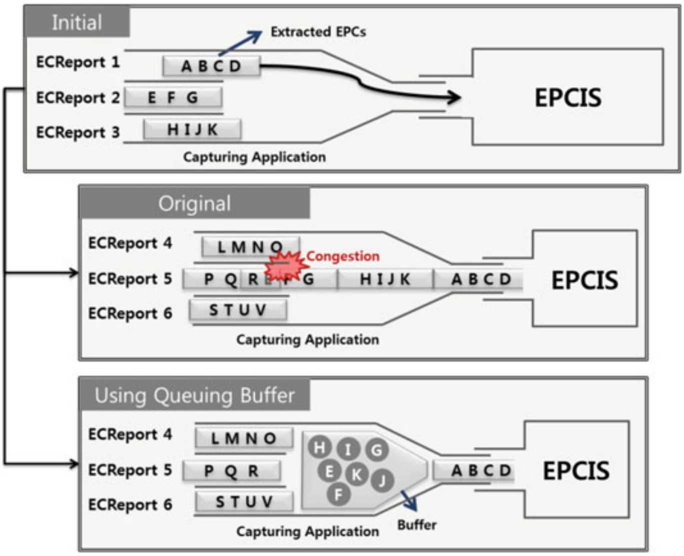


Fig. 3 Preventing bottleneck with Queuing buffer

Fig. 4 Pseudo code for preventing data losses

```
[ALE Adapter in Capturing Application]

ECReport report;
QueuingBuffer queue;

while(is connected with ALE) {
   report = receive ECReport from ALE;
   queue.push(report);
}


[EPCIS Connector in Capturing Application]

while(is queue not empty) {
   report = queue.pop();
   send the report to EPCIS;
}
```

fast, the Capturing Application will be overloaded trying to process the ECReport, as shown in Fig. 3.

This will cause a loss of Event Data or a critical error in the Capturing Application. In addition, if ECReports 4, 5, and 6 occur before the EPCs in ECReport 1, 2, and 3 are processed, the EPCs will be placed in a standby condition and a bottleneck will arise. However, by using the Queuing Buffer, we can prevent the Capturing Application's from being overloaded by periodically emptying the buffer and temporarily storing the ECReports in the buffer. The algorithm in Fig. 3 is expressed in the pseudo code of Fig. 4

Preventing Generation of Duplicated EPCIS Events

Two ECReports are often generated for one EPC according to the Event Cycle of the ALE. Sometimes when an EPC Tag is stopped in front of a Reader, multiple ECReports are generated, which include unnecessary EPCs, until the EPC Tag moves out of the range of the Reader. For this reason, Queuing Buffer analyzes the ECReport Data in the buffer, removes the duplicated EPCs, and sends the EPC to the EPCIS Event Process. The method for removing the duplicated EPCs is illustrated in Fig. 6.

In Fig. 6, if ECReport Data A has the same EPC as ECReport Data B, the EPC in ECReport Data A is removed. The duplicated EPC in ECReport Data A is removed and it is sent to the EPCIS Event Process.

If someone wants to retain the first read event and remove the duplicated data, they can simply change the Queuing Buffer algorithm according to the RFID Application environment.

The algorithm in Fig. 6 is expressed in the pseudo code of Fig. 5. Likewise, ECReports B and C are compared, the duplicated EPC is removed, and ECReport

Fig. 5 Pseudo code for preventing generation of duplicated EPCIS events

```
ECReport report;
QueuingBuffer queue;


while(is connected with ALE) {
   report = receive ECReport from ALE;

queue.push(report);
   queue.removeDuplicatedEPCs();
}

void removeDuplicatedEPCs() {
   List reportList = a list of queue;
   for i = 0 to reportList size {
     for j = i + 1 to reportList Size {
       report[i].remove(same EPC in report[j]);
       }
   }
}
```

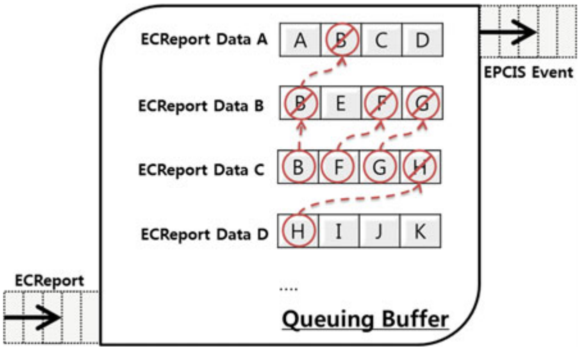


Fig. 6 Queuing buffer method

B is processed by the Capturing Application. This process operates continuously. Consequently, the Queuing Buffer recognizes the EPCIS Event as an EPC only when it moves out of the range of the Reader. Therefore, the Capturing Application can avoid the duplication caused by the Event Cycle period and EPC Tag congestion.

Experiment

Figure 7 shows an experimental environment for comparing a system with and without using the Queuing Buffer Method. Read Point A models an area of a production line where bottlenecks frequently occur, and Read Point B models the release gate where data duplication most frequently occurs.

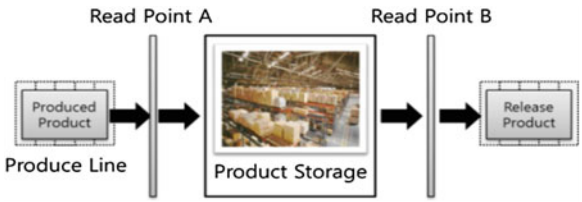


Fig. 7 Experiment environment

Experiment Method

Let us compare two cases: with and without the Queuing Buffer Method. One hundred RFID Tags pass the Read Point A and generate an ECReport. This ECReport is sent to the Capturing Application every 500 ms. We use 500 ms simulate a real production line and assume that three products are produced per second. At Read Point B, release gate congestion is simulated, which generates unnecessary duplicated data artificially by assuming that a vehicle waits at the release gate.

In this paper, we compare applying or not applying the Queuing Buffer.

Experiment Result and Analysis

One hundred RFID Tags are moved from Read Point A to Read Point B according to the suggested experimental method. We get data from the related middleware and summarize it in Table 1. Table 1 shows 12 data losses and 2 data duplications at Read Point A. In addition, at Read Point B, the tags are stopped from moving for several seconds and 62 data duplications occur.

Figure 8 shows a comparison and analysis results based on the experiment data.

Table 1 Experiment result data (size limits)

Section	Read point A		Read point B	
	Original	Applying the method	Original	Apply the method
Processed EPC data (number of tags generating EPICS event)	87	100	143	100
Lost EPC data (number of tags not generating EPCIS event)	13	0	2	0
Number of EPCIS data duplications	0	0	43	0
Number of EPCIS queries	34	5	50	5
Processing time of EPCIS event processor (ms)	407	67	598	68

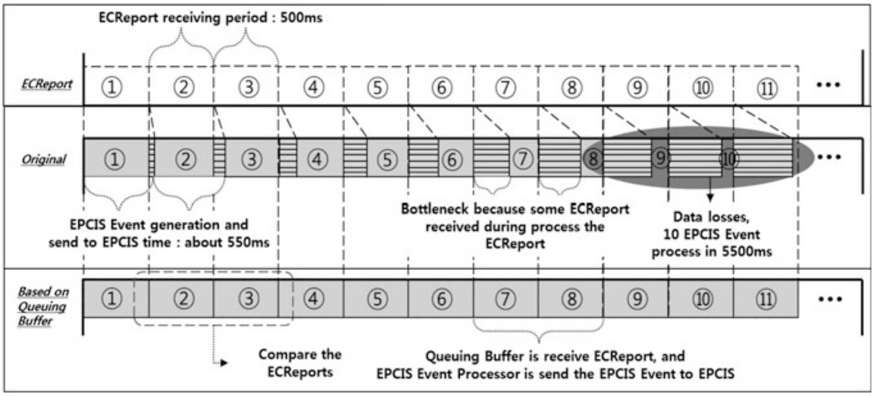


Fig. 8 Experimental result analysis

The average communication time with EPCIS is about 531 ms (Park et al. 2010), and the EPCIS Event processing time is about 12 ms. The ECReport collecting and EPC extracting time is about 8 ms. A total of 550 ms are needs to generate an EPCIS Event. In other words, this is how long a received ECReport is put in a standby condition until the EPCIS Event is generated. Because of this, exceeding the accumulated waiting times for the EPCIS Event processing time causes data losses. However, applying the Queuing Buffer Method sends the data inside the buffer to the EPCIS Event Processor when the queue fills up or the time limit is reached. There are no data losses from bottlenecks.

Read Point B receives the same ECReport from ALE, and generates a total of 162 duplicated EPCIS Events. Like Read Point A, some data losses occur from the bottleneck. However, applying the Queuing Buffer Method can prevent unnecessary EPCIS Events by comparing the generated ECReports and removing the duplicated EPCs.

Table 1 shows the experiment results of using count items. The queue is designed with the capability of storing 11 items. When the queue is full, the Capturing Application processes the first 10 items for an EPCIS Event, and the remaining item is placed into the front. This remaining item does not have duplicated EPCs.

As seen in Table 1, EPCIS communicates 36 and 47 times using the original method but only five times using the Capturing Application that applies the method suggested in this paper. This result shows a decrease in EPCIS communications and a decrease in EPCIS Event processing time.

The next case uses the time interval criterion. In this case, an interval of 1,000 ms is used. The Capturing Application processes the queue every 1,000 ms. Table 2 shows the results with the time interval. There are slight differences from the original case, which is not a big problem because these cannot influence the results of the experiment.

Table 2 Experiment result data (time interval)

Section	Read point A		Read point B	
	Original	Applying the method	Original	Applying the method
Processed EPC data (number of tags generating EPICS event)	88	100	132	100
Lost EPC data (number of tags not generating EPCIS event)	12	0	2	0
Number of EPCIS data duplications	2	0	32	0
Number of EPCIS queries	36	19	47	25
Processing time of EPCIS event processor (ms)	432	247	564	325

Let us compare the results. These two criteria can improve the reliability of the Capturing Application. The results show improved performances compared to the original case. No data losses or duplicated data occur with the Queuing Buffer Method.

However, a comparison of the applications shows some performance differences. The results show that processing the ECReports in the queue when the queue fills up is better than processing them at time intervals.

There are several reasons. Just accessing the queue requires processing time. Thus, more time is needed when using regular time intervals. In addition, this experiment environment produced three products in 1,000 ms. The time interval was 1,000 ms. Therefore, the queue only had three ECReports in each time interval. That is not efficient considering the size limitation. In contrast, when using size limit criterion, the queue is only processed when it fills up with ECReports. Theoretically, in the size limitation case, the queue processed 10 ECReports each time, which is three times the time interval criterion's three ECReports.

Nevertheless, this is a special case. In a real industrial application, products pass read points every 100–1,000 ms. The time interval criterion would work better than in an experimental environment.

The criterion should be carefully chosen. There are performance differences between the criteria according to the RFID system environment.

Conclusions and Future Works

The RFID application system operation is based on a massive amount of data and events. It uses a particular EPC as the parameter for acquiring master and event data. The event data is generated by combining stored ECPIS Event information from many EPCISs. Therefore, a particular EPC must be generated only once from an EPCIS Event.

In this paper, we analyzed the characteristics of an RFID environment and suggested a reliable Queuing Buffer Method that follows the international standard. EPCIS Event generation using this Queuing Buffer Method prevents the generation of unnecessary events by removing the duplicated data collected by ALE, and prevents the loss of EPCIS Events at massive data occurrence points. Hence, it can be applied to various industrial fields that use an RFID infrastructure to acquire reliable event data from an RFID application system. In addition, it can be componentized for application to various middleware based on RFID, and more research is in progress for event generation based on a sensor network.

Acknowledgments This work was supported by u-Food System Development Project of the Korea Food Research Institute.

References

Lewis S (2004) A basic introduction to RFID technology and its use in the supply chain, Laran RFID, White paper

EPCglobal (2007a) Electronic Product Code (EPC): An Overview http://www.idii.com/wp/LaranRFID.pdf

EPCglobal (2007b) EPC information services (EPCIS) version 1.0.1 specification

EPCglobal (2009a) The EPCglobal architecture framework version 1.3

EPCglobal (2009b) Application level events (ALE) specification, version 1.1.1

Juels A (2006) RFID security and privacy: a research survey. IEEE J Sel Areas Commun [0733-8716] 24:381

H.K. launches RFID supply chain project (2005)http://www.rfidjournal.com/article/articleview/1630/1/1

Park S, Kim D, Yeom K et al (2010) The method of effective obtaining distributed information in EPC network environment. J Korean Inst Next Gener Comput August 4

Retailer RFID spending projected to research (2005) http://ww.techweb.com/wire/172303296

Walmart (2005) Supplier information: radio frequency identification uaege http://www.walmartstore.com

Part VI
Mathematical Modeling in Transport and Production Logistics

Dynamic Optimization Model for Planning of Integrated Logistical System Functioning

Iryna V. Morozova, Mykhaylo Ya. Postan and Sergey Dashkovskiy

Abstract In our paper, the dynamic (discrete-time) economic-mathematical model for determination of optimal plane of purchase of raw material and complete set, finished products output by a plant, and their delivery to consumers is proposed. This model is developed for the cases of fixed and random demand at destinations over the given planning horizon. Our approach is based on a generalization of the Wagner-Whitin inventory control model.

Introduction

The logistical process, by definition, may be considered as a system that links enterprise with the suppliers and customers through the distribution channels and transportation systems. Integrated logistical management should provide the effective control for all the elements of a logistical system. The primary concerns of integrated logistical management are the development of optimal production plan and provide materials, work-in-process, finished products optimal inventory control, and optimal transportation plan for the final delivery of a manufactured or

I. V. Morozova
Department of Sea Ports Operation, Odessa National Maritime University, Odessa, Ukraine

M. Ya. Postan (✉)
Department Management and Marketing, Odessa National Maritime University, Odessa 65029, Ukraine
e-mail: postan@ukr.net

S. Dashkovskiy
University of Applied Sciences, Erfurt, Germany
e-mail: dsn@informatik.uni-bremen.de

H.-J. Kreowski et al. (eds.), *Dynamics in Logistics*, Lecture Notes in Logistics, DOI: 10.1007/978-3-642-35966-8_24,

processed product to the customers (Bowersox et al. 1986). To realize such an effective management in practice a corresponding decision support system must be used that is based on operations research methods, for example, inventory control and queueing theories, linear, convex, dynamic, stochastic programming, etc.

Many researchers have paid attention to modeling various logistical systems (Buffa and Miller 1979; Peterson and Silver 1979; Bramel and Simchi-Levi 1997; Dashkovskiy et al. 2005; Brandimarti and Zoterri 2007; Scholz-Reiter et al. 2007; Hutz and Mattfeld 2008; Armbruster et al. 2011). Most part of results obtained in the cited works was concerned with the analysis or optimization only a more or less significant part of integrated logistical system without detailed description of supply, production and transportation processes.

In our paper, the corresponding optimization model is proposed which includes description of main elements of any integrated logistical system. This model is developed for the cases of fixed and random demand at destinations of finished products. Formally, our approach may be considered as a long-standing generalization of the Wagner-Whitin type multi-item inventory model (Bramel and Simchi-Levi 1997).

Optimization Model with Fixed Total Demand Over Planning Horizon

Consider a manufacturing enterprise (plant) which produces the K types of finished product. To manufacture these products the R kinds of material and complete set are used. It is assumed that the matrix

$$A = \|a_{rk}\|, \quad k = 1, 2, \ldots, K; \; r = 1, 2, \ldots, R,$$

of technological coefficients is given, where a_{rk} is an amount of the rth kind of material needed for manufacturing of the kth type of finished product's unit.

A plant purchases all kinds of materials at the R suppliers. The finished products must be shipped to the N destinations. The planning horizon is T (time is measured in discrete units). The total demand for the kth type of finished product at the nth destination over the period T is known and equals to $d_{kn} > 0$ (it may be determined, for example, in result of market's analysis). Taking into account the given demand, plant purchases the materials and manufactures the products.

In addition, we make the following assumptions:

- The market of raw materials is unlimited.
- All ordering of materials and delivering of finished products occurs at the start of each period. Inventories of materials are charged on the amount on hand in the end of each period.
- The lead time is zero; that is, an order arrives as soon as it is placed.
- The time of transportation of any amount of finished product to any destination doesn't depend on this amount.

- The production equipment is absolutely reliable.
- The capacities of production lines of plant are limited only by capacities of warehouses' for storage of raw materials and finished products.

We introduce the following designations:

- Let x_{rt} be the amount of the rth kind of material ordered and purchased in period t, for $t = 1, 2,\ldots,T$.
- Let y_{kt} be the amount of the kth type of finished product which plant plans for output in the end of period t, for $t = 1,2,\ldots,T$.
- Let z_{knt} be the amount of the kth type of finished product planned for delivery to the nth destination in the end of period t, for $t = 1,2,\ldots,T$.
- Let s_{knt} be the sale unit price for the kth type of finished product shipped to the nth destination in period t, for $t = 1,2,\ldots,T$.
- Let p_{rt} be the per unit order cost and K_{rt} be the fixed order cost for the rth kind of material ordered in period t, for $t = 1,2,..,T$.
- Let e_{kt} be the per unit production cost of the kth type of finished product in period t, for $t = 1,2,\ldots,T$.
- Let c_{knt} be the cost of transportation of the unit of the kth type of finished product from plant to the nth destination in period t, for $t = 1,2,\ldots,T$.
- Let $h_{1rt}(h_{2kt})$ be the holding cost per unit of the rth kind of material (of the kth type of finished product) in period t, for $t = 1,2,\ldots,T$.
- Let C_1 (C_2) be the warehouse's capacity for storage of materials (finished products).
- Let $q_{1r}(q_{2k})$ be the initial inventory level of the rth kind of material (of the Kth type of finished product). It is assumed that

$$\sum_{r=1}^{R} q_{1r} \le C_1, \quad \sum_{k=1}^{K} q_{2k} \le C_2.$$

Let $I_{1rt}(I_{2kt})$ be the inventory level of the rth kind of material (of the kth type of finished product) in the end of period t, for $t = 1,2,\ldots,T$.

It is obvious that the following inventory-balanced equations are valid:

$$I_{1rt} = I_{1r,t-1} + x_{rt} - \sum_{k=1}^{K} a_{rk} y_{kt}, \quad r = 1, 2, \ldots, R, \tag{1}$$

$$I_{2kt} = I_{2k,t-1} + y_{kt} - \sum_{n=1}^{N} z_{knj}, \quad k = 1, 2, \ldots, K; t = 1, 2, \ldots, T, \tag{2}$$

where $I_{1r0} = q_{1r}, I_{2k0} = q_{2k}$. From (1) and (2), we obtain

$$I_{1rt} = q_{1r} + \sum_{j=1}^{t} (x_{rj} - \sum_{k=1}^{K} a_{rk} y_{kj}), \quad r = 1, 2, \ldots, R, \tag{3}$$

$$I_{2kt} = q_{2k} + \sum_{j=1}^{t} (y_{kj} - \sum_{n=1}^{N} z_{knj}), \quad k = 1, 2, \ldots, K. \tag{4}$$

Since the total inventory levels $\sum_{r=1}^{R} I_{1rt}$ and $\sum_{k=1}^{K} I_{2kt}$ for any t can't exceed the values C_1 and C_2 correspondingly, from (3) and (4), it follows the necessary conditions:

$$\sum_{r=1}^{R} q_{1r} + \sum_{r=1}^{R} \sum_{j=1}^{t} x_{rj} - \sum_{j=1}^{t} \sum_{r=1}^{R} \sum_{k=1}^{K} a_{rk} y_{kj} \le C_1, \tag{5}$$

$$\sum_{k=1}^{K} q_{2k} + \sum_{k=1}^{K} \sum_{j=1}^{t} y_{kj} - \sum_{k=1}^{K} \sum_{n=1}^{N} \sum_{j=1}^{t} z_{knj} \le C_2, \quad t = 1, 2, \ldots, T. \tag{6}$$

On the other hand, in period t it can't be consumed the rth material and delivered the kth product in amounts more then inventory levels $I_{1,r,t-1}$ and $I_{2,k,t-1}$ correspondingly in the end of period $t - 1$, that is:

$$\sum_{k=1}^{K} a_{rk} y_{kt} \le I_{1,r,t-1}, \quad r = 1, 2, \ldots, R, \tag{7}$$

$$\sum_{n=1}^{N} z_{knt} \le I_{2,k,t-1}, \quad k = 1, 2, \ldots, K;\ t = 1, 2, \ldots, T. \tag{8}$$

From (7), (8), taking into account the relations (3), (4), it follows that

$$\sum_{k=1}^{K} \sum_{j=1}^{t} a_{rk} y_{kj} \le q_{1r} + \sum_{j=1}^{t-1} x_{rj}, \quad r = 1, 2, \ldots, R, \tag{9}$$

$$\sum_{n=1}^{N} \sum_{j=1}^{t} z_{knj} \le q_{2k} + \sum_{j=1}^{t-1} y_{kj}, \quad k = 1, 2, \ldots, K; t = 1, 2, \ldots, T. \tag{10}$$

At last, at the nth destination the kth product must be delivered in amount d_{kn} over the planning horizon, i.e.

$$\sum_{t=1}^{T} z_{knt} = d_{kn}, \quad k = 1, 2, \ldots, K; n = 1, 2, \ldots, N. \tag{11}$$

The expression for total profit of logistical system under consideration over the planning horizon is

$$P = \sum_{t=1}^{T} \left\{ \sum_{n=1}^{N} \sum_{k=1}^{K} p_{knt} z_{knt} - \sum_{k=1}^{K} \left[e_{kt} y_{kt} + h_{2kt} \left(q_{2k} + \sum_{j=1}^{t} y_{kj} - \sum_{j=1}^{t} \sum_{n=1}^{N} z_{knj} \right) \right] - \sum_{r=1}^{R} \left[p_{rt} x_{rt} + K_{rt} \delta(x_{rt}) + h_{1rt} \left(q_{1r} + \sum_{j=1}^{t} x_{rj} - \sum_{j=1}^{t} \sum_{k=1}^{K} a_{rk} y_{kj} \right) \right] \right\}, \tag{12}$$

where $p_{knt} = s_{knt} - c_{knt}; \delta(x) = 1$ if $x > 0, \delta(0) = 0$.

Now we can formulate the optimization problem: it is needed to find out the nonnegative variables x_{rt}, y_{kt}, z_{knt} satisfying the constraints (5), (6), (9)–(11) and maximizing the function (12).

In order that this optimization problem is permissible, from (5), (6), (9)–(11), it follows that the following conditions must be fulfilled:

$$\sum_{k=1}^{K} \sum_{n=1}^{N} d_{kn} \leq (T-1)C_2 + \sum_{k=1}^{K} q_{2k},$$

$$\sum_{k=1}^{K} a_{rk} (\sum_{n=1}^{N} d_{kn} - q_{2k}) \leq q_{1r} + (T-2)C_1, \quad r = 1, 2, \ldots, R.$$

On the other hand, to avoid a trivial situation, obviously we must assume that

$$\sum_{n=1}^{N} d_{kn} > q_{2k}, \quad k = 1, 2, \ldots, K.$$

The optimization problem formulated above may be solved, for example, by dynamic programming algorithm (Hadly 1964) or by the method proposed in (Balinski 1961) which is based on reduction of our optimization model to partly integer linear programming problem.

Model with Random Demand

In practice, demand for finished products is subjected to random fluctuations over the planning horizon. To take into account this circumstance we may generalize the optimization problem presented in Optimization Model with Fixed Total Demand Over Planning Horizon for the case of random demand at destinations. For this purpose let us apply the approach proposed in (Williams 1963).

Now we assume that $d_{kn}(\omega)$ are the continuous mutually independent random variables with the given probability densities $\varphi_{kn}(d)$. Put

$$u_{kn} = \sum_{t=1}^{T} z_{knt}, \tag{13}$$

where u_{kn} is total amount of the kth finished product planned for delivery to the nth destination before realization of random demand $d_{kn}(\omega)$. After its realization, two cases may occur:

1. $u_{kn} < d_{kn}(\omega)$, i.e. demand will not be met;
2. $u_{kn} > d_{kn}(\omega)$, i.e. necessity of storage of the kth product's surplus at nth destination.

We assume that both risks belong to the destinations, i.e. all produced finished products are sold.

Let π_{kn} be the penalty for kth product's deficit at nth destination, and h_{3kn} be the holding cost for storage per unit of kth product at nth destination. Then average total profit of logistical system over the planning horizon is

$$\overline{P} = P - \sum_{k=1}^{K}\sum_{n=1}^{N}\left\{\pi_{kn}\int_{0}^{u_{kn}}(u_{kn}-w)\varphi_{kn}(w)dw + h_{3kn}\int_{u_{kn}}^{\infty}(w-u_{kn})\varphi_{kn}(w)dw\right\}, \tag{14}$$

where P is defined by (12).

It may easily be proven that $\overline{P}$ is concave function in respect u_{kn}. Indeed, taking the second derivative of $\overline{P}$ with respect u_{kn} and applying the Leibnitz rule, we obtain

$$\frac{\partial^2}{\partial u_{kn}^2}\overline{P} = -(\pi_{kn} + h_{3kn})\varphi_{kn}(u_{kn})$$

Since $\pi_{kn} + h_{3kn} > 0$ by definition, the expression in the right-hand side of the last equality is non-positive. Hence, $\overline{P}$ is the concave function.

Thus, we arrive at the following optimization problem in the field of concave programming: maximize the function (14) under constraints (5), (6), (9), (10)–(12), (13) (taking into account non-negativity of control parameters). This problem may be solved with the dynamic programming algorithm, as well.

Generalization for the Case of Multi-phase Transportation

The model given in Optimization Model with Fixed Total Demand Over Planning Horizon may be generalized for the case when finished products have been transported from plant's warehouse to destinations through a set of consequently located intermediate trans-shipment points (phases). For example, first phase may be set of traders and second one may be set of retailers. Below, we consider this case in more details. Let us suppose that finished products are to be transported from plant's warehouse to M destinations through the set of N points of trans-shipment. Let us introduce the additional designations:

- z_{knt} be the amount of the kth type of finished product planned for delivery to the nth point of trans-shipment in the end of period t;
- z_{knmt} be the amount of the kth type of finished product planned for delivery to the mth destination through the nth point of trans-shipment in the end of period t;
- I_{3knt} be the inventory level of kth type of product at nth point of trans-shipment in the end of period t;
- C_{3n} be the warehouse capacity at mth point of trans-shipment.

Obviously, the values I_{3knt} satisfy the following inventory-balanced equations

$$I_{3knt} = I_{3kn,t-1} + z_{knt} - \sum_{m=1}^{M} z_{knmt}, \qquad k = 1, 2, \ldots, K;\ n = 1, 2, \ldots, N; t = 1, 2, \ldots, T. \tag{15}$$

For the sake of simplicity we put $I_{3kn0} = 0$. Then from (15), it follows that

$$I_{3knt} = \sum_{j=1}^{t} \left(z_{knj} - \sum_{m=1}^{M} z_{knmt}\right), \qquad t = 1, 2, \ldots, T; n = 1, 2, \ldots, N; k = 1, 2, \ldots, K. \tag{16}$$

Since the warehouse at nth point of trans-shipment has a finite capacity the following constraints are valid

$$\sum_{k=1}^{K} I_{3knt} \le C_{3n}, \quad t = 1, 2, \ldots, T; n = 1, 2, \ldots, N$$

or taking into account (16)

$$\sum_{k=1}^{K} \sum_{j=1}^{t} \left(z_{knj} - \sum_{m=1}^{M} z_{knmj} \right) \le C_{3n}, \quad t = 1, 2, \ldots, T; n = 1, 2, \ldots, N. \tag{17}$$

Similarly to (7), (8) we have

$$\sum_{n=1}^{N} z_{knmt} \le I_{3kn,t-1}, \quad t = 1, 2, \ldots, T; n = 1, 2, \ldots, N,$$

or (see (16))

$$\sum_{m=1}^{M} z_{knmt} \le \sum_{j=1}^{t-1}\left(z_{knj} - \sum_{m=1}^{M} z_{knmj}\right), \quad t = 1,2,\ldots,T; n = 1,2,\ldots,N; \quad k = 1,2,\ldots,K. \tag{18}$$

At last, in the end of period T all finished products entered to the mth point of trans-shipment must be removed from it, that is $I_{3knT} = 0$ or

$$\sum_{j=1}^{T} z_{knj} = \sum_{j=1}^{T}\sum_{m=1}^{M} z_{knmj}, \quad n = 1,2,\ldots,N; k = 1,2,\ldots,K. \tag{19}$$

The relation (19) may be called the continuity condition by analogy with classical transportation problem (Postan 2006).

Now we compose the objective function for this case. It generalizes the function (12) and has the following form

$$\begin{aligned} P = \sum_{t=1}^{T}\Bigg\{ & \sum_{k=1}^{K}\left[\sum_{n=1}^{N}\sum_{m=1}^{M}(s_{knmt} - c_{knmt})\, z_{knmt} - \sum_{n=1}^{N} c_{knt}\, z_{knt}\right] \\ & - \sum_{k=1}^{K}\left[e_{kt}y_{kt} + h_{2kt}\left(q_{2k} + \sum_{j=1}^{t} y_{kj} - \sum_{j=1}^{t}\sum_{n=1}^{N} z_{knj}\right)\right] \\ & - \sum_{k=1}^{K}\sum_{n=1}^{N} h_{3knt}\left(z_{knt} - \sum_{n=1}^{N}\sum_{m=1}^{M} z_{knmt}\right) \\ & - \sum_{r=1}^{R}\left[p_{rt}x_{rt} + K_{rt}\delta(x_{rt}) + h_{1rt}\left(q_{1r} + \sum_{j=1}^{t} x_{rj} - \sum_{j=1}^{t}\sum_{k=1}^{K} a_{rk}y_{kj}\right)\right]\Bigg\}, \end{aligned} \tag{20}$$

where s_{knmt} is the sale unit price for the kth type of finished product shipped to the mth destination through the nth point of trans-shipment in period t; c_{knmt} is cost of transportation of the unit of the kth type of finished product to mth destination through the nth point of trans-shipment in period t; h_{3knt} is the holding cost for unit of the kth type of finished product at nth destination in period t.

Note that in order to exclude from consideration some destinations it is sufficiently put $c_{knmt} = L$ for them, where L is a large number.

Thus, we arrive to the problem of maximization of function (20) under conditions (5), (6), (9)–(11), (17)–(19) taking into account non-negativity of the variables $x_{rt}, y_{kt}, z_{knt}, z_{knmt}$.

The model presented in this Section may be generalized for the case of random demand at destinations by the same way as it was shown in the Model with Random Demand.

Numerical Results

Let us illustrate the model presented in Optimization Model with Fixed Total Demand Over Planning Horizon for the simplest particular case $K = 2$, $R = 2$, $N = 2$, $T = 2$. For the sake of simplicity we ignore the fixed order cost K_{rt} and put $e_{kt} = 1{,}5, p_{rt} = 0{,}5, h_{1rt} = 0{,}2, h_{2kt} = 0{,}1, \forall\, k, r, t$. The rest initial data are given in the Table 1. The results of the optimization model (5), (6), (9)–(11), (12) solving with the help of *Excel* for some values of demand are presented in the Table 2.

Table 1 Initial data for optimization model

Designation	Numerical value	Designation	Numerical value
C_1	290	p_{212}	2
C_2	330	p_{121}	1
q_{11}	40	p_{122}	2
q_{12}	50	p_{221}	3
q_{21}	100	p_{222}	2
q_{22}	80	a_{11}	0,3
p_{111}	1	a_{21}	0,5
p_{112}	2	a_{12}	0,2
p_{211}	3	a_{22}	0,4

Table 2 Result of calculations

Demand	Optimal supply plan	Optimal production plan	Optimal transportation plan	Maximum value of profit
$d_{11} = 70, d_{21} = 70, d_{12} = 80, d_{22} = 60$	$x_{11} = x_{21} = x_{12} = = x_{22} = 0$	$y_{11} = 50, y_{21} = 50, y_{12} = y_{22} = 0$	$z_{111} = 0, z_{211} = 70, z_{121} = 80, z_{221} = 10, z_{112} = 70, z_{212} = 0, z_{122} = 0, z_{222} = 50$	368
$d_{11} = 60, d_{21} = 50, d_{12} = 50, d_{22} = 55$	$x_{11} = x_{21} = x_{12} = = x_{22} = 0$	$y_{11} = 10, y_{21} = 25, y_{12} = y_{22} = 0$	$z_{111} = 0, z_{211} = 50, z_{121} = 50, z_{221} = 30, z_{112} = 60, z_{212} = 0, z_{122} = 0, z_{222} = 25$	418
$d_{11} = 70, d_{21} = 60, d_{12} = 80, d_{22} = 80$	$x_{11} = x_{21} = x_{12} = = x_{22} = 0$	$y_{11} = 50, y_{21} = 60, y_{12} = y_{22} = 0$	$z_{111} = 0, z_{211} = 60, z_{121} = 80, z_{221} = 20, z_{112} = 70, z_{212} = 0, z_{122} = 0, z_{222} = 60$	366

Conclusions

In our presentation, we proposed the approach to modeling and optimization of integrated logistical system functioning which is based on inventory control theory application. The main idea of our approach is coordination among supply firm, plant, and transport companies at the stage of their joint plans development over the finite planning horizon. The optimization models obtained above are relatively simple and may be implemented in practice with the standard software. But the serious problem here is random variation of demand during the planning horizon. In the further investigations, while developing the optimization models, it is expediently to study the alternative approaches taking into account an uncertainty of demand (fuzzy sets theory, interval arithmetic, etc.). It is interesting also to develop the models describing competition among different integrated logistical systems within the framework of common market. Our approach gives possibility to carry out such a research.

References

Armbruster D, Hendriks MPM, Lefeber E, Udding JT (2011) Structural properties of third-party logistics networks. In: Kreowski H-J et al. (eds) Dynamics in logistics, proceedings of second international conference, LDIC 2009. Springer, Berlin

Balinski ML (1961) Fixed cost transportation problem. Naval Res Log Quart 8(1):41–54

Bowersox DJ, Closs DJ, Helferich OK (1986) Logistical management, 3rd edn. Macmillan Publishing Company, NY

Bramel J, Simchi-Levi D (1997) The logic of logistics: theory, algorithms, and applications for logistics management. Springer, Berlin

Brandimarte P, Zotteri G (2007) Introduction to distribution logistics. Wiley, NY

Buffa ES, Miller JG (1979) Production-inventory systems: planning and control, 3rd edn. Richard D. Irwin, Inc, Homewood

Dashkovski S, Wirth F, Jagalski T (2005) Autonomous control in shop floor logistics: analytic models. In: Chryssolouris G, Mourtzis D (eds) Manufacturing, modelling, management and control 2004. Elsevier Science Ltd, Amsterdam

Hadly G (1964) Nonlinear and dynamic programming. Addison-Wesley Publishing Company, Inc. Reading, Massachusets, Palo Alto

Huth T, Mattfeld DC (2008) Integration of Routing and Resource Allocation in Dynamic Logistic Networks. In: Haasis H.-D. (eds) Dynamics in Logistics. In: Proceedings of first international conference, LDIC 2007. Springer, Berlin

Peterson R, Silver EA (1979) Decision systems for inventory management and production planning. Wiley, NY

Postan MY (2006) Economic-mathematical models of multimodal transport. Astroprint, Odessa

Scholz-Reiter B, Wirth F, Freitag M, Dashkovskiy S, Jagalski T, de Beer C, Rűffer B (2007) Mathematical models of autonomous logistic processes. In: Hűlsman M, Windt K (eds) Understanding autonomous cooperation and control in logistics: the impact on management, information and communication and material flow. Springer, Berlin

Williams AC (1963) A stochastic transportation problem. Opns Res 11, 5: 759–770

A Graph Model for the Integrated Scheduling of Intermodal Transport Operations in Global Supply Chains

Willian Chaves Mates, Enzo Morosini Frazzon, Jens Hartmann and Sérgio Fernando Mayerle

Abstract Integrated scheduling of intermodal transport operations within global supply chains combines flexible land transport schemes with overseas transport running a given timetable. This paper proposes a heuristic scheduling method based on the construction of a cost-weighted graph for each shipment, containing only feasible paths regarding time and capacity. In this way, the scheduling task is formulated as a shortest path problem which can be solved in polynomial time by the well-known Dijkstra's algorithm. The approach is also suitable for larger problem instances, which is demonstrated by means of an example scenario.

Introduction

In the context of global supply chains the integration among partners gains more and more importance (Scholz-Reiter et al. 2007), since the number of partners increases steadily due to the desire to benefit from country-specific advantages and

W. C. Mates · E. M. Frazzon (✉) · S. F. Mayerle
Industrial and Systems Engineering, Federal University of Santa Catarina,
Campus UFSC, Florianópolis, SC 88040-900, Brazil
e-mail: enzo@deps.ufsc.br

W. C. Mates
e-mail: willianmates@deps.ufsc.br

S. F. Mayerle
e-mail: mayerle@deps.ufsc.br

J. Hartmann
BIBA—Bremer Institut für Produktion Und Logistik GmbH, University of Bremen,
Hochschulring 20, 28359 Bremen, Germany
e-mail: hmn@biba.uni-bremen.de

H.-J. Kreowski et al. (eds.), *Dynamics in Logistics*, Lecture Notes in Logistics,
DOI: 10.1007/978-3-642-35966-8_25,

the excellence of the involved partners. Hence, the complexity of global supply chains is increasing (Scholz-Reiter et al. 2010a). The connections between partners should be developed so that they can materialize the competitive advantage of a supply chain (Christopher 2005). In consequence, the transport of shipments between partners, in particular between overseas partners, becomes a crucial task.

This paper proposes a graph theory based heuristic scheduling method for the integrated programming of transport operations within global supply chains. A solution to this scheduling problem assigns to each shipment the transport devices to be used, along with the associated departure, handling, storage and arrival times.

After a review of the currently available methods, it will be shown how to turn the relevant data of the problem into a graph formulation which will be solved iteratively using the known Dijkstra's algorithm. An example scenario for the transportation processes between an original equipment manufacturer (OEM) in Brazil and an Assembler in Germany will be examined to illustrate the method.

Scheduling Transport Operations in Literature

Intermodal transport systems have increasingly made worldwide commercial relationships possible. Due to proceeding globalisation, planning and programming of intermodal transport systems are becoming increasingly relevant (Mattfeld and Kopfer 2003). Since 80 % of all global commercialization of goods occurs by this mode, Berle et al. (2011) claim that the maritime transport mode is a prerequisite for trade at the global level. In this scenario, there is a need for new methodologies for the assessment of global supply chains in a systematic and transparent way (Berle et al. 2011). Different modelling paradigms can be successfully used to describe a supply chain (van Dam et al. 2009).

It is acknowledged that there is an increase in research on intermodal shipments, as well as intermodal transport policies and programming (Scholz-Reiter et al. 2007). The dynamic integration of transport planning at an operational level carries an enormous potential for strengthening the competitiveness of a supply chain. In addition, significant improvements can be obtained through an integrated instead of a sequential programming (Scholz-Reiter et al. 2010a).

Some authors have proposed models for solving the integrated scheduling of a supply chain (e.g. Mula et al. 2010; Macharis and Bontekoning 2004). Nevertheless most of the models consider a supply chain topology network for production and transport planning oriented to the tactical decision level. The purpose of most of these models is the minimization of the total cost of the supply chain (Mula et al. 2010). The demand level, production cost, transport and inventory, and production capacity stand out as shared information. A large variety of OR techniques has been applied, comprising mathematical optimisation, heuristics and meta-heuristics. New techniques have been developed, but for a practically relevant solution of intermodal problems, better techniques are still needed (Mula et al. 2010; Macharis and Bontekoning 2004). It could also be observed that most of the papers focused on solving the integrated scheduling have tried to solve this problem with linear programming or heuristics. In order to improve decision making in dynamic and competitive global environments, the use of resources in logistics systems must be better considered in control systems (Scholz-Reiter et al. 2010a, 2010b; Palekar et al. 1990).

Aiming at developing a model which helps the decision making, this paper will focus on the development of a graph based heuristic which is capable of leading to a result close to the optimum within short computation times. Zhu et al. (2003) developed a graph model which consistently solves traffic grooming problems. In the static case, it can achieve various objectives by using different grooming policies. The authors highlighted the fact that the graph model can be easily adjusted, to cover a broader amount of cases, by manipulating edges and attributing weights. It can be said that the integrated scheduling problem has been studied in the last years and many models have been developed, both conceptual and mathematical. However, with regards to graph theory, not much has been researched.

Integrated Graph Based Scheduling Approach

The main idea of the proposed method is to reformulate the scheduling task as a shortest path problem. In order to do this, a graph $G = (E, V)$ is built, carrying all necessary information only in its set of edges E and nodes V. The following paragraphs present the model assumptions, graph construction and the heuristic scheduling scheme.

Model Assumptions

The model combines land and maritime transport. Land transport via trucks is assumed as being flexibly available at any time, whereas the maritime transport is running a given timetable. The model includes one source (e.g. original equipment manufacturer (OEM)) that provide orders to be transported to sinks (e.g. assembling company ordering goods from the OEM) via several nodes (e.g. ports). The information that characterizes a transport device $s \in S$ is its initial location p_{1s} and destination p_{2s} at times t_{1s} and t_{2s} as well as its cost c_s and capacity k_s. All this information will be assigned to the edges of the graph in section Graph Construction. Furthermore, the ports $i \in I$ are characterized by their location p_i as well as the costs c_i and the time l_i that apply for transferring orders from one vessel to another. An order $j \in J$ features a required transport capacity k^* and a due date τ_j for its delivery to the destination. It is assumed that all orders to be scheduled are available for shipment at time τ^* at the location of the OEM. Note that for all orders $k^*$and τ^* are the same.

As the timetable for maritime transport is preset, there might be waiting times between two subsequent vessels, requiring storage of the cargo and causing additional costs. Given several orders of the assembler, a solution to the scheduling problem is an assignment of all ordered goods to transport devices at specified points in time so that the total cost is minimized. Additionally, the solution has to respect the available capacity of the vessels as well as the due dates of the orders.

Graph Construction

In order to get a formulation of the scheduling task as a shortest path problem all the information mentioned in section Model Assumptions has to be transformed into a graph representation. Most intuitively, the nodes V may represent the ports, distributed in the x-/y-plane according to their geographical position, linked by the edges E standing for the transport device assigned with the travel time, cost and capacity. This approach would lead to a constrained shortest path problem which is NP-complete (Handler and Zang 1980), so it would not be suitable for larger problem instances. In order to overcome this drawback, the graph is built in another way as described in the following. Actually, information about the geographical position of the ports is not relevant in the model since it considers only time and costs, so there is no need to use two axes for the location. Instead, the x-axis is used to create a time-dependent graph as shown in Fig. 1 and only the y-axis is used to assign the locations (ports, OEM, assembler). Consequently a node s of the graph is a pair (t_s, p_s) of a point in time t_s and a position p_s.

The relevant information for a route segment s was indicated as its locations p_{1s} and p_{2v}, times t_{1s} and t_{2s}, cost c_s and capacity k_s. The locations and times can now be stored in two nodes where cost and capacity will be allocated as parameters of the edge between those nodes, as illustrated in Fig. 2. The figure shows a vessel s_1 leaving p_1 at time t_1, and arriving at p_2 at time $t_2 = t_1 + \Delta t$, where Δt is the expected travelling time of s_1 from p_1 to p_2. Analogically, a vessel s_2 leaves the same port at a later time travelling to port p_3.

When a transport device arrives at a port p_i it takes some loading/unloading time l_i until the cargo is ready for the next shipment. This time can also be represented by an edge as shown in Fig. 3. All the costs of the operation (e.g. machine and manpower, port taxes, etc.) are associated to the loading edge, which holds the total cost in its parameters. Depending on the schedule of the maritime transport, there might be a time slot between the time where cargo is ready for shipment at some port and the departure times of the next vessels leaving from here. This gap can be filled by considering it as storage, which can again be represented as an edge with the costs and capacity as assigned parameters (Fig. 3).

Based on a schedule for the maritime transport between OEM and assembler, it is now possible to build a graph for each order taking into account the due date τ_j

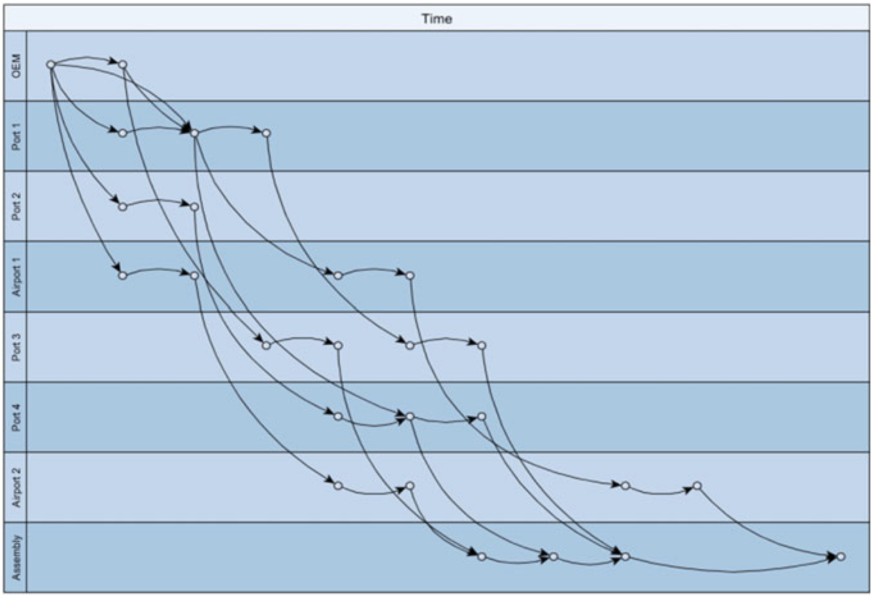


Fig. 1 Time-dependent graph

Fig. 2 Turning vessel information to edges

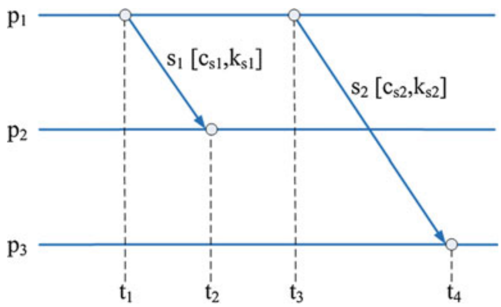
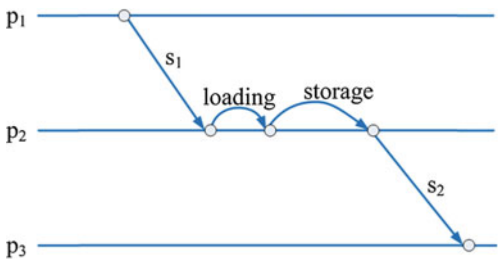


Fig. 3 Connecting subsequent vessels



for the delivery. Due to the time-dependency of the graph, only feasible edges e representing the transport devices s are included into the graph according to the criterion $t_{2s} < \tau_j$. At this stage, the edges of the graph still have two assigned parameters: cost and capacity. The aim is to formulate the initial problem as a shortest (i.e. cheapest) path problem where edges only carry cost information. A solution can then be found in $O(N^2)$ time with Dijkstra's algorithm, where N is the number of nodes in the graph. A heuristic algorithm to realize this situation is presented in the following section.

Heuristic for Scheduling

Based on the timetable for the maritime transport and a set of orders of the assemblers, the graph can be built automatically. This process is shown in Fig. 4. In a first step, the valid time window between the earliest point in time τ^* when the goods can be ready for shipment and the latest due date $max\{\tau_j\}$ is determined. Vessels that leave before the parts are produced or that arrive after the due date are not relevant. The timetable of the maritime transport within the time window is then transformed into edges (blue arrows). Since these edges are not connected, this will be done by some additional edges as follows. When a vessel s arrives at a node (t_{2s}, p_i), a port-specific loading time Δt_p is added (solid green). If there is a vessel $s + 1$ that leaves the port at $t_{1s+1} > t_{2s} + \Delta t_p$, i.e. after the loading is finished, the two points are connected by storage time (dotted green).

At this stage the maritime vessels are linked. In the next step this partial graph is connected to OEM and assembler. It is assumed that the land transport travel times are known and that transport devices are available when needed. So the vessels leaving from ports that have land transport connection to the OEM can be connected with a just-in-time delivery of a truck due to the assumption that storage

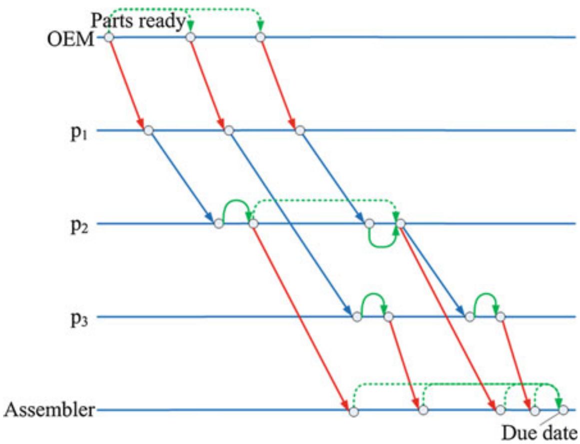


Fig. 4 Building the overall graph step by step

time at the OEM is less expensive than at the port. The same applies to vessels that arrive at ports with land transport connection to the assembler (solid red). Here, the cargo is transported to the assembler directly after the loading process. These operations might still leave time slots between the parts ready point and land transport or between arrival at the assembler and the due date. These gaps are filled with edges representing storage time. In a last step a direct link from OEM to assembler is added with very high cost and unlimited capacity in order to keep the system feasible. Without this edge it would be possible that an order cannot be shipped due to a leak of capacity.

Finally, after the above mentioned steps, an overall graph *G* for a set of orders was built. This does not mean that all the edges in the graph are valid for all orders (e.g. orders with earlier due dates might not be able to be transported by the latest vessels). However, for the formulation as a simple shortest path problem it is required that for the scheduling of an order all edges in the graph are feasible. Thus, for each order the overall graph has to be individually adapted by deleting ineligible edges. In the proposed heuristic the orders were sorted by due date (ascending and descending) and then scheduled one after another. For each order a copy *G'* of the current graph *G* is considered, where all edges arriving at the assembler after the due date are deleted as well as all edges that do not have enough capacity to transport this order. After these operations, any path in the graph is valid and the one with lowest cost can be identified by Dijkstra's algorithm. When this is done, the capacity of this path is reduced in the overall graph *G* to take this into account for the scheduling of subsequent orders. The described method can be realised in an algorithm as shown in Fig. 5.

Fig. 5 Pseudocode for the proposed heuristic

Algorithm: Graph based transport scheduling

Building graph *G*:

initialize relevant time period for scheduling $[\tau^*, max\{\tau_j\}]$;
transfer vessel schedule into partial graph;
for (all edges) {
    add loading time at initial or destination point;
    build possible connections to subsequent edges;
}
add edges representing land transport;
add direct link from source to sink;

Scheduling:

for (all orders) {
    initialize a copy *G'* of *G*;
    delete unfeasible paths for this order in *G'*;
    find shortest paths from source to sink in *G'*;
    reduce capacity of the chosen path in *G*;
}

Computational Analysis

The proposed method aims at performing an efficient intermodal transport planning within global supply chains. In this section it is applied to an intermodal transport problem that connects two production locations in a global supply chain.

Test Scenario

The test scenario consists of an OEM located in Campinas, Brazil and an assembler in Kassel, Germany. The ordered goods are transported via trucks either to the port of Santos or the port of Itaguaí. Both ports offer maritime transport connections to Rotterdam and Hamburg. The last segment of the intermodal transport is done by trucks that connect the ports with the assembler in Kassel, Germany as shown in Fig. 6. The weekly timetable of the maritime vessels is displayed in Fig. 7, which shows for each shipping company the days of departure and the traveling time. The costs of the vessels are assumed to be equivalent to the travel time, i.e. one day of travel time produces costs of one monetary unit (MU).

For the sake of simplicity, the land transport times between ports and OEM or assembler are assumed to be one day, the loading time at all ports is half a day (all at the cost of one MU and with unlimited capacity). The cargo is supposed to arrive as close to the due date as possible, therefore the storage time at the assembler is penalised with a cost of three MU per day, compared to one MU at the OEM.

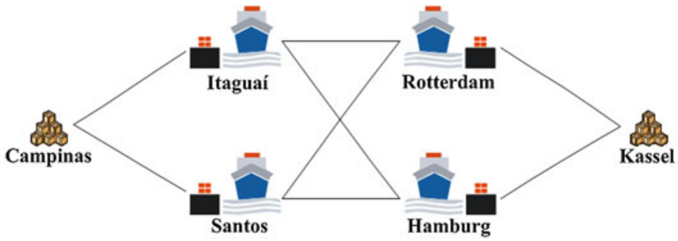


Fig. 6 Test case

Santos -> Hamburg	M	T	W	T	F	S	S
ALIANÇA			18		22		28
CMA CGM			18	31	22		
CSAV		21					
DELMAS			18		22		
HAMBURG SÜD			18		22	21	28
HANJIN SHIPPING						22	
HAPAG-LLOYD			18		22		
MAERSK LINE		21					
MSC		20					
SAFMARINE		21					
UN. ARAB SH. CO			19				
WAN HAI LINES LTD.			50				

Santos -> Rotterdam	M	T	W	T	F	S	S
ALIANÇA			15		19		25
CMA CGM			15		19		
CSAV		15				14	15
DELMAS			15		19		
HAMBURG SÜD			15		19		25
HANJIN SHIPPING						20	
HAPAG-LLOYD			15		19		
MAERSK LINE			15				
MSC		15				14	
SAFMARINE		15					
UN. ARAB SH. CO			17				
WAN HAI LINES LTD.			47				

Itaguaí -> Hamburg	M	T	W	T	F	S	S
CSAV			19				
MSC			19				

Itaguaí -> Rotterdam	M	T	W	T	F	S	S
CSAV			14				
MAERSK LINE			14				
MSC			14				

Fig. 7 Weekly timetable of maritime transport vessels

#scheduled weeks → #orders ↓	1		3		5	
	ascending	descending	ascending	descending	ascending	descending
10	1.522	1,77%	1.721	0,00%	1.522	1,77%
	4,62 [s]	4,76 [s]	5,43 [s]	5,53 [s]	8,48 [s]	8,67 [s]
20	3.276	3,14%	3.472	1,01%	3.044	1,77%
	11,76 [s]	12,4 [s]	12,79 [s]	13,45 [s]	18,58 [s]	20,67 [s]
50	9.762	89,34%	11.397	0,65%	11.317	1,40%
	49,8 [s]	53,95 [s]	60,5 [s]	65,59 [s]	95,79 [s]	110,37 [s]
100	59.594	10,01%	50.275	13,28%	36.079	8,69%
	165,61 [s]	180,93 [s]	215,84 [s]	230,31 [s]	354,07 [s]	396,78 [s]

Fig. 8 Computational results: total costs for ascending order, costs of descending order relative to ascending order, computational time

Results

The proposed method was prototypically implemented in MATLAB 7.12 and applied to the test scenario from section Test Scenario on an Intel Core 2 Duo 3.0 GHz CPU with 4 GB of RAM. In order to show the performance of the program, the number of orders was varied between ten and 100 and the due dates were set so that the scheduling had to be done for a one to three weeks period. The results are shown in Fig. 8. As expected, the computational time increases with a growing number of weeks that have to be scheduled since the graph contains more and more edges. However, despite quintuplicating the number of weeks the computation time only doubled. The second parameter that has influence on the computation time is the number of orders that have to be scheduled. The table shows that the time is growing faster than the number of orders. The quality of the solution may depend on the sequence in which the orders are scheduled. The reason for this is the stepwise reduction of available path capacity which might hinder subsequent orders from taking their preferred path. This was studied with two different sorting schemes, ascending and descending order of due dates. Figure 8 shows that in all cases the ascending order of the due dates was superior or equal to the descending order.

Conclusions and Future Research

This paper introduced a heuristic approach for the scheduling of intermodal transport operations based on graph theory. While an intuitive way of formulating the task by means of a graph would lead to an NP-complete problem, the complexity could be reduced to a shortest path problem, solvable in polynomial time, by building a time-dependent graph.

The heuristic approach allows solving larger problem instances. The results of the application to a test scenario for the scheduling of transport operations between an OEM and an assembler show that the method is suitable to tackle the scheduling task with good performance. The results were achieved by using a prototypical implementation, which was not optimized considering runtime performance. For more significant results to this aspect, a more elaborate code should be implemented.

In this paper the orders that had to be delivered to the assembler were scheduled depending on their due date. The test scenario showed that in most cases an ascending or descending order lead to different total costs. Further improvements might be achieved by studying the influence of this parameter.

In the context of global supply chains an integrated planning of production and transport processes is a promising concept (Scholz-Reiter et al. 2010a). The method proposed in this paper only dealt with transport scheduling with a known point in time when the products are ready for shipment from OEM to assembler. However, the graph based approach appears to be flexibly adaptable to a wide range of scheduling events, so that it seems promising to use it also for the scheduling of production processes and extend the proposed method to an integrated approach.

Acknowledgments This research was supported by CAPES, CNPq, FINEP and DFG as part of the Brazilian-German Collaborative Research Initiative on Manufacturing Technology (BRAGECRIM).

References

Berle Ø, AsbjØrnslett BE, Rice JB (2011) Formal vulnerability assessment of a maritime transportation system. Reliab Eng Syst Saf 96:696–705

Christopher M (2005) Logistics and supply chain management: creating value-adding networks. Financial Times/Prentice Hall, Englewood Cliffs

Handler GY, Zang I (1980) A dual algorithm for the constrained shortest path problem. Networks 10:293–309

Macharis C, Bontekoning YM (2004) Opportunities for OR in intermodal freight transport research: a review. Eur J Oper Res 153:400–416

Mattfeld DC, Kopfer H (2003) Terminal operations management in vehicle transshipment. Transp Res Part A: Policy Pract 37:435–452

Mula J, Peidro D, Díaz-Madroñero M, Vicens E (2010) Mathematical programming models for supply chain production and transport planning. Eur J OR 204:377–390

Palekar US, Karwan MH, Zionts S (1990) A branch-and-bound method for the fixed-charge transportation problem. Manag Sci 36:1092–1105

Scholz-Reiter B, Novaes AGN, Frazzon EM (2007) Innovation, logistics and internationalisation: strategic perspective and cultural aspects. In: Orrico Filho RD, D'Agosto MA (eds) Panorama Nacional de Pesquisa em Transportes: XXI Brazilian Congress of Transportation and Logistics, 2007 Rio de Janeiro. Cybis, H.B.B

Scholz-Reiter B, Frazzon EM, Makuschewitz T (2010a) Integrating manufacturing and logistic systems along global supply chains. CIRP J Manuf Sci Technol 2:216–223

Scholz-Reiter B, Schwindt C, Makuschewitz T, Frazzon EM (2010b) Approach for the integration of production scheduling and inter-facility transportation within global supply chains. In: Hanna MD (ed) 21st annual POMS conference—operations in emerging economies. Vancouver, Canada

van Dam KH, Adhitya A, Srinivasan R, Lukszo Z (2009) Critical evaluation of paradigms for modelling integrated supply chains. Comp Chem Eng 33:1711–1726

Zhu H, Zang H, Zhu K, Mukherjee B (2003) A novel generic graph model for traffic grooming in heterogeneous WDM mesh networks. Trans Networking 11:285–299

Exploitation of Due Date Reliability Potentials: Mathematical Investigation of the Lead Time Syndrome

Mathias Knollmann and Katja Windt

Abstract Due date reliability is a key performance indicator that directly affects customers. Therefore, direct or indirect measures are taken to improve the due date reliability of production processes. These measures induce various problems such as the Lead Time Syndrome. To exploit new potentials regarding due date reliability, this paper presents the mathematical investigation of the Lead Time Syndrome, to point out research topics in this area.

Introduction

Recent developments in research on due date reliability have led to a renewed interest in the so-called Lead Time Syndrome (LTS). The LTS can be summarized as a circle of mistakes: for a targeted improvement of inadequate due date reliability, adjustments of planned lead times are made, which result in actually negative effects on due date reliability. So far, however, there has been little discussion about the formal foundation of the LTS. The original line of argumentation of Mather and Plossl in 1977 was merely logically questioned, but not fundamentally investigated towards the logistical effects in networked production systems. The aim of this paper is to scrutinize the underlying assumptions of the LTS in a more formal way in order to derive research questions that are not yet fully covered. In this respect, the theoretical coherences and the cross-linkage of

M. Knollmann (✉) · K. Windt
Jacobs University Bremen GmbH, Campus Ring 1,
28759 Bremen, Germany
e-mail: M.Knollmann@jacobs-university.de

K. Windt
e-mail: K.Windt@jacobs-university.de

H.-J. Kreowski et al. (eds.), *Dynamics in Logistics*, Lecture Notes in Logistics,
DOI: 10.1007/978-3-642-35966-8_26, © Springer-Verlag Berlin Heidelberg 2013

the logistical effects of the parameter adaptation need to be investigated in further detail. Based on this understanding there is a further need of discussion concerning the appropriate strategy a company should follow in order to avoid the LTS.

We will firstly give an overview of logistical targets and their trade-offs followed by a short introduction into the context of the LTS and the recent history, in addition to previous research results. Afterwards, the underlying scheduling techniques are shortly described. Within the main section of this paper we aim to derive the LTS through mathematical equations according to the logistics operating curve theory by Nyhuis and Wiendahl, to raise questions about the assumptions that have been made to originate the LTS. Therefore, some of these assumptions are discussed subsequently. Finally, as the practical importance of the LTS needs to be discussed, the relevance of the problem will be pointed out, by suggesting adequate utilization of countermeasures against the LTS. This research will serve as a base for future studies to exploit potential improvements of due date reliability, under consideration of the LTS and its interactions.

State of the Art

Logistical Targets and Their Trade-off

Adherence to delivery dates represents the main target from a customer's point of view. But targets of production control like short lead times,[1] low work in process (WIP), appropriate capacity utilization and high due date reliability are conflicting (Nyhuis and Wiendahl 2009) [also known as the 'Dilemma of Operations Planning' (Gutenberg 1951)]. E.g., to ensure appropriate capacity utilization a higher WIP is necessary, which induces higher lead times. Due date reliability is defined as the number of orders produced within a due date tolerance in proportion to the total number of orders (Lödding 2008). Therefore, due date reliability theoretically requires constant lead times with limited variance rather than realistic planned lead times. To solve this dilemma, a certain WIP-level has to be defined for each specific work system within the so-called 'logistical positioning area'. Thereby, one logistical target indicates the primary goal (e.g., appropriate lead times), which affects other target values due to the causal dependencies (Lödding 2008; Nyhuis 2007). The calculation of logistic operation curves of work systems offers the possibility to distinguish between three operating positions (see Fig. 1): underload, transient area and overload state.

Due to the lack of dissemination, the powerful application of logistical operating curves is still not well known in industry. In order to avoid underutilized capacities of bottleneck work systems, a safety factor is often added to planned lead times.

[1] The terms 'throughput time' and 'lead time' are often used synonymously. According to the investigated 'Lead Time Syndrome', only the term 'lead time' is used to simplify it for the reader.

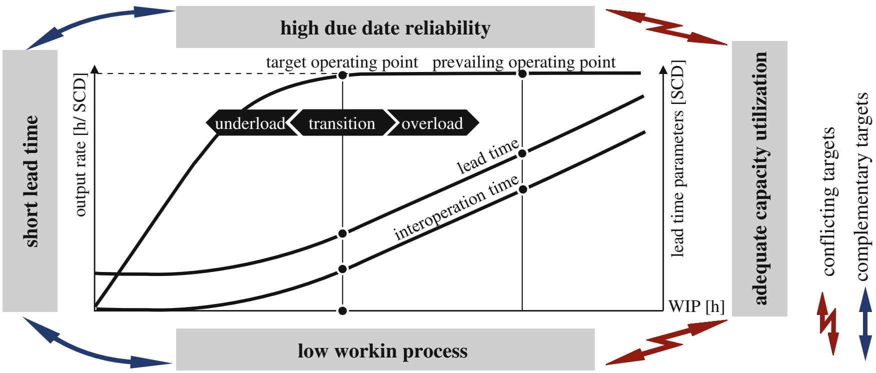


Fig. 1 Logistic positioning [based on (Nyhuis and Vogel 2006; Nyhuis and Wiendahl 2006)]

An earlier order release therewith occurs and causes a higher WIP-level. This means that the work systems states are accidently positioned in the overload area, which results in even longer lead times (Fig. 1, prevailing operating point) (Lödding 2008). This shows, how an adaptation of planned lead times is able to trigger the LTS without adequate consideration of its possible effects.

Specification of the Lead Time Syndrome

The uncertainties in production planning often cause a re-assessment of the production processes and thus may induce the so-called 'Lead Time Syndrome' (Mather and Plossl 1977). The line of argumentation of Mather and Plossl and Wiendahl can be summarized as follows (Fig. 2): If the due date reliability is low, it seems sensible to increase planned lead times, because it seems as if prior planned lead times were set too short to produce on time. Due to the applied backward scheduling, the orders are released earlier and the workloads at the work

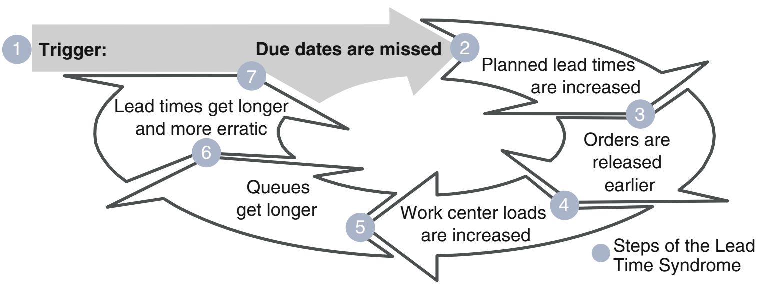


Fig. 2 Lead time syndrome of production control [based on (Mather and Plossl 1977; Wiendahl 1997)]

stations increase. Therefore, the WIP increases and lead times get longer and more erratic. The performed countermeasure reinforces the problem in the end.

Ultimately, the resulting schedule adherence is even worse. To deliver urgent orders in time, rush orders are required. This leads, in theory, to a vicious circle, which continues until the lead times reach a very high level (Nyhuis and Wiendahl 2009, 1997). However, in practice, production planners would intervene early enough to work against too high lead times. Nevertheless, e.g., capacities of bottleneck systems are often not expandable (at least for the short term) and the inequality of planned, mean and real lead times complicate the capacity calculations of scheduling.

Previous Research on the Lead Time Syndrome

The negative influence between logistical values in the sequence of Fig. 2 was firstly discovered by Mather and Plossl and has been coined 'The Vicious Cycle' (Mather and Plossl 1977). Several authors used the line of argumentation to investigate different production planning and scheduling techniques as well as to improve due date reliability (assembly controlling (Wiendahl 1997); logistic positioning (Nyhuis and Wiendahl 2010); production planning and controlling (Schuh 2006); production logistics (Nyhuis and Wiendahl 2009); techniques of assembly controlling (Lödding 2008); logistic company organization (Wiendahl 2008); workload control (Breithaupt et al. 2002). Initially Selçuk et al. analyzed the effects of updating lead times in planning systems and stated that "frequent updates of lead times will lead to uncontrolled production system states with erratic and very long lead times" (Selçuk et al. 2006). Based on more formal studies Selçuk et al. investigated the effects of frequent planned lead time updates under the use of a two-dimensional Markov process (Selçuk et al. 2009). In the context of studying the impact of human and organizational factors on planning stability, Moscoso et al. introduced the term 'planning bullwhip' that subsumes a wide variety of planning instabilities such as the LTS (Moscoso et al. 2011). However, the following questions and coherences are still not sufficiently investigated and answered and thus demand for further investigation:

- Many calculations of production controlling are based on mean values. As the reaction time (until actual values are considered in scheduling calculations) affects all logistical calculations and therefore the processes, the systems reaction behavior and its influences need to be investigated.
- The fundamental assumptions of the LTS and the interactions between logistical targets have not yet been analyzed in this context.
- There is a strong need to define the optimal composition of countermeasures (depending on the prevailing environmental conditions and characteristics) that will work against the causes and effects of the LTS.

Though this list is not exhaustive, it shows that the different scheduling techniques such as Flow Rate Oriented Scheduling, whose calculations strongly depend on the work system response time, have a great impact on the LTS and thus on achieving a high due date reliability. The current research was not specifically designed to evaluate the formal fundamentals of the LTS. To investigate the second research gap in a first step, a derivation of the LTS under use of logistical equations is essential.

Basic Conceptions of Relevant Scheduling Techniques

A main mechanism of the LTS is the rising workload in the processes, which is the result of the underlying scheduling technique. The strong influence of the scheduling techniques to launch and reinforce the LTS demands for a short introduction into scheduling which goal can be described as follows:

> Virtually all manufacturing managers want on time delivery, minimal work in process, short customer lead times, and maximum utilization of resources. (…). The goal of production scheduling is to strike a profitable balance among these conflicting objectives (Hopp and Spearman 2008).

In the context of the LTS, the prevailing scheduling techniques can be subdivided into supporting and mitigating techniques. The following extract of techniques summarizes the relevant ideas of the LTS:

Supporting Scheduling Techniques

Backward scheduling (Kumar 2006; Schönsleben 2012): This type of scheduling is often used in assembly type industries and provides the latest possible start date for orders. Therefore, the order lead times have to be summed up backwards from their latest acceptable completion date.

Forward scheduling (Kumar 2006; Schönsleben 2012): Forward scheduling is often used in job shops that want to produce as fast as possible after an order is received. Hence, the order lead times have to be summed up from the earliest possible starting date (today), to provide the earliest possible completion date.

Mitigating Scheduling Technique

Flow Rate Oriented Scheduling (*FROS*) (Lödding 2008; Nyhuis 2007; Nyhuis and Wiendahl 2009): FROS determines the mean interoperation times under use of the logistical operating curves. Therefore, the operation lead time is separated into operation and interoperation times. The optimal interoperation time that defines the position on the logistical operating curves is calculated using the optimal work system flow rate. For this reason, it is likely to find the optimal position for work systems with the logistical positioning to meet the business objectives, such as adequate capacity utilization and moderate lead times.

Recurring adjustments of the planned lead times within scheduling techniques, like FROS, raise the question of what happens to the capacity utilization at the work systems and what influence the mentioned scheduling techniques have, as maximum output rates remain constant. Moreover, occasional deviations due to WIP fluctuations as well as the impact of the scheduling frequency (interval between lead time adjustments) may induce the LTS, as calculations for capacities and scheduling strongly depend on mean values, which are unequal to real values of work systems.

Investigating the Lead Time Syndrome

Derivation of the LTS Through Logistic Equations

As shown in the preceding chapters, many researchers dealt with the LTS and its causes and effects in theory as well as in practice. To question the underlying assumptions, a more formal research is needed. In recent years, the formal description of the above mentioned logistical targets has made progress. This offers the opportunity to investigate the LTS under use of equations according to the operating curve theory of NYHUIS & WIENDAHL. For lack of space, the main assumptions of the LTS are discussed in detail in the next paper and are marked with a superscripted character X in the left description part. According to Fig. 2, the following equations are subdivided into the seven steps of the LTS:

The result of these interdependent equations is a vicious cycle with a bad due date reliability as initiator. Different assumptions were made to fulfill the idea of the LTS. These assumptions need to be proven regarding their topicality, correctness and completeness, which will be done in the following paper.

Examination of Lead Time Syndrome Assumptions

The line of argumentation of Mather and Plossl postulated different assumptions. Some of these were marked with a superscripted character X in the paper above. Broken down in these characters a more detailed characterization of assumptions is possible to point out the ideas behind the LTS and to enable a critical examination as a foundation for further investigations:

J *Planned lead times* (*Step* 2): Planning and controlling systems often use average values for calculations (e.g., FROS). As real lead times differ from planned and mean values, capacity calculations are necessarily wrong. Moreover, if a process step needs, e.g., 5 days and only 4 days are planned, poor due date reliability is unavoidable. Therefore, the question must be raised, if and if yes, how lead time adjustments necessarily lead into the LTS?

Seven steps of the LTS			
Step 1: Due dates are missed			
Target equation	The relative amount of orders that are produced in a defined time slot around the due date is low. This leads to poor due date reliability (Breithaupt et al. 2002; Windt et al. 2011). A distinction must be made between the mean and standard deviation DR (see Step 6)	$DR = \frac{\text{number of orders with } L_l \leq L \leq L_u}{\text{total number of orders}} \times 100$ $L = \text{Delivery date}_{real} - \text{Due date}_{pl}$ *DR: due date reliability (%) Pl: planned time* *Real: real process time L: lateness (SCD)* *L_l: lower limit of due date tolerance period* *L_u: upper limit of due date tolerance period*	(1) (2)
Step 2: Planned lead times are increased			
Trigger	As it is not possible to produce on time, production planners assume that the real lead time of a product is longer than the planned lead time[J]	$TL_{real} > TL_{pl}$ *TL: lead Time (SCD)*	(3)
Correlation	Lead time can be subdivided into interoperation time (e.g., waiting and transportation) and a value creating operation time (Bechte 1984)	$TL = TIO + TOP$ *TIO: interoperation time (SCD)* *TOP: operation time (SCD)*	(4)
Condition	The real operation time keeps unchanged as well as the maximum output rate[K]	$TOP = \text{Constant}$ (5) $ROUT_{max} = \text{Constant}$ *ROUT: output rate (hrs/SCD)*	(5) (6)
Target equation	Due to an increase of the planned lead time, only the mean planned interoperation time increases, as the planned operation time is constant	$(TL_{pl} + \Delta TL_{pl}) = (TIO_{pl} + \Delta TIO_{pl}) + TOP_{pl}$	(7)
Step 3: Orders are released earlier			
Trigger	A backward scheduling is used to schedule all orders. To define the latest possible start of production, the sum of all operation times has to be calculated. Thereby adherence to release due dates is necessary[L] (Schönsleben 2012)	$LD = \text{Due date} - TL_{pl}$ *LD: launch date (SCD)*	(8)
Target equation	With an increase of the lead time in *step 2*, the production launch is set to be earlier	$\text{Due date} - (TL_{pl} + TL_{pl}) = (LD - TL_{pl})$	(9)

(continued)

(continued)

Step 4: Work center loads are increased		
Trigger	The mean Work In Process is defined as the mean Range multiplied by the mean Output rate (Funnel formula) (Nyhuis and Wiendahl 2009)	$WIP_m = R_m \cdot ROUT_m$ (10) *m: mean value R: range (SCD)* *WIP: work in process (h)* *ROUT: output rate (h/SCD)*
Condition I	Under specified conditions, it is possible to identify the mean Range with the weighted Lead Time, which is calculated by the lead time multiplied by the work content of all operations[N] (Nyhuis and Wiendahl 2009; Yu 2001)	$R_m \cong TL_w$ with (11) $TL_w = \frac{\sum_{i=1}^{n}(TLi \cdot WC_i)}{\sum_{i=1}^{n} WC_i}$ (12) *n: number of operations* *WC: work content TL_w*: weighted lead time
Condition II	Since the work content depends on the Output rate and operation time [both are constant refer to (5) (6)], the work content stays constant	$TOP = \frac{WC}{ROUTmax}$ (13) $\rightarrow WC = constant$ (14)
Target equation	Inserting Eqs. (11) and (12) in Eq. (10), it can be shown that an increased lead time leads to an increased Work in Process with the amount of (17)	$WIP_m = TL_w \cdot ROUT_m$ (15) $= \frac{\sum_{i=1}^{n}((TLi + \Delta TLpl) \cdot WC_i)}{\sum_{i=1}^{n} WC_i} \cdot ROUT_m$ (16) $\rightarrow \Delta WIP_m = \frac{\sum_{i=1}^{n}(\Delta TL_i^{pl} \cdot WC_i)}{\sum_{i=1}^{n} WC_i} \cdot ROUT_m$ (17)
Step 5: Queues get longer		
Target equation	When the mean WIP increases, the mean Range increases automatically with the amount of (19)	$R_m = WIP_m / ROUT_m$ (18) $\rightarrow \Delta R_m = \frac{\sum_{i=1}^{n}(\Delta TL_i^{pl} \cdot WC_i)}{\sum_{i=1}^{n} WC_i}$ (19)
Step 6: Lead times get longer and more erratic		
Standard formula	The Eqs. (20) and (21) show how to calculate mean and standard deviation (Mukherjee and Kachwla 2009)	$TL_m = \frac{\sum_{i=1}^{n} TL_i}{n}$ (20) $TL_s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}[TL_i - TL_m]^2}$ (21) *s: standard deviation*

(continued)

(continued)

Target equation zero	With an earlier order release, the previous measured mean real lead time increases with TL_i^{pl} for each order i (with the above mentioned assumptions)	$\frac{\sum_{i=1}^{n}\left(TL_i+\Delta TL_i^{pl}\right)}{n} = \frac{\sum_{i=1}^{n} TL_i}{n} + \frac{\sum_{i=1}^{n} TL_i^{pl}}{n} = TL_m + \Delta TL_m^{pl}$	(22)
		$TL_s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left[(TL_i + \Delta TL_i^{pl}) - (TL_m + \Delta TL_m^{pl})\right]^2}$	(23)
Target equation I	**With constant** TL_i^{pl}: As calculated in (24) the standard deviation will not change if only a constant TL_i^{pl} is added	$TL_{m,new} = TL_{m,old}$	
		$\rightarrow TL_s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left[TL_i - TL_m\right]^2}$	(24)
Target equation II	**With variable** TL_i^{pl}:	$TL_{m,new} = TL_{m,old} + \frac{\sum_{i=1}^{n} TL_i^{pl}}{n}$	(25)
	If ΔTL_i^{pl} depends on the value of TL_i, the Eqs. (25) and (26) show that lead times not only get longer but also more erratic	$TL_s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left[(TL_i + \Delta TL_i^{pl}) - (TL_m + \Delta TL_m^{pl})\right]^2}$	(26)
Step 7: LTS loop			
Loop Target equation	As the mean lead time increases by the factor of step 2 and the standard deviation of the lead time increases, the due date reliability worsens	$TL_{m,new} > TL_{m,old}$	(27)
		$TL_{s,new} > TL_{s,old}$	(28)
		$DR = \frac{\text{number of orders with } L_l \leq L \leq L_u}{\text{total number of orders}} \cdot 100$	(29)

K *Fixed values* (*Step* 2): It is assumed that the utilized machines in the production processes keep unchanged (e.g., not expandable bottlenecks). Therefore, it is not possible to produce more than $ROUT_{max}$ per day and the real operation time stays unchanged, too. Indeed, in most production systems, it is possible to enable long or short term capacities such as working on weekends.

