Hans-Dietrich Haasis
Hans-Jörg Kreowski
Bernd Scholz-Reiter
Editors

Dynamics in Logistics

First International Conference, LDIC 2007
Bremen, Germany, August 2007
Proceedings

Springer

Dynamics in Logistics

Hans-Dietrich Haasis · Hans-Jörg Kreowski
Bernd Scholz-Reiter

Editors

Dynamics in Logistics

First International Conference, LDIC 2007
Bremen, Germany, August 2007
Proceedings

Springer

Prof. Dr. Hans-Dietrich Haasis
ISL – Institute of Shipping Economics and Logistics
Universitätsallee GW 1A
28359 Bremen
Germany
haasis@isl.org

Prof. Dr. Hans-Jörg Kreowski
University of Bremen
Department of Computer Science
Theoretical Computer Science
Linzer Straße 9a
28359 Bremen
Germany
kreo@informatik.uni-bremen.de

Prof. Dr. Bernd Scholz-Reiter
Universität Bremen
BIBA – Bremen Institute for Production and Logistics
Hochschulring 20
28359 Bremen
Germany
bsr@biba.uni-bremen.de

ISBN 978-3-540-76861-6 e-ISBN 978-3-540-76862-3

DOI 10.1007/978-3-540-76862-3

Library of Congress Number: 2008921363

Typesetting and Production: le-tex publishing services oHG, Leipzig, Germany
Cover Design: Frido Steinen-Broo, eStudio Calamar, Spain

Printed on acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

Preface of the Editors

LDIC 2007 was the first International Conference on Dynamics in Logistics, which was organized by the Research Cluster for Dynamics in Logistics (Log*Dynamics*) at the University of Bremen.

The scope of the conference was concerned with the identification, analysis, and description of the dynamics of logistic processes and networks. The spectrum reached from the planning and modelling of processes over innovative methods like autonomous control and knowledge management to the new technologies provided by radio frequency identification, mobile communication, and networking.

The proceedings of LDIC 2007 consist of the two invited papers and of 42 contributed papers, which are organized into the following twelve subjects:

- general aspects of dynamics in logistics,
- routing in dynamic logistic networks,
- RFID in logistics and manufacturing networks,
- supply chain control policies,
- decentralized decision-making in supply chains,
- the Global RF Lab Alliance: research and application,
- sustainable collaboration,
- knowledge management and service models in logistics,
- container logistics,
- autonomous control in logistics,
- next generation supply chain concepts, and
- logistic process modelling.

We would like to thank the members of the programme committee and the secondary reviewers for their help in the selection process. We are also grateful to Karsten Hölscher for his technical support in running the conference system and in editing the proceedings. Moreover, we would like to express our gratitude to Dieter

Uckelmann and Hendrik Wildebrand for their support in the local organization. Finally, we would like to acknowledge the excellent cooperation with the Springer-Verlag.

Bremen,
November 2007

Hans-Dietrich Haasis
Hans-Jörg Kreowski
Bernd Scholz-Reiter

Committees

Program

George Chryssolouris (Patras, Greece)
Jean-Pierre Émond (Florida, USA)
Bonghee Hong (Pusan, Korea)
Kap Hwan Kim (Pusan, Korea)
Paul George Maropoulos (Bath, United Kingdom)
Lázló Monostori (Budapest, Hungary)
Kulwant Pawar (Nottingham, United Kingdom)
Janat Shah (Bangalore, India)
Tetsuo Tomiyama (Delft, The Netherlands)
Kanji Ueda (Tokyo, Japan)
Sergej S. Uwarow (St. Petersburg, Russia)
Albert Veenstra (Rotterdam, The Netherlands)

Organization

Hans-Dietrich Haasis (Chair)
Michael Hülsmann
Frank Kirchner
Herbert Kopfer
Hans-Jörg Kreowski
Bernd Scholz-Reiter
Klaus-Dieter Thoben
Dieter Uckelmann

Contents

General Aspects of Dynamics in Logistics

The Global RF Lab Alliance: Research and Applications

Invited Papers

Challenges in Design of Heterarchical Controls for Dynamic Logistic Systems

Neil A. Duffie

Department of Mechanical Engineering, University of Wisconsin-Madison, Madison, WI 53706, USA

Abstract Principles of design, development and evolution of distributed systems in several disparate domains are reviewed in this paper with the objective of shedding new light on control design in the domain of logistic systems for production, transportation, etc. The spectrum of available organizational structures is reviewed and the case is made for choosing heterarchical structures. Then, principles for designing control systems and building and evolving organizations with heterarchical structures are reviewed. Finally, the design of web services is reviewed, providing another set of principles that are expected to be useful for designing heterarchical controls for dynamic logistic systems.

1 Introduction

In 1978, Enslow identified a set of characteristics defining distributed data processing systems that offered benefits such as easy adaptation to new functions, easy expansion in both capacity and function, incremental replacement and upgrading of components, and graceful degradation. Key characteristics included a multiplicity of computing resources, physical distribution of the computing resources, and a high degree of autonomy between the computing resources. The mode of operation of such systems would be "cooperative autonomy", not hierarchy or anarchy, in which all computing resources follow a "master plan." It has been over 30 years since the advent of the microprocessor and the advent of Ethernet communication networks jointly created the opportunity for designing and implementing such distributed computing systems for applications such as control of production and transportation logistic systems. Yet, in spite of the vast improvements in recent decades in the technology available, the number of successful implementations of such systems remains surprisingly small.

The resistance to such implementations has many facets relating to reluctance to abandon hierarchical system structures. Fairtlough (2005) observed that "Familiarity and naturalness are undeniable and important advantages of hierarchy" as well

H.-D. Haasis et al., *Dynamics in Logistics*

DOI: 10.1007/978-3-540-76862-3,

clarity and certainty: In a hierarchy, "You know who makes decisions" and "When things go wrong you know who is to blame." Organizational structure, leadership and responsibility are clear in a hierarchy and are recognized as ways to prevent chaos. Designers of controls for production and transportation logistic systems have been heavily influenced in the past by the "command and control" style of military management, the hierarchical corporate structures of their employers and customers, and their lack of familiarity with non-hierarchy in organizations.

The distribution of control in organizations and government is a perennial topic of discussion of philosophers and political scientists, and a number of books have appeared in the popular literature in recent years on topics such as the rapid evolution of the world economy, its increasingly non-hierarchical structure, the adaptability and resilience of highly distributed organizations, and the importance of trust both within organizations and in doing business in the world economy. Many factors have contributed to and enabled this global changeability. Those identified by Friedman (2005) included workflow software, outsourcing, insourcing, supply-chaining, informing and uploading. He noted that there is now a dynamic web-enabled platform for global collaboration and that "Wealth and power will increasingly accrue to those countries, companies, individuals, universities, and groups who get three basic things right: the infrastructure to connect with this flat world platform, the education to get more of their people innovating on, working off of, and tapping into this platform, and, finally, the governance to get the best out of this platform and cushion its worst side effects."

Wiendahl et al. (2007) used the taxonomy illustrated in Fig. 1 to summarize the need for production networks to respond this global changeability and the shift from static logistic systems to dynamic logistic systems that adapt to disruptive forces. They noted, "In the past, product development and order handling were regarded

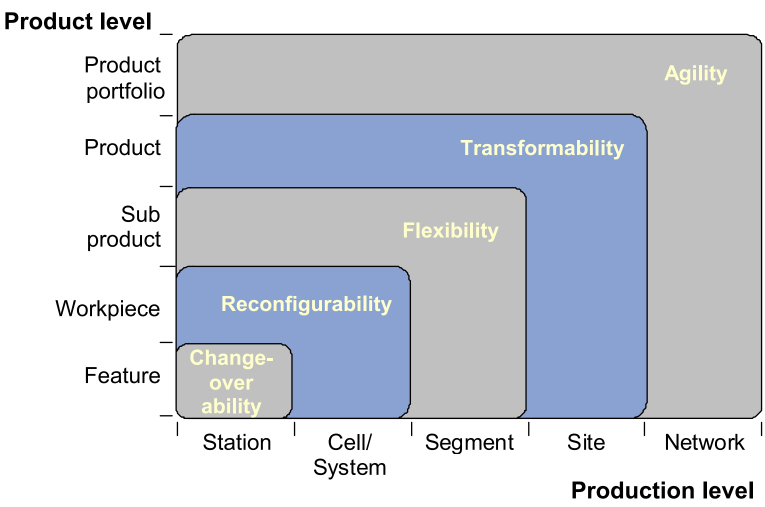


Fig. 1 Types of production changeability (Wiendahl et al. 2007)

as primary processes whereas order fulfillment and distribution were seen more as auxiliary functions. But nowadays, reliable delivery of customized products has the highest priority in globally distributed markets." And, "With decreasing market life of the products, the processes and the equipment also have to produce the next product generation or even the generation after. The buildings have to be adaptable to two and sometime three product generations as well, and even the site must follow new requirements e.g. regarding logistics." Similarly, Windt and Hülsmann (2007) noted, "… today's customers expect a better accomplishment of logistical targets, especially a higher due date reliability, and shorter delivery times. In order to cope with these requirements the integration of new technologies and control methods has become necessary. This is what characterizes the ongoing paradigm shift from a centralized control of 'non-intelligent' items in hierarchical structures towards a decentralized control of 'intelligent items' in heterarchical structures in logistical processes." They note that such intelligent items can include raw materials, components, products, transit equipment and transportation systems.

Ever improving technologies and approaches for designing and implementing non-hierarchical systems such as agents (Monostori et al. 2006), emergent behavior (Ueda et al. 2007), bounded rationality (Selten 2001) and tracking and tracing technologies such as RFID (Schuh et al. 2007) are creating new opportunities for communication, cooperation and collaboration in control of dynamic logistic systems. Also, technologies that are creating virtual, mobile and personal systems are taking all forms of collaboration, enhancing them, and "making the world flatter every day" (Friedman 2005). These developments should encourage system designers to consider and explore system architectures that are not hierarchical in nature.

There are many challenges in the design and evolution of controls for dynamic logistic systems: What could/should be the characteristics of designs of future controls for such systems? From what domains can guidance for design of controls for dynamic logistic systems be drawn? Which fundamental design philosophies are likely to produce trustworthy and profitable logistic systems that can be readily evolved over time and reduce the need for "jump off the cliff" system investments? To address these challenges, principles of design, development and evolution of distributed systems in several disparate domains will be reviewed in this paper with the hope of shedding new light on control design in the domain of logistic systems. The question of what combinations of technologies and methods can be used to implement such controls is important, but will not be focused on here. First, the spectrum of available organizational structures will be reviewed. Then the case for choosing heterarchical structures will be made. Then principles for designing control systems with heterarchical structures for manufacturing machines and cells will be reviewed, followed by a review of principles for development and evolution of organizations. As illustrated in Fig. 2, it is expected that consideration of principles for design and development in these domains will be beneficial in the design and evolution of controls for dynamic logistic systems such as those for production and transportation; furthermore, it is expected that the influence of such design domains will change depending upon the level to be controlled within a production, transportation or other network. Finally, the concept of web services will be reviewed because they provide

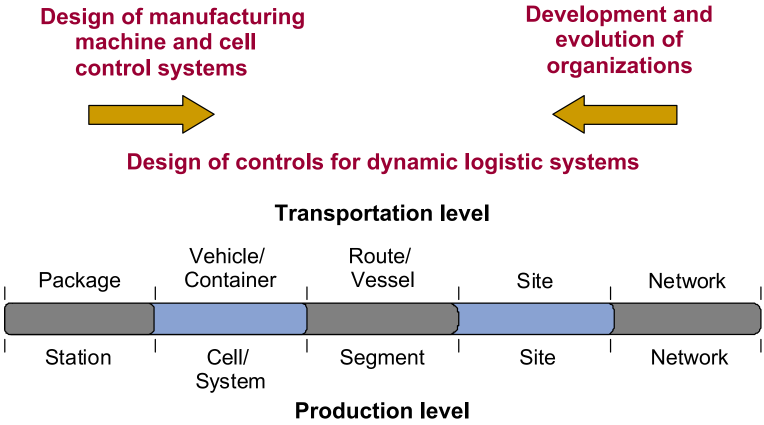


Fig. 2 Design of controls for dynamic logistic systems for applications such as production and transportation can benefit from principles developed for other domains

another set of principles that may be useful, both directly and comparatively, for designing heterarchical controls for dynamic logistic systems.

2 Options for Structuring Controls for Logistic Systems

Fairtlough (2005) identified three ways of structuring organizations to "get things done": hierarchy, heterarchy and responsible autonomy. He noted, "… in real life organizations we find a blend of the three ways. It is the blend that differs from organization to organization." Figure 3 illustrates the spectrum represented by these three organizational structures, together with anarchy, which can be thought of as lack of structure. Control is distributed in all of these structures because there are multiple decision-making entities; however, hierarchies exhibit centralized control,

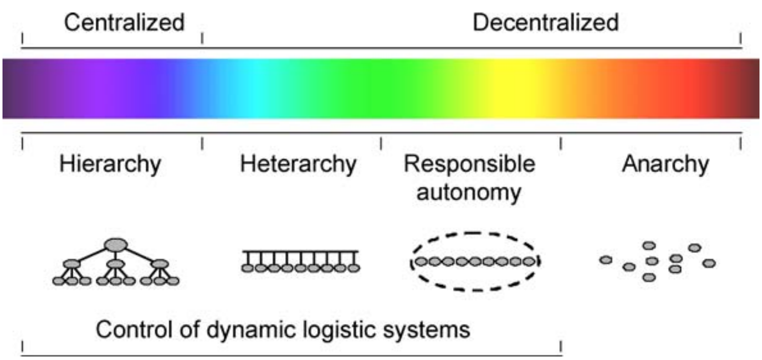


Fig. 3 Spectrum of organizational structures (adapted from Fairtlough, 2005)

whereas heterarchies and responsible autonomies exhibit decentralized control. Because positioning the structure of controls for dynamic logistic systems in the hierarchy, heterarchy, responsible autonomy spectrum is likely to be a major design decision, it is useful to review the characteristics of such organizational structures.

2.1 Hierarchy

A hierarchy is an organizational structure in which each entity of the system (except for a supreme entity) is subordinate to a master entity. In a hierarchy, a master entity makes decisions on behalf of its subordinates, and subordinate entities collaborate with their surrounding entities through their master entity (Fairtlough 2005). Control in a given portion of a hierarchy is centralized in its master entity. An example of a hierarchy is a patriarchal family system.

2.2 Heterarchy

A heterarchy is an organizational structure in which each entity shares the same "horizontal" position of power and authority, none being subordinate to another. In a heterarchy, decision-making is mutual and an entity can act in collaboration with any of its surrounding entities without guidance or permission from a master entity (Fairtlough 2005). An example of a heterarchy is a strategic business alliance in which no one partner dominates the others.

2.3 Responsible Autonomy

A responsible autonomy is an organizational structure in which each entity makes decisions autonomously; however, each entity is accountable for the outcome of its decisions. In a responsible autonomy there are "… clearly defined boundaries at which external direction stops" (Fairtlough 2005). An example of a responsible autonomy is the relationship between businesses and shareholders. The shareholders periodically examine business performance and make changes in management if desired.

2.4 Anarchy

In an anarchy there is no organizational structure for making decisions and each entity has the absolute liberty to act on its own. Modern-day definitions of anarchy refer to a state of disorder, as well as to a state of harmony and "… voluntary cooperation and free association of individuals and groups" (www.m-w.com 2007). The Internet has sometimes been described as an anarchy because of the equality of access to much of it (obtained using a common public protocol) and because much of the content is provided without oversight or governance.

3 Design of Heterarchical Control

As noted above, Enslow proposed in 1978 that distributed computing systems be designed as a "cooperative autonomy," which would not be a hierarchy or an anarchy. In 1982, Duffie noted that distribution in machine control could "... produce modular, expandable and adaptable control systems with a high degree of hardware transparency ... Cost benefits are likely to be accrued through improved flexibility, reliability, and performance of these systems ... The attainment of these benefits depends on organizing the system into a set of autonomous processes supported on a group of processors connected by a communication network." In 1985, Hatvany noted, "Highly centralized and hierarchically ordered systems tend to be rigid, constrained by their very formalism to follow predetermined courses of action. However carefully 'optimized' their conduct may be, it has been shown that this very property of inherent resistance to organizational change itself necessarily leads in due course to catastrophic collapses ... On the other hand, fragmentation of a system into small, completely autonomous units, each pursuing its own selfish goals according to its own, self made laws, is the absurdity of primitive anarchy." Instead, Hatvany proposed that the entities (autonomous units) would be structured as a "cooperative heterarchy" in which the entities would

- Have equal right of access to resources
- Have equal mutual access and accessibility to each other
- Have independent modes of operation
- Strictly conform to the protocol rules of the overall system.

In explaining his concept for cooperation, Hatvany proposed, "... participants of such a heterarchy receive rights in return for assuming obligations. While there is no 'higher level' controller of the system, nevertheless each member must conform to certain rules, in order to obtain certain privileges ..." Each entity would have two sets of goals: "... one set of goals is concerned with the internal conduct of the sub-system and remains within the domain of its autonomy ... another set must always be dominant in each local set of evaluation criteria. This second set of goals is directed towards the optimal overall operation of the system, and it is these that introduce the property of cooperation." While stressing the importance of autonomy, none of these authors envisioned structuring data processing, machinery control, or manufacturing cell control as a responsible autonomy in which there are no requirements for dynamic cooperation and collaboration between entities. It is reasonable to think of this structure as an extreme case of heterarchy in the spectrum shown in Fig. 3. Nevertheless, the teaching of these authors is clearly that in the spectrum of organizational structures, a major emphasis in the design of heterarchical controls should be on autonomy.

3.1 Principles for Partitioning

To assist in designing heterarchical control systems with high levels of autonomy, Duffie (1990) proposed the following set of principles for decomposing and partitioning system functional requirements into a set of quasi-independent, communicating, cooperating entities:

- There is a natural decomposition associated with the system.
- The result of decomposition is a set of quasiindependent entities with relatively weak interactions.
- All communication between entities takes the form of messages transmitted on a network.
- The physical system configuration should be transparent to the entities in the system, and entities should not need to know where other entities reside.
- Time-critical responses should be contained within entities and should not be dependent on time-critical responses from other entities.
- The resilience and viability of individual entities should be a major criteria.

3.2 Principles for Fault Tolerance

Duffie (1990) also proposed the following set of principles for achieving fault tolerance in heterarchical control systems.

- Master/slave relationships should not exist between entities.
- Entities should cooperate with other entities whenever possible.
- Entities should not assume that other entities will cooperate with them.
- Entities should delay establishing relationships with other entities for as long as possible, and should terminate these relationships as soon as possible.
- Information generated by an entity should be retained locally rather than globally, and communicated upon request to other entities.

Several of the principles listed above raise questions about the level of trust that a control system should be designed for. The issue of trust will be discussed in more detail later in this paper.

3.3 Example: Heterarchical Control of Part Production

Duffie designed a number of controls with heterarchical structures for manufacturing machine and cell control (Duffie 1982, Duffie et al. 1986, Duffie et al.

1988). For example, in the system shown in Fig. 4, 12 parts are to be produced using two milling machines and two EDM machines. Each part has "intelligence" in the sense that each part is supported by its own independent software application, running on a computer somewhere in the system, that enables it to communicate with machine controllers via a network. In this simple system there are no due dates and no schedule optimization: when a part enters the system it begins to compete for machines using the method summarized in Fig. 5. When there is contention in the systems (several parts competing for the machines), parts continually execute the request/reservation sequence until they successfully reserve a machine. This is a heterarchical control structure because there is no master scheduler and the part entities autonomously compete for machine resources. Cooperation between parts and machines is enabled by the request/reservation procedure for resource allocation. There is no need for the parts to cooperate directly, but their processing is intrinsically coordinated by the procedure. Relationships are only established when reservations have been confirmed. If, for whatever reason, machines or parts are not sufficiently trustworthy to always fulfill their relationships, then additional communication and fault recovery steps must be added to the procedure.

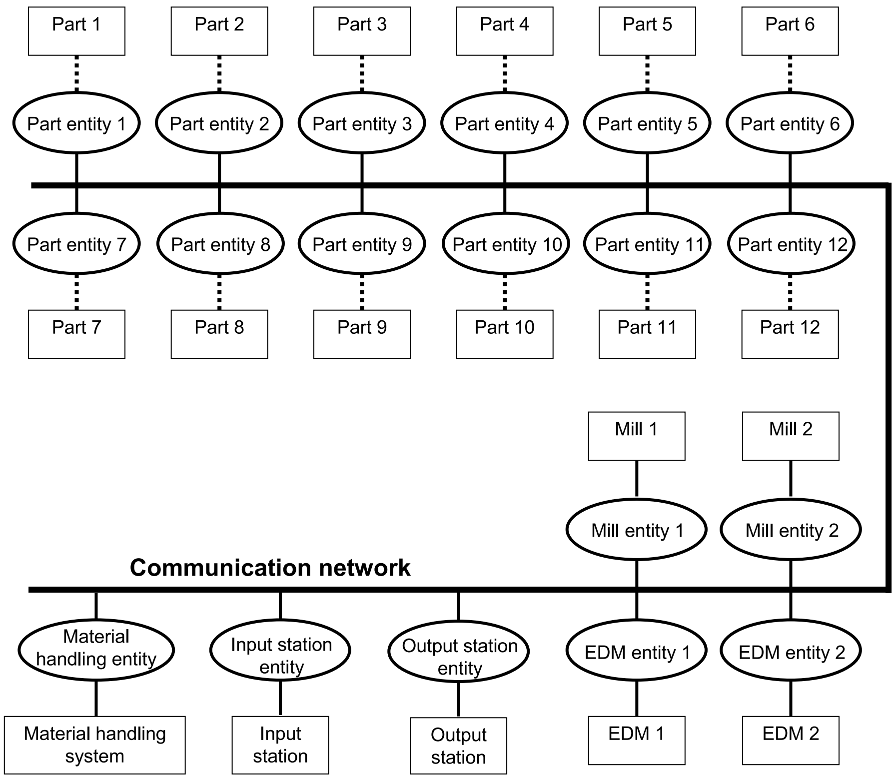


Fig. 4 Part production system with distributed part and machine intelligence

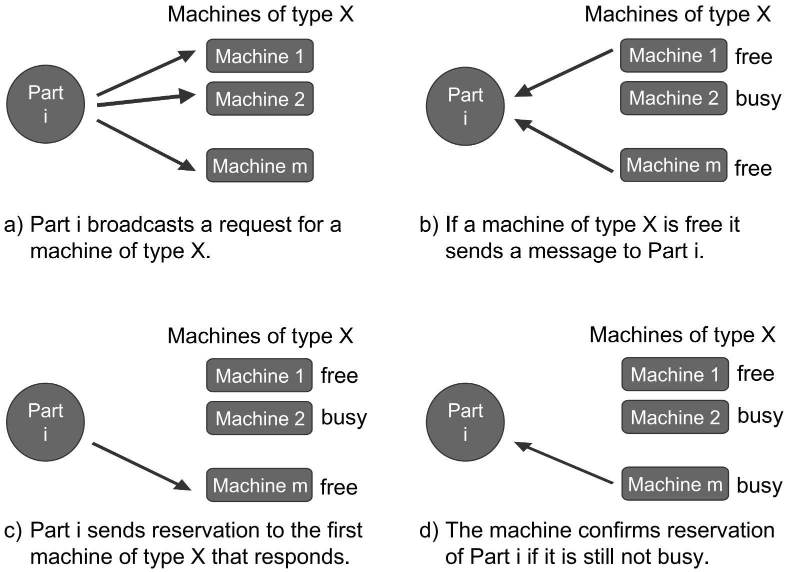


Fig. 5 a) – d) Request/reservation procedure for heterarchical control of part production

3.4 Example: Heterarchical Control of a Multitude of Propulsion Units

A heterarchical control was studied by Li et al. (1995) for a material handling system consisting of multiple, small, passive vehicles driven a high speeds through a multitude of approximately 1-meter long linear induction motor (LIM) propulsion units. The goal was high-speed material handling for applications such as delivery of small parts and tooling. Such a system would require distributed routing and dispatching, branching and merging, and associated collision avoidance. Due to the potentially large number of propulsion units and the short period of time (milliseconds) that a vehicle moving at high speed in the system would pass through and be controlled by a given propulsion unit, it would not be practical to coordinate the propulsion units using a centralized controller. Therefore, the structure shown in Fig. 6 was chosen in which the propulsion units are linked with a communication network with a topology identical to the topology with which the propulsion units are connected. An example of a 20-unit propulsion/com-munication network topology is shown in Fig. 7.

The identical propulsion and communication network topologies allow messages to be transmitted downstream and upstream from a propulsion unit in which a vehicle is (briefly) present for the purpose of avoiding collisions with other vehicles in the network. Messages can be transmitted downstream from the propulsion unit containing a loaded vehicle to locate the propulsion unit at the destination of that vehicle and implicitly map the route to the destination propulsion unit. Similarly, messages can be transmitted upstream from the propulsion unit at a station requir-

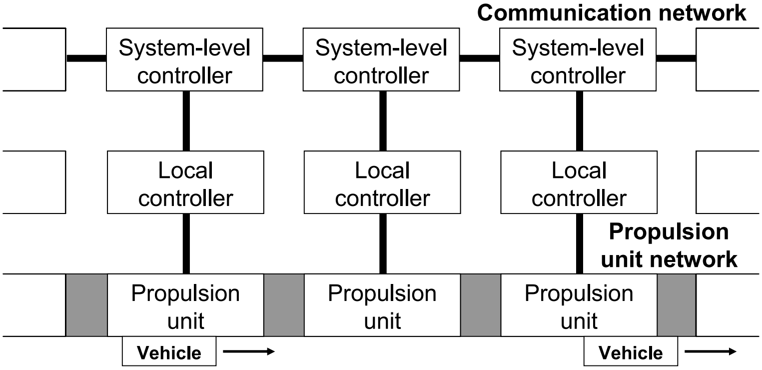


Fig. 6 Identical propulsion unit and communication network topologies enable heterarchical control of vehicles in a high-speed material handling system

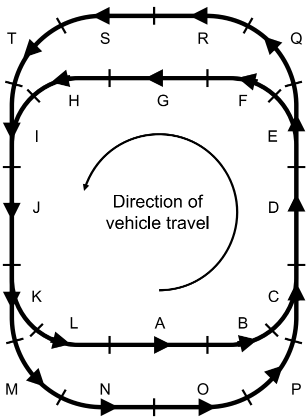


Fig. 7 Example of a material handling system with identical communication and propulsion network topologies

ing service to locate a propulsion unit containing an empty vehicle and implicitly map the route from that propulsion unit to the station.