L *Scheduling technique* (*Step* 3): Mather and Plossl introduced the LTS in 1977 (Mather and Plossl 1977). At that time, simple scheduling techniques as forward or backward scheduling were the prevailing technologies to meet logistical targets. As the backward scheduling provides the latest possible time slot to produce in time, a lower inventory level is aspired (Kumar 2006). To meet due dates and reach a lower inventory, infinite capacity and correct lead times are assumed (Schönsleben 2012). In practice, both assumptions are inapplicable, which leads to a poor due date reliability and—combined with lead time adjustments—into the LTS. The use of backward scheduling in today's production planning will be discussed in the following chapter.

M *Adherence to release due dates* (*Step* 3): The forms of order release can be divided into order release after due date, immediate order release and work in process regulating order release (Lödding 2008). In the LTS, the first form is assumed and its topicality will be discussed in the following chapter. For example, if order release dates are scheduled, but due to the production system released immediately after the receipt of the order, an adaption of planned lead times will not provide a different order release. Therefore, the system would not respond.

N *Lead Time versus Range* (*Step* 4): To derive the interdependency of lead time and Work in Process, the relation between the range value and the lead time value has to be investigated. The result is the substitution of the mean range value in Eq. (15) through a term that includes lead time variables. Nyhuis and Wiendahl show that the equation $R_m \cong TL_{wa}$ applies, if the system is well established and only the long view is observed (Nyhuis and Wiendahl 2009). In a short term observation, it can be shown that the "mean values of the range always change earlier than the values of the throughput time" (Nyhuis and Wiendahl 2009).

P *Constant or variable lead time adjustments* (*Step* 6): A constant adjustment of planned lead times does not change the standard deviation of the mean lead times and therefore has no impact on due date reliability. In contrast to this, a variable adjustment of planned lead times affects the standard deviation of the mean lead times and results in a worsened due date reliability. In spite the fact that both assumptions increase the level of the mean lead time, it is important for future research to investigate the influence of Steps 2–6 on Step 7, to underline this uncoupled line of argumentation.

Relevance in Practice

Many companies try to save money in their production processes and want to improve their due date reliability. Unfortunately, they often do not have direct

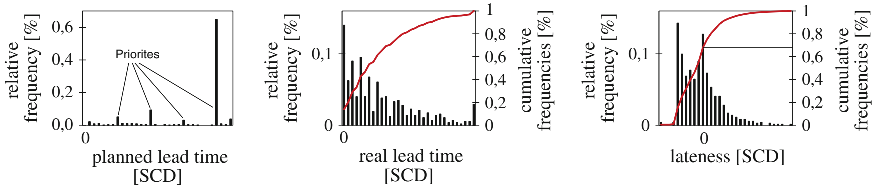


Fig. 3 Gap between supposed deadlines and due date reliability in practice

access to their work system specific operating curves and therefore are not able to position themselves (see Fig. 1) in order to meet their goals. In such scenarios, lead time adjustments can result in the LTS. This effect is even more likely if the LTS and its interactions or the logistical operating curves are unknown. As the main reactions of the production derived from an assumed backwards scheduling, it needs to be examined if the subsequent problems are still relevant today. Even under use of a more sophisticated scheduling technique, due date reliability is often unsatisfactory such as by the recurring adjustments of the planned lead time in FROS, including occasional deviations due to WIP fluctuations:

> Experience has shown that despite complicated scheduling methods the quality of the planned lead time is often unsatisfactory. This is because Production Planning and Control systems seldom offer appropriate supports for maintaining or determining the basis for planning (Nyhuis 2007).

The relevance of this problem in production can be shown in the following example from an industry project. The left chart in Fig. 3 maps the relative frequency of planned order lead times and shows that most orders are planned by means of their priority. In contrast, the middle chart maps the real lead times of orders and shows that despite the priority planning orders are produced as early as possible. As a result, the standard deviation of lateness is high which induces a very low due date reliability, even though the fact that the mean lateness is nearly zero and about 70 % of all orders are delivered in time (Fig. 3, right chart).

The relevance in production planning is given by the fact that ERP software like SAP is used for scheduling, which implies classical backward scheduling (Balla and Layer 2006). If the latest possible starting date is in the past, forward scheduling is used as well as a degressive scheduling that shortens interoperation times to produce in time (Dittrich et al. 2009). Several weaknesses can be listed, e.g., the necessity to calculate with mean values that might have high standard deviations or that it is possible to add a time buffer on lead times to deal with weaknesses in forecasts (Bauer 2009; Dittrich et al. 2009). If planners know about the LTS in a quantifiable manner and consider the assumptions made in the preceding chapters, different approaches are available that will be discussed in the following paper.

Countermeasures

If planners are familiar with the LTS, different planning techniques are already available to improve logistical performance (with regard to the logistical targets). The following list points out some examples:

- Load-oriented order release (Bechte 1984; Burkhalter 2010): The underlying backwards scheduling of orders only assigns priorities. Orders are released under consideration of production load limits and their priorities.
- Compensation of a lengthened lead time (i.e., of one work system) with the reduction of other lead times (i.e., up- or downstream the processes) in order to achieve a constant order lead time.
- Introduction of a customer decoupling point. Lead time adjustments prior to a decoupling buffer would not affect due dates after the buffer.
- Short term increase of capacities to compensate rising work load.

This abstract of ideas and techniques shows that several methods are available to counteract against the different LTS triggers. Therefore, it is not necessary to introduce a completely new method that avoids the LTS. Moreover, the coherences that lead to the LTS need to be investigated in future work to enable the derivation of individual adequate methods that fit the prevailing conditions, since some measures could be inapplicable such as an increase of capacities of not expendable systems. As there has been much research on each method already, a great chance is offered to investigate an optimal composition of measures that is able to work against the effects of the LTS. More research is needed in this field to derive the optimal method composition that fits the prevailing conditions.

Conclusion and Outlook

Even though the LTS was firstly observed in an industry project in 1977, this paper shows the topicality of the problem despite all modern technologies. Moreover, the introduced mathematical investigation enabled an inside view into the logistical effects and showed that many questions are still not answered. However, to reveal potential in the improvement of due date reliability, further research on this topic needs to be done.

Acknowledgments The research of Katja Windt is supported by the Alfried Krupp Prize for Young University Teachers of the Alfried Krupp von Bohlen und Halbach-Foundation.

References

Balla J, Layer F (2006) Produktionsplanung mit SAP APO-PP/DS: Prozesse und Customizing in der Praxis. Galileo Press, Germany
Bauer J (2009) Produktionscontrolling und -management mit SAP ERP: Effizientes Controlling, Logistik- und Kostenmanagement moderner Produktionssysteme. Vieweg+Teubner, Verlag
Bechte W (1984) Steuerung der Durchlaufzeit durch belastungsorientierte Auftragsfreigabe bei Werkstattfertigung, 70th edn. VDI, Verlag
Breithaupt J-W, Land M, Nyhuis P (2002) The workload control concept: theory and practical extensions of Load Oriented Order Release. Prod Plann Control 13:625–638
Burkhalter JP (2010) Steuerung- und Regelungsstrategien für Produktionstechnische Systeme. Dissertation. ETH, Zürich Nr. 18830
Dittrich J, Mertens P, Hau M, Hufgard A (2009) Dispositionsparameter in der Produktionsplanung mit SAP: Einstellhinweise, Wirkungen, Nebenwirkungen. Vieweg+Teubner, Verlag
Gutenberg E (1951) Grundlagen der Betriebswirtschaftslehre, 1st edn. Springer, Berlin
Hopp WJ, Spearman ML (2008) Factory physics, 3rd edn. McGraw-Hill/Irwin, New York
Kumar SA (2006) Production and operations management. new age international (P) ltd. Publishers
Lödding H (2008) Verfahren der Fertigungssteuerung, 2nd edn. Springer, Berlin
Mather H, Plossl G (1977) Priority fixation versus throughput planning. Production and inventory management. J Am Prod Inventory Control Soc, APICS 3rd Q.:27–51
Moscoso P, Fransoo J, Fischer D, Wäfler T (2011) The planning bullwhip: a complex dynamic phenomenon in hierarchical systems. In: Fransoo JC, Wäfler T, Wilson J (eds) Behavioral operations in planning and scheduling. Springer, Heidelberg, pp 159–186
Mukherjee PN, Kachwla TT (2009) Operations management and productivity techniques. Prentice-Hall of India, New Delhi
Nyhuis P (2007) Practical applications of logistic operating curves. CIRP Ann—Manufact Technol 56:483–486
Nyhuis P, Vogel M (2006) Entwicklung eines Modells zur Beschreibung der Termintreue von Fertigungsbereichen. PPS Manag 3:45–48
Nyhuis P, Wiendahl H-P (2009) Fundamentals of production logistics: theory, tools and applications. Springer, Berlin
Nyhuis P, Wiendahl H-P (2006) Logistic production operating curves: basic model of the theory of logistic operating curves. CIRP Ann: Manufact Technol 55:441–444
Nyhuis P, Wiendahl H-P (2010) Logistische Kennlinien: Grundlagen, Werkzeuge und Anwendungen, 3rd edn. Springer, Berlin
Schuh G (2006) Produktionsplanung und -steuerung: Grundlagen, Gestaltung und Konzepte. Springer, Berlin
Schönsleben P (2012) Integral logistics management: operations and supply chain management within and across companies, 4th edn. Auerbach Publications, Boca Raton
Selçuk B, Adan IJ, de Kok TG, Fransoo JC (2009) An explicit analysis of the lead time syndrome: stability condition and performance evaluation. Int J Prod Res 47:2507–2529
Selçuk B, Fransoo JC, De Kok AG, DeKok A (2006) The effect of updating lead times on the performance of hierarchical planning systems. Int J Prod Econ 104:427–440
Wiendahl H-P (1997) Fertigungsregelung: Logistische Beherrschung von Fertigungsabläufen auf Basis des Trichtermodells. Hanser Verlag, München
Wiendahl HP (2008) Betriebsorganisation für Ingenieure, 6th edn. Hanser Verlag, München
Windt K, Nyhuis P, Kolev I et al (2011) Improving due date reliability of steel mills: identification of punctuality potentials, innovation in product and production. In: 21st international conference on production research
Yu K-W (2001) Terminkennlinie: Eine Beschreibungsmethodik für die Terminabweichung im Produktionsbereich. Fortschritt-Berichte. VDI, Verlag

A Multi-Level Programming Model and Solution Algorithm for the Location of Distribution Centers with Uncertainty Demand

Patareeya Lasunon and Raknoi Akararungruangkul

Abstract The tri-level programming model is presented to seek the optimal location for distribution centers. The objective of top level model is to determine the optimal location by minimizing the total logistic cost which consists of the fixed operating cost to open distribution centers, the penalty cost and the transportation cost. The middle level model gives an equilibrium demand distribution center by maximizing the total customer demand (in %) that can be delivered on time and to maximize customer satisfaction in the bottom level model. The proposed algorithm is applied to solve this model to obtain the optimal solution. Finally, the tri-level programming model and its algorithm are demonstrated by a numerical example.

Keywords Multi-level programming · Optimization model · Location problem · Distribution network

Introduction

The supply chain network (SCN) design is an important problem. SCN usually involves with multiple and conflicting objectives such as logistic cost, service level, customer satisfaction, etc. In particular, the facility location and allocation problem is an interesting topic. There are many works in literature considering the

P. Lasunon (✉)
Department of Mathematics, Faculty of Science, Khon Kaen University,
Khon Kaen 40002, Thailand
e-mail: lpatar@hotmail.com

R. Akararungruangkul
Department of Industrial Engineering, Faculty of Engineer, Khon Kaen University,
Khon Kaen 40002, Thailand
e-mail: raxaka1@hotmail.com

H.-J. Kreowski et al. (eds.), *Dynamics in Logistics*, Lecture Notes in Logistics,
DOI: 10.1007/978-3-642-35966-8_27,

Table 1 Mathematical model for the supply chain problem

Years	Problem	Model
2003 (Sun et al. 2008)	Distribution network design	Bi-level programming
2007 (Lasunon et al. 2011)	Supply chain planning	Probabilistic bi-level programming
2007 (Afshari et al. 2010)	Distribution network design with uncertainty demand	Fuzzy programming
2008 (Zhang et al. 2010)	Decision making in supply chain management	Bi-level linear multi objective programming
2008 (Shun and Gao 2003)	Location problem	Bi-level programming
2010 (Roghanian et al. 2007)	Location within network distribution	Multi-objective mixed integer programming
2010 (Chen et al. 2007)	Decision making in supply chain management	Tri-level programming
2010 (Lasunon et al. 2011)	Location within supply chain planning under an uncertainty demand	Stochastic tri-level programming

concept of SCN on the location models, which have been devoted to the development of mathematical models as in Table 1.

In this paper, the location problem of distribution centers involves three kinds of decision-makers are plant managers, distribution managers and customers who have different objective functions. Therefore, the tri-level programming model is used to describe this problem.

Multi-Level Programming Model

The bi-level programming model (Lasunon et al. 2011), which is formulated for a problem in which two decision-makers make decisions successively has much more advantages than the single-level programming model. They are summarized as follows (Shun and Gao 2003).

1. The bi-level programming model can be used to analyze two different and even conflicting objectives at the same time in the decision-making process.
2. The multi-criteria decision-making methods of bi-level programming can reflect the practical problem better.

By extending the bi-level programming, Roghanian et al. (2007) defined the multi-level programming such as the tri-level programming where the second level is itself a bi-level programming. So the multi-level programming models are developed for the multi decision problem and explain the problem much more.

The general formulation of a tri-level programming problem can be stated as follows (Chen et al. 2007):

$$\begin{array}{ll} \min\limits_{x} & F_1(x,y,z) \\ \text{s.t.} & G_1(x,y,z) \le 0, \end{array}$$

Where $y = y(x)$ is implicitly defined by

$$\begin{array}{ll} \min\limits_{y} & F_2(x, y, z) \\ \text{s.t.} & G_2(x, y, z) \leq 0, \end{array}$$

Where $z = z(x, y)$ is implicitly defined by

$$\begin{array}{ll} \min\limits_{z} & F_3(x, y, z) \\ \text{s.t.} & G_3(x, y, z) \leq 0. \end{array}$$

The variables $x \in R^n$, $y \in R^m$, $z \in R^p$ are called the top-level, middle-level and bottom-level variables respectively. Similarly, $F_i : R^n \times R^m \times R^p \to R, \quad i = 1, 2, 3$ are called the top-level, middle-level and bottom-level functions respectively and the vector valued functions $G_i : R^n \times R^m \times R^p \to R, \quad i = 1, 2, 3$ are called the top-level, middle-level and bottom-level constraints respectively. All of the constraints and objective functions may be linear, non-linear, quadratic, fractional, etc.

Many methods have been proposed to solve the multi-level programming problems (Zhang et al. 2010), as shown in Table 2 below.

In this paper first we present a tri-level programming model for location and allocation in distribution network problems. The objectives are to minimize the total logistic cost including fixed operating cost, penalty cost and transportation cost, in order to maximize the total customer demand (in %) that can be delivered on time and maximize customer satisfaction. Lastly, we develop a tri-level algorithm to solve the proposed model. A numerical example further illustrates the application of this algorithm.

Table 2 Methods for multi-level programming problems

1. Extreme point search	K*th*—best algorithm
	Grid-search algorithm
	Fuzzy approach
	Interactive approach
2. Transformation approach	Complement pivot
	Branch and bound
	Penalty function
3. Descent and heuristic	Descent method
	Branch and bound
	Cutting plane
	Dynamic programming
4. Intelligent computation	Tabu search
	Simulated annealing
	Genetic algorithm
	Artificial neural network
5. Interior point	Primal–dual algorithm

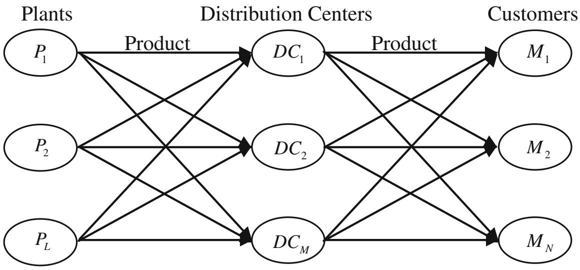


Fig. 1 Supply chains

Problem Statement

As in the Fig. 1, the three-layer distribution network consists of three different types of logistic nodes such as plants, distribution centers and customers.

A tri-level programming model of a distribution network based on several assumptions is introduced. The assumptions in the proposed model in this paper are summarized as follows.

Assumptions:

1. There are the multiple plants, distribution centers, customers and a single product.
2. Total capacity of plants and distribution centers are known in advance.
3. Customer demand is uncertain.
4. Transportation cost per unit is a coefficient of distance between facilities and facilities.
5. There is only on time delivered for customer satisfaction.
6. There is a minimum level of customer satisfaction.

Model Formulation

Variables

$$u_j = \begin{cases} 1, & \text{If distribution center } j \text{ is opened} \\ 0, & \text{Otherwise} \end{cases}$$

x_{ij} Percentage of demand distribution center j supplied by plant i

y_{jk} Percentage of demand customer k supplied by distribution center j

Parameters

L Sets of plants $(i \in L)$
M Sets of distribution centers $(j \in M)$
N Sets of customers $(k \in N)$
a_i Production capacities of plant i
b_j Capacities of distribution center j
c_k Demand of customer k
d_{ij} Distance between plant i and distribution center j
$\tilde{d}_{jk}$ Distance between distribution center j and customer k
s_k Minimum level of customer satisfaction k
f Coefficient of total cost in objective function
t Unit transport cost
o_j Cost of operating distribution center j
p_k Cost of penalty for customer k

Tri-Level Programming Model

The proposed model in this paper can be formulated as follows:

The top level ($L1$):

$$\min_{u_j} \mathrm{F}_1 = f \cdot \left(\begin{array}{l} \sum_{i=1}^{L} j \sum_{j=1}^{M} t \cdot d_{ij} \cdot b_j \cdot x_{ij} + \sum_{j=1}^{M} \sum_{k=1}^{M} t \cdot \tilde{d}_{jk} \cdot c_k \cdot y_{jk} + \\ \sum_{j=1}^{M} o_j \cdot u_j + \sum_{k=1}^{N} p_k \cdot c_k \left(1 - \sum_{j=1}^{M} y_{jk} \right) \end{array} \right)$$

Subject to:

$$\sum_{j=1}^{M} u_j \geq 1 \tag{1}$$

$$u_j \in \{0,1\}, \quad \forall j \tag{2}$$

The middle level ($L2$):

$$\min_{x_{ij}} \mathrm{F}_2 = \frac{\sum_{i=1}^{L} \sum_{j=1}^{M} b_j \cdot x_{ij}}{\sum_{j=1}^{M} \sum_{k=1}^{N} c_k \cdot y_{jk}} + \frac{\sum_{j=1}^{M} \sum_{k=1}^{N} c_k \cdot y_{jk}}{\sum_{k=1}^{N} c_k}$$

Subject to:

$$\sum_{i=1}^{L} b_j \cdot x_{ij} \leq \sum_{k=1}^{n} c_k \cdot y_{jk}, \ \forall j \tag{3}$$

$$\sum_{j=1}^{M} c_k \cdot y_{jk} \leq c_k, \ \forall k \tag{4}$$

$$\sum_{j=1}^{M} b_j \cdot x_{ij} \leq a_i, \ \forall i \tag{5}$$

$$\sum_{k=1}^{N} c_k \cdot y_{jk} \leq b_j \cdot u_j, \ \forall j \tag{6}$$

$$\sum_{j=1}^{M} x_{ij} \leq m, \ \forall i \tag{7}$$

The bottom level (L3):

$$\min_{y_{jk}} \mathrm{F}_3 = \frac{(1-f)}{n} \cdot \sum_{j=1}^{M} \sum_{k=1}^{N} y_{jk}$$

Subject to:

$$\sum_{k=1}^{N} y_{jk} \leq N \cdot u_j, \quad \forall j \tag{8}$$

$$\sum_{j=1}^{M} y_{jk} \geq s_k, \quad \forall k \tag{9}$$

All variable ≥ 0

Objectives:

- F1 is the summation of total logistic cost (transportation cost, fixed operating cost and penalty cost), that is multiplied by weighted coefficient *f*.
- F2 is the summation of total customer demand (in %) that can be delivered on time.
- F3 is the summation of the level of the customer satisfaction that is multiplied $(1-f)$.

Constraints:

1. Ensures that at least one distribution center is opened.
2. Represents the binary restrictions of the decision variable.

3. Indicate that all transportation quantity from plants to each distribution center is enough for all transportation quantity from each distribution center to customer.
4. Ensures that the demand of each customer must be met by supply from some distribution center.
5. Shows that the capacity constrain for each plant.
6. Shows that the capacity constrain for each distribution center.
7. States if plant i satisfies the demand.
8. States if distribution center j satisfies the demand.
9. Implies that there is a minimum level of customer satisfaction.

Tri-Level Algorithm

Based on the tri-level algorithm (Chen et al. 2007; Aryanezhad and Roghanian 2008), the algorithm for the tri-level programming model is summarized as follows:

Step 1: Set $i = 1$, Solve the top level with the simplex method to get a solution (u_1, x_1, y_1).
Step 2: Let $(u^*, x^*, y^*) = (u_i, x_i, y_i)$.

Fix $u = u^*$ in the middle level and solve with bounded simplex method to get a solution $(\hat{x}_1, \hat{y}_1)$.

Step 3: Use bi-level algorithm to solve the middle and bottom level.
3.1 Set $j = 1$ and $(\hat{x}, \hat{y}) = (\hat{x}_j, \hat{y}_j)$.
3.2 Fix $x = \hat{x}$ in the bottom level and solve with bounded simplex method to get a solution $\bar{y}$.

- If $\hat{y} = \bar{y}$ then the optimal solution of the middle and bottom level is $(\hat{x}, \hat{y})$, then go to step 4.
- If $\hat{y} \neq \bar{y}$ then there exist $k \leq n$, where n is the number of total constrains such that $A_k u + B_k x + C_k \hat{y} = b_k$.

Set $T1 = \{(u, x, y) : A_k u + B_k x + C_k y = b_k,\ k = 1, \ldots, n\}$
3.3 Set $j = j + 1$ and solve the middle level on $T1$to get a solution $(\hat{x}_j, \hat{y}_j)$.

- If $(\hat{x}, \hat{y}) = (\hat{x}_j, \hat{y}_j)$ then the optimal solution of the middle and bottom level is $(\hat{x}, \hat{y})$, then go to step 4.
- If $(\hat{x}, \hat{y}) \neq (\hat{x}_j, \hat{y}_j)$ then go to step 3.2

Step 4: If $(\hat{x}, \hat{y}) = (x^*, y^*)$ then the optimal solution of the tri-level is (u^*, x^*, y^*), Stop. Else if $(\hat{x}, \hat{y}) \neq (\hat{x}_j, \hat{y}_j)$ then there exist $k \leq n$ such that $A_k u + B_k \hat{x} + C_k \hat{y} = b_k$. Set $T2 = \{(u, x, y) : A_k u + B_k x + C_k y = b_k,\ k = 1, \ldots, n\}$

Step 5: Set $i = i + 1$ and solve the top level on $T2$to get a solution (u_i, x_i, y_i).

- If $(u^*, x^*, y^*) = (u_i, x_i, y_i)$ then the optimal solution of the tri-level is (u^*, x^*, y^*), Stop.
- If $(u^*, x^*, y^*) = (u_i, x_i, y_i)$ then go to step 2.

Stop.

Numerical Example

In this section, a simple numerical example is presented to illustrate the applications of the tri-level programming model and its algorithm proposed for the optimal location problem. To apply the method, we refer to an example in (7). It is assumed that there are three plants; P_1, P_2, P_3, three distribution centers; DC_1, DC_2, DC_3 and three customers; M_1, M_2, M_3.

The parameters of customer demand, capacity of plants, capacity of distribution centers, customer satisfaction, the overall cost which corresponds to the total cost which consists of the fixed operating cost, the penalty cost and the transportation cost, the distance between the plants and the distribution centers and the distance from the distribution centers to the customers are shown in the Tables 3, 4, 5 and 6.

The algorithm in section Tri-Level Algorithm is coded by using Lingo-12 to solve the above problem, where the customer satisfaction value is 0.6 and the coefficient of total cost $f = 0.5$. The final solution is $u_1 = 0$, $u_2 = 1$, $u_3 = 1$ and the objective values are $F_1 = 1293.525$, $F_2 = 1.886792$, $F_3 = 0.45$. It means that the first distribution center should not be opened, others should be opened. Moreover, the final result of modeling the mentioned case study with $f = 0.5$ with different customer satisfaction values can be shown in the following (Table 7).

Figure 2 demonstrates the total cost with different customer satisfaction values and different coefficients of total cost in objective function.

Figure 2 illustrates that the total cost vary directly with the customer satisfaction level for every f. Therefore, a high level of customer satisfaction requires a high product delivery in order to meet customer requirements resulting in high cost.

Table 3 Maximum capacities of plants, capacities of distribution centers and demands of customers

Maximum capacity	1	2	3
Plants (unit)	300	200	200
Distribution centers (unit)	250	170	280
Customers (unit)	150	180	200

Table 4 Customer satisfaction, penalty cost, opening and operating cost

	1	2	3
Minimum level of customer satisfaction	0.6	0.6	0.6
Penalty cost (per 1 unit)	4.01	3.26	4.29
Opening and operating cost (per times)	414	490	360

Table 5 Distance from plants to distribution centers

From	To		
	Center 1	Center 2	Center 3
Plant 1	256	102	222
Plant 2	173	223	236
Plant 3	234	112	289
Unit transport cost from plant to distribution center is 0.01			

Table 6 Distance from distribution centers to customers

From	To		
	Customer 1	Customer 2	Customer 3
Center 1	271	287	318
Center 2	98	344	291
Center 3	302	104	228
Unit transport cost from distribution center to customer is 0.01			

Table 7 The result of model in different customer satisfaction

Minimum level of customer satisfaction	0.2	0.4	0.6	0.8	1.0
Total cost	1104.23	1159.21	1293.52	1329.79	1422.18
u (1,2,3)	(0,1,0)	(0,1,0)	(0,1,1)	(0,1,1)	(0,1,1)

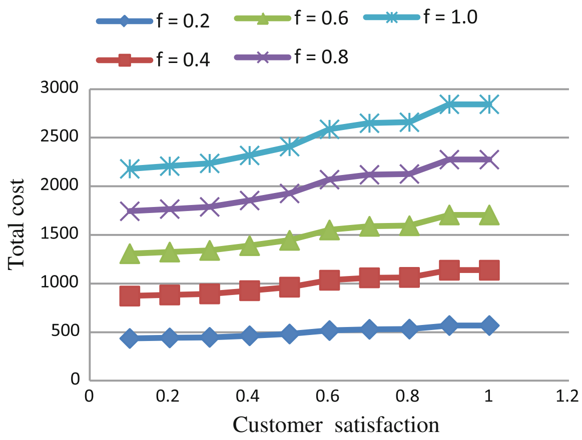


Fig. 2 Total cost versus customer satisfaction at f = 0.1–1

Conclusions

In this paper, a tri-level programming model is proposed to describe the location problem with the uncertainty of customer demand. Its algorithm is developed to solve the problem and implemented in Lingo-12. This model is applied for determining the optimal number and location of distribution centers by minimizing the total logistic cost, maximizing the total customer demand (in %) and maximizing customer satisfaction. Finally, a numerical example shows that the multi-level programming model, which shows its algorithm is feasible and advantageous. It can be used for determining the optimal location of distribution centers uncertainty demand of customers and to maximize customer satisfaction as well as to minimize total logistic cost.

Acknowledgments This research is supported by the Center of Excellence in Mathematics (CEM), Faculty of Science at Mahidol University, Phayathai Campus, Ratchathewi, Bangkok 10400.

References

Afshari H, Amin-Nayeri M, Jaafari A (2010) A multi-objective approach for multi-commodity location within distribution network design problem. In: Proceeding of the international conference of engineers and computer scientists, Hong Kong, 17–19 Mar 2010

Aryanezhad MB, Roghanian E (2008) A bi-level linear multi-objective decision making model with interval coefficients for supply chain coordination. Int Eng Spec Issue 19(1–2):6–74

Chen Z, Zhang D, Li S (2007) An optimization model and its solution algorithm for distribution network design problem with uncertainty demand. In: Proceeding of the IEEE international conference on automation and logistics, Jinan, China, pp 2170–2175, 18–21 Aug 2007

Lasunon P, Remsungnen T, Akararungruangkul R (2011) A stochastic tri-level programming model to minimize total cost in a supply chain planning with uncertainty demand. In: Proceeding of 2nd international conference on logistics and transport, Queenstown, New Zealand, pp 693–701, 16–18 Dec 2011

Lasunon P, Wetweerapong J, Remsungnen T (2011) A new algorithm for solving bi-level linear programming problems. In: Proceeding of 16th annual meeting in mathematics, Khon Kaen, Thailand, 10–11 Mar 2011

Roghanian E, Sadjadi SJ, Aryanezhad MB (2007) A probabilistic bi-level linear multi-objective programming problem to supply chain planning. Appl Math Comput 188:786–800

Shun HJ, Gao ZY (2003) Bi-level programming optimization model for distribution systems in supply chain. J Manage Sci Chin 6(2):66–70

Sun H, Gao Z, Wu J (2008) A bi-level programming model and solution algorithm for the location of logistics distribution centers. Appl Math Model 32:610–616

Zhang G, Jie L, Montero J, Zen Y (2010) Model, solution concept, and K^{th}-best algorithm for linear tri-level programming. Inf Sci 180:418–492

A Method to Estimate the Accumulated Delivery Time Uncertainty in Supply Networks

Mehdi Safaei, Safir Issa, Marcus Seifert, Klaus-Dieter Thoben and Walter Lang

Abstract Today, value creation within the manufacturing industry is realised in supply networks. A supply network is considered as a collaboration of suppliers and a manufacture (Original Equipment Manufacturer-OEM) to reach a common objective. Firms can now decide which parts of the supply network they will manage directly and which they will be out-source activities. One of the supply network's features is outsourcing, and it makes a new source of uncertainty. This uncertainty has an impact on supply network performance. Nowadays, researchers have recognised the importance of analysing the delivery time uncertainty in supply networks. Delivery time uncertainty in the OEM, results from the accumulation of individual uncertainties of the suppliers in the network. The supply networks consist of the combination of basic types of network such as linear, star, tree, etc. The types of network have an essential impact on the accumulation of the individual uncertainties. In this paper, a method to calculate the accumulation of the individual uncertainties based on the network type is presented. Finally, an example is given to show the application of this method to **the** all types of supply network with four nodes.

Keywords Supply network • Delivery time uncertainty • Network type • Probability density function

M. Safaei (✉) · S. Issa · K.-D. Thoben · W. Lang
International Graduate School (IGS), University of Bremen, Bremen, Germany
e-mail: saf@biba.uni-bremen.de

M. Safaei · M. Seifert · K.-D. Thoben
BIBA—Bremer Institut für Produktion und Logistik GmbH, Bremen, Germany

S. Issa · W. Lang
Institute for Microsensors—Actuators and—Systems (IMSAS), University of Bremen, Bremen, Germany

H.-J. Kreowski et al. (eds.), *Dynamics in Logistics*, Lecture Notes in Logistics,
DOI: 10.1007/978-3-642-35966-8_28, © Springer-Verlag Berlin Heidelberg 2013

Introduction

Today's business has rapidly changed and has become more competitive. Organisations more and more realise the effective role of supply networks to compete in the global market and networked economy. Value creation within the manufacturing industry is recognised in the supply networks. Quantities, delivery times, due dates, start times, etc., in the network may change at any time. Hence, the supply network is a dynamic system where these quantities are changing continually. Consequently, supply network systems must be updated accordingly so that decisions are based on dynamic information (Barker 2004).

Nowadays, a firm in the supply network can outsource different functions and signifies different degree of commitment and integration between the company and the contractor. Outsourcing in the supply network makes a new source of uncertainty in delivery time and other quantities and qualities factors. This uncertainty makes the impact on supply network performance by affecting delivery time reliability (Vanany et al. 2009). The importance of delivery time as a strategic weapon has been recognised in the arena of global competition (Lederer 1997; Christopher 2000). The strategic importance of delivery time uncertainty has been recognised (introduced) by many researchers and practitioners, and it has emerged as a key competitive factor in a supply network. Thus, many manufacturers are adopting the use of delivery-time guarantees as part of their market positioning strategy (Urban 2009).

The type of supply network has an essential impact on the accumulation of the individual uncertainties. Therefore, it influences on the overall uncertainty, which the network can guarantee for the final customer (Safaei et al. 2010). Whereas delivery time uncertainty has an impact on supply network performance, study on the impact of the network type on delivery time uncertainty in the supply network is a gap in this area. In this paper, an adapted method to calculate the accumulation of the individual **uncertainties** is introduced and the performance of this method by a numerical example on the supply network with four nodes has been studied. One of the most important of the advantage of this methodology is easy to use in complex networks. Moreover, this method can provide more details such as identification those parts of network, which has the highest potential for improving the total delivery time uncertainty, etc.

This study is part of the ongoing research. The methodology of this paper is based on the Guide to the Expression of Uncertainty in Measurement (GUM), and a literature review on supply networks, network typologies, probability theory, graph theory and Markov properties. Difference effects of delivery time uncertainty of a supplier on the final delivery time according to the network type has been studied by proposing a methodology based on the GUM and probability theory, graph theory. The contribution in this paper is the adaptation of the uncertainty estimation in measurement fields to the logistic field through the supply networks.

In this paper, an identification of the network type impact on the accumulation of the individual delivery time uncertainties for the network with four nodes is presented. Different network types, which can be generated by four nodes, are considered. Section Delivery Time Uncertainty introduces the methodology. In section Method Description, the used method is described and in section Numerical Example and Results, the capability of the method by a numerical example in different possible network types with four nodes is examined. Finally, in section Conclusion, the conclusion of this work is stated.

Delivery Time Uncertainty

The uncertainty of the delivery time in the supply network generally consists of several components, which can be grouped into two categories according to the way in which their numerical value is estimated (JCGM 2008):

Type A: those, which are evaluated by statistical methods (e.g., evaluating by individual observation).
Type B: those, which are evaluated by other means (e.g., evaluating by certificate or from experience).

The purpose of the Type A and Type B classification is to indicate the two different ways of evaluating uncertainty components and is for convenience of discussion only; the classification is not meant to indicate that there is any difference in the nature of the components resulting from the two types of evaluation. Both types of evaluation are based on probability distributions, and the uncertainty components resulting from either types are quantified by variances or standard deviations.

The estimated variance u^2, characterizing a delivery time uncertainty component obtained from a type "A" evaluation by calculating from series of repeated observations of delivery time and the estimate standard deviation u is the positive square root of u^2 (the variance) and for convenience, sometimes called the type "A" delivery time uncertainty. Therefore, a type "A" delivery time uncertainty, is obtained of a probability density function derived from an observed frequency distribution, while a type "B" delivery time uncertainty is obtained from an assumed probability density function based on the degree of belief that an event will occur.

For estimating the delivery time uncertainty of the entire supply network, the impact of individual supplier uncertainties on the overall uncertainty must be understood. The way how the individual uncertainties need to be accumulated depends on the network type (Guiffrida and Jaber 2008). The type of the supply network directly affects the final accumulation of the uncertainty of the delivery time and in this paper, discussed on the networks with four nodes (three suppliers and an OEM).

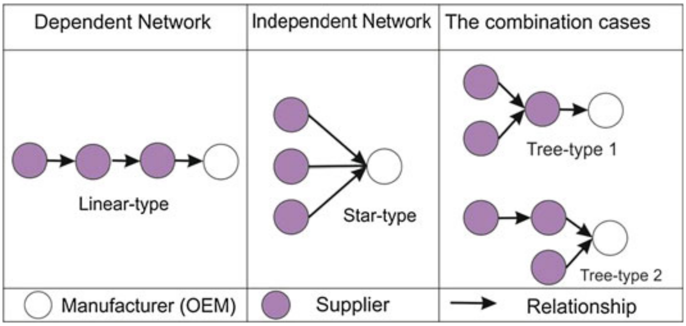


Fig. 1 Possible types of supply network (for manufacturing industries) with four nodes

Figure 1 shows all possible types of supply networks with four nodes. These four types of network could be characteristic in three categories (Independent, Dependent and Combination of independent and dependent networks). If uncertainty in delivery time of a supplier has an effect on another supplier, these two suppliers are dependent (e.g., Linear-type). If the delivery time uncertainty of both suppliers influences on the OEM separately, these two suppliers are independent (e.g., Star-type). Moreover, there are other types of network, which obtained from the combination of the independent and dependent parts. In these types of networks, in some sections of network, suppliers have a dependency and in other parts, they have not (Tree-type 1, 2). In consequence, the model of combination for these types of network will be different.

In the linear-type, the supplier's interaction pattern mainly follows as a chain. Maximum in and out degree for any node (firm) in the network is one and multiple tiers exist, but every firm has exactly one supplier below it (Pathak et al. 2007). In the Star-Type, all suppliers interact with one central node (OEM). Maximum depth of it is one, and it has included of single tier where all suppliers go to a single central firm. The Tree-Type, which is the most prevalent and assumed ''normal'' structure for a supply network, is much like the star topology, except that it doesn't use a central node, and it is a kind of directed acyclic graph (DAG) (Thoben and Jagdev 2001; Pathak et al. 2007). There are two possible kinds of Tree-Type networks.

The confidence interval introduces the timeframe, which a supply network can deliver a certain percentage of orders. The confidence coefficient represents the percentage of orders, which are delivered within the confidence interval. In consequence, the delivery time uncertainty can be defined as the combination of the confidence interval with the related confidence coefficient. Both indexes can be achieved of PDF[1] for each supplier. The time-related objective of planning and controlling strategies in supply networks is to reach a low level of delivery time uncertainty of the entire network—in other words, to reach a low confidence interval in combination with a high confidence coefficient (Safaei et al. 2011).

[1] Probability Density Function.

Method Description

Based on the (JCGM 2008) (Guide to the expression of uncertainty in measurement) the method is explained to calculate the delivery time uncertainty in the supply network with four nodes.

If there are n variables, $x_1, x_2, \ldots, x_n$, which represent delivery time of each suppliers, the relationship function between these variables is expressed as

$$Y = f(x_1, x_2, \ldots, x_n) \tag{1}$$

The basic combined uncertainty formula, which is also called the law of propagation of uncertainty, is (JCGM 2008):

$$u_c^2(Y) = \sum_{i=1}^{n}\sum_{j=1}^{n}\left(\left(\frac{\partial f}{\partial x_i}\right) \times \left(\frac{\partial f}{\partial x_j}\right) \times u\left(x_i, x_j\right)\right) \tag{2}$$

Where u_c, is the combined uncertainty for all parameters (suppliers in our case); $\frac{\partial f}{\partial x_i}$, is the sensitivity coefficient for the variable x_i, which is the partial derivative of the function f par rapport x_i; $u\left(x_i, x_j\right) = u\left(x_j, x_i\right)$ is the estimated covariance associated with x_i and x_j. The degree of correlation between x_i and x_j is characterized by the estimated correlation coefficient (JCGM 2008).

$$r\left(x_j, x_i\right) = \frac{u\left(x_j, x_i\right)}{u(x_i) \times u(x_j)} \tag{3}$$

Where $r\left(x_j, x_i\right) = r\left(x_i, x_j\right)$ and $-1 \leq r\left(x_j, x_i\right) \leq 1$. If the two variables are independent, then $r\left(x_i, x_j\right) = 0$, else have to estimate the correlation coefficient when they are dependent.

Equation 2, can be written then as follow,

$$u_c^2(Y) = \sum_{i=1}^{n}\left(\frac{\partial f}{\partial x_i}\right)^2 \times u(x_i)^2 + 2 \times \sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\left(\frac{\partial f}{\partial x_i}\right) \times \left(\frac{\partial f}{\partial x_j}\right) \times u(x_i) \times u(x_j) \times r\left(x_j, x_i\right) \tag{4}$$

For the example of three suppliers (s_1, s_2, s_3), formula number 4 become:

$$\begin{aligned} u_c^2(Y) = {} & \left(\frac{\partial f}{\partial s_1}\right)^2 u(s_1)^2 + \left(\frac{\partial f}{\partial s_2}\right)^2 u(s_2)^2 + \left(\frac{\partial f}{\partial s_3}\right)^2 u(s_3)^2 \\ & + 2\left(\frac{\partial f}{\partial s_1}\right)\left(\frac{\partial f}{\partial s_2}\right) u(s_1)u(s_2)r(s_1, s_2) \\ & + 2\left(\frac{\partial f}{\partial s_1}\right)\left(\frac{\partial f}{\partial s_3}\right) u(s_1)u(s_3)r(s_1, s_3) \\ & + 2\left(\frac{\partial f}{\partial s_2}\right)\left(\frac{\partial f}{\partial s_3}\right) u(s_2)u(s_3)r(s_2, s_3) \end{aligned} \tag{5}$$

Numerical Example and Results

In order to better understanding the performance of the method, a numerical example is given to show its application in the supply network with four nodes (three suppliers and an OEM). The aim of this example is to clarify the impact of the network types on accumulation of delivery time uncertainties. In this example, each supplier has specific probability density function (PDF) for its delivery time, which is obtained from series of repeated observations. Thus, there are three different PDFs. Our assumption in this example is that the delivery time PDF of each supplier follows the normal PDF.

DT is delivery time and DT_i, is the delivery time of the *i*th supplier. S_1, S_2 and S_3 are our suppliers and M, is the index of manufacturer (OEM). DT_1, DT_2 and DT_3, follow normal PDF with means μ_1, μ_2, μ_3 and standard deviations σ_1, σ_2, σ_3 respectively. The normal PDF is usually denoted by $N \approx (\mu, \sigma^2)$. In this example confidence coefficients of suppliers 1, 2 and 3 are shown by α, β and γ %. The confidence interval of supplier 1, 2 and 3 are shown by $\mu_1 \pm \sigma_1, \mu_2 \pm \sigma_2$ and $\mu_3 \pm \sigma_3$.

Delivery time of supplier 1 is elucidated by DT_1 and its mean is shown by μ_1, and its standard deviation is exposed by σ_1 then they indicated them together with this format: $DT_1 \sim N\left(\mu_1 = 15, \sigma_1^2 = 2\right)$

And similarly for suppliers 2 and 3:

$$DT_2 \sim N\ (\mu_2 = 17, \sigma_2^2 = 4),\ DT_3 \sim N\ (\mu_3 = 5, \sigma_3^2 = 1),$$

These are the indexes of the delivery time of each supplier and all the numbers of the means and standard deviations in this example are based on assumption.

The normal PDF is expressed as follows:

$$g(x) = \frac{1}{\sigma \times \sqrt{2\pi}} \times \left(e^{\frac{-1}{2} \times \left(\frac{x-\mu}{\sigma}\right)^2}\right) \tag{6}$$

x is the input variable which follows as a normal probability density function and in this paper represent the delivery time in days.

Based on the features of each supplier, PDFs diagram of delivery time uncertainty for each of the three suppliers are shown in Fig. 2.

A simple case the relationship function between DT_1, DT_2 and DT_3 is given by:

$$Y = DT_1 + DT_2 + DT_3 \tag{7}$$

First, in the independent case (Star-type), there is no correlation coefficient between each two suppliers and then $r(DT_1, DT_2) = 0$, $r(DT_1, DT_3) = 0, r(DT_2, DT_3) = 0$ (Fig 3).

Equation 2 becomes:

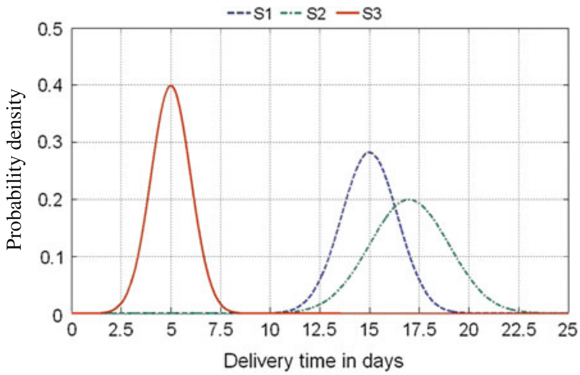


Fig. 2 PDFs diagrams of delivery time uncertainty of suppliers

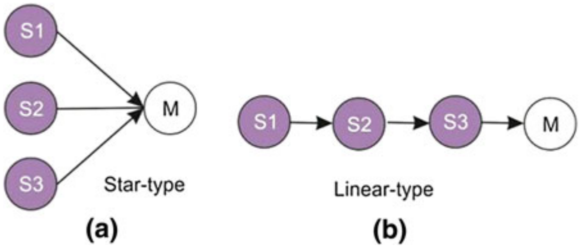


Fig. 3 **a** Star-type of supply network with three nodes. **b** Linear-type

$$\begin{aligned} u_c^2(Y) = {} & u_{DT1}^2 + u_{DT2}^2 + u_{DT3}^2 + 2 \times r(DT_1, DT_2) \times u_{DT1} \times u_{DT2} + 2 \\ & \times r(DT_1, DT_3) \times u_{DT1} \times u_{DT3} + 2 \times r(DT_2, DT_3) \times u_{DT2} \times u_{DT3} \\ = {} & 7\,\text{Days}. \end{aligned}$$

The standard uncertainty is $u_c(Y) = \sqrt{7}$. Because of the independency between all of suppliers, the manufacture will decide base on the maximum deliver time as a mean of combined delivery time $\mu(Y) = \max(\mu_1, \mu_2, \mu_3) = 17$ days.

Delivery time (Star-type) $= \mu(Y) \pm k \times u_c(Y) = 17 \pm \sqrt{7}$ Days ($k = 1$, with 68.2 % of confidence) $\approx$ [14.3, 19.6]

k is the standard deviation coefficient which called it "uncertainty coefficient". The pervious result are given for the 68.2 % of level of confidence. This means that there is 68.2 % sure of the result. It is suggested that the result of uncertainty to be given with 95.4 % of level of confidence i.e., the coverage factor $k = 2$ (for normal distribution), then: μ_{DT} = Mean of Delivery Time, σ_{DT} = Standard Deviation of Mean for Delivery Time. In consequence, the timeframe of delivery time with approximately 4.6 % of the uncertainty (95.4 % of confidence), is obtained when, $k = 2$. Then, delivery time in the independent case (Star-type) $= 17 \pm 2 \times \sqrt{7}$ Days ($k = 2$, with 95.4 % of confidence) $\approx$ [11.7, 22.3], Second, in the dependent case, the high degree of dependency between suppliers if there is a connection is assumed and the correlation coefficient between supplier 1 and 2 is considered maximum, then $r(DT_1, DT_2) = 1$, and it is the same for the correlation coefficient between supplier 2 and 3, $r(DT_2, DT_3) = 1$, but because there is no direct relation between supplier 1 and 3 the correlation coefficient between them assumed zero, $r(DT_1, DT_3) = 0$.

Equation 2 becomes

$$\begin{aligned} u_c^2(Y) = {} & u_{DT1}^2 + u_{DT2}^2 + u_{DT3}^2 + 2 \times r(DT_1, DT_2) \times u_{DT1} \times u_{DT2} + 2 \\ & \times r(DT_1, DT_3) \times u_{DT1} \times u_{DT3} + 2 \times r(DT_2, DT_3) \times u_{DT2} \times u_{DT3} \\ = {} & \sigma_1^2 + \sigma_2^2 + \sigma_3^2 + 2 \times \sigma_1 \times \sigma_2 + 2 \times \sigma_2 \times \sigma_3 = 16.7\,\text{Days} \end{aligned}$$

Then, $u_c(Y) = \sqrt{16.7}$. Because of the dependency between suppliers in this type of network, the manufacturer will decide base on the summation of deliver time as a combined delivery time of suppliers and the estimation of delivery time is $\mu(Y) = \mu_1 + \mu_2 + \mu_3 = 37$ days.

$$\begin{aligned} \text{Delivery time} & = 37 \pm \sqrt{16.7}\ \text{Days}\ (K = 1) \approx [32.9, 41.1] \\ & = 37 \pm 2 \times \sqrt{16.7}\ \text{Days}\ (K = 2) \approx [28.8, 45.2] \end{aligned}$$

Third, The combination of dependent and independent networks, There are two possible tree-types of network with four nodes. In the first one, in this example, both supplier 1 and supplier 2 are before the supplier 3, and they work in the star-type of network with each other and there is just one supplier (supplier 3) before the manufacturer. In this case, by assuming that there is no correlation between supplier 1 and 2 ($r(DT_1, DT_2) = 0$), maximum correlation coefficient, between the supplier 3 and the accumulation of supplier 1 and 2 is considered, then $r(DT_1, DT_3) = 1$ and $r(DT_2, DT_3) = 1$ (Fig. 4).

In this case, Eq. 2 is becomes:

$$\begin{aligned} u_c^2(Y) = {} & u_{DT1}^2 + u_{DT2}^2 + u_{DT3}^2 + 2 \times r(DT_1, DT_2) \times u_{DT1} \times u_{DT2} \\ & + 2 \times r(DT_1, DT_3) \times u_{DT1} \times u_{DT3} + 2 \times r(DT_2, DT_3) \times u_{DT2} \times u_{DT3} \\ = {} & 13.8\,\text{Days} \end{aligned}$$

This type of network consists of two parts, in the first one there is independency between supplier 1 and 2 where the delivery time uncertainty in supplier 3 has the dependency to the accumulation of the individual delivery time uncertainty of suppliers 1 and 2 in the second part. Then, the standard uncertainty is $u_c(Y) =$

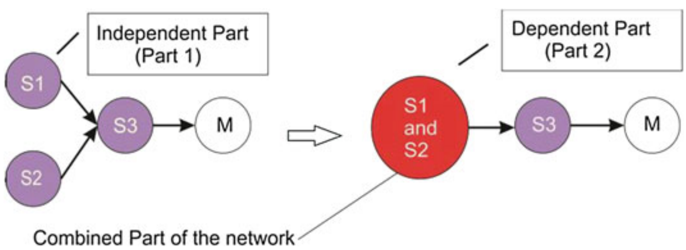


Fig. 4 Tree-type (1) of supply network within four nodes

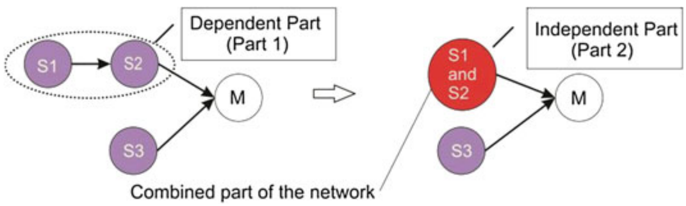


Fig. 5 Tree-type (2) of supply network within four nodes

$\sqrt{13.8}$ and the estimate for the delivery time in the independent part is obtained from the maximum of the delivery times and in the dependent part is obtained from the summation of delivery times. Thus, the delivery time of this network is $\mu(Y) = \max(\mu_1, \mu_2) + \mu_3 = 22$ days.

$$\begin{aligned}\text{Delivery time }(1) &= 22 \pm \sqrt{13.8}\text{ Days }(K = 1) \approx [18.3, 25.7]\\ &= 22 \pm 2 \times \sqrt{13.8}\text{ Days }(K = 2) \approx [14.6, 29.4]\end{aligned}$$

Second form of Tree-Type is shown in the Fig. 5,

In this case, the network consists of two different parts (at first dependent part and then independent part). In the second form of the tree-type, there is dependency between supplier 1 and supplier 2 ($r(DT_1, DT_2) = 1$), and their accumulation of delivery time uncertainty has impact on manufacture directly. After the combination of the supplier 1 and supplier 2, the network becomes similar to the star-type of network with three nodes. Both of suppliers 1 and 2 have no dependency to supplier 3, then $r(DT_1, DT_3) = 0$ and $r(DT_2, DT_3) = 0$.

Equation 2 for the Tree-Type (2) is then,

$$\begin{aligned}u_c^2(Y) = {} & u_{DT1}^2 + u_{DT2}^2 + u_{DT3}^2 + 2 \times r(DT_1, DT_2) \times u_{DT1} \times u_{DT2} + 2\\ & \times r(DT_1, DT_3) \times u_{DT1} \times u_{DT3}\\ & + 2 \times r(DT_2, DT_3) \times u_{DT2} \times u_{DT3} = 12.7\,\text{Days}\end{aligned}$$

In that case, $u_c(Y) = \sqrt{12.7}$. Based on the dependent and independent part in this network, estimation of the delivery time is obtained of the summation of delivery times in the dependent parts and maximum of the delivery times in the independent parts. Therefore, $\mu(Y) = \max((\mu_1 + \mu_2), \mu_3) = 32$.

$$\begin{aligned}\text{Delivery time }(2) &= 32 \pm \sqrt{12.7}\text{ Days }(K = 1) \approx [28.4, 35.6]\\ &= 32 \pm 2 \times \sqrt{12.7}\,\text{Days }(K = 2) \approx [24.9, 39.1]\end{aligned}$$

All results are resumed in the Table 1, values are expressed in days.

As obtained results from the Fig. 6, Table 1, and the discussion parts the star-type network is the best network type with four nodes, with minimum accumulated delivery time uncertainties, which transferred to the OEM. To reach 68.2 % of confidence, in the star-type, a timeframe between 14.3 and 19.6 days is needed;

Table 1 Delivery time uncertainty timeframe

Timeframe for all types	Star-type	Linear-type	Tree-type (1)	Tree-type (2)
Delivery time uncertainty (k = 1)	[14.3, 19.6]	[32.9, 41.1]	[18.3, 25.7]	[28.4, 35.6]
Delivery time uncertainty (k = 2)	[11.7, 22.3]	[28.8, 45.2]	[14.6, 29.4]	[24.9, 39.1]

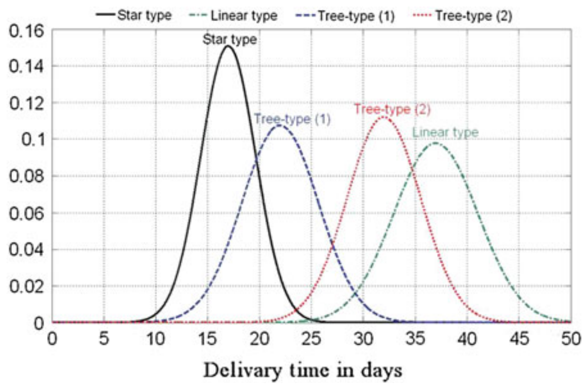


Fig. 6 Accumulation of the delivery type of network in the different network types

approximately near the 5.3 days as an interval confidence, while these results for other networks are much more than this number (e.g., in the linear type this time frame is [32.9–41.1] days; approximately near the 8.2 days as an interval confidence. For the tree-type (1) is [18.3–25.7] days; about the 7.4 days as an interval confidence and in the tree-type (2) this timeframe is between 28.4 and 35.6 days; about 7.2 days interval confidence). The ranking is the same with 95.4 % of confidence. Thus, in the star-type, because of the less delivery time uncertainty transferred to the OEM; it needs less interval confidence to reach to the same percentage of confidence than the others. This method could be use for more difficult network. This methodology gives implications to understand the sensitivity of delivery time uncertainty on types of network in supply networks.

Conclusion

Analysing the effect of individual uncertainties in delivery time on the delivery time uncertainty of OEM is one of the essential factors to improve the reliability in delivery time of the whole supply network. In this paper, the different network types with four nodes are studied, and then their role on delivery time uncertainty by proposed methodology is investigated. This methodology has the ability to use in the more complex network, and it is one of the most important features. Finally, the application of this methodology illustrated in all possible network types with four nodes by a numerical example.

In a supply network with four nodes, there are three categories of network. The first one entitled, "independent networks," all of suppliers have direct relation to the OEM, and the star-type of network is belonged to this category. In the second one, which called "dependent networks," all of suppliers interactions follow as a chain and except for the final supplier (the supplier before OEM), others have not direct relation to the OEM. The last category, included of the combination of the first and second categories, the tree-types networks are belonged to this part.

The final results of the example, represent the priority of the star-type of network than the others. The second position is belonged to the tree-type (2), in the less delivery time uncertainty which transferred to the OEM and respectively tree-type (1) and linear-type are in the next positions. The difference between the tree-type (2) and tree-type (1) and the priority of the number 2 are shown the supply network planners must design the networks, as possible to independency form at least in the final part.

This research was based on the normal probability density functions. The suggestion for the future researches is to investigate in the combination of the different PDFs (e.g., Beta, exponential and so on).

References

Barker J (2004) 'Using dynamic information in the supply network. In: Proceedings of the 8th international conference on manufacturing & management: operations management & advanced technology: integration for Success, Australia, pp 1–10

Christopher M (2000) The agile supply chain: competing in volatile markets. Ind Mark Manage 29(1):37–44

Guiffrida AL, Jaber MY (2008) Managerial and economic impacts of reducing delivery variance in the supply chain. Appl Math Model 32:2149–2161

JCGM Member Organizations (2008) 'Evaluation of measurement data—Guide to the expression of uncertainty in measurement (GUM)-GUM 1995 with minor corrections', JCGM

Lerder PJ (1997) Pricing, production, scheduling, and delivery-time competition. Int J Oper Res 45(3):407–420

Pathak SD, Dilts DD, Biswas G (2007) On the evolutionary dynamics of supply network topologies. IEEE Trans Eng Manage 54(4):662–672

Safaei M, Seifert M, Thoben KD (2010) Proposed a method for increasing the delivery performance in dynamic supply network. World Acad Sci Eng Technol 70:620–623

Safaei M, Seifert M, Thoben KD (2011) Toward a better understanding of the network type impact on delivery time uncertainty in supply networks', MITIP 2011. In: Proceedings of the 13th international conference on modern information technology in the innovation processes of industrial enterprises, pp 1–11

Thoben KD, Jagdev H (2001) Typological issues in enterprise networks. J Prod Plan Control 12(5):421–436

Urban TL (2009) Establishing delivery guarantee policies. Eur J Oper Res 196:959–967

Vanany I, Zailani S, Pujawan N (2009) Supply chain risk management: literature review and future research. Int J Inf Syst Supply Chain Manage 2(1):16–33

An LMI Approach to Exponential Stock Level Estimation for Large-Scale Logistics Networks

Hamid Reza Karimi

Abstract This article aims to present a convex optimization approach for exponential stock level estimation problem of large-scale logistics networks. The model under consideration presents the dependency and interconnections between the dynamics of each single location. Using a Lyapunov function, new sufficient conditions for exponential estimation of the networks are driven in terms of linear matrix inequalities (LMIs). The explicit expression of the observer gain is parameterized based on the solvability conditions. A numerical example is included to illustrate the applicability of the proposed design method.

Keywords Stability analysis · Logistics networks · Estimation · LMI

Introduction

Logistics networks are emerging as a new type of cooperation between and within companies, requiring new techniques and methods for their operation and management (Wiendahl and Lutz 2002). Logistics network design has been referred to as facility location problems in literature. Location theory was first introduced by Weber (1929) who considered the problem of locating a single warehouse among customers to minimize total distance between warehouse and customers. In recent years, a lot of research work has been published and several review papers have sought to categorize and compare different models. The behaviour of such a

H. R. Karimi (✉)
Department of Engineering, Faculty of Engineering and Science, University of Agder, Grimstad, Norway
e-mail: hamidrk@uia.no

H.-J. Kreowski et al. (eds.), *Dynamics in Logistics*, Lecture Notes in Logistics,
DOI: 10.1007/978-3-642-35966-8_29,

network is affected by external and internal order flows, planning, internal disturbances, and the control laws used locally in the work systems to adjust resources for processing orders (Duffie et al. 2008). These networks can exhibit unfavourable dynamic behaviour as individual organizations respond to variations in orders in the absence of sufficient communication and collaboration, leading to recommendations that supply chains should be globally rather than locally controlled and that information sharing should be extensive Helo (2000); Huang et al. (2003); Monostori et al. (2004); Duffie et al. (2008); Nyhuis et al. (2005); Scholz-Reiter et al. (2005); Duffie and Falu (2002). However, the dynamic and structural complexity of these emerging networks inhibits collection of the information necessary for centralized planning and control, and decentralized coordination must be provided by logistic processes with autonomous capabilities (Monostori et al. 2004). Recently, a local input-to-state stability (ISS) of logistics networks was studied in (Dashkovskiy et al. 2012). Their behaviour is described by a functional differential equation with a constant time-delay. An appropriate Lyapunov-Razumikhin function and the small gain condition are utilized to establish some conditions for stability analysis of the network under consideration. Also, in Dashkovskiy et al. (2010) the logistics networks are investigated in view of stability to avoid negative outcomes such as high inventory costs or loss of customers. The local input-to-state-stability (LISS) property and the tool of an LISS Lyapunov–Krasovskii functional are used for the stability investigation. More recently, the problem of local capacity control for a class of production networks of autonomous work systems with time-varying delays in the capacity changes was presented in (Karimi et al. 2011, 2010). Attention was focused on the design of a controller gain for the local capacity adjustments which maintains the work in progress (WIP) in each work system in the vicinity of planned levels and guarantees the asymptotic stability of the system and reduces the effect of the disturbance input on the controlled output to a prescribed level.

On the other hand, the problem of observer design for reconstructing state variables is a more involved issue in systems with any kind of nonlinearity terms. In general, some sufficient conditions for the existence of an observer have been established and computational algorithms for construction of the observers have been presented in the literature, see for instance the references Thau (1973), Trinh et al. (2006), Busawon and Saif (1999). In Zitek (1998), the spectrum assignment method was introduced to design the state observer. Lyapunov stability theory is used to design the state observers for linear time-varying or nonlinear systems, see the references Gu and Poon (2001), Karimi (2008), Gao et al. (2005), Khargonekar et al. (1996), Lien (2004).

In this paper, we contribute to the problem of exponential stability analysis and a convex optimization approach for filtering problem of large-scale logistics networks. The system under consideration presents the dependency and interconnections between the dynamics of each single location. Using a Lyapunov function, new sufficient conditions for exponential estimation of the networks are driven in terms of linear matrix inequalities (LMIs). The explicit expression of the observer gain is parameterized based on the solvability conditions. Finally, numerical results are provided to demonstrate the proposed approach.

The rest of this paper is organized as follows. Section Problem Description formulates the exponential observer problem of the large-scale logistic networks. In section Main Results, a convex optimization approach is given to calculate the observer gain. In section Numerical Results, computer simulations are provided to demonstrate the effectiveness of the proposed scheme. Finally, conclusions are presented in section Conclusion.

The notations used throughout the paper are fairly standard. I and 0 represent identity matrix and zero matrix; the superscript 'T' stands for matrix transposition; $\mathcal{R}^{\boldsymbol{n}}$ denotes the n-dimensional Euclidean space; $\mathcal{R}^{\boldsymbol{n}\times\boldsymbol{m}}$ is the set of all real n by m matrices. Refers to the Euclidean vector norm or the induced matrix 2-norm. $\boldsymbol{diag}\{\ldots\}$ represents a block diagonal matrix. The operator $\boldsymbol{sym}\{\boldsymbol{A}\}$ denotes $\boldsymbol{A}+\boldsymbol{A}^{\boldsymbol{T}}$. The notation $\boldsymbol{P}>\boldsymbol{0}$ means that $\boldsymbol{P}$ is real symmetric and positive definite; the symbol $*$ denotes the elements below the main diagonal of a symmetric block matrix. Finally, given a signal $\boldsymbol{x}(\boldsymbol{t})$, $||\boldsymbol{x}(\boldsymbol{t})||_2$ denotes the $\boldsymbol{L_2}$ norm of $\boldsymbol{x}(\boldsymbol{t})$; i.e., $\|\boldsymbol{x}(\boldsymbol{t})\|_2^2 = \int_{\boldsymbol{0}}^{\infty} \boldsymbol{x}^{\boldsymbol{T}}(\boldsymbol{t})\boldsymbol{x}(\boldsymbol{t})\boldsymbol{dt}$.

Problem Description

Here we consider a mathematical modelling that deal with continuous material flows in the network. Such a network can be described by ordinary differential equations as follows:

$$\dot{x}_i = u_i + \sum_{i,j\neq i} c_{ij} f_j\left(x_j(t)\right) - f_i(x_i(t)) \tag{1}$$

$$y_i = g_i(x_i(t)) \tag{2}$$

where $x_i(t)$ denotes the state of the ith location, $i = 1,\ldots,n$, and $0 \leq c_{ij} \leq 1$ is the share of orders of location j at location i and $\sum_{i=1}^{n} c_{ij} = 1$. $u_i(t)$ is the system's state variable associated with the ith neuron. $f_i(.)$ is the production rate function.

In the sequel, we make the following definitions for the systems (1).

Definition 1 Each function f_i is continuous with $f_i(0) = 0$ and there exists two scalars $\underline{f}_i$ and $\bar{f}_i$ with $\underline{f}_i < \bar{f}_i < \infty$ such that for any $\alpha, \beta \in \mathcal{R}$,

$$\underline{f}_i(\alpha-\beta)^2 \leq [f_i(\alpha) - f_i(\beta)](\alpha-\beta) \leq \bar{f}_i(\alpha-\beta)^2$$

.

We set

$$\underline{E} = diag\left\{\underline{f}_1, \underline{f}_2, \ldots, \underline{f}_n\right\}$$

and $\bar{E} = diag\{\bar{f}_1, \bar{f}_2, \ldots, \bar{f}_n\}$.

Definition 2 (*Exponential stability*) A dynamical system is said to be exponentially stable if there exists constants $\beta(\alpha) \geq 1$ and $\alpha > 0$ such that $|x(t)| \leq \beta(\alpha) sup_{-\tau^* \leq s \leq 0} |x(s)| e^{-\alpha t}$ for any $t \geq 0$. Constant α is said to be the degree of exponential synchronization (Khalil 1992).

According to Definition 2, if a new error $\hat{x}_i(t)$ is defined by $\hat{x}_i(t) = e^{\alpha t} x_i(t),$, then the error dynamics between (1) and (2) can be expressed by

$$\dot{\hat{x}}_i(t) = \alpha \hat{x}_i(t) + \hat{u}_i + \sum_{i,j \neq i} \hat{c}_{ij} \hat{f}_i\left(x_j(t)\right) \tag{3}$$

$$\hat{y}_i = \hat{g}_i(\hat{x}_i(t)) \tag{4}$$

where

$$\hat{f}_j\left(x_j(t)\right) := e^{\alpha t} f_j\left(x_j(t)\right),$$

$$\hat{g}_j\left(\hat{x}_j(t)\right) := e^{\alpha t} g_j\left(x_j(t)\right)$$

and

$$\hat{c}_{ij} := \begin{cases} c_{ij}, j \neq i \\ -1, j = i \end{cases}$$

Remark 1 It is easy to see that by Definition 1, each function $\phi_i(.)$ is continuous with $\phi_i(0) = 0$ and there exists two scalars $\underline{f}_i$ and $\underline{f}_i$ with $f_i < \bar{f}_i < \infty$ such that

$$\underline{f}_i(\alpha - \beta)^2 \leq [\phi_i(\alpha) - \phi_i(\beta)](\alpha - \beta) \leq \bar{f}_i(\alpha - \beta)^2 \tag{5}$$

The main objective of the paper is to design an nth order filter with the following state-space equations

$$\dot{\hat{x}}_{f_i}(t) = \alpha \hat{x}_{f_i}(t) + \hat{u}_i + \sum_{i,j \neq i} \hat{c}_{ij} \hat{f}_j\left(\hat{x}_{f_j}(t)\right) + l_i\left(\hat{y}_i - \hat{y}_{f_i}\right) \tag{6}$$

$$\hat{y}_{f_i}(t) = \hat{g}_i\left(\hat{x}_{f_i}(t)\right) \tag{7}$$

where $\hat{x}_{f_i}(t)$ and $\hat{y}_{f_i}(t)$ are the estimations of $x_{f_i}(t)$ and $y_{f_i}(t)$, respectively, and $\hat{e}_i(t) = \hat{x}_i(t) - \hat{x}_{f_i}(t)$ is the estimation error. The parameter l_i is the observer design objective to be determined. It is required that $\|\hat{e}_i(t)\| \to 0$ as $t \to \infty$.

Now, we obtain the following state-space model, namely filtering error system:

$$\dot{\hat{e}}_i(t) = \alpha\hat{e}_i(t) + \sum_{i,j\neq i} \hat{c}_{ij}\phi_j\big(\hat{e}_j(t)\big) + l_i\psi_i(\hat{e}_i(t)) \tag{8}$$

where

$$\phi_j\big(\hat{e}_j(t)\big) := \hat{f}_j\big(\hat{e}_j(t) + \hat{x}_{f_j}(t)\big) - \hat{f}_j\big(\hat{x}_{f_j}(t)\big)$$

and $\psi_j\big(\hat{e}_j(t)\big) := \hat{g}_j\big(\hat{e}_j(t) + \hat{x}_{f_j}(t)\big) - \hat{g}_j\big(\hat{x}_{f_j}(t)\big)$.

Therefore, the set of Eq. (8) can be represented in compact form as

$$\dot{\hat{e}}(t) = \Lambda\hat{e}(t) + C\phi(\hat{e}(t)) + L\psi(\hat{e}(t)) \tag{9}$$

where $C := \big(c_{ij}\big)$, $L = diag\{l_1, l_2, \ldots, l_n\}$, $\Lambda := \alpha\mathrm{I_n}$, and

$$\phi(\hat{e}(t)) := col\{\phi_1(\hat{e}_1(t)), \phi_2(\hat{e}_2(t)), \ldots, \phi_n(\hat{e}_n(t))\}$$

and

$$\psi(\hat{e}(t)) := col\{\psi_1(\hat{e}_1(t)), \psi_2(\hat{e}_2(t)), \ldots, \psi_n(\hat{e}_n(t))\}$$

Main Results

In this section, sufficient conditions for the solvability of the observer design problem of the logistics network (1–2) are proposed in the following theorem using the Lyapunov method and an LMI approach is developed.

Theorem 1 *Consider the two logistics network* (1) and (2). *The observer error system* (6–7) *can exponentially estimate stock levels if there exist a matrix* $\overline{P}$ *and a positive-definite matrix* P, *such that the following matrix inequality (LMI) is feasible*,

$$\begin{bmatrix} sym\Big\{P\Lambda - \underline{E}(\Lambda_1 + \Lambda_2)\bar{E}\Big\} & PC + \Big(\underline{E} + \bar{E}\Big)\Lambda_1 & \bar{P} + \Big(\underline{E} + \bar{E}\Big)\Lambda_2 \\ * & -2\Lambda_1 & 0 \\ * & * & -2\Lambda_2 \end{bmatrix} < 0 \tag{10}$$

The desired observer gain matrix is given by

$$L = P^{-1}\bar{P} \tag{11}$$

Proof To investigate the stability analysis of the error system (9), we define a class of Lyapunov functions as follows

$$V(t) = \sum_{i=1}^{3} V_i(t) \tag{12}$$

where

$$V_1(t) = \hat{e}^T(t)P\hat{e}(t)$$

$$V_2(t) = -2\sum_{i=1}^{n} \beta_i^{(1)} \int_0^t \Big(\phi_i(\hat{e}_i(s)) - \underline{f}_i\hat{e}_i(s)\Big)(\phi_i(\hat{e}_i(s)) - \bar{f}_i\hat{e}_i(s))ds$$

and

$$V_3(t) = -2\sum_{i=1}^{n} \beta_i^{(2)} \int_0^t \Big(\psi_i(\hat{e}_i(s)) - \underline{f}_i\hat{e}_i(s)\Big)(\psi_i(\hat{e}_i(s)) - \bar{f}_i\hat{e}_i(s))ds$$

Time derivative of $V_i(t), i = 1, \ldots, 3$, along the system trajectory (9) become, respectively,

$$\begin{aligned}\dot{V}_1(t) &= 2\hat{e}^T(t)P\dot{\hat{e}}(t)\\ &= 2\hat{e}^T(t)P(\Lambda\hat{e}(t) + C\phi(\hat{e}(t)) + L\psi(\hat{e}(t)))\end{aligned} \tag{13}$$

$$\dot{V}_2(t) = -2\hat{e}^T(t)\underline{E}\Lambda_1\bar{E}\hat{e}(t) - 2\phi^T(\hat{e}(t))\Lambda_1\phi(\hat{e}(t)) + 2\hat{e}^T(t)(\bar{E} + \underline{E})\Lambda_1\phi(\hat{e}(t)) \tag{14}$$

and

$$\dot{V}_3(t) = -2\hat{e}^T(t)\underline{E}\Lambda_2\bar{E}\hat{e}(t) - 2\psi^T(\hat{e}(t))\Lambda_2\psi(\hat{e}(t)) + 2\hat{e}^T(t)(\underline{E} + \bar{E})\Lambda_2\psi(\hat{e}(t)) \tag{15}$$

where $\Lambda_j = diag\Big\{\beta_1^{(j)}, \beta_2^{(j)}, \ldots, \beta_n^{(j)}\Big\}, \quad j = 1, 2.$

From (12–15), we have

$$\dot{V}(t) \le \pi^T(t)\Xi\pi(t) \tag{16}$$

where

$$\pi(t) := col\{\hat{e}(t), \phi(\hat{e}(t)), \psi(\hat{e}(t))\}$$

and the matrix Ξ is given by

$$\Xi := \begin{bmatrix} sym\{P\Lambda - \underline{E}(\Lambda_1 + \Lambda_2)\bar{E}\} & PC + (\underline{E} + \bar{E})\Lambda_1 & PL + (\underline{E} + \bar{E})\Lambda_2 \\ * & -2\Lambda_1 & 0 \\ * & * & -2\Lambda_2 \end{bmatrix} \tag{17}$$

Then, by considering $\bar{P} := PL$, the LMI (10) is obtained. From (16), we can easily obtain for all $\hat{e}(t) \neq 0$,

$$\dot{V}(t) < -\mathcal{X}\lambda_{min}(P)\hat{e}(t)^2 < 0 \tag{18}$$

where κ is a positive constant.

Based on Lyapunov stability theory, the estimation error (9) is exponentially stable. This completes the proof.

Numerical Results

In this section, we will verify the proposed methodology by giving an illustrative example. We solved LMI (10) by using Matlab LMI Control Toolbox (see the Ref. (Gahinet et al. 1995)). The example is given below.

Example 1 Consider the case of a logistics network with five work stations and actual production rates $f_i(x_i(t)) = \eta_i\left(1 - e^{-x_i(t)}\right)$, where η_i is the maximal production rate and the state $x_i(t)$ is the number of unsatisfied orders at location i. Let us take values $c_{12} = 0.3$, $c_{13} = 0.1$, $c_{14} = 0.3$, $c_{15} = 0.5$, $c_{21} = 0.3$, $c_{31} = 0.1$, $c_{41} = 0.2$, $c_{51} = 0.3$ for the shares of delivered products and $u_i(t) = 0, i = 1, 2, \ldots, 5,$ for external inputs and $\eta_1 = 8, \eta_2 = 5.8, \eta_3 = 3, \eta_4 = 5.2, \eta_5 = 7.8$ for the maximal production rates. The initial values of stock level are given by $x_1(0) = 1, x_2(0) =$

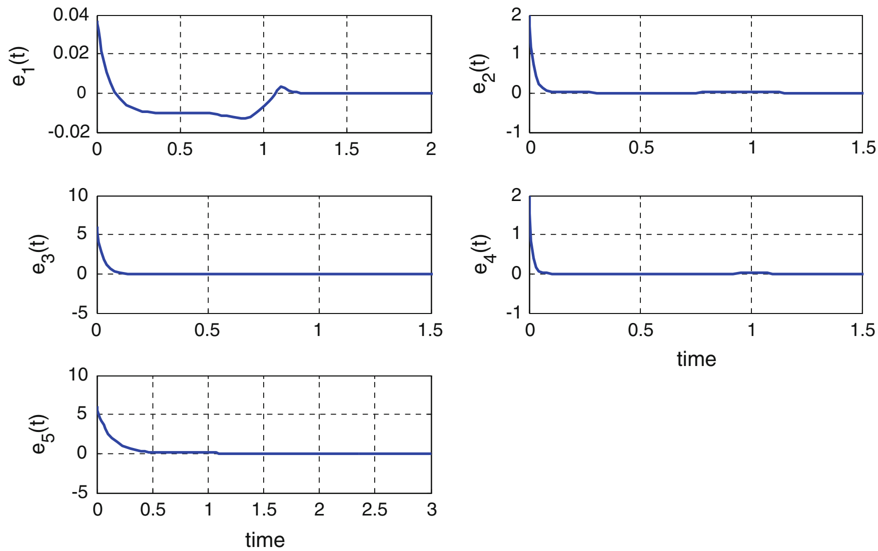


Fig. 1 The estimation errors of stock levels at different locations

$2, x_3(0) = 6, x_4(0) = 2, x_5(0) = 6$. It is also assumed $g_1(x_1(t)) = \sin(0.1x_1(t))e^{-x_i(t)}$ and $g_i(x_i(t)) = |x_i(t)|$ for $i = 2, 3, \ldots, 5$.

It is required to design an observer such that the stock levels of logistics network can be exponentially estimated. To this end, in light of Theorem 1, we solved LMI (10) with $\alpha = 2.0$ and obtained the observer gain as $L = diag\{51.5, 80.24, 78.85, 98.55, 46.55\}$. The simulation results are given in Fig. 1. As one can mention all the stock levels are exponentially estimated well.

Conclusion

The problem of exponential stability analysis and a convex optimization approach for filtering problem of large-scale logistics networks were studied in this paper. The system under consideration presented the dependency and interconnections between the dynamics of each single location. Using a Lyapunov function, new sufficient conditions for exponential estimation of the networks were driven in terms of linear matrix inequalities (LMIs). The explicit expression of the observer gain was parameterized based on the solvability conditions. Finally, numerical results were provided to demonstrate the proposed approach.