For example, suppose that it is desired that a vehicle be located and be sent to Propulsion Unit A in Fig. 7. As illustrated in Fig. 8, a message requesting a vehicle is sent upstream from Propulsion Unit A, and each propulsion unit through which the message passes appends its name and its length to the message to map the route. The message is replicated at each branch, and a message is discarded if it returns to a unit already on the list, including Propulsion Unit A. If the message encounters a propulsion unit containing an available vehicle, an acknowledging message is returned to Propulsion Unit A via the route in the requesting message. Propulsion Unit A can choose a vehicle from the set of acknowledging messages it receives and again uses the route to make a reservation with the vehicle. The route then is passed from propulsion unit to propulsion unit as the vehicle moves toward Propulsion Unit A. This is a heterarchical control because there is no master controller. The propulsion

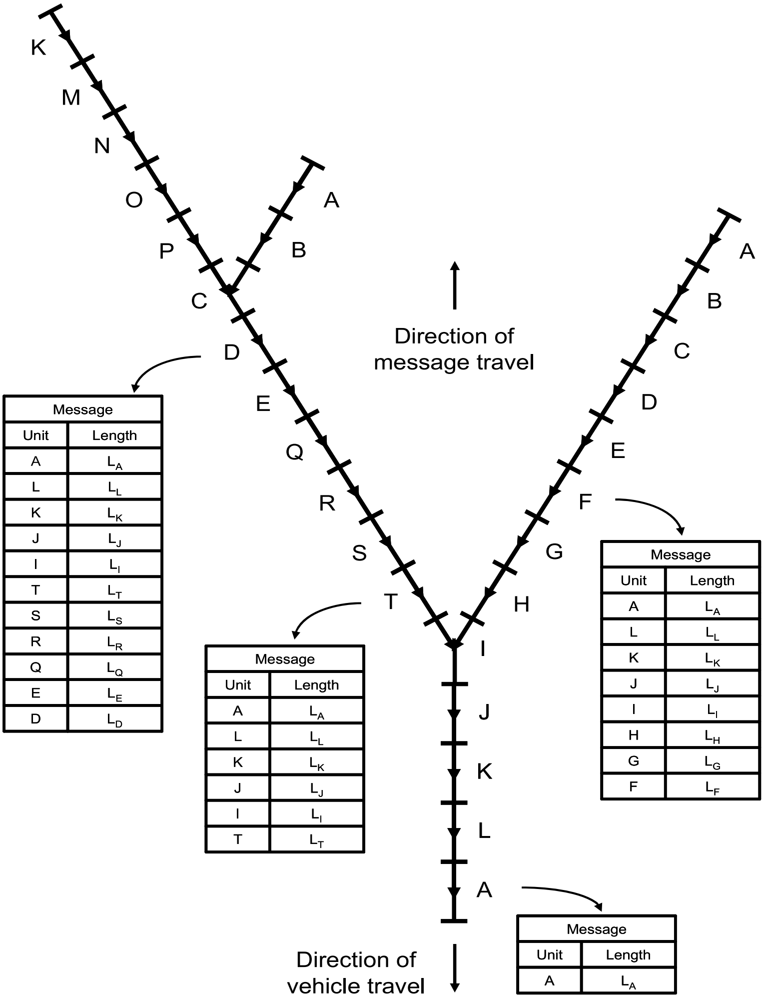


Fig. 8 Dispatching messages are propagated upstream in the propulsion network

unit entities are autonomous, but cooperation is achieved using rapid transmission of messages between neighbors. Responsibility for vehicle control is passed from propulsion unit to propulsion unit as the vehicle enters and leaves a propulsion unit.

4 Developing and Evolving Organizations

Suroweiki (2005) observed, "… under the right circumstances, groups are remarkably intelligent, and are often smarter than the smartest people in them." For this to be the case, Suroweiki noted that groups should have the following attributes: decentralization (group members should be specialized and draw on local knowledge);

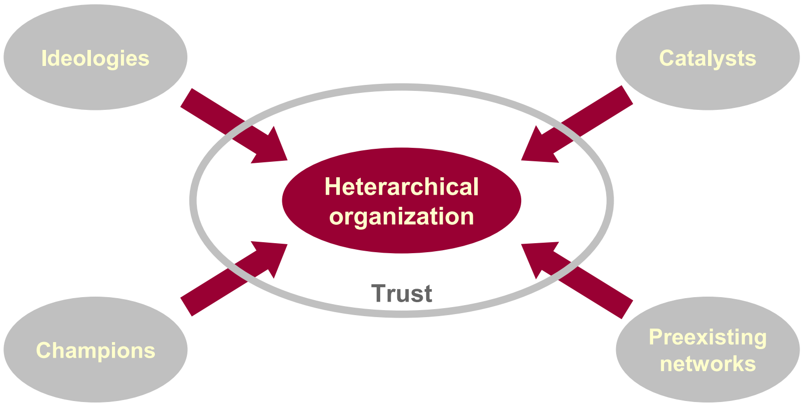


Fig. 9 Key contributors to developing and evolving a heterarchical organization (adapted from Brafman and Beckstrom 2006)

diversity of opinion (members should have private information); independence (the judgments of local members should not significantly influenced by the judgments of members around them); and aggregation (a mechanism needs to exist for turning the local judgments of members in the group into collective decisions). As indicated in Fig. 2, principles for the structuring, development and evolution of organizations (groups of communicating, cooperating, collaborating humans such as governing bodies, professional organizations, clubs and activist groups) can serve as important guides in the design of controls for dynamic logistic systems. Figure 9 identifies some key contributing elements to developing and evolving a successful heterarchical organization. Brafman and Beckstrom (2006) discussed the importance of catalysts, champions and preexisting networks:

Ideologies

"Ideology is the glue that holds decentralized organizations together . . . " and organizations that are successful over a long period of time have powerful ideologies.

Catalysts

"... a catalyst gets a decentralized organization going and then cedes control to the members." A "catalyst" builds connections and networks, building an organization person by person. "Once there's an emotional connection, then and only then is it time to brainstorm and talk strategy." The following attributes are among those noted as being important for a catalyst: genuine interest in others, loose connections, meeting people "where they are," assuming a peer relationship, listening intently, inspiring change without being coercive, and tolerance for ambiguity. Regarding the latter, Brafman and Beckstrom noted, "One day people are excited, the next they're ambivalent. One circle excels, another fails. There's no way to measure results. There's no way to keep track of all the members. There's no way to even know who is doing what, let alone where and when. To an outsider, the chaos might appear overwhelming." However, "... ambiguity creates a platform for creativity and innovation."

Champions

Champions "tend to be more like salesmen than organizers or connectors." A champion inspires the group, but is not the decision maker. Champions are more concerned with strategy and tactics than network building; hence, because both salesman and organizers are necessary, heterarchical organizations with complementary catalysts and champions are more likely to be successful.

Preexisting Network

Most highly successful decentralized organizations have been built upon an existing network. The Internet is a key network upon which decentralized organizations can be developed and evolved, but face-to-face networks remain important. The health of a heterarchical organization can be monitored by examining both the organization's network and its individual entities: Are the entities participating? Is the network growing and evolving? Is it becoming more or less decentralized? How active are the entities? Are they autonomous? What is the nature of their connections with other entities?

Trust

Friedman (2005) observed, "With a flattened hierarchy, you never know what people are going to do. You can't control the outcomes... All you can control is whether people have personal relationships with each other based on trust." Similarly, Reina and Reina (2006) noted that transactional trust, a key element in heterarchical control of dynamic logistic systems, has three main components and can be achieved in organizations by the following actions:

Trust of disclosure

- Share information
- Tell the truth
- Admit mistakes
- Give and receive constructive feedback
- Maintain confidentiality
- Speak with good purpose

Trust of capability

- Acknowledge people's skills and abilities
- Allow people to make decisions
- Involve others and seek their input
- Help people learn skills

Trust of character

- Manage expectations
- Establish boundaries
- Delegate appropriately
- Encourage mutually serving intentions
- Keep agreements
- Be consistent

Trust is important, but is not the only way that cooperation can be achieved. Axelrod (1984) noted, "… the two key requisites for cooperation to thrive are that the cooperation be based on reciprocity, and that the shadow of the future is important enough to make this reciprocity stable." Axelrod also noted that "… there is no need to assume trust between the players: the use of reciprocity can be enough to make defections unproductive." And "Finally, no central authority is needed: cooperation based on reciprocity can be self-policing." Even the simple "TIT FOR TAT" strategy for cooperation was shown by Axelrod to be a highly successful strategy for mutual cooperation because "Its niceness means that it is never the first to defect, and this property prevents it from getting into unnecessary trouble. Its retaliation discourages the other side from persisting whenever defection is tried. Its forgiveness helps restore mutual cooperation. And it clarity makes its behavioral pattern easy to recognize; and once recognized, it is easy to perceive that the best way of dealing with TIT FOR TAT is to cooperate with it."

Evolution of the Structure of Organizations

Early in their development, many organizations may have heterarchical structures; however, they may grow more hierarchical over time as they grow in size and complexity. As noted by Friedman (2005), organizations need to become more heterarchical ("flatter") in response to changes in global changes in economics and societies. Evolution, being at the right point in the spectrum in Fig. 3 at the right time, is important in the development and survival of organizations. As an example, Fig. 10 shows the evolution of a hypothetical professional organization with time, beginning with its formation as a heterarchical organization consisting of many local groups with mainly local interactions, evolving to a more hierarchical organization with many national income-generating components and a national headquarters and staff to manage them, and then evolving to a leaner and more heterarchical organiza-

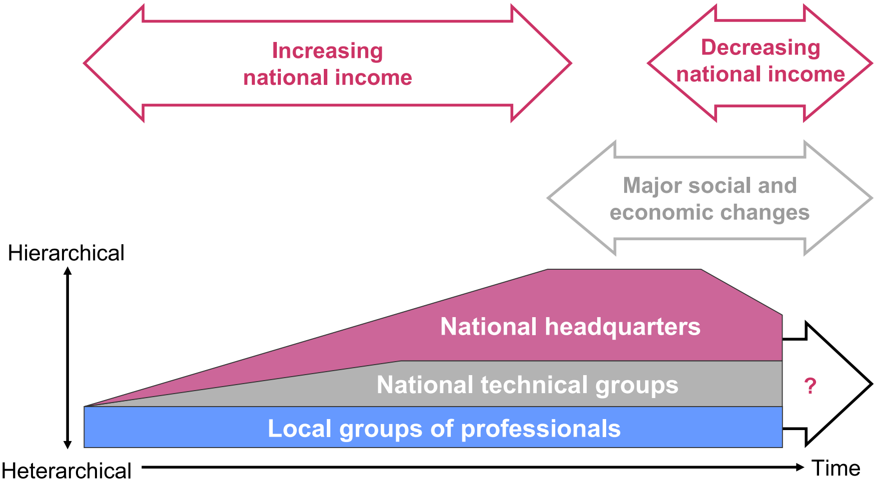


Fig. 10 Example: Evolution of a hypothetical professional organization

tion as national income and influence decreases due to social changes; for example, new generations of professionals join the organization who have new ways of communicating.

5 Design of Web Services

Web service concepts provide another set of principles that can be useful in designing heterarchical controls for dynamic logistic systems. Web services are delivered by distributed service providers and are accessed over a network. These remotely executed services are connected using a service-oriented architecture (OASIS 2006) to form needed logistic processes. If web services are loosely coupled as summarized in Table 1, then interoperability can be achieved, making logistic processes

Table 1 Coupling in web services (adapted from Kaye 2003)

	Tightly coupled	Loosely coupled
Interaction	Synchronous	Asynchronous
Messaging style	RPC	Document
Message paths	Hard coded	Routed
Technology mix	Homogeneous	Heterogeneous
Data types	Dependent	Independent
Syntactic definition	By convention	Published schema
Bindings	Fixed and early	Delayed
Software objective	Re-use, efficiency	Broad applicability
Consequences	Anticipated	Unexpected

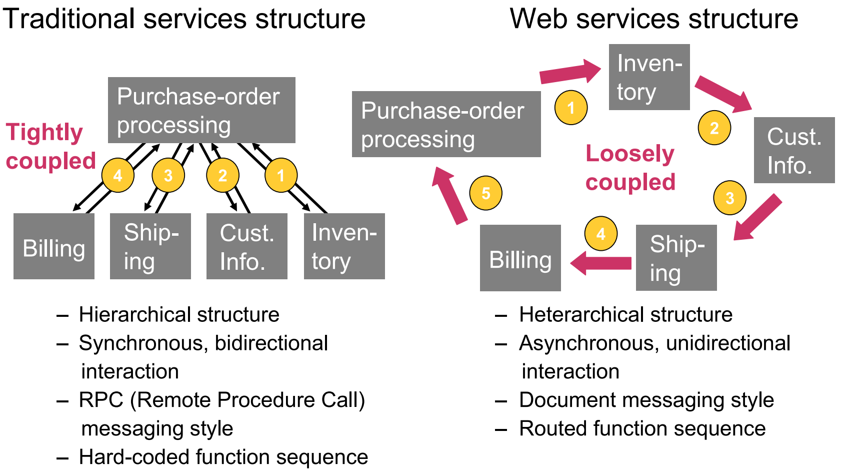


Fig. 11 Traditional and web services structures

changeable and making the services available for unanticipated future applications (by Kaye 2003). As an example, a purchase-order handling process is illustrated in Fig. 11 for both a tightly-coupled traditional services structure and a loosely-coupled web services structure. The traditional structure is hierarchical because the purchase-order processing entity has a master-slave relationship with the service entities. The web services struture is heterarchical because of the autonomy granted to the service entities as orders are processed.

Document Messaging

Specifications needed for logistic processes include the services required and the order in which they are to be executed (for example, the four steps in the purchase order handling process in Fig. 11: checking inventory, followed by obtaining customer shipping and billing information, followed by shipping the order, followed by billing the customer), as well instructions for executing the services (for example, cancelling the order if inventory is insufficient or rerouting the shipment if the primary route is not available). Kaye (2003) discusses several methods for handling such specifications, which can be adapted as follows for logistical processes:

Embed specifications in a centralized logistic control entity in a hierarchical structure.

- Specifications are in a single place
- Does not distribute intelligence
- Tight coupling

Embed specification subsets in specialized entities in a heterarchical structure.

- A single change in a logistic process could require a cascade of changes in multiple entities
- Tight coupling

Embed specifications in documents communicated between specialized entities in a heterarchical structure.

- Requires local interpretation of document schema
- Ideal for Web Services

With the latter, document messaging method, the document contains the schema (a description of the document's structure and content), the services required (including the order in which services are needed), rules for executing the services, initial order data (used in decision-making, scheduling, routing, etc.), and additional data (results) appended by upstream services.

Specialized Logistic Service Entities

Kaye (2003) listed a set of principles for designing service entities that can be adapted as follows for dynamic logistic processes:

- Functional components of logistic processes should be distributed between entities according to principles of modularity.
- Entities should only need to handle that portion of the information in a document that is relevant to the function(s) the entity performs.
- Entities should not need to know details of the function(s) performed by previous or subsequent entities on the document's route.
- Entities should not need to know the overall purpose of a document; an entity should be able to process any document that requires that entity's function(s).

Trust in Web Services for Logistic Processes

Distribution, loose coupling, document routing and asynchronous messaging can create a "fear of betrayal" in logistic processes based on web services that does not exist in the tightly-coupled traditional services structure shown in Fig. 11. In the web services structure, trust is required that the document will be returned after a (possibly long) delay, with information (data appended to the document by the service providing entities) indicating that services have been provided as called for in the document. If trust is low, perhaps due to insufficiently detailed specifications or incapable service providing entities, then tighter coupling can be introduced into the

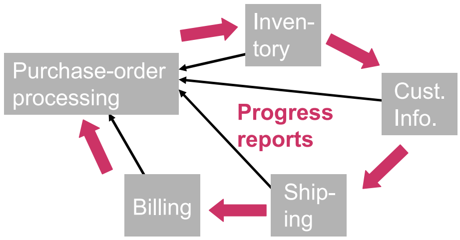


Fig. 12 Tighter coupling in a situation of lower trust

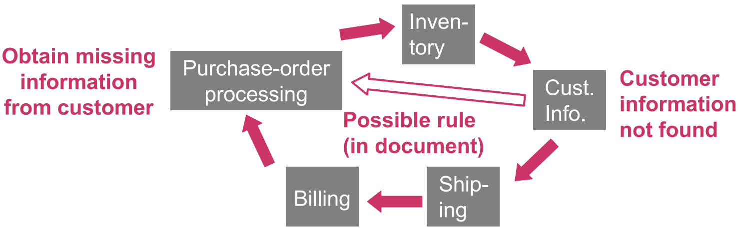


Fig. 13 Improvement of specifications in document, making service trustworthy

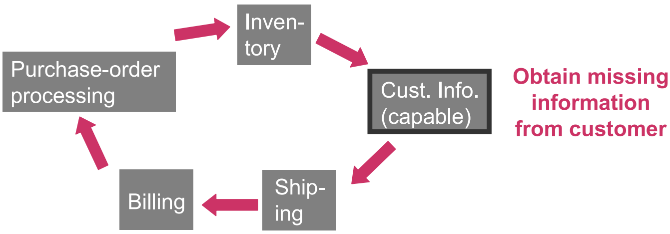


Fig. 14 Replacement customer information service entity with a more capable entity

design (for example, the progress reporting functions shown in Fig. 12). However, this reduces autonomy in the system. There are alternatives that preserve the benefits of loose coupling including inclusion of rules in documents that specify actions to be taken in exceptional situations (for example, as illustrated in Fig. 13, a rule for returning the document to the purchase-order processing entity if proper customer information cannot be found) or replacing of a service providing entity with a more capable entity (for example, as illustrated in Fig. 14, a new entity capable of obtaining missing customer information).

Evolution of Web Services

The replacing a service entity with a more capable service entity is one form of evolution in web services. Another form of evolution is the augmenting of logistic processes with additional services. For example, a new customer follow-up service

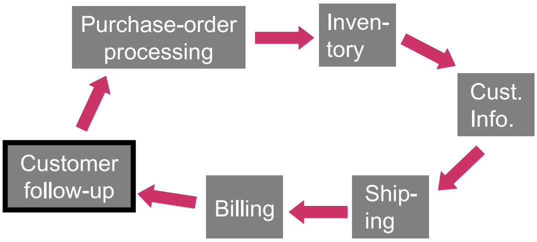


Fig. 15 Addition of a new service to a logistic process

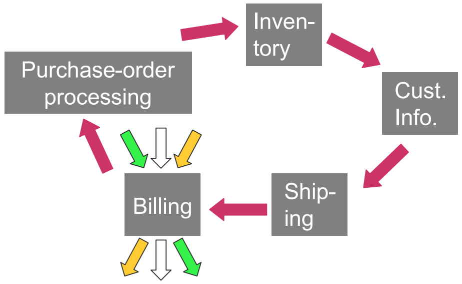


Fig. 16 Broad applicability of a web service can lead to is use by many logistic processes

might be added, as shown in Fig. 15, to assess customer satisfaction. Another form of evolution is the use of a broadly applicable web service by a many logistics processes. For example, if the Billing service is designed so that it is broadly applicable, this may lead to its use by other logistic processes as shown in Fig. 16, increasing the demand for the service.

6 Conclusions

The concepts of cooperative autonomy and cooperative heterarchy, described some 20–30 years ago for applications as diverse as distributed data processing and manufacturing cell control, remain relevant today in the design of controls for dynamic logistic systems. Some designers remain uncomfortable with the loose coupling implicit in heterarchical designs, preferring instead to take traditional tightly-coupled approaches that result in hierarchical, master-slave structures. Other designers may dismiss heterarchical controls because the advantages of a heterarchical often can be recreated in a hierarchical structure (but often with greater complexity and less fault tolerance). Rather than assessing functionality, internal inspection of a system can reveal its structure. Brafman and Beckstrom (2006) note that this can be done by asking questions such as: Are information and control distributed or concentrated? Do entities communicate directly rather than through intermediaries? Is there a clear division of roles? If an entity is removed, is the organization harmed? Are entities self motivated or are they motivated by the organization?

As illustrated in Fig. 17, design of heterarchical controls for dynamic logistic systems can be guided by design principles developed for other domains such as control of manufacturing machines and cells, development and evolution of organizations, and design of web services. However, the requirements for control not only will vary between logistic systems, but also may change with time for a given logistic system. Hence the positioning of a design in the spectrum of heterarchy between hierarchy and responsible autonomy in Fig. 18 depends upon the application and its evolution. Furthermore, as noted by Fairtlough (2005), large complex systems can be structured as a blend of hierarchy, heterarchy and responsible autonomy. As pointed out by both Friedman (2005) and Brafman and Beckstrom (2006), a decentralized organization can evolve and grow quickly and easily, and there is a tendency for such

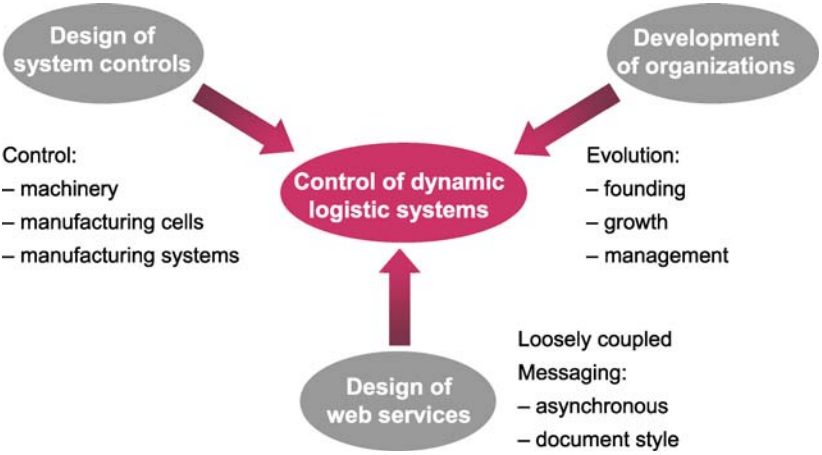


Fig. 17 Some sources of principles for design of heterarchical controls for logistic systems

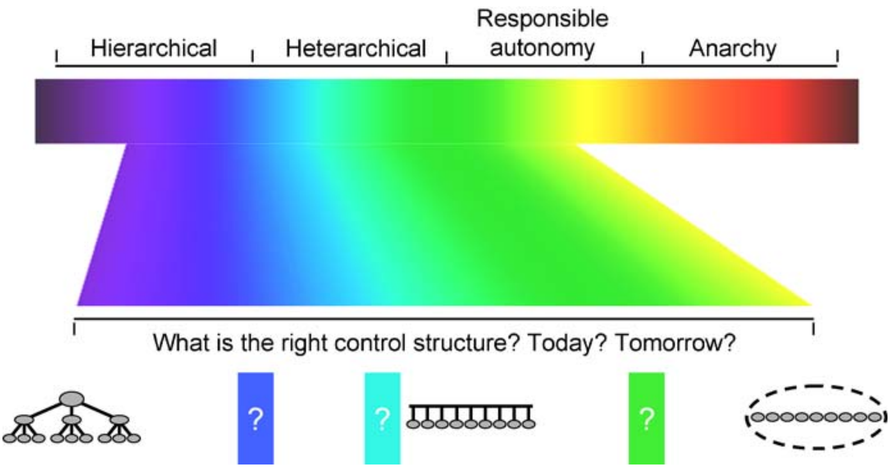


Fig. 18 Requirements for control of dynamic logistic systems can evolve over time

organizations to become more decentralized when under pressure, from changing political or economic conditions for example. However, profit making may become more difficult as a logistic system becomes more decentralized, creating challenges in choosing or evolving to the right structure.

A vision of the future is that systems of autonomous logistic processes, organized in heterarchical structures within networks for production, transportation, etc., rapidly evolve and increase in capability. Their complexity would evolve dynamically rather than be designed, and investments in them would be incremental to minimize cost and risk. Many such logistic systems and their associated logistic processes could emerge, with some disappearing rapidly due to foreseen and unforeseen forces such as globalization, the Internet and local innovation. Profit making could be challenging in delivering logistic processes in such networks because, as noted by Friedman (2005) "The small shall be big" if they can take advantage of Internet-based collaboration tools "… to reach farther, faster, wider, and deeper." Will the designers of controls for tomorrow's logistic systems dare to use heterarchical structures? Will tomorrow's users dare to use them? Time will tell, but the technological developments in recent years in communication networks, embedded processors, service oriented architectures, etc. have removed many obstacles and created many opportunities for designing entirely new heterarchical controls for dynamic logistic systems.

Acknowledgements This work was supported in part by the U.S. National Science Foundation under grant DMI-0646697.

References

Axelrod R (1984) The Evolution of Cooperation, Basic Books, NY

Brafman O and Beckstrom RA (2006) The Starfish and the Spider – The Unstoppable Power of Leaderless Organizations. The Penguin Group, New York

Duffie N (1982) An Approach to the Design of Distributed Machinery Control Systems. IEEE Trans. on Industry Applications, Vol. IA-18, No. 4, pp 435–442

Duffie NA (1990) Synthesis of Heterarchical Manufacturing Systems. Computers In Industry, Vol. 14, pp 167–174

Duffie NA, Chitturi, R and Mou JI (1988) Fault-Tolerant Heterarchical Control of Heterogeneous Manufacturing System Entities. Journal of Manufacturing Systems, Vol. 7, No. 4, pp 315–328

Duffie NA, Piper RS, Humphrey BJ and Hartwick JP (1986) Hierarchical and Non-Hierarchical Manufacturing Cell Control with Dynamic Part-Oriented Scheduling. Proc. of NAMRC-XIV, Minneapolis, MN, May 28–30, pp 504–507

Enslow PH (1978) What is a "Distributed" Data Processing System? IEEE Computer Vol. 11, January, pp 13–21

Fairtlough G (2005) The Three Ways of Getting Things Done: Hierarchy, Heterarchy & Responsible Autonomy in Organizations. Triarchy Press, Dorset

Friedman TL (2005) The World is Flat – A Brief History of the Twenty-First Century. Farrar, Straus and Giroux, New York

Hatvany (1985) Intelligence and Cooperation in Heterarchical Manufacturing Systems. Robotics and Computer Integrated Manufacturing, Vol. 2, pp 101–104

http://www.m-w.com 2007. Merriam-Webster Inc., Springfield

Kaye (2003) Loosely Coupled – The Missing Pieces of Web Services. RDS Press, Marin County, CA

Li LJ, Prabhu VV, Duffie NA and Lorenz RD (1995) Distributed LIM-Based Actuator and Control for an Ultrafast Material Transfer Network. Proc. of SPIE Photonics East '95 Symp., Philadelphia, PA, Oct. 22–26, 1995, Vol. 2595, pp 200–213

Monostori L, Váncza J and Kumara SRT (2006) Agent-Based Systems for Manufacturing. Annals of the CIRP, Vol. 55, No. 2, pp 697–720

OASIS (2006) Reference Model for Service Oriented Architecture 1.0. Organization for the Advancement of Structured Information Standards, Billerica, MA

Reina DS and Reina ML (2006) Trust & Betrayal in the Workplace: Building Effective Relationships in Your Organization. Berrett-Koehler Publishers

Schuh G, Gottschalk S and Höhne T (2007) High Resolution Production Management. Annals of the CIRP, Vol. 56, No. 1, pp 439–442

Selten L (2001) What is Bounded Rationality?. In: Gigerenzer G and Selten R (eds) Bounded Rationality: The Adaptive Toolbox, MIT Press, Cambridge, MA, pp 13–36

Surowiecki J (2005) The Wisdom of Crowds. Anchor Books, New York

Ueda K, Fujii N and Inoue R (2007) An Emergent Synthesis Approach to Simultaneous Process Planning and Scheduling. Annals of the CIRP, Vol. 56, No. 1, pp 463–466

Wiendahl HP, ElMaraghy HA, Nyhuis P, Zäh MF, Wiendahl HH, Duffie N and Brieke M (2007) Changeable Manufacturing – Classification, Design and Operation. Accepted for publication in Annals of the CIRP, Vol. 56, No. 2

Windt K and Hülsmann M (2007) Changing Paradigms in Logistics – Understanding the Shift from Conventional Control to Autonomous Cooperation and Control. In: Hülsmann M and Windt K (eds) Understanding Autonomous Cooperation and Control in Logistics, Springer, Berlin, pp 1–16

Making the Business Case for RFID

Bill Hardgrave[1], **Cynthia K. Riemenschneider**[1], **Deborah J. Armstrong**[2]

[1] Information Systems Department, Sam M. Walton College of Business, University of Arkansas, United States
[2] Department of Management Information Systems, The College of Business, Florida State University, United States

Abstract In June 2003, Wal-Mart asked its top 100 suppliers to begin using radio frequency identification (RFID) tags on pallets and cases beginning January 2005. Since that announcement, the business value of RFID in the consumer packaged goods supply chain has been periodically questioned. Recently, a Wall Street Journal article asserted that RFID is not living up to its hype, and in reality is not providing the promised tangible business value throughout the supply chain. In light of such claims, this paper examines the business value of RFID in the supply chain and presents a model of RFID assimilation which proposes that the creation of business value is dependent upon the depth of assimilation (extent of use). The model is grounded in industry observations of the difficulty of early adopters to fully realize the benefits of RFID assimilation. The model conceptualizes RFID assimilation as occurring in three waves: the first wave of the model is Technology Deployment, the next wave of Data Understanding, and, lastly, the final wave is Business Value Creation. In this paper, the first two waves of the model are explored briefly with the emphasis placed on proven business cases and potential opportunities for RFID to provide business value in the supply chain.

1 Introduction

Radio frequency identification (RFID) is a form of automatic identification that uses radio waves to identify products. Because of its many advantages over barcode (the current standard for automatic identification), many retailers have begun the transition to RFID. Perhaps the most significant sign of this transformation was Wal-Mart's announcement in June 2003 of its intention to have top suppliers begin using RFID tags on pallets and cases by January 2005. The Department of Defense, Target, Albertson's, Best Buy, and others soon followed with their own RFID initiatives.

However, the transition has not been smooth. Many in the industry question the business value (i.e., ROI) of RFID. In a recent *Wall Street Journal* article (circa February 15, 2007) entitled "Wal-Mart's Radio-Tracked Inventory Hits Static," the

H.-D. Haasis et al., *Dynamics in Logistics*
DOI: 10.1007/978-3-540-76862-3,

reporter paints a rather dismal picture of RFID: grumbling suppliers, high costs, no return on investment, and reluctant retailers (McWilliams, 2007). In a similar article in *Computerworld* (June 15, 2006) the author states, "A few years ago, the IT industry was abuzz with the wonders of radio frequency ID technology. It was set to revolutionize everything about business process management. But in the past year, there have been few advances in RFID that would put it center stage" (Gittlen, 2006). If we are to believe the popular press, it would appear then, that the use of RFID is losing traction due to a lack of proven business value creation.

Contrary to these negative reports on RFID, we believe there is a business case for RFID. The use of RFID throughout the supply chain can provide manufacturers, suppliers and retailers unprecedented visibility. This newfound visibility provides many real and potential benefits, such as reducing out of stocks, properly executing promotions, and reducing theft. In this paper, we propose a model of RFID assimilation which suggests that the creation of business value is dependent upon the depth of assimilation (extent of use). The model conceptualizes RFID assimilation as occurring in three waves: the first wave of the model is Technology Deployment, the next wave of Data Understanding, and, lastly, the final wave is Business Value Creation.

2 Model of RFID Assimilation

The three phases of RFID assimilation include technology deployment, data analytics, and business value (see Fig. 1). The first phase involves creating an RFID-enabled environment, which includes the equipment, such as the readers, antennae, and tags. Deploying RFID technology alone will not produce business value. Instead, the technology produces data that, with proper analysis, produces information and insight that can produce business value. As we learn more about the data that is needed, the technology deployment can be improved; the more we learn about the business value that is being produced, the better idea we have about the data needed. Thus, the feedback loops in the model illustrate the need to update the technology deployment and data analytics, as necessary. In this paper, we focus on the last phase – Business Value – but keep in mind that one cannot get to the business value phase without passing through the two earlier phases (briefly described in the next sections).

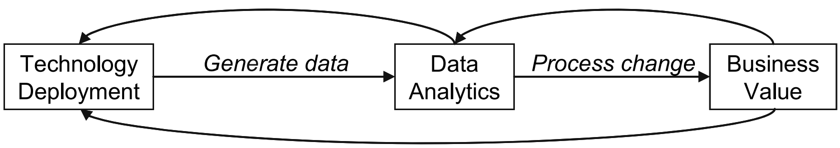


Fig. 1 RFID Model of Assimilation

2.1 Phase 1: Technology Deployment

Unlike other common automatic identification technologies, such as barcodes, RFID uses radio waves to communicate. Compared to barcodes, this offers several advantages, such as: (1) RFID tags are not constrained by a need for line-of-sight; (2) many RFID tags can be read simultaneously; (3) RFID tags can store information regarding the individual item; (4) RFID works well in a dirty environment; and (5) RFID tags can potentially be written multiple times, making them reusable data containers (Raza et al., 1999 and Shepard, 2005).

In its simplest form, an RFID system consists of a tag (attached to the product to be identified), an interrogator (i.e., reader), one or more antennae attached to the reader, and a computer (to control the reader and capture the data). At present, the retail supply chain has primarily been interested in using passive RFID tags. Passive tags receive energy from the electromagnetic field created by the interrogator (e.g., a reader) and backscatter information only when requested for it. The passive tag will remain energized only while it is within the interrogator's magnetic field. Unlike passive tags, active tags have a battery on board to energize the tag. Because active tags have their own power source, they don't need a reader to energize them; instead they can initiate the data transmission process. On the positive side, active tags have a longer read range, better accuracy, more complex re-writable information storage, and richer processing capabilities (Moradpour and Bhuptani, 2005). On the negative side, due to the battery, active tags have limited lifetime, are larger in size and are more expensive than passive tags.

Companies currently have several options when implementing RFID. They can use static RFID portals which create a set read field at discrete choke points such as a dock door or sales floor door. Companies may also use mobile devices such as forklift readers or handheld readers. The choice of technology is dictated by the type of data one wants to capture.

2.2 Phase 2: Data Analytics

What RFID data does a company need to create business value? The current deployment of RFID includes only a small portion of the overall supply chain (specifically, retailer distribution center and store backrooms) using static portals. Trends, however, suggest infusion across the entire supply chain and a move to more mobile devices and fewer static portals.

Given the current deployment, it is important to understand what data is available. A typical retailer distribution center would contain readers at the receiving doors, on the conveyor, and at the shipping doors (see Fig. 2). Read points in a generic store include receiving doors, sales floor doors, backroom, and box crushers (see Fig. 3).

An RFID tag contains an Electronic Product Code (EPC) which is a family of product codes, including such things as the serialized global trade identification

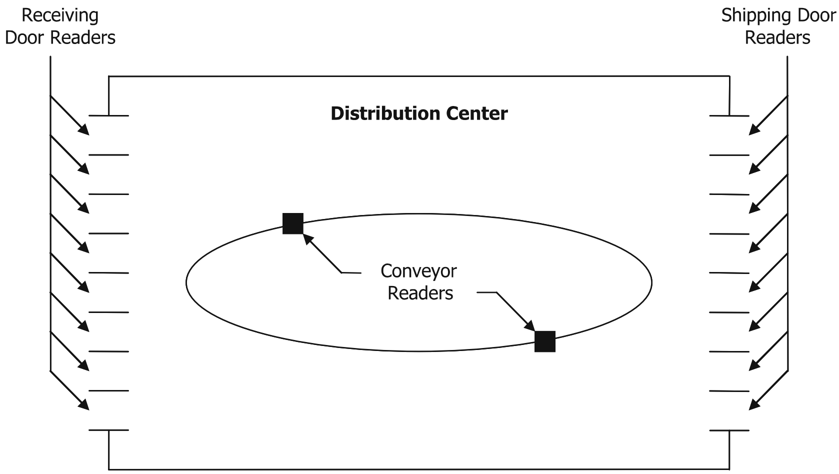


Fig. 2 Case Movement through a Distribution Center

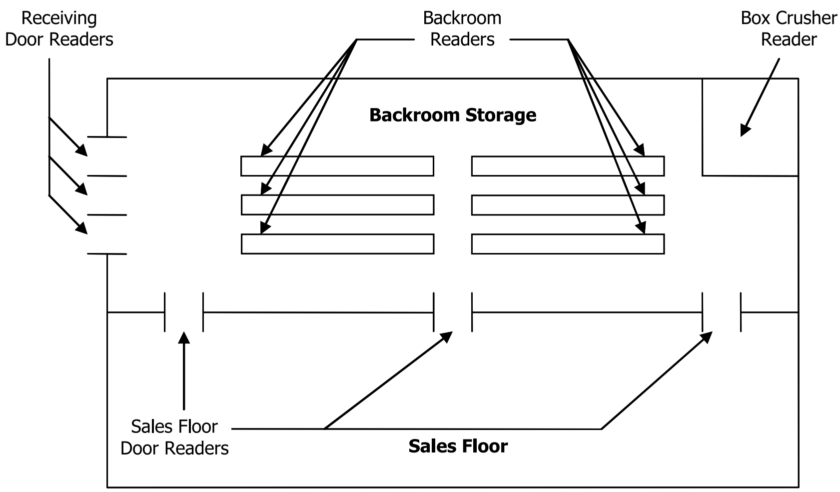


Fig. 3 Case Movement through a Store

number (SGTIN) and the serialized shipper container code (SSCC), among others. SGTIN is the standard identifier for cases of products and allows each case to be uniquely identified with a serial number. In contrast, the typical Universal Product Code (UPC) only provides information about the product but does not uniquely identify each case.

An example SGTIN is 0038010.150853.203. The first seven digits represent the manager number or company prefix. The next six digits represent the item number or object class. The number following the second period is the serial number. In this example it is only three digits, but the serial number can be up to 12 digits.

Table 1 Sample RFID Data

Facility	EPC	Date/time		Reader
DC 123	0038010.150853.203	08-04-07	23:15	inbound
DC 123	0038010.150853.203	08-09-07	7:54	conveyor
DC 123	0038010.150853.203	08-09-07	8:23	outbound
ST 987	0038010.150853.203	08-09-07	20:31	inbound
ST 987	0038010.150853.203	08-10-07	1:12	backroom
ST 987	0038010.150853.203	08-11-07	15:01	sales floor
ST 987	0038010.150853.203	08-11-07	15:47	sales floor
ST 987	0038010.150853.203	08-11-07	15:49	box crusher

When the data is initially received, it will not necessarily be in a user friendly form. The first step of the data analytics is to filter and cleanse the data. The data may show many different reads for one case or pallet of product. Once the data has been filtered and cleansed, it should be integrated with the existing systems of the organization and then interpreted to decipher what the data means. Once the data is interpreted and understood, then actions may be taken that lead to the business value phase. An example of cleansed data with the four key elements of a read – EPC, facility, date/time of read, and reader location – is shown in Table 1.

2.3 *Phase 3: Business Value – Proven*

Ultimately, the true test for RFID is whether or not it creates business value. As described, the technology must be deployed and the data captured as precursors to creating business value. Although initial deployments are limited (partial supply chain, not all suppliers, not all products, etc.), benefits are already being found. Additionally, there are several potential benefits waiting to be tested. The real and potential benefits of RFID for the major supply chain participants – retail stores, distribution centers, and manufacturers – are summarized in Table 2 and examined

Table 2 Business Value for RFID: Real and Potential

	Retailer	Distribution Center	Manufacturer
Proven business value	• out of stocks • manual orders • promotion execution	• receiving • electronic proof of delivery	• asset management • product life cycle tracking
Potential benefits	• shrinkage • inventory accuracy • product rotation • recall management	• recall management • product rotation • shipping accuracy	• recall management • shipping accuracy • reusable containers • targeted use of merchandisers

in the following sections. As the realized benefits have been segment specific, there is potential for future benefits to cross multiple segments of the supply chain. The realized benefits are presented by major supply chain participant category while the potential benefits are discussed more holistically.

Retailer: Out of Stocks

Out of stocks is a major problem for retailers, suppliers, and consumers. Nationally, the average out of stock rate is about 8% (Corsten and Gruen, 2003), which means that about 1 out of every 12 items on a consumer's shopping list is not on the shelf. The result: lost sales and unhappy customers. Although retailers have tried to improve out of stocks for years, the 8% rate has remained relatively stable (Corsten and Gruen, 2003). One of the anticipated benefits of RFID was the reduction in out of stocks. To examine this benefit, in 2005, the University of Arkansas conducted the largest out of stocks study ever. The study included thousands of products in 24 stores of various Wal-Mart formats over a 6-month period. The initial results, released late 2005 (see Hardgrave, Waller, and Miller, 2005), were both positive and extremely encouraging. In the test stores (those that were RFID-enabled), out of stocks dropped 26% during the study period. A follow-up analysis based on the velocity of sales (i.e., daily sales per item) indicated that for those items that sold between 0.1 and 15 units per day, RFID was responsible for a 30% reduction in out of stocks (Hardgrave, Waller, and Miller, 2006). In some categories, such as those selling between 7 and 15 units per day, out of stocks were reduced by more than 60%! Ultimately, this translates into about a 1% sales lift for the retailer and 0.8% sales lift for the supplier. Is 1% important? Last year, Wal-Mart gross sales were $344.6 billion.

Retailer: Manual Orders

In retail stores, store associates will often create a manual order for a product if it cannot be found in the backroom (although it may be there). Unnecessary manual orders affect perpetual inventory accuracy (as the product inventory count is often set to zero before placing a manual order). Unnecessary orders also send false signals up the supply chain about the current demand for product, thus creating the bullwhip effect (Lee, Padmanabhan, and Whang, 1997). The bullwhip effect has adverse effects on both the supplier and the retailer. With RFID, Wal-Mart has reduced the number of unnecessary manual orders by 10% (Sullivan, 2005).

Retailer: Promotion Execution

RFID has proven to be an effective tool for improving promotions (Collins, 2006; Murphy, 2005). In February 2006, Procter & Gamble launched a new product and promotional campaign for the Gillette Fusion razor. As a part of the promotional campaign, Procter & Gamble tagged the Fusion promotional displays in addition to tagging the cases and pallets. They were able to determine if the retailer had placed the promotional display on the floor in a timely manner to correspond to the promotional launch. Additionally, they could assess whether a store had restocked the razors on the sales floor. By having this information, Procter & Gamble was able to either send their own personnel to the retailer or prompt the retailer to restock or put the promotional display on the sales floor. According to Collins (2006), "… improving promotions proved to be an even greater 'sweet spot' for the technology within the company."

Distribution Center: Receiving

Early evidence suggests that RFID can reduce the amount of time to receive product at a warehouse (Katz, 2006). Instead of scanning each case of product individually with a barcode scanner, RFID tagged product can be read automatically at a dock door portal. Gillette reported a reduction from 20 seconds to 5 seconds in pallet receiving at their distribution center due to RFID (Katz, 2006). The process of receiving was not drastically changed (i.e., forklifts unloaded the product as before). The only change was eliminating the need to manually scan the product. Thus, the process became more efficient.

Distribution Center: Electronic Proof of Delivery

RFID can reduce the errors in receiving (in addition to increasing the speed of receiving as previously discussed) via the RFID-enabled process referred to as electronic proof-of-delivery (or ePod) (Mason et al., 2006). With barcode, mistakes are made by misidentifying the quantity and type of product. For example, a pallet of 48 cases of shampoo consisting of 24 regular and 24 scented may be received at the distribution center. The receiving clerk mistakenly identifies the pallet as 48 cases of scented shampoo. Thus, the inaccurate receiving creates an inventory inaccuracy for the receiving company and a perceived overage/underage situation for the supplier.

Manufacturer: Asset Management

With RFID, manufacturers have the ability to better manage their movable assets, such as forklifts, equipment, expensive tools, etc. For example, the University Medical Center in Tucson, Arizona has implemented an RFID asset tracking system for all 2,300 pieces of mobile medical equipment. The location of each piece of equipment within the 8-story hospital is transmitted wirelessly and can be displayed on a virtual map (Philips Launches …, 2006). Now, equipment can be found anytime anywhere; thus, avoiding time loss in locating equipment in emergency situations, for example. Obviously, similar projects in other industries, such as tracking forklifts in a warehouse, are feasible and potentially valuable.

Manufacturer: Product Life Cycle Tracking

An example of product life cycle tracking is the use by Michelin to keep track of large industrial tires to know when they need to be retreaded (Murphy, 2005). RFID can be used to monitor the life cycle of the product, thus improving safety for the truck drivers and enhancing the relationship with the customer.

2.4 Phase 3: Business Value – Potential

Recall Management

With barcode, retailers have no idea where (i.e., which stores) the recalled product is located. Thus, they are forced to look at each store and perhaps pull the product from the shelves at each store. With RFID, cases could be tracked to particular stores. Similar to retail stores, distribution centers and manufacturers would be able to track product at the case level with RFID. A recent occurrence was the pinpointing of the tainted raw material product shipment from China to manufacturing plants which produced poisonous dog and cat food. As recent as April 19, 2007, the Food and Drug Administration has expanded the recall of pet foods due to the tainted wheat gluten and rice protein concentrate imported from China (Pet food recall, 2007).

Product Rotation

RFID data can provide the visibility needed to know whether or not product is being rotated properly (using the first-in-first-out (FIFO) method, for example). Consider the wide variety of products that have expiration dates or are perishable. This insight can, thus, help ensure proper rotation in both the retail store and the distribution center to get the product which expires/perishes the soonest, to the front of the shelf/display.

Shipping Accuracy

Distribution centers and manufacturers often make mistakes by loading product on the wrong truck. With RFID, the system could send an alert (visible, audible, etc.) to the person loading the truck that a mistake was made. This alert could save a company money from shipping and reshipping the same items as well as enhance inventory control.

Inventory Accuracy

The key to good forecasting and replenishment models for the retailer is accurate inventory. Unfortunately, inventory counts (commonly called perpetual inventory) which are calculated by subtracting point of sale from the amount of product received, are inherently wrong. A recent study (Raman, DeHoratius, and Ton, 2001) found that as much as 65% of all perpetual inventory counts are wrong. Thus, forecasting and replenishment models are based on data that is wrong 65% of the time! RFID could provide much better data to determine accurate inventory counts.

Shrinkage

Shrinkage is any loss of product as it goes through the supply chain whether through damage, spoilage, loss or theft. Although RFID cannot eliminate all shrinkage, it can help reduce it with the visibility it provides. For example, a company may know what 5% of its product is disappearing in the supply chain, but have no idea where (retail store or distribution center). RFID would provide the insight into where it was last seen so that a company could determine (for example) that the product made it to the store but never to the shelf (which may suggest that employees were stealing the product). Another example could be of a manufacturer of a seasonal product that shipped 10 cases of the product, but only 9 cases appeared at the store. With bar code tagging, once the product leaves the manufacturer the location of the product within the supply chain is unknown. However, with RFID, quick identification and location of the lost product could provide time and cost savings and ultimately business value.

Reusable Containers

Many manufacturers ship their products in reusable containers, such as milk crates, reusable totes, bins, rolling shelves, etc. Unfortunately, these containers have a tendency to disappear. Each container could be tagged with RFID, thus similar to the shrinkage discussion the manufacturer would know where the container was last seen.

Targeted Use of Merchandisers

Many suppliers send merchandisers to retail stores to bring product from the backroom to the sales floor. Currently, merchandisers methodically work from store to store (perhaps having responsibility for 20 or more stores). Some stores will have merchandise in the backroom and a need to work it to the shelf, others will not. Visits to those stores that do not need it represent a waste of time for the merchandiser. With RFID, the merchandiser could remotely determine (from their home office, for example) if product was in the backroom and if it needed to be placed on the sales floor shelf.

3 Conclusion

This paper has identified three stages of RFID assimilation: technology deployment, data analytics, and business value. The primary focus of the paper was to illustrate proven business value and potential value that companies could glean from using RFID. We have demonstrated the proven business value of RFID by giving specific examples for retailers, distribution centers, and manufacturers that have previously been experienced. Additionally, we identified seven potential benefits for the retailer, distribution center, manufacturer or some combination of the three that is possible with RFID. Some business value from RFID used to improve promotion execution, asset management and electronic proof of delivery has already been achieved by many companies (e.g., Wal-Mart, Gillette, Michelin, and Procter & Gamble). We assert that the identified and unidentified (due to future technology improvements not yet available) potential benefits will expand the business value companies throughout the supply chain receive from RFID.

References

Collins, J., 2006, “P&G Finds RFID ‘Sweet Spot’,” *RFID Journal*, May 3. Available at: http://www.rfidjournal.com/article/articleview/2312/1/1/

Corsten, D., and Gruen, T., 2003, “Desperately Seeking Shelf Availability: An Examination of the Extent, the Causes, and the Efforts to Address Retail Out-of-Stocks,” *International Journal of Retail & Distribution Management*, 31 (11/12), 605–617.

Gittlen, S., 2006, “The Failure of RFID,” *Computerworld*, June 15. Available at: http://www.computerworld.com/action/article.do?command=viewArticleBasic&articleId=9001194&taxonomyId=16

Hardgrave, B., Waller, M. and Miller, R., 2005, “Does RFID Reduce Out of Stocks? A Preliminary Analysis,” White Paper, Information Technology Research Institute, Sam M. Walton College of Business, University of Arkansas. Available at: http://itrc.uark.edu/research/display.asp?article=ITRI-WP058-1105

Hardgrave, B., Waller, M. and Miller, R., 2006, “RFID’s Impact on Out of Stocks: A Sales Velocity Analysis,” White Paper, Information Technology Research Institute, Sam M. Walton College of Business, University of Arkansas. Available at: http://itrc.uark.edu/research/display.asp?article=ITRI-WP068-0606

Lee, H.L., Padmanabhan, V., and Whang, S., 1997, “The Bullwhip Effect in Supply Chains,” *Sloan Management Review*, 38 (3), 93–102.

Katz, J., 2006, “Reaching the ROI on RFID,” *IndustryWeek*, February 1. Available at: http://www.industryweek.com/ReadArticle.aspx?ArticleID=11346

Mason, M., Langford, S., Supple, J., Spears, M., Lee, R., Dubash, J., Roth, L., Subirana, B., Sarma, S., and Ferguson, C. “Electronic Proof of Delivery,” 2006, EPCglobal, Available at: http://www.epcglobalinc.org/news/EPODVignetteApprovedV2.pdf

McWilliams, G. Wal-Mart's Radio-Tracked Inventory Hits Static. *The Wall Street Journal*, February, 15, 2007, B1.

Murphy, C., 2005, "Real-World RFID: Wal-Mart, Gillette, and Others Share What They're Learning," *InformationWeek*, May 25. Available at:http://informationweek.com/story/showArticle.jhtml?articleID=163700955&_loopback=1

"Pet-food recall expanded, Chemical turns up in second ingredient," *Richmond Times Dispatch*, April 19, 2007, Available at: http://www.timesdispatch.com/servlet/Satellite?pagename=RTD/MGArticle/RTD_BasicArticle&c=MGArticle&cid=1173350832895

"Phillips Launches RFID Asset-Tracking for Hospitals," *Medical Technology Business – Europe*, December 14, 2006. Available at: http://www.mtbeurope.info/news/2006/612024.htm

Raman, A., DeHoratius, N., and Ton, Z. 2001, "Execution: The Missing Link in Retail Operations," *California Management Review*, 43 (3), 136–152.

Sullivan, L., 2005, "Wal-Mart RFID Trial Shows 16% Reduction In Product Stock-Outs." *Information Week*, October 14. Available at: http://informationweek.com/story/showArticle.jhtml?articleID=172301246

General Aspects of Dynamics in Logistics

Review of Trends in Production and Logistic Networks and Supply Chain Evaluation

Paul Maropoulos[1], **Maxime Chauve**[2], **Catherine Da Cunha**[2]

[1] Mechanical Engineering Department, University of Bath, United Kingdom
[2] Institut de Recherche en Communications et en Cybernétique de Nantes (IRCCYN), Ecole Centrale Nantes, France

Abstract This paper is a state of the art review of trends, theories and practices in production and logistic networks and the evaluation of supply chains. The paper outlines the progression of concepts from supply chains to the definition of production networks, summarises the trends for virtual companies and outlines recent research in logistics. The use of competence profiling is reviewed for selecting companies for a network, on the basis of core competencies. New methodologies for performance assessment of production networks are also reviewed, together with well established benchmarks arising from the lean and agile methodologies.

1 Introduction

Companies operate in a very competitive, global marketplace and in an uncertain business environment. Globalisation, the evolution of technologies, pressure of cost reduction, and customer expectations are parts of the challenges they have to deal with. Thus, there is an increasing need for flexibility and reactivity [1] and under such circumstances, cost competitiveness becomes a market qualifier. Production and service networks are evolving within those parameters using new technologies. Many organisations have responded to these pressures by restructuring or outsourcing to concentrate on their core competencies, or are planning to do so. The measurement of performance is essential for ensuring competitive operations, but has become more complex and multidimensional. Also, companies have to assess their potential suppliers and partners before building their networks.

This paper outlines the present a state of the art of trends in production and logistic networks, starting from the supply chains and progressing to production networks, including the concept of reverse logistics. The second part of this paper is dedicated to a presentation of the concept of performance evaluation in the context of the new possibilities offered by the evolution of technologies and tools for net-

H.-D. Haasis et al., *Dynamics in Logistics*
DOI: 10.1007/978-3-540-76862-3,

work and partner assessment. Finally, a brief state of the art of management theories in networks will help to establish a benchmark for the development of international competitive production network strategy.

2 From Supply Chain to Production Networks

Nowadays companies cannot exist in isolation. Therefore, a key part of the business and manufacturing strategy for a company lies in the correct selection of its network. The supply chain concept was the first instantiation of a coherent outsourcing strategy and this has recently been replaced by supply networks [2]. But other cooperation concepts can be envisaged, arising from the need for flexibility. From their survey, Garber et al. [3] identified the network design, flexible supplier relationships, simplification processes and supply chain connectivity as the four key factors to reduce cycle time which is a major aspect to achieve production flexibility. Supply chain flexibility and stability as well as the constraints associated with achieving them are major criteria for the performance improvement of supply chains and networks.

2.1 Supply Chain and Supply Chain Management

The scope of the supply chain is hard to define because of the common use of this designation and it is commonly used to include other areas such as logistics or purchasing [4]. As a consequence, difficulties could appear when different companies work together [5] and even within a large heterogeneous company in terms of identifying key staff and developing business competencies. The following definition for supply chain is proposed: a set of more than three companies from supplier to customers, which could include factories, warehouses, retailers and end customers depending on the complexity and the maturity of the chain. The companies in a supply chain are closely linked by upstream and downstream flows of materials, information, finances and services [4, 5]. The Supply Chain definition was born as a consequence of the decision of enterprises to make-or-buy a large range of their components, including major components, and thus having a direct need to manage these new flows [6].

To define Supply Chain Management (SCM), one can use the definition of management by Forrester [7] which is considered as still being highly relevant: “Management is on the verge of a major breakthrough in understanding how industrial company success depends on the interactions between the flows of information, materials, money, manpower, and capital equipment. The way these five flow systems interlock to amplify one another and to cause change and fluctuation will form the basis for anticipating the effects of decisions, policies, organizational forms, and investment choices.” Moreover, the supply chain manager has to strive to achieve

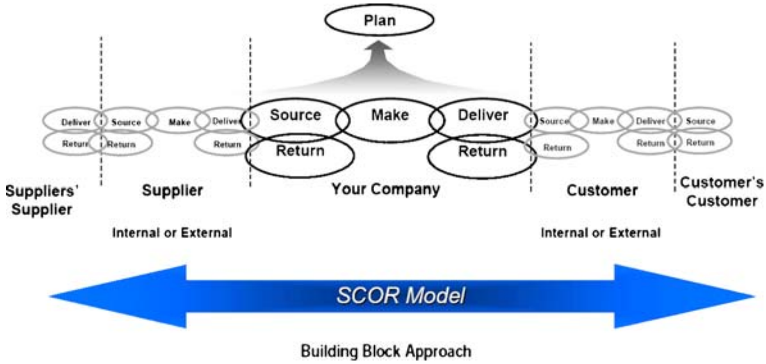


Fig. 1 SCOR Model of Supply Chain [8]

customer satisfaction. The usual actions describing SCM are "Plan, Source, Make, Deliver", but according to the Supply Chain Organisation Reference model (SCOR model) there are five core processes involved as shown in Fig. 1.

The supply chain is the most widely used concept to define a network involved with the selection of suppliers and having a distribution strategy. Within a supply chain, a company is involved in collaboration with other companies, which could be as long and as stable as the quality of the inter-relations and the evolution of the companies' needs.

The supply chain usually evolves to a more complex type of cooperation as a consequence of the evolution of the five flows mentioned above. So after the company is introduced as a link in a supply chain, the concept of supply networks appears as the latest evolution of the common network [2]. Harlan et al. [9] define the supply network as follows: "Supply Networks encompass the mesh and the complexity of networks involving lateral links, reverses, loops, and two way exchanges, and include a broad, strategic view of resource acquisition, development, management and transformation".

2.2 Integration, Virtual Integration

The integration is another approach to build a network. However, the term "integration" is commonly used in the literature, and we can identify several types of integration.