References

Busawon KK, Saif M (1999) A state observer for nonlinear systems. IEEE Trans Autom Control 44:2098–2103

Dashkovskiy S, Karimi HR, Kosmykov M, Mironchenko A, Naujok L (2010) Application of the LISS Lyapunov-Krasovskii small-gain theorem to autonomously controlled production networks with time-delays. In Proceedings of the conference on control and fault-tolerant systems, pp. 765–770, Nice, France

Dashkovskiy S, Karimi HR, Kosmykov M (2012) A Lyapunov-Razumikhin approach for stability analysis of logistics networks with time-delays. Int J Syst Sci 43(5):845–853

Duffie N, Falu I (2002) Control-theoretic analysis of a closed-loop PPC system. Ann CIRP 5111:379–382

Duffie NA, Roy D, Shi L (2008) Dynamic modelling of production networks of autonomous work systems with local capacity control. CIRP Ann Manuf Technol 57:463–466

Gahinet P, Nemirovsky A, Laub AJ, Chilali M (1995) LMI control toolbox: for use with matlab. The MATH Works, Inc., Natik

Gao H, Lam J, Xie LH, Wang CH (2005) New approach to mixed H2/H∝ filtering for polytopic discrete-time systems. IEEE Tran Signal Process 53(8):3183–3192

Gu DW, Poon FW (2001) A robust state observer scheme. IEEE Trans Autom Control 46:1958–1963

Helo P (2000) Dynamic modelling of surge effect and capacity limitation in supply chains. Int J Prod Res 38(17):4521–4533

Huang GQ, Lau JSK, Mak KL (2003) The impacts of sharing production information on supply chain dynamics: a review of the literature. Int J Prod Res 41(7):1483–1517

Karimi HR (2008) Observer-based mixed H_2/H_∞ control design for linear systems with time-varying delays: an LMI approach. Int J Control Autom Sys 6(1):1–14
Karimi HR, Duffie NA, Dashkovskiy S (2010) Local capacity H_∞ control for production networks of autonomous work systems with time-varying delays (regular paper). IEEE Trans Autom Sci Eng 7(4):849–857
Karimi HR, Dashkovskiy S, Duffie NA (2011) Stability analysis of large scale networks of autonomous work systems with delays': dynamics in logistics. Springer, Heidelberg, pp 69–78
Khalil HK (1992) Nonlinear systems. Macmillan, New York
Khargonekar PP, Rotea MA, Baeyens E (1996) Mixed $H2/H_\infty$ filtering. Int J Robust Nonlinear Control 6(6):313–330
Lien CH (2004) Robust observer-based control of systems with state perturbations via LMI approach. IEEE Trans Autom Control 49(8):1365–1370
Monostori L, Csaji BC, Kadar B (2004) Adaptation and learning in distributed production control. Ann CIRP 53(1):349–352
Nyhuis P, Cieminski G, Fischer A (2005) Applying simulation and analytical models for logistic performance prediction. Ann CIRP 54(1):417–422
Scholz-Reiter B, Freitag M, de Beer C, Jagalski T (2005) Modelling dynamics of autonomous logistic processes: discrete-event versus continuous approaches. CIRP Ann Manuf Technol 54(1):413–416
Thau FE (1973) Observing the state of nonlinear dynamic systems. Int J Control 17:471–479
Trinh H, Fernando T, Nahavandi S (2006) Partial-state observers for nonlinear systems. IEEE Trans Autom Control 51(11):1808–1812
Weber A (1909) Uber den Standort der Industrien, – translated as Alfred Weber's theory of location of industries. University of Chicago Press, Chicago
Wiendahl H-P, Lutz S (2002) Production in networks. Ann CIRP 52(2):573–586
Zitek P (1998) Anisochronic state observers for hereditary systems. Int J Control 71:581–599

Part VII
Information, Communication, Risk, and Failure

Characterization of Thermal Flow Sensors for Air Flow Measurements in Transport Containers

Safir Issa and Walter Lang

Abstract Air flow measurements inside containers for sensitive and perishable products effectively participate in improving transport processes. Results of such measurements allow taking preventive actions to maintain the desired temperature during transport trips. Consequently, we can optimize the quality of transported goods and reduce their losses. Thermal flow sensors are chosen for these measurements. This paper introduces an overall characterization of these sensors to prove their suitability for the intended objective. The characterization covers the air velocity range from 0 to 5 m/s, which is the expected range in the container. Results show that the characteristic curve is linear for the ultra low flow range and the minimum detectable air velocity is ca. 0.4 mm/s.

Keywords Characterization · Thermal flow sensors · Transport container

S. Issa (✉) · W. Lang
IMSAS (Institute for Microsensors, Actuators and Systems),
Microsystems Center Bremen (MCB), University of Bremen,
Otto-Hahn-Allee, Bld. NW1 D-28359 Bremen, Germany
e-mail: sissa@imsas.uni-bremen.de

W. Lang
e-mail: wlang@imsas.uni-bremen.de

S. Issa · W. Lang
IGS (International Graduate School for Dynamics in Logistics),
University of Bremen, C/o BIBA Hochschulring 20 D-28359 Bremen, Germany

H.-J. Kreowski et al. (eds.), *Dynamics in Logistics*, Lecture Notes in Logistics,
DOI: 10.1007/978-3-642-35966-8_30,

Introduction

Nowadays, customers demand a constantly increasing amount of sensitive and perishable products. This fact forces producers and marketers to ensure the arrival of these products to their target consumers in a good quality. In case of fruit and vegetables, temperature is the dominant environmental factor that influences their deterioration. It affects their external shape, quality and shelf life. Temperatures either above or below the optimal range for fresh produce can cause rapid deterioration due to some factors such as freezing, chilling injury or heat injury (FAO 2004). Thus maintaining a specific temperature throughout the container during the whole transport process is an essential matter to keep product's quality and to reduce its losses. However, this purpose is very difficult to achieve, due to the fact that air conditioning system should provide homogeneous distribution of air flow inside the container, which in reality is not possible. Additionally, the internal production of heat and moisture generated by fruit and vegetables are supplementary parameters that affect the temperature profile. Furthermore, the internal geometry of containers participates in the non-uniform distribution of temperature. All equipped pallets and boxes can provide only narrow spaces and holes for air current passages. This crucial role of temperature not only encourages researchers to measure it in real time, but also to predict its values in advance. Such prediction allows taking preventive actions to maintain temperature within the allowable limits. One of the essential methods to predict temperature variations is the air flow measurement. The temperature and its distribution are controlled by air flow pattern (Moureh et al. 2009). Thus performing air flow measurements will provide a better understanding of convection transport inside the container and will identify the stagnant zones where the air flow is very poor. Temperatures in these zones are surely higher than expected as the air circulation is not able to convect the generated heat. In the next level, the measurement results will help in the optimization of air circulation and the improvement of efficiency of the air conditioning system, to avoid forming stagnant zones and to obtain more homogeneity in temperature distribution.

In the literature there are some reports about studying air flow patterns for the perishable goods transportation. Zou et al. (2005) develop a computational fluid dynamics (CFD) modeling system of the airflow patterns and heat transfer inside ventilated packages for fresh food. This research is concerned by food packages and not the whole container. Moureh et al. (2009) introduced reports on the numerical and experimental characterization of airflow within a semi-trailer enclosure loaded with pallets in a refrigerated vehicle with and without air ducts. Measurements of air velocities were carried out by a laser Doppler velocimeter in clear regions (above the pallets) and thermal sphere-shaped probes located inside the pallets. The velocimeter is placed outside the vehicle and the measurement is done through a glass window.

The present work is a part of the above referred objective; air flow measurements inside logistic containers by means of thermal flow sensors. It proves the suitability and the capability of the chosen sensors to perform the intended measurements. They are micro sensors developed by IMSAS (Buchner et al. 2006). This choice is based on the multi measurement requirements in containers. Their very small size enables placing them in different positions in the container i.e., on walls, ceiling, between pallets, inside boxes, etc. Moreover, these sensors are able to send data through wireless network by using RFID technologies, and to be integrated within a sensor network such as presented in Jedermann et al. (2009). Additionally, they have a fast response time (Issa et al. 2011).

In this paper, an overall characterization for these sensors is studied. Section Measurement Setup and Characteristic Curves describes the measurement setup and the characteristic curves of the sensors. Section Minimum Detectable Air Velocity focuses on the minimum detectable flow which is investigated by both theoretical calculation and experimental measurements.

Measurement Setup and Characteristic Curves

The investigated thermal flow sensors are based on silicon as substrate material; they consist of a heater and two symmetric thermopiles embedded in silicon nitride membrane. The heater is made of tungsten-titanium, whereas the thermopiles are made of a combination of polycrystalline silicon and tungsten-titanium. In case of zero flow, the heat generated by the heater is distributed uniformly to both sides and no difference in temperature is detected by the thermopiles signals. However, in the case of flow, there is a difference in temperature between the two signals, as the air flow convects a part of the generated heat. This difference is related to the value of flow and will be the interesting part for the sensor characterization. Four sensor configurations are considered in this study TS5, TS10, TS20 and TS50. They have the same membrane area of 1 mm^2 and differ in the distance between heater and thermopiles. This distance is 5 μm for TS5 and 50 μm for TS50.

Two flow ranges are studied to separate the linear part of the characteristic curve from the whole curve. The ultra-low flow is from 0 to 10 Standard Cubic Centimeter per Minute (SCCM) and the low flow is until 1,000 SCCM. They cover the probable air velocity values inside the container from 0 to 5 m/s. In the characterization setup, three mass flow controllers with maximum capacities: 20, 200 and 1,000 SCCM from MKS Company are used. They are connected through pipes to an air supply source from one side and to the sensor air channel from the other side as shown in Fig. 1. These controllers are driven by a control unit MKS through a LabVIEW program. According to this program several mass flow controllers are chosen with an adjustable flow steps. The sensor is operated in constant power mode and the two thermopile's signals are extracted by National Instruments (NI) data acquisition device. The data of sensor's outputs with the related flow values are stored in a PC.

Firstly, the ultra-low-flow is discussed. In this case only one MKS mass flow controller is used with maximum capacity of 20 SCCM. By analyzing the

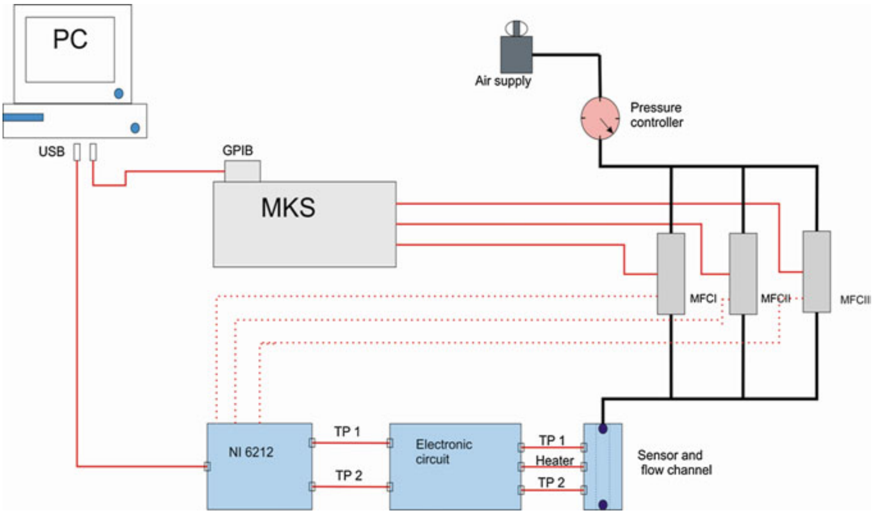


Fig. 1 Measurement setup for characterization of thermal flow sensors

extracted data, we calculate the output voltage difference which is the difference between the two thermopile signals as a function of air flow. The resulting curve is called the characteristic curve. We compare then the curves of the four sensor configurations TS5, TS10, TS20 and TS50. Figure 2 shows this comparison, these curves are all linear as they are in a good agreement with the linear fitting. The R-squared values are all in the range of 0.999. This fact indicates that the fitting degree is very high. We can notice that the sensor sensitivity increases with the distance between the heater and the thermopiles. This gives advantages to select TS50 sensor for future measurements in containers. It is important to mention that these sensors have a small zero offset due to a slight difference between the up- and down- stream thermopiles. When the two thermopiles are perfectly identical, they give the same signal at zero flow, else a small difference between the two signals is noticed, which cause the zero offset. The zero offset values are taken into consideration to allow all curves to start from the same point for comparison.

Secondly, the low-flow case is presented. Here we use 3 mass flow controllers, with maximum capacities 20, 200 and 1,000 SCCM. The characteristic curves for the four sensor configurations are extracted as in the ultra-low flow case. The output voltage difference starts to increase linearly with the air velocity until a certain limit and then continues its increase but in non linear way as in Fig. 3. The function model that specifies the relationship between air velocity and voltage difference is:

$$\Delta E = a\,e^{bu} + c\,e^{du}$$

where ΔE is the output voltage difference of the sensor; u is the air velocity; a, b, c and d are constants to be determined for each sensor configuration. We use the

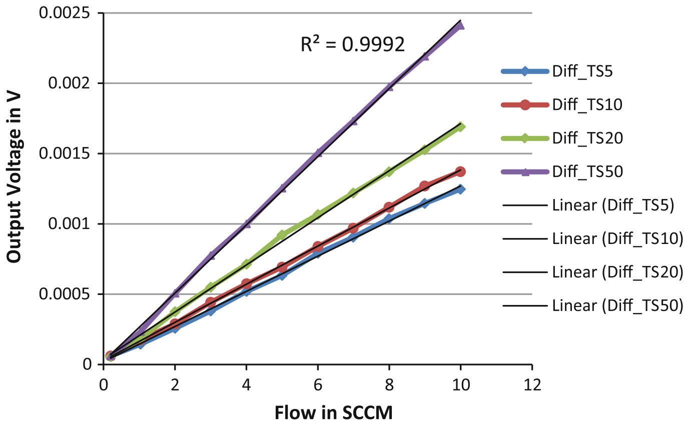


Fig. 2 Thermopiles output voltage difference as function of flow with the linear fitting for the four sensors configurations TS5, TS10, TS20 and TS50

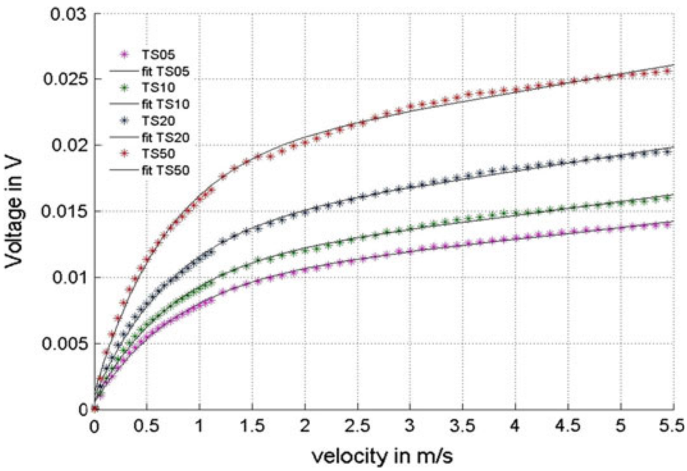


Fig. 3 Characteristic curves with fits of the four sensor configurations: TS5, TS10, TS20 and TS50. Dots are the experimental results and the lines are the fits

MATLAB based function fminsearchbnd to find the suitable fits. Substituting experimental measurement results in the model formula enables estimating the best values of the constants for the best fit. Figure 3 shows the characteristic curves with their fits of the four sensor configurations as function of air velocity. Moreover, Table 1 depicts the constant values of a, b, c and d in addition to the R- squared values which are very close to 1. These results ensure the suitability of the fits.

Table 1 Values of constants for the fitting curves and the R-squared values

	a	b	c	d	R_seq
TS05	0.01	0.0638	0.0095	1.2735	0.9987
TS10	0.0114	0.0647	0.0107	1.33	0.9982
TS20	0.0142	0.0616	0.0131	1.3284	0.9981
TS50	0.0194	0.0539	0.0181	1.4459	0.9973

Minimum Detectable Air Velocity

The minimum detectable flow is a parameter characterizes flow sensors. It becomes very important when these sensors are used for very low-flow applications. This parameter is basically influenced by natural convection and thermopile noises. Firstly, the natural convection is a mechanism in which the fluid motion is generated by density differences in the fluid due to temperature gradients (Incropera and DeWitt 2002). Air surrounding the sensor heater receives heat, becomes less dense and rises. The surrounding, cooler air then moves to replace it. This cooler air is then heated and the process continues, forming convection current. Natural convection air velocities are very small and can be neglected in many cases, especially when there is a (forced) air flow through the sensor's channel. Secondly, the thermopile noises are basically the temperature noise and the thermal noise. The temperature noise is caused by temperature fluctuations in the surrounding atmosphere. We assume that this noise has negligible effect on our calculations. The thermal noise or the Johnson noise is an electrical noise source caused by random motion of electrical charges in the material. The Johnson noise is determined by Johnson (1928):

$$V_{noise} = \sqrt{4 \cdot \mathrm{k_B} \cdot \mathrm{T_{ext}} \cdot \mathrm{R_e} \cdot \Delta \mathrm{f}}$$

where $\mathrm{k_B}$ is the Boltzmann's constant; $\mathrm{T_{ext}}$ is the external temperature; $\mathrm{R_e}$ is the electrical serial resistance and Δf is the frequency bandwidth.

With $\mathrm{k_B} = 1.38066 \times 10^{-23}\ \mathrm{JK^{-1}}$; $\mathrm{T_{ext}} = 323$ K; $\mathrm{R_e} = 200$ KΩ; $\Delta \mathrm{f} = 1{,}000$ Hz. The noise is then 1.89 μV.

To identify the experimental noise level, the sensor output signals are extracted in the zero flow case. The signals are registered through a LabVIEW program for 1000 samples with time interval of 5 s. Figure 4 shows the sensor's output voltage difference as function of time.

We analyze statistically the results to calculate the arithmetic mean and the uncertainty. The arithmetic mean value is 0.1198 mV, assumed to be the zero offset of the sensor. The expanded uncertainty of a measurement U is given by JCGM (2008):

$$U = k \cdot u_c$$

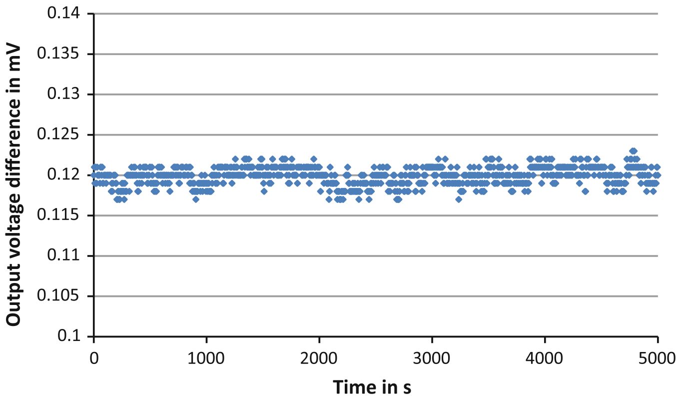


Fig. 4 Sensor output signal (thermopiles voltage difference) versus time in the zero flow case

where k is the coverage factor *and* u_c is the combined uncertainty. Assuming that the sample' values follow a Normal distribution function, then for a level of confidence 95 %, the coverage factor k equals 2 (JCGM 2008). u_c combines all uncertainty components. We assume that the statistically-evaluated components are the dominant ones. Then u_c is restricted to the standard deviation of the sample values. The standard deviation of the results is 0.0011 mV and, the corresponding measurement expanded uncertainty is 2.3 μV. Comparing this value with the thermal noise which is 1.89 μV enables estimating the effect of the other parameters on sensor output signal, in the stagnant flow case. To calculate the minimum detectable velocity, we still need the sensor's sensitivity value. It is defined as the derivative of the output voltage difference with respect to the flow velocity at a zero flow as in the following equation (Kaltsas and Nassioppoulou 1999):

$$S = \left. \frac{\partial V_{diff}}{\partial u} \right|_{u=0}$$

In other words, sensitivity is the slope of the sensor's characteristic curve. For example, sensitivity of TS50 is 0.044 V/(m/s), the minimum detectable velocity is estimated then to be: 2.3 μV/0.044 V/(m/s) = 0.05 mm/s. This value is recognized by the sensor, not as noise.

Experimentally, different methods are investigated to determine the minimum detectable velocity value. The first one is done by using mass flow controller of maximum capacity of 20 SCCM. According to its specifications, the control range is from 2 to 100 % of full scale. The equivalent velocity range is from 2.22 to 111 mm/s as the air-channel area is 1.5 × 2 mm^2. All measurements performed by this controller show that velocities lower than 1 mm/s cannot be identified. This value is 20 times higher than the theoretical estimated value. Therefore, a physical method is established to generate smaller flow rates. This method is shown and presented in Fig. 5.

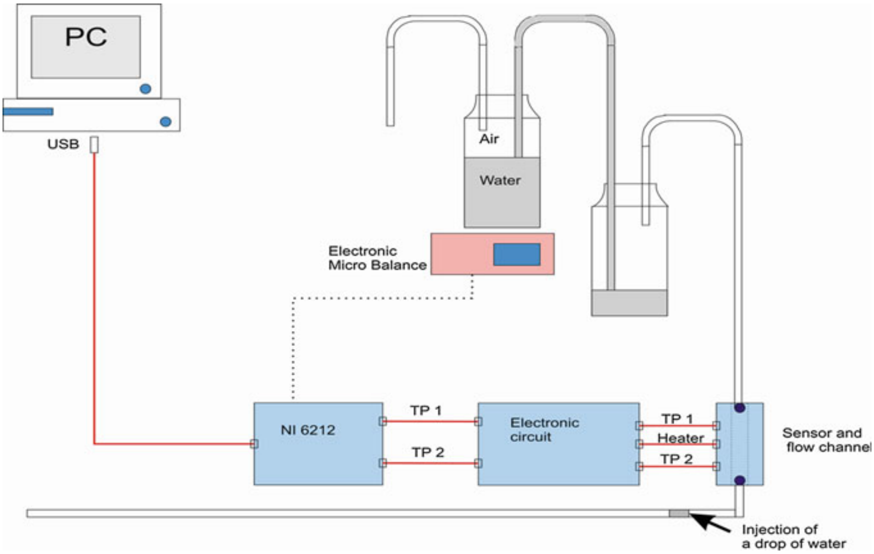


Fig. 5 Setup for generating very small flow rates. The flow is identified by two methods: first by measuring the water flow rate between *two closed bottles*. Second method is by injection a drop of water in the outlet pipe of the sensor and estimating its velocity

We initiate a water flow between two perfectly closed bottles, placed in different height positions. This forces an equivalent air flow from the second bottle to be created. The air flow is guided through a pipe to the sensor air-channel.

As initial condition, the first bottle is half full of water, and the second one is empty. Then we consider three different cases regarding the height positions of the two bottles. In such a way their height differences are: large, moderate and small respectively. In the first two cases, the first bottle is discharged completely into the second one, but of different speeds. Whereas in the third case, water starts to flow slowly between the two bottles until equilibrium is reached. This can happen when the two bottles have the same water level. Figure 6 compares the water mass flow vs. time for the three cases. In the first position, the flow decreases in linear way from 0.15 to 0.09 g/s. In the second position, the flow decreases slowly from ca. 0.08 to 0.015 g/s. In the last case, the velocity has a parabolic curve; it decreases very slowly toward zero. We are particularly interested in the latter case.

The associated LabVIEW program extracts the sensor output synchronized with the balance reading. We calculate the equivalent induced air flow and then the air velocity through the sensor air channel. From these data the curve of the sensor output is plotted vs. the air velocity as in Fig. 7. The sensor is capable to detect very small velocities. However, sensor output fluctuations increase when flow is in the vicinity of zero. This is due to several reasons: sensor noise, balance reading changes and also air humidity. Nevertheless, by this method the sensor can detect velocities less than 0.4 mm/s with uncertainty of sensor output signal of about ± 0.02 mV. This uncertainty value decreases significantly with the increase of air velocity.

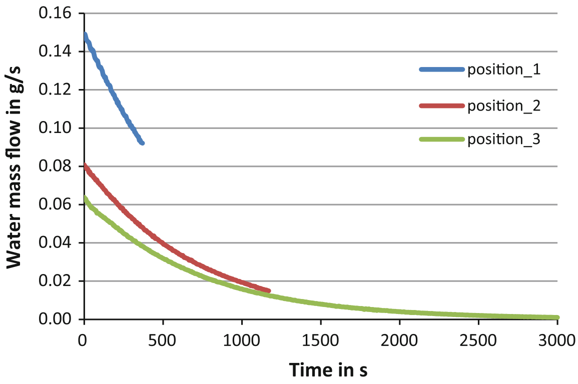


Fig. 6 Comparison of water mass flow between *two bottles* placed in *three different positions* regarding their height difference. The height differences are large, moderate and small respectively

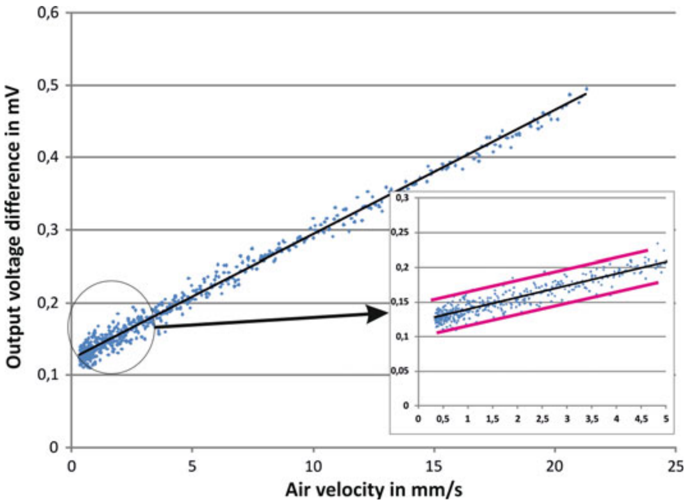


Fig. 7 Sensor output voltage difference as function of air velocity in the ultra low flow range

Finally, a third method is examined, by using the same setup shown in Fig. 5. The air flow in the outlet of the air channel is guided through a long straight pipe. A drop of water is injected into this pipe which will be pushed by the air flow. Thus, knowing its speed enables determining air velocity through the air channel. This can be done easily by measuring the distance traveled by the water drop within short time steps. By this method we can achieve very low speed values, less than 0.1 mm/s, but unfortunately the sensor output signals for such values are in the noise level. For this reason we adapt the result obtained by the second method.

Conclusion

In this study an overall characterization of thermal flow sensors is achieved for the ultra low and low flow ranges. Results show that the characteristic curves of different sensor configurations are linear for the range from 0 to 10 SCCM. Minimum detectable velocity is about 0.4 mm/s with sensor output voltage difference uncertainty of ca. 0.02 mV. This uncertainty value decreases significantly as the air velocity increases. As a result, these sensors are capable to perform very precise air flow measurements inside transport containers. They also fulfill the different requirements imposed by the container conditions. Performing these measurements will help improving transport processes by maintaining quality and reducing losses of the transported products.

References

Buchner R, Sosna C, Maiwald M et al (2006) A high temperature thermopile fabrication process for thermal flow sensors. Sens Actuators, A 130–131:262–266

FAO Corporate Document Repository (2004) The role of post-harvest management in assuring the quality and safety of horticultural produce

Incropera FP, DeWitt DP (2002) Fundamentals of heat and mass transfer, 5th edn. Wiley, Newyork

Issa S, Sturm H, Lang W (2011) Modeling of the response time of thermal flow sensors. Micromachines 2:385–393

JCGM 100:2008. Evaluation of measurement data—guide to the expression of uncertainty in measurement

Jedermann R, Ruiz-Garcia L, Lang W (2009) Spatial temperature profiling by semi-passive RFID loggers for perishable food transportation. Comput Electron Agric 65:145–154

Johnson JB (1928) Thermal agitation of electricity in conductors. Phys Rev 32:97–109

Kaltsas G, Nassioppoulou AG (1999) Novel C-MOS compatible monolithic silicon gas flow sensor with porous silicon thermal isolation. Sens Actuators, A 76:133–138

Moureh J, Tapsoba S, Derens E et al (2009) Air velocity characteristics within vented pallets loaded in a refrigerated vehicle with and without air ducts. Int J Refrig 32:220–234

Zou Q, Opara LU, McKibbin R (2005) A CFD modeling system for airflow and heat transfer in ventilated packaging for fresh foods. J Food Eng 77:1048–1058

Design and Implementation of Location-Based Handoff Scheme for Multimedia Data

Dong Hyun Kim and Jong Deok Kim

Abstract The handoff technology of the IEEE 802.11 refers to the process of transferring an Internet service from one channel connected to the Access Point (AP) to another. In order to provide mobility in wide area like port logistics, a handoff among APs is essential. However, the IEEE 802.11 handoff using hard-handoff may easily lose the connection and it wasters much time on recognizing the connection severance and channel search time for re-connection. In order to solve this problem, there have been several studies on the method of handoff prediction using (RSSI) Received Signal Strength Indication, moving pattern, multi-interface and Global Positioning System (GPS) technology. The handoff prediction methods that only consider one element compensate for the disadvantage of the legacy handoff method, but such handoff prediction methods still have problems. Especially, in case of a lot of mobile stations, the handoff prediction method using GPS causes a lot of data in GPS server and network. Then, if a mobile station handoffs only with distance or location information between APs, it is easy to cause a problem of frequent handoff in the border area. In order to solve such problems, we use GPS information to find out the location of the mobile station and predict the handoff point. Then, to guarantee the bandwidth, we propose the handoff algorithm using direction, velocity, and RSSI of the station. We implement the proposed handoff algorithm and evaluate its performance. As a result, we confirm the superiority of our algorithm by reducing disconnection time by about 3.7 ms. Our location-based handoff scheme improved the communication bandwidth by about 24 % more than that of location-based handoff.

D. H. Kim · J. D. Kim (✉)
Department of Computer Science and Engineering, Pusan National University,
30 Jangjeondong, Geumjeong Gu, Busan 609-735, Republic of Korea
e-mail: kimjd@pusan.ac.kr

D. H. Kim
e-mail: dhkim1106@pusan.ac.kr

H.-J. Kreowski et al. (eds.), *Dynamics in Logistics*, Lecture Notes in Logistics,
DOI: 10.1007/978-3-642-35966-8_31,

Keywords IEEE 802.11 · Location-based handoff · GPS · Port logistics

Introduction

IEEE 802.11 (2007)-based wireless Internet services support high transmission speed as well as convenient network connections, and it is the representative protocol for supporting wide-band wireless connection networks. In addition, mobile devices, such as a smart phones, laptops, PDAs and PMPs equipped with an IEEE 802.11-based WLAN module are spreading rapidly. Hence, users constantly want access to real-time multimedia services, such as VoIP, as well as Internet services.

While users are on the move, to provide these services, a mobile station needs multiple APs and it needs seamless handoff between these APs. However, legacy IEEE 802.11 technology uses the hard handoff method unlike Cellular network. The hard handoff is one in which the channel in the source AP is released and only then the channel in the target AP is engaged. Therefore, the handoff scheme of IEEE 802.11 easily loses connections, which consumes time dropped by data transmission. This time is proportional to the handoff latency; it approximates over 300 ms in the OSI layer 2 (Mishra et al. 2003). This is longer than 150 ms that users using Voice over Internet Protocol (VoIP) recognize a disconnection when they speak over a telephone and it is problem to support the real time multimedia services (ITU-T 1996). Moreover, if a mobile station uses a real-time multimedia service that needs to transmit a tremendous amount of data, then this requires high bandwidth. However, because IEEE 802.11 only considers the Received Signal Strength Indicator (RSSI) as a method of AP selection, the RSSI of AP that is selected by the moving direction of the mobile station may be weakened; therefore, the mobile station cannot get enough bandwidth. This causes the frequent handoff problem (Park et al. 2009). For this reason, we need a scheme that reduces handoff latency for real-time multimedia services to select the best AP considering the moving direction of the mobile station. In this paper, we propose an handoff scheme that predicts the handoff point of a mobile station using the location information of GPS, and the proposed algorithm guarantees high transmission bandwidth according to not only RSSI, but also to the moving speed and moving direction of mobile stations. Our experimental results confirm that the suggested algorithm achieves suitable handoff latency and bandwidth in real-time multimedia services.

Related Work

IEEE 802.11 Handoff

The early IEEE 802.11 standard did not consider the mobility of a mobile station. Then, in the IEEE 802.11, when a mobile station loses connection with the old AP, it spends T_{detect} time to detect the disconnection with the old AP and it begins the connection process to detect the new AP. This procedure is composed of three phases; Scanning Phase, Authentication Phase and Re-association Phase. The procedure sequence is shown in Fig. 1.

IEEE 802.11 standard offers Active Scanning and Passive Scanning methods as a scanning function. In the Active Scanning phase, a mobile station sends a Probe Request frame and receives Probe Response frames that contain MAC addresses.

The mobile stations need to individually scan all channels, since the IEEE 802.11 b/g uses 11 ~ 13 channels, which consume a lot of channel switching time. We define the channel switching time as Probe Response Wait time (T_{probe}). T_{auth} is the time until the mobile station receives an authentication reply from the new AP, and T_{asso} is the time until the mobile station associates the new AP with a connection.

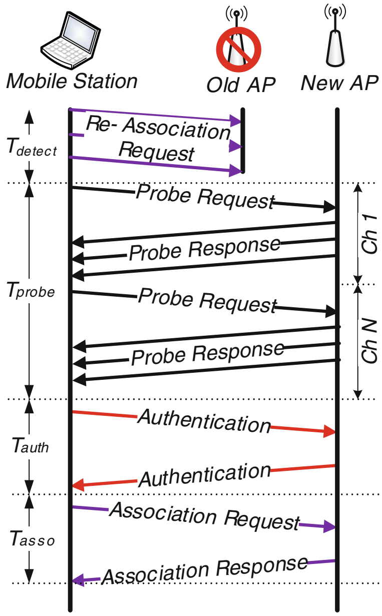


Fig. 1 The sequence of the IEEE 802.11 handoff procedure

Previous Handoff Studies

Previous handoff studies were mostly on reducing the delay time of the channel scanning procedure after a mobile station loses a connection to an AP. For example, the Selective Channel Scanning (Kim et al. 2004) and Improving Channel Scanning (Tsao and Cheng 2007) effectively search for a new channel to connect and reduce the connection time; however, the recognition time of a disconnection (T_{detect}), which has the longest delay time, was not considered.

Hence, several other studies have focused on predicting handoff time using RSSI (Wu et al. 2006, 2007; Mellimi and Rao 2008) moving pattern (Wanalertlak and Lee 2007), multi-interface (Mellimi and Rao 2008) and GPS technologies (Montavont and Noel 2006; Dutta et al. 2006), before the disconnection with the present AP occurs.

The handoff prediction algorithm, using RSSI, performs a handoff to a new AP when the RSSI value on the present AP is over the threshold or when the present AP finds other APs with better RSSI values via channel scanning (Wu et al. 2007). In addition, the handoff prediction algorithm, using the mobile station's moving pattern, decides to select the Next-AP to connect with based on AP trajectory (Wanalertlak and Lee 2007). The handoff prediction algorithm, using the GPS location information, uses the location information of neighboring AP's in the GPS server (Montavont and Noel 2006), or it uses the split method of AP service area (Dutta et al. 2006).

Problems of Related Handoff Studies

We have discussed predicting a handoff using RSSI, moving pattern, multi-interface and GPS in the previous section. Now, we reduce disconnection time using the handoff prediction method; this method still has some problems that need to be addressed. RSSI is an important metric for predicting disconnections as well as for selecting new APs. However, the RSSI value changes in the same location changes as well; it is influenced by the environment; thus, if we only consider RSSI, then the wrong handoff will easily occur. In addition, as shown in Fig. 2, RSSI has characteristics that RSSI and bandwidth values decrease while the distance increases.

Therefore, if a mobile station selects a new AP in the opposite direction against the mobile station's moving direction, then the RSSI value of the mobile station decreases rapidly. As a result, the mobile station cannot have enough bandwidth, and handoff happens again. Therefore, the mobile station must consider the moving direction of a mobile station as well as RSSI.

In order to predict a handoff using moving patterns, we must know the moving pattern of a mobile station. If a mobile station does not have any moving pattern, then the mobile station must search for a channel throughout the channel list, such

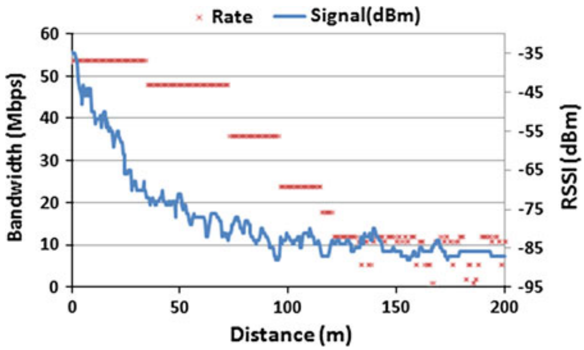


Fig. 2 The relationship between RSSI and bandwidth

as that of an IEEE 802.11 handoff, and an additional disconnection occurs. Also, since a mobile station must store all the moving history, it suffers from the overhead. Finally, in the case that a lot of mobile stations exist, the handoff prediction method using GPS creates heavy overhead for both the server and network. In addition, if a mobile station performs a handoff only at the distance between APs, then frequent handoff problems will occur in the border areas.

Location-Based Handoff Algorithm

We propose the location-based handoff algorithm which uses the GPS information to find out the location of the mobile station and get the guaranteed bandwidth using direction, velocity, RSSI of mobile station.

Considerations of Proposed Location-Based Handoff Algorithm

In order to solve the prediction handoff which only considers the RSSI and distance, we use the expected RSSI with a distance. That is, if an actual RSSI value is less than an expected value of RSSI about distance between mobile station and an AP, we exclude the AP from the handoff list.

We have performed several experiments according to the different speeds of the mobile station to set on the expected RSSI value by distance. Then, we got the Fig. 3 and Eq. (1) which can include a several path loss equations.

$$Y(dBm) = -40.4 - 20\,logD(m) \quad (1)$$

We consider the moving direction, because RSSI value rapidly decreases according to the moving direction. The mobile station is located the same distance from AP1 and AP2. An expected value of RSSI by an Eq. (1) is −40 dBm, therefore

Fig. 3 The RSSI and expected RSSI value based on speed of mobile station and a distance between a mobile station and AP

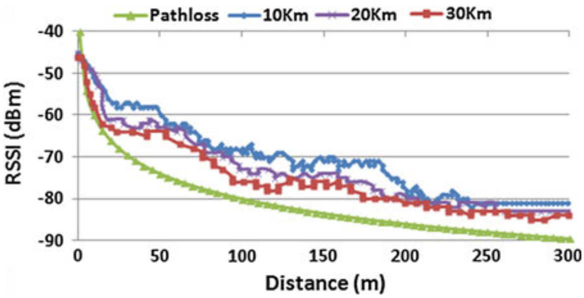
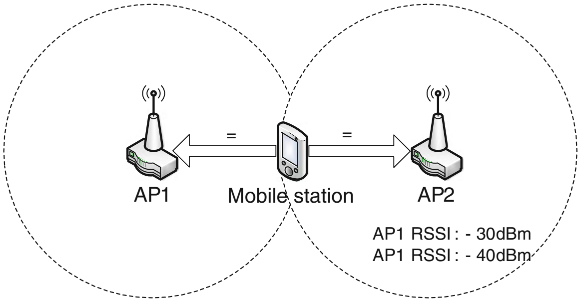


Fig. 4 The handoff considering moving direction and RSSI value



the measured value of RSSI is around −30 and −40 dBm respectively (Fig. 4). Given this situation, the mobile station needs to handoff to the AP1. However, if a mobile station moves to AP2, AP1's RSSI value dramatically decreases, and AP2's RSSI value increases. As a conclusion, a mobile station considers the moving direction when a mobile station handoffs to new AP. IEEE 802.11 handoff and simple location-based handoff do not consider moving speed, (IEEE 802.11 2007; Kim et al. 2004; Tsao and Cheng 2007; Wu et al. 2006, 2007; Wanalertlak and Lee 2007; Mellimi and Rao 2008; Montavont and Noel 2006; Dutta et al. 2006). However, in order to perform efficient handoff, we also consider moving speed.

As shown in Fig. 5, AP is installed at intervals of 50 m, we assume that an AP offers bandwidth to a mobile station up to 150 m. Then, the mobile station to communication with the AP1 moves in the AP8. Handoff occurs in the point A, where it is apart from the AP1 around 150 m. Support that moving speed is 30 km/h, an IEEE 802.11 handoff may connect to the AP5 at the point B after disconnection because of delay time. In the case of existing location-based handoff, a mobile station may handoff to the AP4 in the point A. In this time, the mobile station can be serviced from AP4 and AP5 after 18 and 24 s respectively. If the mobile station moves during 1 h using this moving pattern, the number of handoff time may become 200 and 150 respectively. Assuming that handoff delay is one second, the total duration of disconnection reaches 200 and 150 s respectively, and it takes 5.55 %, 4.16 % of the total 3,600 s respectively. At this time, if a mobile station handoffs to the AP7 with the same condition, the number of handoff may become 100 times and disconnection of 2.77 % occurs.

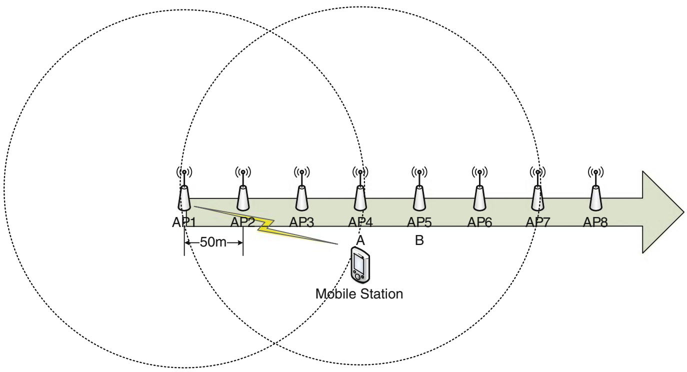


Fig. 5 The handoff considering moving speed

Proposed Location-Based Handoff Algorithm

In this section, we propose a geographical handoff algorithm using distance between mobile station and AP, moving speed, and moving direction. The algorithm is shown in Fig. 6.

A mobile station collects an AP's information from a beacon frame of the AP. $CurrentAP_D$ describes the distance from the connected AP, and $Threshold(D)$ describes the handoff boundary value. If $CurrentAP_D$ is more than $Threshold(D)$, the handoff algorithm is performed. $Threshold(D)$ depends on the bandwidth that guarantees enough service that a mobile station can receive. In other words, because a service such as multimedia video needs a high bandwidth,

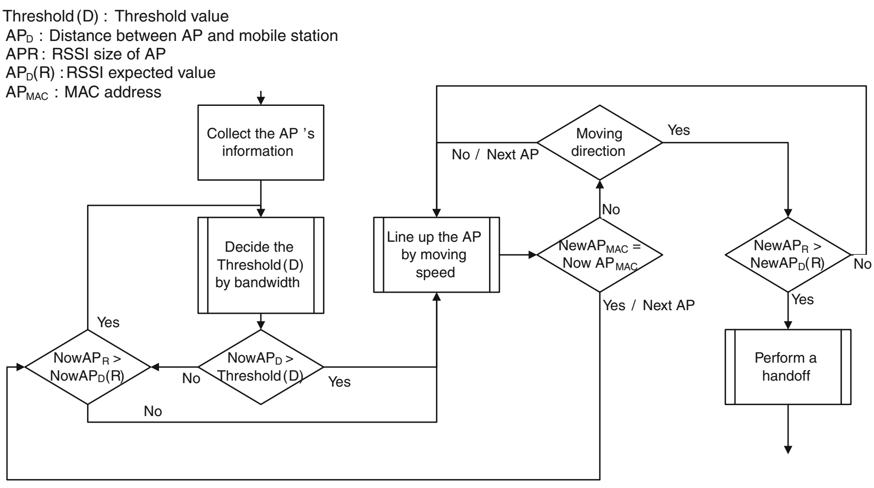


Fig. 6 Proposed location-based handoff algorithm

RSSI value must be high for guaranteeing the high bandwidth. Therefore, *Threshold*(*D*) means the distance that a mobile station can receive enough service from the AP to prevent unnecessary handoff.

When a mobile station needs the handoff by *Threshold*(*D*), it can choose an AP according to the handoff algorithm. First, AP belong to the same moving direction has higher priority than others. Second, the priority of AP is decided by the moving speed of a mobile station. Assuming that moving speed is low, the nearest AP from the mobile station has the highest priority. When moving speed is high, the nearest AP from the *Threshold*(*D*) among APs has the highest priority. In order to guarantee the bandwidth from an AP to a mobile station, we can obtain the expected value ($CurrentAP_D(R)$ and $NewAP_D(R)$) by using Eq. (1).

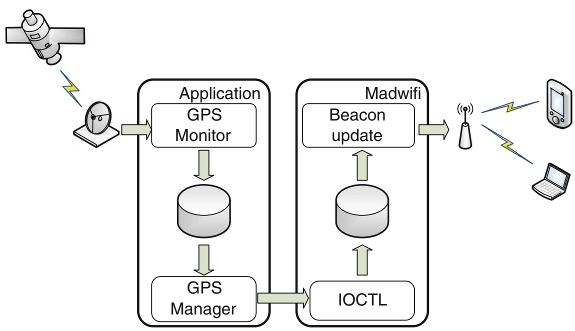
Implementation

In order to evaluate the performance of the proposed algorithm, we implemented a mobile station that performed the location-based handoff algorithm and AP that can transmit the location information via a beacon frame.

Transmission Method of Location Information

The location information is directly transmitted by GPS application program and Madwifi driver. It is shown in Fig. 7. Location information, obtained every 1 s from the GPS, is stored in the location information table by the application program, and the information is then passed to Madwifi via Input Output ConTroL (IOCTL). The Madwifi adds the location information to the end of beacon frame, and the AP transmits the beacon frame to the mobile station over 100 ms intervals. Generally, AP can transmit a beacon frame to the mobile station located out of data communication range, because it transmits a beacon frame to the mobile station at a rate of 1 Mbps.

Fig. 7 Transmission method of location information (AP)



Access Point and Mobile Station

We use the Alix 3C2 of PC Engines as an AP; it has one CPU of 500 MHz AMD Geode LX800, a 256 MB DDR DRAM, two miniPCI slots, two USB ports and a CompactFlash socket, etc. Especially, we use two LAN cards of Atheros chipset to support the service to a mobile station and to communicate between APs; moreover, we use a USB GPS in the USB port. The operating system is Pyramid Linux 1.0b6 of Linux Kernel 2.6.19.2. In order to communicate between APs, we use Optimized Link State Routing Protocol (OLSR), Internet Protocol (IP) Forwarding and Proxy Address Resolution Protocol (ARP). The mobile station uses a USB GPS with Fujitsu E8410 and a PCMCIA LAN card by Atheros chipset. The operating system is Linux Suse 10.2. In order to receive the beacon frame that contains the neighbor AP location information, we use the modified Madwifi driver.

Performance Evaluation

In order to evaluate the proposed handoff algorithm, we have performed two experiments: handoff delay time measurement and transmission bandwidth measurement.

Delay Time of Channel Scanning

In order to reduce delay time during the scanning phase, we have a beacon frame including the location information of an AP and the mobile station to predict the handoff point.

In order to evaluate performance, we design the experiment environment as shown in Fig. 8. When a mobile station moves from AP1 and AP2, we collect the packet on the laptop with AirMagnet to measure the handoff delay time. As a result, we obtained the result as shown in Fig. 9.

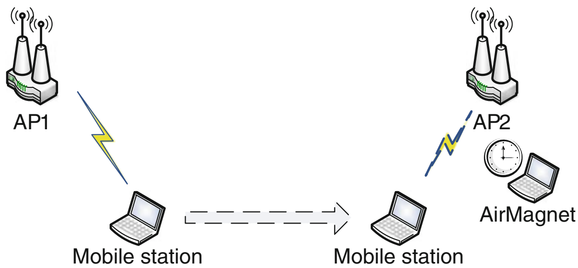
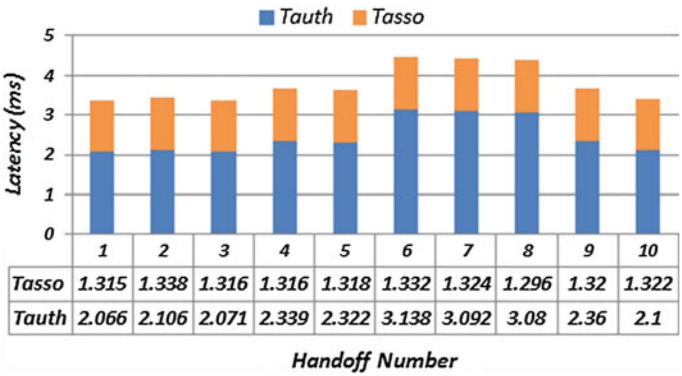


Fig. 8 The experiment environment of handoff delay time



	1	2	3	4	5	6	7	8	9	10
Tasso	1.315	1.338	1.316	1.316	1.318	1.332	1.324	1.296	1.32	1.322
Tauth	2.066	2.106	2.071	2.339	2.322	3.138	3.092	3.08	2.36	2.1

Fig. 9 The layer 2 delay time of proposed algorithm

The average handoff delay time of IEEE 802.11 is T_{probe} (334 ms), T_{auth} (2.3 ms) and T_{asso} (1.5 ms); however, the average handoff delay time of the our algorithm is T_{probe} (0 ms), T_{auth} (2.4 ms) and T_{asso} (1.3 ms).

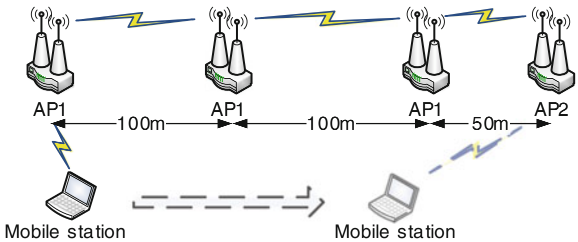
Since we know the exact location of AP, the mobile station must handoff before the connection expires. As a result, the channel-search time can be reduced by 334 ms on average. In addition, disconnection time (5.3 s) is removed by the help of the handoff prediction.

The Comparison of Transmission Bandwidth

AP1, AP2, AP3, and AP4 is located in the 100 and 50 m intervals respectively. It is shown in Fig. 10. Moving from AP1 to AP4, a mobile station uses the Iperf 1.2.0 to measure the bandwidth as Iperf client. Then, we installed the Iperf server in the APs to measure the exact bandwidth. The experiment environment is shown in Fig. 10.

The Fig. 11 represents the comparison result about our proposed location-based handoff and IEEE 802.11 based Madwifi driver's handoff. In case of IEEE 802.11, the bandwidth of a mobile station connected AP1 decreased constantly, because a mobile station connected in AP1 moves from AP1 to AP4. The connection from AP1 is broken in 180 m because the communication range of AP1 is 180 m. Then, a mobile

Fig. 10 The experiment environment of transmission bandwidth



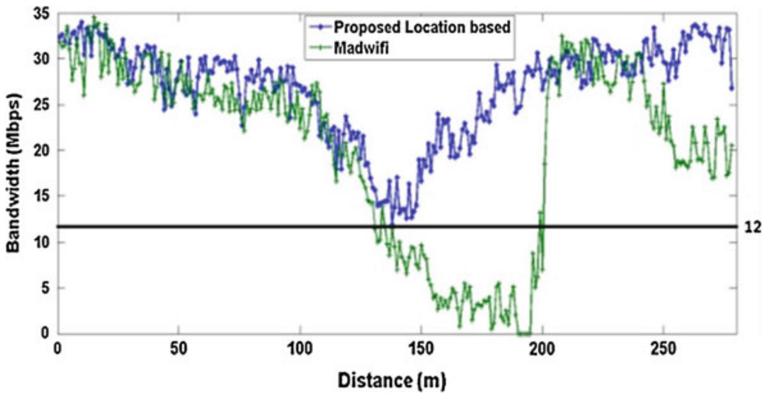


Fig. 11 The comparison of proposed location-based handoff and IEEE 802.11 handoff

station performs a handoff to AP3 which has the strongest RSSI value by channel scanning. The average bandwidth of IEEE 802.11 based handoff was 19.81 Mbps.

The Fig. 12 represents the comparison result about the proposed location-based handoff and the existing location-based handoff. Because the existing location-based handoff considers only distance, handoff was happened each AP and disconnection times happen each handoff. The average bandwidth of existing location-based handoff is 23.62 Mbps more than that of the IEEE 802.11 based handoff. On the other hand, using a proposed location-based handoff, a mobile station can detect the location information of AP by the beacon frame. A mobile station received the location information of AP4 in 60 m. Unlike existing location-based handoff, a mobile station keep a connection with AP1 to 140 m by *Threshold*(D) value that is depended on the moving speed. After that, a mobile station performs a handoff to AP4 by proposed location-based handoff algorithm. At this time, the delay by channel scanning or disconnection did not happen and the average bandwidth was 26.28 Mbps.

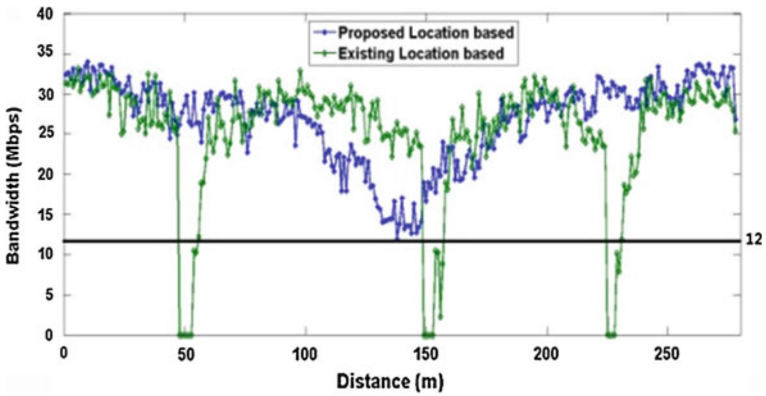


Fig. 12 The comparison of proposed location-based handoff and existing location-based handoff

Conclusion

In order to support the real-time multimedia service such as VoIP, we need the seamless handoff method which can guarantee enough bandwidth in IEEE 802.11 WLAN. Then, we had proposed the location-based handoff scheme. We used the GPS to transmit the location information of AP and we had implemented the proposed location-based handoff algorithm which uses a location information. Then, we measured the delay time and bandwidth to evaluate the performance. As a result, we can eliminate the T_{detect} which occupy a lot of time in the IEEE 802.11 based handoff, and we reduced the T_{probe} and the delay time of 98 % in ISO Layer 2. In addition to, compared with Madwifi drive and existing location-based handoff, the bandwidth of 24.8 and 4.7 % had improved respectively.

Acknowledgments This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2012-0001578).

References

Dutta A, Madhani S, Chen W (2006) GPS assisted fast-handoff mechanism for real-time communication. In: 2006 IEEE Sarnoff Symposium

IEEE 802.11 (2007) Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, IEEE 802.11-2007

ITU-T (1996) Recommendation G.114: one-way transmission time, Feb 1996

Kim H, Park S, Park C, Kim J, Ko S (2004) Selective channel scanning for fast handoff in wireless LAN using neighbor graph. ITC-CSCC 2004, July 2004

Mellimi S, Rao SV (2008) Location based fast MAC handoffs in 802.11. Wireless, mobile and multimedia network 2008

Mishra A, Shin M, Arbaugh W (2003) An empirical analysis of the IEEE 802.11 MAC layer handoff process. ACM mobile computing and communications reviews 33, Apr 2003

Montavont J, Noel T (2006) IEEE 802.11 handovers assisted by GPS information. In: Proceedings of the 2nd IEEE international conference on wireless and mobile computing, networking and communications, June 2006

Park HS, Han SH, Kim JD (2009) Vehicular client roaming and location-based handoff through multiple WLAN APs in a container terminal. ICHIT2009, Aug 2009

Tsao SL, Cheng YL (2007) Improving channel scanning procedures for WLAN Handoffs. Lecture notes in computer science, vol 4809

Wanalertlak W, Lee B (2007) Global path-cache technique for fast handoffs in WLANs. ICCCN 2007

Wu TY, Lai CC, Chao HC (2006) Efficient IEEE 802.11 handoff based on a novel geographical fingerprint scheme. Wireless communications and mobile computing

Wu H, Tan K, Zhang Y, Zhang Q (2007) Proactive scan: fast handoff with smart triggers for 802.11 wireless LAN. IEEE INFOCOM 2007

Reconfiguring Multi-Rate Wi-Fi Mesh Networks with Flow Disruption Constraints

Yafeng Zhou, Sang-Hwa Chung and Han-You Jeong

Abstract Wi-Fi mesh networks (WMNs) have been considered as a promising data-networking infrastructure in port-logistics environments. In such networks, it has been known that user traffic has a periodic characteristic due to diurnal cycle of human activities. To cope with these dynamics, a WMN has to adaptively reconfigure its settings, such as an activated link, an assigned channel, and a route of a flow. However, a careless reconfiguration of a WMN may lead to a heavy disruption of the on-the-fly traffic, hence resulting in the degradation of a WMN performance. Therefore, in this paper, we focus on the *σ-constrained WMN reconfiguration* (*σ-WMNR*) problem defined as follows: Given the current WMN configuration, the goal of σ-WMNR is to find a new WMN configuration suitable for the upcoming traffic estimate under a constraint that the number of disrupted flows must be limited to a predefined fixed value σ. To address this σ-WMNR problem, we propose a heuristic algorithm, called min–max WMN reconfiguration (MM-WMNR). From the numerical results, we show that our approach outperforms the static WMN configuration algorithms by 25 %.

Y. Zhou (✉) · S.-H. Chung
Department of Computer Engineering, Pusan National University,
Pusan, Republic of Korea
e-mail: flyzyf@hotmail.com

H.-Y. Jeong
Institute of Logistics IT, Pusan National University,
Pusan, Republic of Korea

H.-J. Kreowski et al. (eds.), *Dynamics in Logistics*, Lecture Notes in Logistics,
DOI: 10.1007/978-3-642-35966-8_32,

Introduction

A multi-hop extension of Wi-Fi networks, called the *Wi-Fi mesh network* (*WMN*) (Wu et al. 2006), has recently received a considerable attention as a municipal wireless-networking infrastructure, due to its advantages over the existing networks in terms of installation cost, robustness, ease of maintenance, and so on. deployed as a networking infrastructure for supporting many port-logistics applications, for example, delivery of a RFID tag identifier, real-time location system (RTLS), etc. A crucial problem of such networks is to decide which links to activate, which channel to assign to an activated link, and which route to establish for a given flow, which is known as the *WMN configuration*. Most of WMNs employ a *static* WMN configuration, where a WMN configuration does not change once it is fixed.

However, it has been shown that the user traffic has a periodic characteristic due to diurnal cycle of human activities (Meng et al. 2004). Then, a WMN configuration designed for a traffic demand may not be good enough for another. In this case, it would be better to periodically update a WMN configuration according to the dynamic traffic demand, which is known as *WMN reconfiguration*. Given a current WMN configuration, the goal of the WMN reconfiguration is to find a new WMN configuration consisting of link activations, channel reassignments, and route changes. However, a careless reconfiguration of a WMN may lead to a *flow disruption*, where a fraction of user traffic on the fly may experience an excessive delay or be discarded due to the WMN configuration inconsistencies during the reconfiguration. These inconsistencies may include (1) disconnection of all paths traversing a link due to changes in link activations; (2) short disconnection of frame delivery due to changes in channel assignment; and (3) disconnection of frame delivery due to changes of routing. In (Balachandran et al. 2009; Franklin et al. 2010; Kanagasabapathy et al. 2010) the authors present a channel and flow assignment algorithm to control the impact of the changes in channel assignment. However, their algorithm is still limited in the sense that it does not address the other two inconsistencies. To control the extent of all flow disruptions, we define a new WMN reconfiguration problem, called the *σ-constrained WMN reconfiguration* (*σ-WMNR*) problem, as follows: Given the current WMN configuration, the goal of σ-WMNR problem is to find a new WMN configuration suitable for the upcoming traffic estimate under a constraint that the number of disrupted flows during reconfiguration must be limited to a predefined fixed value σ.

In this paper, we present a new heuristic algorithm for the σ-WMNR problem, called the *min–max WMN reconfiguration* (*MM-WMNR*) algorithm, which is decomposed into three sub-algorithms: a *routing discovery algorithm* (*RDA*), a *re-routing algorithm* (*RRA*), and a *channel assignment algorithm* (*CAA*). First, the RDA sequentially extracts a *new flow* (existing in the upcoming traffic estimate only) in the descending order of its estimated bandwidth. Then, the RDA determines its route based on a dual link-cost metric: one for the residual bandwidth of an *activated* link, and the other is for the activation of an *inactive* link. Among the flows which traverses the maximally congested link, the RRA selects one in the descending order of the bandwidth, and then reroute it until the traffic load of the maximally congested link is reduced under the constraint of σ disrupted common flows (existing in both the current and the upcoming traffic estimates). Finally, the CAA assigns a channel to each link in such a way that minimizes the maximum traffic load of each channel. From the numerical results obtained from simulation, we validate that the MM-WMNR algorithm achieves 25 % higher throughput than the static WMN configuration algorithm.

The remainder of this paper is organized as follows: In section Network Model and Problem Definition, we introduce the network model and the problem formulation. Then, we present the details of σ-WMNR algorithm in section Min–Max WMN Reconfiguration Algorithm. In section Numerical Results, we validate the performance of σ-WMNR algorithm with the numerical results. Finally, we conclude this paper in section Conclusion.

Network Model and Problem Definition

In this section, we first describe an application of WMN in port-logistics environments, discuss the WMN model and assumptions and then formulate our σ-WMNR problem.

Applications of WMNs in Port-Logistics Environments

In port-logistics environments, two representative applications are asset tracking and video surveillance. Asset tracking allows you to locate and manage assets. To aim these, RFID data identified using RFID system is delivered to RFID/RTLS middleware via networking infrastructure. Video surveillance enables you to 24 h monitor each corner of the port to prevent theft of goods and enhance container security. To achieve this, video information is delivered to video surveillance recorder via networking infrastructure. In this paper, we focus on the networking infrastructure issue in port-logistics environments.

Figure 1 shows an application of WMN in port-logistics environments. RFID readers identify each tagged asset and deliver the identified information to associated mesh router. Each mesh router then delivers RFID data and video information to mesh gateway. Mesh gateway then deliver these video information to a video surveillance recorder via wired network infrastructure.

Network Model

A WMN is modeled by a graph G = (V, E), where V = {1, 2,…, N} represents the set of mesh routers and E = $\{(m,n)|m,n \in V\}$ V} is the set of wireless links. The set of available channels in a WMN is denoted by $\mathbb{C} = \{1, 2, \ldots, C\}$. Each

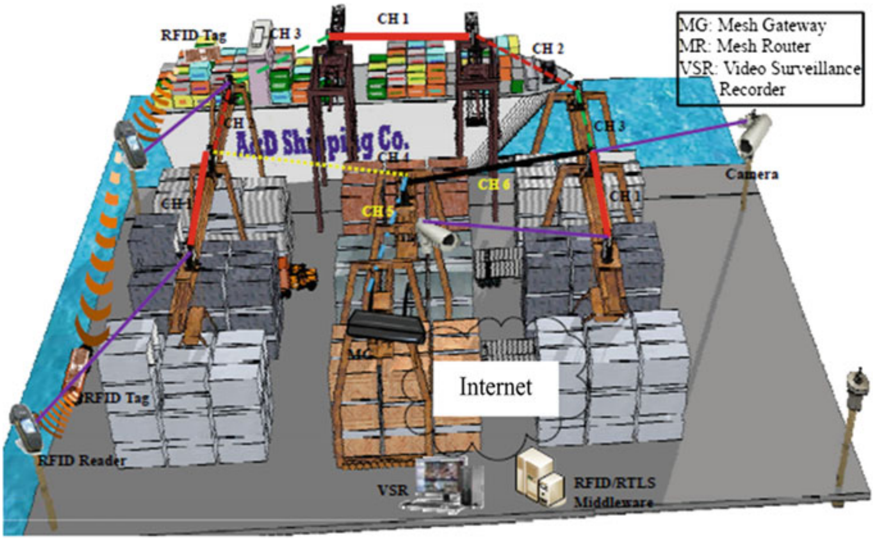


Fig. 1 An application of WMN in port-logistics environments

mesh router is assumed to be equipped with I radios, each of which operates on a distinct channel. Two neighboring mesh routers can establish an active link if each of them has a radio tuned to the common channel.

The data rate of a mesh link (m, n) determined by the distance between mesh router m and n is denoted by r(m, n). For each mesh link $(m, n) \in E$, we define the interference set I(m, n), which contains the potentially interfering mesh links (p, q) to (m, n).

In this paper, we assume the traffic demand varies every T interval. The traffic demand in current interval is represented by a N × N matrix T^C, in which each element t_{sd} $(1 \leq s, d \leq N)$ is the estimated bandwidth for flow f_{sd} originating from source router s to destination router d. Let $F^C = \{f_{sd} | t_{sd} > 0\}$ denote the set of flows of which the traffic intensity is non-zero in T^C. Likewise, T^N denotes new traffic demand in next interval and F^N denotes the set of flows of which the traffic intensity is non-zero in T^N. Let P_{sd} denote the routing path for flow f_{sd}.

We assume CSMA/CA is employed as the MAC protocol in WMNs. In addition, the “saturation throughput” of CSMA/CA is set to 0.68 as shown in (Bianchi 2000).

We assume there is a centralized agent called *reconfiguration server*(*RS*) in WMNs. RS gathers network connectivity information after a WMN is deployed, estimates upcoming traffic, then computes a new WMN configuration and finally sends it to each mesh router. Each mesh router updates the network following received network configuration. To communicate between RS and each mesh router, we assume they have an additional radio operating a default control channel.

Problem Formulation

A part of on-going user traffic maybe disrupted during the WMN reconfiguration phase. For instance (1) deactivation of a link may disconnect all paths which traverse along this link; (2) channel switching leads to “failure” of a link for an interval during which the data transmission on this link stops; (3) routing rediscovery causes the disconnection of a flow until a new route is established for this flow; in addition, the flooding messages utilized by routing discovery process may affect the on-going transmission. We refer to the above cases as flow disruption.

It has been shown in Subramanian et al. (2008), that the channel switching delay is in the order of milliseconds. By contrast, since flooding control messages is employed during routing discovery process, it takes much longer time to establish a path. Deactivation of a link may lead to routing rediscovery. Moreover, the larger the number of current transmission flows which change routes, the larger the number of flows which are disrupted. If multiple routes are required to rediscover, due to collision among control messages, it takes even longer time to complete a routing discovery process. Consequently, the on-going transmission is heavily disrupted, which in turn degrades WMN performance. Therefore, to cope with the flow disruption during WMN reconfiguration process, we restrict the number of disrupted flows to a given threshold σ in this paper.

Based on the above analysis, we define a new WMN reconfiguration problem, called σ-WMNR problem: Given the estimated traffic, the current network configuration, our objective is to find a new network configuration such that (1) the number of flows which change routes mush not exceed a predefined value σ, (2) for each flow $f_{sd} \in F^N$, traffic be routed in proportion to its demand t_{sd}. In other words, we consider the problem of maximizing λ such that for each flow, at least λ fraction of its demand be routed.

Min–Max WMN Reconfiguration Algorithm

To further analyze the causes of flow disruption during WMN reconfiguration, we first divide the flows in T^C and T^N into three groups: (1) deletion flow group (DFG) $F_{DG} = F^C - F^N$; (2) new flow group (NFG) $F_{NG} = F_N - F_C$; (3) common flow group (CFG) $F_{CG} = F_C \cap F_N$. For each flow in NFG, routing discovery is inevitable. In other words, there is little possibility to control the flow disruption. In addition, for each deletion flow in DFG, the routing entry is automatically deleted which has no impact on current flow transmission. On the other hand, the routing rediscovery process for a common flow will disrupt the on-going transmission of this flow due to lack of new route. If every common flow changes its routing path, the network state needs longer time to get steady which in turn will heavily disrupt the on-going transmission of each common flow. Therefore, we restrict the number of routing changes for flows in CFG to a given threshold σ in this paper.

As stated before, our goal is to maximize λ. To achieve this, the maximal interference on any channel should be minimized. The interference on a link (m, n) for a channel i is denoted by:

$$Int(i) = \max_{(m,n)\in E} \frac{A(m,n,i)}{r(m,n)} + \sum_{(p,q)\in E} \frac{A(p,q,i)}{r(p,q)} \tag{1}$$

where A(m, n, i) denotes the aggregated traffic on link (m, n) which is assigned channel i. To minimize the maximal interference on any channel, we need to compute: (1) the aggregated traffic on each mesh link (m, n) denoted by a(m, n); (2) a suitable channel index to each mesh link (m, n) with a(m, n) > 0. To deal with the former issue, route for each flow in T^N should be found. After route is determined for each flow, re-routing of a part of flows may improve the performance. Therefore, the MM-WMNR algorithm is decomposed into three sub-algorithms: *routing discovery algorithm* (*RDA*), *re-routing algorithm* (*RRA*) and *channel assignment algorithm* (*CAA*).

Routing Discovery Algorithm

In this subsection, we propose RDA algorithm to find out route for each new flow ($f_{sd} \in F_{NG}$). As stated before, our goal is to minimize the interference on the most interfering channel. To ensure this, at a first glance, the simplest way is to evenly distribute the traffic load along each mesh link. However, too many active links will increase the congestion level due to limited number of radios and channels. On the other hand, less number of active links will cause some links to be more congested which in turn increase the congestion level of channels. Then too much consideration need to be taken into account when choosing an inactive link to route frames during routing discovery phase.

We first propose a dual link-cost metric in which the cost of a link is determined by the normalized residual link utilization. The utilization on a mesh link (m, n), denoted by u(m, n), is given by

$$u(m,n) = \begin{cases} \dfrac{\sum_{(p,q)\in\{(m,n)\cup I(m,n)\}} \frac{a(p,q)}{|C|}}{r(m,n)}, & if\,(m,n)\,is\,active \\ \eta \times \dfrac{\sum_{(p,q)\in\{(m,n)\cup I(m,n)\}} \frac{a(p,q)}{|C|}}{r(m,n)}, & otherwise \end{cases} \tag{2}$$

In (2), a(m, n) denotes the traffic load on mesh link (m, n) [note that a(m, n) = 0, if (m, n) is inactive]. To differentiate inactive mesh link from active link, a *penalty constant* denoted by l is associated to each inactive mesh link. Essentially, the physical meaning of u(m, n) is the channel utilization of the channel which is assigned to mesh link (m, n). However, the channel information is unknown during this step; we just assume that the aggregated traffic load within a interference set is evenly distributed over each available channel. Based on this assumption, utilization on a mesh link is obtained. Then normalized coefficient is obtained by scaling up the maximal link utilization to a predefined value τ ($\tau < 1$). In addition, the normalized link utilization is computed by multiplying utilization by τ. Finally, the cost of each link is the reciprocal of the residual normalized link utilization.

Algorithm 1 Routing Discovery Algorithm (RDA)

STEP I	Choose a flow f_{sd} with the largest traffic load t_{sd} and which have not found a routing path. Terminate this algorithm, if f_{sd} exists; otherwise, go to STEP II.
STEP II	Based on the traffic load on each mesh link a(m, n), compute u(m, n).
STEP III	Compute the normalized coefficient κ. Through this, the largest normalized utilization is equal to a predefined value τ ($\max_{e \in E} u(m,n) \times \kappa = \tau$).
STEP IV	Obtain the normalized utilization on each mesh link (m, n) denoted by μ(m, n) (μ(m, n) = μ(m, n) * τ). Then the cost of each mesh link is the reciprocal of the residual normalized utilization of the link.
STEP V	Employ Dijkstra's algorithm to find out a routing path with the smallest summation of link-cost.
STEP VI	Update a(m, n) of each link m, n) within P_{sd}, a(m, n) = a(m, n) + t_{sd}.

Based on the dual link-cost metric, RDA is employed to find a path for each new flow. The detailed RDA is presented in Algorithm 1.

Re-routing Algorithm

After the RDA step, route for each flow in the upcoming traffic estimate is computed. However, during RAA step, each flow independently chooses its route based on current network state and does not consider the future routing discovery for remaining flows. This situation may cause uneven distribution of traffic load in the network. To solve this, in this subsection, we present RAA algorithm to change routes for a part of flows:

The RAA is summarized in Algorithm 2

Algorithm 2 Re-routing Algorithm (RRA)

STEP I	Let findFlag(m, n) = 0 (a(m, n) > 0) and findFlow(f_{sd}) = 0. Let num_Change = 0.
STEP II	Choose the link (m, n) with largest μ(m, n) and findFlag(m, n) = 0. Terminate the algorithm, if (m, n) exists; otherwise, go to STEP III.
STEP III	Choose the flow f_{sd} traversing along active link (m, n) and with largest t_{sd}, and findFlow(f_{sd}) = 0. If flow f_{sd} exists, go to STEP IV; otherwise, go to STEP II.
STEP IV	If flow $f_{sd} \in F_C$ and num_Change == σ, go to STEP III; otherwise, go to STEP V.
STEP V	Dijkstra's algorithm is employed to find out a new route for f_{sd}. Find out the active link (p, q) with the largest utilization in the new path. If μ(p, q) < μ(m, n), routing is adjusted; otherwise, the current routing path is kept.
STEP VI	If routing is not adjusted for flow f_{sd}, go to STEP III.

Channel Assignment Algorithm

After the above two steps, a set of eligible links is determined by choosing those links of which the link utilization μ(m, n) is greater than zero. Then CAA assigns channels to each eligible link.

The objective in channel assignment step is to minimize Int(i) for any channel i. Since there is limited number of radios on each mesh router and limited number of channels, to achieve the goal, the channel assignment algorithm should determine: (1) to which active mesh link, a dedicated channel is assigned; (2) to which mesh link, a shared channel is assigned; (3) on which criterion, a channel is chosen.

The CAA is summarized in Algorithm 3

Algorithm 3 Channel Assignment Algorithm (CAA)

STEP I	Initialize findFlag(m, n) to zero for each eligible link.
STEP II	Choose a eligible link (m, n) with largest μ(m, n) and findFlag(m, n) = 0. Terminate the algorithm if (m, n) does not exist; otherwise go to STEP III.
STEP III	Assign the least congested channel i within I(m, n) to (m, n).
STEP IV	Let findFlag(m, n) = 1, go to STEP II.

Numerical Results

In this section, we evaluate the performance of our proposed network reconfiguration algorithm. We first address the experiment parameters, and then discuss the performance of our proposed network reconfiguration algorithm.

Experiment Parameters

30 nodes are randomly placed in 700 × 700 m square area. We use 802.11a PHY layer. Each node is equipped with six radios, and there are 12 available channels.

Since the link rate used by any two neighboring mesh routers for communications is determined by the distance between them, in the experiments, the link rates 6, 18, 36 and 54Mbps are considered and the corresponding transmission ranges are set to 238, 178, 119 and 89 m, respectively. The interference range is set to 300 m.

In the experiments, traffic matrix T is randomly generated as follows: each traffic matrix consists of predetermined number of flows, the source and destination of each flow is randomly chosen among 30 nodes, traffic intensity of each flow is uniformly chosen in interval (0Mbps, 1Mbps). Parameter σ is set to ten and \$\tau\$ is set to 0.8.

Performance of MM-WMNR Algorithm

In this section, we evaluation the performance of our proposed heuristic algorithm based on the numeric results obtained from the experiments. The core idea of the SNC algorithm used in the experiment is as follows: (1) shortest path routing is employed, in which link cost is in inverse relationship to the link rate; (2) channel is assigned to each mesh link according to the number of routed flows.

We first show the effect of different penalty constants. Figure 2 shows the performance impact of different penalty constants. In the experiments, each traffic matrix consists of 174 flows (1/5 of the maximum number of flows in the network) and the each point is the average measurement of 15 different traffic matrices. We observe that as the value of the penalty constant increases, the performance deteriorates. This is due to the fact that the smaller the value of the penalty constant, the more the preference to choose an inactive mesh link to be activated which in turn decrease the maximal link utilization in the network. In additional, the traffic is preferred to aggregate on the higher-rate link in the case of smaller penalty constant, which also decrease the link utilization.

We next show the performance of our MM-WMNR algorithm under different traffic density. Figure 3 plots the network throughput versus traffic density. We compare the obtained scaling factor of MM-WMNR and SNC under 1/5 (174 flows), 2/5(348 flows), 3/5(522 flows), 4/5(696 flows) and 5/5(870 flows) of the maximum number of flows in the network. For high traffic density (4/5, 5/5), the throughput is almost the same between MM-WMNR and SNC; while for low traffic density (1/5), the obtained throughput by MM-WMNR is 25 % higher than

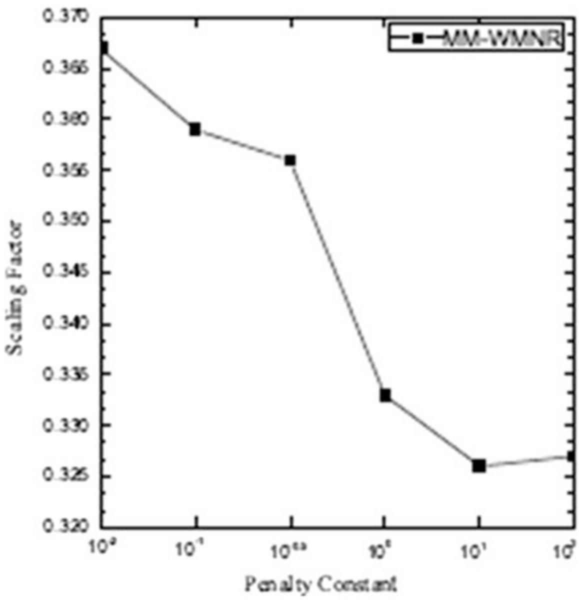


Fig. 2 Performance impact of different penalty constants

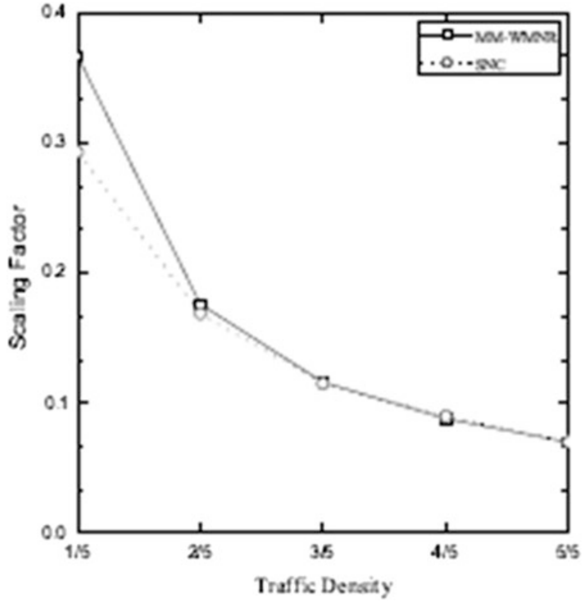


Fig. 3 Network throughput versus traffic density

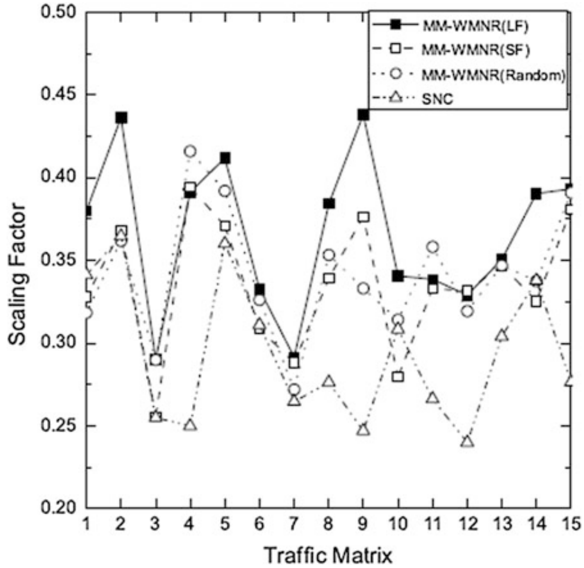


Fig. 4 Overall throughput versus traffic matrices

that of SNC. This is because that most of the mesh links become congested when traffic density is higher, then there is little chance to decrease the congestion level in the network.