The oldest types of integration are the "horizontal and vertical integration". Horizontal integration indicates the development of a firm in its marketplace by taking initiatives to achieve the control of its rivals. Vertical integration is the organization of successive production processes within a single firm, which is the direct or indirect owner of these activities and keeps the profits from production after compensating other claimants [10]. Backward integration is used to describe the situation when the firm owns some or all activities upstream of its core operations. This might be the raw materials, the suppliers of parts and components, and perhaps even the transportation equipment used to deliver the raw materials. Vertical integration can thus encompass the downstream activities (manufacturers, retailers) from a firm's main activities (forward integration), but first and foremost includes the upstream supplier activities of a firm (backward integration). The advantages of vertical integration are obvious [11]: more coordination, easy share of information, a stable relationship, no more negotiation, common design, and the ability of the buyer to use the knowledge and technologies of the supplier. All those aspects can motivate vertical integration. Moreover, a firm can also avoid negotiating with a competitor in terms of purchasing [12]. However in terms of investment, vertical integration utilises a large part of the available capital of the company and the buyer has to deal with the resistance to change and the mistrust from the employees of the companies involved. Authors also warn about quality and capacity risks [12, 13]. With vertical integration, a firm may loose the ability to ensure that its suppliers must be competitive. Furthermore, the capital tied into the vertical integrated company reduces the revenue available for R&D within the complete network, and especially within the integrated suppliers.

The last integration concept is the virtual integration defined by Wang et al. [14] as "the substitution of ownership with partnership by integrating a set of suppliers through information technology (IT) for tighter supply chain collaboration". A major advantage of virtual companies is that they can be set up quickly, they need low capital investment and they can also be easily disbanded when the opportunity has passed. The virtual integration is certainly the most flexibly integrated network and at the same time the most unstable. In order to builb a virtual enterprise it is crucial to establish trust between the partners. Therefore, the contractual arrangements must include rules for information and knowledge sharing and define a win-win scope for all partners [15].

2.3 Joint Venture

The Joint Venture is another entity created by two or more companies, sharing risks, investments and revenues. This kind of integration is usual in energy (Petrol) but also in innovative sector (Sony Ericsson). A joint venture is frequently appropriate between companies from different countries. It could be used to penetrate the markets in counties that employ protective strategies in terms of the acquisition of technologies and knowledge of their companies. It is at the same time a good opportunity to discover a new market and understand a new culture with the help of an experienced partner. The risk in joint ventures is the departure of one of the partners when the company has reach its goals such as knowledge acquisition or gaining understandings and confidence in the market [16]. Hence, at the launch of a joint venture it is important to define appropriate rules and procedures for its governance and interdependency (a joint venture can be dependent, interdependent or independent of the partners) and to build the organisation and decide its management. As a consequence of the complex issues involved, around 50% of today's joint ventures result in failure [17].

2.4 Cluster

The concept of "cluster" was introduced by Porter [18] to indicate a geographical concentration of interconnected companies and institutions. The institutions could be research centres, universities or other governmental organisations. Thus, companies located within a cluster may benefit from innovations, access to employees with the right skills and from gaining a visibility they could not reach otherwise [19]. Companies can also find competitive suppliers as well as customers especially in business-to-business interactions. All the companies of a cluster focus on the same market and are seen as stakeholders in this market [6]. The mix of innovation, competition and cooperation within a cluster can create a real dynamic environment and improve the motivation of employees.

Cluster can relate to new technologies such as the Silicon Valley and also exist in more traditional sectors, such as the wine industry. The companies share the risks of the sector but are also exposed to the quick transmission of competitive pressures. The choice of a cluster for a company is not a neutral decision because of the implantation and benefit of potential subventions, which in turn can oblige the company to remain in the cluster for a certain time. Hence, together with the partner assessment, a company has to assess its membership of a cluster though its key aspects: competition, relationship, innovation, supplier, infrastructure, and market [20].

2.5 Production Networks

The last development in our review of networks is the production network [6]. Production networks are created by the cooperation between several companies which could usually share the same market and target the same customers and suppliers. Redundancies in capabilities are viewed as opportunities to exchange information and resources. On one hand companies create a network with suppliers and in the other hand they reinforce horizontal partnerships. Production networks are characterized by long term and stable network relationships.

2.6 Reverse Logistics

As presented previously, reverse logistics is a current concept in logistics. The definition of reverse logistics by The Council of Logistics Management is as follows: "the process of planning, implementing, and controlling the efficient, cost effective flow of raw materials, in-process inventory, finished goods and related information from the point of consumption to the point of origin for the purpose of recapturing value or proper disposal." [21].

Many barriers prevent the so viewed unnatural direction of flows [22]. The main reasons which lead to return flows are: unsold items (retailers, journal), return of

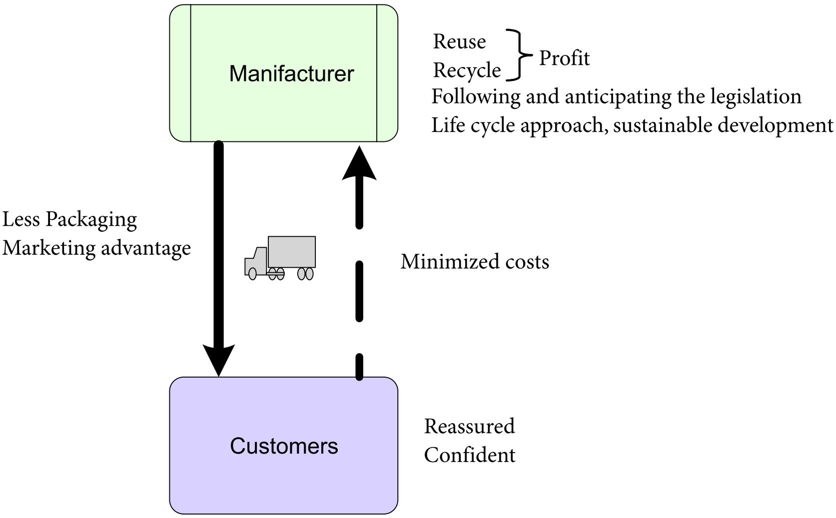


Fig. 2 Impact of managed return logistics

empty packaging, return from dissatisfied online buyers, warranty returns, end of life or expired products but also recycling. However, the most important barrier to this opposite distribution stream is the lack of information. Thus, the concept of reverse logistics seeks to manage, from a strategic point of view, the return flows to gain competitive advantage. Recycling and reuse of materials could be a source of profit [23] and the relevant legislation regarding re-use is likely to become more restrictive. Moreover, a good return management process may offer a strategic advantage, as it improves customer confidence in the quality of the product and enhances the customer perception of services related to the product (Fig. 2).

The implementation of a return logistics policy should be done simultaneously with performance assessment to pilot the process and deal with any lack of data. The choice of the criteria is particularly critical because it influences the behaviours of partners and customers alike [24, 25]. Companies can use the experience of no-for-profit firms [26] to collect and treat the return products. A consideration that has become very important recently is the life cycle approach [27]. Under lifecycle considerations, reverse logistics can be considered as a loop [28] that takes place in a sustainable development context, within a global supply chain.

2.7 Competence Profiling for Company Identification and Appraisal

In response to the virtual enterprise and cluster paradigms, a Competence Profiling Methodology has been developed to address the identified requirements of rapid competence identification, search, and matching of the core competencies of potential company partners [29]. The methodology consists of four main stages [30]:

1. Competence data collection: An on-line questionnaire has been development to assist in collecting and updating competence related information. Companies can load their competence information by registering on the 'Competence Profiling' web-site. The questionnaire aims at revealing the key skills and capabilities of engineering SMEs by focusing on both hard and soft factors, such as:
   - Company key indicators e.g. type of business and key markets
   - Awarded standards
   - Key manufacturing and management process capabilities
   - Core skills
   - Business philosophy
2. Normalizing: To ensure comparability among the various companies and correct use of the on-line tools, training and validation of the provided company data is provided. The 'normalising' stage also allows estimation of the quality of the identified competences. Since some of the information requires subjective judgement, the assistance of experts is considered essential within the normalisation process. Visits to the company premises by experts are organized and further-on validation and assistance is provided in capturing and assessing key competences.
3. Making competence information available for searching: In this stage competence information is stored in the relevant database which enables users to search for the appropriate set of business skills and capabilities to undertake a customer specific project. For this purpose the web-based Competence Search module has been developed.
4. Partnership formulation: In this final stage, companies with appropriate complementary competences are identified and matched to generate viable dynamic partnerships in response to a project need.

The methodology has been successfully tested and validated within the automotive and general engineering sectors [31, 32]. A new and improved development allows implementation and testing in the product development process of the aerospace and defence industries [30]. These sectors are technology driven and focused on demanding and rigorous quality standards.

3 Performance Assessment of Supply Chains and Networks

To move in a positive direction and to assess the "as is" condition of a network, a method for the evaluation of performance is a key aspect of judging business competitiveness. Before outlining a pragmatic approach of performance criteria for performance evaluation, the concept of performance is outlined in the context of supply chains and networks.

3.1 The Concept of Performance

The concept of performance has undergone the same evolution as the methods of production and the market dynamics, from a single measure of the costs associated with production to the assessment of multilevel criteria at a global basis.

One can consider a company as an organization which consumes resources (means, time and capital) to satisfy the needs of the society (prospective customers or any other societal group) by creating value (creation of goods meeting the market needs) and has costs associated with the consumption of resources to create this value. A criterion for economic performance can be defined as the ratio of the consumed resources and the value created. However, to increase performance, a company should not only try to maximize the value created or minimize the cost, but to address both value and cost so as to create useful value [33, 34].

This simple approach to performance is not sufficient to deal with the complexity of today's supply chains and networks. An approach to understanding and measuring performance in global supply networks may be based on the following aspects: pertinence, effectivity, efficiency and effectiveness [35]. These attributes should be appreciated in terms of costs and value in the longer term within the lifecycle of the system. Figure 3 shows a graphic representation of performance as corresponding to the volume of a tetrahedron that is constructed using the above defined attributes.

To consider the dynamic aspects of performance, Berrah [37] proposes three temporal approaches to performance: the first approach is a function of the capacity of adaptation of the production equipment to the market, the second is

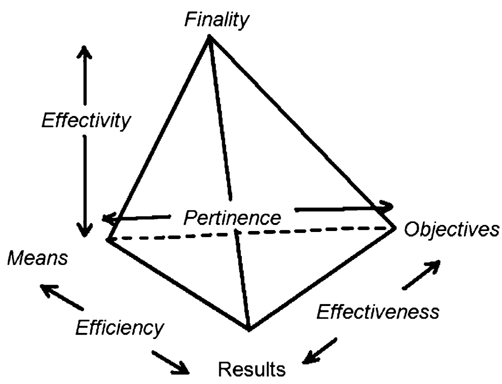


Fig. 3 Tetrahedron of the performances of the systems of production [36]

centred on the lifecycle of the product (design, prototyping, maturity, decline, recycling, and destruction) and the third is centred on the concept of sustainable development.

3.2 An Overall view of Performance Criteria

In literature, we can find a lot of criteria and methodologies that can be used for the assessment of suppliers. Figure 4 presents a global view of the criteria used not only to select a supplier but also to evaluate current networks. The criteria are organised in six areas. The management of the suppliers, the care for the environment, the potential flexibility and the communication abilities of the network are more recent areas which complement the traditional production and strategy approach to performance and competitiveness evaluation. Figure 4 is mainly based on a survey by Kannan and Tan [38] and incorporates views and ideas from other researchers [39, 40, 41, 42, 43, 44, 45].

The growing environmental awareness is a means of preventing litigation risks and offers additional benefits arising from an environment-aware reputation of the firm in the market place [46]. There are two last points to keep in mind during the design of a network. The first is the impact of the network on the customers, as it relates to a key goal of a company to satisfy customers' expectations and establish how customer practices, such as lean methods [47], can influence the network. The second point relates to the impact of the network on the operation of each one of the associated companies [48].

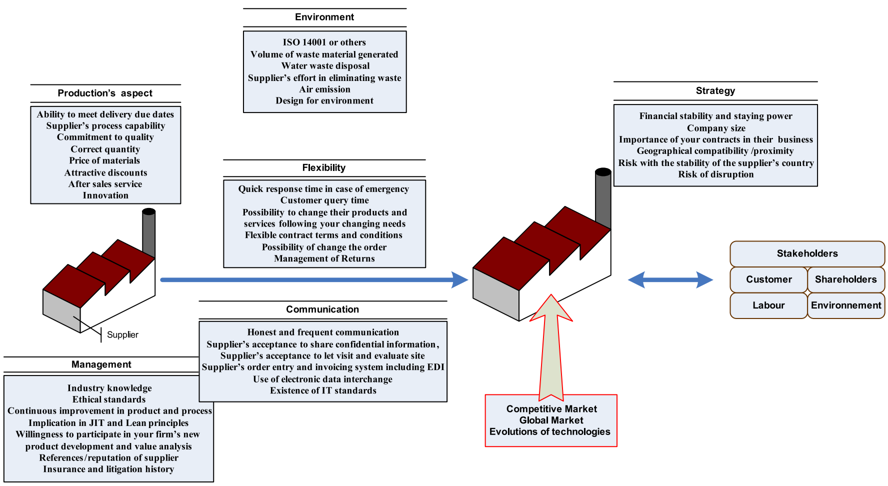


Fig. 4 An overall view of criteria to assess suppliers

3.3 Evaluation

Every aspect of performance identified must be measured using indicators of performance (IP). According to the Association Française de Gestion Industrielle [49], “an IP is a quantified data which measures the exposed criteria of whole or part of a process or a system, compared to a standard, a plan or an objective, determined and accepted within the framework of a strategy of company”. The choice of the performance indicator and its implementation are as important as the choice of the criteria. Performance measurement must be done where the activities take place [50]. Also, the performance evaluation methods must be easily comprehensible and controllable by the people who will use them [50]. The indicators deployed could be as complex as the nature of the performance that is measured. Finally, the “measurement” should be developed into an “evaluation”. This means to compare performance with the objectives and interpret rate of success by someone who is in charge or/and has knowledge of the process, as shown in Fig. 5. It must be remembered that all the processes of measurement and evaluation usually modify the behaviours of the actors so it is important to keep a global view and to have a good interpretation of results. This is a critical point because the potential for adding value by performance measurement depends on deploying the correct procedures for evaluation [51].

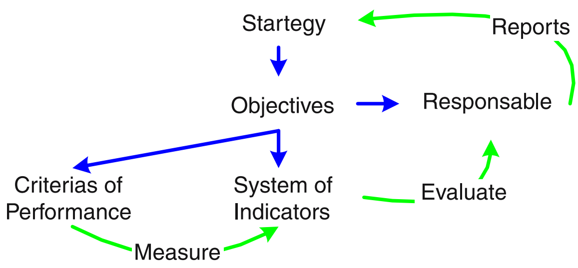


Fig. 5 The process of evaluation

3.4 The Major Role of Communication in an Assessment Process

The aspects of performance communication can be thought off as either being upstream or downstream. In terms of downstream communication, the measurement of performance is reported to the manager piloting the evaluation. In fact, an important aspect of a system of measurement is to achieve effective communication of criteria and results throughout the enterprise [52]. The measurement of performance can also be considered as a downstream communication tool with all the stakeholders of the company, especially the employees. The simple act of measurement will influence their behaviours through motivation and one needs to consider their job satisfaction. Any assessment has to be done carefully and with some explanation in order for it to be accepted and avoid human resistance.

The two major communication vectors for performance evaluation are the French "tableau de bord" and the balanced score card. Chiapello and Lebas [53] identify four targets for the "tableau de bord" method: the piloting manager, the superior through reporting, other involved managers and the others stakeholders as sharing data. The balanced scorecard presents a more strategic approach of the business unit and is covering the following perspectives: financial, customer, internal-business-process and learning and growth [54]. In all cases, the way a company chooses to communicate the evaluation of performance represents its broader strategy in terms of business development and personnel management.

The upstream aspect of communication relates to the collection of data. It is well known that managers face considerable difficulties in obtaining valid and accurate data for performance evaluation from parts of the company or from its suppliers. Similarly, it is reported that lack of information is the biggest barrier in reverse logistics [22]. The quality of data is also important. So the upstream communication is essential to break the barrier between decisional and operational areas or to demonstrate transparency and trust within the partnership.

3.5 Real Time Networks Evaluation Technology: The Radio Frequency Identification (RFID)

The Radio Frequency Identification (RFID) is a technology which allows automatic identification through a tag from a small chip to an RFDI reader in a wireless manner. When compared to the old methods such as bar-coding, RFID can be considered as a technology compatible with the global positioning system (GPS) and ERP and that can use the internet to realize transparent and real time information flows [55] and to avoid the bullwhip effect. According to references [56] and [57], Table 1 presents the advantages and limitations of RFID.

Table 1 Advantages and limitations of RFID

Advantages	Limits	Future
• Real Time data • Visibility of inventory • Visibility of VMI • Labour Cost Reduction • Faster Pick and Pack • Information Sharing • Reduce Stock Out • Reduce Shrink • Delivery Reliablility	• Cost (chip, software) • Risk of hacking • Reluctance to share Data • Different technology	• Pressure of manufacturer and domino effect • Standardisation (ECP) • Linked with ERP • Safety tag • 0.05$ per tag

As Wall Mart did with its suppliers, the pressure of large manufacturing companies can increase the use of RFID within supply chains and networks. The market expects tag costs at 5cts [58], but currently the cost depends on the localisation. Electronic Product Code (EPC), the family of coding schemes, can be considered as the standard tag, even if China wants to develop another one. The use of RFID may present safety risks and the reluctance of being totally transparent in real time could be a safety limitation but it is an opportunity to collect real time data to build a database. It shifts the willingness to share data within a network from a communication aspect to a real time control aspect. Hence, the decision to implement RFID needs to be taken at the start of the network design process and must be included in the collaborative negotiations and the network investment planning.

4 Established Benchmarks for Production Networks

As outlined in the previous sections, the process of performance evaluation requires the existence of clear and unambiguous objectives. The concepts of "lean", "agility" and "leagility" appear to offer such benchmark objectives.

4.1 The Lean Principles in Supply

The concept of lean production is far from new. The principles of lean thinking were developed by Japanese Industry and codified by Womack and Jones [59]. To be lean a company has to fight against "muda" that means waste, especially targeting any human activity which absorbs resources but creates no value for the customer. The strategy of lean is based on the specification of value which can best be defined by the customer. Among the ten usual tools to implement lean production are; value stream and process mapping, 5S, Kaisen, Poka Yoke, JIT and Kanban, SMED and performance measurement and improvement in relation to 6σ projects. Clearly, the lean philosophy includes the development of the supply chain. The fight against waste, the reduction of the cycle time, the stock reduction, the value stream mapping and cost reduction can directly be translated to the logistics area and be applied in relation to warehouse operation, product transportation and delivery [60]. The enterprises that have adopted lean supply chain concepts have improved product quality, reduced inventories, improved customer service and visibility and enhanced supply chain flexibility [61]. When the lean technology is combined with new enterprise integration methods, such as ERP, the results can be globally competitive operations that satisfy the expectations of shareholders and stakeholders alike. The Aberdeen Group identified the positive role played by ERP in the deployment of lean supply chains [61].

When operating within a network, a company should not seek to export operating constraints such as inventory and costs to its partners but, on the contrary, to build a lean network in which all companies can reach the lean objectives. It is reported that it is easier and less risky to achieve lean supply network operation with current, long term suppliers rather than when having to integrate new suppliers into the operation of the network [62]. A lean network is based, first and foremost, on trust and transparency. Customers and suppliers have to share information, such as costs, that they used to keep secret to use in negotiations [63]. Thus, supply chains can build new performance evaluation matrices based on the key criteria of "cost-quality-delivery".

4.2 The need for Agility

Agility is another concept often cited together with lean. But the emphasis for the application of agility is rather different to lean. Lean is predominantly beneficial in mass production as it reduces inventories and achieves doing more with less. On the other hand, agility is a concept ideally suited to mass customization and to uncertain, volatile markets. Agility has been defined as the capacity of an organisation to respond rapidly to a turbulent market, guided by the customer [64, 65] and seeking to add value to customers [66]. As a consequence, an agile supply chain must be close to the market it serves and needs to be based on a well developed IT system between all the partners of the network [65]. The assessment of performance in agile networks is hard to evaluate. Literature [64, 67] gives us some criteria such as, the use of IT for responsiveness, data accuracy, delivery speed, logistics adaptability and trust. A key aspect that influences agile behaviour is the inherent complexity of a network and this should be considered when specifying the production network [68].

4.3 Leagility

Agility is not an evolution of lean and both can coexist within the supply chain and their combined application is called "leagility". Towill and Chistopher [69] showed that lean and agile can play a combined role in the operation of a network but not in the same place and at the same time. Thus companies can adopt lean and agile in the same place/market but at a different time or at the same time but in different marketplaces depending on the fluctuation of markets, the nature of the products and the needs of the customers for customisation or mass production. Therefore, it is crucial to choose a supply chain strategy according to business needs [70] that takes into account the nature of the product and the target market [71], so as to best address the customer requirements. A well combined mixture of agility and lean production is the best way to provide products of value to the customer, at a globally competitive cost, within a flexible supply chain.

5 Conclusions

This paper has reviewed the diverse types of supply and production networks a company can join or build. Competence profiling has been identified as a methodology suitable for partner selection for network membership, on the basis of company core competencies. The assessment of performance and the use of established benchmark practices in industry that can define clear objectives for performance evaluation were also reviewed.

The necessity of having a clear strategy regarding the identification of the target market, the needs of customers and the evolution of products appear to be crucial factors for the specification of correct performance evaluation objectives. Information management through ERP, RFID and rapid information sharing and communication of performance measurement within the network are vitally important for the success of performance evaluation.

Both the lean and the agile concepts offer a range of objectives that can be used as performance benchmarks, thus driving the performance evaluation process within a production network. The choice of suitable measures is entirely based on the type of production, the complexity of the product and the characteristics of the marketplace.

References

1. Sanchez AM, Perez MP (2005) Supply chain flexibility and firm industry, International Journal of Operations and Production Management, 25 : 681–700
2. Rajamani D (2005) Intelligent Supply network, The University of Texas at Dallas, Center for Intelligent Supply Networks, Presentation
3. Garber R, Sarkar S (2007) Want a more flexible supply chain? Supply Chain Management Review Jan/Fev 2007 : 28–34
4. Larson P, Halldorsson A (2002) “What is Supply Chain? And, where is it?” The Journal of Supply Chain Management 38 : 36–44
5. Menzer JT (2001) Supply Chain Management, Sage Publications, Thousand Oaks
6. Wiendahl HP, Lutz S (2002) Production Networks, Institute of Production Systems and Logistics (IFA), Annals of the CIRP vol 51/2, University of Hannover
7. Forrester JW (1958) Industrial Dynamics: a major breakthrough for decision makers, Harvard Business Review 36 : 37–66
8. Supply Chain Council (2006) Supply Chain Operations References (SCOR Model), SCOR Overview SAPICS 2006, Peter Baker
9. Harlan MC, Lamming RC, Zheng J, Johnsen TE (2001) A Taxonomy of supply Networks, Journal of Supply Chain Management, Fall 2001
10. Riordan MH (2005) Competitive effects of vertical integration, Columbia University, Discussion Paper No.: 0506-11
11. Porter ME (1980) Competitive Strategy, The Free Press, New York
12. Heavner DL (2004) Vertical Enclosure: Vertical Integration and the reluctance to purchase from a competitor, Journal of Industrial Economics 52 : 179–199
13. Wu CL (2003) A Study of Collaborative Forecasting mechanism for Panel Maker In Taiwan TFT-LCD Industry, Master thesis, National Sun Yat-sen University

14. Wang ETG, Tai JCF, Wei HL (2006) A virtual Integration Theory of Improved Supply Chain Performance, Journal of Management Information Systems 23 : 41–64
15. Ouzizi L, Anciaux F, Portmann MC, Vernadat F (2006) A model for cooperative planning within a virtual enterprise, International Journal of Computer Integrated Manufacturing 19 : 197–209
16. Yeung DWK, Petrosyan L (2006) Dynamically stable corporate joint ventures, Automatica 42 : 365–370
17. Bamford J, Ernst D, Fubini DG (2004) Launching a world class Joint Venture, Harvard Business review, 82 : 90–100
18. Porter ME (1998) Clusters and the new economics of competition, Harvard Business Review 76 : 77–90
19. DeWitt T, Giunipero LC, Melton HL (2006) Clusters and supply chain management: the Amish experience, International Journal of Physical Distribution & Logistics Management 36: 289–308
20. Steine C, Schiel H, Mietzner (2006) Merging a Firm-centred and a Regional Policy Perspective for the Assessment of Regional Clusters: Concept and Application of a "Dual" Approach to a Medical Technology Cluster, European Planning Studies 15(2) : 235–251
21. Dale SR, Tibben-Lembke RS (1998) Going Backwards, Reverse Logistics Trends and Practices, Reverse Logistics Executive Council
22. Ravi V, Shankar R (2004) Analysis of interactions among the barriers of reverse logistics, Department of Management Studies, Indian Institute of Technology Delhi
23. Stock J, Speh T, Shear T (2002) Many Happy (Products) Returns, Harvard Business Review 80 : 16–17
24. Bernon M, Cullen J (2007) An integrated approach to managing reverse logistics, International Journal of Logistics: Research and Applications 10 : 41–56
25. Breen L (2006) Give me back my empties or else! A preliminary analysis of customer compliance in reverse logistics practice, Management Research News 29 : 532–551
26. Reyes PM, Meade LM (2006) Improving Reverse Supply Chain Operational Performance: A transhipment Application Study for not-for-profit organizations, The Journal of Supply Chain Management, Winter 2006 : 38–47
27. Chouinard M, D'Amours S, Aït-Kadi D (2005) Integration of reverse logistics activities within a supply chain information system, Computers in Industry 56 : 105–124
28. Waller DL (2003) Operations Management: a supply chain approach, 2nd edn, Thomson
29. Armoutis N, Matthews PC, Lomas C (2005) Partner Profiling to Support Agile Design, International Conference on Changeable, Agile, Reconfigurable and Virtual Production, Sept 22–23, Munich, Germany, 272–276
30. Armoutis N, Maropoulos PG, Lomas C (2006) Establishing Agile Supply Networks Through Competence Profiling, 3rd International Conference in Digital Enterprise Technology, 18–20 September 2006, Setubal, Portugal
31. Armoutis N, Bal J (2003) Building the Knowledge Economy: Issues, Applications, and Case Studies, chapter E-Business through Competence Profiling, IOS Press : pp 474–482
32. Bal J, Armoutis N (2005) E-Challenges for Manufacturing Regions and their SME's, e-Challenges 2005, Ljubljana
33. Ernult J (2005) Le pilotage de la performance, Cahiers du CEREN vol 10 : 24–31
34. Lorino P (2003) Méthodes et pratiques de la performance, Éditions d'Organisation, 2003
35. Senechal O (2004) Pilotage des systèmes de production vers la performance globale, Mémoire d'habilitation à diriger des recherches, Université de Valenciennes et du Hainaut Cambresis
36. Bescos PL, Dobler P, Mendoza C, Naulleau G (1995) Contrôle de gestion de Management, Montchrestien edn, Collection Entreprendre, Guide des techniques et de la décision
37. Berrah L (2002) Les indicateurs de performance : concepts et applications, Cépadues edn
38. Kannan VR, Tan KC (2002) The Journal of Supply Chain Management, "Supplier Selection and Assessment: their impact on business performance?" Fall 2002
39. Angerhofer BJ, Angelides MC (2006) A model and a performance measurement system for collaborative Supply Chains, Decision Support Systems 42: 283–301

40. Bharadway S (2006) Supplier Visibility Translates to Higher Performance, Supply Chain Management Review, October 06 : 57–58
41. Gunasekaran A, Patel C, McGaughey RE (2004) A framework for supply chain performance measurement, International Journal of Production Economics 87: 333–347
42. Humphreys PK, Wong YK, Chan FTS (2003) Integrating environmental criteria into the supplier selection process, Journal of Materials Processing Technology
43. Katsikeas CS, Pararoidamis NG, Katsikea E (2004) Supply source selection criteria: The impact of supplier performance on distributor performance, Industrial Marketing Management
44. Nakashima K, Nose T, Kuriyama S (2006) A new approach to environmental-performance evaluation.
45. North Rizza M (2006) Supplier Performance Management Can Make a Difference, Supply Chain Management Review, October 06, pp 13–15
46. Lash J, Wellington F (2007) Competitive advantage on a warming planet, Harvard Business Review, 85 : 95–102
47. Womack J, Jones D (2005) Lean Consumption, Harvard Business Review, 83 : 58–68
48. Hammant J (2000) From vertical integration to virtual integration: Barriers to information flows and how to overcome them, Enterprise Integration, March 2000
49. Association Française de Gestion Industrielle (1992) Evaluer pour évoluer, les indicateurs de performance au service du pilotage, ouvrage collectif AFGI
50. Kaplan RS, Norton P (1998) Le tableau de bord prospectif, pilotage stratégique : les 4 axes du succès, Les Editions d'Organisation
51. Neely A, Bourne M, Kennerley M (2003) Dysfunctional Performance through Dysfunctional Measures, Cost Management, Sept/Oct.2003 : 41–45
52. Iribarne P (2006) Les tableaux de bord de la performance,2nde edn, Dunod
53. Chiapello E, Lebas M (2001) The Tableau de Bord, a French Approach to Management Information
54. Kaplan RS, Norton DP (1996) The Balanced Scorecard. Harvard Business School Press
55. Zhou S, Ling W, Peng Z (2007) An RFID-based remote monitoring system for enterprise internal production management, The International Journal of Advanced Manufacturing Technology 33 : 837–844
56. Chuang ML, Shaw WH (2005) How RFID will impact Supply Chain Networks, Engineering Management Conference
57. Chopra S, Sodhi MMS (2007) Looking for the Bang from the RFID Buck, Supply Chain Management Review: May/June 2007 : 34–41
58. Fish LA, Forrest WC (2007) A Worldwide Look at RFID, Supply Chain Management Review, April 2007 : 48–55
59. Womack J, Jones D (1996) Lean Thinking: Banish Waste and Create Wealth in your Cooperation, Simon and Schuster
60. Jones DT, Hines P, Rich N (1997) Lean Logistics, International Journal of Physical Distribution & Logistics Management 27 :153–173
61. Buxton M, Jutras C (2006) The Lean Supply Chain Report: Lean Concepts Transcend Manufacturing through the Supply Chain, The Lean Supply Chain Benchmark Report, Aberdeen-Group
62. Simpson DF, Power JF (2005) Use the supply relationship to develop lean and green suppliers, Supply Chain Management: an International Journal 10 : 60–68
63. Lamming R (2003) Beyond Partnership: strategies for innovation and lean supply, The manufacturing practitioner series
64. Agarwal A, Shankar R, Tiwari MK (2007) Modeling agility of supply chain, Industrial Marketing Management 36 : 443–457
65. Christopher M (2000) The Agile Supply Chain Competing in Volatile Markets, Industrial Marketing Management 29 : 37–44
66. Smith JB, Colgate M (2007) Customer value creation : a practical framework, Journal of Marketing Theory and Practice 15 : 7–23

67. Swafford PM, Ghosh S, Murphy NN (2006) A framework for assessing value chain agility, International Journal of Operations and production management 26 : 118–140
68. Arteta BM, Giachetti RE (2004) A measure of agility as the complexity of the enterprise system, Robotics and Computer Integrated Manufacturing 20 : 495–503
69. Towill D, Christopher M (2002) The Supply Chain Strategy Conundrum : to be Lean Or Agile or to be Lean And Agile?, International Journal of Logistics:Research and Application 5 : 299–309
70. Narasimhan R, Swink M, Kim SW (2006) Disentangling leanness and agility: an empirical investigation, Journal of Operations Management 24: 440–457
71. Agarwal A, Shankar R, Tiwari MK (2006) Modeling the metrics of lean, agile and leagile supply chain : an ANP-based approach, European Journal of Operational Research 173 : 211–225

Dynamic Data Mining for Improved Forecasting in Logistics and Supply Chain Management

Richard Weber, Jose Guajardo

Department of Industrial Engineering, University of Chile, Chile

Abstract Supply Chain Management relies heavily on forecasts, e.g. of future demand or future prices. Most applications, however, use static forecasting models in the sense that past data is used for model construction and evaluation without being updated adequately when new data becomes available. We propose a dynamic forecasting methodology and show its effectiveness in a real-world application.

1 Introduction

The efficient management of supply chains is based – among other elements – on reliable forecasts of e.g. future demand for certain products, e.g. (Aburto, Weber 2007). The respective forecasting models are usually constructed and evaluated based on past data which is split in training, validation, and test data. Once such a model has been built, it is applied to forecast future demand usually without major adaptation. We present a dynamic forecasting methodology based on regression models where we propose Support Vector Regression for model construction. Our methodology, however, is independent of the particular regression method, i.e. other regression methods can be used within the proposed framework.

Section 2 provides the basics of Support Vector Regression. In Sect. 3 we develop the proposed forecasting methodology. Section 4 presents the results derived by our methodology as well as using alternative approaches. Section 5 concludes this work and points at future developments.

2 Support Vector Regression

Here we describe the standard Support Vector Regression (SVR) algorithm, which uses the ε-insensitive loss function, proposed in (Vapnik 1995). This function allows a tolerance degree to errors not greater than ε. The description is based on the terminology used in (Smola, Schölkopf 1998).

H.-D. Haasis et al., *Dynamics in Logistics*
DOI: 10.1007/978-3-540-76862-3,

Let $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)$, where $\mathbf{x}i \in \Re^n$ and $yi \in \Re\,\forall i$, be the training data points available to build a regression model. The SVR algorithm applies a transformation Φ to the original data[1] from the initial *Input Space*, to a generally higher-dimensional *Feature Space* (F). In this new space, we construct a linear model, which corresponds to a non-linear model in the original space:

$$\Phi : R^n \to F, \quad w \in F \qquad f(x) = \langle w;\, \Phi(x) \rangle + b$$

The goal when using the ε-insensitive loss function is to find a function that fits current training data with a deviation less or equal to ε, and at the same time is as flat as possible. This means that one seeks to minimize the norm $\|\mathbf{w}\|^2$ (Smola, Schölkopf 1998). Introducing slack variables ξ_i, ξ_i^* which allow error levels greater than ε, we arrive at the formulation proposed in (Vapnik 1995), which is known as the primal problem of the SVR algorithm (see (1)). The objective function takes into account generalization ability and accuracy in the training set, and embodies the structural risk minimization principle (Vapnik 1998). Parameter C determines the trade-off between generalization ability and accuracy in the training data, and parameter ε defines the degree of tolerance to errors.

$$\begin{aligned} &\text{Min}\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}(\xi_i + \xi_i^*) \\ &y_i - \big\langle w, \Phi(x_i)\big\rangle - b \le \varepsilon + \xi_i \\ &\big\langle w, \Phi(x_i)\big\rangle + b - y_i \le \varepsilon + \xi_i^* \\ &\xi_i, \xi_i^* \ge 0, \quad i = 1, 2, \ldots, l \end{aligned} \tag{1}$$

3 The Proposed Forecasting Methodology

To build forecasting models, one has to deal with different issues such as feature selection, model construction, and model updating. In this section we present a framework to build SVR-based forecasting models, which takes into account all these tasks. First, a general description of this framework is provided; then, we present details about SVR model updating. The issues feature selection and model construction have been introduced e.g. in (Guajardo et al. 2006).

3.1 General Framework of the Proposed Methodology

The first step consists of dividing the data into training, validation and test subsets. Training data will be used to construct the model, validation data is used for model and feature selection, and test data is a completely independent subset, which is

[1] When the identity function is used, i.e. $\Phi(x) \to X$, no transformation is carried out and linear SVR models are obtained.

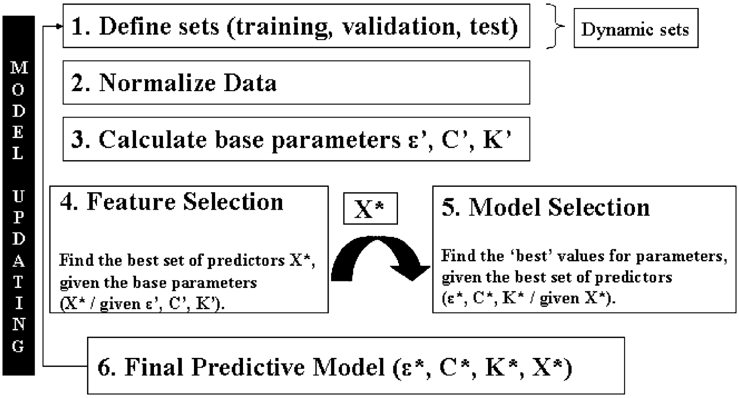


Fig. 1 SVR-UP Forecasting Methodology

used to provide an estimation of the error level that the model would have in reality. As we perform model updating, these subsets will be dynamically redefined over time.

The proposed methodology can be summarized as follows: first, calculate base parameters (ε', $C'$and kernel parameter K') that 'work well' under some general conditions. Next, and using these base parameters, perform feature selection using a wrapper method with forward selection strategy to obtain the best set of predictor variables X^*. Finally, using predictors X^*, return to the problem of model construction performing grid search (Chang, Lin 2001) around base parameters to get the optimal parameters ε^*, C^*, K^*. Thus, at the end, the predictive model is determined by parameters ε^*, C^*, kernel function K^* and predictors X^*. The final element in this framework is model updating. Figure 1 provides a general view of our framework for SVR-based forecasting.

3.2 Model Updating within the Proposed Methodology

Let $\{(x_1, y_1), \ldots, (x_l, y_l)\}$, where $x_i \in \Re^n$ and $y_i \in \Re$, be the training data points available to build a regression model for predicting the values of the time series defined by y_i. We suppose the series to be characterized by seasonal patterns (cycles) of length c. In logistics this would be relevant if we want to predict seasonal sales pattern, e.g. higher sales volumes at the end of each month.

The Static Case

The static approach sequentially splits the data into training, validation and test sets. We represent the static configuration of these 3 subsets by using a matrix as shown in Fig. 2 where the index tri represents the position of the initial training data point

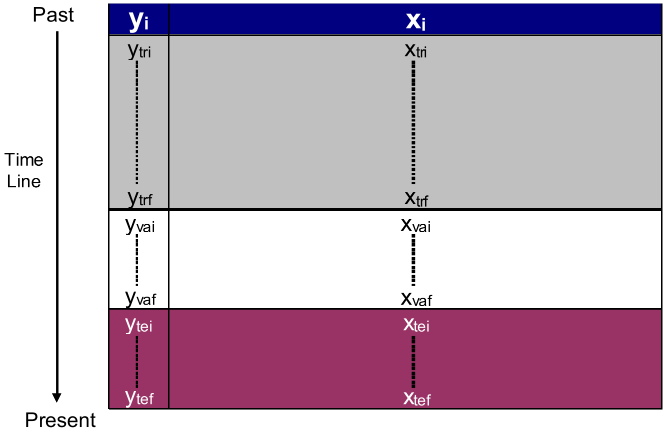


Fig. 2 Static Configuration

and the index trf the position of the final training data point. The same definition is considered for validation (vai, vaf) and test indexes (tei, tef).

Since in the static case data is sequentially divided into the 3 subsets, tri will be set equal to 1 and tef will be set equal to m. The values of trf and vaf are defined to obtain the desired proportion of data belonging to each subset.

The Dynamic Case

We designed our model updating strategy to deal with time series with seasonal patterns. We will refer to a complete seasonal pattern as a cycle; examples are e.g. monthly data with yearly seasonality, where a cycle is defined by a year.

First, we have to identify the length of the cycles of the series. For this purpose, graphical inspection, autocorrelation analysis or spectral analysis can be carried out. Then we define a test set containing at least two cycles. This test set is divided into subsets, each one containing a single cycle. Let there be k cycles of length c belonging to the test set in the static case, i.e., $\text{tef} - \text{tei} + 1 = k \cdot c$ number of observations contained in the test set. In the static case, we construct just one model to predict all the observations contained in the test set, as well as for future observations, maintaining this model throughout the entire procedure unchanged. The main idea of our updating approach consists of developing different predictive models for each cycle of the test set, as well as for any future cycle by incorporating the most recent information for model construction.

Next, we have to define the configuration of the training and validation sets for predicting the first cycle of the test set. Figure 3 displays such configuration. As can be seen, training data contains two parts where the first part is a set of past (or historical) data and the second part contains the most recent information. The idea of integrating the most recent information into the training set is to obtain

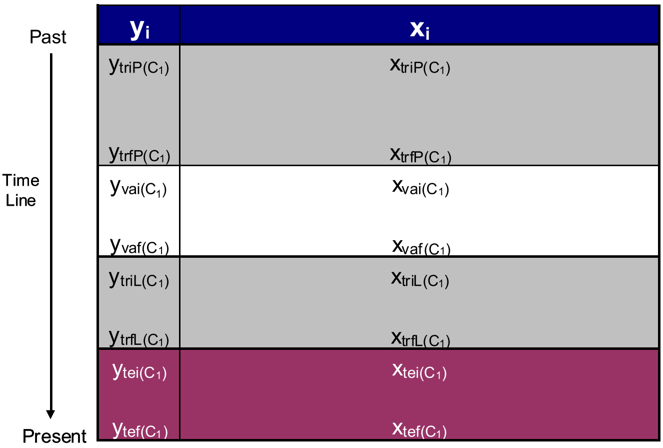


Fig. 3 Predicting the first cycle (C1)

more accurate estimations by taking into account the most recent patterns in model construction, while the proportion of data belonging to the training and validation is kept stable over time.

To predict the second cycle of the test set, we add the first cycle of the test set (this data is now part of the past) to the training set, and build a different predictive model having the configuration shown in Fig. 4. By doing so, we ensure again that most recent information has influence in model construction.

As we want to keep the proportion of data belonging to the training and validation sets stable over time, we move data from the historical part of the training set to the validation set. The amount of data to be moved is determined according to the

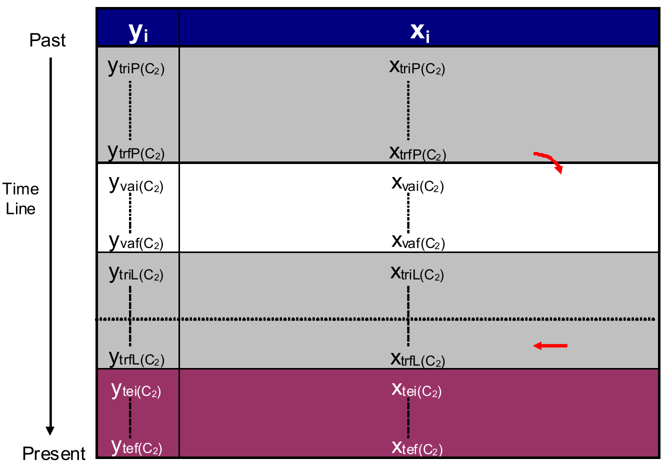


Fig. 4 Predicting the second cycle (C2)

original proportion of data in each subset. This same strategy will be applied to build predictive models for the rest of the cycles of the test set, as well as for any future observations.

4 Experiments and Results

We applied the proposed methodology to a real-world sales prediction problem, where a company wants to predict the number of units that will be sold the following week, for 5 different products. Consequently, we analyzed 5 time series representing weekly sales data, during the period from January 2001 to September 2004 each.

Besides the application of our methodology using support vector machines (SVR-UP), we utilized the same framework but now with neural networks (NN-UP). Also, we developed ARMAX models to have a standard forecasting method as benchmark.

Table 1 shows the mean absolute error (MAE), by applying the 3 forecasting methods for predicting one period ahead sales for the five products. Similar results have been obtained using mean absolute percentage error (MAPE) and root mean squared error.

Table 1 Mean absolute error in test set (underlined: best result for each row)

Product	ARMAX	NN-UP	SVR-UP
P1	<u>292</u>	350	342
P2	347	368	<u>283</u>
P3	103	<u>89</u>	96
P4	<u>268</u>	288	284
P5	328	280	<u>264</u>
Average	275	275	<u>254</u>

5 Conclusions and Future Works

We presented the methodology SVR-UP for time series forecasting, which combines feature selection, regression model construction using Support Vector Machines, and model updating.

We have applied the proposed methodology using SVR as well as neural networks as regression models to a sales forecasting problem and showed its superior performance compared to a standard ARMAX approach. Major advantages of the proposed methodology are expected when dealing with dynamic phenomena, where the performance of a forecasting model could be significantly improved by model updating.

Especially in complex supply chains which face a dynamic environment such an adaptive forecasting procedure could be extremely important.

Acknowledgements Support from the Millennium Institute "Complex Engineering Systems" and the Fondecyt project 1040926 is greatly acknowledged.

References

Aburto, L., Weber, R. (2007): *Improved Supply Chain Management based on Hybrid Demand Forecasts*. Applied Soft Computing 7, 136–144

Chang, C., Lin, C. (2001): *LIBSVM: a library for support vector machines*. Software available at http://www.csie.ntu.edu.tw/cjlin/libsvm

Guajardo, J., Weber, R., Miranda, J. (2006): A Forecasting Methodology Using Support Vector Regression and Dynamic Feature Selection. Journal of Information & Knowledge Management 5, No. 4, 329–335

Smola, A., Schölkopf, B. (1998): *A Tutorial on Support Vector Regression*. NeuroCOLT Technical Report NC-TR-98-030, Royal Holloway College, U of London, UK

Vapnik, V. (1995): *The nature of statistical learning theory*. Springer, New York

Vapnik, V. (1998): *Statistical Learning Theory*. John Wiley and Sons, New York

Introducing Bounded Rationality into Self-Organization-Based Semiconductor Manufacturing

Tomomi Kito, Kanji Ueda

Research into Artifacts, Center for Engineering (RACE), the University of Tokyo
Kashiwanoha 5-1-5, Kashiwa, Chiba 277-8568, Japan
e-mail: {kito,ueda}@race.u-tokyo.ac.jp

Abstract This study proposes a design approach for self-organization-based systems by introducing bounded-rational agents. The validity of the proposed model of bounded rationality is validated in comparison to the proposal of information localization, using a simulation of self-organization-based semiconductor manufacturing, which is a typical complex system. The effectiveness of the proposal is demonstrated also under changing situations.

1 Introduction

Today's manufacturing industries face strong trends of diversification of cultures, individualization of lifestyles, globalization of activities, and growing consideration of the natural environment. These trends can be summarized as growing complexity and dynamics in manufacturing environments (Wiendahl et al. 1994). It is difficult to accommodate such complexity and dynamics using existing ideas with a top-down type of control provided by centralized systems. For that reason, many new ideas based on bottom-up approaches have been proposed for next-generation manufacturing. The self-organization model is one such approach: it adopts emergent synthesis methodologies (Ueda et al. 2001) including self-organization, a biologically inspired idea; moreover, its effectiveness under complex and uncertain situations has already been demonstrated (Ueda et al. 1997, Ueda et al. 2002).

Instead of flexibility and adaptivity resulting from the agents' autonomy, a self-organization-based system has a substantive aspect: local competition might arise among agents because of their respective decision-making processes, which are based on local information. Such local competition might degrade the system performance. The existing effective ways to solve this problem are those which pro-

H.-D. Haasis et al., *Dynamics in Logistics*
DOI: 10.1007/978-3-540-76862-3,

vide divisions of roles or rules to avoid competition, or which give agents information about the whole system to generate cooperation among agents. However, each of them requires top-down control in the form of providing rules or roles, or information related to the whole system. The authors urge that introducing bounded rationality as a characteristic of agents constitutes a new approach to solving this problem without top-down control and using only local information.

The authors have proposed a model of bounded-rational decision-making (Kito et al. 2006). The present study verifies the validity of the proposed model of bounded rationality using a simulation of self-organization-based semiconductor manufacturing, which is a typical complex system. Especially in this paper, the effectiveness of introducing bounded rationality is demonstrated through comparison to other proposals of information localization (Kuraoka et al. 2006), the details of which will be described later.

The next section describes the concept of introducing bounded rationality into multi-agent systems and the proposed model of bounded-rational decision-making. Section 3 describes the model of self-organization-based semiconductor manufacturing, with its complexity and ways to solve problems that occur in the system. Section 4 shows simulation results and discusses the effectiveness of introducing bounded rationality.

2 Introducing Bounded-Rational Agents

The authors propose that introducing bounded rationality into constituent agents of multi-agent systems can solve local competition among agents and improve the system performance.

Bounded rationality was originally coined as a concept by Simon (Simon 1956) to indicate that the decision-making behavior of human beings cannot conform to the ideal of *perfect rationality* because of limitations of cognitive capabilities. In the engineering field, the use of the term *bounded rationality* follows the tradition of Simon, i.e., it has the same meaning as *resource-bounded*: limited calculation time and computational capacity. Artificial agents are designed based on the pursuit of optimization under time and capacity constraints.

The authors emphasize that bounded rationality has another aspect: human beings originally have no firm incentives for being rational. We specifically examine this "incompleteness of incentives" as a key to resolving local competition.

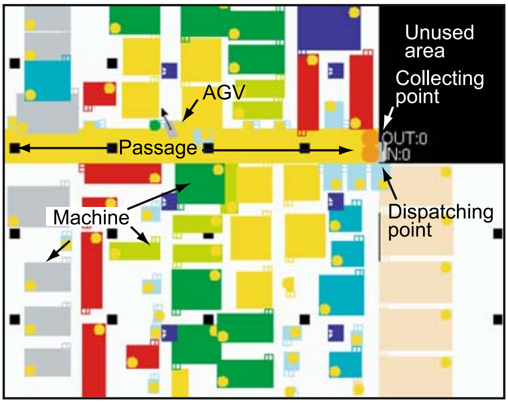
In this paper, the bounded-rational decision-making is modeled as follows: "When the input of the decision-making function matches the criterion, choose the alternative that is derived as rational behavior using not all the input information but a fraction of it." The input of the decision-making function is perceptual information of the environment and the agent's internal state.

3 Self-Organization-Based Semiconductor Manufacturing Model

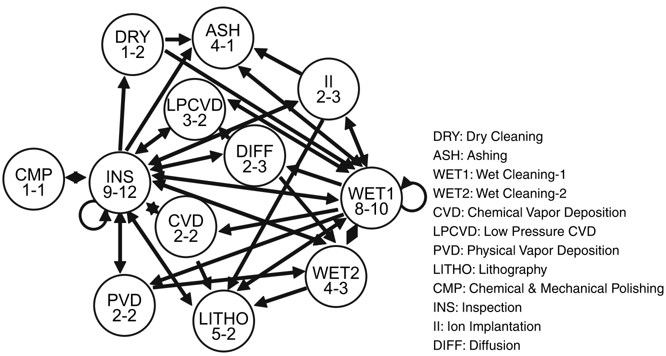
3.1 Complexity of Semiconductor Manufacturing

Computer simulations of self-organization-based semiconductor manufacturing system have been carried out. Semiconductor manufacturing systems are representative examples of complex systems: they include hundreds of machines; even if only a single kind of product is to be produced, production of semiconductor elements necessitates a complicated process flow in which the same kinds of machines are used repeatedly in different production steps.

Figure 1a) shows the floor layout of the semiconductor manufacturing system used for the simulations described in this study. Products supplied from the dis-



a) Floor layout



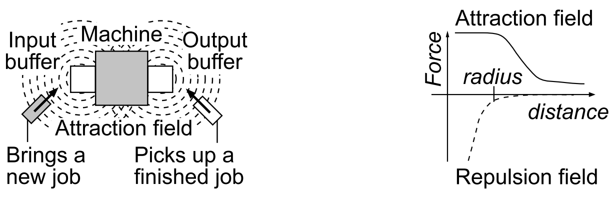
b) Process flow network

Fig. 1 Floor layout and process flow network in a semiconductor manufacturing system patching point are carried by AGVs toward appropriate machines and are processed. After processing, a product is carried toward a next machine for next process. Consequently, the finished products are carried by AGVs toward the collection point. Figure 1b) shows the process flow network of this system, in which nodes mean process types (i.e. machine types) and the arcs (arrows) mean input-output relations. Numbers below process types indicate the input-output number between processes. The complexity of the process flow is apparent in this figure.

3.2 Self-Organization-Based Model

A self-organization process is defined as a process that reflects the global behavior that emerges from local interactions among the system elements (Vaario et al. 1998). Here, the elements of self-organization-based semiconductor manufacturing are the processing machines and AGVs as transporters of products. In manufacturing, self-organization means that production is carried out through local matching between the capabilities of the machines and the product requirements. No global control is exercised.

Potential fields are employed (Vaario et al. 1998) to realize local interaction as local matching between capabilities and requirements. This field is the superposition of attraction and repulsion fields: each machine generates an attraction force according to its capabilities to obtain products. An AGV senses the accumulated attraction field of machines' input buffers with capabilities that match the product that it is transporting (Fig. 2(a)). Finally, the AGV gives the product to an input buffer of a machine. If an AGV is not transporting a product, it senses the attraction fields of machines' output buffers or the dispatching point to pick up a product. Each AGV also generates a repulsion force and senses those forces generated from other AGVs, thereby avoiding mutual collision. Typical shapes of attraction and repulsion fields are shown in Fig 2(b).



(a) Local interactions **(b)** Field shape

Fig. 2 Potential-field based local interaction among machines and AGVs

Using this model, semiconductor manufacturing processes are self-organized. The simulation setting, which was determined based on real manufacturing data, is as follows:

- Machines: There are 12 types of machines. Their total number is about 80.
- AGVs: The number of AGVs is 4.
- Products: A product consists of one lot of wafers, which require about 300 process steps. It successively activates the production requirement with the process. Some products arrive at the floor on a fixed schedule.

3.3 Local Competitions in Self-Organization-Based System

Local competition that occurs among agents in the self-organization-based semiconductor manufacturing system is, for instance, a situation in which multiple AGVs gather at one machine to receive a product simultaneously, or a situation in which an AGV machine generating an attraction field has its movement disturbed by another machine generating the same kind of attraction field.

Such situations occur because potential fields pervade the floor; spatial restriction is not considered. Local competition might be avoided and the system performance might therefore improve if movements by which each AGV moves to a neighbor machine emerge through self-organization processes.

There are mainly two ways to consider and introduce spatial restriction in this self-organization-based semiconductor manufacturing system:

1. Limiting information generation: machines limit the area in which their generated potential fields spread.
2. Limiting information usage: AGVs limit the area in which they sense potential fields.

In fact, (1) has been proposed as "information localization" (Kuraoka et al. 2006). This study proposes the idea of introducing bounded rationality, which is a way of realizing (2). In the next subsection, the models of these two ways for introducing self-organization-based semiconductor manufacturing are described.

3.4 Introducing Spatial Restriction

Limiting information generation: Information Localization

Each machine i generates an attraction field $Fi(l)$; l is the distance from the machine i. Actually, $F_i(l)$ is defined as $F_i(l = \exp - l^2/\lambda_i^2$, where λ is the influence parameter. In the simulation, the following four variables are used as λ: 10, 17, 28,

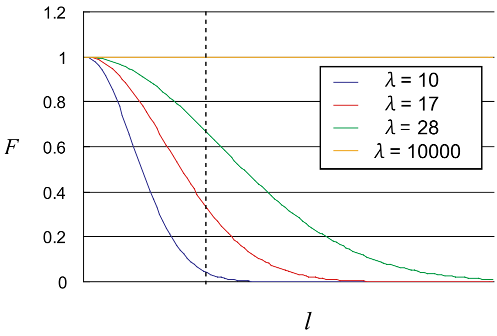


Fig. 3 Information Localization

and 10000 (Fig. 3). The situation in which $\lambda = 10000$ is equivalent to that in which spatial restriction is not considered.

Limiting information usage: Model of Bounded Rationality

AGVs are modeled as decision-making agents. A rational AGV moves in the direction of the attraction force if it senses an attraction field. It performs no action if it senses no attraction field. Three types of bounded rationality are implemented:

(Type-0) the AGV consistently limits the distance range of the sensing area of attraction fields.

(Type-1) It limits the sensing area of attraction fields if the number of times the AGV carries products is greater than the average number of times the other AGVs carry them.