We finally compare the performance of our MM-WMNR algorithm and SNC. For MM-WMNR algorithm, we compare the performance of three different flow choosing strategies which is used in the RDA step: *large traffic load flow first* (*LF*), *small traffic load flow first* (*SF*) and random. In this experiment, the constant penalty is set to 0.01 ($\tau = 0.01$). Figure 4 plots the network throughput across traffic matrices. We observe that MM-WMNR with LF shows the best throughput, SNC shows the worst throughput. Moreover, the throughput gain is almost the same for MM-WMNR with SF strategy and with Random strategy.

Conclusion

In this section, we evaluation the performance of our proposed heuristic algorithm based on the numeric results obtained from the experiments. The core idea of the SNC algorithm used in the experiment is as follows: (1) shortest path routing is employed, in which link cost is in inverse relationship to the link rate; (2) channel is assigned to each mesh link according to the number of routed flows.

Acknowledgments This work was supported by the Grant of the Korean Ministry of Education, Science and Technology. (The Regional Core Research Program/Institute of Logistics Information Technology).

References

Balachandran A, Franklin AA, Murthy CSR (2009) Integrating traffic estimation and dynamic channel reconfiguration in wireless mesh networks. In: International conference on high performance computing (HiPC), Kochi, India

Bianchi G (2000) Performance analysis of the IEEE 802.11 distributed coordination function. JSAC 18(3):535–548

Franklin AA, CSR Murthy, Balachandran A, Marina M (2010) Demand based state aware channel, reconfiguration algorithm for multi-channel multi-radio wireless mesh networks. In: Proceedings of IEEE INFOCOM, San Diego, CA, USA

Kanagasabapathy A, Franklin AA, Murthy CSR (2010) An adaptive channel reconfiguration algorithm for multi-channel multi-radio wireless mesh networks. IEEE Trans Wireless Commun 9(10):3064–3071

Meng X, Yuan Y, Wong S, Lu S (2004) Characterizing flows in large wireless data networks. In: Proceedings of ACM MobiCom'04, Philadelphia, Pennsylvania, USA

Subramanian A, DAS S, Gupta H, Cao J (2008) Minimum interference channel assignment in multi-radio wireless mesh networks. IEEE Trans Mob Comput 7(12):1459–1473

Wu H, Tan K, Chen J, Zhang Q, Yang F, Zhang A (2006) Distributed channel assignment and routing in multi-radio multi-channel multi-hop wireless networks. IEEE JSAC, Special Issue on Multi-Hop, Wireless Mesh Networks 24(11):1972–1983

Maintenance Work Order Based on Event-Driven Information of Condition-Monitoring Systems—A Genetic Algorithm for Scheduling and Rescheduling of Dynamic Maintenance Orders

Marco Lewandowski, Oliver Siebenand, Stephan Oelker and Bernd Scholz-Reiter

Abstract By being able to calculate a machines remaining lifetime, event-driven maintenance actions can be initiated and scheduled ad-hoc into the existing working plan. Today's maintenance management systems work mainly based on regular and cyclic services. This paper describes a new methodology to prioritize event-driven maintenance workloads driven by information from condition monitoring systems. It proposes an approach to rebuild an existing working plan automatically exploiting genetic algorithms. Accordingly, the paper gives a valuable insight regarding requirements for the integration of condition monitoring systems into maintenance operation and a possible approach to solve this issue.

Introduction

Scheduling problems are part of the many problems of logistics operation (i.e. Karger et al. 1998). All problems have one thing in common that tasks must be allocated chronologically to scarce resources. Regarding their form they are very diverse. The variants differ significantly in their complexity.

Through the progressive use of new technologies, methods and strategies in the maintenance of machinery and equipment, new scheduling methods have to be developed. By implementing and using condition monitoring techniques and systems in machines or certain components, the change from reactive maintenance

M. Lewandowski (✉) · S. Oelker · B. Scholz-Reiter
BIBA—Bremer Institut für Produktion und Logistik GmbH,
Hochschulring 20, 28359 Bremen, Germany
e-mail: lew@biba.uni-bremen.de

O. Siebenand
REHAU AG + Co, Rehau, Germany

H.-J. Kreowski et al. (eds.), *Dynamics in Logistics*, Lecture Notes in Logistics,
DOI: 10.1007/978-3-642-35966-8_33,

strategies to preventive ones has been facilitated. Current maintenance research focuses on failure prediction for special components or the whole machine based on the online monitoring of machinery data.

However, the subsequent step has been largely ignored. The automated scheduling of maintenance actions based on the findings of the condition monitoring systems and taking into account existing maintenance plans will be discussed in this paper using a genetic algorithm.

The remaining paper is organized as follows. In the next section, an overview of topical aspects regarding maintenance management and condition monitoring is provided in order to define the requirements for the algorithm to be developed. Then, the proposed approach and method is sketched out and essential new aspects are detailed. First results of the application are provided in section Results. Concluding remarks and future research directions are finally presented.

Theory of Maintenance Management and the Role of Condition Monitoring

Objectives of Maintenance Planning and the Use of Condition Monitoring

Efficiency in processes in production and logistics have reached a high relevance. Modern methods and technologies for instance strongly influence and rationalize core processes. The ongoing standardization supports the continuous automation of processes. According to that, the degree of saturation regarding certain branches is very high, so that the optimization of supporting processes like the maintenance of technical equipment will play an important role to gain further efficiency (Mobley 2002; Mobley et al. 2008).

Accordingly, the main objectives and constraints of a maintenance plan have to be known in order to enable efficient processes. The aspects regarding the preparation of the plans must be taken into account and form a tense triangle according to Fig. 1. Minimizing the number of maintenance operations precludes maximizing the utilization of wear tolerances. In contrast, component failures and possible consequential damage have to be minimized. The following figure illustrates an enhanced tense triangle focusing on the one hand on the planning objectives and on the other hand the additional constraints for scheduling and planning.

Therefore ideal maintenance time must be determined to come close to the theoretical optimum, which is formed by the superposition of the cost function of a preventive maintenance and the cost function by a failure. The frequency of preventive maintenance actions accordingly influences the failure probability and reliability. Condition monitoring technology attempts to calculate the times that are ideal for preventive measures before a failure occurs. Additionally, if good data is provided, actions can be performed as late as possible.

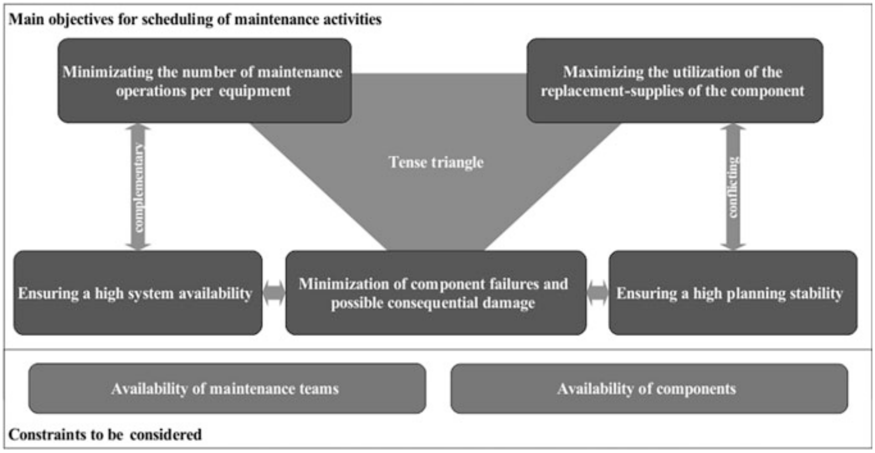


Fig. 1 Objectives and conditions in the preparation of maintenance plans

Condition Monitoring

Regarding condition monitoring systems, modules and components of machinery are elementary equipped with sensors. The aim is to appraise the condition of these elements by analyzing the sensor data. In case of any abnormalities, corresponding actions can be taken to rebuild the condition of the particular machine (Mobley 2002; Mobley et al. 2008; Byrne et al. 1995). Moreover, continually processing of the condition data promises to provide sound estimates of the residual life and improves the certainty in planning for maintenance. The impulses for the further process chains are consequently given by the machine itself so that such systems can be considered as a contribution to the paradigm shift to autonomous cooperating logistic processes (Scholz-Reiter et al. 2004). By this means, the systems support adaptive logistics processes including autonomous capabilities for the decentralized coordination of autonomous logistics.

Data Acquisition

Data acquisition is the basis for the further prognosis and diagnostic procedures. Two groups of data can be distinguished: the event data and condition monitoring data. The former are records of incidents of the past, such as preventive replacement of components, failures, adjustments or repairs (Heng 2009). The significance of these data, particularly for predicting the failure behavior, is often underestimated (Jardine et al. 2006; Vlokl et al. 2002). The reason for this is the fact that the monitoring of the state via condition monitoring technology might not detect changes over a long period.

The raw data obtained from the state monitoring requires further processing. The types of sensor data can be grouped into data that consist of single values (oil condition, temperature, pressure, humidity), wave-like data (pressure, acoustic measurements) and multi-dimensional data (thermography and X-ray images) (Jardine et al. 2006).

Condition Diagnosis and Prognosis

The condition diagnosis and prognosis is based on the obtained and processed sensor data and provides the basis for the scheduling decision. The state forecast for a predictive scheduling is critical. Compared with the diagnosis, the development of forecasting methods in connection with condition monitoring have not progressed so far (Heng 2009; Jardine et al. 2006; Dragomir et al. 2009; Goode et al. 2000). However, a variety of statistical, AI and model-based approaches have been applied, but the results quality varies (Caesarendra et al. 2010). The existing methods differ between classical reliability approaches that are only based on event data, approaches that are only based on state data, and those that combine both data (Dragomir et al. 2009). The latter are able to make long-term predictions and thus provide an improved basis for optimal scheduling. Extensive literature reviews of forecasting methods are provided e.g. by Heng (2009), Jardine et al. (2006) Dragomir et al. (2009), Caesarendra et al. (2010).

Scheduling

It is possible to describe the problem as a machine scheduling problem (Brucker 2001). Disposable maintenance personnel is considered as a machine $M_j(j = 1,\ldots,m)$, taking the components to be repaired as orders $J_i(i = 1,\ldots,n)$. The relationship between orders and machine leads to a single-machine problem (one maintenance team), a parallel machine problem (multiple maintenance teams) or a shop problem (maintenance measure consists of several operations that are performed by different teams).

Developed Method

The developed method is an approach for scheduling maintenance workload based on variant data sources, among others by including forecast data through condition monitoring system as well as statistical data. The developed algorithm is a genetic algorithm using specific operators for the actual problem. The following Fig. 2 illustrates the general idea reaching from the input data to the final schedule at minimum total costs. The several steps of the flow chart are explained in the following sections.

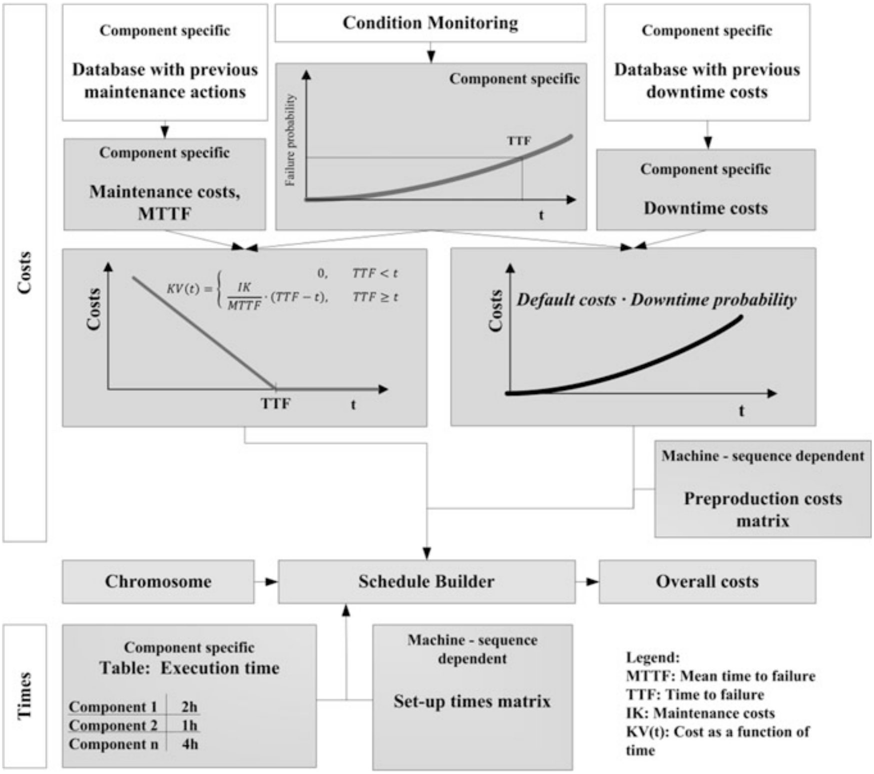


Fig. 2 General idea of the schedule building method and the determination

Objective

The target is to minimize total deviation costs incurred as a result of earliness or tardiness by the respective dates. The optimal replacement date d_i can be derived from the projected date of a failure using a condition monitoring system. If the completion date of an order is C_i, the earliness is $E_i = \max\{0, d_i - C_i\}$ and the tardiness is $T_i = \max\{0, C_i - d_i\}$. Assuming a weighting for the costs associated with the deviation, the earliness receives the factor w_i^E and the tardiness the factor w_i^T. Each maintenance action J_i can be part of a family $fml_i \subseteq \{1,\ldots,0\}$ whereas sequence dependent preproduction costs between an order i for machine g and a following order i + 1 for machine h is held in a matrix S and added to the cost term.

$$Z = \min \sum_{i=1}^{n} \left(\max\left\{w_i^E \cdot (d_i - C_i),\ w_i^T \cdot (C_i - d_i)\right\} + S_{gh}\right) \qquad (1)$$

Input

On the input side, essentially the cost functions have to be considered regarding the earliness and tardiness aspects. The costs of preventive actions are derived from the component and material costs, personnel costs and opportunity costs that arise due to the downtime in the maintenance period. The calculation is done as follows:

$$MC = CMC + ET \cdot PC \cdot TS \cdot T + DA \cdot (ET/IU) \tag{2}$$

MC: Maintenance costs [€], CMC: Component- and material costs [€], ET: Execution time [h], PC: Personnel costs [€/(Persons·h)], TS: Team size [Persons/ Team], T: Number of teams [Teams], DA: Depreciable amount [€/year], IU: Intensity of use [h/year].

To identify the costs in the event of a failure is difficult because they are based on estimates and uncertainties. A frequently chosen approach for the quantification of risks is the use of Failure Mode and Effects Analysis (FMEA). Here, a risk priority number of different probabilities is being formed. In the context of maintenance the occurrence probability of failures increases over time, as well as the risk priority number. Based on a cost-based risk priority number (Rhee and Ishii 2002) possible component failures can be assessed in monetary terms. Multiplied by the increasing probability of the failure, an imputed cost curve results. The uncertainties associated with the estimation can also be taken into account using an appropriate treatment via Monte Carlo analysis for instance.

Also groups of maintenance tasks at the same facility should be considered. In this case deviations from the optimal maintenance times have to be taken into account. In the calculation model, the costs savings based on grouping are compared to the deviations of the optimal maintenance time.

Processing

Based on the cost functions, a genetic algorithm is developed to solve the problem.

Representation of the Solution

The order of genes in the chromosome vector determines the order in the scheduling. The orders are for example coded with M003C004, which describes the machine assignment (M) and the component (C). From the machine assignment one gets the setup times, from the number part, for example, one gets the completion time.

Initializing

The initialization of the initial generation is done by generating a random permutation. If an available solution to a problem already exists, the existing schedule has to be simply updated. The initial generation is based on the existing plan and a quick rescheduling for dynamically changing ad hoc order situations is possible.

Fitness Evaluation

The algorithm provides satisfactory results when on the one hand, the reality is depicted in the objective function as accurately as possible and on the other hand, the input parameters can be determined with reasonable cost.

To reduce the complexity of the calculation by a continuous function, matrices were generated, which specify for each component (line) the dispatching of order costs on a given day (column). In the algorithm, the corresponding values for the evaluation are read from the matrices. The determination of the costs is based on the preventive cost matrix, the failure cost matrix and the preproduction cost matrix. Moreover, machine-sequence dependent setup-times and component dependent execution times are provided.

Termination Criterion

During the algorithm test, a fixed total number of generations in the range of 100–1,000 were chosen as the termination criterion. In addition, a fixed number of generations with no further improvement could serve as an alternative termination criterion. This is particularly the case with existing schedules.

Selection

The selection is made by a tournament selection (Poli 2005). The omission of an elaborate sorting is a big advantage because the chromosomes are very long. Random participants are taking part in the tournaments. The drawing of participants is without replacement until all individuals are selected. The two best individuals of a competition have the opportunity for recombination. Another advantage of the method compared to a fitness-proportional selection is, that even with the progressive approximation of the results, a differentiation between the solutions takes place (Saravanan 2006).

Recombination

For recombination, the order crossover OX3 (Cotta et al. 1998) is used. The method offers a good compromise between the important criteria, the preservation of absolute positions and the preservation of blocks. The recombination takes place in pairs, until the parents have formed a number of offspring, which corresponds to the competition size of the tournament selection.

Mutation

As standard, the inversion mutation operators, the exchange mutation and the insertion mutation are used (Kleeman and Lamont 2007; De Falco et al. 2002). In addition to these, four operators with problem-specific mutation methods have been developed. These are intended to foster in particular the group building.

Orders of the same group attract and already received connections are difficult to be separated again. When shifting jobs, they are shifted as a whole block without having to cut other conglomerates. The selection, of the following four mutation procedures, is random.

The **block fusion operator** is a targeted form of the displacement mutation. First, a random position is selected. The choice covers not only the gene, but the entire group, which is combined here. It will be randomly inserted to the right or left until the next order from the same machine.

The **inter-block gene exchange operator** swaps orders between the blocks. It is a targeted insertion mutation. In this way genes are given the opportunity to move into a group where they fit better.

The **block-shift operator** allows the transfer of a group, without constraining a merger. The random insertion is varied in such a way, that no other batch is cut.

The latter approach is the **block fragmentation**. It allows orders to "breakout" from the chromosome and in turn it allows other genes to be stored here.

Generational Change

The generational change takes the form of elitism. The generation of parents is completely replaced by the generation of children, but the best individual remains unchanged, to obtain the information it contains.

Output

An undelayed schedule is generated by the utilized algorithm. This corresponds to real conditions where maintenance teams operate at full capacity. Theoretically it is possible that a not permissible active or semi-active schedule is optimal. This is the case when existing overcapacity during maintenance results in a premature execution of orders. To solve this problem the algorithm should not be used to generate many small gaps in the availability calendar. Instead, visualization is provided showing clearly the areas where a large number of orders are scheduled early. The algorithm is also the place where additionally conditions, like the limited availability of equipment or spare parts can be considered.

Results

A manually generated case example is being used for the evaluation. There are 13 machines each with 20 components. Furthermore in this case there are three maintenance teams involved. 260 orders have to be arranged in chromosomes. The chosen representation allows possibilities for 3, 8×10^{516}. The computing times have been recorded. The behavior is proportional to the product of the number of generations and the population size.

The population size has been varied between 100 and 900 in four tests. The striven for target of the most possible low overall costs has already been reached from 500 individuals. Cost differences in cases with higher populations are caused by set up costs. A population of 700 or 900 individuals has a clearly positive influence on a low level of changing occurrences. In a population setting of 100 individuals the improvement ends at approximately the 100th generation. This is an indicator for a completely equalized population. This means that all consecutive iteration steps are useless. In the case of 900 individuals there are improvements taking place until a generation number of 300.

In all tests a fixed generation number has been used as the termination criterion. It shows that the interdependence between result and number of generations is only valid until a certain point. This effect is caused by the equalization of the individuals. A fixed amount of iteration steps without any improvement could be a reasonable termination criterion.

Robustness of the results is an important criterion for the quality evaluation of the solution. The results show that this target is achieved. Patterns show that the solutions are meaningful and are also an indicator for their high robustness.

Discussion and Outlook

Scheduling problems in practical situations are very complex. Usually these problems cannot be solved by classical optimization algorithms in a suitable time. Thus, we used genetic algorithms as an alternative which can be adapted to the special needs of scheduling problems in very flexible ways. The task of (re-) scheduling maintenance jobs based on relevant condition monitoring information can be solved by genetic algorithms as well.

The results of the example solution can be applied to many similar problems, for instance JIT machine capacity planning with sequence-based set-up times and job families. Stocks of spare-parts or materials are further aspects that could be easily integrated within the solutions.

Considering the range of self-organized algorithms (Scholz-Reiter et al. 2004), genetic algorithms are only the first attempt. The application of self-organization for the described problem would be thinkable.

References

Brucker P (2001) Scheduling algorithms. Springer, Berlin

Byrne G, Dornfeld D, Inasaki I et al (1995) Tool condition monitoring (TCM)—the status of research and industrial application. Ann CIRP 44(2):541–567

Caesarendra W, Niu G, Yang BS (2010) Machine condition prognosis based on sequential Monte Carlo method. Expert Syst Appl 37(3):2412–2420

Cotta C, Troya JM (1998) Genetic forma recombination in permutation flowshop problems. Evol Comput 6(1):25–44

De Falco I, Della Cioppa A, Tarantino E (2002) Mutation-based genetic algorithm: performance evaluation. App Soft Comput 1(4):285–299

Dragomir EO, Gouriveau R, Dragomir F et al (2009) Review of prognostic problem in condition-based maintenance. Annual conference of the prognostics and health management society

Goode KB, Moore J, Roylance BJ (2000) Plant machinery working life prediction method utilizing reliability and condition-monitoring data. Proc Inst Mech Eng Part E J Process Mech Eng 214(2):109–122

Heng ASY (2009) Intelligent prognostics of machinery health utilising suspended condition monitoring data. Queensland University of Technology, Brisbane

Jardine AKS, Lin D, Banjevic D (2006) A review on machinery diagnostics and prognostics implementing condition-based maintenance. Mech Syst Signal Processing 20(7):1483–1510

Karger D, Stein C, Wein J (1998) Scheduling algorithms. In: Atallah MJ (ed) Algorithms and theory of computation handbook. CRC-Press, Boca Raton

Kleeman MP, Lamont GB (2007) Scheduling of flow-shop, job-shop, and combined scheduling problems using MOEAs with fixed and variable length chromosomes. In: Dahal KP, Tan KC, Cowling PI (eds) Evolutionary scheduling. Springer, Berlin

Mobley RK (2002) An introduction to predictive maintenance, 2nd edn. Butterworth-Heinemann, Amsterdam

Mobley RK, Higgins LR, Wikoff DJ (eds) (2008) Maintenance engineering handbook, 7th edn. McGraw-Hill handbooks. McGraw-Hill, New York

Poli R (2005) Tournament selection, iterated coupon-collection problem, and backward-chaining evolutionary algorithms. In: Wright AH, Vose MD, Jong KAD, Schmitt LM (eds) Foundations of genetic algorithms 8th international workshop, pp 132–155

Rhee SJ, Ishii K (2002) Life cost-based FMEA incorporating data uncertainty. In: Proceedings of DETC. ASME2002 design engineering technical conferences. Montreal, Canada

Saravanan R (2006) Manufacturing optimization through intelligent techniques. Taylor & Francis, Boca Raton

Scholz-Reiter B, Windt K, Freitag M (2004) Autonomous logistic processes: new demands and first approaches. In: Monostori L (ed) 37 the CIRP-International seminar on manufacturing systems, pp 357–362

Vlokl PJ, Coetzeel JL, Banjevic D et al (2002) Optimal component replacement decisions using vibration monitoring and the proportional hazards model. J Oper Res Soc 53(2):193–202

A Procedure for the Selection of a Supply Network Risk Mitigation Strategy in Relational Arrangements

José B. S. Santos Júnior, Sérgio A. Loureiro and Orlando F. Lima Júnior

Abstract A proper management of the relationships among the partners in a supply network is key factor for increasing the competitive advantage in a global environment. The dynamic nature of these relationships, influenced by the heterogeneous mix of partners, increases the complexity of risk profiles and requires an appropriate approach for risk management. Supply network risk management (SNRM) consists of identification of risks, their evaluation and implementation of risk mitigation strategies in a supply network environment. This paper presents an additional procedure to select the appropriate strategy to mitigate risks in a SNRM framework. This procedure complements the Tummala and Schoenherr's SNRM framework, including a model which identifies the behavior of relationships between network partners. A test case from the literature is used to validate this approach. Finally, the results are compared with two different works that used the same case study.

Introduction

Nowadays, relationships between companies are seen as opportunities to gain complementary assets (technologies and skills). Thus, the management of relationships between organizations has become a key factor in the new economy.

J. B. S. Santos Júnior (✉) · S. A. Loureiro · O. F. Lima Júnior
Learning Lab in Logistics and Transport, Civil Engineering, Architecture and Urban Design School at the University of Campinas, Campinas, São Paulo, Brazil
e-mail: jbened@gmail.com

S. A. Loureiro
e-mail: loureiro@fec.unicamp.br

O. F. Lima Júnior
e-mail: oflimaj@fec.unicamp.br

H.-J. Kreowski et al. (eds.), *Dynamics in Logistics*, Lecture Notes in Logistics,
DOI: 10.1007/978-3-642-35966-8_34, © Springer-Verlag Berlin Heidelberg 2013

The rapid growth of different models of relational arrangements between companies such as inter-organizational networks, joint ventures, coalitions, extended enterprises, virtual enterprises, consortia, partnerships and alliances, emphasizes this relevance (Loureiro and Lima Júnior 2010).

The dynamic nature of relationships between several links in the supply networks (SN), in a global environment, requires an appropriate approach for risk management.

Supply network risk management (SNRM) consists of identification and evaluation of risks and consequent losses in a global environment, and implementation of appropriate strategies, through a coordinated approach among network partners, to mitigate these losses and to ensure the SN outcomes (Manuj and Mentzer 2008).

This work aims to present an additional procedure to select the appropriate strategy to mitigate risks in a generic supply network risk management framework. This procedure focuses on different behaviors in relational arrangements between SN partners and extent the Money-Power-Trust model (MPT Model) developed by Loureiro and Lima Júnior (2010).

This chapter is structured as follows: after the introduction, section Risk Management and Relational Arrangements shows the main concepts regarding risk management and relational arrangements. Section Procedure for Selection of Mitigation Strategy in SNRM Framework presents the complementary procedure for SNRM and a test case to validate this model. Finally, in section Conclusions and Future Research, conclusions and opportunities of future research are outlined.

Risk Management and Relational Arrangements

Risk Management

The topic of risk management has its origin in economic and financial fields. However, this concept has been expanded to other disciplines, for example, supply chain management.

Nevertheless, the definitions of risks in different disciplines are harmonic in presenting three issues: the unpredictability—the difficulty to identify and make an assertive prediction of the occurrence of an event; the decision-making—how to make the most appropriate decision in this uncertainty environment; and finally, the potential for losses—how to estimate the possible effects and losses associated with a particular event (Ritchie and Brindley 2007).

The risk and uncertainty definitions used in this work are: risks as the expected outcome of an uncertain event; uncertainties as an absence of information regarding a specific event that reflects the ambiguity surrounding the decision, its causes and possible solutions (Manuj and Mentzer 2008).

Basically, potential sources of risks are identified in accordance with: environment characteristics, industry characteristics, network configuration, network partners, organization strategy, problem specific variables and decision making (Christopher et al. 2011).

Based on these sources, some authors proposed classifications for risks. Tang (2006) classifies risks according to: risks related to operation (uncertainties in network, such as variability in demand, disruptions during the supply process, etc.); risks related to natural disasters (hurricanes, earthquakes, etc.); and, risks caused by direct human actions (war, financial bubbles, etc.). Olson and Wu (2010) present the risks sources according to two main categories: internal sources, including all the risks related to internal SN process (e.g. available capacity in supply, production and transportation process); and external sources: risks related to external events like changes in political and financial systems, disasters in nature (floods, epidemics, etc.) and changes in market regulation and competitors. This work will use the typology presented by Olson and Wu (2010).

All these sources of risks show vulnerability of supply networks in a global environment. To address these risk exposures, using adequate approaches for risk management in SN is necessary (Hallikas et al. 2004).

Essentially, a generic SNRM framework considers four steps: risks identification, analysis, evaluation and monitoring. In the first step all risks for supply chain are determined. In the analysis phase, a deep understanding of risk identification is required. The purpose of the evaluation step is to define the most appropriate management response for each risk or combination of risks. Finally, in the last step, a monitoring and control procedure has to be implemented to manage risks (Khan and Burnes 2007; Ritchie and Brindley 2007). Table 1 summarizes frameworks of SNRM from literature.

This work is based on the framework presented by Tummala and Schoenherr (2011). This framework has three phases. Phase 1 approaches three steps: risk identification, which involves the determination of potential supply chain (SC) risks; risk measurement, which determines the effects of potential SC risks and their magnitudes; and, in the last step, risk assessment, which determines the likelihood of each potential risk.

In phase 2, this framework presents two steps: risk evaluation, which defines the risk exposure and the risk acceptance for each potential risk; and risk mitigation and contingency plan, which involves the development of the mitigation strategies and the response actions.

Finally, in phase 3, in risk control and monitoring step, the action plans are monitored and corrective actions can be implemented, if it necessary.

This framework encompasses a more robust approach, considering the components and stages from the models presented by Ritchie and Brindley (2007), Matook et al. (2009), Cucchiella and Gastaldi (2006), Kleindorfer and Saad (2005).

Table 1 Frameworks of supply network risk management

Source	Framework basis
Kleindorfer and Saad (2005)	Four steps: 1. Specify the nature of underlying hazards leading to risks 2. Risk assessment 3. Manage risks 4. Mitigate risks
Cucchiella and Gastaldi (2006)	Six stages: 1. Analyze supply chain 2. Identify sources of uncertainty 3. Examine subsequent risks 4. Manage risks 5. Individualize the most adequate strategy 6. Implement supply chain risk strategy
Ritchie and Brindley (2007)	Five components: 1. Risk sources and profile 2. Risk and performance drivers 3. Risk and performance consequences 4. Risk management responses 5. Risk performance outcomes
Matook et al. (2009)	Five stages: 1. Risk identification 2. Risk assessment 3. Report and decision of risk 4. Risk management responses 5. Risk performance outcomes
Tummala and Schoenherr (2011)	Three phases: Phase 1: risk identification, measurement and assessment Phase 2: risk evaluation, mitigation and contingency plans Phase 3: risk control and monitoring

Strategies to Risk Mitigation in Supply Networks

Several authors suggest different strategies to risk mitigation in the SN context. These strategies encompass actions to avoid the disruption in the whole SN operations (supply, production and distribution processes). Focusing on the supply processes, Christopher et al. (2011) propose some strategies to mitigate risk sources: redesign the supply network, considering multiple sourcing; collaboration between sourcing parties, building trust in relationships; agility, giving quick response to unpredictable changes in demand or supply; and, creating a sourcing risk management culture, by monitoring and managing the risk profile of the SN.

Relational Arrangements

Different approaches and theories have been applied to the study of inter-company relationships. Three representative perspectives are: transaction cost, resourses based and industry structure. The common feature of these theories and approaches is the importance given to elements related to power and trust. Several classifications have been proposed based on these theories and approaches. Most of them have focused on the dyadic relationships (supplier–consumer).

All these classifications try to rank relationships between inter-companies through a scale ranging from arm′s length to vertical integration. The importance of the relationship increases according to the level of formalization of the relationship, the number of interactions between organizations, and the level of information and communication exchanged between them. However, companies are involved in several different relationships at the same time in order to manage their position in the business network (Bengtsson and Kock 2000). Some of them are characterized as competition, or cooperation or collaboration.

Competition means the conflicting relationship associated with market power (e.g. the values associated with the brand), industry attractiveness, knowledge management and ethical issues of the organization (Rajagopal 2007).

Cooperation refers to the level of willingness to work among the organizations, which is performed by the division of labour between the participants, whether subordinated or not, to the hierarchical coordination (Payan 2007; Peña and Arroyabe 2002). While collaboration is a coordinated and synchronized activity, the effort to build and maintain shared operations involves several aspects like: communication, trust, commitment, information sharing and joint planning (Barratt 2004).

Nowadays a competition-cooperation dichotomy guides the relationships between organizations (Balestrin and Verschoore 2008). Formal and informal mechanisms of reward and punishment have been adopted by many links of production networks to accomplish a better integration (Ballou 2006). These mechanisms are based on power and trust.

Power in intercompany relationship is the company ability to influence on intentions and actions of other companies. The power exercised by an agent in the supply network is directly related to his position in the SC and the segment that he belongs to (Williams and Moore 2007, Stadtler 2009).

Trust is an expectation of positive results (or non-negative) based on the expected actions of other part in an interaction characterized by uncertainty (Sahay 2003).

MPT Model

Lima (2007) proposes it is necessary to analyze the combination of three different forces in relational arrangements of supply chains: money, power and trust. For the

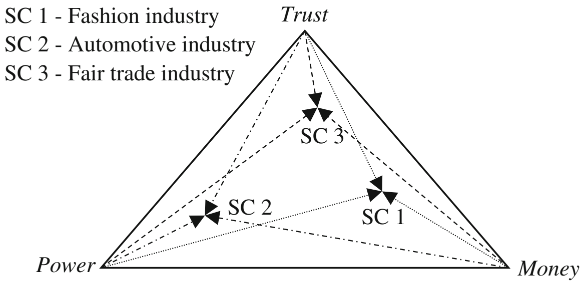


Fig. 1 MPT model (Lima 2007)

Table 2 Characteristics of the relational arrangements in supply networks

Elements	Forces		
	Money	Power	Trust
Duration	Discrete	Short and medium term	Long term
Attitude	Independent, doubtful	Independent and competitive	Open, confident and cooperative
Communication	Virtually non existent	Intensive, but predominantly unidirectional	Complex and bi-directional
Information	Owner	Owner	Shared
Planning and Goals	Individual, short-term	Individual, short-term	Group, long term
Benefits and Risks	Individual	Individual	Shared
Troubleshooting	Disintegrated	Top-down	Mutual integrated

author, these forces are responsible for the behavior of agents in the SC. Figure 1 illustrates the MPT (money, power, trust) model proposed.

Thus, in addition to the relationships based on trust and power, some agents have the predominance of money exchange for goods or services between them.

Benton and Maloni (2005) emphasize the perception described above. According to the authors, the relationships in supply chains vary from one dimension of discrete operations to integrated operations. The elements related to interactions between agents are: communication, planning and goals, benefits and risks, information sharing and troubleshooting.

Table 2 shows the combination between the set of the elements presented by Benton and Maloni (2005) and the forces proposed by Lima (2007).

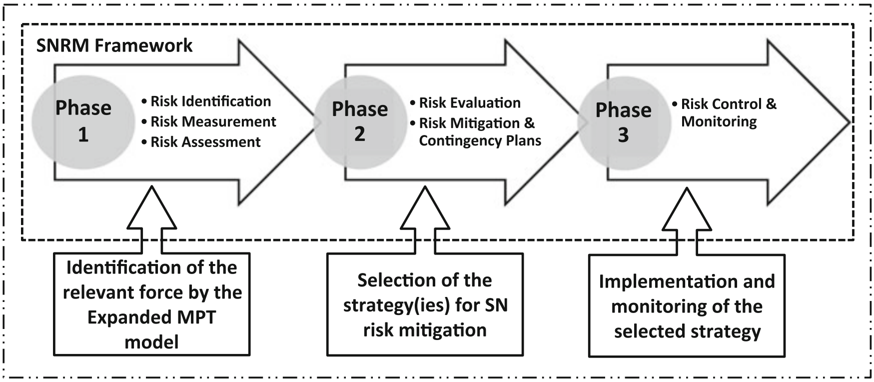


Fig. 2 Procedure for selection of risk mitigation strategy

Procedure for Selection of Mitigation Strategy in SNRM Framework

This work proposes a procedure to assist the selection of mitigation strategy in the SNRM framework, specifically in the step "risk mitigation and contingency plan" (phase 2 of the Tummala and Schoenherr's framework). In this step the most suitable strategy must be identified and selected to mitigate SN risks. The procedure suggests the application of the MPT model, extended for risk management approach, to identify the predominant force in relationships to support the selection of mitigation strategy for risks.

During the risk assessment (the last step of Phase 1 in the framework of Tummala and Schoenherr 2011) it is necessary to identify what force is the most relevant in the SN relationships (according to Table 2). Afterwards, the most suitable(s) strategy(ies) must be selected. The whole procedure is summarized in Fig. 2.

The selection of mitigation strategy results in risk-taking choice for the operation in a SN. This choice has a strong correlation with investments and resource allocation in a SN. So, a structured procedure that aids the selection of the most appropriate mitigation strategy could spare resources in a SN.

Thereafter it is presented the MPT model extended for risk management and a test case applied to validate this procedure. This test case used one of the companies analyzed in the exploratory study by Zsidisin et al. (2000). These authors identified the mitigation strategies used for some industries in the purchasing department.

MPT Model Expanded for Risk Management

The MPT model developed by Lima (2007) was adapted considering the risk management approach. This extended model shows the combination of three

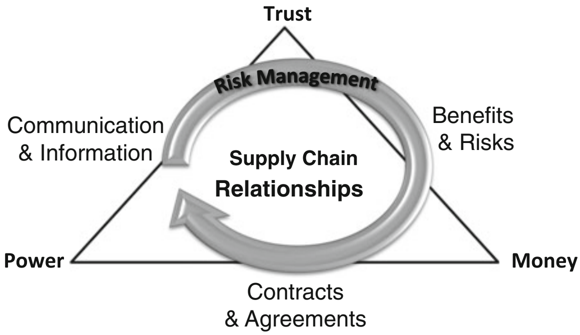


Fig. 3 Extended money, power and trust model (adapted from Lima 2007)

different forces (money, power and trust) and three classes of elements (communication and information, benefits and risks, and contracts and arrangements) to understand the relationships among supply network partners (agents). These forces and elements are responsible for the behavior of agents in the SN. The resulted behavior influences the selection of mitigation strategy in a SN. Figure 3 presents the MPT model extended.

Test Case

In order to validate this approach, the procedure was applied for one company in the exploratory study by Zsidisin et al. (2000). The company selected was from the aerospace industry with the following characteristics: large size (annual revenue over USD 10 billion), with an expressive number of suppliers, performs risk management through the use of qualitative and quantitative models, and based the risk analysis on the capacity of the suppliers.

From these assumptions and with the additional information given by the exploratory study, it was identified that the trust force was the most relevant issue in the relationships between the company and its suppliers. In this case, based on the previous discussion regarding risk management and relational arrangements, a suggested mitigation strategy is to increase the visibility in the relationships, through an intensive collaboration (e.g. sharing production and supplier orders information).

Christopher et al. (2011) highlight this strategy option for this segment of industry, by creating a risk management culture in SN. Actually, the company adopted a mitigation strategy to closely work with the supplier to mitigate risks.

Conclusions and Future Research

This paper presented a procedure to aid the selection of mitigation strategy in supply networks, applying the MPT model in a risk management framework. This approach demonstrated a good potential in the case tested, although it is necessary to validate this model in other situations. Suggestions for future research: (a) Increase the analysis in different supply chain process (e.g. production, transportation, etc.) as well as in different types of relational arrangements (collaborative, competitive or cooperative); (b) Apply this approach in a global supply network, considering the external risks in this environment; and (c) Develop a simulation model to allow the generation of multiple scenarios considering different mitigation strategies.

Acknowledgments This research was supported by Capes as part of the Brazilian-German Collaborative Research Initiative on Manufacturing Technology (BRAGECRIM).

References

Balestrin A, Verschoore J (2008) Redes de cooperação empresarial: estratégias de gestão na nova economia. Bookman, Porto Alegre

Ballou RH (2006) The evolution and future of logistics and supply chain management. Produção 16:375–386

Barratt M (2004) Understanding the meaning of collaboration in the supply chain. Supply Chain Manage Int J 9:30–42

Bengtsson M, Kock S (2000) "Coopetition" in business networks—to cooperate and compete simultaneously. Ind Mark Manage 29:411–426

Benton WC, Maloni M (2005) The influence of power driven buyer/seller relationships on supply chain satisfaction. J Oper Manage 23:1–22

Christopher M, Mena C, Khan O, Yurt O (2011) Approaches to managing global sourcing risk. Supply Chain Manage Int J 16(2):67–81

Cucchiella F, Gastaldi M (2006) Risk management in supply chain: a real option approach. J Manuf Technol Manag 17(6):700–720

Hallikas J, Karvonen I, Pulkkinen U, Virolainen VM, Tuominen M (2004) Risk management processes in supplier networks. Int J Prod Econ 90(1):47–58

Khan O, Burnes B (2007) Risk and supply chain management: creating a research agenda. Int J Logistic Manage 18(2):197–216

Kleindorfer PR, Saad GH (2005) Managing disruption risks in supply chains. Prod Oper Manage 14(1):53–68

Lima OF (2007) Logistics services performance. Sabatical Studies Bornemouth: Bournemouth University, UK

Loureiro SA, Lima Júnior OF (2010) Integrated modeling and simulation methods of operations applied to problems: analysis and practical applications. In: XVI ICIEOM—International conference on industrial engineering and operations management, São Carlos, Brazil, pp 112–121

Manuj I, Mentzer JT (2008) Global supply chain risk management strategies. Int J Phys Distrib Logistics Manage 38(3):192–223

Matook S, Lasch R, Tamaschke R (2009) Supplier development with benchmarking as part of a comprehensive supplier risk management framework. Int J Oper Prod Manage 29(3):241–267

Olson DL, Wu DD (2010) A review of enterprise risk management in supply chain. Kybernetes 39(5):694–706

Payan JM (2007) A review and delineation of cooperation and coordination in marketing channels. Eur Bus Rev 19:216–233

Peña NA, Arroyabe JCF (2002) Business cooperation: from theory to pratice. Palagrave Macmillian, New York

Rajagopal A (2007) Competition versus cooperation: analysing strategy dilemma in business growth under changing social paradigms. Int J Bus Environ 1:476–487

Ritchie B, Brindley C (2007) An emergent framework for supply chain risk management and performance measurement. J Oper Res Soc 58(11):1398–1412

Sahay BS (2003) Understanding trust in supply chain relationships. Ind Manage Data Syst 103:553–563

Stadtler H (2009) A framework for collaborative planning and state-of-the-art. OR Spectr 31:5–30

Tang CS (2006) Perspectives in supply chain risk management. Int J Prod Econ 103:451–488

Tummala R, Schoenherr T (2011) Assessing and managing risks using the supply chain risk management process—SCRMP. Supply Chain Manage Int J 16(6):1–26

Williams Z, Moore R (2007) Supply chain relationships and information capabilities: the creation and use of information power. Int J Phys Distrib Logistics Manage 37:469–483

Zsidisin GA, Panelli A, Upton R (2000) Purchasing organization involvement in risk assessments, contingency plans, and risk management: an exploratory study. Supply Chain Manage Int J 5(4):187–197

Optimization of Failure Behavior of a Decentralized High-Density 2D Storage System

Kai Furmans, Kevin R. Gue and Zäzilia Seibold

Abstract To meet the requirement of flexibility in intralogistics systems, the GridFlow 2D storage system has been developed which is able react to fluctuating processing volumes and layout changes. It consists of multiple autonomous conveyor modules, FlexConveyors, each equipped with a controller. The decentralized control can be described with a set of rules for the communication by message passing between the controllers of the FlexConveyors. In this paper an adapted control algorithm is presented which enables the conveyor modules to react to occurring failures in neighboring modules. It has been theoretically proved that the presented algorithm prevents system deadlock. Also the impact of occurring failures on the system performance has been examined through a Monte Carlo experiment using discrete event simulation.

Introduction

Economic changes like globalization, e-commerce and economic instability—just to name a few—have shifted the conditions and requirements of logistics systems

K. Furmans · Z. Seibold (✉)
Institute for Material Handling and Logistics,
Karlsruhe Institute of Technology, Karlsruhe, Germany
e-mail: zaezilia.seibold@kit.edu

K. Furmans
e-mail: kai.furmans@kit.edu

K. R. Gue
Department of Industrial and Systems Engineering,
Auburn University, Alabama, USA
e-mail: kevin.gue@auburn.edu

H.-J. Kreowski et al. (eds.), *Dynamics in Logistics*, Lecture Notes in Logistics,
DOI: 10.1007/978-3-642-35966-8_35,

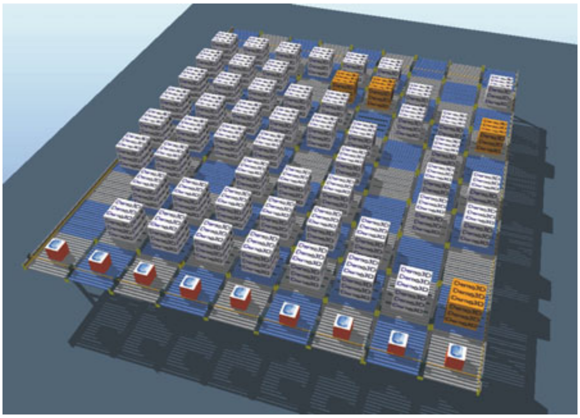


Fig. 1 The gridflow storage system

in recent years. As part of material handling systems, modern storage systems should combine both high performance and flexibility. Decentralized control enables the realization of flexible, automated material handling systems (Furmans et al. 2010; Windt and Hülsmann 2007).

To meet the new requirements of high flexibility, the GridFlow system has been developed consisting of multiple, autonomous modules being able to fulfill the tasks of a storage system (Gue and Furmans 2011) (Fig. 1). Each module is equipped with electronics to act independently of a central controller; the control of the GridFlow is decentralized.

GridFlow unifies two different approaches to turn a storage system into a high-performing and space-efficient system. On one hand, the design guarantees high throughput and high density at the same time. On the other hand, it follows the current trend towards flexibility in material flow systems by featuring decentralized control. Single modules can easily be plugged and unplugged in order to adapt to fluctuating volumes and changing layout requirements.

Besides flexibility, another advantage of decentralized control is system reliability: In centralized systems, the failure of the central controller causes a complete system breakdown, whereas, in decentralized systems, the failure of modules does not necessarily affect the function of the other modules.

The research presented in this paper introduces an algorithm enabling conveyor modules in a GridFlow system to react to failures and to resolve blockings in the material flow.

In the following section the general concept of GridFlow is presented. The next section describes the problem of failures in GridFlow and presents the adapted algorithm. We show that the system will not deadlock, and then describe the impact of failures on system performance.

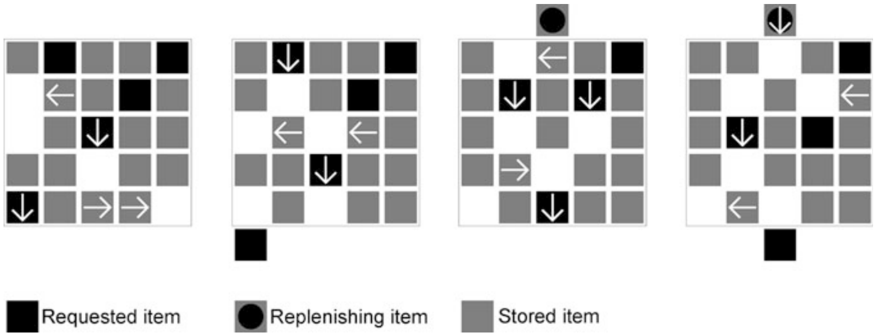


Fig. 2 Example movement in a 5 × 5 gridflow system with 4 empty cells (*white spaces*)

System Description

The GridFlow system examined in this paper is designed in 2D with incoming and outgoing goods. To enable movement of the goods, there must be at least one empty cell per row in the initial setup. Figure 2 shows an example set of movement steps in a 5 × 5 GridFlow system. As outgoing goods are retrieved to the South of the storage space, requested items move southward to the retrieval row. Also incoming goods, the replenishing items, move southward and fill up the storage space.

To enable southward movement of requested and replenishing items, empty conveyor cells (white spaces) must be south of them. Stored items in the way of requested or replenishing items try to move east- and westward to create empty cells. In Fig. 2, white arrows specify the movement of items during four cycles. An item can move to a directly neighboring cell during one cycle.

The decision of which items are moved is made by communication between the controllers in the conveyor units. The conveyors are considered to be agents capable of making decisions based on perceptions and performing resulting actions (Woolridge 2009). Each cycle consists of four process steps, which start in each conveyor unit concurrently.

North–South Negotiation

Communication is conducted in North–South direction. Conveyor units occupied with requested or replenishing items make a decision about conveying. If the South neighbor is empty, the conveyors commit to convey down. If the South neighbor is occupied with a stored item, it will try to transport it east- or westward.

East–West Negotiation

Communication is conducted in East–West direction. The conveyors of each row decide which stored items to transport east- or westward. Occupied conveyors south of requested or replenishing items try to become empty and therefore initiate a message flow following a Request-Willing-Commit logic. Empty cells are needed to enable East–West movement of stored items. Basically, requested and replenishing items located north of the row compete for empty cells.

Transportation

The conveyors perform the conveying action chosen during the negotiation phases. Active conveyors update their state according to the movement of the conveyed items.

Retrieval, Replenishment and Requesting

During this step outgoing goods are retrieved, new items are placed in the replenishment row and new requests are put into the system. Affected conveyors update their state.

To control the system replenishment and the North–South distribution of empty cells, a target row is assigned to each item. The target row defines the storage location to which the item must move. As long as the item has not reached its target row, it is considered to be a replenishing item.

Problem Statement

High reliability is one of the advantages of a decentralized system. In order to maximize the reliability of GridFlow, each conveyor must be able to handle failures occurring both to itself and to neighboring conveyors. In this research, only failure types affecting the negotiations between the conveyor units are considered. Therefore, the algorithm has been adapted to the following two failure types: failed conveyors and failed ports.

If a complete conveyor unit fails, the carried item cannot be retrieved. To prevent a system deadlock, the control algorithm has to detect and to resolve blockings of row sections in the storage space. The system is deadlocked if one conveyor having the objective to forward the carried item cannot fulfill its task at any future cycle. The objective of the algorithm is to keep the system working until the failure is repaired and to guarantee that as many requested items as possible can be retrieved.

The following section describes the modification of the algorithm and the countermeasures taken to prevent system deadlock. We then demonstrate that the modified algorithm successfully prevents deadlock, and show the performance of a system with failed conveyors and failed ports. The results can give insights into how reliable a single conveyor unit should work and how quickly failures should be repaired.

The Control Concept

A conveyor can either fail mechanically, which means that it can communicate its breakdown to its neighbors, or the controller can fail, which interrupts

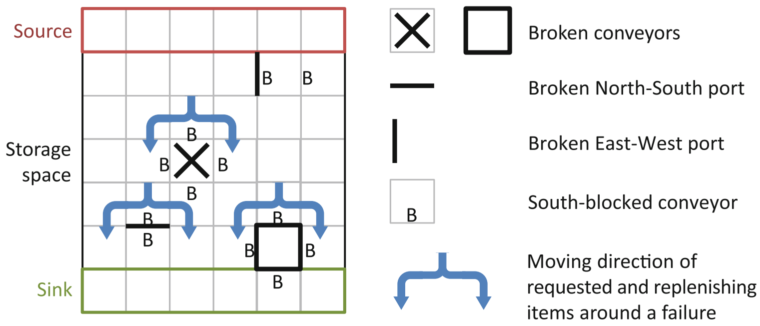


Fig. 3 Schematic representation of a system with failures

communication. Both cases have the same consequences on the control problem, only failure detection is done in a different way.

A failed port interrupts message passing and hinders communication between the two adjacent conveyor units. The neighbors connected by this port do not get any reply to messages. A failed port is comparable to a passive port at the border of the system because it interdicts conveying in this direction. In the figures below, passive ports are represented by a black line (see Fig. 3).

As shown in Fig. 3 a failed conveyor corresponds to a conveyor whose four ports (North, East, South and West) are unresponsive or passive. The conveyor is blocked on four sides and the currently placed item cannot be moved until the failed conveyor is repaired. In order to generalize, every failure type is represented by a passive port in the control algorithm.

A passive port in the middle of the storage space blocks movement of items: A failed East–West port divides a row into two sections and blocks East–West movement necessary to locate empty cells where they are needed. A failed North–South port hinders South-movement of requested and replenishing items. By implication, requested or replenishing items on conveyors blocked to the South must move east- or westward before being able to move southward. A conveyor always has one of the states shown in Table 1.

Table 1 Conveyor states

	Conveyor state	Occupied by	Objective to forward the item
□	Empty	–	–
■	S-Requesting	Requested or replenishing item	Southward
■ (grey)	Occupied	Stored item	–
◆ (grey)	EW-Requesting	Stored item	East- or westward
◆	S-EW-Requesting	Requested or replenishing item	East-or westward in direction of closest escape to the South

In the following section, message types are introduced which are necessary to detect and resolve the blocking of a row section. The complete algorithm also contains other message types and rules explained in communication protocols depending on negotiation states and message types (Krothapalli and Deshmukh 1999).

North–South Negotiation

↓R	Request	This message is triggered at the beginning of North–South negotiation
↑B	Blocked	This message is triggered if a request cannot be satisfied because of a passive port

During North–South negotiation, the S-Requesting conveyors send request messages to the South looking for an empty cell. Assuming that the successful reception of a request must be confirmed, the lack of feedback implies a passive port. If a S-Requesting conveyor detects a passive port, its state turns into S-EW-Requesting. Figure 4 shows an example.

A S-EW-Requesting conveyor also checks if the closest escape to the South is in the East or the West. If the South-blocked conveyor discovers that there is neither an escape for the carried item in the East nor the West, it does not try to move the item anymore. In this pathological case with a conveyor blocked in three directions, it acts as if it were completely failed.

East–West negotiation

R →	Request	This message is triggered at the beginning of East–West negotiation
B →	Blocked	This message is triggered if a conveyor receives a request and movement of the carried box to the opposite side is blocked
RB →	Resolve blocking	This message is triggered if a S-EW-Requesting conveyor is blocked on both sides

During East–West negotiation, the conveyors decide which items to move east- or westward. In a system with failures they should additionally detect blocked row sections and take actions to resolve the blocking.

An EW-Requesting conveyor tries to forward the carried item east- or westward to create an empty cell for a requested or replenishing item from the North. To minimize transportation of items, it selects the direction of the closest empty cell enabling East–West movement.

A S-EW-Requesting conveyor also tries to forward the carried item east- or westward. But in contrast to the EW-Requesting conveyor, it should forward the

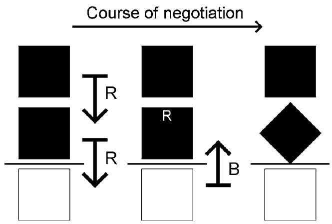


Fig. 4 Example for North-South negotiation

item in direction of the closest escape to the South. This direction has been determined at the end of North–South negotiation. Only if the escape is equidistant in both directions, the conveyor selects the direction of the closest empty cell during East–West negotiation.

At the beginning of East–West negotiation, EW-Requesting and S-EW-Requesting conveyors initiate Request-messages in one or both directions concurrently. A row section is blocked if the objective of a conveyor to move the carried item east- or westward cannot be satisfied.

To detect passive ports, the successful transmission of a Request-message from conveyor to conveyor must be confirmed. Again, the lack of feedback implies a passive port. A row section is blocked if there is no Empty or S-Requesting conveyor; the requesting conveyor will receive feedback that movement in this direction is not possible. If the row section is blocked in both directions, the conveyor must resolve the blocking by requesting a stored item to move southward.

In the example in Fig. 5 three conveyors must check if the movement of the carried item is blocked long-term. Two of them actually detect and resolve a blocking: The EW-Requesting conveyor requests the carried item to move southward and the S-EW-Requesting conveyor on the left side requests the stored item on the neighboring conveyor to move southward.

Demonstration of System Liveness

Assuming that the retrieval row is empty at the beginning of each cycle, it is only necessary to prove that every requested and replenishing item can be transported to the next row. This is the case if every conveyor with an objective can forward the carried item at some future cycle. Instead of preventing blocked sections, the presented algorithm should guarantee that blockings of row sections are detected and resolved.

Obviously, there are combinations of multiple failures blocking the retrieval of requested items, for example a complete row of failed North–South ports. In this case a conveyor carrying a requested item cannot fulfill its objective and gives it up; it acts as completely failed.

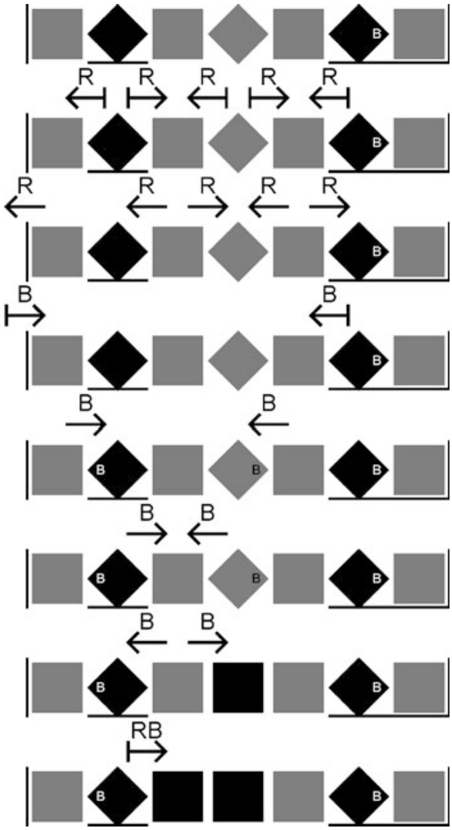


Fig. 5 Example for East–West negotiation

Three questions have to be answered to demonstrate that the described algorithm guarantees that a requested item can be retrieved if there is a path of functioning conveyors and ports to the retrieval row:

- When is a row section blocked?
- How are blocked row sections detected?
- How are blockings of row sections resolved?

Every conveyor has one of the states described in the Table 1. An empty or S-Requesting conveyor indicates that a row section is not blocked because it enables East–West movement.

When Is a Row Section Blocked?

A row section is blocked if a conveyor with an objective cannot forward the carried item in the required direction. By implication, a row section is blocked if there is an EW-Requesting or S-EW-Requesting conveyor but no empty or S-Requesting conveyor in the conveying direction of the requesting conveyor.

How are Blocked Row Sections Detected?

Each conveyor with the objective to move its item east-or westward initiates requests in the potential conveying directions. A request is forwarded by occupied conveyors until it is:

- stopped by an empty or S-Requesting conveyor, because these states indicate that movement is not blocked long-term.
- stopped by an EW-Requesting or S-EW-Requesting conveyor, because this conveyor already sent requests by itself.
- answered with a message that movement is blocked if there is a passive port.

A Blocked-message is forwarded through the complete row until it is stopped either by a conveyor with the indication that the row section is not blocked or by a conveyor that is blocked on the opposite side.

How Are Blockings of Row Sections Resolved?

A blocking can be resolved by turning a conveyor state from Occupied or EW-Requesting into S-Requesting. The affected item will move southward and leave an empty cell to enable East–West movement. If an EW-Requesting conveyor receives Blocked-messages from both sides, it turns into S-Requesting. If a S-EW-Requesting conveyor is blocked in both directions, it triggers a message to resolve the blocking which is forwarded to the closest conveyor occupied with a stored item.

The message rules defined in communication protocols guarantee the described behavior. Every conveyor with the objective to forward a carried item can fulfill its task at some future cycle. Consequently, every requested item that is not blocked on a failed conveyor can be retrieved, and the system does not deadlock.

System Performance Analysis

GridFlow has been modeled in an agent-based, discrete-event simulation environment. Items enter the system by appearing at random cells in the replenishment row and leave the system by disappearing in the retrieval row. In this study the system performance is determined by a CONstant Work In Process analysis (CONWIP), i.e. the number of items and the number of requested items in the system are constant.

For the performance analysis of a system with failures, a high system density has been chosen: The storage space stores 144 items on 12 rows and 13 columns, which corresponds to a density of 92.3 %. The work in process varies from 10 to 130 requests in steps of 10. In order to fulfill the conditions for the CONWIP analysis, the simulation does not exactly correspond to the system behavior in reality: If a requested item is blocked on a failed conveyor, another request is generated in order to keep a constant number of processed requests in the system.

The objective of the study is to simulate the system performance with realistic conditions. The results demonstrate how reliable the conveyor units should be and how frequently failures should be repaired.

Every conveyor unit and every port fails at a certain rate. The reciprocal value of the failure rate defines the "Mean Time Between Failure" (MTBF) indicating how long a conveyor/port works without failure after being brought into service or after being repaired. In the experiments, MTBF varies from 800 to 8,000 operating hours. The probability of a failure occurrence is assumed to be exponentially distributed: A conveyor/port fails with the same probability at all times independently from the last failure occurrence. Basically, every cycle in GridFlow can be understood as Bernoulli trial with 1/MTBF as probability of failure for every conveyor module.

Failed units are not repaired immediately but at a periodically defined time, for example, at the end of a working shift. As the algorithm has been adapted to failures, the system does not need to shut down immediately after failure occurrence but continues to operate with the failure until the next scheduled repair. The "Time Between Repair" (TBR) varies from 1 to 3 shifts. It describes the definite time between two scheduled repairs in contrast to the commonly used term "Mean Time To Repair" (MTTR) describing the statistically distributed time between a failure and its repair. A failure can easily be repaired by exchanging the failed conveyor unit with a working one.

Because of the statistical failure occurrence, the interference of multiple failures is highly influencing the result: for example, two failed North–South ports neighboring in the same row have a significantly greater impact on the throughput than two failed North–South ports in the same column. Therefore, a Monte Carlo simulation with multiple experiments has been conducted for each combination of MTBF, TBR and work in process. The throughput and the number of resolving actions are averaged over the different values of WIP.

Figure 6 shows the average throughput compared to a system without failures. As expected, the throughput decreases with increasing failure probability (1/MTBF) and decreasing repair frequency (1/TBR). The figure also shows that

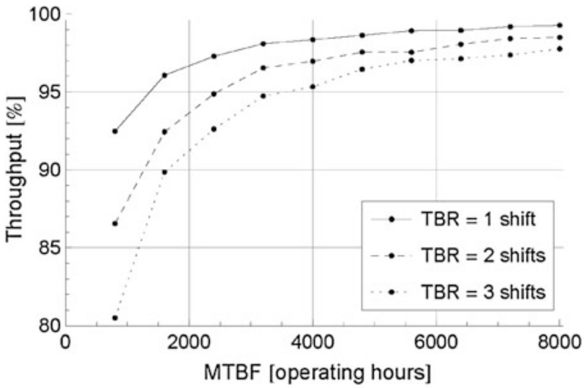
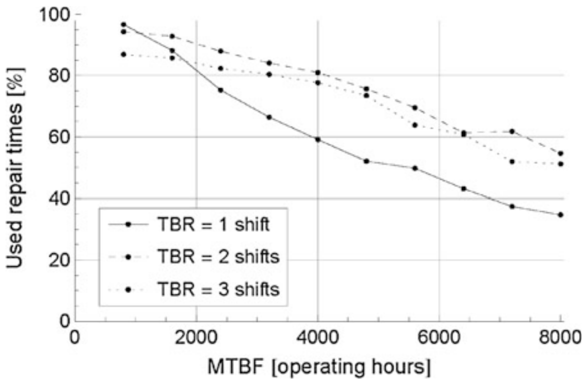


Fig. 6 System performance with statistical failure occurrence

Fig. 7 Usage of repair times



even with a repair frequency equal to 3 shifts, a good system performance can be achieved despite failures. If, for example, the throughput should not decrease more than 5 % compared to a system without failures and if the repair frequency is 3 shifts, a conveyor unit should work at least 4000 operating hours without failure.

Figure 7 gives insights into the necessity of repairs. It illustrates how often the scheduled repair time is required. For higher MTBF, it is not necessary to schedule a repair after each working shift.

Conclusion

The algorithm for the GridFlow system has been successfully adapted to failures in any occurring combination. It has been shown that system deadlock is prevented.

The simulation of a working GridFlow with statistical failure occurrence has shown that good system performance can be realized despite failures. Since it is not required to repair a failure immediately after its detection, high system availability is achieved. Because a failure can be repaired by exchanging the failed unit, a relatively high repair frequency has been assumed. Realizing a high technical reliability would decrease the required repair frequency.

References

Furmans K, Schönung F, Gue KR (2010) Plug-and-work material handling systems. In: Proceedings of the international material handling research colloquium 2010

Gue KR, Furmans K (2011) Decentralized control in a grid-based storage system. In: Doolen T, Van Aken E (eds) Proceedings of the 2011 industrial engineering research conference 2011

Krothapalli NKC, Deshmukh AV (1999) Design of negotiation protocols for multi-agent manufacturing systems. Int J Prod Res 37(7):1601–1624

Windt K, Hülsmann M (2007) Changing paradigms in logistics—understanding the shift from conventional control to autonomous cooperation and control. Springer, Berlin Heidelberg

Woolridge M (2009) An introduction to multiagent systems. Wiley, New York

Part VIII
Autonomous Control

Distributed Reinforcement Learning for Optimizing Resource Allocation in Autonomous Logistics Processes

Jan Ole Berndt and Otthein Herzog

Abstract This paper examines multi-agent coordination for resource allocation tasks in autonomous logistics processes. It identifies requirements for the learning of optimal behavior in a multi-agent setting. Based on a real-world logistics application, the paper distinguishes between single resource allocation by independent agents and joint activities by agent teams. For both cases it introduces adaptations of the Q-learning algorithm and evaluates their convergence as well as their scalability for large scenarios. The results demonstrate that the known conditions for the convergence of multi-agent reinforcement learning are insufficient. This leads to the identification of an additional requirement for convergence in this paper.

Introduction

Growing interrelations between logistics processes and their inherent dynamics result in a high complexity of logistics operations. The application of conventional centralized planning and control approaches to these processes suffers from that intrinsic complexity. Hence, decentralized methods become necessary which employ autonomous actors representing logistics entities and objects (Hülsmann et al. 2006). These autonomous entities can be implemented as intelligent software agents to model logistics networks as multi-agent systems (MAS). Those systems

J. O. Berndt (✉) · O. Herzog
Center for Computing and Communication Technologies,
Intelligent Systems Department, University of Bremen, FB3/TZI/Intelligent Systems,
P.O. Box 330 440, 28334 Bremen, Germany
e-mail: joberndt@tzi.de

O. Herzog
e-mail: herzog@tzi.de

H.-J. Kreowski et al. (eds.), *Dynamics in Logistics*, Lecture Notes in Logistics,
DOI: 10.1007/978-3-642-35966-8_36,

enable simulations, evaluations, and the actual application of autonomous control in logistics (Schuldt 2011).

When developing the aforementioned approaches, coordination of autonomous entities is a challenging task. On the one hand, decentralized decision-making decreases the computational effort for logistics process control by reducing the number of parameters to take into account. That is, each entity uses only locally available information for its decisions. On the other hand, to achieve the desired overall system performance, autonomous agents have to coordinate their actions.

This paper examines multi-agent coordination for resource allocation tasks in autonomous logistics. Specifically, it applies distributed learning to achieve optimal allocations. Its structure is as follows. First, it introduces a real-world example for distributed resource allocation and suggests its transition to autonomous logistics. Second, the paper identifies multi-agent coordination challenges, discusses available coordination mechanisms, and motivates the application of reinforcement learning to optimize MAS performance. Subsequently, it introduces two adaptations of the Q-learning algorithm for both individual and joint resource allocation. An empirical evaluation assesses their capability to converge to an overall optimum and their scalability for large scenarios. As a result, an additional requirement for convergent multi-agent reinforcement learning is identified. Finally, a concluding summary of this paper and directions for future research are given.

Resource Allocation in Autonomous Logistics

Onward container carriage serves as a real-world example for logistics resource allocation throughout this paper. This process covers the haulage of shipping containers from the port of discharge to their receivers as the final part of their overseas transport. The scenario and its transition to autonomous logistics has been subject to a thorough investigation (Schuldt 2011) providing the foundation for this example.

Autonomous control can be applied to this process by attaching an intelligent agent to each container, means of transport, and storage facility. These agents represent the logistics objects in a MAS. In the onward carriage process, they have to coordinate the following logistics functions of their represented objects: First, an agent acting on behalf of a container must identify a suitable storage facility as its destination. Subsequently, it has to select a means of transport to move the container to the chosen storage facility. Both activities require the agent to allocate resources from appropriate service providers with sufficient capacities.

In order to maximize the utilization of service providers, groups of containers should be handled together (Schuldt and Werner 2007). Therefore, agents can establish teams for conducting joint activities (Schuldt 2011). Each of these teams names one agent as a representative for all agent interactions. This *team manager* has to identify service providers which meet the requirements of the whole team. If no single service provider has enough capacities, a team manager has to split its resource allocation over several ones in order to fulfill the joint objective of its team.

Coordination Challenges and Related Work

The aforementioned example scenario covers some typical problems of distributed agent coordination. As there are several single agents or teams of agents pursuing different objectives, each agent acting on its own or as a team manager has to cope with the following challenges:

1. Other agents' activities can interfere with the agent's own actions.
2. The agent can only partially observe its environment.

In the decentralized setting of autonomous control, agents act simultaneously. Given limited capacities of the logistics service providers, the outcome of an agent's allocation request depends on the parallel actions of its fellow agents (Challenge 1). Thus, in order to successfully reach its logistics objective, each agent has to take these activities into account when deciding which service provider to select.

However, each agent has only limited information about other agents' existence, their capabilities, and their behavior (Challenge 2). This is due to the decomposition of the overall process into local decision-making contexts in autonomous control. Therefore, multi-agent coordination mechanisms have to circumvent conflicting agent activities while preserving the principle of decentralization in order to optimize the overall logistics performance.

In their concurrent activities, the agents compete for the highest individual benefits due to the mutual interference of their actions. However, the partial observability and the principle of distribution in autonomous logistics make it difficult to apply existing coordination approaches for competitive environments. For instance, the well-known *minimax* principle (Neumann 1928) and its generalization for settings with more than two players (Luckhart and Irani 1986) requires each agent to be aware of the others' possible actions and their respective payoff values. The lack of this information prevents these game theoretical methods from being applied. Multi-agent *negotiations* over re-allocations, in which agents make deals about possible improvements of a given allocation (Endriss et al. 2006), cannot be applied, either. This is because in order to identify advantageous re-allocations, the agents have to be aware of the current allocation of all resources. Finally, *combinatorial auctions* (Cramton et al. 2006; Ramezani and Endriss 2010), in which an auctioneer collects all other agents' preferences being expressed as bids on combinations of resources, contradict the concept of problem decomposition in autonomous control. As there is only one single auctioneer which determines the winners of the auction, this is equivalent to computing an optimal allocation centrally (Ramezani and Endriss 2010).

Due to the lack of explicit information for multi-agent coordination, the agents must coordinate their activities implicitly. That is, they individually have to identify their most beneficial actions by deducing missing information from observing the outcome of their own activities. Deducing action or state values from observed payoffs is known as *reinforcement learning* (RL) (Sutton and Barto 1998). For the case of one single learning agent, a well-understood RL algorithm is Q-learning (Watkins and Dayan 1992). This algorithm estimates expected rewards (action payoffs) as Q-values $Q(s, a)$ for each state and action. It uses an update rule to refine its estimation when observing a reward $R(s, a)$ for action a in state s. If each state action pair is sampled infinitely often, the Q-values converge to the unobservable true values Q^* (Watkins and Dayan 1992). This enables the learning agent to select its activities according to their expected payoff values. Thus, as the values converge, it can identify the optimal action for each state.

However, in distributed resource allocation, the outcome of an agent's actions depends on the combined actions of all agents. Therefore, each agent has to identify its most beneficial action, notwithstanding the missing information about the other agents' choices or additional environmental influence factors. Lacking this information, it cannot observe the state s as in conventional Q-learning. In fact, the agents do not even know the underlying state space. Instead, they only have to assess the expected payoff for each action without taking the state into account. Omitting it leads to Q-values $Q(a)$ and the following simplified update rule (Claus and Boutilier 1998):

$$Q(a) \leftarrow Q(a) + \lambda \cdot (R(a) - Q(a)).$$

For every learning rate $0 < \lambda \leq 1$, this rule exhibits the convergence property of conventional Q-learning (Claus and Boutilier 1998; Watkins and Dayan 1992).

This update rule, however, only covers single agent learning. In fact, for multiple learners which act simultaneously, the convergence property of the Q-values does not hold in general (Claus and Boutilier 1998). This is due to the fact that the agents act in a non-stationary environment as the outcome of each agent's current action depends on all other agents' actions which can vary over time.

To guarantee convergence in a multi-agent setting, an agent can either adapt its action selection strategy to other agents' activities (Tesauro 2004) or the agents optimize a shared global reward function (Kemmerich and Kleine Büning 2011). The former requires an agent to estimate its fellow agents' actions in order for itself to select those actions which provide the highest expected payoff. However, as the other agents' activities are not observable, this method is not applicable to the considered resource allocation problem. Contrastingly, the latter technique solely considers an agent's own activities. However, it is only suitable for cooperative MAS in which all agents pursue a joint overall objective. This assumption only holds for members of the same team of agents, whereas agents from different teams compete for limited resources. Additionally, it requires a global observer to calculate the common reward value. This contradicts the principle of distribution in autonomous logistics.

Distributed Reinforcement Learning in Autonomous Logistics

Following from the previous section, specialized multi-agent Q-learning is not applicable to the considered autonomous logistics process. Moreover, the single agent algorithm generally cannot guarantee convergence in a distributed setting. Nevertheless, (Claus and Boutilier 1998) identify two additional conditions for the convergence of the latter in scenarios containing two agents:

1. All possible actions of an agent have a non-zero probability of being selected.
2. The agents use asymptotically exploitive action selection strategies.

Condition 1 ensures the infinite sampling of all agent actions for $t \rightarrow \infty$. Furthermore, it prevents the agents from executing strictly correlated explorations. That is, no combination of actions becomes impossible to occur. This is a necessary extension of the infinite sampling requirement for single agent Q-learning: In a multi-agent setting, each combination of actions must be executed infinitely often, as an action's payoff depends on all other actions triggered concurrently. Condition 2 requires the agents to pursue a *decaying* exploration strategy. This decreases the probability of concurrent exploration activities over time. Hence, the agents become less likely to disturb each other's actions and the Q-values can settle.

The following sections exploit these conditions to apply the stateless Q-learning technique to autonomous logistics, thereby extending the considerations in (Claus and Boutilier 1998) to settings with arbitrary numbers of agents. First, Independent Agent Interaction introduces a Q-learning algorithm for multiple agents which independently allocate resources. In contrast to that, Team Interaction focuses on joint resource allocation for teams of agents. Finally, an empirical examination of the convergence to an optimal allocation and the scalability for large scenarios is given in the Evaluation and Discussion of both cases.

Independent Agent Interaction

In the onward container carriage scenario, the agents representing the containers form the set AG of agents attempting to allocate resources from the available service providers. All agents allocate their resources simultaneously. Each agent requests one single unit of a resource at a time and each service provider offers exactly one unit. Therefore, in order to maximize their allocation performance, the agents have to find a bijective mapping of service providers to service consumers.

Independent agent interaction means that each agent $i \in AG$ allocates its required resources on its own. The set A^i consists of all allocation actions of agent i. Depending on all concurrent actions, an agent's attempt $a^i \in A^i$ to allocate a resource is either successful or it fails. When observing the respective result, the agent evaluates its action by calculating a reward value $R^i(a^i)$.

Algorithm 1. Distributed Stateless Q-learning ($\forall i \in AG$)

Input: agent action set A^i, learning rate λ, exploration rate ε, decay rate δ

Initialize $t \leftarrow 0$ and $\forall a^i \in A^i : Q_0^i(a^i) \leftarrow 0$
loop
  $maxA_t^i \leftarrow \arg\max_{a^i \in A^i} Q_t^i(a^i)$
  Select $A_t^i \subseteq A^i$ with $P(A_t^i = maxA_t^i) = 1 - \varepsilon$ and $P(A_t^i = A^i) = \varepsilon$
  Randomly select $a_t^i \in A_t^i$
  Execute action a_t^i and observe reward $R_t^i(a_t^i)$
  for all $a^i \in A^i$ **do**

$$Q_{t+1}^i(a^i) \leftarrow \begin{cases} Q_t^i(a^i) + \lambda \cdot (R_t^i(a^i) - Q_t^i(a^i)) & \text{if } a^i = a_t^i \\ Q_t^i(a^i) & \text{otherwise} \end{cases}$$

  end for
  Update $\varepsilon \leftarrow \varepsilon \cdot \delta$ and $t \leftarrow t + 1$
end loop

Algorithm 1 defines the distributed Q-learning approach for the independent agent interaction case. In order to ensure an asymptotically exploitive action selection with non-zero probabilities, it employs a decaying ε-greedy strategy. That is, an agent selects its next action from the set of those with the highest Q-values $maxA^i$ with a probability of 1-ε (with $0 < \varepsilon \leq 1$). If there are several best actions, the agent chooses randomly among them. With a probability of ε, the agent selects its next action randomly out of all actions in A^i. In the course of time, ε becomes discounted by a decay factor $0 < \delta \leq 1$. Therefore: $\lim_{t \to \infty} \varepsilon = 0$ and $\forall\, t < \infty: \varepsilon > 0$.

This selection strategy leads to high exploration rates in the beginning of the learning process which decrease over time. Thus, once an agent has identified a successful action, it increasingly tends to stick to that action. In order to achieve this behavior, the agent rewards a successful action with a value of one while a failure receives a value of zero. Consequently, the Q-value for a repeatedly successful action will approximate one and the value for a failing action will converge to zero. In conjunction with initial Q-values of zero for all actions, this leads to a purely explorative behavior in the beginning and in case of repeated failures as $maxA^i = A^i$ if $\forall\, a^i \in A^i$: $Q^i(a^i) = 0$. However, as soon as an agent observes a successful action, it will switch to the ε-greedy strategy to exploit this knowledge.