(Type-2) It limits the sensing area of attraction fields if the distance the AGV moved is greater than the average distance the other AGV moved.

4 Simulation Results and Discussion

4.1 Comparison Between Information Localization and Information-Use Limitation

The results of introducing information localization and introducing (Type-0) bounded rationality were compared, as shown respectively in Table 1 and Table 2.

Table 1 System performance with information localization

	$\lambda = 10$	$\lambda = 17$	$\lambda = 28$	$\lambda = 10000$
WIP (Lot)	71.46	68.84	70.96	73.77
TAT (day)	6.44	6.19	6.40	6.65
Mileage (m)	4786.9	4696.8	4730.1	4802.4

Table 2 System performance with (Type-0) AGVs

	$d = 30$	$d = 40$	$d = 50$	$d = 60$
WIP (Lot)	71.11	68.10	72.00	72.68
TAT (day)	6.50	6.17	6.53	6.60
Mileage (m)	4724.8	4684.3	4698.0	4784.3

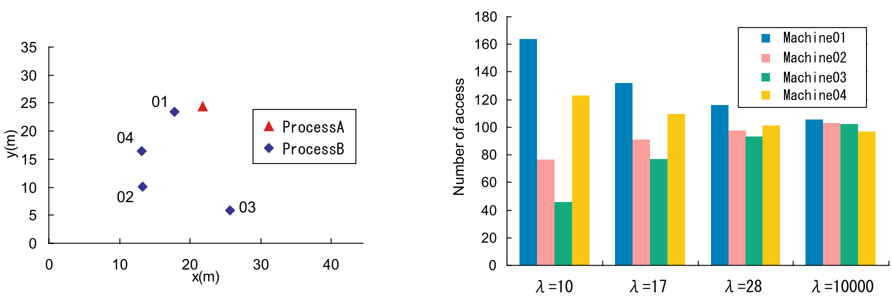
Here, WIP means Work In Process. Also, TAT is the short turn-around-time of the products; Mileage is the carried distance of products.

As Table 1 shows, the system performance of the case in which machines generate attraction fields to be spread throughout the floor ($\lambda = 10000$) is worse than the case in which machines generate fields locally ($\lambda = 17, 28$). In addition, $\lambda = 10$ is too small to realize production processes effectively.

The results presented in Table 2 show a similar tendency: the system performance of the case in which the distance range of the AGVs' sensing areas are limited as $d = 40$ or 50 is higher than in the case in which that range is not so limited or too limited ($d = 30, 60$).

These results indicate that both ways of considering spatial restriction in the self-organization-based system are effective for improving the system performance. Figure 4 shows the result, which represents one reason for improvement of the system performance. Figure 4(a) displays the positions of the four machines, which are dealing with the same process (Process B), and the position of the machine for Process A, which is the previous process of Process B. Figure 4(b) shows the differences of the total number of accesses from the machine for process A toward each of the four machines for Process B, according to λ. As might be readily apparent, the greater the degree of information localization, the greater the difference of access numbers among machines. By considering spatial restriction, the movement by which products are carried toward nearer machines would emerge; consequently, the system performance would be improved.

The two methods of (1) information localization and (2) introduction of (Type-0) bounded rationality are similar in terms of limitation of information for AGVs. On



(**a**) Positions of machines (**b**) Numbers of accesses

Fig. 4 (a)–(b) Positions of machines for processes A and B, and the number of accesses from a machine for process A to four machines for process B the other hand, their difference is: (1) information is incomplete and agents are so rational that they use all information; and (2) information is complete but agents are bounded rational, such that they use some information. In the real world, although information providers try to provide complete information, it will not always be complete for agents because of uncertainty or dynamic changing of the environment. Introducing bounded rationality is a more realistic way than information localization, and has more flexibility and feasibility.

4.2 Introduction of Bounded Rationality

Table 3 shows results of the simulation by introducing (Type-1) or (Type-2) AGVs. In those simulations, each of the four AGVs senses only those fields within a 40 (m) radius if the input of the decision-making function matches the criterion. As the table shows, the system performance improved according to evaluation indexes compared to the results presented in Table 1 and Table 2. Although all the AGVs have identical simple decision-making functions, they sometimes sense a distant field and sometimes not. Such various behaviors offer the potentiality to accelerate a kind of role-sharing, and thereby improve the system performance.

The effectiveness of bounded rationality under changing environments was also verified. Simulations were executed in which machines break down with a given probability. Table 4 shows the results of the case of $\lambda = 10000$, $\lambda = 17$, and the case with Type-1 or Type-2 agents. As Table 4 shows, the overall system performance worsened. Compared to the case of $\lambda = 10000$, the case of $\lambda = 17$ maintains the performance. This result indicates that information limitation would be effective under changing situations. Moreover, in the case of introducing (Type-1) or (Type-2) bounded-rational agents, the system performance can be retained at a higher level than in the case of $\lambda = 17$. These results imply that, because bounded-rational agents choose their actions according to situations of the environment and their internal

Table 3 System performance with (Type-1) or (Type-2) AGVs

	Type-1	Type-2
WIP (Lot)	67.66	67.81
TAT (day)	6.00	6.11
Mileage (m)	4498.5	4492.3

Table 4 System performance under changing environments

	$\lambda = 10000$	$\lambda = 17$	Type-1	Type-2
WIP (Lot)	98.41	78.75	77.53	77.44
TAT (day)	8.87	7.11	7.00	6.96
Mileage (m)	4927.2	4789.2	4600.1	4587.4

states, they can be more flexible or adaptive to changing environments than when considering information localization.

5 Conclusion

This study proposes a design approach for self-organization-based systems by introducing bounded-rational agents. This proposal is effective to solve the problem of local competition that occurs among agents. Bounded rationality was modeled from the standpoint of not using all information that the agent can perceive when the input of the decision-making function matches the criterion.

In self-organization-based semiconductor manufacturing simulations, the proposed model of bounded rationality and another proposed model of information localization were implemented. The simulation results show that both modes of information localization and introduction of bounded rationality can be effective to improve the system performance. Moreover, these results imply that bounded rationality might be the key to endowing the system with added flexibility and adaptivity to solve local competition under changing situations.

To deal with multi-agent systems in the real world, such as complex manufacturing systems, with environments and structures which are complex and uncertain, the proposed idea is that introducing bounded-rational agents can be an effective approach to derive adaptive solutions under incompletely understood conditions while simultaneously obviating the need for top-down control or provision of global information and rules.

References

Kito, T., Fujii, N., Ueda, K., 2006, "Design of Manufacturing Systems by Introducing Bounded-Rational Agents", CIRP Journal of Manufacturing Systems, Vol.35, No.2, pp.151–160.

Kuraoka, H., Fujii, N., Ueda, K., 2006, "Emergence of Cooperation in a Network Environment Introducing Social Distance," Proceedings of European Conference on Complex Systems (ECCS'06).

Simon, H.A., 1956, "Rational Choice and the Structure of Environments," Psychological Review, Vol.63, pp.129–138.

Ueda K., Vaario, J., Ohkura, K., 1997, "Modeling of Biological Manufacturing Systems for Dynamic Reconfiguration," Annals of the CIRP, Vol.46, No.1, pp.343–346.

Ueda, K., Markus, A., Monostori, L., Kals, H.J.J., Arai, T., 2001, "Emergent Synthesis Methodologies for Manufacturing," Annals of the CIRP, Vol.50, No.2, pp.535–551.

Ueda, K., Fujii, N., Hatono, I., Kobayashi, M., 2002, "Facility Layout Planning Using Self-Organization Method," Annals of the CIRP, Vol.51, No.1, pp.399–402.

Vaario, J., Ueda, K., 1998, "An Emergent Modeling Method for Dynamic Scheduling," Journal of Intelligent Manufacturing, No.9, pp.129–140.

Wiendahl, H.-P., Scholtissek, P., 1994, "Management and Control of Complexity in Manufacturing," Annals of the CIRP, Vol.43, No.2, pp.533–540.

Routing in Dynamic Logistics Networks

Travel Time Estimation and Deadlock-free Routing of an AGV System

Hyo Young Bae, Ri Choe, Taejin Park, Kwang Ryel Ryu

Department of Computer Engineering, Pusan National University,
Busan, 609-735, Korea
e-mail: {nextpage,choilee,parktj,krryu}@pusan.ac.kr

Abstract Concurrent operation of multiple automated guided vehicles (AGVs) in a limited space such as a container terminal results in a high probability of occurrences of collisions, deadlocks, and livelocks; all these occurrences decrease the productivity of the AGVs. Therefore, it is important to assign a route to each AGV in a manner that prevents these occurrences. Moreover, the accelerated motion of an AGV makes routing more difficult. In this paper, we propose a method of avoiding collisions, deadlocks, and livelocks using an occupancy area reservation (OAR) table. We also propose a method to estimate the travel time of an AGV with accelerated motion in order to select the most efficient AGV route. The experimental results indicate that our proposed method is effective in increasing the productivity of AGVs when they are operated in a limited space such as a container terminal.

1 Introduction

In recent years, an increasing number of container terminals have adopted automated systems, and the interest in automated guided vehicle (AGV) systems is also growing. Therefore, several researchers are investigating efficient routing and dispatching for AGVs.

AGVs, being unmanned vehicles, require a routing process that is capable of generating travel routes from their current position to their destination. In particular, the routing process should be capable of determining the travel route with the lowest cost in order to improve efficiency. It is also important to control the motion of multiple AGVs in order to avoid collisions, deadlocks, and livelocks, which paralyze traffic flow and cause damage to vehicles and containers. Thus, traffic control and lowest-cost routing have been the major concerns in AGV systems.

Many previous researches are based on the zone control scheme (Evers et al., 1996; Kim et al., 1991; Rajeeva et al., 2003; Reveliotis, 2000; Yeh et al., 1998). The zone control scheme generates a graph by dividing all possible guide paths into

H.-D. Haasis et al., *Dynamics in Logistics*

DOI: 10.1007/978-3-540-76862-3,

zones and linking them to each other. In this scheme, it is easy to manage traffic control and to apply the shortest path algorithm of graph theory to routing process. However, it has a limitation with regard to the degrees of freedom to set a travel path because it is based on a fixed layout. Furthermore, it cannot take into account the accelerated motion of an AGV because it assumes path costs as constants.

Another research is based on the grid control scheme that generates flexible paths and increases space utilization to enhance the efficiency of AGVs (Kim et al., 2004). The grid method divides space into rectangular grids smaller than an AGV and represents the travel path as a sequence of occupied grids.

The grid method can set up flexible routes; thus, it is compatible with free-roaming AGVs. When an AGV turns, unlike the zone control scheme, the sequence of occupied grids cannot follow the direction of the motion of an AGV. So more sophisticated control algorithm is required for collisions and conflicts.

We solved the problem of discordance between the sequence of areas and the motion of an AGV when it turns by using more flexible grids divided by the line of flow of an AGV. The proposed method avoids collisions, deadlocks, and livelocks by using an occupancy area reservation (OAR) table to control the travel time of an AGV. Further, we proposed a method to estimate the travel time of AGVs by taking into consideration accelerated motions and the interference due to other AGVs for generating low-cost paths to be used in real-time applications easily.

The rest of the paper is organized as follows. Section 2 explains our traffic control method for preventing collisions, deadlocks and livelocks. Section 3 describes our method for estimating the travel time of an AGV considering interferences. Section 4 provides our experimental setting and results. Finally, Sect. 5 presents our concluding remarks and discusses future direction of our work.

2 AGV Traffic Control

AGV traffic control is broadly divided into two major parts—creating a route and generating a travel schedule in order to prevent problems such as collisions, deadlocks, and livelocks.

2.1 Route Creation

An AGV route can be a curved line from the starting point to the end point. It could also be defined as a sequence of areas divided by the route at regular intervals. In this study, we refer to these areas as occupation areas (OAs). In order to prevent collisions, an AGV has to reserve OAs in advance to move forward. If it is unable to reserve the required OA, it waits at its current position until the reservation is allowed. Realizing AGV traffic control by using larger OAs is simpler, however, it

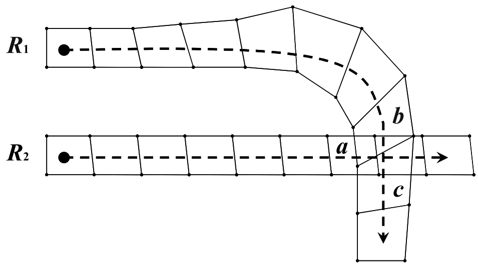


Fig. 1 Overlapping OAs belonging to different routes

reduce space utilization by occupying area unnecessarily. To improve the efficiency of space utilization, the OAs are divided by the line of flow of an AGV. As shown in Fig. 1, the OAs are assigned independently to each AGV route. Hence, to avoid collisions, it is necessary for an AGV to verify the reservations of all OAs on other routes that overlap with the OA which the AGV is to reserve. For example, if we assume that an AGV traveling on route R_2, as shown in Fig. 1, is going to reserve the OA labeled a, then it is required to check if b and c of R_1 are reserved by any other AGV before reserving a.

2.2 AGV Travel Scheduling

This study introduces an OAR-table-based method for preventing deadlocks and livelocks among AGVs. The OAR table records when an AGV reserves and passes through each OAs, and is created just before the time of departure of an AGV by reference on the OAR tables of other AGVs. A traveling AGV is allowed to reserve OAs for only as long as specified by the OAR table's storing reservation time.

Figure 2 shows a deadlock/livelock-free routing algorithm. For convenience, we assume that an AGV moves forward as far as the unit length of an OA and the length of the AGV is twice that of the OA. $\mathrm{OAR}(i)$ is the duration for which an AGV occupies OA_i; $\mathrm{OAR}_e(i)$, the moment at which a vehicle completes its occupation of OA_i; $\mathrm{UAR}_s(i)$, the moment at which a vehicle is prohibited to occupy OA_i; $\mathrm{UAR}_e(i)$, the moment at which a vehicle is allowed to occupy OA_i again; and $\mathrm{canReserve}_i(s,e)$, the function that verifies whether the duration for which the occupation of OA_i is disallowed overlaps with (s, e).

There are three preconditions for an AGV deadlock to occur. Essentially, multiple AGVs are not allowed to reserve the same OA simultaneously. Occupying several OAs, an AGV tries to reserve another OA. Lastly, an AGV cannot reserve an OA that has been reserved earlier by another AGV.

Our proposed method allocates the reservable time of each OA when it makes an OAR table of an AGV k. In this process, if the reservable time of a certain OA for k_i overlaps with that for other AGVs which depart earlier, k postpones the reservable time of the OA. Thus, we can conclude that no deadlocks occur when the routes are

```
Input: OA_end-last OA on the route

t = 0
for i = OA_first to OA_end
        t = t + 1
        if canReserve_{i+1}(t, t+2)
                i = i + 1
                OAR(i) = (t, t+2)
        else
                OARe(i) = t+2
                while !canReserve_i(OAR(i))
                        i = i-1
                        OARe(i) = t+2
                t = UARe(i)
                OARe(i) = t+2

Output: OAR table
```

Fig. 2 Deadlock-free routing algorithm

generated based on the rules suggested in the above-stated method. Furthermore, in the worst-case scenario, an AGV waits just until another AGV that was assigned routes earlier passes through the intersection of routes. That is, the proposed method prevents livelocks.

3 Travel Time Estimation Algorithm

With regard to generating AGV routes, a route with the shortest travel time improves the efficiency of AGVs significantly. Due to the accelerated motion of an AGV, its travel time may vary even though it travels the same distance. A traveling AGV is always subject to interference from other AGVs, thereby resulting waiting time in order to prevent collisions and deadlocks. To calculate the travel time of an AGV with precision, we need to consider the two closely connected factors carefully.

3.1 Travel Time Estimation in Accelerated Motion

In this study, we assume that an AGV travels with uniform acceleration. We define the initial velocity to be v_0; final velocity, v; acceleration,a; driving distance, d; and the distance required for an AGV to accelerate from v_0 to v, $d_a = (v^2 - v_0^2)/2a$. Then, the travel time can be simply calculated as follows:

if $d \geq d_a$,

$$t = \frac{v - v_0}{a} + \frac{d}{v} - \frac{v}{2a} + \frac{v_0^2}{2av} \tag{1}$$

else,

$$t = -\frac{v_0}{a} + \sqrt{\frac{v_0^2}{a^2} + \frac{2d}{a}}. \tag{2}$$

Thus, the travel time of an AGV can be estimated using Eq. (1) and Eq. (2) when the distance between the starting, end, and turning points and the rotatory distance of an AGV are known.

3.2 Travel Time Estimation Considering Interference

According to Sect. 2.2, an AGV is programmed to wait at the OA before collision and deadlock spots until other AGVs pass through in order to prevent collisions and deadlocks. Due to this, each traveling AGV is interfered by other AGVs, thereby resulting in delays. It is required to consider delays due to interference from other vehicles in order to estimate travel time of an AGV more precisely. This paper suggests the following method**.**

For an AGV k, P_k is an array of intersecting points where the routes of k and those of other AGVs intersect each other. These points are sorted by their distance from the starting point, i.e., the point nearest to the starting point of k is listed first in the array. Further, $T(i,j)$ is the travel time of k between points i and j. $T_s(i)$ is the moment at which an AGV approaches i, and $T_e(i)$ is the moment at which an AGV passes through i. Figure 3 shows the algorithm designed to estimate AGV travel time.

An OAR table records the time that an AGV actually reserves OAs and passes through them. This table makes it easier to find $T_s(i)$ and $T_e(i)$ by searching for an OA that includes the intersection point i marked in the table.

```
Input: P_k
t = 0, i = 0
for each p_i ∈ P_k
        t = t + T(i, p_i)
        i = p_i
        if T_s(p_i) < t < T_e(p_i)
                t = T_e(p_i)
t_out = t
Output: t_out–estimated travel time
```

Fig. 3 AGV travel time estimation algorithm

4 Experimental Results

4.1 Experimental Setting

In order to demonstrate the efficiency of the proposed algorithm, a simulation experiment is set up with the following configuration. The target container terminal consists of one berth, three quay cranes (QCs), and seven blocks with seven automated transfer cranes (ATCs). The QCs and ATCs are assumed to be able to deal with up to fifty containers per hour at maximum. Table 1 shows the specification of AGVs used in this experiment.

Table 1 AGV specification

Forward speed	Cornering speed	Acceleration	Deceleration
4 m/s	2 m/s	0.64 m/s^2	1.55 m/s^2

4.2 Results

This study measures the error between the actual travel time and the estimated one to verify the accuracy of the algorithm stated in Sect. 3. Since the interference among AGVs influences travel time estimation, the error estimation experiment considers two approaches: (1) taking into consideration all the attributes of an AGV model, such as acceleration and deceleration, except interference variables and (2) taking into consideration all variables. This could elucidate how interference affects travel time with greater clarity. Table 2 lists the travel time estimation accuracies for both these approaches. The experiments on the driving routes has been conducted 20,000 times. Considering interference, the travel time estimation error decreased by 1.9%.

Table 2 Travel time estimation accuracy

	Absolute error	Percent error	Mean squared error
All variables considered	84.6 s	11.5 %	126.7
All variables considered, except interference	99.0 s	13.4 %	152.4

Another simulation experiment is performed to determine the reduction in travel time when the travel time estimation algorithm is applied to AGV routing. The number of AGVs varies from 6 to 24 by three. In each experiment, an AGV carries out the discharging process for 24 hours and repeats the same process 5 times. We averaged the results.

We use three routing algorithms—shortest estimated travel time (SETT), shortest travel distance (STD), and random path selection (Random). The SETT method uses the travel path selected by the proposed method. The STD method selects the path with the shortest travel distance. The Random method selects one of available paths at random. We compare the results of each simulation.

Figure 4 shows the travel time and distance of an AGV traveling from a block to a quay. The travel time and distance are the averaged time and distance for an AGV traveling from a block to a QC. In the case of the travel time tests, the SETT method yields the shortest time. The difference between the results obtained from the SETT and STD methods is minimum when the number of AGVs is six. This difference increases as the number of AGVs increases, and becomes maximum with 15 AGVs (6.6%). This difference, however, decreases slightly beyond that point**.** This is because the interference increases as the number of AGVs increases. As the interference increases, the SETT method outperforms the other method. It is no wonder because our method is the only method that considers AGV interference. Meanwhile, when more than 18 AGVs are employed, the number of vehicles traveling in an apron increase; this could decrease the availability of selectable routes thereby decreasing the effectiveness of interference control. On the other hand, In the case of the travel distance tests, the results of the SETT and STD methods show little difference. Comparing these two methods with the Random method the difference in the obtained distances increases significantly. The difference is maximum when six AGVs are used. On increasing the number of AGVs beyond six, this difference decreases gradually.

For AGVs traveling from a quay to a block, the maximum number of AGVs allowed for simultaneous departure is three; this is due to the job cycle time of a QC. That is, the travel from the quay to the block generates less interferences and bottle necks compared with the opposite direction. Therefore, the experimental results show little difference between the SETT and the STD methods.

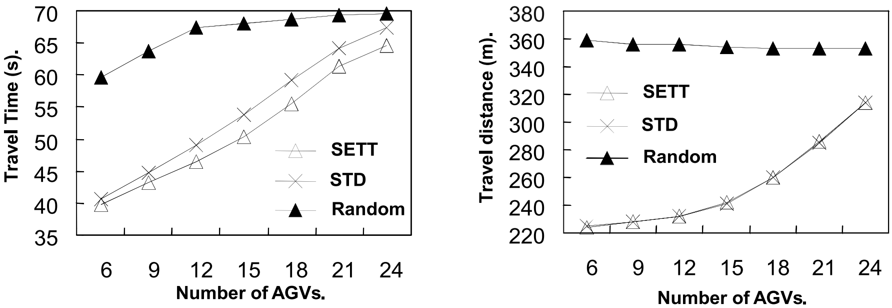


Fig. 4 Average travel distances and times of AGVs traveling from a block to a QC

5 Conclusions

In this study, we suggest approaches to improve the efficiency of AGV systems, an AGV routing process to prevent deadlocks, and methods for travel time estimation considering acceleration, deceleration, and the interferences due to other AGVs. The results from the simulation experiments show that the average travel time decreases gradually when the estimation is performed with considering interference.

As future research, it would be necessary to apply the proposed travel time estimation method to estimate the entire duration of container transportation tasks for the optimization of AGV job scheduling.

Container terminal facilities face certain uncertainties—errors or unexpected malfunctions of sensors and other equipments. Therefore, the suggested approaches also need further research to cater to the need of trends by formulating an improved system to predict unexpected accidents, modify existing travel schedules, or establish new ones in order to adjust to an uncertain environment.

Acknowledgements This work was supported by the Korea Research Foundation Grant funded by the Korean Government (MOEHRD) (The Regional Research Universities Program/Research Center for Logistics Information Technology).

References

Evers, J. J. M. and Koppers, S. A. J. (1996), Automated guided vehicle traffic control at a container terminal, Transportation Research-A 30, pp. 21–34.

Kim, C. W. and Tanchoco, J. M. A. (1991), Conflict-free shortest-time bidirectional AGV routing, International Journal of Production Research 29(12), pp. 2377–2391.

Kim, K. H., Jeon, S. M., and Ryu, K. R. (2004), A deadlock detection method for automated guided vehicles in port container terminals, Proceedings of the 33rd international conference on computers and industrial engineering.

Rajeeva, L. M., Wee, H.-G., Ng, W.-C., and Teo, C.-P. (2003), Cyclic deadlock prediction and avoidance for zone-controlled AGV system, International Journal of Production Economics 83, pp. 309–324.

Reveliotis, S. A. (2000), Conflict resolution in AGV systems, IEEE Transactions 32, pp. 647–659.

Yeh, M. S. and Yeh, W. C. (1998), Deadlock prediction and avoidance for zone-control AGVs, International Journal of Production Research 36(10), pp. 2879–2889.

Integration of Routing and Resource Allocation in Dynamic Logistic Networks

Thomas Huth, Dirk C. Mattfeld

Technical University Braunschweig, Chair for Information Systems and Decision Support, Braunschweig, Germany
{t.huth,d.mattfeld}@tu-bs.de

Abstract This paper gives an integrated view on two important issues in freight transportation. One is the pickup and delivery problem, the other is the resource allocation problem. In the pickup and delivery problem, re-usable resources are transported by trucks between locations. Additionally, these charge carriers have to be balanced in the transportation network on a rolling planning horizon. We introduce a mathematical formulation of this problem integrating routing and allocation. In a dynamic environment, integration strategies come along with advantages and disadvantages. We discuss these strategies and issues of their application.

1 Introduction

Hub-to-hub transportation problems, become increasingly important in order to exploit advantages by consolidating freight (Grünert and Sebastian, 2000). Crainic (Crainic, 2003) focuses on evolutions and consequences in the transportation sector. Special attention receives modeling of transportation systems, network design as well as resource allocation issues (Golden and Assad, 1988), (Crainic and Laporte, 1997), (Crainic and Laporte, 1998), (Crainic et al., 2007).

On a tactical level, the allocation and repositioning of resources is an important research area. This contribution combines this problem with an operational issue, the routing problem. An integrated view promises better results because of synergy effects.

Because of the multi-period situation, the problem becomes dynamic. Under these circumstances, a trade off between the more long term allocation and the more short term vehicle routing has to be found (Psaraftis, 1988). The integrated view on the swap trailer problem can suggest some approaches to deal with the operational and tactical horizon.

The paper is organized as follows. Section 2 describes the special properties of the problem. In Sect. 3 we formulate the mathematical model of the swap trailer problem by integrating a pickup and delivery model and a set partitioning model.

H.-D. Haasis et al., *Dynamics in Logistics*
DOI: 10.1007/978-3-540-76862-3,

Afterward, in Sect. 4 some properties occurring in a dynamic case are discussed and approaches to deal with such situations are presented. A conclusion in Sect. 5 closes the contribution and gives an outlook on our further research.

2 Problem Description

Today, large parcel service providers operate in hub-and-spoke networks (Grünert and Sebastian, 2000), (Huth and Mattfeld, 2007a). Customers are located in particular service regions and are assigned to hubs. We consider the daily hub to hub transports occurring between all hubs. Transports are performed by swap trailers, i.e. vehicle independent units comparable with containers, which have a standardized size and can be easily deposited. One transportation request is represented by one swap trailer. A trailer truck can transport at most two swap trailers, a sufficient number of swap trailers is available. Between every pair of hubs, one or more swap trailers are shipped by one or more trucks. Transports are performed by 3rd party carriers paid on the basis of distances only.

One task is the construction of routes, connecting hubs. Another task is to provide swap trailers in a sufficient number at the hubs over time. Swap trailers have to be allocated according to the demand of the forthcoming periods, determined by the forecasted requests. To satisfy the demand at a hub, the currently available swap trailers minus outgoing requests plus incoming requests are considered. Additionally, deadheads may be required to meet the demand. The origin hub of deadheads is subject of selection by the parcel service provider.

Next to deadheads, we consider that empty swap trailers can be transported as a by-product of routing (if there is capacity available). In contrast to transportation requests, we refer to these as allocation requests. This way, costly deadheads may be saved by integrating allocation requests in the preceding routing phase. Therefore, the selection of supply hubs is a degree of freedom for the planner. Hence, entrainment of empty swap trailers, detours for loaded swap trailers and still deadheads can occur.

These phases in the course of time are shown in Fig. 1. A regular shippers day begins with the loading of swap trailers in the evening. By night the transportation requests have to be routed by carriers and can be unloaded at the delivery hubs in the morning. If there is a lack of empty swap trailers in the next period (the demand exceeds the available amount), deadheads during the day may be necessary.

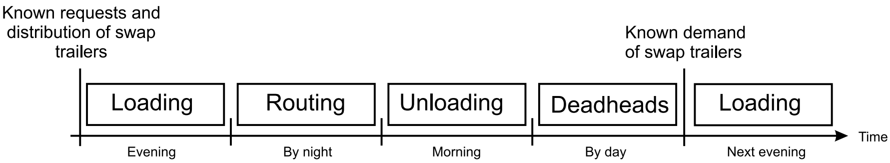


Fig. 1 The swap trailer problem in the course of time

3 Mathematical Model

We model the hubs as nodes and distances between hubs as weighted arcs. A request denotes a swap trailer and a sequence of nodes, traveled by a truck, determines a route. The assignment of requests to trucks leads to a solution of the problem. The goal is to fulfill all requests in a distance minimal sequence of nodes. This is can be formulated as **multi-depot pickup and delivery problem** according to (Savelsbergh and Sol, 1995).

The demand of swap trailers is known for the next period. For the allocation decision the supply node for a demand of a empty swap trailer in the next period has to be chosen. This can be formulated as a **set partitioning** problem. A decision variable describes the selection of supply nodes. The decision, how to treat them (deadhead, detour or entrainment), is now up to the routing problem. For literature about the allocation sub-problem see (Crainic et al., 1993), (Holmberg et al., 1998).

This section provides the mathematical formulation of the model. We introduce the adapted pickup and delivery model of Savelsbergh and Sol (Savelsbergh and Sol, 1995) to formulate the routing problem and integrate the allocation problem. A current survey on pickup and delivery models can be found in (Parragh et al., 2006a), (Parragh et al., 2006b).

The swap trailer problem is defined on a graph G containing nodes N and arcs E. All hubs are located in N. The requests $a \in TA$ consist of a pickup location R^+ and a delivery location R^-. $R = R^+ \cup R^-$ and $R \subseteq N$. Allocation requests (possible allocations to match the future demand) form the set TA_{empty} and transportation requests form the set TA_{load}. $TA = TA_{empty} \cup TA_{load}$. The trucks $k \in K$ have an origin depot D^+ and a destination depot D^-. $D = D^+ \cup D^-$ and $D \subseteq N$.

Pickup and delivery nodes: The set of pickup/delivery locations are determined by the pickup/delivery nodes of all requests.

$$R^+ := \bigcup_{a \in TA} R_a^+ \qquad R^- := \bigcup_{a \in TA} R_a^-$$

Load: The required capacity for a swap trailer counts one. The capacity of a truck trailer decreases at a pickup location and increases at a delivery location.

$$q_a = \sum_{i \in R_a^+} q_i = -\sum_{i \in R_a^-} q_i = 1 \qquad \forall a \in TA$$

Depots: In order to express that each node starts and ends in arbitrary nodes, each node is associated with a depot. The start depots of trucks are located in the origin nodes k^+ and the end depots are located in the destination nodes k^-.

$$D^+ := \{k^+ \mid k \in K\} \qquad D^- := \{k^- \mid k \in K\}$$

Decision variables: Sequence decision variable x_{ij}^k holds if truck k travels from node i to node j. Selection decision variable y_a holds if request a is accepted for routing and vector y finally forms the request set to be routed. Assignment decision variable z_a^k holds if request a is assigned to truck k.

$$\begin{aligned} x_{ij}^k &\in \{0,1\} && \forall i,j \in R \cup D, k \in K \\ y_a &\in \{0,1\} && \forall a \in TA \\ z_a^k &\in \{0,1\} && \forall a \in TA, k \in K \end{aligned}$$

Objective function: Minimization of the total traveled distance of all trucks between request and depot nodes with the distance coefficient c_{ij}.

$$(MODEL-GPDP) \sum_{i \in R \cup D} \sum_{j \in R \cup D} \sum_{k \in K} x_{ij}^k c_{ij} \rightarrow \text{min!} \quad (1)$$

subject to:

Truck constraints: A request a is assigned to exactly one truck k (2). Origin and destination node of truck k are the depots (3 and 4). Furthermore, (5) binds the two decision variables and ensures, that a truck k leaves or enters a node l if an request exists (z_a^k holds) and this request is choosen for routing (y_a holds). We describe the selection task later in the model.

$$\sum_{k \in K} z_a^k = 1 \qquad a \in TA \quad (2)$$

$$\sum_{j \in R^+ \cup D^-} x_{k^+ j}^k = 1 \qquad \forall k \in K \quad (3)$$

$$\sum_{i \in R^- \cup D^+} x_{ik^-}^k = 1 \qquad \forall k \in K \quad (4)$$

$$\sum_{i \in R \cup D} x_{li}^k = \sum_{i \in R \cup D} x_{il}^k = z_a^k * y_a \qquad \forall l \in R_a^+ \cup R_a^- ;\ a \in TA;\ k \in K \quad (5)$$

Time constraints: The auxiliary variable for time balance d_i describes the point of time, node i is visited. The departure time at the pickup node p is less or equal to the arrival time at the corresponding delivery node q (6). Time at pickup node i plus travel time t_{ij} between i and j is the arrival time at node j (7) if this arc is traveled. One may linearize this term by multiplying both sides with the decision variable x_{ij}^k. The next constraints set up start time to zero and prevent negative times at every node (8 and 9).

$$d_p \leq d_q \qquad \forall p \in R_a^+, q \in R_a^-, a \in TA \quad (6)$$

$$x_{ij}^k = 1 \Rightarrow d_i + t_{ij} \leq d_j \qquad \forall i,j \in R \cup D, k \in K \quad (7)$$

$$d_{k^+} = 0 \qquad \forall k^+ \in D^+ \quad (8)$$

$$d_i \geq 0 \qquad \forall i \in R \cup D \quad (9)$$

Capacity constraints: The auxiliary variable for load balance equation l_i states the used capacity at node i. Equality (10) ensures, that the load at node i plus the request at node i is equal to the load at node j if the arc between these nodes is traveled. All loads must be less or equal the truck capacity and negative loads are not permitted.

$$x_{ij}^k = 1 \Rightarrow l_i + q_i = l_j \qquad \forall i,j \in R \cup D, k \in K \tag{10}$$

$$0 \leq l_i \leq 2 \qquad \forall i \in R \cup D \tag{11}$$

Set partitioning constraint: e_{ab} is the set partitioning matrix with all necessary tasks $b \in T$. These set of tasks fulfill the transportation requests (TA_{load}) and satisfy the demand of empty swap trailers (TA_{empty}). The allocation requests can be fulfilled from several supply nodes but to fulfill a demand of one swap trailer is just one task. Consequently, only one supply node has to be chosen. Because of the bijective relationship between a and b in a transportation request, there is only one possibility to choose y_a. In contrast, the set partitioning constraints (12) prohibit selecting more allocation requests than needed. The constraint ensures, that each task $b \in T$ is fulfilled exactly once.

$$\sum_{a \in TA} e_{ab} y_a = 1 \qquad \forall b \in T \tag{12}$$

Binary variables:

$$x_{ij}^k \in \{0,1\} \qquad \forall i,j \in R \cup D, k \in K \tag{13}$$

$$y_a \in \{0,1\} \qquad \forall a \in TA \tag{14}$$

$$z_a^k \in \{0,1\} \qquad \forall a \in TA, k \in K \tag{15}$$

Multi request property: Firstly, there is more than one request per node. We have to adapt the data structure to incorporate this property. The most basic way is the insertion of virtual nodes. These nodes will be located at the same coordinates and must be integrated for every extra request at the same node. This can be achieved with an enlargement of the distance matrix, shown in Table 1. For the second request from Node 1 to Node 2, an extra Node $1'$ (pickup) and Node $2'$ (delivery) will be generated.

Secondly, the distance matrix can be adapted to incorporate precedence constraints. Basically, one may forbid direct connections between delivery nodes and pickup nodes of the same request and in this sequence. In a preprocessing step, these distances can be set infinite, shown in the third part of the table.

Depot location property: The depot locations and the request locations are within the set of hubs, so we must adapt the distance matrix once again. This way, a model adaptation can be avoided. Every hub can be a depot and a request node. For the simple example given in Table 1, the corresponding distance matrix with depots is given in Table 2. Depot 1 is located at Node 1 and Depot 2 is located at Node 2. Direct connections between depots are changes between trucks with zero costs, underlined in this example.

Table 1 Adaptation of the distance matrix to incorporate one more request in node 1

i/j	1	2	i/j	1	2	$1'$	$2'$	i/j	$1+$	$2-$	$1'+$	$2'-$
1	∞	c_{12}	1	∞	c_{12}	0	c_{12}	$1+$	∞	c_{12}	0	c_{12}
2	c_{21}	∞	2	c_{21}	∞	c_{21}	0	$2-$	$\underline{\infty}$	∞	c_{21}	0
			$1'$	0	c_{12}	∞	c_{12}	$1'+$	0	c_{12}	∞	c_{12}
			$2'$	c_{21}	0	c_{21}	∞	$2'-$	c_{21}	0	$\underline{\infty}$	∞

Table 2 Adaptation of the distance matrix to incorporate the depots

i/j	1	2	$D1$	$D2$	i/j	1	2	$D1$	$D2$
1	∞	c_{12}	0	c_{12}	1	∞	c_{12}	$\underline{\infty}$	$\underline{\infty}$
2	c_{21}	∞	c_{21}	0	2	c_{21}	∞	c_{21}	0
$D1$	0	c_{12}	∞	$\underline{0}$	$D1$	0	$\underline{\infty}$	∞	0
$D2$	c_{21}	0	$\underline{0}$	∞	$D2$	c_{21}	$\underline{\infty}$	0	∞

As applied for the multi request property, a preprocessing step can be applied. Direct connections between depot and delivery node respectively between pickup node and depot are not allowed and underlined in the second part of Table 2. The introduced pickup and delivery model is now able to operate with this structured data.

This integration approach enables simultaneous decision making and thus synergies by considering both problems at once. But the disadvantage is obvious. The *MODEL – GPDP* can be applied on a rolling planning horizon in order to meet the requirements of a dynamic environment. However, effects of balancing allocation requests between periods cannot be achieved this way and requests in further periods cannot be anticipated. So we have a myopic solution and neglect events in further periods. One may expect an advantage by applying tactical knowledge, so we propose a coarse-grained model for allocation decisions (Powell, 2003) on a tactical level.

4 Strategy for a Dynamic Environment

Within a multi-period environment, flows satisfying demands at all nodes and in all periods has to be determined. This is achieved by a multi-stage transportation model, similar to the dynamic transportation problem proposed by (Bookbinder and Sethi, 1980). The output of the model are transportation flows within a time-space network. These flows are transformed into allocation requests. Through passing these requests to the *MODEL – GPDP*, they will be routed.

In contrast to the classical transportation problem, the decision variable x needs a third index for the time dimension. So x_{ij}^t is a natural number that describes the flow of empty swap trailers from node i to node j in period t. In Period 0 the initial amount of empty swap trailers has to meet the demand at this node because there is no preceding period to deliver these nodes. A certain stock of resources r_i^t and a demand b_i^t is attributed to a node i in period $t \in T$.

$$(MODEL - msTP) \sum_{t \in T} \sum_{i \in N} \sum_{j \in N} x_{ij}^t c_{ij} \rightarrow \text{min!} \tag{16}$$

Subject to:

$$r_i^{t+1} = \sum_{h \in N} x_{hi}^t \qquad \forall i \in N, t \in T \tag{17}$$

$$r_i^t = \sum_{j \in N} x_{ij}^t \qquad \forall i \in N, t \in T \tag{18}$$

$$r_i^t \geq b_i^t \qquad \forall i \in N, t \in T \tag{19}$$

$$r_i^t, x_{ij}^t \in \mathbb{N} \qquad \forall i,j \in N, t \in T \tag{20}$$

The objective is to minimize the distance of all flows in the network (16). The balancing constraint (17) ensures that the flows going into node i equals the available resources of node i in period $t+1$. The available resources determine the outgoing flows (18). Furthermore, there must be at least as many empty swap trailers as needed in node i (19). r_i^t and x_{ij} are natural numbers (20).

The solution on this tactical level for the next period goes into the $MODEL-GPDP$ without the selection vector $\mathbf{y}$ in constraint (5) and (14) because of the presetting of these allocation requests. Afterward, we have to distinguish four different types of requests in the solution.

1. Transportation requests of the next period TA_{load} which must be fulfilled.
2. Allocation requests in TA_{empty} which are needed for demand in the next period. Deadheads must be realized for these requests.
3. Allocation requests in TA_{empty} which can be transported without deadheads. These are requests, carried on trucks in unity with transportation requests TA_{load}.
4. Allocation requests in TA_{empty} which are needed in a forthcoming period but nut in the next period. These requests can be delayed for fulfillment in later periods.

This distinction can be made after the routing decision. Only the first three sets must (1. and 2.) and should (3.) be fulfilled. An interesting set of requests is the fourth one. An examination, concerning the time of fulfillment, has to be made. These costly deadheads can be delayed if the following reason is considered. The forecasts are uncertain, which has two consequences: Deadheads may be cheaper fulfilled in forthcoming periods or deadheads must not be executed because the corresponding requests do not appear.

The advantage of the $MODEL-GPDP$ turns into the weakness of this approach. A preceding determination of requests which is a consequence of the problem integration biases the routing decision. The decision space shrinks up but we integrate tactical knowledge from the transportation model into the routing model on the operational level. An appraisement of the two proposed integration strategies can be found by computational tests.

Exact solutions for this $\mathcal{NP}$-hard problem cannot be found in reasonable time. A local search procedure can solve the integrated problem. Recent advantages in local search can be found in (Ropke and Pisinger, 2006). Summarizing the results obtained in (Huth and Mattfeld, 2007b), we can state that the $MODEL-msTP$ in conjunction with the pure GPDP (without selection) outperforms the $MODEL-GPDP$, but both integration strategies dominate the sequential decision making with routing and deadheads described in Fig. 1.

5 Conclusion

The swap trailer problem is a highly relevant problem in the transportation sector. Either, the pickup and delivery sub-problem and the allocation sub-problem are hard to solve. We propose an integrated view on these two problems in order to achieve a mathematical formulation.

The problem becomes more challenging within a multi-period environment. The demand of empty swap trailers in forthcoming periods should be taken into account (Cheung and Powell, 1996), (Powell, 1987). A forecasting component may cover this aspect, but the question is: How to tackle the allocation sub-problem? Decisions in the current period affect future periods. For this reason, an approach is needed, taking into account information from the tactical level. Therefore, a more long-term view for this part of the problem is needed (Psaraftis, 1988). We propose a multi-stage transportation problem on a tactical level and integrate resulting flows into the swap trailer model. Moreover, a weak point in this consideration is the anticipated demand of the forecasting component. Depending on the correctness or the degree of reliance of the forecasts, deadheads in one period could be accomplished as entrainments in preceding periods or could be dropped out. A research issue is the advantage of flexible solutions to overcome imprecision of forecasts and respect the long-term aspect of the allocation.

References

Bookbinder, Sethi (1980) The dynamic transportation problem: A survey. Naval Research Logistics 27

Cheung RM, Powell W (1996) Models and algorithms for distribution problems with uncertain demands. Transportation Science 30(1):43–59

Crainic T (2003) Long-haul freight transportation. In: Hall R (ed) Handbook of Transportation Science, Kluwer Academic Publishers

Crainic T, Laporte G (1997) Planning models for freight transportation. European Journal of Operational Research 97(3):409–438

Crainic T, Laporte G (eds) (1998) Fleet Management and Logistics. Kluwer Academic Publishers

Crainic T, Gendreau M, Dejax P (1993) Dynamic and stochastic models for the allocation of empty containers. Operations Research 41(1):102–126

Crainic T, Di Francesco M, Zuddas P (2007) An optimization model for empty container reposition under uncertainty. In: Sixth Triennial Symposium on Transportation Analysis (TRISTAN)

Golden B, Assad A (eds) (1988) Vehicle Routing: Methods and Studies. Elsevier Science Publishers B.V. (North Holland)

Grünert T, Sebastian HJ (2000) Planning models for long-haul opertions of postal and express shipment companies. European Journal of Operational Research 122(2):289–309

Holmberg K, Joborn M, Lundgren J (1998) Improved empty freight car distribution. Transportation Science 32(2):163–173

Huth T, Mattfeld D (2007a) Dynamics of the multi-depot pickup and delivery problem. In: Sixth Triennial Symposium on Transportation Analysis

Huth T, Mattfeld D (2007b) Integration of routing and resource allocation in a dynamic logistics network. submitted to Transportation Research Part C: Emerging Technologies

Parragh S, Doerner K, Hartl R (2006a) A survey on pickup and delivery models part i: Transportation between customers and depot. Tech. rep., University of Vienna

Parragh S, Doerner K, Hartl R (2006b) A survey on pickup and delivery models part ii: Transportation between pickup and delivery locations. Tech. rep., University of Vienna

Powell W (1987) An operational planning model for the dynamic vehicle allocation problem with uncertain demands. Transportation Research Part B 21(3):217–232

Powell W (2003) Dynamic models of transportation operations. In: Kok A, Graves S (eds) Supply Chain Management: Design, Coordination and Operation, Handbooks in Operations Research and Management Science, vol 11, Elsevier Science Publishers B.V., pp 451–516

Psaraftis H (1988) Dynamic vehicle routing problems. In: Golden B, Assad A (eds) Vehicle Routing: Methods and Studies, Elsevier Science Publishers B.V. (North Holland), pp 223–248

Ropke S, Pisinger D (2006) An adaptive large neighborhood search heuristic for the pickup and delivery problem with time windows. Transportation Science 40(4):455–472

Savelsbergh M, Sol M (1995) The general pickup and delivery problem. Transportation Science 29(1):17–29

Dynamic Vehicle Routing with Drivers' Working Hours

Asvin Goel

MIT-Zaragoza International Logistics Program, Zaragoza Logistics Center
Avda. Gómez Laguna 25, 1^a Planta, 50009 Zaragoza, Spain
e-mail: agoel@zlc.edu.es

Abstract applications, restrictions to drivers' working hours have only received very little attention in the vehicle routing literature. In this paper we consider a rich dynamic vehicle routing problem in which transportation requests arrive dynamically and the new EU drivers' working hour regulations, that entered into force in April 2007, must be considered. We present two algorithms for solving the problem under investigation and perform computational experiments that clearly indicate that regulations concerning drivers' working hours can have a significant impact on the performance of algorithms which are competitive when drivers' working hours are not considered.

1 Introduction

In April 2007 new regulations on drivers' working hours came into effect in the European Union. According to the new regulations motor carriers must organise the work of drivers in a way that drivers are able to comply with these regulations. Furthermore, carriers can be made liable for infringements committed by the drivers. Therefore, carriers must be able of generating vehicle tours considering information regarding driving times, breaks and rest periods. The dynamic nature of road transport creates a challenge as actual and planned driving and rest periods may not coincide. Fleet telematics systems allow to capture the vehicle's position and the begin and end of driving and rest periods. As described in Goel and Gruhn (2005) information obtained by a typical commercial off-the-shelf fleet telematics system can be automatically transferred to the carrier's order & fleet management system, allowing for dynamic vehicle routing considering drivers' working hours. As there is usually only little time to manually resolve infeasibilities arising due to discrepancies between real-life problem and model representation, models used in dynamic vehicle routing must narrow this gap as much as possible. In this paper we consider a rich dynamic vehicle routing problem in which transportation requests arrive dynamically, i.e. execution of routes as well as load acceptance and routing decisions

H.-D. Haasis et al., *Dynamics in Logistics*
DOI: 10.1007/978-3-540-76862-3,

must be carried out simultaneously. This problem extends the General Vehicle Routing Problem (GVRP) introduced by Goel and Gruhn (2007) by considering drivers' working hours as presented in Goel and Gruhn (2006).

This paper is organised as follows. First we give an overview over related literature in Sect. 2. In Sect. 3 we present the General Vehicle Routing Problem (GVRP). In Sect. 4 we describe the regulations concerning drivers' working hours and how they can be considered within the GVRP. In Sect. 5 we present a Reduced Variable Neighbourhood Search algorithm and a Large Neighbourhood Search algorithm for the dynamic General Vehicle Routing Problem with Drivers' Working Hours (GVRP-DWH). In Sect. 6 we evaluate the algorithms presented in the previous sections and report computational results.

2 Related Literature

A comprehensive discussion of dynamic vehicle routing can be found in Psaraftis (1988) and Psaraftis (1995). Dynamic real-life problems often require rich models, in most of the literature on dynamic routing problems, however, some simplifying assumptions are made. For example, in the dynamic full-truckload PDP, which recently has received increasing attention, see Fleischmann et al. (2004), Yang et al. (2004), and Powell et al. (2000), each vehicle can only carry one transportation request at a time and cannot load further shipments until all currently loaded shipments are unloaded. The only works known to the author considering rich VRP in a dynamic context are presented by Savelsbergh and Sol (1998) and Goel and Gruhn (2007). This paper extends the General Vehicle Routing Problem presented in the latter paper by considering drivers' working hours as presented in Goel and Gruhn (2006). For a more comprehensive discussion of literature related to the General Vehicle Routing Problem the reader is referred to Goel and Gruhn (2007) and for further information on approaches considering drivers' working hours to Goel and Gruhn (2006).

3 The General Vehicle Routing Problem

In the General Vehicle Routing Problem (GVRP) a transportation request is specified by a nonempty set of pickup, delivery and/or service locations which have to be visited in a particular sequence by the same vehicle, the time windows in which these locations have to be visited, and the revenue gained when the transportation request is served. Furthermore, some characteristics can be specified which constrain the possibility of assigning the transportation requests to certain vehicles due to compatibility constraints and capacity constraints. At each of the locations some shipment(s) with several describing attributes can be loaded or unloaded. In contrast to many other commonly known routing problems, not all transportation requests have to be assigned to a vehicle. Instead, a so-called make-or-buy decision is necessary to determine whether a transportation request should be assigned to a self-operated vehicle (make) or not (buy). A fleet of heterogeneous vehicles is available to serve the transportation requests. The vehicles can have different capacities, as well as different travel times and travel costs between locations. The vehicles can transport shipments which require some of the capacity the vehicle supplies. Instead of assuming that each vehicle becomes available at a central depot, each vehicle is given a start location where it becomes available at a specific time and with a specific load. Furthermore, the vehicles do not have to return to a central depot and for each vehicle a final location is specified, which has to be reached within a specific time and with a specific load. Each vehicle may have to visit some locations in a particular sequence between leaving its start and reaching its final location. All locations have to be visited within a specific time window. If the vehicle reaches one of these locations before the begin of the time window, it has to wait. A tour of a vehicle is a journey starting at the vehicles start location and ending at its final location, passing all other locations the vehicle has to visit in the correct sequence, and passing all locations belonging to each transportation request assigned to the vehicle in the correct respective sequence. A tour is feasible if and only if for all orders assigned to the tour compatibility constraints hold and at each point in the tour time window and capacity restrictions hold. The objective is to find distinct feasible tours maximising the profit, which is determined by the accumulated revenue of all served transportation requests, reduced by the accumulated costs for operating these tours.

4 Drivers' Working Hours

Since April 2007 drivers' working hours in the European Union are regulated by EC regulation 561/2006. According to the new legislation motor carriers must organise the work of drivers in such a way that drivers are able to comply with the regulations and are made liable for infringements committed by the drivers. The most important regulations according to the new legislation are:

- After a driving period of four and a half hours a driver shall take an uninterrupted break of not less than 45 minutes, unless he takes a rest period. During a break a driver must not drive or undertake any other work.
- The daily driving time between the end of one rest period and the beginning of the following daily period shall not exceed 9 hours. A daily rest period is any period of at least 11 hours during which a driver may freely dispose of his time.
- In a multiple manned vehicle, a driver may take a break on the moving vehicle whilst another driver is driving. The daily rest period in which the vehicle must be stationary may be reduced to 9 hours.
- The weekly driving time shall not exceed 56 hours.
- A weekly rest period shall start no later than 144 hours after the end of the previous weekly rest period.

According to these regulations, the state of a driver at any point in a tour can be represented by a label (l = *ARR, WDT, DDT, NDT*), where *ARR* denotes the arrival time at the point, *WDT* the weekly driving time, *DDT* the daily driving time, and *NDT* the non-stop driving time until reaching the point. The label for the starting point of a tour depends on the actual driving and rest periods of the driver. For all succeeding points in the tour a label l is feasible if there exists a schedule of driving, breaks and rest periods for the partial tour until that point which is in compliance with EU legislation such that the label l represents the resulting driver state and that the arrival time is within time windows. Obviously, there are different ways how to schedule driving and rest periods given a specific tour, and therefore, a point in the tour of a vehicle may have different potential feasible labels. Goel and Gruhn (2006) showed how, given a specific tour, all potential schedules for driving, breaks and rest periods can be determined efficiently by eliminating unnecessary labels using a domination criterion.

The General Vehicle Routing Problem with Drivers' Working Hours (GVRP-DWH) extends the GVRP by the requirement that there must exist a feasible label for each point in the tour of every vehicle. Given any sequence of locations specifying a partial tour, the labels for these locations can be determined as described in Goel and Gruhn (2006).

5 Solution Approaches

The Reduced Variable Neighbourhood Search algorithm and the Large Neighbourhood Search algorithm presented by Goel and Gruhn (2007) for the GVRP can also be used for the GVRP-DWH as long as driver labels are recalculated after every modification of a tour. Both algorithms start from an initial solution obtained by applying the auction insertion algorithm that is used in the cited paper. Then, the improvement methods try to find new solutions while incorporating modified problem data arriving dynamically. In the following, the two approaches are briefly described.

5.1 Reduced Variable Neighbourhood Search

The Reduced Variable Neighbourhood Search (RVNS) algorithm changes the neighbourhood structure during the search by selecting different neighbourhoods defined by the following elementary neighbourhood operators:

- The INSERT-operator randomly chooses an unscheduled order and inserts it to the tour of the vehicle with lowest incremental costs.
- The REMOVE-operator randomly chooses a scheduled order and removes it from the tour it is assigned to.

- The RELOCATE-operator randomly chooses a scheduled order and removes it from the tour it is assigned to. The order is inserted to the tour of the vehicle with lowest incremental costs.
- The REPLACE-operator randomly chooses an order in the tour of a vehicle and removes it from the tour. Then, another order is inserted to the tour from which the first order was removed.
- The SWAP-operator randomly chooses two scheduled orders and inserts them to the respective other tour.

Given the neighbourhood structures defined by these operators we can now describe the RVNS algorithm as follows:

1. **Initialisation:** Find an initial solution; choose a stopping condition
2. **Repeat** the following until the stopping condition is met:
   a. **Shaking:** Generate a new solution in one of the neighbourhoods of the current solution
   b. **Move or not:** If the new solution is feasible and better than the current solution, replace the current solution by the new solution

The choice of the neighbourhood in the first step of the iteration is made randomly.

5.2 *Large Neighbourhood Search*

The Large Neighbourhood Search (LNS) algorithm is based on repeated removals and re-insertions and can be described as follows:

1. **Initialisation:** Find an initial solution s; choose a stopping condition
2. **Repeat** the following until the stopping condition is met:
   a. **Remove:** Choose a number k and remove k orders from their tours
   b. **Re-insert:** Generate a new solution by applying an insertion method
   c. **Move or not:** If the new solution is feasible and better than the current solution, replace the current solution by the new solution

In the first step of the iteration the number of orders to be removed is randomly chosen between 2 and 30. Furthermore, the orders to be removed are also chosen randomly. In the second step the auction method is chosen for re-insertion.

6 Evaluation

In order to evaluate our algorithms we generated test problems similar to the ones used in Goel and Gruhn (2007). A heterogeneous vehicle fleet has been generated where some of the vehicles have refrigerated cargo bodies and some are manned by two drivers. We assume that all vehicles are en-route when planning starts and each vehicle becomes available somewhere around the central depot. All vehicles eventually have to return to the central depot. We randomly generated shipments according to the same frequency distribution for pickup and delivery locations used in Goel and Gruhn (2007). Transportation requests have been generated by choosing one full or half truckload shipment or by combining two half truckload shipments with identical pickup or delivery location. Some of the transportation requests require a vehicle with refrigerated cargo body, some require a vehicle manned by two drivers. Travel distances are based on the direct distances, however are multiplied by 1.3 in order to consider deviation occurring in road transport. Travel costs are proportional to the travel distance. Vehicles with refrigerated cargo bodies and vehicles manned by two drivers are more expensive than vehicles with standard cargo bodies and those manned by one driver. The revenue of the transportation requests is set to double the costs of the cheapest vehicle capable of transporting the shipments. That is, the shippers are not only willing to pay for the transport itself, but also for the return trip to the start location.