Team Interaction

In the team interaction case, several agents jointly allocate their required resources. Each team is represented by a single team manager agent. Therefore, the set of service consumers equals to the set of those managers $TM \subseteq AG$. Furthermore, each team manager $i \in TM$ has to allocate resources for all team members. Thus, the team action set A^I for k team members consists of all combinations of allocation actions of size k: $\forall a^I \in A^I : a^I \subseteq A^i$ with $|a^I| = k$.[1]

It is possible to apply Algorithm 1 to this case as well. In order to do this, one has to replace all single agent actions $a^i \in A^i$ with team actions $a^I \in A^I$. It follows that each agent learns Q-values $Q^I(a^I)$ for those *composite* actions. Only the reward function requires an adaptation. A composite action can be partially successful, as each single component action can either succeed or fail, independently from the others. Hence, the team reward function R^I aggregates the rewards for all single actions in a^I: $R^I(a^I) = \sum_{a^i \in a^I} R^i(a^i)$. Consequently, in the case of $k = 1$ for all teams, the team interaction approach is equivalent to single agent interaction.

However, there are two major drawbacks of this method for team interaction. Firstly, the set A^I grows very large, even for small numbers of agents. Thus, learning Q-values $Q^I(a^I)$, $\forall a^I \in A^I$ slows down the convergence process significantly. Secondly, evaluating these composite actions as a whole by $R^I(a^I)$, an agent cannot distinguish between their successful and their ineffective parts. This leads to a lack of incentives to explore alternative actions in case of partial failures, as it cannot identify them. This will tempt the agent to be satisfied with a suboptimal option.

Algorithm 2 aims at overcoming these drawbacks by decomposing the team actions into single agent ones. This enables an agent to retain the Q-function Q^i and, thus, to restrict the number of observable actions to $|A^i|$. In order to achieve this, the agent does not aggregate the reward values when observing the result of a composite action. Instead, it handles each of the component actions independently and performs the aggregation for the composite action's Q-value: $Q^I(a^I) = \sum_{a^i \in a^I} Q^i(a^i)$. Again, for $k = 1$, this is equivalent to Algorithm 1.

Apart from restricting the number of considered Q-values, the decomposition of composite actions has the advantage that ineffective actions can be clearly identified. Therefore, an agent can explore the action space systematically in search for replacements of those actions. In fact, it can even exploit similarities among different composite actions if they share one or more components, as performing similar actions increases the number of observations for these common parts. Thus, the action decomposition does not only avoid the mentioned slowdown of the learning process. Instead, it even has the benefit of speeding it up.

---

[1] In the following, the lowercase i denotes a team manager, while the uppercase I refers to the team represented by that agent. Thus, a team action a^I is a set $\{a^i_1, \ldots, a^i_k\}$ of k single agent actions.

Algorithm 2. Q-learning with Action Decomposition ($\forall i \in TM$)

Input: team size k, agent action set A^i, learning rate λ, exploration rate ε, decay rate δ

Initialize $A^I \leftarrow \binom{A^i}{k}$, $t \leftarrow 0$,and $\forall a^i \in A^i : Q_0^i(a^i) \leftarrow 0$
loop
  $maxA_t^I \leftarrow \arg\max_{a^I \in A^I} \left(\sum_{a^i \in a^I} Q_t^i(a^i)\right)$
  Select $A_t^I \subseteq A^I$ with $P(A_t^I = maxA_t^I) = 1 - \varepsilon$ and $P(A_t^I = A^I) = \varepsilon$
  Randomly select $a_t^I \in A_t^I$
  for all $a^i \in A^i$ **do**
    if $a^i \in a_t^I$ **then**
      Execute action a_t^i and observe reward $R_t^i(a_t^i)$
      $Q_{t+1}^i(a^i) \leftarrow Q_t^i(a^i) + \lambda \cdot (R_t^i(a^i) - Q_t^i(a^i))$
    else
      $Q_{t+1}^i(a^i) \leftarrow Q_t^i(a^i)$
    end if
  end for
  Update $\varepsilon \leftarrow \varepsilon \cdot \delta$and $t \leftarrow t + 1$
end loop

Evaluation and Discussion

The evaluation of the proposed distributed Q-learning methods is conducted using the multi-agent-based simulation system Plasma (Schuldt et al. 2008). The experiments follow the scenario description in the section on Independent Agent Interaction. That That is, there are equal numbers of service providers and service consumers and the latter have to find a bijection between both sets. If the resource allocation attempts of two or more agents $i,j,\ldots$ collide, only the action $a^{\min(i,j,\ldots)}$ is successful whereas the others fail. In the team interaction case, the set of service consumers is partitioned into n teams of equal size k with $k \cdot n = |A^i|, \forall i \in TM$. Each experiment consists of 500 simulation runs with the following parameters: $\lambda = 0.5$, $\varepsilon = 0.05$, and $\delta = 0.99$.

Figure 1 shows the average overall allocation performance and the corresponding standard deviation for the single agent interaction case with 10, 100, and 1,000 service providers and an equal number of service consumers. The performance is measured as the fraction of successful allocation requests compared to the number of all interactions. With more than 80 %, it is remarkably high already in the beginning. The agents' exploration efforts lead to some disturbances during the first 100 time steps. Afterwards, the agent interactions settle and the performance approaches the optimum. The standard deviation supports this observation. The performance of several simulation runs deviates the most in the beginning. As the average values

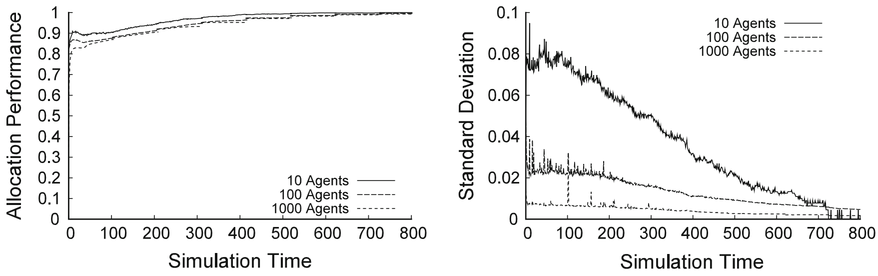


Fig. 1 Resource allocation performance and its standard deviation for single agent interaction

converge to the optimum, the deviation approaches zero. Therefore, the agents reliably find an optimal allocation of resources in the course of time.

Interestingly, the performance is very similar for all three numbers of agents. There is almost no difference between the performance of 100 and that of 1,000 agents after 100 time steps. The standard deviation even decreases for higher numbers of agents. Thus, the distributed stateless Q-learning algorithm scales very well for large numbers of concurrently learning agents.

The evaluation of the team interaction case covers the consideration of composite actions as a whole as well as the action decomposition approach. Figure 2 depicts the performance and standard deviation for 5 teams of 3 agents each (15 agents) and for 50 teams of 20 agents (1,000 agents). The results show that the performance of considering composite actions does not converge to the optimum and consistently deviates over different simulation runs. That is, the agents fail to reliably find an optimal resource allocation. Contrastingly, the action decomposition approach exhibits almost the same properties as the single agent interaction method. It even lessens the disturbances in the beginning and converges slightly faster. Thus, this approach allows for reliable convergence of distributed Q-learning in team interaction and scales very well for large multi-agent systems.

The results of this evaluation show the ability of Q-learning to converge, notwithstanding the non-stationary environment in a multi-agent system. However, the observations for team interaction show that the conditions from (Claus and Boutilier 1998) are insufficient. In fact, as an additional requirement, an

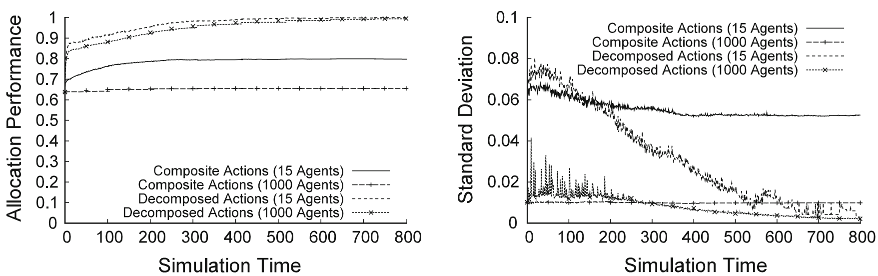


Fig. 2 Resource allocation performance and its standard deviation for team interaction

agent's actions must be *disjoint*. Otherwise, the agent cannot reliably rule out sub-optimal ones. The composite actions experiment with 15 agents validates this conclusion. In this case, the number of actions is $|A^I| = 455$. As the independent agent interaction converges for more than twice this number of actions, the failure is not caused by too many options, but by the inability to identify ineffective components of a composite action.

Conclusions and Outlook

This paper has proposed the application of reinforcement learning for multi-agent resource allocation in autonomous logistics. It has identified the conditions for distributed Q-learning to converge to an overall optimum in a multi-agent setting: non-zero action selection probabilities and asymptotically exploitive exploration strategies. Two algorithms have been introduced, for independent agent interaction and joint resource allocation respectively, which satisfy these conditions.

The evaluation of these algorithms has shown that distributed Q-learning does indeed converge to an optimal allocation of resources. Moreover, the approach scales up remarkably well for large multi-agent scenarios. Additionally, the results have led to the identification of a third convergence condition, especially in joint interaction settings: the disjointness of actions. This is a novel finding of which system designers have to be aware when creating distributed learning applications.

However, there are still questions open for future research. For example, the evaluation scenario for resource allocation only considers the success or failure of an agent action. Future work should include multiple criteria for evaluating the results of an agent's activities. Thus, the integration and aggregation of several logistics key performance indicators will be a necessary step towards a more realistic setting for distributed reinforcement learning in autonomous logistics.

Furthermore, the requirement of asymptotic exploitation prevents the application of reinforcement learning to dynamic scenarios. However, logistics processes must be able to adapt to a changing environment (Berndt 2011). Hence, future research will consider either how to identify changes in an agent's environment in order to adjust the exploration rate accordingly or how to circumvent that requirement.

References

Berndt JO (2011) Self-organizing supply networks: autonomous agent coordination based on expectations. In: Filipe J, Fred A (eds) ICAART 2011, vol 2. SciTePress, Rome, Italy, pp 104–113

Claus C, Boutilier C (1998) The dynamics of reinforcement learning in cooperative multi-agent systems. In AAAI 1998, Madison, USA, pp 746–752

Cramton P, Shoham Y, Steinberg R (eds) (2006) Combinatorial Auctions. The MIT Press, Cambridge, USA

Endriss U, Maudet N, Sadri F, Toni F (2006) Negotiating socially optimal allocations of resources. JAIR 25:315–348

Hülsmann M, Scholz-Reiter B, Freitag M, Wycisk C, De Beer C (2006) Autonomous cooperation as a method to cope with complexity and dynamics? A simulation based analyses and measurement concept approach. In: Bar-Yam Y (ed), ICCS 2006, The New England Complex Systems Institute, Boston, USA

Kemmerich T, Kleine Büning H (2011) A convergent multi-agent reinforcement learning approach for a subclass of cooperative stochastic games. In: Knudson M, Vrancx P, Grześ M (ed), ALA@AAMAS 2011, pp 75–82

Luckhart C, Irani KB (1986) An algorithmic solution of N-person games, vol 1. In AAAI 1986, Morgan Kaufmann, Philadelphia, USA, pp 158–162

Neumann JV (1928) Zur theorie der gesellschaftsspiele. Math Ann 100:295–320

Ramezani S, Endriss U (2010) Nash social welfare in multi-agent resource allocation. In: David E, Gerding E, Sarne D, Shehory O (eds) Agent-mediated electronic commerce. Springer, Heidelberg, Germany, pp 117–131

Schuldt A (2011) Multi-agent coordination enabling autonomous logistics. Springer, Heidelberg, Germany

Schuldt A, Gehrke JD, Werner S (2008) Designing a simulation middleware for FIPA multi-agent systems. In: Jain L, Gini M, Faltings BB, Terano T, Zhang C, Cercone N, Cao L (eds) WI-IAT 2008. IEEE Computer Society Press, Sydney, pp 109–113

Schuldt A, Werner S (2007) Towards autonomous logistics: conceptual, spatial, and temporal criteria for container cooperation. In LDIC 2007, Springer, Bremen, Germany, pp 311–319

Sutton RS, Barto AG (1998) Reinforcement learning: an introduction. The MIT Press, Cambridge, USA

Tesauro G (2004) Extending Q-learning to general adaptive multi-agent systems. Adv Neur In 16:871–878

Watkins CJCH, Dayan P (1992) Q-learning. Mach Learn 8(3–4):279–292

Improving Returnable Transport Equipment Management with an Autonomous Approach

Rebecca Aggarwal and Ming K. Lim

Abstract Challenges of returnable transport equipment (RTE) management continue to heighten as the popularity of their usage magnifies. Logistics companies are investigating the implementation of radio-frequency identification (RFID) technology to alleviate problems such as loss prevention and stock reduction. However, the research within this field is limited and fails to fully explore with depth, the wider network improvements that can be made to optimize the supply chain through efficient RTE management. This paper, investigates the nature of RTE network management building on current research and practices, filling a gap in the literature, through the investigation of a product-centric approach where the paradigms of "intelligent products" and "autonomous objects" are explored. A network optimizing approach with RTE management is explored, encouraging advanced research development of the RTE paradigm to align academic research with problematic areas in industry. Further research continues with the development of an agent-based software system, ready for application to a real-case study distribution network, producing quantitative results for further analysis. This is pivotal on the endeavor to developing agile support systems, fully utilizing an information-centric environment and encouraging RTE to be viewed as critical network optimizing tools rather than costly waste.

Keywords Intelligence · Product-centric · Radio-frequency identification · Returnable transport equipment · Autonomous

R. Aggarwal (✉) · M. K. Lim
Engineering Systems and Management, Aston University,
Birmingham B4 7ET, UK
e-mail: aggarwar@aston.ac.uk

H.-J. Kreowski et al. (eds.), *Dynamics in Logistics*, Lecture Notes in Logistics,
DOI: 10.1007/978-3-642-35966-8_37,

Introduction

The circulation of returnable transport equipment (RTE) within logistics networks continues to grow as the use of items such as cages, pallets, and totes becomes more prominent throughout the supply chain. A recent survey carried out by Bridge, a EU sponsored project, highlighted an increase in demand of companies which own or rent RTE whereby the current figure stands at 67 % (France and partners 2007). RTE management operations can be of a large scale with millions of units in distribution globally, flowing through an array of industries, for example, Container Centralen, a RTE purchasing and hiring company, monitors and controls 3.5 million units worldwide (Centralen 2011). The multifarious nature of RTE and the scale of the networks provide challenges, therefore, developing and implementing strategies to achieve optimal movement and storage of RTE is imperative in order for companies to retain the level of agility required to remain competitive.

The management of RTE is currently prone to problems such as high costs, shrinkage and losses (Rosenau et al. 1996; Johansson and Hellström 2007; France and partners 2007; Ilic et al. 2009). These problems can significantly decrease the return on assets of RTE, affecting network viability and increasing overall supply chain costs. Difficulties of RTE management arise due to the inadequate support systems which are currently failing, leading to vast inefficiencies. One study reported a loss of up to 50 % of RTE throughout the network (Ilic et al. 2009). Losses incur equipment and product costs which have knock-on negative effects until the equipment is replaced. Poor RTE management leads to inefficient networks, lower RTE utilization rates and a decrease in network optimization of storage and movement of goods.

A solution to managing RTE is the implementation of radio-frequency identification (RFID) technology. RFID is an auto-id technology in the form of tags, which can send signals to readers, of any tagged items. RFID can provide visibility of the RTE network preventing losses and slow returns. Companies such as Walmart, Tesco and Container Centralen cite authentication of RTE and reduction in paperwork as some of the positive outcomes (Collins 2005, 2006; Wessel 2011). However, it can be difficult for retailers to achieve a positive return as the quantity of RFID installation required downstream is higher incurring higher costs however, the reduction of pallet losses and the removal of manual processes, lowers the overall cost of moving pallets encouraging positive gains (Bottani and Bertolini 2009). However, further exploration of RFID application improvements to RTE network management needs to be conducted in order to solidify results and encourage adoption. This paper seeks to investigate the application of RFID to reduce RTE problems and in addition, when combined with other technologies, optimize the RTE and the goods flowing through the network.

This paper is organized as follows, in section Returnable Transport Equipment Management, there is a description of RTE functions with an exploration of problem reduction followed by a description of RFID enabled improvements which can lead to supply chain optimization. In section Product-Centric Approaches, the concept of product-centric approaches is explored through analyzing the literature of "intelligent" products and "autonomous" objects highlighting where the concept of "intelligence" should be placed. In section Unlocking RTE as a Network Optimizing Tool, the paper proposes a dynamic information-centric environment embedded with advanced software agents, able to conduct decisions on behalf of the RTE, creating network agility able to cope with the aggressive and changeable logistics environments of today. This moves beyond the current practices of RFID application to RTE, creating network optimizing units, encouraging supply chain practitioners and researchers to view RTE not as a network waste but a critical network optimizing tool. Finally, a conclusions and further research section stresses the importance of further exploration of agent based modeling (ABM) within the context of RTE management improvement.

Returnable Transport Equipment Management

RTE Function

RTE can include; pallets, crates, totes, boxes, trollies, dollies and stillages with multiple functions which increase as the equipment is integrated further into efficiency enhancing supply chain roles (Logicacmg 2004). Firstly, the most important function of RTE is to move, store and handle products to the right place, at the right time and in the right condition. Secondly, the equipment needs to protect the goods and act as a stable packaging unit. Thirdly, the RTE needs to be reusable with supporting logistics systems in place to action efficient returns in order to compete with one way disposable packaging particularly as environmental concerns increase (Prendergast 1995). However, it is in the latter function, inadequate support systems, which is leading to a failure of RTE management and a draining of resources throughout logistics networks hence, further investigation into system improvements is being investigated further in this paper.

Effective RTE Management

There are multiple levels of strategy required to achieve an effective RTE network as illustrated in Fig. 1. Once RTE management problems have been resolved and an approach of continuous network improvement implemented, the next stage is to develop a path of optimization. A technology which can support RTE network improvement, particularly through the virtue of its visibility enhancing capability and technological potential, is RFID.

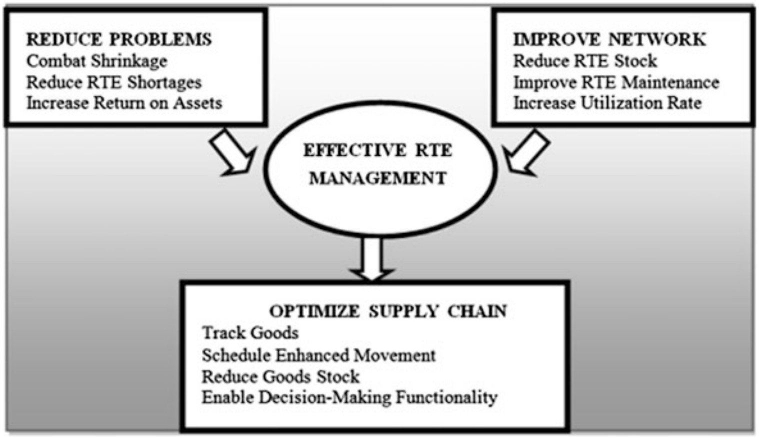


Fig. 1 Diagram illustrating effective RTE management

RTE Problem Reduction

(1) *Combat Shrinkage*: Substantial monetary losses can occur throughout the RTE cycle whereby lack of visibility is cited as the main factor (Johansson and Hellström 2007). In the automotive sector alone, it is projected that RTE losses can cost companies such as Mercedes-Benz, $750 million dollars annually as one piece of equipment can cost up to $2,000 (Wasserman 2010). RFID can be applied to each RTE, projecting a unique identification, which increases RTE network visibility and in some case studies has resulted in the elimination of shrinkage completely (Hellstrom 2009).

(2) *Reduce RTE Shortages*: Unequal management of supply and demand leads to shortages which increases the level of inefficiencies as transportation and storage processes stall. Tracking can provide visibility to ensure that the network holds an adequate number of RTE to satisfy demand, prevent bottlenecks, enhance flows and achieve a higher level of on time deliveries.

(3) *Increasing Return on Assets*: RTE can be expensive in comparison to a one flow disposable packaging system as additional costs are tied up in return flows (Rosenau et al. 1996). The reusable nature of RFID tags compliments RTE networks creating a strong return on investment case for the application of the technology (Bacheldor 2006b). RFID can improve RTE utilization, encouraging optimal movement patterns, leading to a higher return on equipment as illustrated by a case study conducted at Arla Foods where the estimated payback period for RFID application to RTE was 14 months (Hellstrom 2009).

Network Improvement

(1) *Reduce RTE Stock*: RTE can be expensive, making investment costly and inefficient, particularly, when lying idle, empty or in storage. Therefore, in order to improve performance of the network, it is essential to keep the RTE stocks as low as possible. Packaging Logistics Services, aim to reduce the level of safety stock whilst still maintaining RTE availability through the application of RFID to 2.5 million RTE whereby potentially a 30 % reduction can be achieved (Wasserman 2010; Swedberg 2009).
(2) *Improve RTE Maintenance*: Maintenance of RTE offers a proactive approach, preventing breakages and ensuring a continuation of cycles. Therefore, introducing regular checks to clean and maintain the equipment is essential in order to have a healthy RTE lifecycle as RFID provides monitoring capability.
(3) *Increase RTE Utilization*: The increased rate of utilization for each piece of equipment can enable return on asset improvements therefore, enhancing viability of RTE network implementation. Cycle time reduction, through bottleneck prevention and the minimization of idle and empty time is essential. RFID application enhances visibility which improves RTE scheduling (Ilic et al. 2009). Once the RTE network is improved then the conversion into network optimizing units can begin.

Optimize Supply-Chain

(1) *Track Goods*: There are two aspects of monitoring, the RTE and the goods. CHEP applied RFID to crates in order to monitor the crates for their own system and the goods within the crates for the customer (Bacheldor 2006a). RFID has the storage and rewritable capability to facilitate both. Therefore, each tag has a double monitoring function every time it is scanned which is essential when taking in account overall network optimization moving beyond RTE.
(2) *Schedule Enhanced Movement*: RFID enhanced visibility improves tracking of both the RTE and goods which DHL takes advantage of in order to encourage faster speeds of delivery, utilizing more accurate and reliable data of 1.3 million pallets tagged for the Metro Group (Wessel 2008). Further improvements such as reduction of labor and manual errors and speedier unloading and loading times has improved the schedules of movement and storage producing higher rates of customer satisfaction (Swedberg 2010, 2011a).
(3) *Reduce Goods Stock*: Stock reduction and the injection of agility are essential and the RFID enabled real-time data can be utilized to improve forecasting

and replenishment systems whilst monitoring the storage levels and implementing stock reduction policies.

(4) *Enable Decision-Making Functionality*: In large networks it can be effective to decentralize the decision-making functions so that better solutions are achieved within the environment (Lee and Whang 1999). When decisions are made using local information it can be challenging to achieve goals on a global scale optimizing the whole network (Lu et al. 2012). However, it is possible to create levels of autonomy within networks so that decisions are made on a local level whilst optimization on a global scale are achieved (Hülsmann et al. 2008). The possibility of advanced decision-making is explored further from a RTE perspective by considering product-centric approaches in order turn RTE into network optimizing units.

Product-Centric Approaches

In a bid to improve flows, research in product-centric approaches in supply chain applications have been increasing (Rönkkö et al. 2007; Baïna et al. 2009; Bardaki et al. 2011). Baïna et al. (2009) use the approach to resolve product based data inaccuracies which cripples the supply chain and stress the importance of having full and complete information throughout the network in order to increase intelligence.

Bardaki et al. (2011) emphasize the significance of RFID technology in object tracking as the information retrieved offers full visibility, initializing an product-centric approach. Although the amount of data related inefficiencies can be amplified due to the increased quantity of data transactions, logistics service providers are encouraged to move away from the current location based model to a more product-centric approach increasing responsiveness and customizability (Rönkkö et al. 2007; Holmström et al. 2009). There are currently two prominent paradigms within the area of product-centric approaches to supply chain management improvements, intelligent products which focuses on the information side and autonomous objects which explores the decision-making elements, both areas are explored further.

Intelligent Products

Increasing the "intelligence" of items through a product-centric approach within supply chain management, centers on product information management and tracking with full visibility (Holmström et al. 2009; Yang et al. 2009; Valckenaers et al. 2009; Kiritsis 2011; Meyer et al. 2009; Kärkkäinen et al. 2003). Holmström et al. (2009) focus their attention on creating change by implementing systems based on continuous monitoring of products throughout the supply chain rather than being scanned at set locations. RFID can provide the capability needed to retrieve and transfer the relevant information whilst updating its own status. Kiritsis (2011) builds product based intelligence via four levels, whereby the application of RFID tags is essential to enable combination with sensors, higher memory and provide advanced communication for integration with other technologies. Other literature within the paradigm of "intelligence" has taken the definition further to consider agent application (Meyer et al. 2009; Yang et al. 2009; Valckenaers et al. 2009). Valckenaers et al. (2009) recognize that in order to encompass the notion of intelligence there needs to be a decision-making element. They state that an intelligent product needs to be combined with an intelligent agent. Yang et al. (2009) embody intelligent product theory by attaching "intelligent data units" to each product. The purpose of the unit is to provide lifecycle data where each unit is packed with sensors, memory and communication aspects. The information is then transferred to a software agent in order to process the relevant information for lifecycle management. The intelligence of the product is created through a data information retrieval function and a software agent for data analysis.

This paper emphasizes that a viable product-centric approach requires both an information-centric approach which, can be enabled by RFID, and intelligent agents to represent each unit to enable decision-making functionality. The concept of decision-making in agents is taken further in the literature under autonomous objects.

Autonomous Objects

Autonomous control is defined by Windt et al. (2008) as "the ability of logistic objects to process information, to render and to execute decision on their own". The literature around intelligence focuses mainly on the collection, process and transferal of information with the notion of decision-making however, the thread of autonomy takes this further, although it is in early stages. This moves beyond "intelligent" products as they make their own decisions based on the relevant information collected from the environment and critically the learning functions which can be implemented. A key attribute of autonomous control is the decentralization of decision-making to the products. Windt et al. (2008) explores decentralized decision-making, which improves efficiency of production logistics, enhancing responsiveness of machines and improving system flexibility. This highlights the potential of autonomous objects in the supply chain and the ability to enable them to act independently. Currently much of the research is conceptual and in early development. Further research is required into the potential within this area in preparation for applications of these concepts to real world cases.

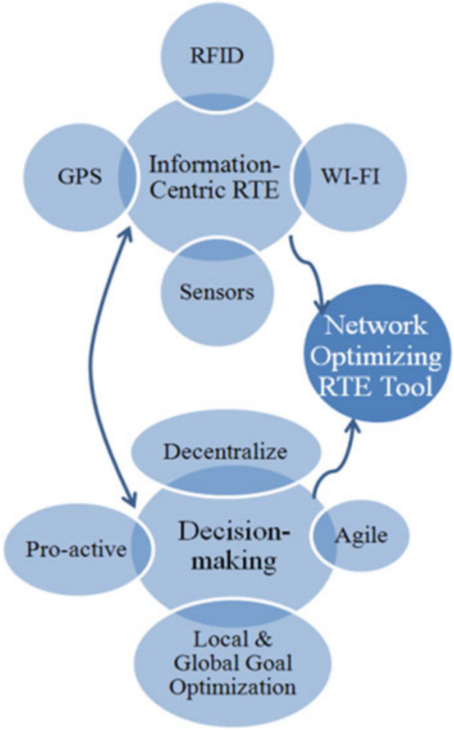


Fig. 2 Diagram illustrating the essentials for an RTE optimizing tool

After reviewing the literature of autonomous and intelligent objects, the two key requirements that need to be added to the RTE network in order to facilitate network optimization and RTE network improvement are decision-making functionality and information-centric RTE, as illustrated in Fig. 2. The relationship between the two areas is discussed further in the section Unlocking RTE as a Network Optimizing Tool.

Unlocking RTE as a Network Optimizing Tool

RTE play a prominent and pivotal role within logistics networks providing a constant contact point to the items being handled. To transform RTE into a network optimizing tool, improving product and RTE flows, invoking the concepts of "intelligence" and "autonomy", two approaches need to be considered simultaneously; information-centric RTE and software agents with a high level of decision-making capability as illustrated in Fig. 2. Both functions are intertwined, affecting each other's ability and potential. The quality of decision-making is directly related to the information fed into the environment therefore, both are important in relation to real-time RTE management and network optimization.

Information-Centric RTE

The quality, quantity and placement of information are crucial in order to create a network optimizing RTE tool. RFID is an essential technology towards achieving full product tracking and visibility as when tagged to each RTE can offer item level tracking which can be further enhanced when combined with GPS, GPRS and WI-FI capability offering full wireless network visibility (Bardaki et al. 2011; Swedberg 2011b; Kärkkäinen et al. 2003). Increasing the information capability around products can enhance decision-making process efficiency, for example, lowering lead-times and improving vehicle delivery, which can lead to improved profits through increasing customer satisfaction and decreasing labor costs (Klein and Thomas 2009; Kim et al. 2008). However, further extension of decision-making functionality is required in order to fully utilize access to rich and large amounts of real-time data hence the exploration into the field of autonomous technology application. A decentralized decision-making concept is considered in order to increase network flexibility and responsiveness, utilizing the real-time knowledge gathered throughout the chain via RFID enabled RTE. Information is only useful if it is visible and retrieved for its function and purpose. By using RFID, agents can be fed continual pools of information at reasonable costs. The key is to extract the right data and install the appropriate functions to utilize and manipulate the data, aiding efficient decision-making. RFID technology is crucial in order to offer unique identification and automated communication to different parts of the supply chain (Kärkkäinen et al. 2003). In a RTE network, the automation of communication is essential in developing agent capability. Therefore, an important attribute of RFID enabled RTE management is the automation of data retrieval and transfer. This is particularly crucial in order to facilitate the decision-making functions.

Decision-Making

The complexity of supply chains continue to increase, challenging optimization, particularly, in a centralized decision-making format. The decision-making process tends to be centralized in order for management to remain in control and save costs. However, the dynamic, distributed and intricate nature of supply chains lends to a structure which favors a more decentralized approach enhancing agility. Agent based models can be structured with the agility to cope with the dynamic and complex nature of supply chains (Ying and Dayong 2005; Davidsson et al. 2005). Especially when software agents are combined with RFID technology, "intelligence" can be increased, and the data, information exchange and communication between the agents and environment, improved (Dias et al. 2009). Hence the combination of RFID and distributed agents can be an effective way of initializing a product-centric approach (Rönkkö et al. 2007).

A RTE-centric approach to supply chain management means that each RTE has an effective stake in the optimality of the supply chain. Therefore, every decision carried out by each RTE would either hinder or improve the network. A positive outcome is reliant on the representation of each RTE and the capability of the decision support functions. These may be contradictory and therefore, agents need to be packed with negotiating ammunition in order to resolve any conflicts of RTE scheduling which may arise. Information of the product and its environment is just as important as the decision-making support functions of the agents. The agents can only utilize their decision-making capabilities based on the information fed into them. As Fig. 2 illustrates, in order to achieve a network optimizing RTE tool, where both RTE management problems can be reduced and the wider network improved and optimized, data needs to be fed from information-centric RTE. This information based decision-making can be used to introduce agility and a pro-active nature.

Agents can be built with learning functions to process the other agent's behavior and the interactions with the environment (Panait and Luke 2005). Logistically, the activities such as scheduling can be observed and decisions can be made by other agents proactively. In addition, if there are unexpected changes and scenarios the knowledge gained from the learning functions can enable less damage to be caused to the outcome. Supply chains and RTE networks need to be agile in order to respond to unexpected changes. A multi-agent system can offer a path to agility for supply chains through the high-level coordination which can be programmed for agent interactions (Lou et al. 2004). In addition, the combination of real-time data gathered from RFID enabled RTE with efficient decision-making of software agents on a RTE level, can enable increased network agility. The combination of information-centric RTE and intelligent agents programmed with rules and communication capability instills network flexibility leading to the creation of autonomous RTE.

Conclusion and Further Research

RFID technology has been used to improve the management of RTE in certain areas such as loss reduction, process re-engineering and tracking. However, the potential of information-centric approaches combined with ABM to RTE management has been neglected. This paper highlights the potential of aligning information-centric RTE with decentralized and autonomous object based decision-making functions. RFID, which in some cases is already implemented on RTE, can be used further to provide valuable information, enhancing the environment of the software agents representing the RTE. The improved environment along with the appropriate level of algorithmic intensity can enable the agents to conduct research based on real-time knowledge, improving the decision-making and producing better solutions for scheduling. The addition of autonomy through agents enables the RTE to develop from supply chain burdens to network optimizing tools particularly, if the levels of agents are functioning to optimize not only the RTE but the goods to be transported, increasing the level of agility.

Further research is required under the ABM paradigm in relation to RTE network management with the development of a conceptual model followed by validation. Simulations will be conducted in comparison to a real-world case study and an assessment of the viability of ABM to represent RTE. A program of agents representing RFID enabled RTE will be linked in real-time to make effective and efficient decisions, competent for action in today's challenging logistics networks.

References

Bacheldor B (2006a) CHEP RFID-enables reusable containers for auto industry [Online]. RFID J Available: http://www.rfidjournal.com/article/view/2859 Accessed 06/04/2011 2011

Bacheldor B (2006b) Reusable pallet and container coalition studies RFID [Online]. RFID J Available: http://www.rfidjournal.com/article/view/2434 Accessed 09/08/2011 2011

Baïna S, Panetto H, Morel G (2009) New paradigms for a product oriented modelling: case study for traceability. Comput Ind 60:172–183

Bardaki C, Kourouthanassis P, Pramatari K (2011) Modeling the information completeness of object tracking systems. J Strategic Inf Syst (in press, corrected proof)

Bottani E, Bertolini M (2009) Technical and economic aspect of RFID implementation for asset tracking. Int J RF Technol Res Appl 1:169–193

Centralen C (2011) CC at a glance [Online]. Available: http://www.container-centralen.com/en/about%20cc/cc%20at%20a%20glance.aspx Accessed 02/09/2011 2011

Collins J (2005) Europe finds its own path to ROI [Online]. RFID J Available: http://www.rfidjournal.com/article/purchase/1933 Accessed 06/04/2011 2011

Collins J (2006) Tesco revises RFID plans [Online]. RFID J Available: http://www.rfidjournal.com/article/articleview/2243/1/1/ 19/02/2010

Davidsson P, Henesey L, Ramstedt L, Törnquist J, Wernstedt F (2005) An analysis of agent-based approaches to transport logistics. Transp Res Part C Emerg Technol 13:255–271

Dias JCQ, Calado JMF, Osorio AL, Morgado LF (2009) RFID together with multi-agent systems to control global value chains. Annu Rev Control 33:185–195

France G, Partners W (2007) Returnable transport items: requirements to improve reusable asset management. In: IST-2005-033546 ECCN (ed) Building radio frequency identification for the global environment, BRIDGE WP09. Available at http://www.bridge-project.eu/index.php/bridge-public-deliverables/en/

Hellstrom D (2009) The cost and process of implementing RFID technology to manage and control returnable transport items. Int J Logistics Res Appl Leading J Supply Chain Manage 12:1–21

Holmström J, Kajosaari R, Främling K, Langius E (2009) Roadmap to tracking based business and intelligent products. Comput Ind 60:229–233

Hülsmann M, Grapp J, Li Y (2008) Strategic adaptivity in global supply chains—competitive advantage by autonomous cooperation. Int J Prod Econ 114:14–26

Ilic A, Ng JWP, Bowman P, Staake T (2009) The value of RFID for RTI management. Electron Markets 19:125–135

Johansson O, Hellström D (2007) The effect of asset visibility on managing returnable transport items. Int J Phys Distrib Logistics Manage 37:799

Kärkkäinen M, Holmström J, Främling K, Artto K (2003) Intelligent products–a step towards a more effective project delivery chain. Comput Ind 50:141–151

Kim J, Tang K, Kumara S, Yee S-T, Tew J (2008) Value analysis of location-enabled radio-frequency identification information on delivery chain performance. Int J Prod Econ 112:403–415

Kiritsis D (2011) Closed-loop PLM for intelligent products in the era of the internet of things. Comput Aided Des 43:479–501

Klein T, Thomas A (2009) Opportunities to reconsider decision making processes due to auto-ID. Int J Prod Econ 121:99–111

Lee H, Whang S (1999) Decentralized multi-echelon supply chains: incentives and information. Manage Sci 45:633–640

Logicacmg (2004) Making waves: Rfid adoption in returnable packaging. LogicaCMG Nederland BV, The Netherlands. Available at http://www.can-trace.org/portals/0/docs/logicacmg_rfid_benchmark_study.pdf

Lou P, Zhou Z-D, Chen Y-P, Ai W (2004) Study on multi-agent-based agile supply chain management. Int J Adv Manuf Technol 23:197–203

Lu SYP, Lau HYK, Yiu CKF (2012) A hybrid solution to collaborative decision-making in a decentralized supply-chain. J Eng Tech Manage 29:95–111

Meyer GG, Främling K, Holmström J (2009) Intelligent products: a survey. Comput Ind 60:137–148

Panait L, Luke S (2005) Cooperative multi-agent learning: the state of the art. Auton Agent Multi Agent Syst 11:387–434

Prendergast GP (1995) The EC directive on packaging and packaging waste: current status and logistical implications. Logistics Inf Manage 8:10–17

Rönkkö M, Kärkkäinen M, Holmström J (2007) Benefits of an item-centric enterprise-data model in logistics services: a case study. Comput Ind 58:814–822

Rosenau WV, Twede D, Mazzeo MA, Singh SP (1996) Returnable/reusable logistical packaging: a capital budgeting investment decision framework. J Bus Logistics 17:139–165

Swedberg C (2009) PLS uses RFID to track pallets, containers [Online]. RFID J Available: http://www.rfidjournal.com/article/view/5043 Accessed 06/04/2011 2011

Swedberg C (2010) Axios MA launches tagged pallets and real-time tracking solution [Online]. RFID J Available: http://www.rfidjournal.com/article/view/8014 Accessed 06/04/2011 2011

Swedberg C (2011a) Monroe Americana tracks reusable containers [Online]. RFID J Available: http://www.rfidjournal.com/article/view/8316/ Accessed 06/04/2011 2011

Swedberg C (2011b) TrakLok offers hybrid system for tracking containers [Online]. RFID J Available: http://www.rfidjournal.com/article/view/8504 Accessed 26/08/2011

Valckenaers P, Germain BS, Verstraete P, van Belle J, Hadeli K, van Brussel H (2009) Intelligent products: agere versus essere. Comput Ind 60:217–228

Wasserman E (2010) Automakers seek fast ROI from RFID [Online]. RFID J Available: http://www.rfidjournal.com/article/purchase/7460 Accessed 06/04/2011 2011

Wessel R (2008) Metro group, DHL to roll out RFID in France [Online]. RFID J Available: http://www.rfidjournal.com/article/view/4277/ Accessed 7/04/2011 2011

Wessel R (2011) Container Centralen says RFID provides 'substantial' cost savings [Online]. RFID J Available: http://www.rfidjournal.com/article/view/8885/1 Accessed 21/10/2011

Windt K, Böse F, Philipp T (2008) Autonomy in production logistics: identification, characterisation and application. Rob Comput Integr Manuf 24:572–578

Yang X, Moore P, Chong SK (2009) Intelligent products: from lifecycle data acquisition to enabling product-related services. Comput Ind 60:184–194

Ying W, Dayong S (2005) Multi-agent framework for third party logistics in E-commerce. Expert Syst Appl 29:431–436

Stability Analysis Scheme for Autonomously Controlled Production Networks with Transportations

Sergey Dashkovskiy, Michael Görges and Lars Naujok

Abstract In this paper, we present a scheme for the stability analysis of autonomously controlled production networks with transportations. We model production networks by differential equations and discrete event simulation models (DES) from a mathematical and engineering point of view, where transportation times are considered in the models as time delays. Lyapunov functions as a tool to check the stability of networks are used to calculate stability regions. Then, this region is refined using the detailed DES. This approach provides a scheme to determine stability regions of networks with less time consumption in contrast to a pure simulation approach. In presence of time delays, new challenges in the analysis occur, which is pointed out in this paper.

Introduction

Production networks are used to describe company or cross-company owned networks with geographically dispersed plants (Wiendahl and Lutz 2002), which are connected by transport routes. One of the approaches to handle such complex systems is to shift from centralized to decentralized or autonomous control.

S. Dashkovskiy (✉)
Department of Civil Engineering, University of Applied Sciences Erfurt, Erfurt, Germany
e-mail: sergey.dashkovskiy@fh-erfurt.de

M. Görges
BIBA-Bremer Institut für Produktion und Logisitk GmbH at the University of Bremen, Bremen, Germany
e-mail: goe@biba.uni-bremen.de

L. Naujok
Centre for Industrial Mathematics, University of Bremen, Bremen, Germany
e-mail: larsnaujok@math.uni-bremen.de

H.-J. Kreowski et al. (eds.), *Dynamics in Logistics*, Lecture Notes in Logistics, DOI: 10.1007/978-3-642-35966-8_38,

In this paper, we consider a certain autonomously controlled production network scenario consisting of six interconnected plants focus. In the context of production networks, the concept of autonomous control enables intelligent logistic objects such as parts and orders, for example, to decide about routes through the system autonomously. The concept of autonomous control aims at affecting the systems performance positively (Windt and Hülsmann 2007).

Different autonomous control methods have been developed in the literature (Scholz-Reiter et al. 2011b). In this paper, the local rational autonomous control method queue length estimator (QLE) is considered for autonomous decision making on the network level and on the shop-floor level. On the shop-floor level, the QLE enables parts to choose a workstation according to local information about their current workload. In contrast, parts using this method on the network level estimate the waiting times at succeeding production plants. According to this method, they will choose the next possible production plant with the shortest estimated waiting time.

Stability of a production network means that the work in progress (WIP) remains bounded over time. Instability, by means of an unbounded growth of the WIP, may cause high inventory costs, downtimes of machines or loss of customers, for example. The implementation of autonomous control methods can lead to instability of the system (Windt 2006; Philipp et al. 2007). Hence, it is necessary for logistic systems to derive parameters, which guarantee stability.

For the stability analysis, we provide a dual approach: a mathematical and an engineering point of view. Based on retarded functional differential equations, Lyapunov-Razumikhin functions (Teel 1998) are used to calculate a stability region, which includes parameters for which the network is stable (Dashkovskiy and Naujok 2010c; Dashkovskiy et al. 2010a). These stability parameters are implemented into a more detailed microscopic model, where all plants are represented by a complete shop floor. This microscopic view, models the scenario with the help of a discrete event simulation (DES) tool. Using this approach, the calculated stability region will be refined. The advantage is that we first apply the mathematical theory to find in a very fast way those parameters, where stability is guaranteed. A refinement is performed by simulations, in order to enlarge the set of parameters, which guarantee stability. This scheme provides an identification of a stability region with less time consumption in contrast to a pure simulation approach, where the time needed for the simulations increases exponentially by increasing the number of plants, parts and machines.

An existing work on stability analysis for production networks without transportations can be found in Scholz-Reiter et al. (2011a), where the mentioned scheme was firstly introduced. Here, we adopt this approach to networks with transportations using Lyapunov tools for networks presented in Dashkovskiy and Naujok (2010c) and applied in Dashkovskiy et al. (2010a, 2011a). In presence of transportations, i.e., time delays, the dynamics of the network is much more complex than the dynamics without transportations. The worst-case approach of the mathematical analysis this leads to a rough calculation of the stability region that is then calculated more precisely with help of DES. The identification of stable or unstable behavior of the network using the DES for the refinement of the stability region is a challenging task and the abort criterion used in Scholz-Reiter et al. (2011a) needs to be adapted for networks with time delays.

Modeling

For the illustration of the novel stability analysis approach a particular production network has been chosen, which is described in this section.

The production network in Fig. 1 consists of six geographically distributed production locations, which are connected by transportation routes. In this paper we consider the material flow between the locations, described in Fig. 1 by arrows. In this scenario the $x_i(t) \in R$ for $i = 1,\ldots, 6$ represents the WIP of the ith location at time t, where $t \in R_+$ and R_+ denotes all positive real values. In the rest of this paper for the ith production location we write *subsystem i*. The network of all six subsystems we name simply whole system.

Each plant of the network is represented by a complete shop floor scenario. It consists of three parallel production lines. Every line has three workstations and an input buffer in front of each workstation. The structure allows the parts to switch lines at every stage. The decision about changing the line is made by the part itself by internal control rules. This rule on the shop floor level is the QLE.

Subsystem 1 gets some raw material from an external source, denoted by $u(t) \in R$ and some material from subsystem 6. The material will be processed with a certain production rate $\tilde{f}_1$. Then, a truck loads the processed parts and transports them to the subsystem 2 or 3, according to the QLE. The transportation time from subsystem i to j is denoted by τ_{ij}. The parts will be processed with the rate $\tilde{f}_2$ or $\tilde{f}_3$ and sent to subsystem 4 or 5, according to the QLE. After processing the parts with the rates $\tilde{f}_4$ and $\tilde{f}_5$ they will be sent to subsystem 6 and processed there $\tilde{f}_2$with the rate $\tilde{f}_6$. Then, 90 % of the production will be delivered to some customers outside of the network and 10 % of the production of subsystem 6 will be sent back to subsystem 1. This can be interpreted as recycling of the waste produced in subsystem 6, for example.

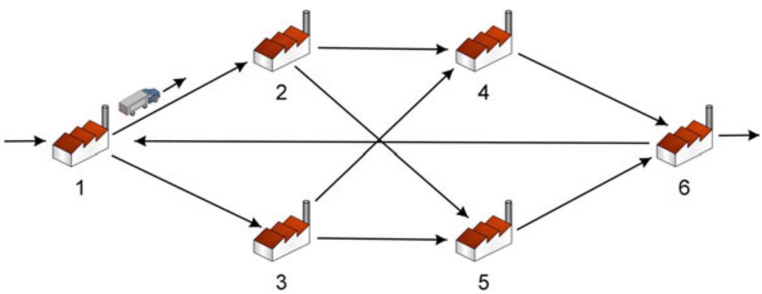


Fig. 1 The particular production network

There are two levels of aggregation and modeling. The macroscopic view focuses on the network level, which consists of production plants only. On the microscopic level, the network is represented more detailed. In addition to the macroscopic view, the microscopic view represents the plants as a set of interconnected machines. In the following these two views will be described.

Aggregated View Using ODEs

In the macroscopic approach we provide our description and analysis from a mathematical point of view. The internal structure on the shop floor level of all subsystems is ignored. All subsystems are autonomously controlled by means of an autonomous adjustment of the production rates. As in Scholz-Reiter et al. (2011a), the production rate for the subsystem i can be modeled by

$$\tilde{f}_i(x_i(t)) := \alpha_i(1 - \exp(-x_i(t))), i = 1, \ldots, 6,$$

where $\alpha_i \in R_+$ is the (constant) maximal production rate of the subsystem i. Note that one can choose any other rate, which fits to a certain scenario. $\tilde{f}_i$ converges to α_i, if the WIP of the subsystem i is large and $\tilde{f}_i$ tends to zero, if the WIP of the subsystem i tends to zero. Accordingly, a huge influx of raw material causes an increase of the production rate close to the maximum, whereas less influx of raw material leads to a production rate, which is almost zero.

When modeling the system by retarded differential equations, we assume that the processed material will be transported at time t to a subsystem according to the QLE and arrives at the succeeding subsystem at the time $t + \tau_{ij}$, where τ_{ij} can be interpreted as transportation time needed for the transportation from subsystem i to j. Here, no delay within the production process is implemented. Note that one can use variable transportation times τ_{ij} instead of constant ones such as state- or time-dependent variables. For example, disturbances on the transport routes can be taken into account choosing variable τ_{ij}.

A retarded differential equation describes the rate of change of the WIP along the time. We model the network by retarded differential equations as follows:

$$\begin{aligned}
\dot{x}_1(t) &:= u(t) + \frac{1}{10}\tilde{f}_6(x_6(t - \tau_{61})) - \tilde{f}_1(x_1(t)),\\
\dot{x}_2(t) &:= \tilde{c}_{12}\tilde{f}_1(x_1(t - \tau_{12})) - \tilde{f}_2(x_2(t)),\\
\dot{x}_3(t) &:= \tilde{c}_{13}\tilde{f}_1(x_1(t - \tau_{13})) - \tilde{f}_3(x_3(t)),\\
\dot{x}_4(t) &:= \tilde{c}_{24}\tilde{f}_2(x_2(t - \tau_{24})) + \tilde{c}_{34}\tilde{f}_3(x_3(t - \tau_{34})) - \tilde{f}_4(x_4(t)),\\
\dot{x}_5(t) &:= \tilde{c}_{25}\tilde{f}_2(x_2(t - \tau_{25})) + \tilde{c}_{35}\tilde{f}_3(x_3(t - \tau_{35})) - \tilde{f}_5(x_5(t)),\\
\dot{x}_6(t) &:= \tilde{f}_4(x_4(t - \tau_{46})) + \tilde{f}_5(x_5(t - \tau_{56})) - \tilde{f}_6(x_6(t)),
\end{aligned} \tag{1}$$

where $\tilde{c}_{ij}$ represent the QLE. The external input is chosen according to fluctuations as $u(t) := AV(\sin(t) + 1) + 5$, where $AV \in R_+$.

$x_i(t)$ may also represent other relevant parameters of the system, e.g., the number of unsatisfied orders. One can extend or change the given production network to describe any other scenario that can be more large and complex. It is possible to perform a stability analysis for the extended system.

Detailed View Using DES

By using a DES approach, a more detailed modeling is performed. Due to the lower aggregation level and the discrete nature of this modeling approach some parameters from the aggregated differential equation based model have to be adjusted. The DES represents the flows of materials by discrete parts passing through the network. This requires an adjustment of the input rate in plant 1. In the DES model the arrival rate $u(t)$ is cumulated. Whenever this cumulated arrival rate reaches an integer value a part enters the system at the corresponding time point t. A second adjustment concerns the production rates of all production plants. In the detailed view the plants represent a shop-floor scenario with 3×3 machines. Due to the parallel machines offered by the shop-floor, the production rate of a plant has to be distributed to these parallel machines. In the case at hand each work station j in the plant i has a maximal production rate of $\alpha_{ij} = \frac{\alpha_i}{3}$.

Stability Analysis

In this section, the scheme and the used tools of a stability analysis are described.

We consider nonlinear dynamical systems of the form

$$\dot{x}(t) = f(x^t, u(t)), \tag{2}$$

which are called retarded functional differential equations (RFDE), where $x^t \in C([-\Delta, 0], R^N)$ is defined by $x^t(\tau) := x(t + \tau)\,, \tau \in [-\Delta, 0]$. Δ denotes the maximal involved delay and $C([-\Delta, 0], R^N)$ denotes the Banach space of continuous functions defined on $[-\Delta, 0]$ equipped with the norm $||x||_{[-\Delta,0]} := \max_i \max_{t\in[-\Delta,0]} |x_i(t)|$. We denote the Euclidian norm in R^{n} by $|\cdot|$ and the essential supremum norm for essentially bounded functions u in R_+ by $||u||_\infty$. $u \in R^M$ is the external input of the system, which is an essentially bounded measurable function and $f : C([-\Delta, 0], R^N) \times R^M \to R^N$ is a nonlinear and locally Lipschitz continuous functional to guarantee that the system (1) has a unique solution $x(t)$ for every initial condition $x_0 = \xi$ for any $\xi \in C([-\Delta, 0], R^N)$. An interconnected system is described by RFDEs of the form

$$\dot{x}_i(t = f_i(x_1^t, \ldots, x_n^t, u_i(t)), \tag{3}$$

$i = 1, \ldots, n,$ where $x_i^t \in C([-\Delta, 0], R^{N_i})$, $u_i \in R^{M_i}$ and $f_i : C([-\Delta, 0], R^N) \times R^{M_i} \to R^{N_i}$. Defining $N := \sum_{i=1}^n N_i$, $m := \sum_{i=1}^n M_i$, $x = (x_1^T, \ldots, x_n^T)^T$, $u := (u_1^T, \ldots, u_n^T)^T$ and $f := (f_1^T, \ldots, f_n^T)^T$, (3) can be written in the form (2).

We define local input-to-state stability (LISS) for each subsystem of (3). For system (2), the definition of LISS can be found in Dashkovskiy et al. (2010a). The used classes of functions can also be found in Dashkovskiy et al. (2010a).

Definition 1 The ith subsystem of (3) is called LISS, if there exist constants ρ_i, $\rho_j^i, \rho_i^u > 0$, γ_{ij}, $\gamma_i \in K_\infty$ and $\beta_i \in KL$, such that for all initial functions $||x_i||_{[-\varDelta,0]} \leq \rho_i$, $||x_j||_{[-\Delta,\infty)} \leq \rho_j^i$, $j \neq i$ and all inputs $||u_i||_\infty \leq \rho_i^u$ it holds

$$|x_i(t)| \leq \max\left\{\beta_i\left(||\xi_i||_{[-\varDelta,0]}, t\right), \max_{j \neq i} \gamma_{ij}\left(||x_j||_{[-\varDelta,\infty)}\right), \gamma_i\left(||u_i||_\infty\right)\right\} \tag{4}$$

$\forall t \in R_+$. γ_{ij} and γ_i are called (nonlinear) gains. Note that, if $\rho_i, \rho_j^i, \rho_i^u = \infty$ then the ith subsystem is ISS.

LISS and ISS, respectively, mean that the norm of the trajectories of each subsystem is bounded. Furthermore, we define the *gain matrix* $\Gamma := (\gamma_{ij})$, $i,j = 1, \ldots, n$, $\gamma_{ii} = 0$, which defines a map $\Gamma : R_+^n \to R_+^n$ by

$$\Gamma(s) := \left(\max_j \gamma_{1j}(s_j), \ldots, \max_j \gamma_{nj}(s_j)\right)^T, \; s \in R_+^n. \tag{5}$$

To check, if the whole network has the ISS property a small gain condition is needed, which is of the form

$$\varGamma(s) \ngeq s, \; \forall s \in \mathbb{R}_+^n \backslash \{0\}. \tag{6}$$

Notation $\ngeq$ means that there is at least one component $i \in \{1, \ldots, n\}$ such that $\Gamma(s)_i < s_i$. A local version of the small gain condition (LSGC) can be found in Dashkovskiy and Rüffer (2010b). A useful tool to verify LISS for time-delay systems are Lyapunov-Razumikhin functions (LRF) or Lyapunov–Krasovskii functionals (LKF), see Teel (1998) and Dashkovskiy and Naujok (2010c). In this paper, we use LRF, but the analysis considering LKF is similar. For systems of the form (2) one can find the definition of LRF for example in Teel (1998) and for systems of the form (3) LRFs are defined in Dashkovskiy and Naujok (2010c). With these definitions we quote the following:

Theorem 1 Consider the interconnected system (3). Assume that each subsystem has an LISS-LRF V_i, $i = 1, \ldots, n$. If the corresponding gain-matrix Γ satisfies the LSGC, then the whole system of the form (2) is LISS.

The proof can be found in Dashkovskiy and Naujok (2010c) with corresponding changes according to the LISS property. This Theorem completes the mathematical part of the stability analysis and the identification of the stability region of the network, which is the set of parameter constellations guaranteeing LISS: One has to find a LISS-LRF for each subsystem of (3) and to check, if the LSGC is satisfied. From these conditions the stability region will be identified in a first step. From Theorem 1 we know that the whole network possesses the LISS property.

In a second step the identified stability region will be refined using DES. By LRFs and the LSGC, we obtain a rough estimation of the stability region. The DES can investigate the system and its behavior in a more detailed way. The drawback of the identification of the stability region based only of the DES approach is that one has to simulate all possible combinations of free systems variables. By a linear growth of the number of subsystems and the variables this leads to an exponential growth of time needed for the simulation runs such that a determination of the stability region in an acceptable time is not possible.

The advantage of the presented approach in this paper is that the identification of the stability region using the mathematical approach is possible in a short time, where only few parameter constellations are left for investigation in view of stability. This can be performed by the DES and the stability region of networks can be identified with less time consumption in total in contrast to an approach only based on simulation runs. An illustration of the scheme can be found in Scholz-Reiter et al. (2011a).

Stability Evaluation

In this section, we determine the stability region of the scenario, introduced in section Modeling from a mathematical point of view and refine it using the DES.

Determining Stability Regions

We choose $V_i(x_i) = x_i, i = 1, \ldots, 6$ as the LRF candidates for the subsystems, define the Lyapunov gains for the first subsystem by

$$\chi_u(u(t)) := -\ln\left(1 - \frac{u(t)(||u||_\infty + 0.1 \cdot \alpha_6)}{||u||_\infty (1-\varepsilon_u)\alpha_1}\right), 1 > \varepsilon_u > 0,$$

$$\chi_{61}(||V_6^d(x_6)||) := -\ln\left(1 - \frac{||u||_\infty + 0.1 \cdot \alpha_6}{(1-\varepsilon_{61})\alpha_1}(1 - \exp(-||V_6^d(x_6)||))\right), 1 > \varepsilon_{61} > 0.$$

where $||V_i^d(x_i(t))|| := \max_{s \in [t-\Delta, t]} |V(x(s))|$. The gains of the other subsystems are chosen accordingly and by similar calculations as in (Dashkovskiy and Naujok 2010c), we can show that $V_i(x_i)$ are the LRFs of the subsystems. All subsystems are LISS and the gain-matrix consisting of χ_{ij} satisfies the small-gain condition. All the calculations are skipped, because of the limited space.

To guarantee that the Lyapunov gains are well-defined, we get conditions for α_i: $\alpha_1 > ||u||_\infty + 0.1\alpha_6; \alpha_2 > \alpha_1; \alpha_3 > \alpha_1; \alpha_4 > \alpha_2 + \alpha_3;\ \ \alpha_5 > \alpha_2 + \alpha_3;\ \ \alpha_6 > \alpha_5\ + \alpha_4$ from which we get that subsystem 1 is LISS with $\rho_1^u := \alpha_1 - \alpha_6 > ||u||_\infty$. Note that these conditions are derived only for the particular scenario. For other scenarios, one may get other stability conditions. Subsystems 2–5 are ISS. If these conditions are satisfied for any input $u(t)$, we get by Theorem 1 that the whole system is LISS, i.e., the WIP of the whole system is bounded. Note that the choice of the input $u(t)$ can be arbitrary. The higher $||u||_\infty$ is, the higher the maximal production rates are needed to guarantee the fulfillment of the conditions for α_i and to guarantee stability.

Refining Stability Regions by Using a DES Model

The abort criterion defines unstable states of the network as follows: a simulation run is considered to be unstable whenever the WIP starts to grow persistently in a predefined time interval about 10 % (Scholz-Reiter et al. 2011a). On the basis of this abort criterion, an approach is proposed which reduces the maximal production rate of all plants in different simulation runs in steps of 1 % until the abort criterion is satisfied. The reduced maximal production rates are the results of the refinement. Due to the autonomous decision making on the network level the quantities shipped between the locations may vary. A static approach, which reduces the production rates of all plants uniformly seem not to be suitable. Thus, the approach used in this paper extends the static approach presented above. The new refinement procedure has to take into account that the shipment between plants is related to the production rates of preceding plants. Thus, the refinement procedure varies the maximal production rate of each plant separately.

In order to provide a systematic refinement of the production rates, an approach based on an algorithmic scheme is proposed: The refinement algorithm aims at finding the smallest maximal production rates for all plants guaranteeing stability. In the beginning the algorithm and its parameters are initialized. It starts with an arbitrarily network configuration, which is considered to be stable. The algorithm consists of two iterative loops. The inner iteration reduces stepwise the maximal production rates of the plants at one network stage S and starts the simulation. This iteration will be repeated until the abort criterion is satisfied. After this, the counter S is set to the next network stage and the inner loop will be repeated again, until the counter S equals the number of the last production stage. In this case the counter of the second iteration *i* is increased and the counter S is set again to the first network stage. The second outer iteration causes the repetition of the inner iteration until a pre-defined amount of iteration steps are reached.

Results of the Stability Analysis

We set $\tau_{ij} = 1$. The results of the analysis, i.e., the identification of stability regions, are the same for other values of τ_{ij}. Figure 2 shows the results of the mathematical stability analysis for the production plant 1 against the arrival rate amplitude variation AV. The figures for the other plants are similar. Within the mathematically identified stability region, stability of the network can be guaranteed. However, below the border of this region stable systems behavior neither can be guaranteed nor negated. In this area, the stability of the network is investigated by simulations. For the refinement of the stability region the algorithm described above is applied, where the calculated bounds of the stability region are the initial values. The simulated stability region of Fig. 2 is the result of the refinement. The simulation model has stable behavior above the simulative derived stability border. Below this border, unstable behavior was observed.

Comparing the results of the mathematically determined and the refined stability regions, it can be noticed that gap between the mathematical and the simulated results grows with increasing AV. For AV = 60, the difference between the simulated and the calculated bounds of stability is 67 %. This can be explained by the usage of the worst case within the mathematical stability property ISS, namely the supremum norm. In particular, for oscillating inputs, like in the case at hand the maximal value is used to derive stability parameters. By increasing AV this leads to bigger mathematical bounds of stability. Figure 3 illustrates the results of the simulation based refinement more detailed. It depicts the WIP against time of all plants for a simulation run with AV = 4. The left column of Fig. 3 shows the results for the stable situation ($\alpha_1 = 10.4$, $\alpha_2 = 5.41$, $\alpha_3 = 5.41$, $\alpha_4 = 5.2$,

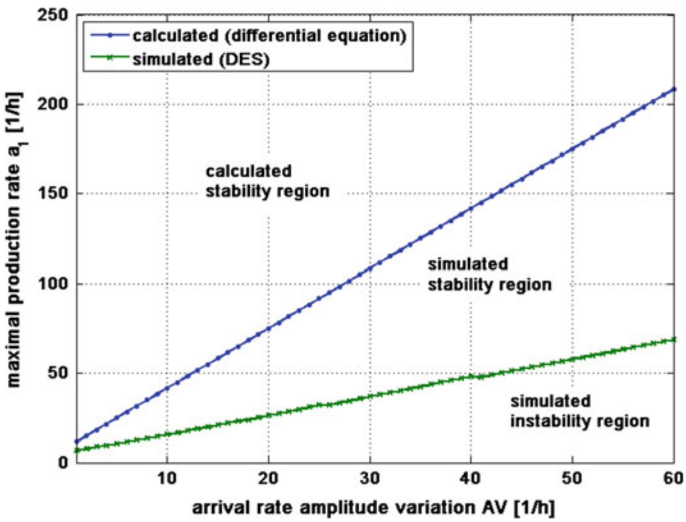


Fig. 2 Stability region of plant 1 for increasing AV

$\alpha_5 = 5.2$, $\alpha_6 = 12.13$). By contrast the right column shows the results of an unstable situation ($\alpha_2 = 4.98$ and $\alpha_3 = 4.98$). The results of the right column are determined by reducing the maximal production rates of network stage by 1 %.

In the unstable situation, the maximal production rate of plant 2 and plant 3 is not sufficient to process the incoming material. The WIP of this plant starts to grow consequently and the network is considered to be unstable. This example shows that the mathematically determined stability region can be refined to a sharp and precise bound of stability. Note that the oscillating behavior is caused by the presence of time delays.

Summarizing these results it shows the symbiotic character of the dual approach combining mathematical stability analysis and simulation. The simulation of the model leads to sharp and accurate stability borders. However, without a properly chosen start configuration a time intensive trial and error approach is necessary. Here the mathematical theory helps to find parameter constellations which guarantee the stability of the network. These results are used as start parameters in the simulation model. This reduces the range of possible parameters to test in the simulation approach. Accordingly, the presented stability analysis scheme can be performed more efficient compared to a trial and error approach.

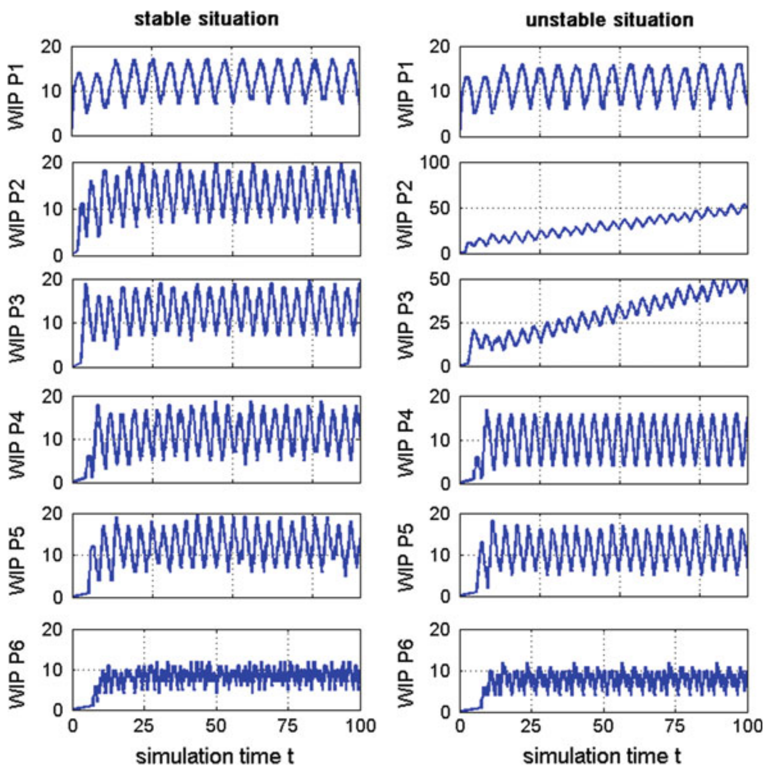


Fig. 3 Stable and unstable situation for AV = 4

Summary and Outlook

This paper presented an approach for the stability analysis of autonomously controlled production networks with transportations. Tools from mathematical stability theory were combined with the simulation of dynamic systems, which has the advantage of less time consumption in contrast to a pure simulation approach to identify stability regions. The approach has been applied to an exemplarily autonomously controlled production network in order to identify parameter constellation which guarantee the stability of the entire network.

Future research will focus on applying this combined approach to more complex scenarios with different autonomous control methods. The presented approach can be used for the identification of regions of effective or optimal behavior of the network in view of economic or logistic goals.

Acknowledgments This research was supported by the German Research Foundation (DFG) as part of the Collaborative Research Centre 637 "Autonomous Cooperating Logistics Processes—A Paradigm Shift and its Limitations".

References

Dashkovskiy S, Karimi HR, Kosmykov M, Mironchenko A, Naujok L (2010a) Application of the LISS Lyapunov-Krasovskii small-gain theorem to autonomously controlled production networks with time-delays. In: Proceedings of SysTol, Nice, France, pp 765–770, 06 Oct 2008

Dashkovskiy SN, Rüffer BS (2010b) Local ISS of large-scale interconnections and estimates for stability regions. Syst Control Lett 59(3–4):241–247

Dashkovskiy S, Naujok L (2010c) Lyapunov-Razumikhin and Lyapunov-Krasovskii theorems for interconnected ISS time-delay systems. In: Proceedings of the 19th MTNS 2010, Budapest, Hungary, pp 1179–11845–9 July

Dashkovskiy S, Görges M, Kosmykov M, Mironchenko A, Naujok L (2011) Modeling and stability analysis of autonomously controlled production networks. Logistics Res 3(2):145–157

Philipp T, de Beer C, Windt K, Scholz-Reiter B (2007) Evaluation of autonomous logistic processes—analysis of the influence of structural complexity. In: Hülsmann W (ed) Understanding autonomous cooperation and control, Springer, Berlin, pp 303–324

Scholz-Reiter B, Dashkovskiy S, Görges M, Naujok L (2011a) Stability analysis of autonomously controlled production networks. Int J Prod Res 49(16):4857–4877

Scholz-Reiter B, Görges M, Jagalski T (2011b) Modeling and analyzing integrated autonomously controlled production and transport processes. In: Proceedings of the 21st international conference on production research. ICPR 21, CD-Rom, p 6

Teel AR (1998) Connections between Razumikhin-type theorems and the ISS nonlinear small gain theorem. IEEE Trans Automat Control 43(7):960–964

Wiendahl HP, Lutz S (2002) Production in networks. Ann CIRP Manuf Technol 51(2):1–14

Windt K, Hülsmann M (2007) Changing paradigms in logistics—understanding the shift from conventional control to autonomous cooperation and control. In: Hülsmann W (eds) Understanding autonomous cooperation and control, Springer, Berlin, p 4–16

Windt K (2006) Selbststeuerung intelligenter Objekte in der Logistik. In: Vec M, Hütt M, Freund A (eds) Selbstorganisation—Ein Denksystem für Natur und Gesellschaft, Böhlau, Köln, pp 271–314

Towards the Learning Behaviour and Performance of Artificial Neural Networks in Production Control

Bernd Scholz-Reiter and Florian Harjes

Abstract Recently, artificial neural networks have proven their potential in manifold production related tasks. At this, the learning ability is both their greatest strength and greatest weakness. The performance in the application is highly related to the learning process. Only sufficient training data in combination with a suitable configuration of the learning parameters leads on to useful results. Nevertheless, due to the general structure of artificial neural networks, the learning behaviour not necessarily correlates with the later functional behaviour. This paper considers the learning process and the application performance of four common network architectures in two production related applications; control and prediction. The evaluation of the network performance takes place by means of a generic shop floor model.

Introduction

The handling of today's complex and dynamic production processes requires the continuous advancement of production planning and control systems. In this context, the adaption of methods from the field of artificial intelligence came into focus (Harjes et al. 2011). Artificial neural networks, for example, are applicable for prediction as well as for planning and control purposes (Mandic and Chambers 2001;

B. Scholz-Reiter · F. Harjes (✉)
BIBA—Bremer Institut für Produktion und Logistik GmbH at the University of Bremen, Hochschulring 20, 28359 Bremen, Germany
e-mail: haj@biba.uni-bremen.de

B. Scholz-Reiter
e-mail: bsr@biba.uni-bremen.de

H.-J. Kreowski et al. (eds.), *Dynamics in Logistics*, Lecture Notes in Logistics,
DOI: 10.1007/978-3-642-35966-8_39,

Scholz-Reiter et al. 2004). Unfortunately, the transfer from the academic sector into practice still faces some obstacles. In the case of neural networks, the learning process is often difficult, as a direct correlation between a satisfying training behaviour and good results during operation is not always guaranteed. At this, universal indications for the selection of the proper network architecture and exact topology are missing. The same is the case for the training process, where the training method, the related parameter settings and the composition of the training data have to be determined.

This paper contributes to the evaluation of common neural network architectures, especially in the context of production control. Therefore, it considers the learning and application behaviour of four common network architectures in a shop floor environment. At this, the paper is structured as follows.

Section Introduction gives a short introduction to the topic and objective of the paper. Section Artificial Neural Networks introduces neural networks in general and describes the evaluated architectures in particular. The following section The Generic Shop Floor Model deals with the generic shop floor model that underlies the performance evaluation. The training of the considered networks and the corresponding results are described in section Training. The discussion of the experimental results of the simulation runs takes place in section Application Performance before the paper ends with a summary in section Summary.

Artificial Neural Networks

Artificial neural networks imitate the structure and functionality of nervous systems in nature (Haykin 2008). Similar to the nerve cells of the natural model, artificial neural networks consist of interconnected neurons. At this, the neurons are structured in layers depending on their function (Dreyfus 2005). Between the single neurons and layers, weighted links represent nerve connections with different strength. The number of layers as well as the direction and functionality of the links depends on the type of neural network. In general, an artificial neural network comprises two layers, an input and an output layer. In between, an unlimited number of so called hidden layers are possible.

Neural networks offer a fast data processing and a comparatively small modelling effort (Dreyfus 2005). At this, the artificial neurons act as small processing units, which receive and forward data via the internal links. In general, the incoming values are summed up at a neuron and the result serves as the basis for the calculation of a so called activation. The comparison with a threshold decides, if the activation is further processed into an output or not. For both the activation and the output, every neuron has a specific corresponding activation and output function. The choice of these functions is, besides the structure, a major characteristic of a network type (Chaturvedi 2008). In many cases, the output function is the identity, so that the output equals the activation. It is further possible to choose a threshold of zero, so that an information processing takes place for every input (Mandic and Chambers 2001).

The main advantage of neural networks is their ability to learn. At this, they can learn from exact examples, given feedback concerning the correctness of an output or even completely autonomous. The first approach is called supervised learning and denotes a learning process using pairs of input and output data (Chaturvedi 2008). During the presentation of corresponding input and output values, the neural networks adjusts its weights in a way that every input pair generates the desired output. At this, the weight adaption follows a specific learning algorithm. This algorithm determines the mode, the error is calculated, the order, the adaption takes place and so forth. Common learning algorithms are, for example, the Delta- and Hebb-Rule or the backpropagation algorithm and its variations (Lawrence and Giles 2000).

The learning process finishes, when a determined number of learning cycles is complete or the obtained learning error falls below a given limit. To ensure the generalisation of the network, a check of the learn success by means of a validation dataset takes place after the initial training phase. The validation denotes a test with so far unknown input and output data. If the network achieves a similar training error for the validation data, it is in all probability able to handle all input correctly that features the same internal correlations as the training data. If it fails the validation, the training data is either inappropriate to model the problem, the network is supposed to solve or the data pool is not extensive enough. In some cases, it is also possible, that a neural network merely memorises a set of training data. This phenomenon is called overfitting (Lawrence and Giles 2000).

The second approach, training with feedback, is called reinforcement learning. At this, the neural network only receives a response, if the output for a given input is correct or not. In contrast to the supervised learning, reinforcement learning does not require knowledge concerning the exact output. Aside from this, the training process follows the same steps, including the validation (Haykin 2008).

Finally, unsupervised learning (also known as self-organised learning) denotes a training process without any feedback. Neither an exact output nor a hint concerning the correctness of an output is given. In this case, the neural network tries to detect possible correlations within the input pattern autonomously (Dreyfus 2005).

Feed-Forward Networks

Feed-forward networks, also called Multi Layer Perceptrons (MLPs), constitute a comparatively simple type of neural networks, consisting of fully connected layers without any loops. Common activation functions for this network type are sigmoid functions, such as the hyperbolic tangent or the logistic function (España-Boquera et al. 2007). The training mostly takes place in a supervised or reinforcement-based proceeding. As learning algorithms, especially backpropagation and its variations quick- and resilient propagation are common.

Due to their simple structure, they are applicable to manifold tasks, such as signal diagnosis or modelling (Pawlus et al. 2011). Further, they are comparatively easy to construct and promise a fast learning process. In the field of production control, feed-forward networks are, amongst other things, applied to control and prediction tasks (Hamann 2008). Therefore, they serve as a reference in this paper for the training and operation performance of neural networks in this application area.

Radial Basis Function Networks

Radial basis function (RBF) networks have a structure quite similar to feed-forward networks. But in contrast to the feed-forward approach, RBF networks have a fixed link weight of +1 between the input layer and the first hidden layer. In addition, the neurons of the hidden layer only have radially symmetric activation functions, such as the Gaussian function (Lee et al. 1999). Mathematically, these Gaussian activation functions act as a kind of supporting points for the approximation of multidimensional functions. Due to the use of radial functions for activation purposes, the learning procedure splits into two steps. The first is a nonlinear adaption of the hidden layer. This comprises the selection of suitable supporting points for the applied Gaussian function. The selection process can take place randomly or supervised. In this case, supervised means a shift of the starting points to areas within the input space, which feature a broad representation within the training data. It is also possible, to spread the starting points equally over the solution space of the function. The second step of the general learning procedure is a linear optimization of the output layer. RBF networks come into operation for face recognition (Er et al. 2002) or circuit fault diagnosis (Catelani and Fort 2002).

Cascade-Correlation Networks

Cascade-correlation (cascor) networks belong to the class of the so called constructive networks (Fahlman and Lebiere 1990). The architecture changes during the learning process. At the beginning, a cascade-correlation network only contains an input and an output layer, which are fully connected with each other. The learning process starts with a traditional learning algorithm, in the most cases quick- or resilient propagation. If this initial training does not lead to a network error below the desired value, the special cascade-correlation algorithm comes into operation. At this, the algorithm gradually extends the existing network with additional hidden layers. These layers consist of only one neuron each. A complete learning iteration is called a cascade and comprises the following steps (Vamplev and Ollington 2005).

The learning algorithm integrates one or more units, the candidate(s), into the network and connects them with the input and possible hidden neurons, but not with the outputs. Now, the candidate neurons train to maximize the correlation between their activation and the residual error of the training set. The unit with the maximum correlation is added to the network and connected to the outputs. At this, the new links to the output neurons have random weights, the existing connections to the input and possible hidden units are frozen (Fahlman and Lebiere 1990). A cascade ends with a training of the random weights, using a common training algorithm again. The cascade-correlation network runs through additional cascades, until the network reaches the desired error level. Cascade-correlation networks are, for example, suitable for pattern recognition (Rivest and Shultz 2002).

Elman Networks

Elman networks are partially recurrent networks predestined for prediction purposes (Elman 1990; Harjes et al. 2011). Basically, they consist of a feed-forward network with an additional special layer. This so called context layer saves the neural activation of previous states and therefore enables the network to take past events into account. The influence of past states depends on the connection weight between the hidden layer and the context layer. In general, the weight is fixed and lies between 1 (strong influence) and 0 (no influence) (Elman 1990).

The Generic Shop Floor Model

The evaluation of the four neural network types takes place in a shop floor environment by means of a generic shop floor model. The model contains 12 machines arranged in 5 specialised workshops. At this, every machine is technically different, which leads to different setup- and processing times for every work-piece.

A simulation period lasts 30 days and comprises a start-up and a phasing-out phase of 2.5 days each. During the simulation, 7 different work-pieces run through the model. They are equally distributed over 3,150 orders and combined in homogeneous lots with a size between 1 and 4 pieces. The work-piece types 1, 2, 6 and 7 run successively through every workshop of the model. In contrast, the types 3, 4 and 5 have a variable machining sequence, which leads to possible backflows between the workshops 1, 2 and 3. The redistribution of work-pieces between the workshops follows an inventory based approach, basing on simple priority rules (Scholz-Reiter and Hamann 2008).

At this, a work-piece is redistributed to the workstation with the lowest inventory level and the shortest setup- and/or processing time for the work-piece. All three values are considered in correlation to the desired inventory levels, which are 42 min for all machines of the first workshop, 63 min for the complete second and 84 min for the third workshop. As inventory controllers, artificial neural networks with a feed-forward architecture come into operation. Structurally, an expert control network exists for every work piece type and every possible redistribution between two workshops. The detailed explanation of the control approach is given in (Scholz-Reiter and Hamann 2008). During the experiments, specific neural controllers are replaced with the networks trained for the evaluation.

For the test runs of the Elman networks, a simplified version of the model comes into operation. In order to make the effects of the redistribution decisions on the inventory levels and capacities of subsequent machines or workshops more comprehensible, the smaller model version contains 8 machines on 4 workshops. However, the general functionality is similar to the greater model.

Training

For comparison purposes, the training process comprises two network variants per architecture. At this, every network variant goes back to the same sets of training and validation data, containing 1,000 input–output pairs each. Every input pair consists of setup and processing times as well as inventory values from the material flow model, while the output represents the associated desired redistribution decision, following the approach introduced above.

The two considered feed-forward networks differ in the number of neurons in the hidden layer. The first network has 10 (FF_1 in Table 1), the second features 50 hidden neurons (FF_2). As learning algorithm, resilient propagation finds a use. Experiments with the underlying backprogagation approach, quickpropagation and the backprogagation variant with additional momentum term show insufficient results. At this, the learning curves oscillate, the training requires a huge number of learning cycles (often 1,000 or more) or does not converge with regard to the learning error (sum of squared error, SSE).

Table 1 Learning and validation results

Network	Algorithm	Learn. error	Val. error	No. cycles
FF_1	Res. prop.	~1 %	55 %	60
FF_2	Res. prop.	~1 %	55 %	100
RBF_1	B.- backprop.	20 %	80 %	700
RBF_2	B.- backprop.	5 %	65 %	700
CC_1	Cascor	n.n.	n.n.	n.n.
CC_2	Cascor	n.n.	n.n.	n.n
El_1	Res. prop.	5 %	15 %	600
El_2	Res. prop.	8 %	14 %	300

Applying the resilient propagation algorithm, both network topologies obtain satisfying results. The learning process of the smaller network converges after 60 cycles with an error close to zero (~ 1 %). Surprisingly, the validation error is with a value of more than 50 % incomprehensibly high (Table 1). The larger network with 50 hidden neurons shows a quite similar learning and validation behaviour. The only difference is a slower conversation, the network requires 100 cycles to reach its minimal learning error.

The two radial basis function networks feature 5 (RBF_1 in Table 1) and 50 hidden neurons respectively (RBF_2). In contrast to the feed-forward variants, the batch backpropagation algorithm renders the best training results for the RBF networks. At this, the smaller network converges to a learning error around 20 % and a validation error of nearly 80 %. Both values arrive after 700 cycles (Table 1). The larger network runs through the similar number of cycles with a learning error of 5 % and a validation error of 65 % (Table 1).

As the cascade correlation networks constitute a constructive approach, the structure of the trained network changes during the learning process. Therefore, the number of hidden neurons depends on the maximum number of candidate units that are available for the algorithm. This paper considers one network with 1 candidate (CC_1 in Table 1) and one with 8 candidate units (CC_2). In contrast to the other network types, the cascor networks both do not converge. They oscillate with high learning and validation error values of nearly 100 % (Table 1). Normally, these results are insufficient for a possible transfer in the practical application. To show the partially elusive coherence between the learning behaviour and the successful practical application, both networks are nevertheless considered in the performance evaluation.

The partially recurrent Elman networks have one hidden layer with 10 neurons and a corresponding context layer with the same size. Their main difference is the number of input neurons. One network has four inputs (El_1 in Table 1), the other one comprises one more input (El_2). Both networks are trained for prediction tasks within a shop floor environment. The network with four inputs is responsible for the prediction of the capacity utilization of single workstations. The second network predicts the inventory level of workstations. At this, the number of input neurons depends on the number of values considered for the prediction. Both networks use the resilient propagation algorithm for the training. The result for the capacity network is an average training error of 5 % with a validation error of 15 %. The second network achieves a training error of 8 %, while the validation data results in an error of 14 % (Table 1). The length of the learning process is around 600 cycles for the capacity predictor, the inventory predictor reaches its minimal error values after 200 cycles. Altogether, the learning and validation results show a great variance, especially regarding the difference between the learning and validation data sets. Further, the number of passed cycles reaches from 60 to 700 Table 1.

Table 2 Evaluation results (control networks)

Network	DD30 (%)	DD15 (%)	ATT (time)	σTT (time)
FF_1	85.13	48.51	8:00:53	1:47:37
FF_2	86.39	57.07	7.54:13	1:39:09
RBF_1	80.96	43.77	7:18:32	1:43:37
RBF_2	81.33	48.19	22:09.24	5:08:54
CC_1	96.95	75.42	7.09:44	0:59:10
CC_2	96.20	76.63	6:55:20	0:59:39

Application Performance

Control Networks

The evaluation of the network‘s application performance takes place by means of the generic shop floor model introduced in section The Generic Shop Floor Model. At this, the comparison bases on the characteristic values of work-piece type 3, which is one of the types with a varying machining sequence and a corresponding high complexity of the related material flow. For evaluation purposes, the newly trained feed-forward, RBF and cascor networks act as relocation controller only for this work-piece type.

The adherence for delivery dates, the average throughput time (ATT in Table 2) and the standard deviation of the throughput time serve as performance indicators. The adherence to delivery dates is considered in two tolerance ranges, both with the exact delivery date as the centre ±15 % (DD15), respectively 30 % (DD30) of the average throughput time of work-piece 3. Table 2 depicts the overall results of the control networks. The adherence to due dates is given in percent, the average throughput time and its deviation in real time (hh:mm:ss).

Similar to the training and validation results, the performance for all three network types shows a great variance. While all networks are able to handle the redistribution control in general, the single values indicate a distinct advance for the cascor networks. The adherence to delivery dates is in both tolerant ranges between 11 and 19 % higher than the next best results (FF_2). The average throughput time shows a similar behaviour, the cascor approach achieves times up to 1 h less than the other networks. The standard deviation finally confirms the performance advance with a distance of up to 40 min. This is a surprising result, as the training and validation results are insufficient. A clear and comprehensible coherence between training and application performance is not recognizable.

Prediction Networks

The evaluation of the Elman networks takes place by means of a smaller version of the shop floor model. Both networks are responsible for the prediction of capacity utilization and inventory level of single workstations within the shop floor. The results of workstation 1 in workshop 3 (ws_{13}) serve as basis for the evaluation.

Within the prediction horizon of 4 h (half a work shift), the capacity predictor reaches a maximal prediction error of 6 %. The average deviation between the predicted and the actual value is around 3 %. The inventory predictor reaches even better results with a maximal of 3.2 % and a deviation of 1.5 %. All values result from a simulation period of 20 h containing five single predictions.

In contrast to the networks evaluated in the control application, the predictive networks are able to confirm their good training and validation results. Similar to their slight training error, the prediction error in the application field is clearly under the critical value of 10 %.

Summary

This paper evaluates the training and validation process as well as the application performance of four common types of artificial neural networks in a production environment. At this, the application comprises two production related tasks; control and prediction. Whereas the control networks strongly differ in their learning and application performance, the networks for prediction purposes are able to repeat their slight training error also in the practical application.

The control networks show a partly incomprehensible behaviour, the results from the practical environment reveal no clear correlation to the training results. Surprisingly, the cascor networks fail during the initial training but obtain the best performance during the experiments. This behaviour can be seen as typical for neural networks and emphasizes the difficulty of finding universal rules for the structure, training and application of neural networks in general and for production related tasks in particular.

Acknowledgments This research is funded by the German Research Foundation (DFG) as part of the project Automation of continuous learning and examination of the long-run behaviour of artificial neural networks for production control, index SCHO 540/16-1.

References

Catelani M, Fort A (2002) Soft fault detection and isolation in analog circuits: some results and a comparison between a fuzzy approach and radial basis function networks. IEEE Trans Instrum Meas 51(2):196–202

Chaturvedi D (2008) Artificial neural networks and supervised learning. In: Chaturvedi D (ed) Soft computing: techniques and its applications in electrical engineering. Springer, Berlin, pp 23–50

Dreyfus G (2005) Neural networks methodology and application. Springer, Berlin

Elman J (1990) Finding structure in time. Cognitive Sci 14(2):179–211

Er M, Wu S, Lu J, Toh H (2002) Face recognition with radial basis function (RBF) neural networks. IEEE Trans Neural Networks 13(3):697–710

España-Boquera S, Zamora-Martínez F, Castro-Bleda M, Gorbe-Moya, J (2007) Efficient BP algorithms for general feedforward neural networks. In: Mira J, Álvarez JR (eds) Lecture notes in computer science—bio-inspired modeling of cognitive tasks. Springer, Berlin, pp 327–336

Fahlman S, Lebiere C (1990) The cascade-correlation learning architecture. In: Touretzky DS (ed) Advances in neural information processing II, Morgan Kauffman Publishers Inc., San Francisco, pp 524–532

Hamann T (2008) Lernfähige intelligente Produktionsregelung. Gito Verlag, Berlin

Harjes F, Scholz-Reiter B, Kaviani Mehr A (2011) Elman networks for the prediction of inventory levels and capacity utilization. Int J Appl Math Inf 10 5(4) pp 283–290

Haykin S (2008) Neural networks and learning machines, 3rd edn. Prentice Hall, New Jersey

Lawrence S, Giles C (2000) Overfitting and neural networks: conjugate gradient and backpropagation. Como, Italy, pp 114–119

Lee C, Chung P, Tsai J, Chang C (1999) Robust radial basis function neural networks. IEEE Trans Syst Man Cybern B Cybern 29(6):674–685

Mandic D, Chambers J (2001) Recurrent neural networks for prediction: learning algorithms, architectures and stability (adaptive and learning systems for signal processing, communications and control series). Wiley, Hoboken

Pawlus W, Robbersmyr K, Reza Karimi H (2011) Performance evaluation of feed forward neural networks for modeling a vehicle to pole central collision. In: Mastorakis N et al (eds) Recent researches in geography, geology, energy, environment and biomedicine. WSEAS Press, Corfu Island, pp 467–472

Rivest F, Shultz T (2002) Application of knowledge-based cascade-correlation to vowel recognition. Hawaii, USA, pp 53–58

Scholz-Reiter B, Hamann T (2008) The behaviour of learning production control. CIRP Ann Manuf Technol 7(1):459–462

Scholz-Reiter B, Hamann T, Höhns H, Middelberg G (2004) Decentral closed loop control of production systems by means of artificial neural networks. Budapest, Hungary, pp 199–203

Vamplev P, Ollington R (2005) On-line reinforcement learning using cascade constructive neural networks. In: Khosla R, Howlett RJ, Jain LC (eds) Lecture notes in computer science—knowledge-based intelligent information and engineering systems. Springer, Berlin, pp 562–568

Part IX
Global Supply Chains and Industrial Application

A Framework of the Forces Influencing the Adaptation of the SCOR Model to the Situation of the Manufacturing Industry in Developing Countries

Fasika B. Georgise, Klaus-Dieter Thoben and Marcus Seifert

Abstract In the current dynamic competitive market, it is supply chains rather than companies that compete. Models such as the SCOR model enable the supply chain wide modeling, evaluation and improvement of processes. However, this approach is tailored to the needs of western, developed industries. Especially, the early processes' industry products (as a source raw of material and simple products) come from developing countries; those industries have totally different conditions and constraints. In consequence, the early processes' industry in the supply chain can often not be considered in a model. Many cross-cultural studies provide evidence that, because of cultural and operating environment differences, successful models, approaches and practices in one country need to be adapted for effective use in another country. Based on our literature analysis, the forces that influence the DCs' supply can be classified into four principal capability's areas. These are: (1) technical & IT, e g. outdated technology & lack integrated computerized systems; (2) supply chain relationship, e.g. reluctant to adopt partnership & alliance based relationship; (3) national infrastructure, e.g. inadequate physical infrastructure like road, rail and (4) organizational & managerial, e.g. obsolete functional based model and working culture difference. The paper presents a framework of the forces that influencing the adaptation of SCOR model to suit to developing country's operating situations.

F. B. Georgise (✉)
International Graduate School for Dynamics in Logistics (IGS),
University of Bremen, Bremen, Germany
e-mail: geo@biba.uni-bremen.de

F. B. Georgise · K.-D. Thoben · M. Seifert
Bremen Institute for Production and Logistics-BIBA GmbH,
Hochschulring 20 28359 Bremen, Germany
e-mail: tho@biba.uni-bremen.de

M. Seifert
e-mail: sf@biba.uni-bremen.de

H.-J. Kreowski et al. (eds.), *Dynamics in Logistics*, Lecture Notes in Logistics,
DOI: 10.1007/978-3-642-35966-8_40,

Introduction

Currently, many manufacturing industry in developing countries (MIDC) are undergoing both technological modernization and transition. In order to facilitate the changes, the MIDC has started program to improve their industry performance and competence using different modern management philosophies and techniques such as total quality management (TQM), supply chain management (SCM), just-in-time (JIT), flexible manufacturing system (FMS), performance measurement systems (PMSs), balanced scorecard (BSC), supply chain reference operations (SCOR) model, business process reengineering (BPR) and lean manufacturing. Like other management innovations, the SCOR model emanated from a developed country. It has been adopted by practitioners and managers in other developed and is beginning to be adopted in less-developed countries such as Africa (Abdelsalam and Fahmy 2009; Irfan et al. 2008). However, it is not sure that innovations such as the SCOR model can be applied in the different context and environment of less-developed countries. The attempt to implement might face a variety of problems that could not occur in developed countries. On the other hand, these modern tools are becoming increasingly expensive to build from scratch and implement in the developing country's scenario. Consequently, an excellent experience from the developed world that has shown success stories could be transferred for application that satisfies some of the requirements, and then extended, tailored for local developing country's requirements (Stephens 2001; Hieber 2002; Huan et al. 2004). The experience of developed and developing countries has shown the importance of adapting existing and already tested model to each particular organization and context (Becker et al. 2007; Han et al. 2002; Galazzo 2006; Waal 2007). This general lesson applies to less-developed countries (LDCs) like Africa as well. Adaptation can potentially be used to reduce new research development time and costs.

There are a lot of efforts are actually being attempted on adoption and implementation modern management techniques and philosophies in developing countries. They might potentially prove to be failures for different reasons, which may not be experienced in developed countries. Different practical researches and experiences that show the models and practiced that have shown success in the developed world may fail in developing countries (Sinkovic et al. 2011; Ohemeng 2010; Kureshi 2010; Han et al. 2002; Rutkowski 2010; Walsham and Sahay 1999).Their investigation has resulted pave a way to start new research on framework for further adaptation works. Walsham and Sahay (1999) has studied a number of Geographical information systems (GISs) that were being implemented in order to simplify the administration of one of the public sector activities in India. The GIS technology assumed that the users were in need of map-based solutions. Cultural obstacles like absence of maps had never been seen as a problem constructed barriers for the implementation of the new technology. Thus, in order to have successful adaptation, structural changes, as well as changes in the actual training, are necessary (Walsham and Sahay 1999). The other interesting research by Kureshi (2010) was investigated & discussed the concept of supply chain integration for its great potential to improve the competitiveness of both large and small businesses in Pakistan. The Motorola Five-Stage Model of Customer/Supplier Partnership Development by can be benchmarked by large buyers in Pakistan's manufacturing industry to develop a more "culturally competent" model that could work in Pakistani business environment. Finally, a modification of Motorola Five-Stage Model of Customer/Supplier Partnership Development was presented for benchmarking in manufacturing sectors of developing economies.

The above discussions are a few examples of the dangers of one sizes fit all strategies implemented in different context. From the recent research by Massachusetts Institute of Technology (MIT) has also shown the same results even in developed countries context. One size doesn't fit all! The Supply Chain 2020 research project provides an excellent example of this type of approach to best practices. This means that the term 'best' may only apply when the whole system of tailored practices is greater than the sum of the parts (Lapide 2006). This paper aims to offer the framework of the forces influencing the adaptation of SCOR model to the manufacturing industry in developing countries. By exploring the challenges and constraints of supply chain practices of the MIDC, this study helps researchers and practitioners to understand better how and why the framework helps to facilitate and affect the SCOR model adaptation. After this brief introduction, the paper follows short explanation of the SCOR model. Secondly, the barriers and challenges of the supply chain in the developing countries are presented. Then, the discussion of the framework of the forces model adaption is presented, including its parts. Finally, conclusion and future research direction are given.

SCOR Model

The Supply Chain Operations Reference (SCOR) model provided by the Supply Chain Council specifies inter-organizational business processes and their information flows. There are five main processes characterizing the SCOR model: Plan, Source, Make, Deliver and Return (Fig. 1). The main processes are defined on the top level (level one). Level one is basis for setting performance competition. On the second level these main processes are clustered into process categories depending on the underlying process model. The second level is configuration level. The third level is a process element level, which defines a company's ability compete successfully in its chosen market.

SCOR model is the first one available designated to measure supply chain performance and logistics impact (Bolstorff 2003) across the boundaries of individual organizations. It is at its growing stage of its life cycle and is enjoying the leverage of becoming the industry standard (Huang 2004). This process reference model was also integrated with the well-known concepts of business process reengineering, benchmarking and best practice. It is a model that links business

Fig. 1 Original SCOR V10.0 model (SCC 2010)

processes, key performance indicators (KPI), and best practice. It was developed to be configurable and aggregates a series of hierarchical process components that can be used as a common language for enterprises to describe the supply chains and communicate with each other (Huang et al. 2005; SCC 2010).

Supply Chain Challenges and Barriers in Developing Countries

Recently, there were a tendency of increasing researches to identify the supply chain challenges and barriers in developing countries. The authors' have reviewed different research papers on the potential supply chain challenges and barriers. Table 1 summarizes the literature reviewed for our research. The table explains the researches done on different countries scenarios and their respective challenges and barriers. We have tried to explain some research results and summarize the findings collectively on this section.

The following sample research papers have discussed on barriers and challenges of the DCs. Khalifa et al. (2008) has investigated the challenges and issues faced by Egyptian industrial zones. Based on supply chain deficiency stages, a framework has developed. The proposed framework has indicated supply chain challenges in developing countries in general, Egypt in particular. It demonstrates that most industrial supply chains at Egyptian industries are suffering from poor supply chain design. The other research by Easton and Zhang (2002), which focuses on issues related to supply chain planning and design. This study has emphasized the importance of planning activities and its role in SCM. These activities are considered the basis for other supply chain activities. Sinkovic et al. (2011) were researched the problems encountered when a performance management system such as the BSC is implemented in the culture of a developing country. They have advised managers need to be aware of the problems if they are to be successful in implementing modern management techniques in the DC. National and organizational cultures influence the degree of difficulty of implementing performance measurement systems such as the Balanced Scorecard.

Table 1 Summary of literature on challenges and barriers for SC in DC

Contribution	Authors	Country
SC Challenges and barriers in Africa	Naude and Badenhorst-Weiss (2011), Hamisi (2011), Darroch (2001), Ohemeng (2010), Ruteri and Xu (2009), Msimangira and Tesha (2009), Msimangira (2003), Khalifa et al. (2008), Waal (2007), Davidson (2007), Voordijk (1999)	Tanzania, Botswana, Egypt, South Africa, Eritrea, General Africa Countries
Critical success factors and challenges to develop SC in Asia and Arabic Countries	Flores et al. (2008), Swaminathan (2007), Babbar (2008), Tippayawong et al. (2010), Ong and The (2008), Ali et al. (2008), Falah et al. (2003), Borade and Bansod (2010), Razzaque (1997)	India, Thailand, Malaysia, Saudi Arabia
Challenges in building global SC	Doss et al. (2010), Kauffman et al. (2005), Holmes et al. (2006), Kauffman and Crimi (2005), Gargeya et al. (2001)	General to developing countries

Based on our literature review, generally, Developing countries' supply chain challenges and barriers fall into four principal capabilities areas.

- Technical and IT: Implementation information technology & company's ability to react to different market segment with their technological & operational capabilities,
- Supply chain relationship: Ability to adopt partnership or alliance based relationships, which promote communication and information sharing,
- National infrastructure: This segment consists of infrastructure & resource factors, cultural variations, arbitrage and leverage opportunities, government incentives and regulations,
- Organizational and managerial: The ability of the companies to handle challenges related to management style & model, organizational culture and skilled & professional man power. Table 2 summarizes the challenges and barriers on accordingly.

Framework of the Forces for the Adaptation

Adapting the SCOR model offers an organization a new way of modeling and improving their supply chain due to well documents model building blocks to change their business operations. The SCOR model intended to be used in supply chain's scenarios, so it would have an impact transcends organizational

Table 2 Summary of supply chain challenges and barriers in developing countries

Technical and IT	National infrastructure	SC Relationship	Organizational and managerial
Different production standards and manual production operation	ICT and physical Infrastructure like road, rail	Lack of commitment, willingness and ability to invest into chain	Lack of compatible goal and strategy
Lack of QM concepts and documentation	Complexities associated with global sourcing	Lack of willingness to work together	Obsolete functional based model
Time delay and logistics problems	High transportation and logistics costs	Lack of mutual trust and mutual dependence	Working cultural difference
Lack of needed competencies/ skilled	Low research and development works	Equitable sharing of risks & rewards	Centralization of operational decision
State of technology in use: Outdated technology	Lack of financial resource, process equipment and technologies	Industry is reluctant to adopt partnership and alliance based relationships	Poor management practices on documentation and data and information handling
Lack of integrated computerized system to links with supplier	Cultural barriers, political instability and corruption	Difficult to find local suppliers	Ignorance for level inventory at the supply chain

Source Meixell and Gargeya (2005), Khalifa et al. (2008), Swaminathan (2007), Flores et al. (2008), Msimangira and Tesha (2009), and Zhang (2002), Hamisi (2011), Darroch (2001), Barbbar (2008), Hernandez et al. (2007), Kureshi (2010), Ruteri and Xu (2009), Holmes et al. (2006), Razzaque(1997), Babbar (2008), Stewart anf Mohamed (2002), Han et al. (2002), Al Falah et al. (2003)

boundaries. Due to the challenges and barriers faced by developing countries, which have been discussed above and complexity of model adaptation, careful analysis of the capabilities has to be done.

A few researchers and professionals have done different works on technological and management techniques adoption and adaptation for developing countries. To facilitate this activity, they have developed and employed different Framework. Al Riyadh et al. (2009) have developed an integrated model which overcomes some limitations of background theories like Technology Adoption Model (TAM), Technology-Organization-Environment (TOE) Framework, Institutional Theory (Dimaggio and Powell 1983) and institutional Intervention Theory. The study has finally developed a conceptual framework of e-banking adoption by SMEs in Bangladesh by integrating all pertinent parameters under umbrellas; these are internal factors, external factors and support institutions. Internal factors include perceived credibility and organizational capabilities. External factors include ICT industries readiness, regulatory support, financial industries readiness, and institutional pressure. They have also identified five support institutions: government, donor agencies, financial institutions, IT supported institutions and resource centers.

The works of Kurnia and Peng (2008) have assessed on the electronic-commerce readiness in developing countries with special emphasis on Chinese's grocery. They have developed the e-commerce readiness framework. Consequently, e-commerce readiness has three dimensions: organizational readiness, industrial readiness and national readiness. With their research, they have addressed e-commerce readiness at all three levels forms the basic conditions of e-commerce adoption by any organization. Upping and Oliver (2011) have developed a contingency model to adapt and investigate accounting changes in Thai Public Universities. It consists of four components: (1). External pressures for change; (2). Internal pressure for change. (3). Barriers to change; (4). Facilitators of change, which capture the varieties identified in previous studies like Luder (1992). The research by Waggoner et al. (1999) has developed a framework identified four generic categories of forces that can be said to shape the evolution and change of organizational performance measurement systems. These categories are: (1) internal influences, e.g. power relationships and dominant coalition interests; (2) external influences, e.g. legislation and market volatility; (3) process issues, e.g. manner of implementation and management of political processes; and (4) transformational issues, e.g. degree of top-level support and risk of gain and loss for change.

The other contribution on the changes in performance measurement systems is developed a framework by Munir et al. 2009 for the banking sector. A framework for the changes in PMS is influenced by a number of macro-level factors (e.g. economic, technological, socio and political), which in turn place pressure on organizations in various forms (mimetic, coercive and normative). The framework has proposed an examination factor stimulating change in performance measurement systems in the organization and the strategic responses to change efforts.

The identified research opportunity is the adaptation of SCOR model to the MIDC. For our fruitful analysis, we have developed a Framework that basis the previous research experience and can address the challenges & barriers of the supply chain. It will begin literature reviews of the available material on the SCOR model, PMSs, manufacturing industries in developing country and other relevant topics. For this research, there will be three phases of analysis as shown on our Framework. The first involves the analysis of SCOR model with the barriers and challenges of the manufacturing industry (MI) environment. The first step provides the input information about the general characteristics of the MIDC supply chains in relation to the SCOR model. The Fig. 2 is shown below summarizes the Framework.

The second phase involves the analysis of the existing business process practices in all five SCOR model processes (Plan, Source, Make, Deliver and Return) and compare the results with current SCOR model. This analysis provides the input characteristics to use in the development of the adapted SCOR model business processes. The third phase uses the characteristics of the adapted business

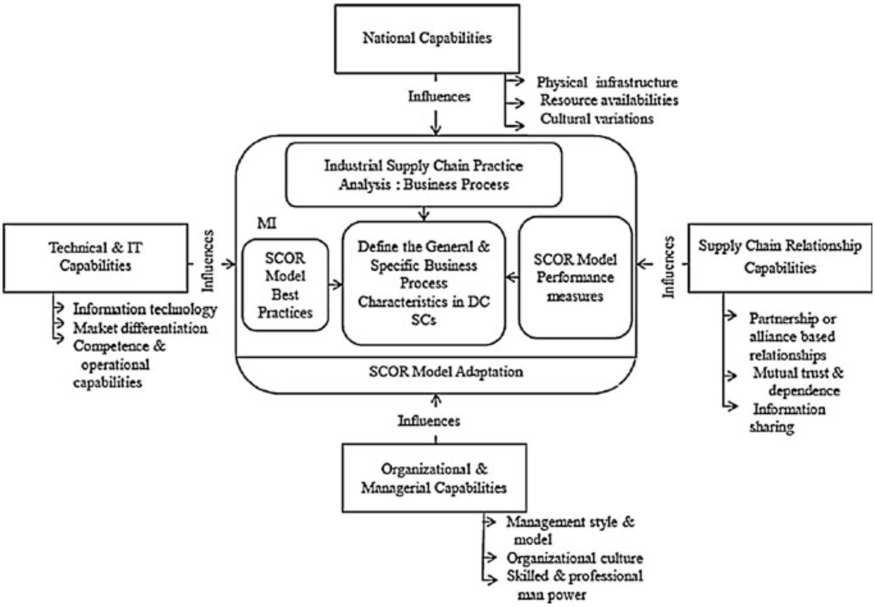


Fig. 2 Framework of forces for SCOR model adaptation

model to develop and implement a new Adapted SCOR model with their respective performance measures and best practices.

To determine the common characteristics, the research must analyze the MIDC as generally as possible. Generalization allows for assessment of multiple scenarios occurring cross-industry, insuring the characteristics apply to the general operations shared by the industry and not a specific industry function. Finally, the result of customization was the establishment of business process elements for the manufacturing industry in developing countries. At the same time, each business process element included suitable performance metrics to evaluate supply chain performance.

Conclusion and Future Direction

In this paper, many supply chain challenges and barriers that can impact the adaptation of SCOR model have been examined. Significantly, the synthesis of the literature has yielded a framework relevant to the adaptation of the SCOR Model. Since, the literature shows that successful evaluation and selection of models and approaches for supply chain modeling and improvement to fit local requirements remains problematic (Sinkovic et al. 2011; Waal 2007; Davidson 2007). The theoretical contribution of this paper is to explain; how is an adaptation of SCOR model in the MIDC can be performed by considering the internal and external operating situations. It is believed that the proposed framework useful in two ways. First, it is a step towards more prudent explanation how SCOR model adaptation will be performed in the MIDC. The framework discussion indicates that the four capabilities: technical and IT, supply chain relationship, national Infrastructure and Organizational and Managerial plays important roles in the adaptation of the model in the context of developing world. The SCOR model is inadequate because they do not incorporate or consider the existing four capabilities of the MIDC satisfactorily. Second, it highlights many important factors for SCOR model adaptation in DC context, which may enable a researcher to identify explanation and find a possible way to facilitate adaptation work more thoroughly. The findings of this study offer additional insights into the current the supply chain barriers and challenges in developing countries.

Future studies apply the framework to examine the current supply chain practices and devise ways for further adaptation in developing countries. Clearly, there is much to be researched about the supply chain practices in the DC. Therefore, there are a number of venues for future research. First, assessing the supply chain practices can lead to appropriate business process adaptation. The outcomes of this study would be used for document preparation and validate the current situations, barriers and challenges for the SCOR model direct application. Second, the development of suitable performance metrics and best practices to measure and improve the adapted business processes. This study represents an ongoing research that looks into supply chain practices, challenges and barriers in developing countries and finds a way out the possibility of adapting the SCOR model to suit the MIDC. The research will contribute towards an adaptation of the SCOR model that suit social, economic and technical factors of manufacturing industry in developing countries. This research work will therefore provide a contribution to knowledge on how adaptation of the SCOR model could be done to suit the manufacturing industry in developing countries.

References

Abdelsalam HM, Fahmy GA (2009) Major variables affecting the performance of the textile and clothing supply chain operations in Egypt. Int J Logistics Res Appl 12(3):147–163

Ali R Md, Jaafar HS, Mohamad S (2008) Logistics and supply chain in Malaysia: Issues and challenges. In: EASTS International symposium on sustainable transportation incorporating Malaysian Universities transport research forum conference (MUTRFC08). Universiti Teknologi Malaysia, Skudai, Johor Bahru

Babbar S (2008) Organizational factors affecting supply chains in developing countries. Int J Commer Manage 18(3):234–251

Becker J, Knackstedt R, Stein A (2007) Extending the supply chain operations reference model: potentials and their tool support. In: Proceedings of the European conference on information systems (ECIS). St. Gallen S. pp 1827–1838

Bolstorff P (2003) Measuring the impact of supply chain performance, a step by step guide of using the supply Chain Council's SCOR Model. Chief Logistic Officer, pp 6–11

Borade AB, Bansod SV (2010) Study of vendor-managed inventory practices in Indian industries. J Manuf Technol Manage 21(8):1013–1038
Darroch MAG (2001) Challenges facing agribusiness supply Chains in South Africa Beyond 2000. Agrekon: Agric Econ Res Policy Pract South Africa 40:505–521
Davidson NE (2007) Performance measures can transform export-led initiatives into competitive strength: a wake-up call for the developing countries of Africa. Newcastle Bus Sch Res Working Papers 2:48–74
DiMaggio PJ, Powell WW (1983) The iron cage revisited: Institutional isomorphism andcollective rationality in organizational fields, Am Sociol Rev 48(2):147–160
Doss AN, Bekki JM, Henderson M, Akhavan O, Tsang J (2010) Modeling supply chains for new ventures in developing countries. In: National Collegiate Inventors & Innovators Alliance (NCIIA), 14th annual conference, California, pp 25−27
Easton RJ, Zhang TB (2002) Supply chain in Asia: challenges and opportunities. http://www.accenture.com/kren/Documents/PDF/asia_supply.pdf
Falah KA, Zairi M, Ahmed AM (2003) "The role of supply- chain management in world class manufacturing" An empirical study in the Saudi context. Int J Phys Distrib Logistics Manage 33(5):396–407
Flores M, Cherian M, Canetta L (2008) Scenarios to collaborate with emerging markets: initial focus India. In: Camarinha-Matos LM (ed) IFIP International federation for information processing, pervasive collaborative networks, vol 283. Willy Picard, Springer, Boston, pp 599–610
Galazzo N (2006) Using SCOR model in a real collaborative sourcing scenario. VIB-ABCAL_PICS, Supply Chain Award 2006
Gargeya VB, Birdwell J, Martin R (2001) Challenges in managing the global supply chain. Forum Empresarial 6(2):22–45
Hamisi S (2011) Challenges and opportunities of Tanzanian SMEs in adapting supply chain management. Afr J Bus Manage 5(4):1266–1276
Han D, Kwon IG, Bae M, Sung H (2002) Supply chain integration in developing countries for foreign retailers in Korea: Walmart experience. Comput Industrial Eng 43:111−121
Hieber R (2002) Supply chain management a collaborative performance measurement approach. Zürich, vdf Hochschulverlag AG an der ETH Zürich: 154
Hernandez R, Reardon T, Berdegue J (2007) Supermarkets, wholesalers, and tomatogrowers in guatemala. Agr Econ 36:281–290
Holmes JS, Pineres SAG, Kiel LD (2006) Reforming government agencies internationally: is there a role for the balanced scorecard? Int J Publ Adm 29(12):
Huang SH, Sheoran SK, Keskar H (2005) Computer assisted supply chain configuration based on supply chain operations reference (SCOR) model. Comput Ind Eng 48(2):377–394
Huan SH, Sheoran SK, Wang G (2004) A review and analysis of supply chain operations (SCOR) model. Supply Chain Manage: Int J 9(1):23–29
Irfan D, Xiaofei X, Chun DS (2008) A SCOR reference model of the supply chain management system in an enterprise. Int Arab J Inf Technol 5(3):
Kauffman RG, Crimi TA (2005) A best-practice approach for developing global supply chains. In: Institute of Supply Management's (ISM's) 90th annual international supply management conference & educational exhibit, San Antonio, TX, pp 8−11
Khalifa N, White A, EISayed A (2008) Supply chain challenges in developing countries: cross industry case studies, In: IEEE SMC UK & RI 7th Conference on cybernetic intelligent systems, Middlesex University, London
Kureshi N (2010) Supply chain integration in SMEs—A model for developing economies. Proceeding of the internationalization of SMEs conference, Pau, France. http://pk.linkedin.com/pub/nadeem-kureshi/13/934/170
Kurnia S, Peng F (2008) Electronic commerce readiness in developing countries: the case of the Chinese grocery industry. In: Proceedings of the 41st annual Hawaii international conference on system sciences, Hilton Waikoloa Village Resort Waikoloa, Big Island, Hawaii
Lapide L (2006) MIT's SC2020 project: the essence of excellence, SC Manag't Review

Luder KG (1992) A contingency model of governmental accounting innovations in thepolitical-administrative environment. Res Gov Non-Profit Account 7:99–127
Meixell MJ, Gargeya VB (2005) Global supply chain design: a literature review andcritique. Transport Res E 41(6):531–550
Msimangira KAB, Tesha CP (2009) International supply chain practices in developing countries: a study in Tanzania. In: Conference Proceedings of the 20th Annual Production and Operations Management Society conference held in Orlando, Florida, USA
Msimangira KAB (2003) Purchasing and supply chain management practices in Botswana. Supply Chain Manage Int J 8(1):7–11
Munir R, Perera S, Baird K (2009) Changes in performance measurement systems in banking sector: an analytical framework. In: Performance measurement association conference, University of Otago, Dunedin, New Zealand
Naude MJ, Badenhorst-Weiss JA (2011) Supply chain management problems at South African automotive component manufacturers. South Afr Bus Rev 15(1):70−99
Ohemeng F (2010) The dangers of internationalization and "one-size-fits-all" in public sector management: Lessons from performance management policies in Ontario and Ghana. Int J Publ Sect Manage 23(5):456–478
Ong TS, The BH (2008) Factors influencing the design and use of performance measurement systems in the Malaysian electrical and electronics industry. Int J Eco Man't 2:437–457
Razzaque MA (1997) Challenges to logistics development: the case of a third world country-Bangladesh. Int J Phys Distrib Logistics Manage 27(1):18–38
Md Riyadh Al N, Akter S, Islam N (2009) The adoption of E-banking in developing countries: a theoretical model for SMEs. Int Rev Bus Res Papers 5(6):70−80
Ruteri JM, Xu Q (2009) Supply chain management and challenges facing the food industry sector in Tanzania. Int J Bus Manage 4(12):70−80
Rutkowski K (2010) Best practices in logistics and supply chain management. In: Donald W (ed) The case of central and Eastern Europe, global logistics: new directions in supply chain management, 6th edn. Kogan Page Limited, London
SCC (2010) Supply Chain Operations Reference Model (SCOR), version 10.0
Sinkovic G, Currie DM, Bevanda V (2011) From IT to BSC: the importance of culture in implementing a performance measurement system in Croatia. Indian J Eco Bus 10:39–55
Stephens S (2001) Supply chain reference model version 5.0: a new tool to improve supply chain efficiency and achieve best practice. Inf Syst Front 3(4):471–476
Stewart R, Mohamed S (2002) Barriers to implementation information technology indeveloping countries. In: Proceedings of the council for research and innovation in building andconstruction-working group 107 1st International conference creating a sustainableconstruction industry in developing countries, Spier, Stellenbosch, South Africa, pp 11–13
Swaminathan JM (2007) Managing supply chain operations in India: pitfalls and opportunities. Int Series Oper Res Manage Sci 98:137–154
Tippayawong KY, Patitad P, Sopadang A, Enkawa T (2010) Factors affecting efficient supply chain operational performance of high and low technology companies in Thailand. Manage Sci Eng 4(3):24–33
Upping P, Oliver J (2011) Accounting change model for the public sector: adapting Luder's model for developing countries. Int Rev Bus Res Papers 7(1):
Voordijk H (1999) Obstacles and preconditions for logistics and manufacturing improvements in Africa—a case study. Int J Oper Prod Manage 19(3):
de Waal AA (2007) Is performance management applicable in developing countries? The case of a Tanzanian college. Int J Emerging Markets 2(1):69–83
Waggoner D, Neely AD, Kennerley MP (1999) The forces that shape organizational performance measurement systems: an interdisciplinary review. Int J Prod Econ 60–61:53–60
Walsham G, Sahay S (1999) GIS for district-level administration in india: problems and opportunities. MIS Q 23(1):39–66

The Impact of Near Sourcing on Global Dynamic Supply Chains: A Case Study

Anna Corinna Cagliano, Alberto De Marco and Carlo Rafele

Abstract Although near sourcing is a valid alternative to global sourcing as a way to improve supply chain (SC) efficiency, there is a lack of studies aimed at understanding the reasons underlying the adoption of near sourcing strategies and at evaluating the related benefits. With the purpose of contributing to the discussion about the role of near sourcing in current dynamic SCs, the present work analyses the case of an Italian retailer that turned the global store furniture procurement process into near sourcing. Switching from Far East suppliers to a continental one enables a SC reengineering that decreases transportation and inventory carrying costs and assures economic viability. The work provides a methodological reference to compare global and near sourcing policies and to calculate the associated savings. The approach may be adapted to investigate different products, services, and business sectors.

Keywords Supply chain management · Purchasing · Global networks · Near sourcing

A. C. Cagliano (✉) · A. De Marco · C. Rafele
Department of Management and Production Engineering, Politecnico di Torino,
Corso Duca degli Abruzzi 24, 10129 Torino, Italy
e-mail: anna.cagliano@polito.it

A. De Marco
e-mail: alberto.demarco@polito.it

C. Rafele
e-mail: carlo.rafele@polito.it

H.-J. Kreowski et al. (eds.), *Dynamics in Logistics*, Lecture Notes in Logistics,
DOI: 10.1007/978-3-642-35966-8_41,

Introduction

Global sourcing, meaning proactively integrating and coordinating materials, processes, technologies, and suppliers across worldwide purchasing, engineering, and operating locations (Trent and Monczka 2003), is an effective way to gain competitive advantage in today's dynamic supply chains (SCs). There are multiple reasons for building an international supplier base: the lower operations costs in low-wage countries, the availability of more advanced manufacturing technologies, the increased availability of potential vendors for selected items, and the possibility of entering new markets and seizing fiscal opportunities (Quintens et al. 2005). However, growing labour costs, the volatility of currency exchange rates, and the awareness that global suppliers may cause inflexibility and non-responsiveness are recently pushing firms to seek alternatives to global sourcing. In such a context, near sourcing can be defined as manufacturing or procuring products and services from foreign suppliers located in geographical areas close to the buyers' facilities and customers and able to offer low prices (Christopher and Holweg 2011).

Due to a substantial lack of analysis about the advantages of near sourcing, this policy is still little applied. The first examples of near sourcing strategies can be found in industries where transportation costs have been largely affected by the increased oil price, such as furniture, apparel, footwear, and steel (Lynch 2008; Allon and Van Mieghem 2010). In addition, some multinational companies like Caterpillar and Ford have recently moved back their production facilities to the US and Mexico because of salary increase and currency strengthening in Far East countries and encouraged by incentives from local governments to invest in manufacturing activities (Cappellini 2011).

In order to contribute to the debate on how near sourcing may increase SC efficiency, this work discusses the case of an Italian retailer that changed a global sourcing strategy into near sourcing of the furniture of flagship retail stores.

The paper is organised as follows. Significant literature about the topic is presented in section Literature Background, whereas section The Case Study details the case study. Implications of the work and conclusions are discussed in section Conclusion.

Literature Background

Although global sourcing allows to focus on core competencies and to improve profitability, efficiency, and flexibility, it is also characterised by relevant drawbacks.

The geographical distances increase transportation costs and long order cycle times cause scarce supplier flexibility and responsiveness to demand changes. Due to the lack of buyer–supplier proximity, global sourcing is incompatible with just in time (JIT) philosophy as distances usually do not allow frequent deliveries of small orders and rapid problem solving (Humphreys et al. 1998). Suppliers may sometimes not meet the promised quality standards (Berman and Swani 2010). Furthermore, different languages, currency, business practices, and heterogeneous cultural and legal environments make buyer–supplier relationships risky (Rao 2004; Wilkinson et al. 2005). Finally, economic and financial events may put high cost pressure on suppliers, which is in turn transferred to their customers.

Thus, considerable costs are incurred with global sourcing strategies and price benefits may be offset. However, companies largely underestimate the hidden costs of global sourcing (Lowson 2001; Weidenbaum 2005; Lampel and Bhalla 2008). A recent study reports that additional sourcing costs are on average 50 % out of the purchasing product price, while they are often perceived to be just 25 % (Platts and Song 2010).

Near sourcing strategies can be adopted to optimise labour, material, and fully landed costs with risk, speed to market, and flexibility (Shister 2008). The reduced lead times allow companies to implement fast inventory turns and JIT policies. Also, the increased flexibility enables to better address last-minute changes in customer demand. Canada and Mexico are the most popular near sourcing destinations for US firms, while eastern European countries play the same role for companies located in Western Europe (Edgell et al. 2008; Lacity et al. 2008; Thelen et al. 2010). Central and eastern Europe countries benefit from lower labour costs than in western Europe, even if more expensive than in traditional far-east locations. However, geographical and cultural ties, partially common language skills, and the availability of trained professionals make central and Eastern Europe suppliers very attractive to western companies (Meyer 2006).

Despite a very rich literature about benefits and limitations of different sourcing locations, the reason why some companies review their global purchasing strategies and move towards near sourcing is still a poorly researched topic.

To this end, the present work discusses a business case in order to better understand the impacts of near sourcing on SC efficiency.

The Case Study

Company Presentation

The case company manages the product lifecycle from design to distribution. Products are sold in company-owned or company-leased retail stores that are outfitted with finishes and furniture compliant to a suitable design to facilitate sales and enhance the brand loyalty of the customers (Cagliano et al. 2011).

The SC of the retail store furniture aims to equip the brand stores with customized pieces of furniture, such as counters, shelves, drawers, dummies, and signs that are purchased from suppliers located in eastern China, then transported and stored at a centralised warehouse in Italy, and finally shipped for installation at the various European retail store locations.

Recently, several inefficiencies have arisen in the system of the European corporate-owned retail stores with exponential growth of the furniture centralised inventory and increasing transportation cost from the Chinese suppliers' facilities. Because of these factors, the case company asked the authors' research group to understand the root causes of the cost increase and to come up with new procurement policies.

The Traditional Purchasing Process

The purchasing process traditionally applied by the case company may be described as follows. First, based on basic design guidelines provided by the Image Office, a store template layout is issued by the Engineering and Design Department to allow the Procurement Department to release a standard order of furniture. Such order, that will be named "buy-to-stock" (BTS) order, contains a preset amount of pieces of furniture whatever the actual layout of a specific shop floor area will be: in fact, the actual layout will be disclosed on a later time. The standard order based on the template layout is necessary to face a three-month long lead time period for the furniture to be manufactured and transported from eastern China to Europe, which results to be a longer time than the two-month long period elapsing from the point in time when the actual material take-off is available to the date of store opening. Usually, detailed design drawings and material take-offs are 80 % compliant to the template layout quantities, so that the remaining 20 % of the furniture can be purchased at that time. Because only approximately two months are left, the Procurement Department purchases the rest of the necessary equipment from a vendor with manufacturing facilities located in eastern Europe, which assures one-and-a-half month long delivery lead time from the date the order is released. This second kind of order will be named "buy-to-order" (BTO) order. It is interesting to remark that in this specific furniture SC the decoupling point (Hoekstra and Romme 1992) is set when the detailed store layout is issued. As a matter of fact, before the release of final drawings and take-offs the material flow is driven by forecasts, while it is driven by the actual quantity of furniture that is needed after such a moment.

Case Study Methodology

The approach developed in collaboration with the case company's management can be summarised by the following four steps.

- *SC organisation and purchasing process mapping*. Data about orders placed and deliveries completed were collected during a period of time of 21 months.
- *Inventory management model*. Based on the analysis of the criticalities of the current process and the identification of all constraints and system variables, an inventory management model was created to simulate the effects of different purchasing and SC management strategies in order to optimise the operations costs. One and a half year holding period is considered. The quantities of material to be ordered from the Chinese suppliers are strictly related to the demand forecast, which equals the projected store openings schedule multiplied by the average unit quantity of furniture to be procured for the template store, which is approximately 35 cubic meters. The material is assumed to be shipped via high-cube forty-foot equivalent unit containers. Also, five cost components are considered, namely purchasing, order, transportation, inventory carrying, and additional orders and stock-out costs. Silver et al. (1998). The purchasing costs represent the average prices for pieces of furniture charged by both the Chinese and the eastern European suppliers. The order costs include the company's expenses for order issuing, insurance coverage, communication, and quality check. The transportation costs are made up of both costs for shipping from the suppliers to the centralised warehouse and expenses for moving the furniture from the warehouse to retail store locations. Transportation costs include custom border expenses. Inventory carrying costs comprise human resource expenses and overhead costs associated with occupancy, interest on working capital, and product obsolescence. Finally, additional orders and stock-out costs are related to the expenditure for orders placed to the eastern European supplier, that are required to fulfil the remainder material to integrate a store template layout with detailed furniture, and to the costs for replacing, again from the eastern European vendor, late Chinese deliveries. Detailed numerical values for the different costs involved in the inventory management model are available from the authors.
- *Risk analysis*. A risk analysis of the selected best SC arrangement was conducted to assess the main factors impacting on the future viability of the strategy to switch from global to near sourcing. A 2 year time horizon, consistent with the forward outlook time span of the company's business planning process, is assumed. Moreover, the risk analysis is performed on a country level: social, economic, and political determinants of uncertainty influencing sourcing decisions (Zsidisin 2003) are considered and reflected by sources of economic, inflation, monetary, and country risks. Such risks are investigated by comparing the values of selected indicators of the sourcing countries.
- *Result analysis*. Results were analysed and interpreted.

Table 1 Total costs under different inventory policies

Inventory management policy	Lot-for-lot (as-is)	Fixed period	EOQ	W–W
Average inventory level [m^3]	229	627	322	229
Max–min inventory level [m^3]	593	630	694	593
Total costs [€]	3,175,205	3,674,055	3,279,524	3,033,613

As-is Scenario

The inventory management model was here used to simulate monthly orders issued according to a lot-for-lot order policy (Boyer and Verma 2010) and a BTS approach with 80 % of a new store furniture sourced from China and the remaining 20 % from eastern Europe.

The first column of Table 1 shows the total costs of the as-is scenario. The numerical value is basically determined by two aspects. First, inaccurate orders, based on the standard store template layout, lead to the procurement of pieces of furniture that are then not used for equipping a store, thus increasing the inventory level in the centralised warehouse. Also, the high store-specificity of the material causing rapid obsolescence and the relevant quantity of safety stock necessary to avoid potential stock-outs, due to the long Chinese shipping time, contribute to raise the inventory holding costs.

Second, shipping costs from China have been growing during the last 2 years due to increased fuel price (United Nations Conference on Trade and Development 2010).

To-be Scenarios

Some alternative scenarios, respectively characterised by fixed period, economic order quantity (EOQ), and Wagner-Within (W–W) inventory management policies (Silver et al. 1998), were simulated and compared with the as-is scenario (Table 1).

The fixed period policy reveals to be the most expensive one with the highest average level of inventory. The inventory level significantly decreases with an EOQ policy, because of frequent orders, but this scenario brings an increase in total costs compared to the as-is situation as well as the greatest difference between the maximum and the minimum inventory level in the holding period. Out of the simulations, the W–W approach results to be the most convenient. In fact, it presents both the lowest total cost and the lowest average inventory and max–min inventory levels. However, the case company chose to keep the lot-for-lot policy because the W–W method is more time-consuming to apply and the current order policy is just slightly more expensive than the W–W one while having the same operational inventory performance.

Table 2 Total costs by changing supply chain structure

	As-is	BTO-Eastern Europe	BTO-Eastern Europe w/o Warehouse
Total costs [€]	3,175,205	2,549,848	2,287,627
Savings [€]	–	625,357	887,578

Therefore, what significantly contributes to decrease costs is not the adopted inventor management policy but more likely the way the SC is structured. In particular, it became interesting to study the case when furniture is completely ordered based on the detailed needs of a new planned store. This might happen only if either the time required for the Engineering and Design Department to issue the detailed material take-off is shortened or the supply lead time period is reduced to meet the Engineering and Design Department timeline. However, it was impossible for the case company to shorten the time required to issue the detailed layout, so the first option was discarded.

A BTO-Eastern Europe scenario, implying sourcing from a continental nearer supplier to reduce the lead time, was simulated. Table 2 compares the resulting SC total costs versus the as-is situation showing that BTO is an appropriate strategy. As a matter of fact, it allows total cost savings up to about 20 %, with a relevant contribution of the reduction in additional orders and stock-out expenses.

The relatively short distance between the supplier's manufacturing facilities and the European retail stores allows to ship directly to the stores with no need for an intermediate storage in the distribution warehouse. In addition, the E.U. location allows to avoid customs duties and delays, and the high quality of the products by the eastern European vendor requires little inspection. Therefore, the previous BTO-Eastern Europe scenario was simulated with direct shipments. The associated results are presented in the last column of Table 2: the relevant cost savings compared to the as-is scenario, due to decreased shipping costs and null inventory carrying costs, make the BTO-Eastern Europe without Warehouse scenario the most appropriate solution.

The case study highlights that near sourcing is attractive because it enables to reduce transportation and inventory holding costs, which counterbalance the less expensive price of goods charged by the global suppliers located in the Far East. To be more precise, the actual advantage of near sourcing is given by its ability to activate changes to the SC organisation, rather than by the lower cost per se. In fact, no significant savings are estimated to be gained with near sourcing compared to global sourcing in those scenarios that do not make structural changes to the SC.

Risk Analysis

With the aim of understanding the robustness of the best solution against potential future market changes and transitions, a 2 year risk analysis is performed. The present work focuses on the indicators of economic, inflation, monetary, and

Table 3 Risk sources and indicators

Risk source	Indicator
Economy	Gross domestic product growth rate
	Labour cost index
Inflation	Consumer price index
	Production price index
Monetary	Currency exchange rate
Country	OECD Ranking

country risks reported in Table 3. Their numerical values were collected through official sources, such as Eurostat and the Organisation for Economic Co-operation and Development (OECD) (Eurostat 2010; OECD 2010), for both China and Eastern Europe.

Based on the prospect indicators for the two geographical areas, the total yearly sourcing costs under the as-is and the BTO-Eastern Europe without Warehouse scenarios were quantitatively estimated (Table 4).

The total sourcing costs in the as-is scenario are increasing in the next 2 years due to the growth of the purchasing and shipping expenses as a consequence of rising Production price index (PPI) and Labour cost index (LCI). The total estimated costs in the BTO-Eastern Europe without Warehouse scenario are still significantly lower than in the as-is one and they are likely to remain steady because of the moderate increase in PPI and LCI. Also, the currency risk in Eastern Europe is more moderate than the one induced by the forecasted re-appreciation of the Chinese RMB against the euro in the near future. Thus, the strategy of switching to the eastern European supplier and arranging direct delivery to stores seems to be validated in the medium-term.

However, the assumptions and results of the risk analysis are strongly influenced by the economic, financial, and social development in an extremely dynamic international context. Therefore, the viability of the suggested policy should be periodically reviewed in order to check its consistency with the latest changes. Also, sourcing areas that may emerge as possible alternatives to the current suppliers should be taken into account.

The shorter SC lead time introduced by the near sourcing strategy provides the case company with cost savings and enables it to make major changes to the purchasing approach, so that a buy-to-stock SC can be transformed into a JIT system. In other terms, the value of the near supplier is inherent with the fact that

Table 4 Future yearly costs according to risk analysis

	As-is		BTO-Eastern Europe w/o Warehouse	
	Year 1	Year 2	Year 1	Year 2
Total costs [€]	2,929,271	3,383,815	1,495,218	1,584,140

its manufacturing facilities are located within regional boundaries, thus allowing for short transportation delays and augmented flexibility. Such advantages offer chances for SC reengineering which makes near suppliers be competitive against low price global vendors.

Conclusion

The current dynamic SC environment characterised by continuously changing economic, social, and political conditions has significantly reduced the potential of global sourcing, especially when the lead time represents a crucial operational factor. In this context, near sourcing allows to implement SC strategies that are not available with global sourcing, such as JIT and short time-to-market, thus contributing to optimise costs. These factors, together with lower transportation expenses, greatly offset the more expensive product prices paid to near suppliers.

In this case study the near sourcing strategy triggered a SC reengineering effort that proved to be effective. Therefore, the company's management chose to implement the BTO-Eastern Europe without Warehouse scenario for the purchasing of store furniture.

The methodology proposed in the present work is intended to be a guideline for SC professionals analysing alternatives to global sourcing. In particular, the discussed approach and case study may be interesting for those industries in search for increased SC flexibility since the time variable is the most important competitive factor. This is for example the case of electronic and high-tech sectors. Also, industries that have recently approached near sourcing as a way to control the costs associated with global strategies may find the proposed methodology beneficial.

As a future research line, the conceived evaluation framework might be extended for application to different manufacturing and service sectors.

References

Allon G, Van Mieghem JA (2010) Global dual sourcing: tailored base-surge allocation to near—and offshore production. Manag Sci 56(1):110–124

Berman B, Swani K (2010) Managing product safety of imported Chinese goods. Bus Horiz 53(1):39–48

Boyer K, Verma R (2010) Operations and supply chain management for the 21st century. Cengage Learning, Mason

Cagliano AC, De Marco A, Rafele C, Volpe S (2011) Using system dynamics in warehouse management: a fast-fashion case study. J of Manuf Technol Manag 22(2):171–188

Cappellini M (2011) Incentivi e flessibilità, produrre in Occidente ritorna vantaggioso. Il Sole 24 Ore 6 June: 23

Christopher M, Holweg M (2011) Supply chain 2.0 managing supply chains in the era of turbulence. Int J Phys Distrib Logist Manag 41(1):63–82

Edgell J, Meister GE, Stamp N (2008) Global sourcing trends in 2008. Strateg Outsourcing An Int J 1(2):173–180
Eurostat (2010) http://epp.eurostat.ec.europa.eu/portal/page/portal/eurostat/home. Accessed 30 Nov 2010
Hoekstra S, Romme J (1992) Integral logistic structures. Developing customer-oriented goods flow. Industrial Press, New York
Humphreys P, Mak KL, Yeung CM (1998) A just-in-time evaluation strategy for international procurement. Supply Chain Manag An Int J 3(4):175–186
Lacity MC, Willcocks LP, Rottman JW (2008) Global outsourcing of back office services: lessons, trends, and enduring challenges. Strateg Outsourcing An Int J 1(1):13–34
Lampel J, Bhalla A (2008) Embracing realism and recognizing choice in IT offshoring initiatives. Bus Horiz 51(5):429–440
Lowson RH (2001) Offshore sourcing: an optimal operational strategy. Bus Horiz 44(6):61–66
Lynch DJ (2008) Transport costs could alter world trade. USA Today 13 August. http://www.usatoday.com/money/industries/manufacturing/2008-08-11-cargo-costs-oil_N.htm. Accessed 21 Sep 2011
Meyer T (2006) Near shoring to central and Eastern Europe. Dtsch Bank Res, E-conomics 58. http://www.dbresearch.com/PROD/DBR_INTERNET_EN-PROD//PROD0000000000201757.pdf. Accessed 21 Sep 2011
OECD (2010) http://www.oecd.org/. Accessed 30 Nov 2010
Platts KW, Song N (2010) Overseas sourcing decisions-the total cost of sourcing from China. Supply Chain Manag An Int J 15(4):320–331
Quintens L, Matthyssens P, Faes W (2005) Purchasing internationalisation on both sides of the Atlantic. J Purch Supply Manag 11(2–3):57–71
Rao MT (2004) Key issues for global IT sourcing: country and individual factors. Inf Syst Manag 21(3):16–21
Shister N (2008) Executive overview: near-sourcing. world trade Mag 3 January. http://www.worldtrademag.com/Articles/Column/BNP_GUID_9-5-2006_A_10000000000000226880. Accessed 5 July 2011
Silver EA, Pyke DF, Peterson R (1998) Inventory management and production planning and scheduling, 3rd edn. Wiley, New York
Thelen ST, Honeycutt ED Jr, Murphy TP (2010) Service offshoring. Does perceived service quality affect country-of-service origin preference? Manag Serv Qual 20(3):196–212
Trent RJ, Monczka RM (2003) Understanding integrated global sourcing. Int J Phys Distrib Logist Manag 33(7):607–629
United Nations Conference on Trade and Development (2010) Review of Maritime Transport. http://www.unctad.org/en/docs/rmt2010_en.pdf. Accessed 10 June 2011
Weidenbaum M (2005) Outsourcing: pros and cons. Bus Horiz 48(4):311–315
Wilkinson B, Eberhardt M, McLaren J, Millington A (2005) Human resource barriers to partnership sourcing in China. The Int J of Hum Resour Manag 16(10):1886–1900
Zsidisin GA (2003) Managerial perceptions of supply risk. J Supply Chain Manag 39(1):14–25

Simulation-Based Analysis of Integrated Production and Transport Logistics

Enzo Morosini Frazzon, Joarez Pintarelli, Thomas Makuschewitz and Bernd Scholz-Reiter

Abstract The purpose of this paper is to present an approach for the integration of production and transport logistics in one original equipment manufacturer (OEM) and to compare its performance with a sequential procedure. Herein, delays in fulfilling delivery due dates in both non-disturbed and disturbed (with oscillating transport time) scenarios will be analysed through a simulation model. The major findings demonstrate that the proposed integrative approach outperforms the sequential one in the absorption of perturbations that can delay the production and transport of orders in production facilities along a supply chain.

Introduction

Production and transport scheduling are mostly carried out sequentially due to their complexity and current lack of appropriate heuristics for supporting a desirable integration on the operational level. The unbalanced and unstable

E. M. Frazzon (✉) · J. Pintarelli
Industrial and Systems Engineering, Federal University of Santa Catarina Campus UFSC, Florianópolis, SC 88040-900, Brazil
e-mail: enzo@deps.ufsc.br

J. Pintarelli
e-mail: joarezpj@gmail.com

T. Makuschewitz · B. Scholz-Reiter
BIBA—Bremer Institut für Produktion und Logistik GmbH, University of Bremen, Hochschulring 20, 28359 Bremen, Germany
e-mail: mak@biba.uni-bremen.de

B. Scholz-Reiter
e-mail: bsr@biba.uni-bremen.de

H.-J. Kreowski et al. (eds.), *Dynamics in Logistics*, Lecture Notes in Logistics,
DOI: 10.1007/978-3-642-35966-8_42, © Springer-Verlag Berlin Heidelberg 2013

integration of manufacturing and transport systems may weaken the competitiveness of supply chains. Especially within dynamic environments, production and transport systems must be properly integrated so that efficiency, responsiveness and flexibility could be achieved and sustained (Scholz-Reiter et al. 2010). Indeed, local decisions cannot only depend on the efficiency of the individual processes at different locations, but rather take into account the behaviour of linked decision systems.

In this paper, an approach for the integration of production and transport logistics in supply chains is analysed using a discrete-event simulation model. The approach is based on a generic framework where the supply chain is structured into a chain of operational planning entities (Scholz-Reiter et al. 2010). The test case comprises one original equipment manufacturer (OEM) as well as the delivery transport to customers. The paper is structured as follows: section Literature Review reviews relevant literature; and the computational simulation-based analysis is presented and tested in section Computational Experiment and Analysis. The paper ends with some conclusions and potential implications.

Literature Review

Sequential and hierarchical schemes for production scheduling and transport planning have been deployed with consistent performance for stable surroundings. Nevertheless, when dealing with dynamic environments, integrative schemes are necessary. Recent approaches for the integration of production and transport systems do not consider current capabilities, the resource utilisation level and transit-/lead-times (Scholz-Reiter et al. 2009). This limitation has special relevance in supply chains, where components of production and logistics must be properly integrated so that efficiency, responsiveness and flexibility can be achieved and sustained (Scholz-Reiter et al. 2010).

Resources and their employment level have to be better considered in production and transport systems so that decision making in the dynamic and competitive environment of supply chains is enhanced. The problem of coordinating supply chain stages can be handled by a monolithic (central) approach, where the schedules are determined simultaneously, or a hierarchical and sequential approach (Sawik 2009). The central approach is usually not practicable in real-world situations due to unfeasible requirements in terms of information availability and communication capabilities.

Even though sophisticated heuristic approaches (e.g. Wang and Cheng 2009a; Lin et al. 2008; Huang and Yang2008; Valente and Alves 2007; Raa and Aghezzaf 2008; Cheung et al. 2008; Hwang 2005) achieved exceptional results in handling isolated scheduling tasks—either production or transport—they are not able to materialise the competitiveness obtained by a combined view of production and transport systems. By utilising the combined flexibility of both systems, challenges triggered by a dynamic changing environment (e.g. perturbations) can be better handled. Therefore, an integrated alignment of production and transport scheduling at the operational level holds a great potential for strengthening the competitiveness of supply chains (Scholz-Reiter et al. 2010).

Several insights and concepts for the integration of production and transport have been developed in recent years (e.g. Cohen and Lee 1988; Chandra and Fisher 1994; Haham and Yano 1995; Thomas and Griffin 1996; Fumero and Vercellis 1999; Sahin et al. 2008). However, most of these concepts focus either on the strategic or tactical level (Chen 2010). Papers that also deal with detailed schedules for the transport can be classified according to the objectives of applied mathematical programs and heuristics. One group only considers the lead time for production and transport of orders (e.g. Potts 1980; Woeginger 1998; Lee and Chen 2001; Hall et al. 2001; Geismar et al. 2008). The second group takes lead times and associated costs into account (e.g. Chen 1996; Cheng et al. 1996; Wang and Cheng 2000; Hall and Potts 2003; De Matta and Miller 2004; Chen and Vairaktarakis 2005; Pundoor and Chen 2005; Chen and Pundoor 2006; Stecke and Zhao 2007). Although the determination of detailed schedules for the production and transport represents a good achievement, the routing of the utilised transport vehicles has to be properly considered. This challenge is only addressed by a few authors (e.g. Li et al. 2005; Geismar et al. 2008).

The problem of balancing the production and delivery scheduling so that there is no backlog and costs are minimised is addressed by Pundoor and Chen (2009). Li et al. (2008) studied a coordinated scheduling problem of parallel machine assembly and multi-destination transport in a make-to-order supply chain. Their approach decomposes the overall problem into a parallel machine scheduling sub-problem and a 3PL (third-party logistic provider) transport sub-problem. By means of computational and mathematical analysis, the 3PL transport problem is shown to be NP-complete. Therefore, heuristic algorithms are proposed to solve the parallel machine assembly scheduling problem.

The establishment of collaborative relationships among supply-chain partners is a requisite for iteratively aligning independent entities in supply chains. Nevertheless, approaches for structuring this collaboration still lack the ability to be implemented (Scholz-Reiter et al. 2010). Specifically in regard to production and transport systems, a comprehensive scheme for handling this integration on the operational level does not exist. Building scheduling approaches that integrate supply, production and distribution and could also deal with various machine processing environments embodies an important research challenge (Wang and Cheng 2009b).

Centralised solutions for the production scheduling and transport planning processes along supply chains are not applicable from a practical perspective due to overwhelming eyesight and communication requirements. On the operational level, these processes are currently carried out sequentially, due to their complexity and current lack of appropriate heuristics for supporting a desirable integration. Considering that the performance of a supply chain could be significantly improved—in terms of both service level and costs—by applying an integrated instead of sequential scheduling schemes on the operational level (Chen and Vairaktarakis 2005), a generic approach for the integration of production scheduling and transport planning in supply chains has been proposed by Scholz-Reiter et al. (2010). This generic approach embraces a chain of operational planning entities that perform the production and transport scheduling problem (PTSP), as well as a mechanism for supporting the alignment between these entities.

Supply chains are composed of a chain of production stages, starting at the suppliers of raw material, followed by several production facilities, and ending at the OEM. These production stages, as well as the final customers, are linked by transport systems. The proposed operational planning entities comprise the production scheduling and transport planning of one facility along the supply chain (Fig. 1).

Therefore, one entity carries out the scheduling for one production facility and associated transport to the subsequent entity of final customers. The scheduling of the orders is based on the order delivery dates, which are provided by upstream planning and align the activities carried out at subsequent entities. Furthermore, in order to ensure the on-time delivery of orders, the entities have the flexibility to contract external production processing or transport capacity. Each entity strives to achieve a certain service level in regard to the on-time delivery of orders and to minimise the costs for production and transport (Scholz-Reiter et al. 2009).

A scheduling scheme at the operational level needs to run in a successive way. This is motivated by the arrival of new orders, unexpected perturbations, as well as variations of current capabilities within the production and transport systems. In the time intervening between iterations, capabilities and employment level of involved production and transport system may change due to either planned events like maintenance of a machine, or a transport device, as well as perturbations like the breakdown of a machine or the flooding of a road. Therefore, the iteration time

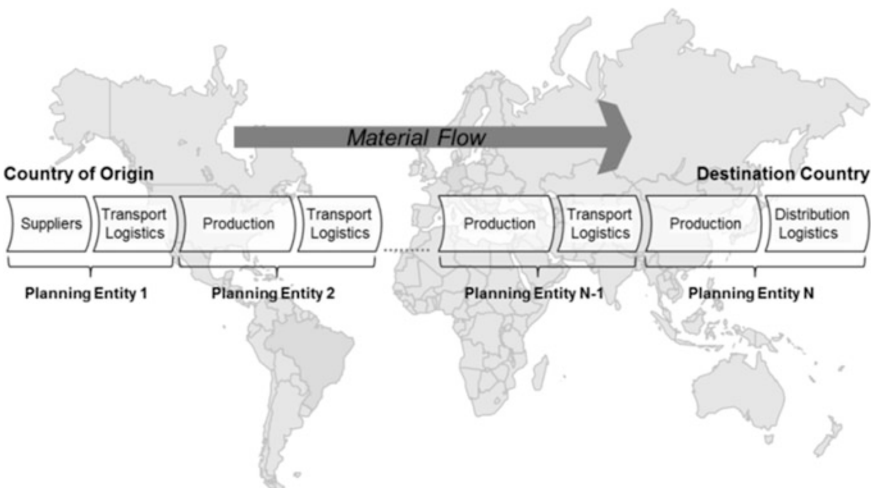


Fig. 1 Chain of planning entities on the operational level (Scholz-Reiter et al. 2009)

should be reduced in order to maximise the adaptability of the supply chain to dynamics. With the acceleration of these feedback loops, an on-line optimisation mechanism for supply chain priorities will emerge.

In general, the above literature is dedicated to be applicable for special settings. Therefore, no generic approach for the integration of production scheduling and transport planning along supply chains exists. This means that they are not suitable for a generic and fully integrated structure of a supply chain; do not consider disturbances or a rolling time horizon.

The design of integrated processes on the operational level of supply chains is a pressing challenge for both practitioners and scientists. On the sequence, a simulation model for analysing the integration of production and transport logistics within one original equipment manufacturer (OEM) taking part in a supply chain will be proposed, formalised and tested in a specific test scenario (section Computational Experiment and Analysis).

Computational Experiment and Analysis

Different departments within supply-chain partners usually perform the scheduling of production and transport by making locally-bounded decisions. As a drawback, the obtained results may be locally optimal but do not pay attention to the requirements of connected systems over the supply chain. In this section, a computational experiment using a simulation model for analysing the proposed integration approach (Scholz-Reiter et al. 2009) is presented.

Model Structure

The applied production scheduling is based on a heterogeneous open flow-shop with several consecutive production levels. Each production level consists of several machines, which feature an order-type specific processing time and processing cost. All orders have to be processed at one machine at each production level. Furthermore, orders can be processed externally in a comparatively short time and at a high cost.

After the production process, the orders are assigned to tours for the transport to the final customer. If at least five orders are assigned to a tour, this tour is conducted. In this case, fixed and variable costs occur. The variable costs depend on the duration of the tour. All considered tours start at the production facility and end at the final customer. A new tour can be conducted as soon as a transport device becomes available. Each tour has a limited transport capacity that cannot be exceeded. Disturbances which affect production or transport processes can be introduced by adjusting the reliability of model elements and the dynamicity of market demand.

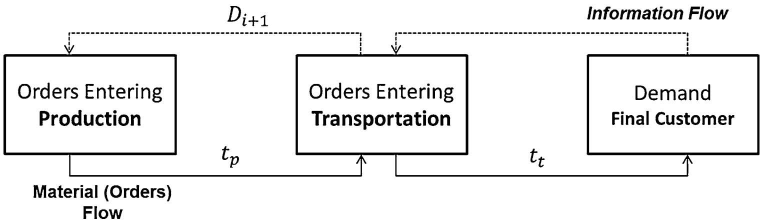


Fig. 2 Model structure for one entity—sequential scheduling set up

The models (Figs. 2 and 3) represent the production and transport execution level of one entity within the chain of planning entities. They are structured as a discrete-event simulation model with two sub-models: the production facility and the transport. When dealing with sequential scheduling (Fig. 2), the demand $\mathbf{D_{i+1}}$ is communicated to the stock of orders ready for delivery. Each order contains the orderID, the orderType, as well as its due-date **d**. The due-dates of orders are distributed equally along the week and are calculated considering the average of transport and production lengths of the last week as well as the demand of the current week. The time when the orders should be ready to be delivered is calculated considering the transport time $\mathbf{t_t}$ taking into account the average of transport lengths of orders shipped in the previous week. Sequentially, the stock of orders ready for delivery triggers the order release in the production system. The order release time is calculated considering the production time $\mathbf{t_p}$ by taking into account the average production time of the previous week. As soon as 5 orders are ready to deliver, the transport device executes the transport to the final customer.

In the case of integrated scheduling (Fig. 3), the demand $\mathbf{D_{i+1}}$ from the final customer triggers directly the order release in the production system. Each order contains the orderID, the orderType, as well as its due-date d, which is derived in the same way as specified for sequential scheduling. The order release time is calculated considering the production $\mathbf{t_p}$ and transport duration $\mathbf{t_t}$, based on the average of transport and production times of the orders produced and transported in the last week. As soon as 5 orders are ready to deliver, the transport device executes the transport to the client facility.

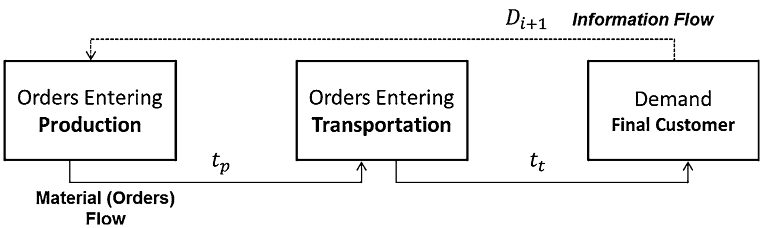


Fig. 3 Model structure for one entity—integrated scheduling set up

In both cases, early deliveries are allowed and orders are sent to the production facility at least 24 h before the first order of the week enters in the production facility. Both set ups described above will be employed in a test case in the next section.

Test Case

In this section, the simulation model for one planning entity within a supply chain is applied to a test case. The test case consists of one OEM located in Brazil. The transport to final customers is performed using land transport. A production process, which was described by Scholz-Reiter et al. (2005), is carried out at the production facility. The structure of the material flow within the production facility and the shipment to customers are shown in Fig. 4. The simulation period of the test case is one year.

The processing times of the three different order types for each machine are given by Scholz-Reiter et al. (2005). The processing costs are proportional to the required time of processing. External processing of orders triggers costs of 12 (production and transport). Externally processed orders return directly to the finished products buffer. Each land transport device has a maximal transport capacity of 10 units. As soon as 5 or more units are ready, the tour is conducted. The disturbed scenario includes transport time oscillations. The simulation model of the production and transport scheduling has been implemented in SIMIO version 3.48.6267 and the computation was carried out on a Core i7 2.8 GHz quad-core computer with 12 GB of RAM. The elapsed computational time for the simulation of the referred scenario was around 15 s.

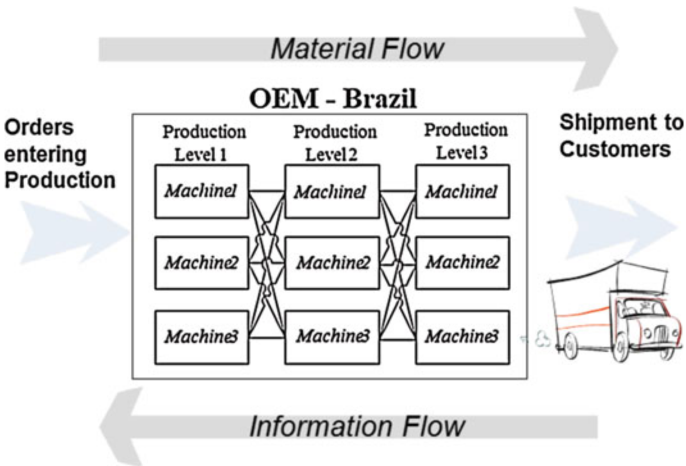


Fig. 4 Structure of the test case scenario

Results

The difference between due dates and actual delivery (moving average of the last 50 deliveries) for the sequential (dark grey line) and integrated (light grey line) set-up considering the disturbed scenario are displayed in Fig. 5. This scenario is characterised by an oscillation in transport time according to a sine function with a period of 3 months and amplitude of 5 h (the initial value of transport time is 7.5 h). The black line embodies the behaviour of the system in the stable situation, i.e. with no oscillations in transport time. The horizontal axis represents the operating hours (1–8760 h) along a year and the vertical axis represents the delay (difference between due-dates and actual delivery) for each order requested by the client, measured in hours. A positive value means that the entity arrived late, and a negative value indicates an early delivery, which is admissible.

The analysis of the difference between due-dates and actual delivery shows that the integrated scheduling reduces the occurrence of late deliveries, which is highly desirable. Nevertheless, it could be also identified a tendency for early deliveries. The results indicate that the proposed integrative approach outperforms the sequential one in disturbed situations. This means that it could absorb not only disturbances originated in transport processes but also oscillations in production processes. Our computational analysis indicates three major findings: (1) costs can be significantly reduced and lead-times can be shortened by properly combining the flexibilities of production and transport systems; (2) if there is a peak on the utilisation of the production system, the available time for processing of orders, which can be delivered in short time, can be extended; (3) in a situation where the transport process requires more time than anticipated, the scheduling of production can be rearranged so that orders are early available for transport.

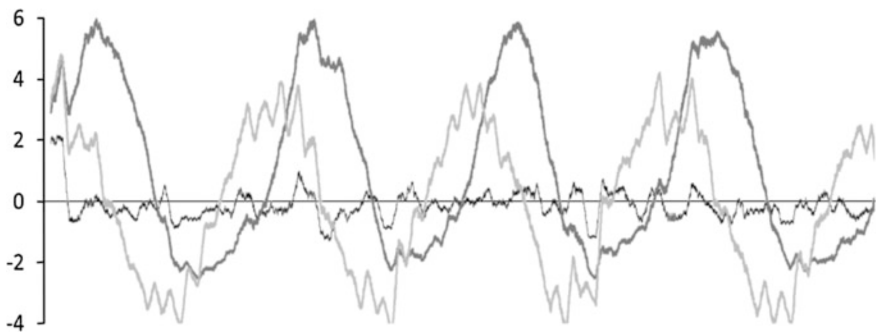


Fig. 5 Difference between due-dates and actual delivery along a year of operations

Discussion and Implications

In this paper, we analysed by means of simulation an approach for the integration of production and transport scheduling that fosters a sustainable management of production and transport systems along whole supply chains (Scholz-Reiter et al. 2010). A simulation model for the case of one entity along a supply chain was formulated. This formulation can be applied on a rolling time horizon and takes dynamic changing capabilities of the transport into account. It was possible to identify that the integrated scheduling can handle oscillations in the production and transport process by constantly checking the amount of time spent to process and deliver orders. The more frequent this checking occurs, the less time it will take to eliminate the discrepancies. However, too much checking's along a period of time could lead to an attempt by the system to correct the random oscillations, which are stochastic. The following topics of research could be pursued in the future: application of sensitivity analysis for identifying the relation between system's perturbations and its performance, development and implementation of more elaborated heuristic decision rules triggering the production and transport processing; development of quasi-real-time scheduling methodologies; and, pursuing empirical descriptive research in different real-world scenarios.

Acknowledgments This research was supported by CAPES, CNPq, FINEP and DFG as part of the Brazilian-German Collaborative Research Initiative on Manufacturing Technology (BRAGECRIM). This research was also supported by a grant from Simio LLC which provided the simulation software used in the computational analysis.

References

Chandra P, Fisher ML (1994) Coordination of production and distribution planning. Eur J Oper Res 72:503–517

Cheng TCE, Gordon VS, Kovalyov MY (1996) Single machine with batch deliveries. Eur J Oper Res 94:277–283

Chen Z-L (1996) Scheduling and common due date assignment with earliness-tardiness penalties and batch delivery costs. Eur J Oper Res 93:49–60

Chen ZL (2010) Integrated production and outbound distribution scheduling: review and extensions. Oper Res 58(1):130–148

Chen Z-L, Vairaktarakis GL (2005) Integrated scheduling of production and distribution operations. Manage Sci 51:614–628

Chen Z-L, Pundoor G (2006) Order assignment and scheduling in a supply chain. Oper Res 54:555–572

Cheung BK-S, Choy KL, Li C-L, Shi W, Tang J (2008) Dynamic routing model and solution methods for fleet management with mobile technologies. Int J Prod Econ 113(2):694–705

Cohen MA, Lee HL (1988) Strategic analysis of integrated production-distribution systems: models and methods. Oper Res 36:212–228

De Matta R, Miller T (2004) Production and inter-facility transport scheduling for a process industry. Eur J Oper Res 158:72–88

Fumero F, Vercellis C (1999) Synchronized development of production, inventory, and distribution schedules. Transp Sci 33:330–340

Geismar HN, Laporte G, Lei L, Sriskandarajah C (2008) The integrated production and transport scheduling problem for a product with short Lifespan. INFORMS J Comput 20(1):21–33

Haham J, Yano CA (1995) The economic lot and delivery scheduling problem: Powers of two policies. Transp Sci 29:222–241

Hall NG, Lesaoana M, Potts CN (2001) Scheduling with fixed delivery dates. Oper Res 49:134–144

Hall NG, Potts CN (2003) Supply chain scheduling: Batching and delivery. Oper Res 51:566–584

Huang R-H, Yang C-L (2008) Overlapping production scheduling planning with multiple objectives—an ant colony approach. Int J Prod Econ 115(1):163–170

Hwang H-S (2005) An integrated distribution routing model in multi-supply center system. Int J Prod Econ 98(2):136–142

Lee CY, Chen ZL (2001) Machine scheduling with transport considerations. J Sched 4:3–24

Li C-L, Vairaktarakis G, Lee C-Y (2005) Machine scheduling with deliveries to multiple customer locations. Eur J Oper Res 164:39–51

Li K, Sivakumar AI, Ganesan VK (2008) Analysis and algorithms for coordinated scheduling of parallel machine manufacturing and 3PL transport. Int J Prod Econ 115(2):482–491

Lin BMT, Lu CY, Shyu SJ, Tsai CY (2008) Development of new features of ant colony optimization for flowshop scheduling. Int J Prod Econ 112(2):742–755

Potts CN (1980) Analysis of a heuristic for one machine sequencing with release dates and delivery times. Oper Res 28:1436–1441

Pundoor G, Chen ZL (2005) Scheduling a production-distribution system to optimize the trade-off between delivery tardiness and distribution cost. Naval Res Logistics 52:571–589

Pundoor G, Chen ZL (2009) Joint cyclic production and delivery scheduling in a two-stage supply chain. Int J Prod Econ 119(1):55–74

Raa B, Aghezzaf E-H (2008) Designing distribution patterns for long-term inventory routing with constant demand rates. Int J Prod Econ 112(1):255–263

Sahin F, Robinson EP, Gao L–L (2008) Master production scheduling policy and rolling schedules in a two-stage make-to-order supply chain. Int J Prod Econ 115(2):528–541

Sawik T (2009) Coordinated supply chain scheduling. Int J Prod Econ 120(2):437–451

Scholz-Reiter B, Freitag M, de Beer C, Jagalski T (2005) Modelling dynamics of autonomous logistic processes: discrete-event versus continuous approaches. Ann CIRP 55(1):413–417

Scholz-Reiter B, Frazzon EM, Makuschewitz T (2009) Integration von Produktions- und Transportsystemen entlang globaler Supply Chains. Ind Manage 6:35–38 (in German)

Scholz-Reiter B, Schwindt C Makuschewitz T, Frazzon EM (2010) An approach for the integration of production scheduling and inter-facility transport within global supply chains. In: Proceedings of POMS annual conference 2010, Vancouver

Stecke KE, Zhao X (2007) Production and transport integration for a make-to-order manufacturing company with a commit-to-delivery business mode. Manuf Serv Oper Manage 9(2):206–224

Thomas DJ, Griffin PM (1996) Coordinated supply chain management. Eur J Oper Res 94:1–15

Valente JMS, Alves RAFS (2007) Heuristics for the early/tardy scheduling problem with release dates. Int J Prod Econ 106(1):261–274

Wang G, Cheng TCE (2000) Parallel machine scheduling with batch delivery costs. Int J Prod Econ 68:177–183

Wang X, Cheng TCE (2009a) Logistics scheduling to minimize inventory and transport costs. Int J Prod Econ 121(1):266–273

Wang X, Cheng TCE (2009b) Heuristics for parallel-machine scheduling with job class setups and delivery to multiple customers. Int J Prod Econ 119(1):199–206

Woeginger GJ (1998) A polynomial-time approximation scheme for single-machine sequencing with delivery times and sequence-independent batch set-up times. J Sched 1:79–87

Logistics Integrated Product Development in the German Automotive Industry: Current State, Trends and Challenges

Ingrid Göpfert and Matthias Schulz

Abstract Due to the increasing importance of logistics for the success of a company, automotive OEMs have started integrating logisticians into their product emergence processes to facilitate optimal systems design by incorporating logistics requirements. In this continuing process, many of the early issues have been improved by now; others will have to be solved in the future. This article contains a record of today's most relevant problems with possible solutions based on circa 25 interviews carried out with automotive representatives of German OEMs and suppliers. It also includes a reference model of the automotive product emergence process and examples of best-practices in Design for Logistics.

Keywords Design for logistics · Integrated product development · Automotive development process

Integrated Product Development in the Automotive Industry

Logistics becomes increasingly important for companies around the world and especially in the automotive industry, which is known for finding innovative logistics solutions to problems that later occur in other industries as well (Göpfert and Grünert 2009, pp. 129–130). The rising significance of logistics service

I. Göpfert · M. Schulz (✉)
Lehrstuhl für ABWL und Logistik, Am Plan 2, 35037 Marburg, Germany
e-mail: matthias.schulz@wiwi.uni-marburg.de

I. Göpfert
e-mail: goepfert@wiwi.uni-marburg.de

H.-J. Kreowski et al. (eds.), *Dynamics in Logistics*, Lecture Notes in Logistics,
DOI: 10.1007/978-3-642-35966-8_43, © Springer-Verlag Berlin Heidelberg 2013

components such as delivery time, reliability and flexibility becomes apparent in every stage of the value-adding process. For example, since the mid-1990s the delivery time of a Mercedes SLK was shortened from two years to 12–16 weeks (Alicke 2005, p. 49).[1] Ford's *Just-in-Time* supplier deliveries are carried out with a reliability of more than 95 % (Knorst 2004, p. 81). Delivery of Volkswagen spare parts to the workshops is carried out within 24 h throughout Germany.

Optimal design of logistics systems requires a fit-for-logistics product. The reason for this is that the product structure determines the structure of the production and therefore the potentials of systems optimization (Dombrowski et al. 2006, p. 723; Koether 2007, p. 14; Pawellek et al. 2007, p. 1). According to the interviewed managers, aspects such as high proliferation, wide dimensions and sensitive surfaces are especially hindering with regard to logistics operations. Since it is commonly assumed that at least 70 % of the costs and customer value are determined during product development (Andreasen 2005, p. 300; Ehrlenspiel et al. 2007, p. 13; Nevins and Whitney 1989, pp. 2, 3), the respective processes bear a high leverage potential to facilitate the design of sophisticated transformation systems.

In this paper, current challenges in Integrated Product Development are examined using the example of logistics integration in the product emergence process of the German Automotive Industry. The following analysis is based on a thorough literature review as well as circa 25 interviews that were carried out between April 2010 and September 2011. The respondents included experts from the German automotive manufacturers Audi, BMW, Daimler, Opel and Volkswagen as well as selected suppliers and an engineering service provider. Development engineers were covered as well as logisticians from a number of different divisions (inbound logistics, production logistics, after-sales logistics…) alongside representatives from contiguous departments (e.g. procurement) to get a comprehensive view on the value chain.

Various success stories connected with the implementation of "Design for Manufacture" (DfM) into different companies (Andreasen et al. 1988, pp. 73–148; Schraft and Bäßler 1986, pp. 127–130; Witte 1986, p. 100) let suggest that the integration of logistics bears similar potential. DfM increasingly became subject to academic publications in the early 1990s (Richter 2009, p. 432). Some of those works contained elements of a "Design for Logistics" (DfL), which is defined as the consideration of logistics aspects during product development through the use of existing degrees of freedom (Becker and Rosemann 1993, pp. 212–235; Pahl and Beitz 1988, pp. 177, 194; Schulte Herbrüggen 1991, pp. 247–259). DfL was then considered a small part of DfM (Sabisch and Tintelnot 1997, p. 69) and received little further attention in literature at that time (Dowlatshahi 1999, p. 62).

[1] Long delivery times used to be a sign that customers cherished the car so much that they were willing to wait for it (Reichhart and Holweg 2004, p. 43). Today, long waiting times are inacceptable for vehicle purchasers: After the accident in Fukushima in 2011, Toyota lost many customers because the delivery of their cars rose from 4–6 weeks to 3 months (Hawranek and Wagner 2011, p. 66).

However, with the extension of logistics' focus beyond company borders and thus the ascension of logistics towards a strategic supply chain management perspective, DfL gains new significance in both practice and economics theory (Göpfert and Schulz 2010, p. 41).

Hence, around the year 2000, automotive OEMs began integrating logisticians into their development processes to use said leverage in logistics systems design and thus a holistic optimization of the value chain. Investments in plant structure are now made in collaboration between factory delegates and logisticians from the development team and product design is influenced by logisticians. The desired success showed itself in lower process costs and improved service quality (e.g. increased reliability) in many examples: In the past, the fuel tank and the filler pipe were often welded together by the 1st-tier-suppliers and then delivered to the assembly line as a composite structure which had positive impacts on the assembly processes. However, due to its size and complex form, package density for this part was extremely low, which resulted in high transportation costs that over-compensated the benefits. The influence of logisticians made possible that most OEMs today procure the parts separately and weld them together in their factories. One manufacturer reported that high storage costs occurred in the after-sales division because a lot of just slightly different exhaust pipes of considerable size had to be made available. The logisticians thereupon designed and proposed a universal pipe that could be tailor-made to the specific car by screwing on end blinds of various length and shape (postponement, cf. Fig. 1). These examples also show that such decisions are often trade-off situations, so that different goals have to be considered and carefully weighed against each other.

Such achievements brought forward the integration of logistics in many ways. Today, logistics is accepted as an equal member of the simultaneous engineering team; crucial information that had to be procured with great effort in the early days of DfL is now provided automatically. However, since many processes (e.g. decision making) did not change since the 1990s, many problems still persist.

Fig. 1 End blinds for an exhaust pipe (*courtesy of* BMW product management parts)

The Product Emergence Process in the German Automotive Industry

The Product Emergence Process (PEP)[2] is initiated by a top-management decision based on the company strategy, important trends and technological progress. The assignment to complement the product portfolio is given alongside rough information on the strategic intent aligned with the new vehicle.

Despite differing company-specific labeling, the PEP in the German Automotive Industry can generally be delineated as consisting of four main stages. The first one is the **Target Definition**, in which the car is described from a customer's perspective. A reference product, often the predecessor, is announced by the project leader along with desired improvements or differences. Important activities for logistics include the definition of requirements, the choice of location, the determination of the degree of vertical integration and the need for investments as well as the definition of the main modules. During the next stage, often referred to as Concept Development or **Fuzzy Front End**, all the important specifications (main dimensions, body concept, design attributes, quality attributes, selection of innovations, target costs…) are defined from a technical perspective. Examples of logistics' assignments are the planning of investments, the selection of the most important suppliers (Forward Sourcing), the packing simulation, and the crude logistics system design. The actual engineering of parts is carried out during **New Product Development**. The interaction of the components is tested using prototypes; the concept is approved and transferred to series production maturity. In this stage, most suppliers of parts and tools are selected and the logistics system is fine-tuned. In the **Ramp-Up** phase several trial series are produced that increasingly resemble the later series production with regard to the tools that are used, the readiness of the procured supplier parts, the batch size etc. Occurring problems in the processes are fixed by readjusting process descriptions or part designs. Important tasks for logistics include change management, fine-tuning of provision processes and reliably managing the supply of up-to-date externally procured parts at minimum costs (Schneider 2008, pp. 51–52; Scholz-Reiter et al. 2005). Figure 2 displays an overview of the PEP including the main milestones.[3]

Challenges in Integrated Product Development

Logisticians are integrated from the beginning of the PEP, however, **other departments often dominate the negotiations** in case of conflicts of interest. Since many German companies produce premium cars, design/sales are considered

---

[2] The PEP includes all activities from the first idea to a series-ready product (Rausch 2006, p. 20). The results of the PEP consist of the product and process documentations (e.g. process sheets, CAD-files, bills of material…), (Stanke and Berndes 1997, p. 18).

[3] For a more detailed analysis cf. Göpfert and Schulz (2010, pp. 41–45); (2011, pp. 7–9); (2012).

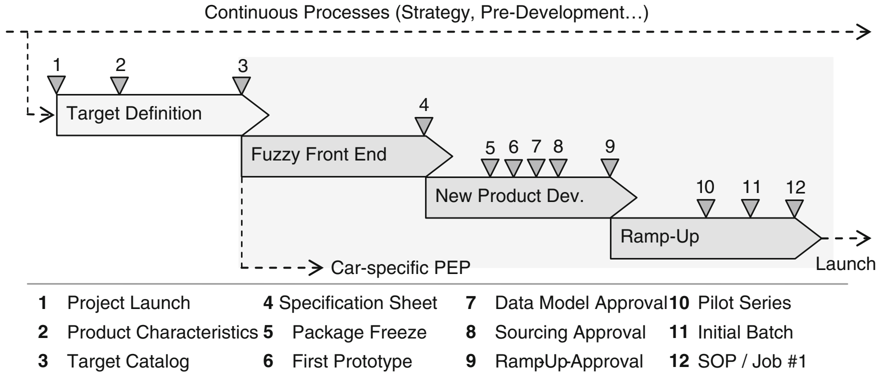


Fig. 2 A reference macro model for the product emergence process in the German automotive industry including essential milestones *Source* Adapted from (Göpfert and Schulz 2010, p. 43)

of crucial importance and thus usually allowed to veto any decision. On the cost side, purchasing and manufacturing account for most of the expenses while logistics' part is below 10 %. According to the interview partners, a two-digit share of these can be saved during the PEP. Cost reductions are usually the only way for logisticians to argue with other functional divisions. Although every OEM has implemented a sophisticated production system, the included principles (flow, work cycle…) are not by themselves accepted as an argument to influence design unless evaluated financially.

In negotiations, a **holistic optimization** is aimed at, which does not necessarily comply with logistics' goals. In case of platform components with high economies of scale, supplying a specific plant with body parts from another factory may mean a large number of transports over the life cycle of a product, but is often still cheaper than procuring a second set of tools or facilities. Many representatives emphasized that they do not have to succeed in every negotiation as long as decisions are made in awareness of the consequences for their department.

Usually, developers react to requirements from logistics openly and with very creative ideas as long as they are mentioned at the right time. This is made difficult by the fact that **only few logisticians are assigned to the car development**—often, one logistician has to represent his specific department (e.g. inbound logistics) in all teams of the OEM's vehicle project(s), which renders him unable to attend every team session (there are up to 50 such "simultaneous engineering teams" per vehicle project, Hackenberg 2007, p. 792). Some OEMs reported they often work reactively to prevent harmful decisions instead of actively influencing product design; however, all logisticians use to make their own suggestions on how their requirements could be implemented.

Despite an in-house production depth of less than 30 % (Dudenhöffer 2002, p. 6; Garcia Sanz 2007, p. 5; VDA 2008, p. 78), **suppliers' logisticians** having an impact on the product design is described as 'improbable'. There are manifold reasons for this: Firstly, many suppliers are appointed late in the New Product

Development stage. Secondly, considering the suppliers that are introduced early, logistics experts are not involved in the most crucial stages of the PEP. Lastly, transports are often within the responsibility of the OEM and thus not a key interest of the supplier.

Some companies **divide their procurement costs into the component's price and logistics costs** (e.g. transport, customs duty). While the sourcing decision is still based on the landed costs (i. e. the sum of both), this practice helps identifying economization opportunities. However, various risks arise depending on the costing system of the specific company: Sometimes, the costs of the product are paid by the procurement division while the logistics department covers the transportation costs. If two suppliers' offers are equal, the procurers might try to influence negotiations so that the supplier is chosen that weight their budget less. In other cases, there is a fixed target for the transportation costs that does not differentiate between distances. If an Asian supplier is chosen over a local one, the target costs are unreachably low, which means that the assigned logistician has to explain the deviations to his superiors. While the raise in costs is usually approved without problems, the responsible logistician may still feel like he is admitting a failure on his task. Thus, conditional costing systems are required. Also, risks connected with high distances (traffic holdups, high replenishment times) are not included in most OEMs' costing systems but—at the utmost—in the suppliers' rankings. These risks, including rising Diesel prices and unfavorable legislation, need to be taken into account.

Communication between logisticians and developers can bear some difficulties. One aspect is the project focus: While developing engineers think in terms of functionality, constructed space and development costs, logisticians however try to optimize transportation costs, packing materials and supply processes. In the past, developers were often unaware of the full amount of variants and its effects on the manufacturing processes (e.g. material stored at the assembly line). This situation is improving; however, some problems still need to be resolved. Another aspect concerns differences in the way of working that people with diverse job backgrounds usually display (Kuhn et al. 2002, p. 31). While logisticians and development engineers share certain aspects like system thinking, usage of information systems and a high sense of responsibility, differences persist with regard to interpersonal skills or stress tolerance (Kapoun 1991, pp. 13–14). This is intensified by the fact that developers usually studied engineering, while logisticians mostly have an economic background.

Figure 3 contains an overview of the most important issues pointed out by the interviewed logistics experts. The following paper contains possible solutions to some of the above mentioned problems and challenges.

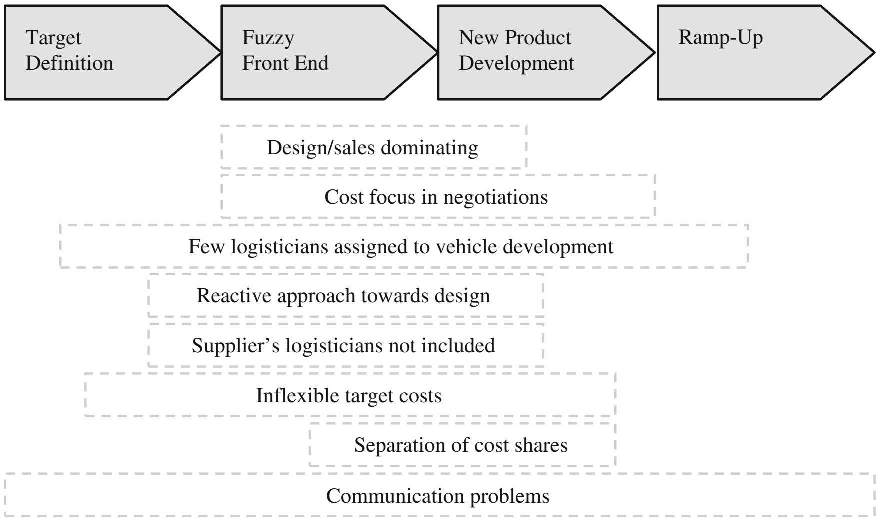


Fig. 3 Challenges in different stages of the product emergence process

Possible Solutions

If logistics' target system can be divided into market-oriented goals (short lead times, high delivery reliability, high flexibility…) and company-oriented goals (low costs, inventory, high degree of capacity utilization), the latter seem to dominate in the automotive industry (Pawellek et al. 2005, p. 73; Seeck 2010, p. 31). However, considering the rising degree of outsourcing, vehicles from different manufacturers increasingly resemble each other, as many components from one supplier are used by different OEMs simultaneously (Daecke 2009, p. 76; Mercer 2004, p. 27). Thus manufacturers will need to find new means to stand out in the perception of the customer (image, design, financial services…), one of which is logistics service (Baumgarten 2001, pp. 31–32; Ihme 2006, p. 264; Thoben 2002, p. 18).[4] If the car's specification sheet contained detailed requirements on the logistics service towards the customer, logisticians would be able to enter **negotiations** with arguments connected to customer value instead of cost savings, which would strengthen their position. For example, with regard to short lead times, when selecting suppliers from different locations, influencing variables would include outstanding logistics service and not only the transportation costs.

The problem of **insufficient manpower** in vehicle projects is currently approached by logistics experts by working together with other departments. Involved procurers or development engineers often possess a certain degree of logistics know-how and recognize decisions or situations that are crucial for the

[4] For a detailed analysis of the value-enhancing effects of logistics service cf Göpfert (2005, pp. 113–114).

design of the logistics systems. Thus, contiguous divisions usually inform the undermanned logistics departments about important upcoming decisions or current developments (one example would be spare parts logisticians working together with repair technicians). More logisticians can be assigned to the development teams if the benefits rise accordingly (e.g. higher savings or the service improvements mentioned above).

Another aforesaid problem was that single, inflexible targets that separate part costs and logistics costs often promote partisan politics. This is currently addressed at one OEM by increasingly setting the **target costs** for a part—and thus the incentives for procurement—on the landed costs only. Regarding the firm's organization, the logistics division is either located in the procurement department or the production department. Most issues with the separation of cost shares into part costs and logistics costs were pointed out by representatives from a company where logistics is a part of production. This suggests that structural changes in the organization might help solve the problem; however, since the sample contains only a few companies, definite correlations cannot be derived with certainty.

Not all OEMs reported to have **communication difficulties**, and those that did mentioned noticeable advances in the last years. Those occur both from a mutual respect and good will as well as from basic knowledge in the respective discipline. While logisticians have to develop a certain technical know-how (reading engineering drawings, understanding physical constraints), development engineers need to be aware of the processes they influence. As of now, they are evaluated based on the targets mentioned above with no regard to logistics costs or service. Incentives based on the target achievement of other divisions such as manufacturing and logistics might help increase awareness. Some OEMs try to sensitize them incrementally, for example by first addressing the problem of product proliferation. In the beginning, suitable aspects can be derived from complementary targets. Engineering goals cover aspects of quality (e.g. product functionality), time (e.g. development time) and costs (e.g. development costs), (Yang and El-Haik 2009, pp. 19–20). Logistics' goals include enhancing the value of the transferred objects, maintaining adaptivity and reducing logistics costs (Göpfert 2005, pp. 110–116). Figure 4 contains examples of shared sub-targets that promote traditional goals from both functional divisions.

These objectives focus on both service improvements as well as cost reductions. Supply chain excellence can improve the product development performance by efficient collaboration with suppliers which leads to a quicker delivery of prototypes, improved quality of processes (resulting in a higher quality of subassemblies), and better synchronization of product development and ramp-up processes. This allows companies to improve their profitability and cope with the rising number of product launches due to an increase in model variety (Seuffert 1994, p. 101) as well as it provides flexibility in a dynamic environment both before and after the start-of-production. On the other hand, the products themselves are improved in terms of costs and quality, e.g. by decreasing the number of errors (wrong deliveries, damaged goods) or improving capacity usage throughout the production network. In the beginning, logisticians and product developers can use

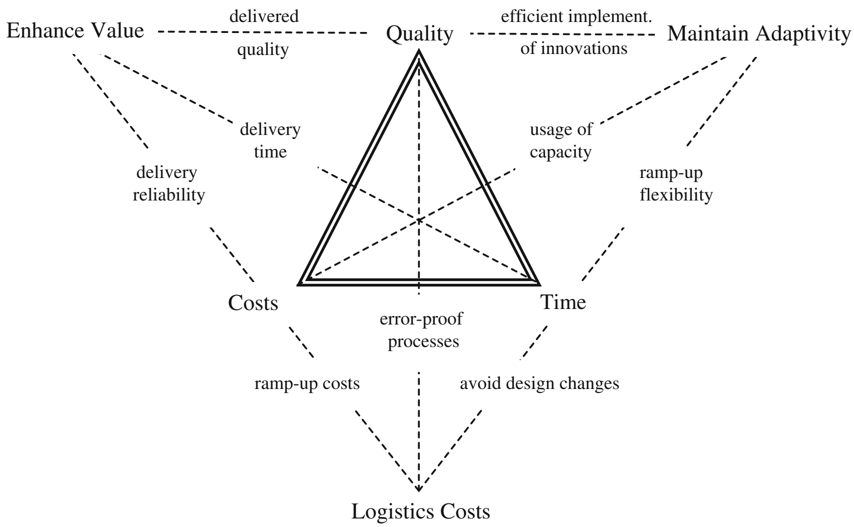


Fig. 4 Examples of complementary targets in product development and logistics

these **shared goals** to find a common language and improve collaboration. On this basis, product developers can understand which prerequisites are needed to improve logistics service, thus enhancing customer value in the above mentioned manner.

Summary

After design for manufacturability has been a common practice for several years, the integration of logistics into product development started around the year 2000 and has since improved steadily. Such integration is usually a continuing process (similar developments can currently be observed with the integration of environmental aspects) in which many problems have to be addressed in a certain succession: In the early days, gaining acceptance and getting access to the relevant information were the most important issues. Today's challenges focus on the negotiations in the simultaneous engineering team sessions, e. g. supplier selection, variant management and part geometry. Some of the most important problems have been pointed out in this article alongside possible solutions and best-practice approaches.

References

Alicke K (2005) Planung und Betrieb von Logistiknetzwerken—Unternehmensübergreifendes supply chain management. Springer, Berlin

Andreasen MM (2005) Concurrent engineering—Effiziente integration der Aufgaben im Entwicklungsprozess. In: Schäppi B, Andreasen MM, Kirchgeorg M, Radermacher FJ (eds) Handbuch Produktentwicklung. Carl Hanser, München, pp 293–315

Andreasen MM, Steen K, Lund T (1988) Design for assembly. Springer, Berlin

Baumgarten H (2001) Trends und Strategien in der Logistik—Die Entwicklung und die Zukunft der Logistik. In: Baumgarten H (ed) Logistik im E-Zeitalter—Die Welt der globalen Logistiknetzwerke. FAZ-Buch, Frankfurt (Main), pp 9–32

Becker J, Rosemann M (1993) Logistik und CIM. Springer, Berlin

Daecke J (2009) Nutzung virtueller Welten zur Kundenintegration in der Neuproduktentwicklung—Eine explorative Untersuchung am Beispiel der Automobilindustrie. Gabler, Wiesbaden

Dombrowski U, Schulze S, Vollrath H (2006) Logistikgerechte Produktentwicklung als Grundlage eines optimalen Logistikkonzepts. In: Z. für wirtschaftlichen Fabrikbetrieb 101 (12):723–727

Dowlatshahi S (1999) A modeling approach to logistics in concurrent engineering. Eur J Oper Res 115(1):59–76

Dudenhöffer F (2002) Neue Wachstums-Branche—Die Automobil-Zulieferindustrie. Automot Eng Partnersh 5(2):4–11

Ehrlenspiel K, Kiewert A, Lindemann U (2007) Kostengünstig entwickeln und konstruieren—Kostenmanagement bei der integrierten Produktentwicklung, 6th edn. Springer, Berlin

Garcia Sanz FJ (2007) Ganzheitliche Beschaffungsstrategie als Gestaltungsrahmen der globalen Netzwerkintegration in der Automobilindustrie. In: Garcia Sanz FJ, Semmler K, Walther J (eds) Die Automobilindustrie auf dem Weg zur globalen Netzwerkkompetenz—Effiziente und flexible Supply Chains erfolgreich gestalten. Springer, Berlin, pp 3–23

Göpfert I (2005) Logistik Führungskonzeption, 2nd edn. Franz Vahlen, München

Göpfert I, Grünert M (2009) Logistiknetze der Zukunft—Das neue Hersteller-Zulieferer-Verhältnis in der Automobilindustrie. In: Göpfert I (ed) Logistik der Zukunft—Logistics for the Future. Gabler, Wiesbaden, pp 127–166

Göpfert I, Schulz MD (2010) Future integrated product and logistics development using the example of the German car industry. In: Blecker T (ed) German-Italian conference on the interdependencies between new product development and supply chain management. TuTech, Hamburg, pp 39–50

Göpfert I, Schulz MD (2011) Supply chain design during the early stages of the automotive development process. In: Sinanesi A, Pero M, Blecker T (eds) Proceedings of the 2nd conference on the interdependencies between new product development and supply chain management (GIC-PRODESC 2011). Milan, Politecnico de Milano, pp 5–12

Göpfert I, Schulz MD (2012) Zukünftige Neuprodukt- und Logistikentwicklung am Beispiel der Automobilindustrie. In: Göpfert I (ed) Logistik der Zukunft—Logistics for the Future, 6th edn. Gabler, Wiesbaden (forthcoming)

Hackenberg U (2007) Simultaneous engineering und Projektmanagement im Produktentstehungsprozess. In: Braess HH, Seiffert U (eds) Vieweg Handbuch Kraftfahrzeugtechnik, 5th edn. Vieweg, Wiesbaden, pp 790–805

Hawranek D, Wagner W (2011) Die Tücken des Systems. Spieg 65(16):66–67

Ihme J (2006) Logistik im Automobilbau—Logitikkomponenten und Logistiksysteme im Fahrzeugbau. Carl Hanser, München

Kapoun J (1991) Konstruktion und Logistik. Plan + Produktion 39(6):12–15

Knorst P (2004) Ford—Lean Manufacturing und Supply. In: Corsten D, Gabriel C (eds) Supply chain management erfolgreich umsetzen—Grundlagen, Realisierung und Fallstudien. Springer, Berlin, pp 73–86

Koether R (2007) Technische Logistik, 3rd edn. Carl Hanser, München
Kuhn A, Wiendahl HP, Eversheim W, Schuh G, Laakmann F, Wiesinger GF et al (2002) Schneller Produktionsanlauf von Serienprodukten. Forschungszentrum Karlsruhe, Karlsruhe
Mercer Management Consulting (2004) Future automotive industry structure (FAST) 2015—die neue Arbeitsteilung in der Automobilindustrie. VDA, Frankfurt (Main)
Nevins JL, Whitney DE (1989) Concurrent design of products & processes—a strategy for the next generation in manufacturing. McGraw-Hill, New York
Pahl G, Beitz W (1988) Engineering design—a systematic approach. Springer, Berlin
Pawellek G, O'Shea M, Schramm A (2005) Logistikgerechte Produktentwicklung—logistics driven product development. Konstruktion—Z. für Produktentwicklung und Ingenier-Werkstoffe (Sonderdruck aus Heft 3):71–76
Pawellek G, O'Shea M, Schramm A (2007) Neue Werkzeuge zur logistikgerechten Produktentwicklung in der Automobilindustrie. Z. für wirtschaftlichen Fabrikbetrieb 102(6):1–6
Rausch H (2006) Management von Unschärfen in der Produktentstehung—Dargestellt an einem Beispiel aus der Automobilindustrie. Josef Eul, Köln
Reichhart A, Holweg M (2004) Build-to-order—impacts, trends and open issues. In: Parry G, Graves A (eds) Build to order—the road to the 5-day car. Springer, London, pp 35–53
Richter M (2009) Prozessmanagement in der Produktentwicklung. In: Bullinger HJ, Spath D, Warnecke HJ, Westkämper E (eds) Handbuch Unternehmensorganisation—Strategien, Planung, Umsetzung, 3rd edn. Springer, Berlin, pp 431–448
Sabisch H, Tintelnot C (1997) Integriertes Benchmarking für Produkte und Produktentwicklungsprozesse. Springer, Berlin
Schneider M (2008) Logistikplanung in der Automobilindustrie—Konzeption eines Instruments zur Unterstützung der taktischen Logistikplanung vor 'Start-of-Production' im Rahmen der Digitalen Fabrik. Gabler, Wiesbaden
Scholz-Reiter B, Höhns H, König F (2005) Intelligentes Änderungsmanagement für die Produktanlaufphase in Produktionsnetzwerken. In: Wildemann H (ed) Synchronisation von Produktentwicklung und Produktionsprozess—Produktreife, Produktneuanläufe, Produktionsanlauf. TCW, München, pp 111–136
Schraft RD, Bäßler R (1986) Möglichkeiten zur Umsetzung des montagegerechten Konstruierens. In: VDI (ed.) Montagegerechtes Konstruieren—Herausforderung für Entwicklung und Fertigung. VDI, Düsseldorf, pp 119–131
Schulte Herbrüggen H (1991) Modellanalyse von Materialflusssystemen für eine kundennahe Produktion. Josef Eul, Bergisch Gladbach
Seuffert D (1994) Innovative Produktionsmethoden in der Automobilindustrie. Nomos, Baden Baden
Seeck S (2010) Erfolgsfaktor Logistik—Klassische Fehler erkennen und vermeiden. Gabler, Wiesbaden
Stanke A, Berndes S (1997) Simultaneous engineering als Strategie zur Überwindung von Effizienzsenken. In: Bullinger HJ, Warschat J (eds) Forschungs- und Entwicklungsmanagement. B. G. Teubner, Stuttgart, pp 15–28
Thoben KD (2002) Extended Products—Die Erweiterung des traditionellen Produktbegriffs. In: Goch G, Heeg FJ, Hirsch BE, Scholz-Reiter B, Thoben KD (eds) Produktentwicklung—Von der Produktgestaltung bis zur Fertigungsplanung—Berichte aus Praxis und Forschung. Mainz, Aachen, pp 13–26
VDA (2008) Jahresbericht 2008. VDA, Frankfurt (Main)
Witte KW (1986) Montagegerechte Produktgestaltung als Aufgabe von Konstruktion und Fertigungsvorbereitung. In: VDI (ed.) Montagegerechtes Konstruieren—Herausforderung für Entwicklung und Fertigung. VDI, Düsseldorf, pp 97–118
Yang K, El-Haik BS (2009) Design for six sigma—a roadmap for product development. McGraw-Hill, New York

Efficient and Safe Transporting of Steel Sheets with Permanent Magnets

Johannes Schröck and Bernd Orlik

Abstract In ports many steel product are handled. A method to improve the number of handled steel sheets per time unit and magnify the safety of the handling system a transporting device based on magnetic forces can be used. For this purpose permanent magnets are adequate, since their created field is static and so on the created forces are static. To cast off the sheets an additional field created by a common coil is used. Such a system is simulated and designed to carry three sheets of ST235 with a weight of 8,500 kg and a size of 4.40 × 12 × 0.02 m. To proof this kind of a concept a scaled down model is created and successfully tested.

Introduction

Today steel products are usually handled with mechanical lifting equipment, which carry the load with hooks, chains, ropes and straps hold. These spreaders are relatively heavy and unwieldy, but assure by quality and reliability. Another method is to handle the steel products not with mechanical lifting devices, but with electro-magnetic carrying devices. In the ISUS project such a lifting device is developed. With this device several steel plates can be transported at one transport cycle. The challenge in this project is the reliability of the system and on the other hand, the simultaneous handling of steel sheets without additional materials. As a result, the number of handled sheets may be increased. Also an advantage can be that the storage timber can be saved.

J. Schröck · B. Orlik (✉)
IALB-Institute for Electrical Drives, Power Electronic and Devices,
University of Bremen, Otto-Hahn-Allee 1 | NW 1,
28359 Bremen, Germany
e-mail: b.orlik@ialb.uni-bremen.de

H.-J. Kreowski et al. (eds.), *Dynamics in Logistics*, Lecture Notes in Logistics,
DOI: 10.1007/978-3-642-35966-8_44, © Springer-Verlag Berlin Heidelberg 2013

Physical Basics

The basic physical effect of this system is a magnetic force. Magnetic forces are generated in different ways. The following figure shows schematically how magnetic forces can be generated (Fig. 1).

The simplest method to generate a magnetic force is the electromagnet. In an electromagnet a conductor is supplied with current. In this way a magnetic field is created. The outcome of this is that in the air gap between the metal material and the electromagnet a force is created. This force (Zastrow 2004; Marinescu 1996) is given by formula 1.

$$F = \frac{1}{2\mu_0} B_l^2 A_l. \tag{1}$$

The formula is a result of the conversation of the energy. First the magnetic energy stored in the air gap is calculated. This energy is transformed into mechanical energy, which creates the viewed force.

From the formula it is recognizable that the force is proportional to the square of the magnetic field in the air gap and linearly of the contact area between the magnet and the transported steel material. The formula is independent of the source, which creates the magnet field in the air gap.

Since in ports appear increased safeties regulations, such a common electromagnet can't be used to capsize sheets. In case of a blackout or technical failing such a system can harm persons and goods. A permanent magnet can also create the needed field. A disadvantage of using a permanent magnet is that the field is permanent. It can't be "switched off" like an electro magnet. This is also an advantage for this project. No energy is needed to join the steel material and the

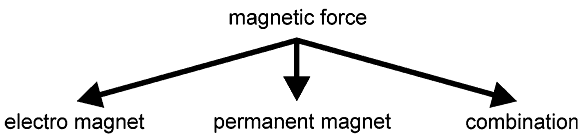


Fig. 1 Magnetic force

Fig. 2 Principle magnetic circuit

transporting device. To cast off the material a combination of these two effects are used. In the following figure the principle magnet circuit is shown (Fig. 2).

In this figure in black are the steel products shown. In red and green the permanent magnet. In grey the yoke with a coil. For this geometry with formula 2 the magnetic field in the air gap between the yoke and the steel material is given. The formula is a result of the "amperes law" and the Maxwell equations.

$$B_L = \frac{B_r + \frac{nI\mu_0}{l_m}}{\frac{A_l}{A_M} + \frac{l_l}{l_m} + \frac{l_{Fe}A_l}{l_m\mu_{rFE}A_{Fe}}} = \frac{B_r + \frac{nI\mu_0}{l_m}}{f(geometry)}. \quad (2)$$

In this formula three important characteristic can be seen. In the denominator only geometric quantities are found. In the numerator the remanence of the permanent magnet and ampere turns of the coil are found. Because the current has a direction, the flux linkage can compensate the magnet and the numerator can be regulated to zero. Then it is possible to cast off the steel material. The flux linkage can also be used to create a greater electromagnetic field, so that the resulting force can be increased. This system is also called an A-stable permanent magnet. The advantage of this system is, that the magnetic force for transportation of the material is produces by permanent magnets. Only when the transported material should be removed energy is needed. The A-stable system needs only a static counter field to compensate the field of the magnet. That means a permanent current must flow through the coil to cast off the material.

State of the Art

On the market there are hoisting tools, which use a magnetic force to transport steel products. A system is successfully used in transporting steels slabs with a weight of 36 t. This system generates the necessary holding force also with a Bi-stable permanent system. This system is not comparable with the system presented in this paper, since only one huge object is lifted. The problem here is to carry three separate metal sheets. The outcome of this is that between the sheets due the surface roughness an air gap is created, which weakens the magnetic flux. Therefore, this system can't be used for this purpose.

Furthermore there are solutions which transport sheets with electromagnets. These are used in forklifts and also have backup batteries to avoid failure, but the system become bigger and heavier. These systems can also transport multiple sheets with one transport cycle. However, the sheets must be pact in an additional step. This system can't be used in ports, because in ports the vertical height is too high. When the system fails and the sheets are starting to glide, it is impossible to estimate the drop splash area. When a spreader system fails, the drop off area is clear, so on emergency it can be calculated where the sheets drop.

The considered system also has an additional advantage; it saves the storage timber. For disembarkation the plates the must be a space created between the packet to ensure that the hooks can hold the sheets. So the new system in this paper has a high potential, and it can be the succeeding model for the existing models.

Simulation

In ports appear increased safety regulations, so a common electromagnet can't be used to capsize sheets into ships. In case of a blackout or technical failing such a system can harm persons and goods. Therefore a solution is with permanent magnets are needed. So these kinds of systems are considered and simulated with Flux. Flux is a numeric finite-element-program, which is specialized to solve electromagnetic problems. It can solve 2D and 3D problems. The focus in the simulation is on the created magnetic force, so all not relevant structure, like the connecting structure from magnetic lifting device to the Traverse, are not considered.

In the simulation some boundary conditions appear. Because permanent magnets are used, the safety factor is assumed with three. One sheet has a weight approx 8,500 kg and three sheets should be transported at once. The material of the sheets is ST235. Also the complete lifting device structure, which carries the magnetic circuit, is build with ST235. So the needed force on sheet number three must be 250 kN or more. The first two sheets are not observed directly. When the third sheet is transported correctly, the two sheets above it will be also transported correctly.

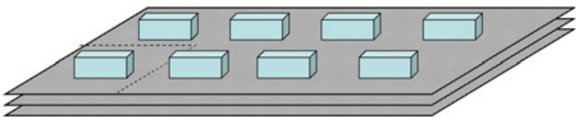
For redundancy reason there will be eight identical separate lifting units. The structure is displayed in Fig. 3.

Since the structure is symmetric and the eight units are identical, only an eighth of the complete structure will be simulated. In this way calculation time is reduced. Therefore the eight lifting devices are identical; they deliver the same force, so the minimum force per lifting device is approx 32 kN. If this force is reached, the safety factor of the system is three. In addition to this the depth of one device is defined to 1 m, to get independent from the depth.

Additional following simplification is done. The air gaps between the magnetic lifting device and sheet number 1 and the air gap between sheet number 1 and 2 are not explicit modeled. They are approximated with a virtual air gap by adding an additional magnetic resistance to the system of equations. This simulation detail is shown in Fig. 4:

Only the interesting air gap between sheet number 2 and number 3 is modeled exactly. It has a high of 0.5 mm. So the air gap is very fine; so many elements are needed to model the gap correctly. This air gap is exactly needed to calculate the

Fig. 3 Simulation area



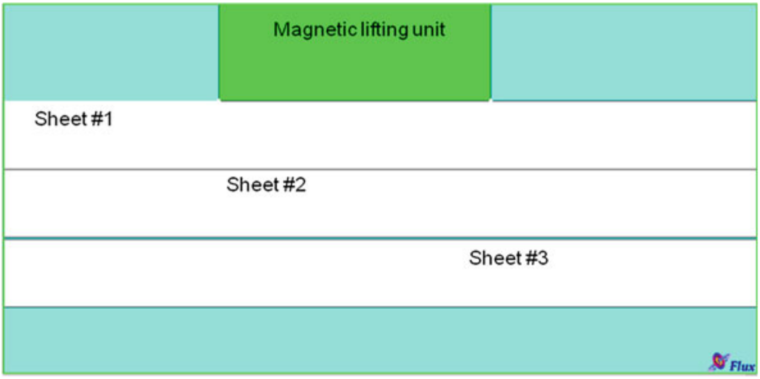


Fig. 4 Simulation details

force at sheet number 3 correctly. The force is calculated by the energy stored in the air gap, referring to formula 1.

In Fig. 5 the basic design is shown. All other designs are developed from this design.

This figure shows the direct output from Flux. The lines represent the flux lines of the magnetic field. Colored the magnetic flux density is displayed. The stream lines are used to ensure that the simulation worked probably. Out of the flux density the force on the third sheet is calculated. The third sheets are only flooded enough, when the upper plates is saturated. So the circuit has to be "over" designed, so that enough flux goes through the desired sheet.

After variation the variants two of the can afford the needed force of 32 kN on the third sheet. The two designs are shown in Fig. 6.

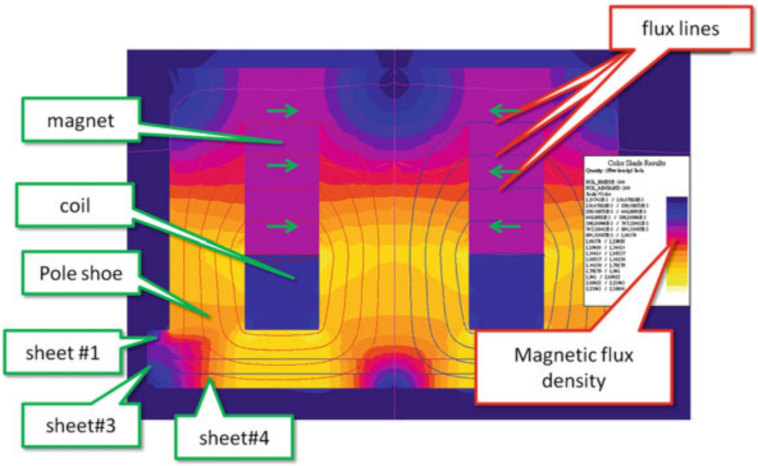


Fig. 5 Principle results of flux

Fig. 6 Schematic results

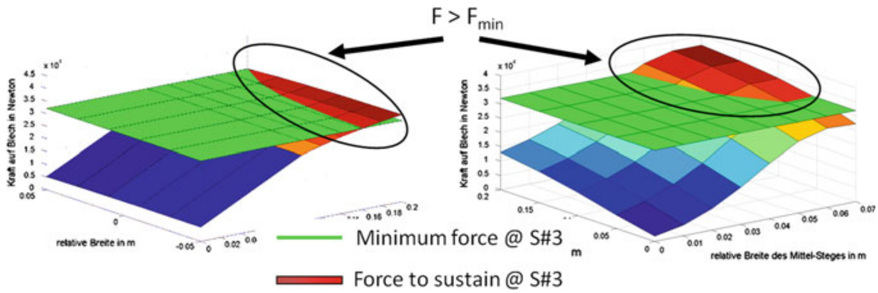


Fig. 7 Magnetic force

In black the permanent magnets are shown. The arrow represents the magnetization direction of the magnets. In green the structure of the magnetic lifting device is displayed. The permanent magnets occupy a relatively large volume and relatively little material is needed to lead the magnetic flux. The two variants are varied in their complete width an as a second parameter the width of the middle connector is varied relative to the basic variant. The results are shown in Fig. 7.

In the two areas the minimum force of 32 kN is reached. The minimum force is displayed in green. So the design with the lowest weight out of both variants is chosen and compared to each other. The results are shown in the following Table 1.

Design 2 has some advantages, it is lighter and smaller, but need a larger coil. So it is possible to increase the conductor diameter, so the needed current is lower. This is also seen on the power dissipation. This amount is converted to heat and must be dissipated. The coil must only energize, when the sheets dropped off. When the logistic circle is known, the total amount of energy can be calculated. The safety factor of three is reached. Even when 5 out of 8 fail, the system will survive with a safety factor of one. In this consideration no failure probability is considered, this could be done in a further step.

Table 1 Simulation results

	Design 1	Design 2
Volume PM	0.0564 m^3	0.0456 m^3
Volume iron	0.11256 m^3	0.0932 m^3
Mass (PM + Iron)	10.7 t (= 8 × 1.34 t)	8.8 t (= 8 × 1.1 t)
Coil (Cu)	1.12 t (= 8 × 140 kg)	1.8 t (= 8 × 235 kg)
Current density of coil	9 A/mm^2	5 A/mm^2
Power dissipation	~20 kW	~10.5 kW

Realization of the Concept

After the successful simulation a model based on simulation results is constructed. With this model it is possible to ensure the function of previously computed system. This model is shown in Fig. 8.

This model is fully functional and similar to the real process. Sheets can be loaded and unloaded to a ship located at the right hand side of the model. The operator can select how many sheets he wants to pick up. Then he can actuate the Traverse and store the plates in the ship or on land in the designated container. The model has some of the same safety features as the future Traverse in the port. For example, the Traverse can detect, if it straight up in the air or on the ground. Only when the Traverse is on the ground, the picked up sheets can be released from the magnetic lifting device. Furthermore, there are two speeds, with the Traverse can be moved. Thus, a precise positioning of the Traverse over the container or the ship is possible.

Furthermore, for the operator multiple operating devices are available, so he can obverse the state of the Traverse. It is contemplated that this information later in the real Traverse the operator are presented. If the emergency stop switch actuated, the entire electricity of the system is switched off, and so a blackout is simulated. The model behaves exactly as predicted in the technical concept, the

Fig. 8 Scaled down model

crane can not proceed and the plates remain on the magnetic lifting device. After restoring power to the crane, it can go to its regular operation. With the construction of the model, it was possible to detect various influencing factors by testing the process.

Conclusion

A system is designed, which can reliable transport three metal sheets at one transport cycle. For safety reasons permanent magnets are used to create a static magnetic force. To compensate the field of the permanent magnets a common coil is used. In simulation a static force is approx 32 kN per magnetic lifting device on the third sheet is calculated. With the eight devices of the complete system can lift a weight of 256 kN. This force appears on the third sheet. When fewer components are carried this static force will be raised, this could be interesting for further consideration. The principle of the concept was successfully tested on a scaled down model. It is possible to transport sheets from a store facility to a ship. The used Traverse creates the static transport force completely passive. Even when the complete power supply of the Traverse is shut down, the sheets remain on the magnetic lifting unit and the Traverse can't move. After connecting the Traverse back to the power supply the Traverse can operate normally again. Additional to this no storage timber is needed. The sheets are located directly on each other. So more sheets can be stored in the same transport unit and the environment will be conserved.

In further steps the force measurement must be catch up. Also in a further step the Traverse could be redesigned, so that it is complete portable and can be used with several cranes all over the world.

References

Cassing W et al (2008) ThyssenKrupp techforum 1. Accessed 31 Oct 2011 http://www.thyssenkrupp-magnettechnik.com/pdf/Elektromagnete_Lang_d.pdf
Fisher U et al (2008) Tabellenbuch Metall. Verlag Europa Lehrmittel, Haan-Gruiten
Marinescu (1996) Elektrische und magnetische felder. Springer, Berlin
Nordmann R (2007) Mechtronik im Maschienbau 2. TU-Darmstadt, Skriptum
Woko Magnet-und Anlagenbau GmbH (2011) Firmenprospekt, Duisburg
Zastrow D (2004) Elektrotechnik, vieweg, Wiesbaden

Part X
Considerations for a Future Internet of Things

The Effective Method for Distributed Information Gain in EPC Network

Taewoo Nam, Byeongsam Kim and Keunhyuk Yeom

Abstract Automatic object identification and distributed computing technology have been used to realize a ubiquitous computing environment. Radio Frequency Identification (RFID) technology has been applied to many business areas such as logistics and supply chains. EPCglobal, which is currently focusing on developing a worldwide standard for RFID, has proposed an open global standard named the EPCglobal network (EPC Network). The amount of communication between each component affects the performance of the EPC Network considerably, because RFID application systems get information from sub-systems of the EPC Network, consisting of the Application Level Event (ALE), Object Naming Service (ONS), EPC Information Service (EPCIS), and EPCIS Discovery Service (EPCIS DS). Therefore, RFID application developers should consider how to communicate with one another, as this is crucial for achieving good performance of RFID application systems in an EPC Network. This paper proposes a method for obtaining distributed information effectively in an EPC Network, and the use of this method results in a reduction of 78 % in the maximum processing.

Keywords EPC network · EPCIS · RFID application

T. Nam · K. Yeom (✉)
Department of Computer Science and Engineering, Pusan National University, 30 Jangjeon dong, Geumjeong Gu, Busan 609-735, Republic of Korea
e-mail: yeom@pusan.ac.kr

T. Nam
e-mail: kaluas@pusan.ac.kr

B. Kim
Food Safety and Distribution Research Group, Korea Food Research Institute, 1201-62 Anyangpangyo-ro, Bundang Gu, Seongnam, Gyeonggido, Republic of Korea
e-mail: bskim@kfri.re.kr

H.-J. Kreowski et al. (eds.), *Dynamics in Logistics*, Lecture Notes in Logistics,
DOI: 10.1007/978-3-642-35966-8_45,

Introduction

Radio Frequency Identification (RFID) is the technology that involves the identification of people and objects using radio frequency transmission (Juels 2006). This technology has been applied to many business areas, including logistics and supply chains. RFID can recognize many tags at a time and can also recognize large amounts of stored information through memory. The nonprofit organization EPCglobal, which is currently focusing on developing a worldwide standard for RFID, has proposed an open global standard named the EPCglobal Network (EPC Network) (GS1 EPCglobal).

The EPC Network provides an architecture that aggregates and filters EPCs and references product information. The application of the EPC network standard is effective when developing RFID application systems, such as those for distribution and logistics (EPCglobal 2009a).

The amount of communication between components affects the performance of an EPC Network considerably because RFID application systems get information from many components, consisting of many sub-systems such as the Application Level Event (ALE) (EPCglobal 2009b), Object Naming Service (ONS) (EPCglobal 2008), EPC Information Service (EPCIS) (EPCglobal 2007), and EPCIS Discovery Service (EPCIS DS) (EPCglobal 2009a). This is particularly true when the number of EPCs from an ALE increases, as the amount of communication between components also increases. For this reason, research for better RFID application systems is necessary to obtain information efficiently when implementing the EPC Network standard.

It is possible to improve the performance of an RFID application by reducing the amount of communication between the system and components when obtaining distributed information. For this, we propose (and verify using experiments) a method to reduce the amount of communication with the EPCIS component that is used most often for obtaining master data and event data.

This paper is organized as follows. Section EPC Network Architecture describes the process of getting distributed information. Section Performance Analysis of Distributed Information Acquisition Based on EPC Network presents a performance analysis for obtaining distributed information based on an established EPC network. Section Method for Obtaining Effective Distributed Information presents the method for obtaining distributed information efficiently and an estimation of the performance improvement. Section Experiment and Result presents an analysis of the method proposed in this paper by using experiments with existing data.

EPC Network Architecture

The ISO/IEC, the International standards organization, and EPCglobal, the organization specializing in RFID standardization, have encouraged the use of RFID technology since the late 1990s. The EPC Network is the RFID standard in the

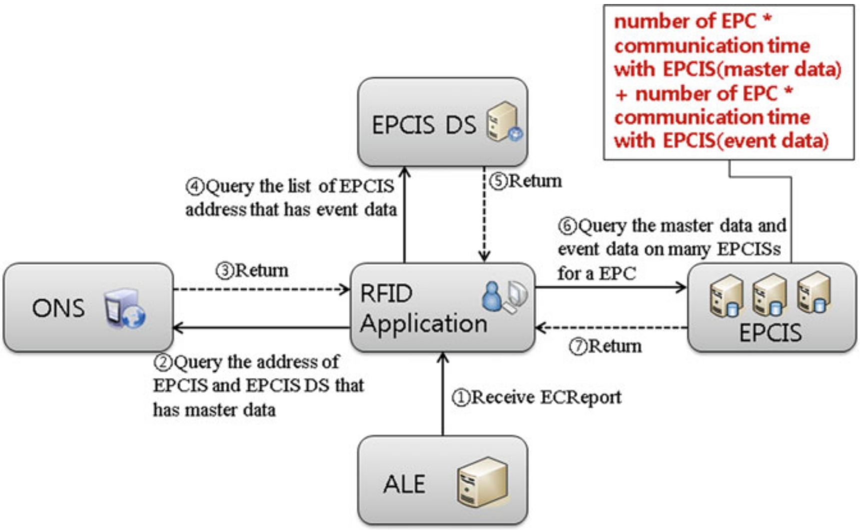


Fig. 1 EPC network architecture and flow of information gain

industrial world. The EPC Network is an EPC information distribution/management infrastructure that provides master data and event data. It provides EPC technology that realizes an identification system for assigning identification codes to identify each object, and RFID technology that can identify high-speed objects from a distant location (GS1 EPCglobal).

Figure 1 illustrates the simple structure of the EPC Network architecture and shows the process of obtaining distributed information in an EPC Network environment. When readers identify an RFID tag, the ALE collects and filters the EPC. The RFID system gets the EPCIS and the EPCIS DS address through ONS based on EPC data aggregated from the ALE; the EPCIS address list that has event data for objects is also obtained through the EPCIS DS. The RFID system gets master data and event data for an object from each EPCIS through these processes. There are two kinds of EPCIS data: event data and master data. Event data, which are created in the process of carrying out business processes, are captured through the EPCIS Capture Interface and then made available for queries through the EPCIS Query Interfaces. Master data are additional data that provide the necessary context for interpreting the event data.

Performance Analysis of Distributed Information Acquisition Based on EPC Network

Types of Distributed Information Obtained in EPC Network

When obtaining distributed information in an EPC Network, the communicating sub-systems are changed in accordance with the information to be obtained. Therefore, it is necessary to carry out a performance analysis, along with the types of distributed information obtained from the EPC Network from the EPC Network analysis. Table 1 lists the types of distributed information and the sub-systems that use the information in an EPC Network.

Table 1 contains three categories according to the types of distributed information: master data, event data, and master data + event data. There are two possible cases for each category, namely, static EPCIS and dynamic EPCIS. The static EPCIS applies when the address of the sub-system with which communication is to be initiated is known. Dynamic EPCIS obtains the address from the ONS when the fixed communication address is not known. Therefore, when using static EPCIS, all EPCs obtained through the ALE physically communicate with the same sub-systems, because they know the addresses of the sub-systems. Dynamic EPCIS needs to obtain the EPCIS address containing the master data through ONS, and the dynamic EPCIS DS also has to obtain the EPCIS DS address using ONS.

Performance Analysis

In this research, we analyzed the process time of an RFID system with different types of distributed information obtained in an EPC Network and expressed the results in numerical formulas. Before carrying out the performance analysis, we need to define variables for calculating the communication time and process time for each sub-system. Table 2 lists these variable definitions.

Table 1 Type of distributed information acquisition

Types of distributed information		Sub-systems that use the information
Master data	Static EPCIS	ALE, EPCIS
	Dynamic EPCIS	ALE, ONS(IS), EPCIS
Event data	Static EPCIS DS	ALE, EPCIS DS, EPCIS
	Dynamic EPCIS DS	ALE, ONS(DS), EPCIS DS, EPCIS
Master data + event data	Static EPCIS, static EPCIS DS	ALE, EPCIS DS, EPCIS
	Static EPCIS, dynamic EPCIS DS	ALE, ONS(DS), EPCIS DS, EPCIS
	Dynamic EPCIS, static EPCIS DS	ALE, ONS(IS), EPCIS DS, EPCIS
	Dynamic EPCIS, dynamic EPCIS DS	ALE, ONS(DS, IS), EPCIS DS, EPCIS

Table 2 Variable definitions for performance analysis

Variables	Descriptions	Variables	Descriptions
T	Total processing time	T'_{ONS_IS}	$T_{ONS\_INIT} + T_{ONS\_S}$
N_{EPC}	Amount of EPC obtained through communication with ALE	T_{DS_INIT}	Preparation time for communication with EPCIS DS
N_{IS_DS}	Amount of EPCIS for event data	T_{DS}	Communication time with EPCIS DS for EPCIS address that has event data
T_{ALE}	Communication time with ALE	T'_{DS}	$T_{DS\_INIT} + T_{DS}$
T_{ONS_INIT}	Preparation time for communication with ONS	T_{IS_INIT}	Preparation time for communication with EPCIS
T_{ONS_DS}	Communication time with ONS for EPCIS DS address	T_{IS}	Communication time with EPCIS to obtain master data and event data
T'_{ONS_DS}	$T_{ONS\_INIT} + T_{ONS\_DS}$	T'_{IS}	$T_{IS\_INIT} + T_{IS}$
T_{ONS_IS}	Communication time with ONS for EPCIS address that has master data	T_P	Information process time without communication time

Table 3 lists the numerical formulas for all information process times based on the variables defined in Table 2.

In order to collect distributed information in an EPC network, it is necessary to communicate with the sub-systems as often as EPC information is obtained from ALE (Table 3). With an increase in communication, the system performance deteriorates. Therefore, the performance of the RFID system can be improved by decreasing the amount of communication with EPC Network sub-systems.

Method for Obtaining Effective Distributed Information

Performance Improvement Method Through Group Query as per EPCISs

This paper proposes a method for improving system performance by first grouping EPCs as per the EPCIS for the RFID system. This method is as follows. Acquire all EPCIS addresses through ONS and EPCIS DS in the RFID system. Then, group EPCs as per EPCISs and query the EPCISs for master data and event data for several EPCs at the same time to acquire distributed information. By decreasing the frequency of EPCIS communication, we can improve the processing time of the RFID system. Table 3 illustrates the results of this method.

The distributed information flow as per EPCISs is shown in Fig. 2.

As shown in Fig. 2, we obtain EPCIS addresses for master data and event data through the ONS and EPCIS DS. With this information, we eliminate overlapped EPCIS address in the RFID system. Then, we place a query to the EPCIS after grouping the information for each EPC. For example, if there are 100 EPCs and all

Table 3 Information processing time in EPC Network

Types of distributed information		Formula of process time
Master data	Static EPCIS	$T = T_{ALE} + T_{IS\_INIT} + \sum_{n=1}^{N_{EPC}} T_{IS_n} + T_p$
	Dynamic EPCIS	$T = T_{ALE} + T_{ONS\_INIT} + \sum_{n=1}^{N_{EPC}} (T_{ONS\_IS_n} + T'_{IS_n}) + T_p$
Event data	Static EPCIS DS	$T = T_{ALE} + T_{DS\_INIT} + \sum_{n=1}^{N_{EPC}} (T_{DS_n} + \sum_{k=1}^{N_{IS\_DS_n}} T'_{IS_k}) + T_p$
	Dynamic EPCIS DS	$T = T_{ALE} + T_{ONS\_INIT} + \sum_{n=1}^{N_{EPC}} (T_{ONS\_DS_n} + T'_{DS_n} + \sum_{k=1}^{N_{IS\_DS_n}} T'_{IS_k}) + T_p$
Master data + event data	Static EPCIS, static EPCIS DS	$T = T_{ALE} + T_{DS\_INIT} + T_{IS\_INIT} + \sum_{n=1}^{N_{EPC}} (T_{IS_n} + T_{DS_n} + \sum_{k=1}^{N_{IS\_DS_n}} T'_{IS_k}) + T_p$
	Static EPCIS, dynamic EPCIS DS	$T = T_{ALE} + T_{ONS\_INIT} + T_{IS\_INIT} + \sum_{n=1}^{N_{EPC}} (T_{ONS\_DS_n} + T_{IS_n} + T'_{DS_n} + \sum_{k=1}^{N_{IS\_DS_n}} T'_{IS_k}) + T_p$
	Dynamic EPCIS, static EPCIS DS	$T = T_{ALE} + T_{ONS\_INIT} + T_{DS\_INIT} + \sum_{n=1}^{N_{EPC}} (T_{ONS\_IS_n} + T'_{IS_n} + T_{DS_n} + \sum_{k=1}^{N_{IS\_DS_n}} T'_{IS_k}) + T_p$
	Dynamic EPCIS, static EPCIS DS	$T = T_{ALE} + T_{ONS\_INIT} + \sum_{n=1}^{N_{EPC}} (T_{ONS\_IS_n} + T_{ONS\_DS_n} + T'_{IS_n} + T_{DS_n} + \sum_{k=1}^{N_{IS\_DS_n}} T'_{IS_k}) + T_p$

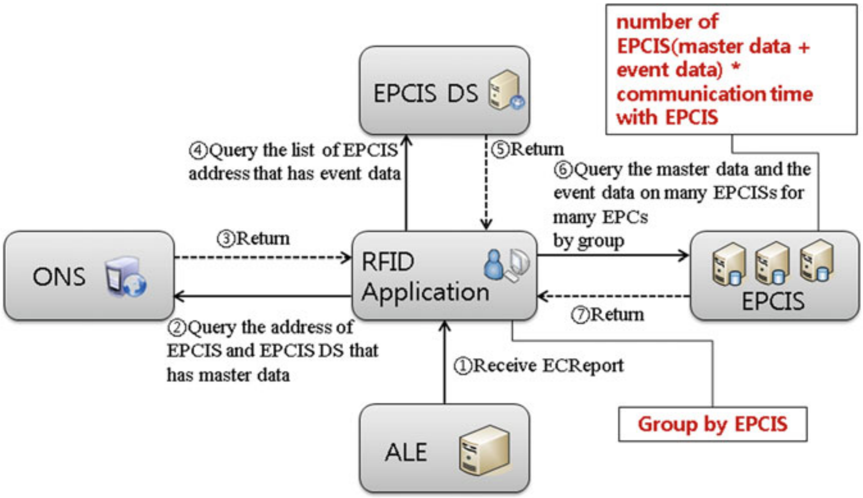


Fig. 2 Obtaining distributed information after grouping query by EPCISs

products have their master data registered in the “A” EPCIS, it is possible to obtain the master data for the 100 products in one communication.

With the information we obtain after grouping EPCs into EPCISs, we can improve the performance of the RFID system by reducing the frequency of overlapping EPCIS communication.

Performance Analysis

In section Types of Distributed Information Obtained in EPC Network, we analyze the entire process time with numerical formulas with respect to the types of distributed information obtained from EPCISs in an EPC Network. In the case in which distributed information is obtained from EPCISs, new variables must be defined, because EPCIS numbers are being changed in accordance with the redundancy of EPCIS addresses from the ONS and EPCIS DS.

We define the EPCIS numbers as NIS_MS and N′IS_MS. The former refers to the number of communications with the EPCIS to obtain the master data, and the latter is the number of communications for obtaining the event data. Using these two variables and the variables suggested in section Performance Analysis, we can obtain the results listed in Table 4.

In existing methods, the frequency of communication with the EPCIS increases as the EPC numbers from the ALE increase. However, the method that this paper suggests shows that communication with the EPCIS occurs as many times as there are EPCIS numbers after eliminating overlapping EPCIS information, as listed in Table 4.

Table 4 Information processing time after applying the group query method as per EPCIS

Types of information		Formula of process time
Master data	Static EPCIS	$T = T_{ALE} + T'_{IS} + T_p$
	Dynamic EPCIS	$T = T_{ALE} + T_{ONS\_INIT} + \sum_{n=1}^{N_{EPC}} T_{ONS\_IS_n} + \sum_{n=1}^{N_{IS\_MS}} T'_{IS_n} + T_p$
Event data	Static EPCIS DS	$T = T_{ALE} + T_{DS\_INIT} + \sum_{n=1}^{N_{EPC}} T_{DS_n} + \sum_{n=1}^{N'_{IS\_DS}} T'_{IS_n} + T_p$
	Dynamic EPCIS DS	$T = T_{ALE} + T_{ONS\_INIT} + \sum_{n=1}^{N_{EPC}} (T_{ONS\_DS_n} + T'_{DS_n}) + \sum_{n=1}^{N_{IS\_DS}} T'_{IS_n} + T_p$
Master data + event data	Static EPCIS, static EPCIS DS	$T = T_{ALE} + T_{DS\_INIT} + T'_{IS} + \sum_{n=1}^{N_{EPC}} T_{DS_n} + \sum_{n=1}^{N'_{IS\_DS}} T'_{IS_n} + T_p$
	Static EPCIS, dynamic EPCIS DS	$T = T_{ALE} + T_{ONS\_INIT} + T'_{IS} + \sum_{n=1}^{N_{EPC}} (T_{ONS\_DS_n} + T'_{DS_n}) + \sum_{n=1}^{N'_{IS\_DS}} T'_{IS_n} + T_p$
	Dynamic EPCIS, static EPCIS DS	$T = T_{ALE} + T_{ONS\_INIT} + T'_{DS} + \sum_{n=1}^{N_{EPC}} T_{ONS\_IS_n} + \sum_{n=1}^{N_{IS\_MS}} T'_{IS_n} + \sum_{n=1}^{N_{IS\_DS}} T'_{IS_n} + T_p$
	Dynamic EPCIS, static EPCIS DS	$T = T_{ALE} + T_{ONS\_INIT} + \sum_{n=1}^{N_{EPC}} (T_{ONS\_IS_n} + T_{ONS\_DS_n} + T'_{DS_n}) + \sum_{n=1}^{N_{IS\_MS}} T'_{IS_n} + \sum_{n=1}^{N'_{IS\_DS}} T'_{IS_n} + T_p$

Table 5 Formula for performance Improvement expectations

Type of information	Expectations of performance improvement
Master data (static EPCIS)	$\sum_{n=1}^{N_{EPC}-1} T'_{IS_n}$
Master data (dynamic EPCIS)	$\sum_{n=1}^{N_{EPC}-N_{IS\_MS}} T'_{IS_n}$
Event data	$\sum_{n=1}^{N} T'_{IS_n}$, when $N = \sum_{n=1}^{N_{EPC}} N_{IS\_DS_n} - N'_{IS\_DS}$

Using the numerical formulas listed in Tables 3 and 4, we can obtain a theoretical expected value of performance improvement by comparing the existing method with the new method proposed in this paper.

We can divide the theoretical expected value of performance improvement into two categories—static EPCIS and dynamic EPCIS—for both master data and event data.

Obtaining master data in a static EPCIS is possible in a single communication with the EPCIS, regardless of the number of EPC numbers that exist, because all EPCs communicate with one EPCIS.

Obtaining master data in a dynamic EPCIS requires communication with the EPCIS. At this point, we calculate the EPCIS number by subtracting the EPCIS numbers we need from the collected EPC numbers.

When obtaining event data, the EPCIS numbers we need are different according to each EPC and can be obtained by communicating with the EPCIS. In this case, the EPCIS is the number of whole EPCIS numbers, except for those that need to be communicated. Table 5 lists the numerical formulas for this.

Experiment and Result

Definition of EPCIS Redundancy

Once we place a group query to the EPCIS using the method suggested in this paper, the performance values change according to the group queries, because there are many possible ways to group EPCs obtained from the ONS and EPCIS. Therefore, we need to define EPCIS redundancy.

EPCIS address redundancy can be divided into two cases: master data acquisition and event data acquisition.

R_{d_ISMS} is the EPCIS redundancy in the case of master data acquisition. It refers to the overlapping rate of EPCIS required to obtain EPC master data from the ALE, as shown in (1). If there are 100 EPCs from the ALE and we acquire the master data without grouping, we need to communicate with the EPCIS 100 times. In this case, the value of R_{d_ISMS} is 0 %. However, if there is only one EPCIS address from the ONS, we only need to communicate with one EPCIS. Therefore, the value of R_{d_ISMS} becomes 99 %.

$$R_{d\_ISMS} = \frac{N_{EPC} - N_{IS\_MS}}{N_{EPC}} * 100\ (\%) \tag{1}$$

$$R_{d\_ISDS} = \frac{\sum_{n=1}^{N_{EPC}} N_{IS\_DS} - N'_{IS\_DS}}{\sum_{n=1}^{N_{EPC}} N_{IS\_DS}} * 100\ (\%) \tag{2}$$

R_{d_ISDS} is the EPCIS redundancy in the case of event data acquisition. This refers to the overlapping rate of EPCIS, as shown in (2). It features the difference of EPCIS numbers according to each EPC.

Equation (2) shows that if a product, P1, has its event data at A-B–C-D EPCIS and another product, P2, has its event data at E–F–C-D EPCIS, we can obtain the information in six communications. In this case, the value of R_{d_ISDS} is 25 %.

We can use (1) and (2) to define the redundancy range. The value of R_{d_ISDS} is 0 % when the EPCIS numbers are the same in both the existing method and in the proposed method.

$$\lim_{N_{EPC} \to \infty} R_{d\_ISMS} = 100\ (\%)\ (\text{if},\ N_{IS\_MS} = 1) \tag{3}$$

$$\lim_{N_{EPC} \to \infty} R_{d\_ISDS} = 100\ (\%)\ (\text{if},\ N'_{IS\_DS} = 1) \tag{4}$$

In addition, from (3) and (4), it can be inferred that as the value of NEPC tends toward infinity, the redundancy tends toward 100. Therefore, the range can be considered to be 100 % when the communication frequency with the EPCIS is 1 for both master and event data acquisition. Therefore, the EPCIS redundancy range can be defined as being between 0 and 100 %.

Experiment Data

We extracted the experiment data by measuring the communication time between actual sub-systems in the EPC Network. EPC Network sub-systems such as ALE, ONS, EPCIS DS, and EPCIS use the standard EPC Network system. We measured the communication time using Business Aware Framework (BizAF) (Kim et al. 2007), which provides information to an RFID system by communicating with the EPC Network sub-systems. WebService is used for communicating with each sub-system, and the EPC numbers collected at one time is fixed at 100. We performed a total of 100 experiments, using one ALE, ONS, and EPCIS DS, and four EPCISs. The average time of each sub-system was recorded as experiment data.

Table 6 lists the experiment data. The T_{IS} value is much higher than the other values in Table 6. This result supports the method proposed by this paper; therefore, we fixed the TIS value at 200 ms, shown as T_{IS_INIT}. The T_{ALE} value

Table 6 The experimental data

Variables	Value (ms)
T_{DS_INIT}	163
T_{ONS_INIT}	31
T_{ONS_IS}	101
T_{IS}	1360
T_{DS}	235
T_{ONS_DS}	61
T_{IS_INIT}	200
N_{IS_DS}	5

was not considered because it is applied equally in all cases. The T_P value was also not considered because it changes in every case.

The Proof of Concept in the EPCglobal US Conference 2005 presented an example of product movement flow (Tearnen 2005). The flow from Sunshine Supplies to Best Distribution to Logistic Provider to Acme Retail showed a traditional product distribution flow from the manufacturer to the distributor, retailer, and finally to the customer (Laudon et al. 2006). In general, distribution event data follows these four steps (McFarlane et al. 2003). Therefore, this study assumed one EPCIS at every distribution step, and we fixed the N_{IS_DS} value at 4.

Experiment Result

We performed an experiment using the method developed in this paper to attempt to decrease the process time using the experiment data presented in section Experiment Data. The experiment is concerned with a change in the entire process time by incorporating EPCIS redundancy and the method developed in this study. The aim of the experiment is to obtain master data and event data in a dynamic EPCIS. We selected dynamic EPCIS because of its high process time.

We fixed the N_{EPC} value at 100 for this experiment. Figure 3 shows the results of the experiment.

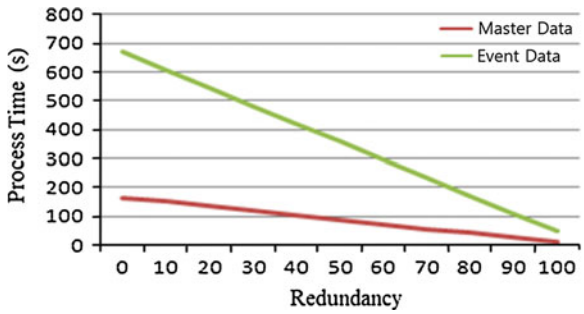


Fig. 3 Changes in processing time

As can be seen from Fig. 3, the processing time when obtaining both the master data and event data decreases as the EPCIS redundancy values change. In the case of master data, the process time diminishes by 92.96 % when the redundancy goes from 0 to 100 %. In the case of event data, the process time diminishes by 92.91 %.

In our research, we performed the experiment to measure the change in process time by increasing the EPC numbers collected at one time from 1 to 1,000. Figure 4 shows the results of this experiment.

When obtaining master data and event data, the process time decreases by 39 % when the EPCIS redundancy is 50 %. When the EPCIS redundancy is 100 %, the process time decreases by 78 %. Depending on the EPCIS redundancy, the degree of change in the overall process time differs. Although the EPCIS redundancy is 0 %, the process time is identical to that of the existing method.

Regardless of EPC numbers, the process time decreases at a regular rate, following the EPCIS redundancy. This is because this experiment uses average subsystem values, and the communication with the existing ONS and EPCIS DS increases with the EPC.

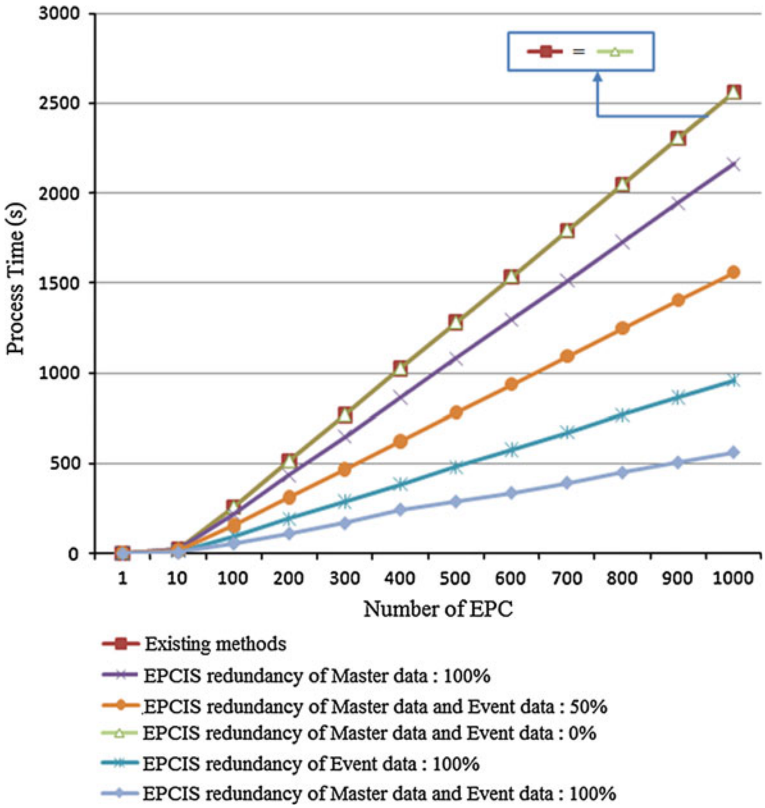


Fig. 4 Changes in processing time compared to existing methods

So far, our research shows that our method for obtaining distributed information reduces the frequency of EPCIS communication for the most used sub-systems in an EPC network. As a result, the performance of the RFID system improves.

Conclusion and Future Work

This paper proposed a method for effectively obtaining distributed information by placing EPCIS group queries. With an EPCIS group query, it is possible to decrease the frequency of communication when there are identical EPCISs. This contributes to RFID system improvement.

In addition, we defined EPCIS redundancy and experimented with process time changes according to redundancy. The experiment showed that the overall process time decreased as the EPCIS redundancy values changed.

Further studies are needed to analyze the decrease in the process time when collecting data in an actual environment. In addition, after measuring the TP value and considering the difference between the proposed and existing methods, the difference between TP and the communication time needs to be analyzed. Further studies should try to establish a method for decreasing the frequency of communication with ONS and EPCIS DS.

Acknowledgments This work was supported by u-Food System Development Project of the Korea Food Research Institute.

References

EPCglobal (2007) EPC Information Services (EPCIS) version 1.0.1 specification

EPCglobal (2008) Object Naming Service (ONS) specification version 1.0.1

EPCglobal (2009a) The EPCglobal architecture framework version 1.3

EPCglobal (2009b) Application Level Events (ALE) specification version 1.1.1

GS1 EPCglobal (http://www.epcglobalinc.org)

Juels A (2006) RFID security and privacy: a research survey. Ann IEEE J Sel Areas Commun 24:381–394

Kim S, Moon M, Kim S et al (2007) RFID business aware framework for business process in the epc network. In: Annals of proceedings of the 5th international conference on software engineering research and applications (SERA 2007), pp 468–475

Laudon J, Ken C (2006) Annals of management information systems: managing the digital firm 9th edition prentice hall: 134–135

McFarlane D, Sheffi Y (2003) Impact of Automatic Identification on Supply Chain Operations. Ann Int J Logistics Manag 14:1–17

Tearnen P (2005) EPCglobal US proof of concept. In: Annals of EPCglobal US conference

Man–Machine-Interaction in the Field of Logistics: Example "Internet of Things"

Lars Windelband and Georg Spöttl

Abstract The introduction of ICT-systems (ICT: Information- and communication technologies) has led to fundamental changes in numerous fields of the world of work in recent years. Above all the networking with the aid of ICT-applications (e.g. safety systems in cars, airplanes etc.) has created many possibilities to monitor complex systems and processes and to control them more and more without human interference. Weyer (Autonomie und Kontrolle. Arbeiten in hybriden Systemen am Beispiel der Luftfahrt. Technikfolgenabschätzung—Theorie und Praxis Nr. Dortmund, 2007, p. 35) comments this development as follows: „Ein besonderes Merkmal der jüngsten Entwicklung ist zudem das scheinbar unaufhörliche Vordringen autonomer technischer Systeme, die immer mehr zu Mitspielern in derartigen Netzwerken geworden sind" (Translation: A special characteristic of this recent development is the apparently continuous permeation of autonomous technical systems which have increasingly become players in such networks). The emergence of "hybrid constellations, permeated by human actors and (semi)autonomous machines" (cf. ibid.) is one of the consequences of this development. As it will be shown later, the autonomy of ICT-systems has been stepwise increasing while the role of the human actors is taking a back seat. The triggers are new high-tech-developments in the field of RFID (Radio-Frequency Identification) (Want et al. IEEE Pervasive Computing, 2006, 5:25–33) and the "Internet of Things" [The term is attributed to Kevin Ashton who used the expression "Internet of Things" for the first time in 1999 (Ashton, RFID Journal. http://www.rfidjournal.com/article/view/4986, 2009). The "Internet of Things" describes the combination of the real world with the virtual world of

L. Windelband (✉) · G. Spöttl
ITB—Institute Technology and Education, University of Bremen, Am Fallturm 1 28359 Bremen, Germany
e-mail: lwindelband@uni-bremen.de

G. Spöttl
e-mail: spoettl@uni-bremen.de

H.-J. Kreowski et al. (eds.), *Dynamics in Logistics*, Lecture Notes in Logistics,
DOI: 10.1007/978-3-642-35966-8_46, © Springer-Verlag Berlin Heidelberg 2013

information technology by means of automatic identification technologies, realtime locating systems, sensors and actuators (Uckelmann et al. Architecting the Internet of Things. Springer, Berlin, 2011, p. 2). As for the field of logistics, the Institut für Materialfluss und Logistik – IML – Institute for Material Flow and Logistics defines the "Internet of Things" as the autonomous, self-controlled transport of logistic objects from the sender to the consignee (cf. ten Hompel in Internet der dinge: status, perspektiven, aspekte der aktuellen RFID-Entwicklung. Vortrag, 2005)]. A new level of technical development has thus been reached by this autonomous control of corporate processes. It exerts a crucial influence on the social and economic forms of organization and on the organization of work. Contrary to machines which carry out tasks and processes determined and controlled by human beings, the networked systems in the "Internet of Things" are able to make their own decisions. The implementation of the "Internet of Things" therefore marks a new level of the division of work in the field of man–machine-interaction.

Status of the "Internet of Things"

The historical development of the "Internet of Things" is not a linear, individual technological development but rather the convergence of histories of different approaches that have been emerging parallelly over a long period of time. More or less all is about RFID- and auto-identification technologies, "Ubiquitous Computing" and mobile communications networks. Ubiquitous Computing was adopted in the late 1990s by the movement of "Ambient Intelligence" and was put into a context with mobile communication. This is the first clear hint to a convergence of individual development strands towards an "Internet of Things" (cf. Cetin 2005).

Strictly speaking the "Internet of Things" is no independent technology but a summary or combination of various technologies which complement one another. This encompasses identification technologies, communication technologies, sensorics, compact information processing systems and energy supply technologies.

Identification technologies (barcode, RFID), digital product memories[1], the intelligent networking of products and the autonomous acting with the aid of special software agents and assistant systems are considered to be the potential application spectrum of the "Internet of Things" in the field of logistics. Further fields of application are logistic control and tracking systems as well as the self-organized transport of logistic objects by in-firm and external transport networks (cf. ten Hompel and Bullinger 2007). A considerable number of research projects

[1] In this case objects are equipped with "Smart Labels" combining RFID with their own power supply and a set of various sensors. Thus the products are able to compile data from their environment. The state of the goods, the state of freshness, storage temperature, origin etc. can be controlled and read out at any time (cf. Brand et al. 2009, p. 106f).

have been implemented for the promotion of these application possibilities. "ICT for Logistics and Service Tasks" represents an important focus within the program ICT 2020 of the German Federal Ministry of Education and Research (BMBF) with a target-oriented promotion of the aspect of the "Internet of Things". But also other fields of promotion offer relevant activities, e.g. the Federal Ministry for Economy and Technology (BMWi). Projects with emphasis on "convergent ICT" (Next Generation Media) are promoted within the field of logistics.

The equipment of objects, rooms and machines with different technologies for the perception of the environment, data storage, communication and autonomous acting forms the technological basis of the "Internet of Things" (cf. Dworschak et al. 2011, p. 2). Above all RFID is increasingly applied. RFID-systems can read out data contactless and without intervisibility. These devices consist of a scanner and a transponder and/or a "tag", i.e. a chip with an antenna. The object equipped with the device can thus be unambiguously identified and the exact path of the individual products can be tracked. RFID-systems are combined with a set of sensors that can be embedded in the products themselves. These sensor data can be automatically updated with the aid of RFID. Due to increased demands with respect to flexibility and mobility, the development of the RFID-systems and sensors tends to wireless systems that require a radio-based transmission of information.

Many experts are convinced that the "Internet of Things" will lead to a considerable optimization of logistic processes. With the aid of active sensors in transport agents, a so-far unknown close-meshed process control can be realized which will considerably facilitate trouble-shooting and the identification of error sources. The organization of logistic processes can thus be sustainably improved and optimized in a cost-effective way. Examples for the application are control solutions for goods, vehicles and transport containers. If necessary, a seamless documentation of the transport process can help to file claims for damages caused by the unprofessional handling of the goods and by the opening of containers. In the long term, this may lead to an improvement in quality of transport services. Another side effect of this complete process control and documentation would enable the customers to access information on the status of their transported goods within the logistic chain at any time. This could lead to a competitive advantage as soon as short delivery cycles are envisaged or in case of logistic chains and their various and often critical transit stations (e.g. country borders, customs) which can be unambiguously identified and tracked.

The current research projects concentrate on the implementation of these objectives. One example is the project "Intelligent Container"-*collaborative research centre* 637 (CRC 637) which is being carried through in cooperation with the University of Bremen. The "Intelligent Container" plans and controls itself entirely autonomously. It determines its own optimal transport path by evaluating external information (traffic and market information) and its internal conditions (environmental impact on the goods) and makes its own decisions (cf. Jedermann and Lang 2008, p. 24f). Further examples: A pilot project run by the logistics service provider Schenker monitors and documents the movement and the opening of containers with the aid of RFID-TAGs. The containers are registered during different phases: at first at the liability transit points in ports, later in the container terminals themselves. This information is shown in the associated systems. Another project located at the University of Dortmund deals with the control hubs of automatic luggage conveying systems at major airports-called "routing agents". The agents decide independently where the pieces of luggage are bound to. All pieces of luggage get special "Smart Labels" that enable them to communicate with the control hubs. Thus they can request the necessary resources on their own (Gaßner et al. 2009 p. 21). Many of these projects are pilot applications or are still in their start-up period. A forecast for the implementation of these technologies in practice is therefore still quite difficult. On the other hand these research activities are already pointing out the potential of these technologies. Manifold support programmes have been created in Germany and Europe. In Europe, a research network called "European Research Cluster on the Internet of Things" (IERC) has already been established (IERC 2011).

The "Internet of Things" offers numerous possibilities for an improvement of the efficiency and effectiveness of logistic processes. These technology-based developments are dominated by a technology-centered automation strategy, i.e. the ambition to develop and implement zero-defect systems and to eliminate the human being as a possible error source. The balance between man and machine is a crucial factor and has so far played a subordinate or even no role at all in the development of the individual technologies and products.

Research-Questions and -Design for the Diffusion of the "Internet of Things"

There are first activities in Germany dealing with the concrete effects of the technological developments towards an "Internet of Things" in the world of work and in skilled work. They are embedded in initiatives of early recognition and the research network FreQueNz,[2] funded by the Federal Ministry for Education and Research (Bundesministerium für Bildung und Forschung).

Altogether three studies with emphasis on the "Internet of Things" are being funded in the following application fields:

- Logistics,[3]
- Industrial production and
- Smart House.

[2] FreQueNz is a research network in Germany, funded by the Federal Ministry for Education and Research. Various institutions are contributing to the early recognition of qualification requirements with their project work.

[3] The two authors direct the research activities in the field of logistics within the framework of the study on early recognition FreQueNz (cf. Windelband et al. 2011).

The present article concentrates above all on the transfer of the "Internet of Things" into the real work processes at skilled worker level. The survey is based on the leading questions within the so-called FreQueNz study: To which extent is skilled work in the field of logistics already confronted with the "Internet of Things" and what are the consequences for the organization of skilled work?

In order to answer the research question, three research steps were selected which apply the qualitative vocational educational scientific instruments of early recognition (cf. Windelband and Spöttl 2003):

1. First step: Analysis of the current status of the diffusion of distribution logistics with the "Internet of Things" in the sectors Foods and Automobile Industry.
   Selection of sectors: The two sectors were selected during a preliminary study. The identification of the practice areas concentrated on those already applying the "Internet of Things".
2. Second step: Analysis of innovative companies and research projects for the implementation and realization of the "Internet of Things" in skilled work.[4]
   Selection of companies for case studies: Companies were identified that applied or carried out pilot studies on "Internet of Things" technologies and gave their permission for case studies. Innovative implementation projects should be accessed[5] with the aid of case studies in the research institutions.
3. Third step: Deepening qualitative analyses (expert interviews) for the identification of development trends and the diffusion of the "Internet of Things" in skilled work and possible consequences for qualification.
   Selection of experts: Apart from the prerequisite to be an expert for the "Internet of Things", the selection of experts was based on further parameters: high practice orientation, cooperation in future-oriented projects as well as a focus on one of the two application areas-Foods and the Automobile Industry.

An instrument for the measurement of the "depth of diffusion" of the "Internet of Things" was developed in step two. The instrument encompasses six characteristics which help to identify the depth of diffusion, i.e.: technology, energy supply, connectivity, information processing, aggregation level and location of the intelligence.[6] In more detail:

- With regard to "technology", Level 0 indicates that no technology at all is in use (therefore no indication in the table-this applies for all characteristics of Level 0). Level 1 makes use of an auto-ID (identification), e.g. a RFID or barcode system. Level 2 describes the use of sensors, i.e. for example a sensor network. Level 3 eventually applies embedded systems, i.e. equipment with decision-making components.

---

4 Six companies and four research institutes were surveyed.

5 A case in the application fields of Food or the Automobile Industry is represented by a research project focusing on development and implementation objectives on the "Internet of Things" in the field of distribution logistics.

6 Following Meyer et al. (2009).

- The characteristic of “energy supply” is completely neglected at Level 0, i.e. no energy is supplied. Information is e.g. read out manually. Level 1 comprises systems which make use of induction for their energy supply, e.g. the reading out and description of passive RFID-transponders. Level 2 is characterized by systems that are supplied with energy from an accumulator and by using active RFID-transponders. Systems of Level 3 have a self-sustaining energy supply.
- The characteristic “connectivity” is again inexistent at Level 0, i.e. the system does not communicate. Level 1 describes the manual reading out of information, i.e. the system communicates only passively. At Level 2 the system communicates on demand, e.g. in case of a signal sent to the control centre or a supervisor during a certain event, e.g. if a measuring value is exceeded. Level 3 systems continuously communicate with other systems and are “online” at all times.
- The characteristic “information processing” indicates how the object deals with information. At Level 0 no information is processed. For most of the systems this also means at the same time that no data at all are compiled by the object. At Level 1, information is both compiled and stored, however, not further processed as e.g. in a temperature logger. At Level 2 information may be relayed to other instances without being processed by the object itself. These applies e.g. for telematics systems which can emit a status message in case of certain events. At Level 3 the object is capable to make a decision. This applies e.g. for embedded systems.
- The “level of aggregation” of a technology at Level 0 is inexistent. At Level 1 the technology is placed on the packing, for example on a cardboard box or a container. Level 2 describes a technology at object level, i.e. at the level of a finished product. In terms of vehicles, this technology would be applied per car. Level 3 systems are applied at component level, in vehicles e.g. on a seat.
- The “location of intelligence” at Level 0 is inexistent. At Level 1, the intelligent systems are distributed within a network. A possible application scenario would be the already described sensor network. Level 2 is marked by intelligence at object level, i.e. on the final product, whereas Level 3 describes various locations for intelligence, i.e. different sub-system are used at different levels. They are, however, interlinked into a system.

Research Results

Although it was obvious that companies are increasingly dealing with the “Internet of Things”, the grade of diffusion was mostly comparatively low. The evaluation of the surveys and the case studies showed that hardly any company had reached the second or even the third level. Above all the objects are so-far not autonomously communicating with each other and do not exert any direct influence on the flows of commodity. With regard to the surveyed fields of practice in the companies, i.e. the concrete fields of application of the "Internet of Things", only a low diffusion of the Internet of Things-technologies could be identified, mostly just reaching Level 1 according to Fig. 1.

In spite of this generally relevant result, some companies could be identified in the field of automobile logistics that already had basic technologies available which are required for the implementation of the "Internet of Things". These companies apply at least barcodes or make use of RFID. The data are mainly read out manually with an adequate (mobile or stationary) barcode reader. Only in companies with RFID-technology the data readout took place automatically. Here the objects carried a data storage unit, a prerequisite for networking as stipulated by the "Internet of Things". In the majority of cases a company-specific ICT-system is customized for the purposes of the company. In most cases this system was specially developed for the designated objective. No company at all makes comprehensive use of this technology or leaves decisions to an algorithm receiving the required data from integrated objects. The data are used for the optimization of the logistics chain (often just up to the company gate). They help to monitor and optimize processes and to control product flow paths and inventories.

The information and communication systems are being used in a variety of forms. They support the employees of the operative area in all companies. The work steps, the required objects etc. are shown in a display and the operation is mostly confined to simple data entries (number of parts, barcodes etc.) in predefined fields.

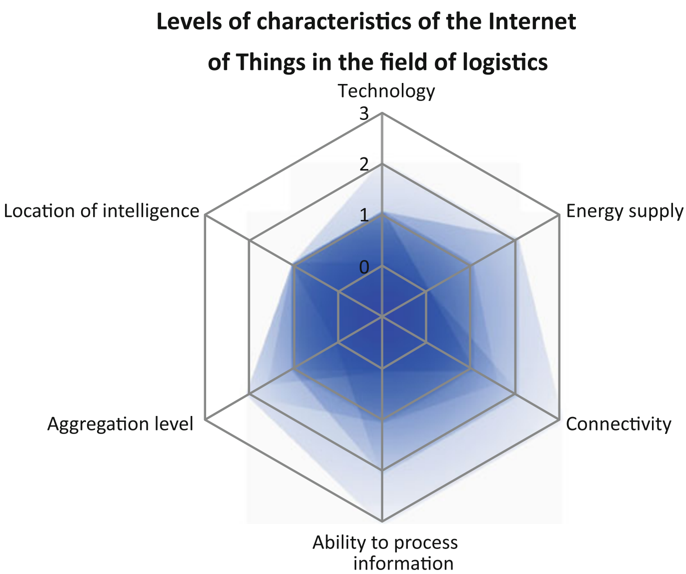


Fig. 1 Characteristics and diffusion levels of the "Internet of Things" in the surveyed logistics companies

In the field of foods distribution the application of this technology was highly divergent. Today, all companies use telematics systems which keep dispatchers and drivers in constant contact via a network. The dispatchers' computers are connected with the onboard Personal Digital Assistant (PDA) or notebooks in the vehicles. These onboard computers mostly have a variety of integrated functions. They are used for the organization of the tours and can also be applied for sending e-mails, as a mobile phone or as navigation systems. The complex tools (fleet management systems) which are applied by the dispatchers are also used for tour planning, organization and control and to determine the exact position of the transport vehicles and goods.

Figure 1 shows a radial diagram of the characteristics of the six case study companies with respect to their application of technology. It is obvious that hardly any company has reached the second or even the third level. As for the field of application of logistics it is underpinned that Internet of Things-technologies are so-far mainly used within a company ("Intranet of Goods"). The costs are again seen as the major obstacle for an investment in new systems. A negative cost-benefit-assessment, the limited readability of RFID-chips in metallic environments and liquids, as well as a lack of uniform standards hamper an investment in new technologies for the "Internet of Things".

It cannot be excluded that more characteristics will emerge in the future. The three levels reveal how far the diffusion can proceed if the development for the implementation follows the "vision" of the "Internet of Things". However, this does not mean that Level 3 stands for the "ideal level" for a logistics chain across several companies. This strongly depends on the conditions of the logistics process. On Level 3 and across all six criteria the process would proceed automated and with an autonomous decision-making. This level comes very close to the "Internet of Things" with an autonomous and self-controlled transport. It is, however, not necessarily desirable for every company. The consequences for Warehouse Management Systems or Transportation Management Systems cannot yet be predicted and this question could not be answered satisfactorily by experts.

As a summary it can be noted that although the technology of the "Internet of Things" has already been well developed and tested, the implementation is still in its infancy. Some of the reasons which were named by companies as obstacles for the implementation are: nonspecific costs, doubts with regard to data security, cooperation efforts for companies within a logistic chain, dynamics of distribution processes.

Consequences of the Implementation of the "Internet of Things" for Skilled Worker-Level

In spite of the so-far minor diffusion of the "Internet of Things" into logistic processes, some companies could be identified that are in fact witnessing sustainable changes in their workflow due to this technology. Also the employees' work tasks were subject to changes. These companies mainly focus on the automation of work cycles and the digitalization of information. A technology mix consisting of digitalized documents, the detection of geographic positions via telematics, special onboard terminals for status monitoring and an integrated flow of information via the Internet support these companies in delivering more efficient services. Two major developments could be identified:

1. On the one hand the technology was applied in order to automate processes and to reduce the error frequency during work processes. The work tasks and the respective requirement profile at the level of skilled work thus have often been simplified. This resulted in the fact that lower qualified personnel could be quickly deployed at lesser wages and without an extended on-the-job training.
2. On the other hand technologies for the "Internet of Things" were used to optimize work cycles. At the same time this has changed the fields of tasks for the employees. As the implementation freed up capacities of the employees, they could take over other tasks instead. This resulted in the fact that the fields of tasks became more comprehensive and above all more diverse. Such companies normally relied on very well trained skilled workers and reacted to the change of tasks with further qualification measures.

If the development aims at an automation of the processes, technology takes over decision-making processes and carries out technologically predefined work routines via program-controlled cycles. In this case the operative tasks of skilled workers are simplified with the exception of malfunctions and other problems within the system itself. The latter must still be solved by highly-qualified skilled personnel. The surveyed companies showed no uniform strategy for the shaping of work organization in order to cope with these contrarian tasks.

With respect to the development to optimize the work processes with the aid of the "Internet of Things", the companies always aimed at taking their employees' strong points into consideration. This resulted in the fact that decision-making processes were not left to the system but to human beings. As a rule the employees got supporting information on process optimization but were given the chance to co-shape the processes. The skilled workers must have a high process-specific knowledge in order to be able to intervene in case of problems or to make decisions. This development offers skilled workers the chance to make use of the technologies as assistance systems.

So far the "Internet of Things" has just been applied in partial processes and often only within a company. As soon as it comes to a comprehensive networking for the optimization of the flow of goods via the "Internet of Things", the challenges for the skilled workers have to be increasingly verified. This will be necessary in order to decide which role the skilled workers could play during the implementation, the optimization and the maintenance and which qualification profiles they should have.

Future Prospective on the Interaction between Man and Machine

The “Internet of Things” further increases the complexity of technical structures of information and communication technologies. Networked information objects which could be regarded as virtual machines, are linked to a network and developed further into web-services. As soon as new products are integrated, this results in a completely new dimension of machines in the Internet. The Internet becomes the “Internet of Things” (cf. Schmidt et al. 2008, p. 107f). With the aid of the “Internet of Things”, every object can be enhanced or equipped with digital information from the Internet and thus be guided and controlled. This process does not only create new economic values, it also increases the value of the information available in the Internet. Companies will finally depend on accessing these new value potentials and to master the resulting challenges (ibid. p. 109).

The development towards autonomously acting systems will turn the relationship of man and machine upside down. An implementation of this technology will have a considerable impact on the world of work. Two main target courses must be further investigated in the future:

- How should the relationship between man and ICT-systems be shaped so that human beings can adapt themselves to situations generated by ICT-systems?
- How should the balance of intelligence between man and machines be equilibrated?

The survey shows that the future flow of goods can be increasingly optimized above all with the aid of RFID-technologies and that processes and products can be better backtracked. There is still a long way to go towards the automation of entire processes or process steps. Here and there partly automated processes can be identified. Two major development strands for the future are already emerging:

- The development of expert systems as a tool for qualified skilled workers (tool scenario);
- Limitation of the autonomy of accomplished skilled workers by the emergence of advanced technology in plants and machines in the field of logistics (automation scenario).

Should the development proceed in the direction of the tool scenario and should human beings keep their freedom to shape, this technology could be employed as a kind of assistance system.

In the other direction the work of the “Internet of Things” is entirely controlled by technology. Skilled workers are not provided with information and do not have any responsibilities. The resulting competency gap entails that a shaping of work processes is only purposeful if guided by the “Internet of Things”.

One example for this development is the simplification of the planning activities for dispatchers in forwarding companies with the aid of technology that could result in the vehicle drivers taking over some selected tasks from the dispatchers.

Man and machine thus interact as authorized acting units that control each other reciprocally. The driver could carry out the dispatcher's tasks directly at the customer's or en route in his or her vehicle. The technology provides the driver with adequate aids on site (data forms, selection of driving jobs). He or she would not simply work off his or her jobs but could even intervene in the order sequence according to the particular situation. The driver could thus directly react to traffic imponderabilities or customer wishes and could shape his or her tour in an optimal way. At the same time the technology could offer support with the autonomous acquisition of new orders (cf. Ruth 2011).

These considerations are opposed by the statement that the "driver should above all be functional": "He or she should not hit any obstacle, he or she should be nice to customers, he or she should adhere to his or her driving schedule." (Statement of a company manager). Drivers should adhere to specifications given by the dispatchers or by respective technical systems. A critical reflection or discussion is not desirable. This would mean a development towards the automation scenario. The technology or the "Internet of Things" guides the skilled workers. The work of the "Internet of Things" is entirely controlled by technology. Skilled workers are not provided with information and do not have any responsibilities. The resulting competency gap entails that a shaping of work processes is only purposeful if guided by the "Internet of Things".

Consequently one of the central questions remains: How will the future profile of skilled work in logistics develop as soon as technologies such as the "Internet of Things" are entering the scene? The results from the case study do not yet show an unambiguous trend, even if a development towards a (semi)automation of the processes is above all favored by the research institutes.

Assuming that there will be a development towards more (semi)automation in the short- and medium term, the field of tasks of employees is likely to change. Routine tasks will decline and more (difficult) special cases will take their place. This phenomenon is derived from the so-called automation paradox. In an automated system, the employees only have to interfere in case of difficult malfunctions. These problems, however, often call for a higher qualification which the employees are no longer able to develop. This situation which is already reality in so-called high-tech-fields (e.g. airplane maintenance) is likely to spread out to all employment fields along with the possibilities of comprehensive networking.

References

Ashton K (2009) That 'Internet of Things' thing. RFID J http://www.rfidjournal.com/article/view/4986 (05.01.2012)

Brand L, Hülser H et al (2009) Internet der dinge - perspektiven für die logistik. Zukünftige Technologien Consulting Nr 80

Cetin S (2005) Geschichte der RFID-Technologie. OBJEKTspektrum, RFID/2005

Dworschak B, Zaiser H, Martinetz S, Windelband L (2011) Logistik als Anwendungsfeld des „Internets der Dinge"in der BMBF-Früherkennung. In: FreQueNz Zusammenfassung Internet der Dinge in der Logistik. Herausgegeben von FreQueNz, pp 2–4. http://www.frequenz.net/uploads/tx_freqprojerg/ Summary_logistik_final.pdf (23.02.2011)

Gaßner K, Bovenschulte M, Hrsg (2009) Internet der Dinge-Technologien im Anwendungsfeld RFID/Logistik. In: Das "Internet der Dinge". Die Informatisierung der Arbeitswelt und des Alltags -Erläuterung einer neuen Basistechnologie. Düsseldorf Nr. 176:15–29

IERC (2011) Webseite des IoT European research cluster. http://www.internet-of-things-research.eu (05.01.2012)

Jedermann R, Lang W (2008) Intelligenter container sorgt für frische Lebensmittel. Impulse aus der Forschung 1/2008. Universität Bremen

Meyer GG, Främling K et al (2009) Intelligent products: a survey. Comput Ind 60:137–148

Ruth K (2011) A culture-based view on innovation communities with some marginal notes on the possible future of open innovation. J Mach Eng 11:36–43

Schmidt BF, Kündig A, Bütschi D et al (2008) Digitalisierte Wirtschaft: Praktisch kein Stein bleibt auf dem anderen. Die Verselbstständigung des Computers, Zürich, pp 99–116

ten Hompel M (2005) Das internet der dinge: status, perspektiven, aspekte der aktuellen RFID-Entwicklung. Vortrag. Dortmunder Gespräche 2005: Fraunhofer Symposium RFID, Dortmund

ten Hompel M, Bullinger H-J (2007) Internet der Dinge. Springer, Verlag

Uckelmann D, Harrison M, Michahelles F (2011) An architectural approach towards the future internet of things. In: Uckelmann D, Harrison M, Michahelles F(Hrsg.) Architecting the internet of things. Springer, Berlin

Want R, Res I et al (2006) An introduction to RFID technology. IEEE Pervasive Comput 5:25–33

Weyer J (2007) Autonomie und Kontrolle. Arbeiten in hybriden Systemen am Beispiel der Luftfahrt. Technikfolgenabschätzung—Theorie und Praxis Nr. 2, 16. Jg., Dortmund

Windelband L, Spöttl G (2003) Research handbook—instruments for early recognitions of qualification need. Paper 1. biat-Reihe Nr. 21, Flensburg

Windelband L et al (2011) Zukünftige qualifikationsanforderungen durch das internet der Dinge. Zusammenfassung. FreQueNz. Bonn: bmbf 4–9

Sensor Reader Emulator for Integrating Sensor Networks into EPCglobal Network

Joo-Ho Lee, Oh-Heum Kwon and Ha-Joo Song

Abstract Sensor networks and RFID (Radio Frequency Identification) systems are two essential components of pervasive computing. In recent sensor applications, sensor networks and RFID systems usually co-exist and their data are interrelated to each other. In this paper, we propose a novel middleware system—SRE (Sensor Reader Emulator)—that emulates standard reader functionality so as to seamlessly integrate sensor networks into the RFID based EPCglobal Network. In SRE, every sensor nodes is assigned an EPC (Electronic Product Code) and its data is delivered in the form of a standard reader message. RFID applications in higher levels can access the sensor nodes as if they are sensor tags. Therefore applications using sensor networks and RFID systems can be developed more efficiently.

Keywords Sensor reader emulator · Sensor network · RFID · Electronic product code · Reader protocol

J.-H. Lee · O.-H. Kwon · H.-J. Song (✉)
Department of Information Technology Convergence and Application, Pukyong National University, Daeyon-dong Nam-gu Busan, South Korea
e-mail: hajoosong@pknu.ac.kr

J.-H. Lee
e-mail: noenemy5@nate.com

O.-H. Kwon
e-mail: ohkwn@pknu.ac.kr

H.-J. Kreowski et al. (eds.), *Dynamics in Logistics*, Lecture Notes in Logistics,
DOI: 10.1007/978-3-642-35966-8_47,

Introduction

Sensor networks and radio frequency identification (RFID) systems are the basic infrastructure of pervasive computing. A sensor network is a network of numerous sensor nodes (motes) that have sensing, data processing, and communication facilities (Finkenzeller 2010; Akyildiz 2002; Sarma 2004). Sensor nodes capture environmental data such as temperature, humidity and pressure and send them to the host via base station. RFID readers periodically scan tags in their coverage, collect the EPCs in the tags, and send them to the host. As the sensor based applications increase, it is expected that both sensor networks and RFID systems are mixed in sensor based applications (Zhang and Wang 2006; Sung et al. 2007).

A sensor tag is a special type of RFID tag that has sensing functionality powered by its own battery. Therefore a sensor tag can be considered as a small sensor node with very limited communication functionality and little computing power. Figure 1 shows the analogy between the sensor tags and sensor nodes. A sensor reader is a type of RFID reader that can read sensing values from sensor tags as well as EPCs in the tags as shown in Fig. 1a. A sensor reader scans all the tags in its field via RFID air protocols and then reports the EPCs and sensing values of the sensor tags in form of a reader message that has been standardized by EPCglobal Reader Protocol 1.1 (RP) and Lower Level Reader Protocol 1.0 (LLRP). Our approach is to enable a sensor

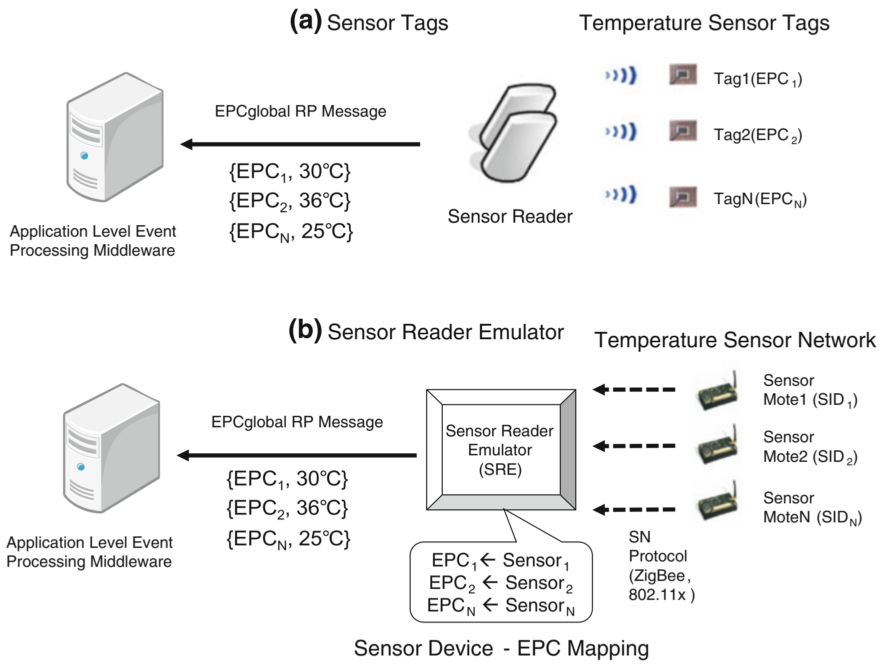


Fig. 1 Sensor reader emulator concept

node in a sensor network to be accessed just like a sensor tag as it is described in Fig. 1b. To this end, we propose Sensor Reader Emulator (SRE) that acts like a sensor reader. It gathers sensing values from the sensor nodes via the sensor network specific communication protocol such as ZigBee. It then reports the sensed data in form of RFID messages. In the proposed system, every sensor node is assigned a network wide unique identifier (sensor node identifier, SID), and the SID is mapped to a preconfigured EPC.

We have adopted EPCglobal RP as the interface for the proposed system, since it provides a better abstraction of an RFID reader with high level functionality than the EPCglobal LLRP. EPCglobal RP provides comprehensive communication messages for RFID readers. RP also give specification about the abstract internal structure of a reader that supports tag data filtering and event smoothing.

In following sections, we start a brief comparison of related researches, and then explain in detail about the proposed system. In section SRE Using MicaZ Sensor Network, we show an example application of the proposed system using MicaZ sensor nodes. Then we conclude in section Conclusion.

Related Works

Standard for Host and Reader Interface

EPCglobal Inc. is a standard organization for RFID technology (EPCglobal 2007). It suggests two types of reader communication protocol standards—RP and LLRP—between a reader and a host (EPCglobal 2006; Floerkemeier and Fleisch 2008; Floerkemeier et al. 2007). RP specifies the functionality, internal structure, and message formats of an RP compliant reader. RP assumes that a reader is composed of four subsystems as shown in Fig. 2. The read subsystem is an acquisition subsystem which gathers tag data using RF protocols. It can also perform tag selection by reading preferred tag patterns. Tag data collected by the read subsystem go through the second subsystem called event subsystem. The functionality of an Event Subsystem is to smooth the tag reads by eliminating abrupt losses of tag reads. Output subsystem manages a message buffer and selects only desired parts of tag data. Communication subsystem is in charge of sending messages to the host with the specified message format and transport binding (MTB). Currently RP has specification for text and XML message format. Read Subsystem and Output Subsystem are controlled by trigger objects called read triggers and notification triggers. RP also defines functionality and interface of essential objects (RP objects) that implement the subsystem. They are Sources, Read Points, Read triggers, Notification triggers, Tag Selectors, Data Selectors, and so on. An RP compliant reader has to implement these objects so as that it can properly operate regarding to the RP messages.

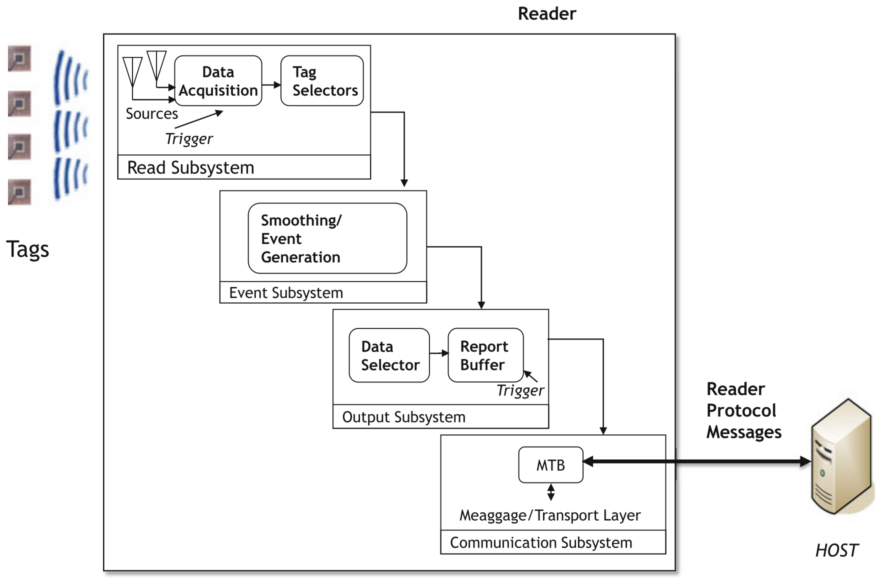


Fig. 2 Internal structure of an EPCglobal Reader Protocol 1.1 compliant reader (EPCglobal 2006)

Integration of Sensor Networks and RFID Systems

Sensor networks and RFID systems have common goals that both systems try to capture automatically what happens in the real world using wireless communication. The difference between them is that sensor networks focus on sensing while RFID systems are to identify objects and to automatically capture what happen in real world (EPCglobal 2007). RFID systems have been evolved to serve a globally inter-connected infrastructure called EPCglobal Network. However, sensor networks are configured for a specific application without sharing information between the applications. There have been several research works on integrating sensor networks and RFID systems (Zhang and Wang 2006; Sung et al. 2007; Choi et al. 2010; Uckelmann et al. 2011). Zhang and Wang (2006) proposed three types of system architectures that integrates sensor networks and RFID readers. The approach is sensor network oriented and does not consider integration of EPCglobal Network. Sung et al. (2007) and Uckelmann et al. (2011) suggested high level architecture of an RFID and sensor network integration in perspective of EPCglobal network, but it does not suggest any concrete method to implement the architecture. In Choi et al. (2010), we have proposed a system that enables a sensor device to act like an RFID tag so that it can naturally integrated into EPCglobal Network. Hence, in this paper we extended the system proposed in Choi et al. (2010) to support sensor networks.

Architecture for an EPC Enabled Sensor Network

Architecture of the Proposed System

Figure 3 shows architecture of the proposed system. SRE Server which is the container of SREs resides between an RFID middleware and sensor networks. This is the position where the RFID readers are placed in a typical RFID application. ALE (Application Level Event) (EPCglobal 2009; Hoag and Thompson 2006) middlewares are connected to SREs via EPCglobal reader protocols (Hoag and Thompson 2006). Currently only the XML/TCP MTB (message and transport binding) is supported. SREs can be managed by administration interface that can be accessed via a web-based or a java management extension (JMX) console. The core component of SRE Server is the SRE that emulates the functionality of a sensor reader. An SRE instance acts just like a sensor reader for a sensor network. SRE Manager deals with creating and deleting SRE instances while the SRE server is running. We use M-Let class loader for dynamic loading and downloading of SRE objects, since it contains classes that can be modified or added on the fly. SRE Manager and SRE instances are implemented as Java MBean (Managed Bean)

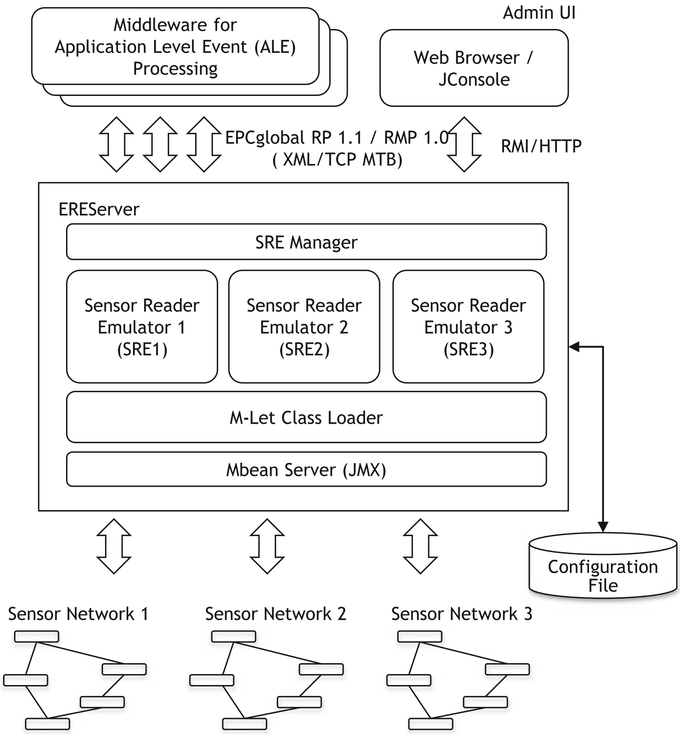


Fig. 3 SRE architecture

objects so that they can be managed by MBean Server of JMX, and provide flexibility for management interface. An SRE instance is allocated to a sensor network and connected to the ALE middleware. From the perspective of ALE middlewares, SREs are just treated as regular RFID readers that can be connected via EPCglobal reader protocols. Configuration file is an XML document that contains a configuration for an SRE Server.

Since the interface to sensor nodes are quite difference from one another and there is no standard interface to access the data from and to the sensor networks, we introduce an adaptor approach. Figure 4 shows the internal structure of an SRE. An SRE consists of two modules, RP module and adapter module. The RP module implements objects that confirm to the EPCglobal reader protocol specification. We do not explain further about the objects in RP module as they are developed according to the EPCglobal reader protocol specification except when it is needed for the better understanding of the proposed system. Sensor adapter provides a gateway interface to and from the base node. The main functionality of

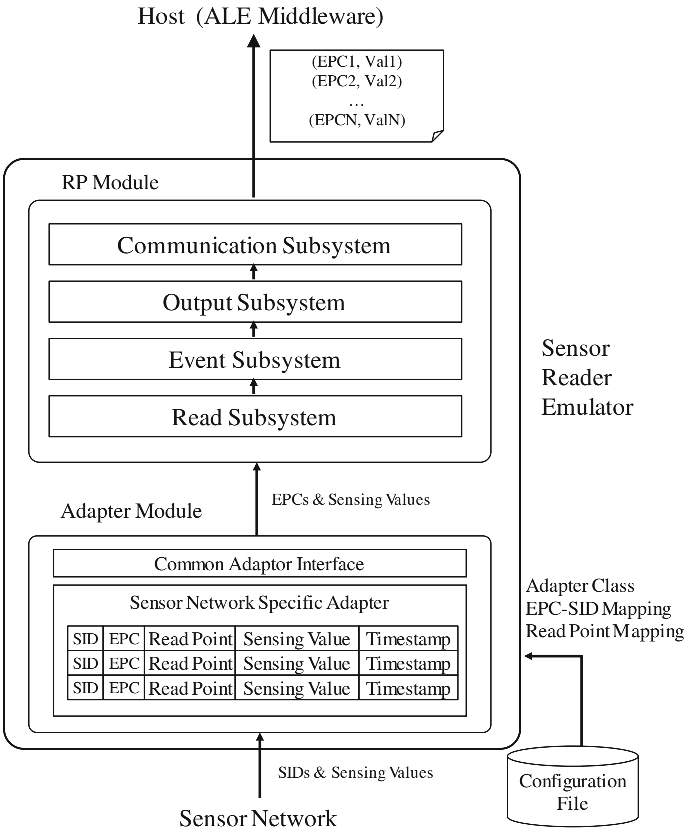


Fig. 4 Internal sturcture of an SRE

a sensor adaptor is to receive data from sensor nodes with their identifiers in sensor network domain, and to deliver the data with the associated EPCs to the RP subsystem. RP subsystem that implements the four subsystems described in RP specification performs event generation and delivers the events to the host in form of RP messages. An adaptor is expected to implement the common adapter interface of which the functionality is similar to the hardware abstraction layer (HAL) in operating systems. Followings are some important methods that adapters have to implement. These methods are defined as abstract methods of the SRE adapter class from which all adapter classes should be derived.

- Initialize: initializes adapter and connection to the base station and device driver. This method also prepares internal data structures to keep EPCs for each sensor nodes and their sensing values. The configuration data are delivered as a map object.
- Finalize: closes connection to sensor network and releases all the resources.
- Activate: starts reading sensing values from the sensor network. This method is called after initialization has been successfully done.
- Deactivate: stop reading sensing values from the sensor network. This method is called before finalization.

To create an adapter, following three configuration data are needed.

- Connection configuration: This is the information needed for the adaptor to connect the base station. Typical examples can be IP address and port number or serial communication port number to the base station. The configuration data can include parameters to change the operation mode of the sensor network.
- Array of (SID, EPC, Read Point name) pairs: The mapping information for each sensor node in the sensor network.
- Refresh cycle: Time interval at which sensing values should be read from the sensor network.

All the sensing values are maintained as a Java string objects in the adapter because currently SRE does not perform any operation on sensing values. Sensor networks are usually connected by serial cable such as RS 232C. Therefore pure java adapters can be implemented on platforms where third party java library for serial ports are provided. We adopt GNU's rxtxSerial library to connect SRE with MicaZ sensor network. For a sensor network that can be accessed only by proprietary device driver and that do not provide a java library, adapters can be implemented using JNI (java native interface) in which a platform dependent C or C++ interface is applicable. In Choi et al. (2010), we have tested SRE with several legacy sensor devices (not a sensor network) using this type of adapters, and found them work properly.

Assigning Read Points

In RP specification, a Read Point is a reader component from which tags or barcodes are captured. An RFID antenna or a barcode scanner is a typical example. A Read Point usually gives a hint for the physical location of tags being read. Therefore in SRE server the administrator can assign Read Point names to the sensor nodes along with their EPCs. The assignment can be done for individual sensor nodes or for the sensor adapter. In latter case, all the sensor nodes in a sensor network should have the same Read Point. If a sensor network has only one Read Point and many sensor nodes, we can assign more specific Read Point names for the sensor nodes. Specific a Read Point name overrides the sensor network name. As it is same with regular RFID readers, an SRE administrator has to carefully select Read Point names so that they give appropriate meaning to the application layers.

Data Gathering Sequence

According to the EPCglobal reader protocol, tag scanning is initiated by the 'Read Triggers'. A Read Trigger is periodically wakes up the source object that scan tags in its coverage using its tag data acquisition system. Figure 5 shows the sequence in which sensing values are delivered from base station to the source object in RP module. Sensing values are periodically gathered from the sensor network and kept in the adapter. Since it takes long time to gather sensing data from the sensor nodes, we have made the sensing values cached in the adapter for the fast access by the RP module. If fast access is not important, the adapter can be configured to gather sensing values only when it is requested. In this case the refresh cycle

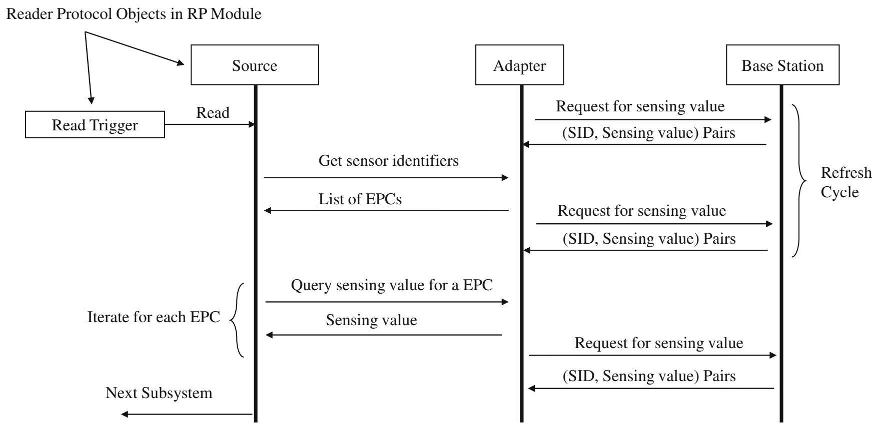


Fig. 5 Sequence of delivering sensor values

should be set to ‘0’. Similar operation can be also archived by setting the Read Trigger to ‘continuous mode’ which means reporting tag data whenever they are available as it is specified in EPCglobal reader protocol. In continuous mode, the latest sensing values are reported to the RP module as soon as they are gathered by the adapter.

Data Gathered and Generated

EPC RP specifies a variety of data that may be delivered from the tag when a tag data is scanned by a reader. The data are represented as ‘tag fields’. Table 1 shows how tag fields of EPCglobal RP are supported by SRE.

In an RFID readers, some tag fields are assigned when the reader has been manufactured. Those tag fields are ‘readerEPC’ and vender fields. Since SRE is an emulator, it should be given when a new SRE instance is created by the administration tool. Other tag fields like ‘readerName’, ‘sourceName’ can be controlled

Table 1 Tag data fields supported in SRE

Tag data fields used in SRE

Field name	Description (RP)	Description (SRE)
EventTriggers	Name of the read trigger that initiated the read	Same as RP
EventType	Name of event	Same as RP
EventTimeTick	Time at which event is generated	Same as RP
EventTimeUTC	Time at which event is generated (UTC)	Same as RP
ReaderEPC	Reader’s EPC	Configured for each SRE instance
ReaderHandle	Name of reader handle	Configured for each SRE instance
ReaderName	Name of the reader	Configured for each SRE instance
ReaderRole	Reader’s role name	Configured for each SRE instance
ReaderNowTick	Time at notification in clock tick	Same as RP
ReaderNowUTC	Time at notification in UTC	Same as RP
TagType	Tag type	‘Gen2′
TagID	Tag ID in binary format	Configured tag id for the sensor node
TagIDasPureURI	Tag ID in Pure Identifier	Configured tag id for a sensor node in PureURI form
TagIDasTagURI	Tag ID in Tag URI form	Configured tag id for a sensor node in TagURI form
SourceName	Name of the source	Same as RP
SourceFrequency	Read frequency	Read Cycle being used in SRE
SourceProtocol	Name of the air protocol used	‘SRE’
NotifyChannelName	Name of notification channel	Same as RP
NotifyTriggerName	Name of the notification Trigger	Same as RP
<Vendername>:<Any>	Any field vender created	Configured when SRE is created

by the RP messages as they are specified in EPC RP. The vender field is an extension tag field for the reader manufacturer's use. SRE use this field for the delivery of sensing values. According to the RP standard, a vender field name should have vender name prefix with a ':' character as a delimiter. SRE use 'lit' as the vender name prefix. Therefore if the administer has given a field name of 'temp' while creating an SRE instance, actual field name found in the RP message is defined as 'lit:temp'. The tag fields related to the time is either originated from the sensor node or generated by the event subsystem in RP module in SRE.

Event Generation

Tag events are generated by the event subsystem in RP module that operates in accordance with the specification in EPCglobal RP. Event subsystem supports event smoothing by removing redundant reporting of the same tags and temporary loss of tags. In SRE, to read sensing values from sensor nodes is same as to read sensing values from sensor tags. Therefore, though sensing values are not read from some sensor nodes for a short period of time, Event Subsystem will continue to report the previous values as long as the failure duration does not exceed a certain period of time. We expect that this kind of smoothing operation will not occur so frequently as it does with RFID tags, since sensor networks use more reliable communication protocols such as ZigBee than an RFID reader does.

SRE Using MicaZ Sensor Network

In this chapter we show an example application of SRE using MicaZ sensor network that operates Tiny OS. In our experiment, we used six sensor nodes, one for a base station and the others for sensing temperature (Fig. 6). Communication between the sensor nodes are done via ZigBee protocol that has been configured in star topology. The link layer uses 2.4 GHz 802.15.4 protocol. The base station is connected to the host computer via serial cable (USB). The cable is used both for installation of programs to sensor motes and for the data gathering. Among the several sensor modules that are available on the MicaZ board, we have adopted the temperature sensor module for test implementation. The sensor motes are all configured to report observed temperature values to the base station every 100 ms. This report cycle has to be set carefully with regard to the 'read cycle' of the source object in the RP module. If the reporting cycle is too short compared with the read cycle of the source object, it would spend too much energy on getting and transmitting data that will be simply invalidated, since the new data will replace the old before they are delivered to the source object. Therefore we set the read cycle of the source object close to the sensor network's report cycle. The base station is configured to send the received date to the host as soon as it receives

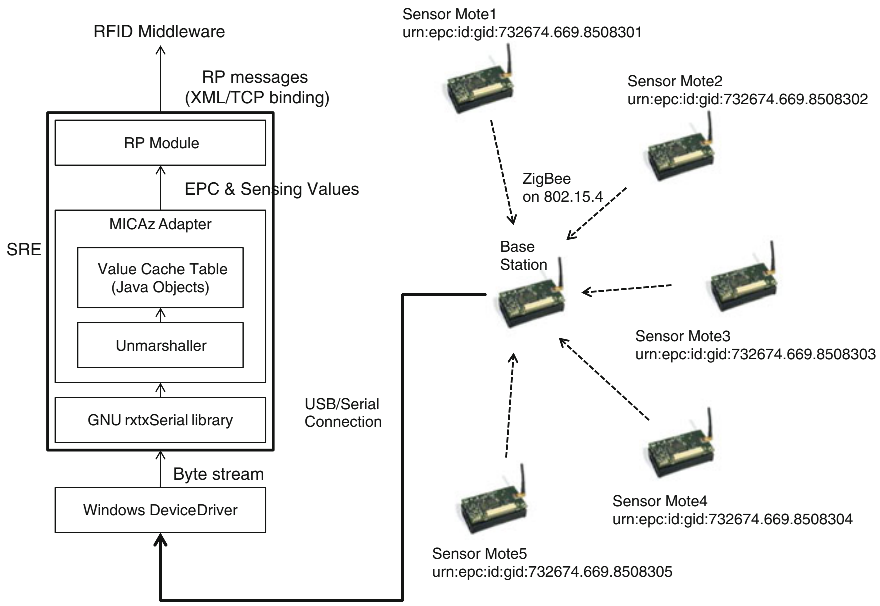


Fig. 6 SRE implementation using MicaZ sensor nodes

them from the motes. We use the GNU rxtxSerial class library to receive data sent from the base station. Data frames are composed of 8 bit octets and parity checking is not applied. MicaZ adapter parses the data frames, identifies the owner of the data using the sensor node identifier and converts byte stream into Java string objects. As the ATmega128 microprocessor in MicaZ sensor motes represents data in little-endian form, the values have to be transformed into double numbers and then into string objects. This can be regarded as an un-marshalling step that has to be done in a sensor network specific way. The sensing values in the cache table are all kept as Java string objects since there is no data processing steps involved at the adapter level and the values are converted into characters before they are submitted in either text or XML messages to the host.

Figure 7 shows an example RP message generated by SRE. According to the RP specification reader messages can be delivered in two formats. One is text and the other is XML. This is called message and transport binding (MTB). SRE supports all possible combinations of MTBs such as XML/TCP, Text/TCP, XML/HTTP, and Text/HTTP. The specific MTB that a middleware and SRE will use is determined during the handshake step. Handshake is a part of initial connection process in which a middleware and a RP compliant reader negotiate with each other for later communication. In Fig. 7, we present only one tag data since others are all the same except that tag's EPC and the sensing value are different. This example is a 'Notification' message that is sent from readers to the middleware when any tags are found that meets the filtering condition in the field.

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<notification xmlns=”urn:epcglobal:xsd:1” xmlns:lit=”http://www.rclit.com”>
  <id>1</id>
  <reader>
    <readerEPC>3500000000AABBCCDDEEFF5</readerEPC>
    <readerName>SRE</readerName>
    <readerRole/>
  </reader>
  <notifyTriggerName/>
  <notifyChangeName/>
  <readReport>
    <sourceReport>
      <sourceInfo>
        <sourceName>source0</sourceName>
      </sourceInfo>
      <tag>
        <tagID>3500B2E0200029D00081D39F</tagID>
        <tagIDAsPureURI>urn:epc:id:gid:732674.669.8508319</tagIDAsPureURI>
       <tagIDAsTagURI>urn:epc:tag:gid-96:732674.669.8508319</tagIDAsTagURI>
        <tagEvent>
          <eventType>evObservered</eventType>
          <time>
            <eventTimeUTC>2011-12-23T05:04:06.328KST</eventTimeUTC>
          </time>
        </tagEvent>
        <other>
          <lit:temp>300.35999328643084</lit:temp>
          <lit:readPoint>ANT2</lit:readPoint>
        </other>
      </tag>
     …

    </sourceReport>
  </readReport>
</notification>
```


Fig. 7 An RP message generated from SRE with MicaZ sensor network

Conclusion

Sensor applications that include both sensor networks and RFID systems are getting widely used. To provide an integrated view of sensors from both technologies, we suggest a sensor reader emulator (SRE) that provides a sensor tag abstraction for the sensor nodes. Once the sensor nodes are given their own EPCs, they can be accessed just like an ordinary sensor tags. Therefore development of sensor application can be archived more efficiently. We verified the feasibility of the proposed approach by developing an application that uses MicaZ sensor nodes. We are extending SRE to provide a host interface with lower level reader protocol which is another standard reader interface from EPCglobal Inc.

Acknowledgments This work was supported by the Grant of the Korean Ministry of Education, Science and Technology (The Regional Core Research Program/Institute of Logistics Information Technology).

References

Akyildiz I (2002) Wireless sensor networks: a survey. Comput Netw 38(4):393–422

Choi S-H, Kim T-Y, Kwon O-H, Song H-J (2010) Sensor reader emulation system for accessing sensor device through EPCglobal reader protocol. J KIISE: Comput Pract Lett 6(8):825–887

EPCglobal (2006) Reader protocol standard version 1.1

EPCglobal (2007) The EPCglobal architecture framework

EPCglobal (2009) The application level events (ALE) specification, Part I: core specification

Finkenzeller K (2010) RFID handbook: fundamentals and applications in contactless smart cards, radio frequency identification and near-field communication, 3rd edn. Wiley, New York

Floerkemeier C, Fleisch E (2008) RFID applications: interfacing with readers. Softw IEEE 25(3):67–70

Floerkemeier C, Lampe M, Roduner C (2007) Facilitating RFID development with the Ac-cada prototyping platform. Fifth annual IEEE international conference on pervasive computing and communications workshops (PerComW'07), pp 495–500

Hoag JE, Thompson CW (2006) Architecting RFID middleware. Internet Comput IEEE 10(5):82–92

Sarma S (2004) Integrating RFID. Queue 2(7):50–57

Sung J, Sanchez TL, Kim D (2007) EPC sensor network for RFID and USN integrated infrastructure. In: Fifth annual IEEE international conference on pervasive computing and communications

Uckelmann D, Harrison M, Michahelles F (2011) Architecting the internet of things. Springer, Berlin, pp 1–24

Zhang L, Wang Z (2006) Integration of RFID into wireless sensor networks: architectures, opportunities and challenging problems. 2006 Fifth international conference on grid and cooperative computing workshops, pp 463–469

A Design of Distributed Discovery Service of EPCglobal

Gihong Kim, Bonghee Hong and Joonho Kwon

Abstract For global tracing of RFID tagged object, the Discovery Service (DS) is one of the core services of EPCglobal Network Architecture. A DS provides a means to locate all the list of EPCIS addresses that an EPC object has visited before. The EPCglobal has not yet specified the standard for the DS. So this paper proposes an object based distribution of a DS for standard. An Object Based Discovery Service (OBDS) is composed of several unit-DSs and a UDLS. A unit-DS is in charge of some objects and a UDLS enables to discover the right unit-DS among distributed unit-DSs for global tracing of an object. We analyze the performance of several approaches and ODBS.

Introduction

RFID enables automatic identification, and moreover, global tracing of tagged object in supply chain environment (Roy 2006). The EPCglobal (2011) is the de facto international standard for RFID, but there is only introduction but no

G. Kim · B. Hong (✉)
Department of Computer Engineering,
Pusan National University, Busan, Republic of Korea
e-mail: bhhong@pusan.ac.kr

G. Kim
e-mail: buglist@pusan.ac.kr

J. Kwon
Institute of Logistics Information Technology,
Pusan National University, Busan, Republic of Korea
e-mail: jhkwon@pusan.ac.kr

H.-J. Kreowski et al. (eds.), *Dynamics in Logistics*, Lecture Notes in Logistics, 571
DOI: 10.1007/978-3-642-35966-8_48, © Springer-Verlag Berlin Heidelberg 2013

concrete standard for a Discovery Service (DS) (EPCglobal Inc 2011) that provides a means of global tracing of tagged objects.

In this paper, we show some naive approaches for Discovery Service. Firstly in the centralized approach, the data is managed by only one Discovery Service, so the main advantage of centralized DS is their simplicity. But the scale is often limited by the capacity of the server. Centralized DS architecture suffers from serious performance and scalability problems. Next is hierarchical approach dividing DS to several local DSs and root DS that is region based distribution of DS. The local DS covers the some region regardless of kind of object. Root DS in this approach is bottleneck of entire architecture and there are problem of coupling of data. So we propose an object based distribution of DS. The design concept of this architecture is that unit-DS is not binding with region but object itself. We solve the problem of coupling of data and this approach outperforms other approaches.

In the section Related Works, we show some related works, and in section Design of Discovery Service, first we analyze some naive approaches for Discovery Service and propose a new distributed architecture for Discovery Service. In section Analytic Comparison, we show analytic comparison of several approaches for Discovery. Section Conclusions and Future Works is conclusion and future works.

Related Works

EPCglobal Network Architecture

Parities in a supply chain deliver a physical object such as a trade product identified with Electronic Product Codes (EPCs) (EPCglobal Inc 2010) to another one. Here, physical object exchange consists of such operations as shipping, receiving, and so on. The EPCglobal architecture framework (EPCglobal Inc 2009; Verisign 2005) defines EPC physical object exchange standards to make sure that if one company delivers a physical object to another company, the latter can identify the EPC of the physical object and interpret it properly.

The EPCglobal architecture framework also defines interface standards for end users to share EPC data. When exchanging data with each other, end users are the primary beneficiaries of the EPCglobal architecture framework.

EPC Information Service (EPCIS) (EPCglobal Inc 2007) consists of both interfaces (EPCIS Query Specification) and data itself (EPCIS Event Specification) for data exchange. EPCIS data is information that trading partners share to gain more insight into what is happening to physical objects in locations outside their own firewalls. EPCIS is the primary entry point for data exchange between companies or parties who are the EPCglobal subscribers.

Object Name Service

Object Name Service (ONS) (EPCglobal Inc 2008) can be a simple lookup service that takes an EPC as input, and produces the address of an EPCIS which the item's EPC begins with (such as the manufacturer EPCIS) as output. ONS does not contain actual data about the EPC. It only contains the network address of services that contain the actual data. ONS uses the internet's existing Domain Name System (DNS) for looking up (resolving) information about an EPC. DNS can be a single global lookup service conceptually but is implemented as a hierarchy of lookup services; what's more, DNS is a hierarchically organized system that roughly follows the hierarchy found in the domain-name. Thus, the item's EPC must be converted into a format that DNS can understand to use DNS to find information about an item.

ONS works as follows. When an ONS Client application wants to know about a product, and then use ONS to locate the EPCIS service, the EPC should be converted into a domain-name first. Then the application queries the DNS for NAPTR records for that domain name and receives several records, which each of them conforms to the specified rules. The DNS resolver is responsible for executing the query procedure. At last, the application uses one or more of the records returned to locate an appropriate service depending on the service that the ONS Client desires.

Discovery Service

Discovery Service (DS) provides means to locate list of EPCIS addresses that EPC object has visited in the most general situations arising from multi-party supply chains, in which several different organizations may have relevant data about an EPC but the identities of those organizations are not known in advance.

ONS and DS coexist and serve different roles in the EPCglobal architecture. As described above, the Object Name Service (ONS) is a lookup service useful to find the address of the manufacturer EPCIS. However, a Discovery Service returns EPCIS addresses that object has visited. Unlike ONS, generally a Discovery Service contains pointers to entities other than the entity that originally assigned the EPC code.

Discovery Service should provide a means to allow parties to mutually identify and authenticate each other and to share information necessary for authorizing access to EPCIS service listings and EPCIS data. Currently, any standard is not specified for Discovery Service such as storage, exchange, access authorization, or reporting of EPC observation data provided by EPCIS. Some paper (Kurschner et al. 2008; Germany 2005) provides a high-level description of the EPCglobal DS architecture.

Design of Discovery Service

Naive Approaches: Region Based Distribution

This section shows several naive approaches to design Discovery Service. First approach is centralized approach. The data is managed by only one Discovery Service in the centralized system. The main advantage of centralized DS is their simplicity. Because all data is centralized in one DS, this DS is easily managed and has no doubt of data consistency. But, as the number of data and users grows, the scale is often limited by the capacity of the server. Centralized DS architecture suffers from serious performance and scalability problems (Vanalstyne et al. 1995).

Next is hierarchical distributed model of Discovery Service. In this model, a root-DS is divided by local DSs and a local DS is in charge of several EPC Information Service (EPCIS). Main factor of this distribution is that the local DS is connected to the some EPCISs according to its regional relationship. So if any observation of objects is occurred in the some EPCIS, the observation is reported to its connected local DS that is parent node of EPCIS. Detailed data insert and search examples are described in Figs. 1, 2.

There are two approaches for this hierarchical distributed DS whether the root-DS stores or not. According to these approaches, it is similar algorithm for insert and search but there are quite differences for performance.

We abbreviate the method of not storing the data in the root-DS because the algorithm is very similar to the method of storing the data in the root-DS. The Fig. 1 shows insert example. The trajectory of object "a" is A → B → C → D and "b" is F → D → C → A. The observation of an object is reported to each local DS. For example, as object "a" is observed in the EPCIS A, the observed

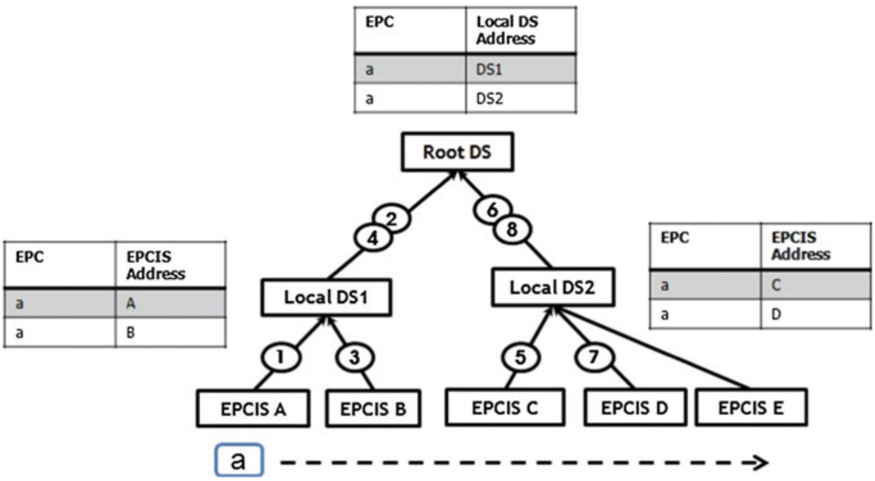


Fig. 1 Insert example

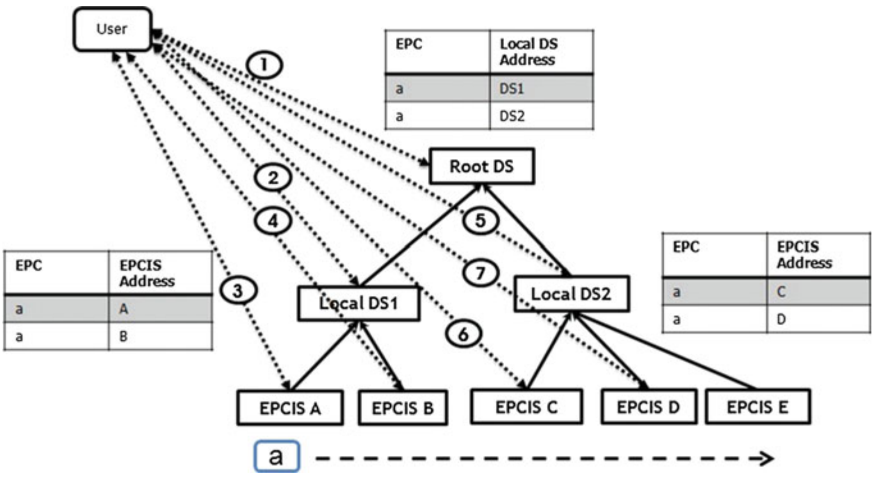


Fig. 2 Search example

information is reported to the local DS1 and also reported to the root-DS. A root-DS should stores all the observation information of all the objects, as you can see in the Fig. 1.

Figure 2 shows search example of the first approach. Users access the DS by accessing application. Firstly accessing application should access the root-DS, because all the child information is stored in the root-DS. If user wants global tracing information of object “a”, the root-DS is firstly accessed, and indirectly address the next level node (local DS) where user can retrieve the next point information. The answer of the root-DS is local DS1 and DS2. Next step is user queries to local DS1 and 2. Also the local DS has indirect address (EPCIS) where the real observed information is. The EPCIS has real observed information of objects. In the Fig. 2, user can retrieve the answer that observation information of object “a” is in the EPCIS A and B by the local DS1. Similarly the information of object “a” is in the EPCIS C and D by the local DS2. Finally user can retrieve the observation information of object “a” in the EPCIS. If user wants all the tracing information of object “a”, user should query to EPCIS A, B, C and D in sequential order.

Solution: Object Based Distribution

In the previous section, we examine the naive approaches such as local DS is in charge of some EPCISs. That is local DS is assigned to the fixed EPCISs on the basis of locations. Like the Fig. 1, if any object is observed in EPCIS A, the observed information should be reported to only to its bound local DS1. For example Korean company EPCISs are binding to only one Korean local DS. At first glance it seems to be right binding, but it is not an appropriate choice. In this chapter, we'll explain why this is not suitable for the Discovery Service architecture.

Design Concept

In this section, we propose an Object Based distributed architecture for Discovery Service. The key design concept is decoupling of observed information. This architecture proposes several unit-DSs which are independent of each other, because all the observed data of one object are stored in only one unit-DS. Therefore user does not need to search on the several unit-DSs. The design concept of this architecture is that unit-DS is not binding with EPCIS but object. So EPCIS can report the data to any unit-DSs, but data of one object is stored on only one unit-DS. Additional component is the Unit-DS Lookup Service (UDLS) that finds appropriate unit-DS address where the global tracing information of object what user wants is stored. The UDLS answers to user on which unit-DS can I find global tracing information of this object.

Components

- **Unit-DS**: This component stores whole global tracing information of some objects, so there is no dependency between each unit-DS. If user wants to find global tracing information of some object, user accesses only one unit-DS. The information of unit-DS is mapping table that mapping object to list of observed EPCISs. If an object is observed on any EPCIS, the pair of object and observed EPCIS address is inserted to right unit-DS in the inserting step. So later user can query to unit-DS where I can find the global tracing information of this object, and unit-DS returns the visited EPCIS address list. Table 1 shows the methods of unit-DS.

Table 1 Methods of unit-DS

Method name	Return value	Parameter
GetEPCISAddress (return list of EPCIS addresses that stores global tracing information of object that user wants	List of URL (list of EPCIS address)	URI of EPC (what object user want to trace)
RegisterEPCObservation (register the EPCIS address where the object is observed)	Void	URI of EPC (observed object) URL (what EPCIS observed the object)

Table 2 Method of unit-DS lookup service (*UDLS*)

Method name	Return value	Parameter
GetUnitDSAddress (return unit-DS address that stores global tracing information of object that user wants)	URL (address of unit-DS)	URI of EPC (what object user want to trace)z
RegisterUnitDSAddress (register the unit-DS address where user can find the tracing information of an object)	Void	URI of EPC (object code) URL (unit-DS address)

- **Unit-DS Lookup Service (UDLS)**: The Unit-DS Lookup Service (UDLS) is newly proposed. In the previous naive approaches there is an entry point that is a root-DS, but in this architecture there is no entry point. User can find a right unit-DS through the UDLS. If user already knows what right unit-DS of that object is, the UDLS is unnecessary. But the UDLS is used by general users who do not know appropriate unit-DS address. The information of the UDLS is mapping table that mapping object into unit-DS address, so user can find appropriate unit-DS address through the UDLS. This information is very static and there is very little update. So at first time if user downloads the mapping table of a UDLS, there is no more access to a UDLS except update of UDLS information. Table 2 shows method of a UDLS.

Insert and Search

In this section, insert and search examples are introduced. In the Fig. 3 the object "a" is observed in the EPCIS A, and EPCIS A asks where the right unit-DS address for object "a" is. The answer is unit-DS 1 for the object "a", and then EPCIS A stores the information to the unit-DS 1. Likewise, the object "b" is stored in unit-DS 3.

The UDLS information is mapping table of object to appropriate address of unit-DS. That information is pre-registered by system administrator. Update of UDLS information is very rare unless the unit-DS IP is changed. So this architecture has room for improvement. If each EPCIS downloads the UDLS information only at the first time, it does not need access to the UDLS every time. You can see the example in the Fig. 3.

Next is search example of Object Based Discovery Service. If user wants to trace object "a", the user firstly accesses to the UDLS, because the information of next point is stored in the UDLS. Unit-DS 1 is the next access point to search the tracing information. In the unit-DS 1, all the tracing information is stored in one database, so user needs to access only one unit-DS unlike naive approaches. At last user can retrieve the observation information of object "a" in the each EPCIS. Similarly search mechanism can be improved as shown in Fig. 4. Users may download the UDLS-data, so that there is no need for access to the online UDLS.

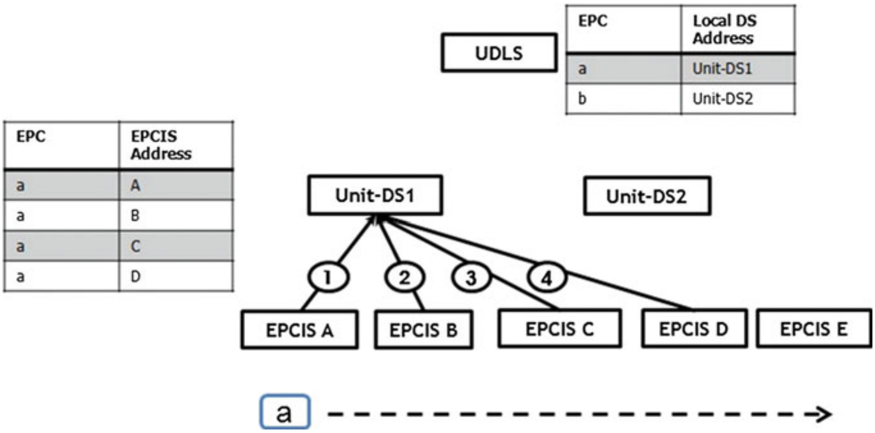


Fig. 3 Insert example of Object Based Discovery Service

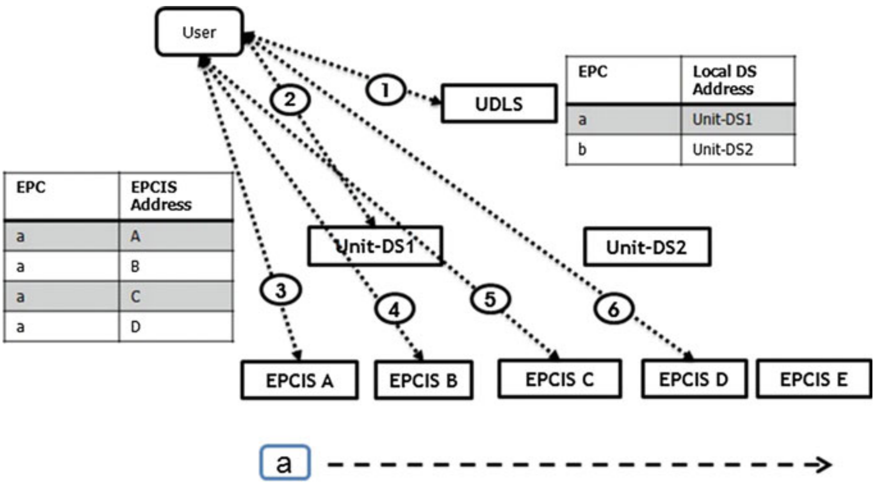


Fig. 4 Search example of Object Based Discovery Service

Later a synchronization mechanism should be proposed when the UDLS information is updated.

Analytic Comparisons

This section shows not experimental but analytic comparison of naive approaches and new solutions in the Table 3. As you see in the previous section, the proposed solution (ODBS 1) has room for improvement that is each EPCIS downloads the

Table 3 Analytic Comparison of Insert/Search/Data Cost

	NAIVE 1	NAIVE 2	OBDS 1	OBDS 2
Distribution	*Tree distribution* • Coupling of data	*Tree distribution* • Coupling of data	*Horizontal distribution* • Decoupling of data	*Horizontal distribution* • Decoupling of data
Root-DS/ UDLS	*Root-DS* • Storing the data in the root-DS	*Root-DS* • Not storing the data in the root-DS	*UDLS* • User does not download the UDLS info.	*UDLS* • Only at first time user downloads the UDLS info.
Insert cost	*High* • One insert to the root-DS	*Low* • No access to the root-DS	*Medium* • One search to the UDLS	*Low* • No access to the UDLS
Search cost	*High* • One search to a root-DS • Search to several local DSs	*Very high* • One search to a root-DS • Search for whole local DSs • Search to several local DSs	*Medium* • One search to UDLS • One search to unit-DS	*Low* • No search to UDLS • One search to unit-DS
Data space (root-DS/ UDLS)	*Very high* • Serial level data of EPC	*Low* • No data in the root-DS	*Medium* • Product level data of EPC	*Medium* • Product level data of EPC

UDLS information only at the first time, therefore it does not need access to the UDLS every time (ODBS 2).

As you can see in the Table 3, in the NAIVE1, accessing to a root-DS on every operation is bottleneck of the entire architecture not only in insert but also in search. The NAIVE 2 is very good for insert but very poor in search because of additional search for whole local DSs all over the world. On the other hand, ODBS 1 has moderate performance both in insert and search. Furthermore, performance of ODBS 2 is improved by downloading the UDLS information in advance.

Conclusions and Future Works

The Discovery Service is the core service of EPCglobal Network Architecture for global tracing of RFID tagged object. The standard of the Discovery Service is not yet specified by the EPCglobal. The DS should have high scalability because the size of data is very huge and users grow rapidly.

In this paper, we proposed a scalable architecture for Discovery Service. We established the entire architecture for Discovery Service that is Object Based

Discovery Service (OBDS). We solved the problem of data coupling and we showed the OBDS outperformed the naive approaches in the insert and search by the analytic comparison in the section Analytic Comparisons.

There are some future works. If the information of a UDLS is updated, we should propose a synchronization mechanism of a UDLS information and user-downloaded information. And next is advanced solution of a UDLS. Because the data of UDLS is composed of two parts, we could propose a dedicated two dimensional index structure.

Acknowledgments This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology(2012014046).

References

EPCglobal Inc (2007) EPC information services version 1.0.1 http://www.epcglobalinc.org/standards/epcis/epcis_1_0_1-standard-20070921.pdf

EPCglobal Inc (2008) ONS standard v. 1.0 http://www.epcglobalinc.org/standards/ons/ons_1_0_1-standard-20080529.pdf

EPCglobal Inc (2009) The EPCglobal architecture framework version 1.3 http://www.gs1.org/gsmp/kc/epcglobal/architecture/architecture_1_3-framework-20090319.pdf

EPCglobal Inc (2010) EPC tag data standard version 1.5 http://www.epcglobalinc.org/standards/tds/tds_1_5-standard-20100818.pdf

EPCglobal Inc (2011) Discovery service working group discovery service version 1.0 http://www.epcglobalinc.org/apps/org/workgroup/subscriber/download.php/15448/SAG_Discovery_Services_Opt-In_Charter_Final.doc

EPCglobal Standard (2011) http://www.epcglobalinc.org/. Accessed 15 Nov 2011

Germany G (2005) Internet der Dinge. Management Information. Das EPCglobal Netzwerk. Tech Rep

Kurschner C, Condea C, Kasten O, Thiesse F (2008) Discovery service design in the EPCglobal network towards full supply chain visibility. Internet Things 4952:19–34

Roy W (2006) An introduction to RFID technology. IEEE Pervasive Comput 5(1):25–33

Vanalstyne M, Brynjolfsson E, Madnick S (1995) Why not one big database? Principles for data ownership. Decis Support Syst 15(4):267–284

Verisign (2005) The EPC network: enhancing the supply chain http://www.verisign.com/static/002109.pdf