For our test cases we generated $|\mathcal{V}|$ vehicles and transportation request which become known at some time during our simulation of 10 hours. In the beginning only $|\mathcal{O}_0 vert$ orders are known to the carrier and every hour $|\mathcal{O}_t|$ new orders become known. For all orders we set the length of the time windows at each location to the same value τ, i.e. either 2 hours or 12 hours.

In our simulations the algorithms were only allowed 60 seconds of computing time per timestep (representing one hour in our simulation scenario) on a personal computer with Intel Pentium 4 processor with 3.00 GHz. All decisions taken in a preceding timestep may be revised in a subsequent timestep. However, at the end of each timestep all unscheduled transportation requests are assumed to be rejected or subcontracted by external carriers, i.e. whenever a transportation request is not assigned to the tour of a vehicle at the end of a timestep the transportation request is removed from the model. With the beginning of a new timestep the start location of every vehicle is updated in order to consider the vehicle's movement. Whenever the first location belonging to a transportation request is visited by a vehicle the transportation request is removed from the model and the corresponding locations are added to the sequence of locations the vehicle must visit before reaching the final location of the tour. New transportation requests are added to the model in the beginning of each timestep and are inserted to the tours by the auction method before the Reduced Variable Neighbourhood Search or Large Neighbourhood Search method is invoked.

Table 1 shows the results of our computational experiments on the instances p1 to p24. The average values of the objective function values obtained by multiple runs of our heuristics are listed in column AVG (f) and the mean absolute deviation in column AVG(Δ). As we can see the RVNS does not perform well for the GVRP-DWH instances and is significantly outperformed by the LNS algorithm for all problems. As shown in Goel and Gruhn (2007) the RVNS algorithm is competitive with the LNS algorithm for the GVRP. The reason why RVNS performs poorly in the presence of drivers' working hour regulations is probably the increased complexity of

Table 1 Results for the GVRP-DWH

Problems					RVNS		LNS	
No.	$\lvert\mathcal{V}\rvert$	$\lvert\mathcal{O}_0\rvert$	$\lvert\mathcal{O}_t\rvert$	τ	AVG(f)	AVG($\triangle$)	AVG(f)	AVG($\triangle$)
p1	100	300	20	2	91742.14	694.17	137381.34	2865.52
p2	100	300	20	2	86726.90	470.92	124830.97	584.05
p3	100	300	20	2	98827.95	840.22	135249.10	674.58
p4	100	300	20	2	103292.76	308.12	148652.44	1992.39
p5	100	300	20	12	155682.01	1071.18	202496.91	1468.62
p6	100	300	20	12	162604.68	635.21	203922.78	4081.44
p7	100	300	20	12	157589.90	1367.47	204054.91	3319.60
p8	100	300	20	12	164516.86	1894.68	215337.51	2429.59
p9	250	750	50	2	303372.62	1586.12	373465.09	5445.90
p10	250	750	50	2	297855.81	911.99	371404.20	2991.09
p11	250	750	50	2	279893.95	1113.82	361869.71	2308.78
p12	250	750	50	2	286444.81	1482.61	359553.26	6354.11
p13	250	750	50	12	484703.12	758.67	563105.14	4622.74
p14	250	750	50	12	457151.68	837.54	518417.30	4000.02
p15	250	750	50	12	456574.40	1088.74	532497.43	2976.30
p16	250	750	50	12	477127.29	1027.74	542029.30	7128.13
p17	500	1500	100	2	604958.04	803.52	670920.94	13598.94
p18	500	1500	100	2	630450.01	1148.93	691416.96	1400.10
p19	500	1500	100	2	629969.09	2712.67	713357.25	6686.20
p20	500	1500	100	2	591035.71	1963.82	665278.51	2206.50
p21	500	1500	100	12	930469.72	1310.01	999714.67	201.88
p22	500	1500	100	12	949612.72	1292.05	986565.17	6189.27
p23	500	1500	100	12	925413.76	3168.22	1004896.40	4767.71
p24	500	1500	100	12	950116.36	1186.17	1039078.49	4496.73

the problem and the higher likeliness of a disconnected search space if neighbourhoods are small. For $k = 2$ the neighbourhood structures of LNS and RVNS are very similar as all neighbourhood operators used by the RVNS method can be interpreted as a combination of at most two removals and subsequent insertions. For most cases, however, $k > 2$ and the larger average size of neighbourhoods used by the LNS algorithm is probably the reason why LNS significantly outperforms RVNS in the presence of drivers' working hour regulations.

7 Conclusions

In this paper we studied the dynamic General Vehicle Routing Problem with Drivers' Working Hours (GVRP-DWH) which extends the dynamic General Vehicle Routing Problem by considering the new EU legislation concerning driving times, breaks, and rest periods. We presented two heuristics for the GVRP-DWH and performed computational experiments. The computational experiments clearly indicate that regulations concerning drivers' working hours can have a significant impact on the performance of algorithms which are competitive when drivers' working hours are not considered. As carriers, according to the new EU legislation, must organise the work of drivers in such a way that drivers are able to comply with the regulations, it is inevitable to consider these regulations in vehicle routing. The LNS approach is well suited for tackling the additional complexity due to the bigger size of the neighbourhoods and the lower likeliness of getting trapped in a local optimum without chance of escaping to a better solution.

References

European Union, "Regulation (EC) No 561/2006 of the European Parliament and of the Council of 15 March 2006 on the harmonisation of certain social legislation relating to road transport and amending Council Regulations (EEC) No 3821/85 and (EC) No 2135/98 and repealing Council Regulation (EEC) No 3820/85," Official Journal of the European Union L 102, 11.04.2006, 2006.

B. Fleischmann, S. Gnutzmann, and E. Sandvoß. Dynamic vehicle routing based on on-line traffic information. Transportation Science, 38(4):420–433, 2004.

A. Goel and V. Gruhn. Integration of telematics for efficient management of carrier operations. In Proceedings of the IEEE International Conference on e-Business Engineering, p. 404–408, 2005. IEEE Society. ISBN 0-7695-2430-3.

A. Goel and V. Gruhn. Drivers' working hours in vehicle routing and scheduling. In Proceedings of the 9th International IEEE Conference on Intelligent Transportation Systems, p. 1280-1285, 2006. IEEE Society. ISBN: 1-4244-0094-5.

A. Goel and V. Gruhn. A General Vehicle Routing Problem, European Journal of Operational Research, doi:10.1016/j.ejor.2006.12.065. 2007.

W.B. Powell, M.T. Towns, and A. Marar. On the value of optimal myopic solutions for dynamic routing and scheduling problems in the presence of user noncompliance. Transportation Science, 34(1):67–85, 2000.

H.N. Psaraftis. Dynamic vehicle routing problems. In B.L. Golden and A.A. Assad, eds., Vehicle routing: Methods and studies, p. 233–248. North-Holland Amsterdam, 1988.

H.N. Psaraftis. Dynamic vehicle routing: Status and prospects. Annals of Operations Research, 61:143–164, 1995.

J. Yang, P. Jaillet, and H. Mahmassani. Real-time multi-vehicle truckload pickup-and-delivery problems. ransportation Science, 38(2):135–148, 2004.

RFID in Logistics and Manufacturing Networks

A Survey of RFID Awareness and Use in the UK Logistics Industry

Johann Riedel[1], **Kulwant S. Pawar**[1], **Stefano Torroni**[2], **Emilio Ferrari**[2]

[1] Centre for Concurrent Enterprise, Nottingham University Business School, Jubilee Campus, Wollaton Road, Nottingham, NG8 1BB, UK.
e-mail: {Johann.Riedel,Kul.Pawar}@Nottingham.ac.uk
[2] Department of Mechanical, Metallurgical Engineering (DIEM), University of Bologna, Viale Del Risorgimento 2, 40136 Bologna, Italy.
e-mail: storroni@libero.it, emilio.ferrari@mail.ing.unibo.it

Abstract The aim of this paper is to assess the readiness of UK logistics companies to adopt RFID technology, trying to understand whether, and when, it will be adopted in large scale. The first barrier to a new technologies adoption is awareness – are the potential users aware of the new technology and its potential benefits? Once potential users are aware of a new technology more practical and economic questions come to the fore – difficulty of implementation, availability, standards, cost, privacy, compatibility with existing business processes, etc. The paper investigates these issues and contrasts them against the interesting features and achieved benefits of RFID. A model of RFID diffusion is presented which shows that a high level of awareness (81%) is not yet matched by adoption (17%). However, diffusion has just overtaken the inflection point of the S-shaped curve that models the diffusion of an innovation, and is now approaching the stage in which the adoption rate is rapidly increasing. Adoption is being led by large companies. Cost, lack of practical knowledge about application, difficulty of demonstrating ROI and difficulty of attaching tags were the major barriers to adoption, followed by technological issues (integration, reliability, lack of standards). To sum up, in supply chain management, it is likely that UHF passive tags will be the most commonly used, and that pallets, cases and carton tagging, rather than item tagging, will be the most popular application of RFID technology. Moreover, it is predicted that, in the early phase, big companies will adopt it, paving the way to wider diffusion once return on investment is demonstrated, best practices developed and costs fall.

1 Introduction

Radio Frequency Identification (RFID) is basically a technology that allows to remotely identify an object using a radio link. In recent years, retail and manufacturing companies have become interested in applying RFID technology to their business processes, being attracted by the number of potential benefits achievable, such as improved supply chain visibility and improved process efficiency. Wal-Mart, Albert-

H.-D. Haasis et al., *Dynamics in Logistics*
DOI: 10.1007/978-3-540-76862-3,

sons and Target, three of the four largest retailers in the United States, have already required their major suppliers to deliver all pallets and cases with Radio-Frequency Identification (RFID) tags, starting in 2005. In Europe other big organizations such as Tesco and Metro are following the way.

The predictable impact of RFID technology on the supply chain is enormous. Supply chain management deals with the flow of goods and information, from suppliers to customers, aiming to deliver the maximum satisfaction of customers at the lowest price. In this context, RFID system represents the new way to complement materials with information, and the natural evolution of the barcode technology.

During November 2005, the "RFID awareness and use" survey was carried out in an effort to establish the readiness of UK logistics companies to adopt RFID technology and to find out whether and when it will be adopted on a large scale.

The aim of this survey was to assess the readiness of UK logistics companies to adopt RFID technology, trying to understand whether and when it will be adopted in large scale. The first barrier to a new technologies adoption is awareness – are the potential users aware of the new technology and its potential benefits? Once potential users are aware of a new technology more practical and economic questions come to the fore – difficulty of implementation, availability, standards, cost, compatibility with existing business processes, etc.

1.1 Objectives

The key issues to be explored are:

1. What is the awareness degree of RFID technology
2. What is the level of diffusion of the technology
3. What are the barriers to RFID adoption
4. How strategic do companies consider the priority of adopting RFID
5. Is there is a relationship between the awareness/use and the size of the company
6. Is there is a relationship between the awareness/use and the main business of companies
7. Whether, in the logistics sector, the kind of benefits achievable through RFID adoption, really address the business problems that companies currently have or not
8. Whether the cost is the main issue to the adoption or, even if the technology was cheap companies would not employ it
9. How the companies consider their approach to adopting new technologies

1.2 Sample Selection

The logistics industry and in general sectors with important supply chains are expected to be the most interested in RFID and to represent the fastest growing market for RFID application. The focus has thus been on companies involved in the haulage of goods, because of the key role they play in the introduction of RFID. UK based companies involved in the haulage of goods were chosen to participate in a post mail administered survey; the questionnaire was circulated to a total of 302 organisations.

In this research, an online database called Financial Analysis Made Easy (FAME) was used to select the companies to be surveyed. FAME is a database that contains information on 2.8 million companies in the UK and Republic of Ireland, 1.9 million of which are illustrated in a detailed format and the remaining are smaller businesses. FAME is run by Bureau van Dijk Electronic Publishing (BvDEP) and is updated nearly daily. For the top 1.9 million companies the reports typically include detailed contact information, company and activity details, a number of financial information such as profit and loss account, balance sheet items, cash flow and ratios, and names of current directors. The depth and accuracy of information coupled with its online availability were the main reasons why FAME was chosen to select the companies.

According to FAME, nearly 2,900 companies are engaged in UK and Republic of Ireland in the haulage of goods. This figure is the result of a query that includes only the companies whose main activity code is included in the following standard industrial classification (SIC, 2003) selection: 6010, 6024, 6100, 6220, 6311, 6312, 6321, 6322, 6323, 6340, 6411, and 6412.

RFID technology is still very expensive and it is unlikely that small organizations have the resources needed to implement it and accept the risk of adopting a technology not yet diffused. Thereby, we focused mainly on medium to very large size companies, which generally speaking might exert a greater influence on the RFID diffusion. It was decided to focus on UK companies because it is arguable that they are ahead in Europe in adopting RFID technology (De Jonge, 2004).

The questionnaire was designed to be as short as possible in order to encourage respondents to return it. It was 3 pages long and included 10 main questions plus a few other ones aimed to collect data about the respondent and the company.

Table 1 The Sample's business type distribution

Main business	Number	Percentage
Logistics	101	33,4
Freight forwarding	71	23,5
Warehouse and Distribution	59	19,5
Haulage contracting	27	8,9
Parcels and mail distribution, Couriers	27	8,9
Cargo handling	6	2,0
Technology provider	5	1,7
Other	6	2,0
Total	302	100,0

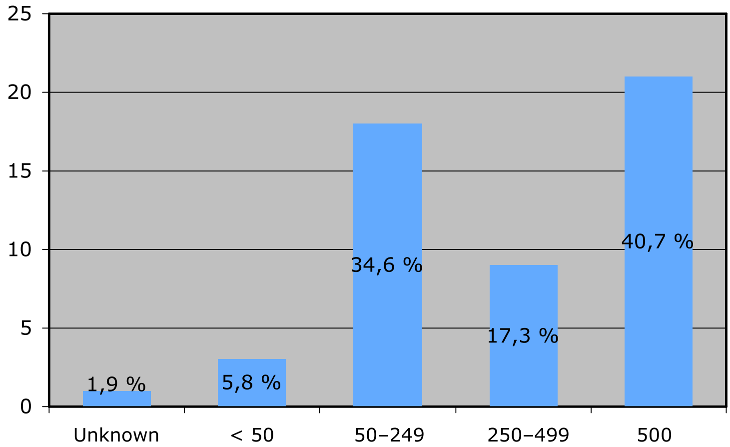


Fig. 1 Distribution of respondents by company employee size

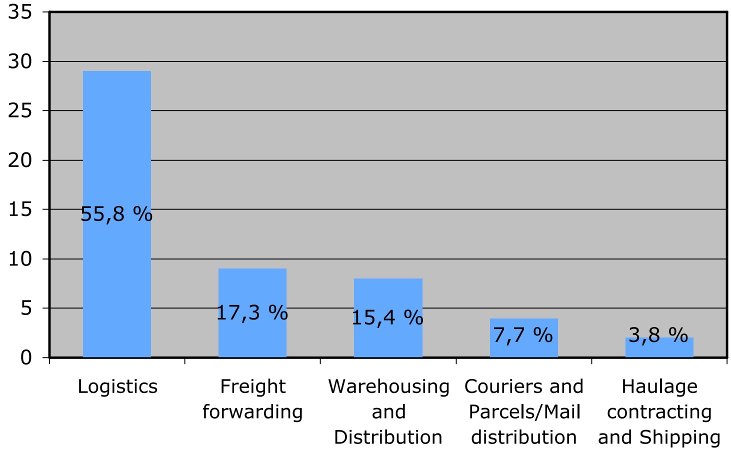


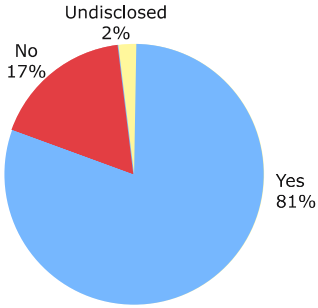
Fig. 2 Distribution of respondents by main business

Fifty two organizations responded to the survey, giving a response rate of 17.2 percent. Figures 1 and 2 show the distribution amongst the respondents for this survey, by company (employee) size and main business, respectively.

2 Degree of Awareness of RFID

81 percent of respondents said they were aware of RFID technology. Compared to previous surveys, it seems that companies are generally becoming more aware of RFID technology. This high percentage is encouraging and could anticipate the rapid diffusion of the technology in the logistics sector. The awareness degree is distributed almost equally among companies in the sector, independently of the main

Fig. 3 RFID Awareness



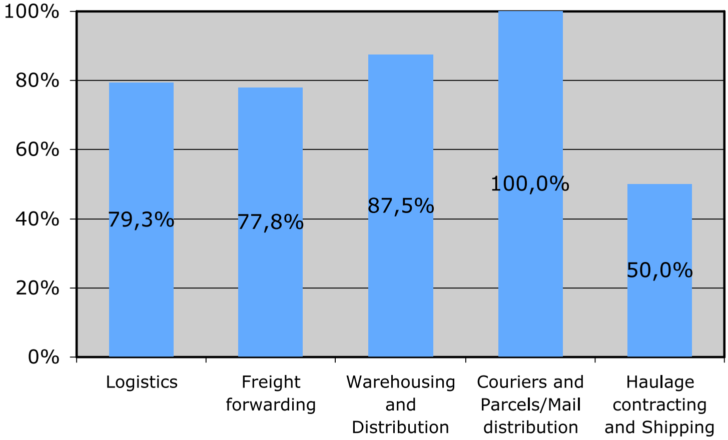


Fig. 4 RFID Awareness by business sector

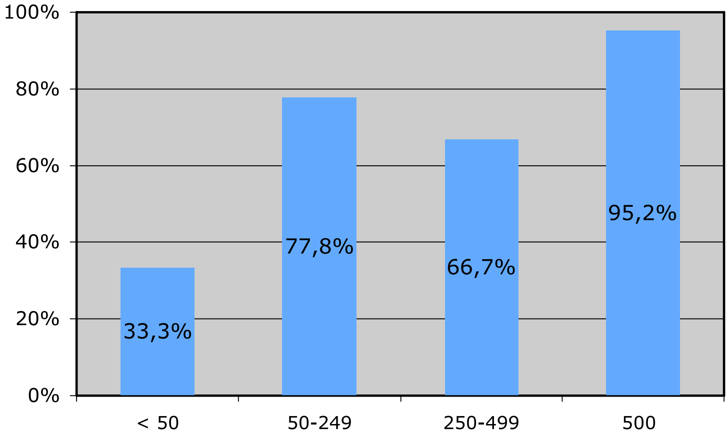


Fig. 5 RFID Awareness by company size

business; whereas it seems to be related to company size: it almost doubles going from small to large companies.

3 RFID Adoption and Diffusion

A single-choice question was used to find out in which stage companies are at the moment regarding RFID adoption. Seven alternatives were given (we have never investigated its applicability, we are carrying out an investigation into its applicability, at the moment this technology does not apply to us, we have already planned to invest in this technology in the near future, we are carrying out a pilot/testing project, we have adopted it in a limited area of our business, we have adopted it widely). In formulating this list it was attempted to identify what the stages/process through which a company passes when deciding to adopt an innovation, taking into account the work of three well-known researchers. Rogers (1995) breaks this process down into five main stages: in the first one "the organization is exposed to the innovation but lacks complete information about it". Zaltman (1973) states that an organization has two ways to detect a performance gap, (1) realizing that its performance is not satisfactory, thereby it has a look around to find a solution, or (2) becoming aware of a potential innovation in its environment, thereby realizing that by its introduction it could improve its performance. After the first stage, the organization "becomes interested in the new idea and seeks additional information about it". West (1990) defines these two stages as the phase in which the organization recognises the potential benefits of deploying the innovation.

Both Rogers and West agree that in the following stage the organization evaluates the applicability, for instance through a cost/benefits analysis or a feasibility study, at the end of which the decision to adopt or reject is taken. If adoption is chosen, the organization plans to test the innovation and allocates the required re-sources, both human and capital. In the final stage, if the test gave a good result, the innovation is deployed in a limited area of its business and only afterwards is it widely deployed, becoming part of the organization's system.

From the point of view of RFID technology adoption, half of the respondents said that they were not currently adopting the technology or had never investigated its applicability, 21% of respondents said they were carrying out an investigation and 17% were using RFID technology. Comparing these results with those from previous surveys, it seems that the diffusion level has risen.

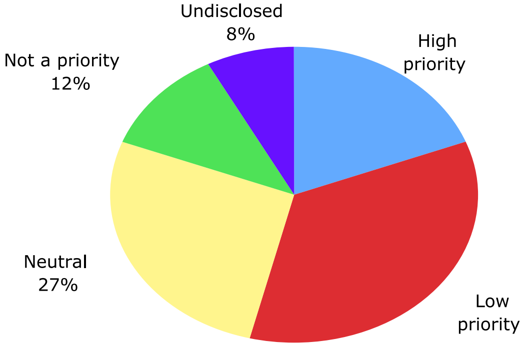


Fig. 6 Priority of RFID adoption

Only two out of nine companies that are using RFID technology consider themselves as technology innovators, whilst the majority consider themselves as first followers and 22% as slow adopters.

4 Modelling RFID Diffusion

Moreover, thinking of how strategic a priority companies consider RFID adoption, 53% of respondents said it was a priority, in opposition to 39% that said it was not.

Referring to the adoption and diffusion of innovations model, it is possible to position the logistics sector at a particular stage of the S-curve (see Fig. 7).

As a result of the above considerations, it seems that for UK logistics companies, RFID technology has already left the first stage and is moving along the section of the curve where the rate of adoption is rapidly increasing.

Furthermore, cross tabulating the adoption by the main business and company size, it is possible to notice that large companies are leading the adoption process.

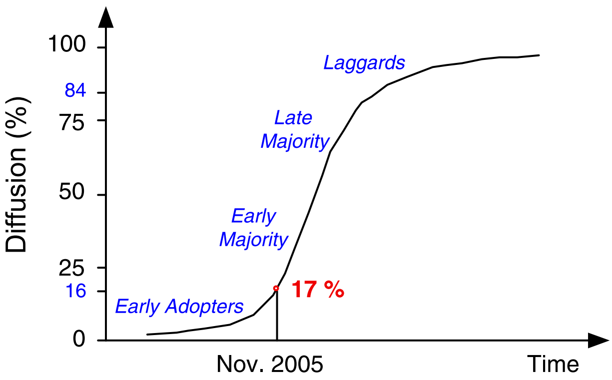


Fig. 7 RFID Adoption S-Curve

5 Barriers to RFID Adoption

The challenges in RFID adoption are from both the economic and technological point of view. Firstly the initial investment is very high, because it is not only a problem of the tag price. A company that wants to implement a full functional system incurs costs for readers, printers, middleware and software, plus it is likely that it will need consultancy support, employee training, system tuning and trials. Further cost comes from the need to redefine and develop new internal processes and to dedicate labour to manage the change. In the early period the system might not be very accurate and reliable and the company might incur many first-time mistakes. Moreover the system needs to be continuously powered and maintained (Fig. 8).

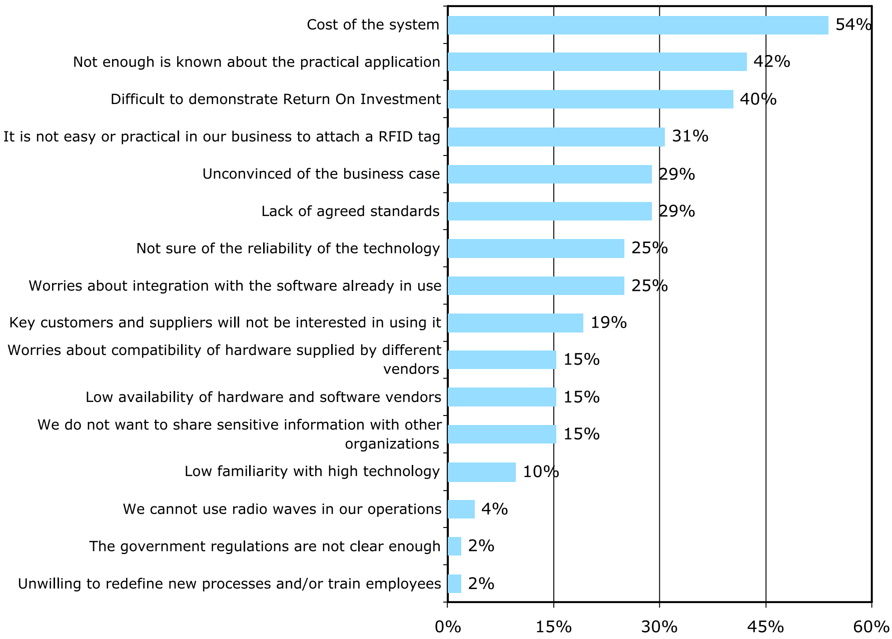


Fig. 8 Barriers to RFID Adoption

Next, because of the lack of long-term experience in using RFID systems in commercial environment, companies face difficulties to demonstrate the business case and the return on investment, and it is not yet possible to reduce cost by imitating industry best practices.

From the technological point of view, the lack of interoperability and communication standards, and the different frequency bands available in different countries, are the main issues. Furthermore, especially in the UHF band, as the technology is relatively immature, there are concerns about the reading accuracy and the system reliability. Also vendors' lack of RFID system expertise is an issue, especially in the domain of integration with enterprise software already in use, and the way the massive stream of data will be managed. Moreover, companies have to be prepared to share information with their partners, which is not always the case.

Next, the RFID system itself is not a plug and play technology, indeed it is quite complex and it is arguable that companies with limited resources would not be prepared to set it up. Another big issue is that manufacturers have to rethink the packaging of their products in order to include the tag, and in some environments placing it in the correct position could be not easy.

1	Cost of the system
= 2	Convenience of the investment (ROI, business case) Practical issues (practical application, ease of attaching tag)
4	Technological issues (standards, reliability, integration)
5	Other actors choice (customers/suppliers interest)

Money matters are in the first place, which agrees with analogous results from previous surveys, where costs and being able to demonstrate return on investment were mentioned as the most important obstacles. But the economical aspect, which appears to be vital, is not the key point that makes the difference between adoption and rejection.

As a proof, two thirds of respondents said they would adopt RFID technology if it was as cheap as Barcodes, which in absolute terms is a big percentage, but also reveals that companies are not stopped only by the required investment; 13% of the respondents said they would not deploy the technology even if it was cheap, and 21% were unsure (Fig. 9).

In the adoption of RFID technology, there is a big "network effect" that may encourage companies to invest in it as soon as it will become more common. Referring to the S-shaped curve of adoption presented above, this might be interpreted as a further signal that the inflection point has just been overcome and that in the next few years the diffusion would increase with an increased rate.

Going back to the barriers, the availability of equipment by different vendors and the compatibility between them seems not to be an important barrier. Further-more, it appears that companies are not worried about sharing sensitive information with their partners: this is an important result, as it indicates that companies are prepared to benefit the most from RFID technology in the future.

Lastly, redefining new work processes and training employees do not seem to be a barrier, nor is the impracticality of using radio waves in operations. Government regulations have been mentioned only by one respondent, which may indicate that UK companies are confident of the support by Government, in opposition to what French or Italian companies may feel.

Cross tabulating the barriers by the business type, it is revealed that for warehousing and distribution companies, apart from the costs, major barriers were practical and technical issues rather than money matters. For instance both "unconvinced of the business case" and "difficult to demonstrate ROI" are less mentioned compared to the responses given by the whole sample, whilst all the options that refer to deployment are mentioned more. This, result for warehousing combined with the higher RFID awareness degree compared to the average of the responses, might indicate that warehousing companies have actually a better understanding of the technology and from the economical point of view are more convinced of the busi-

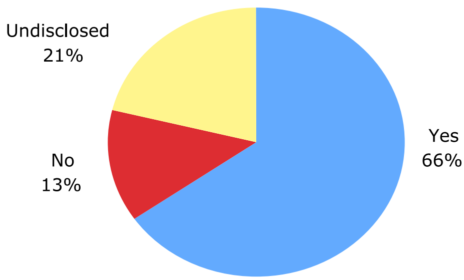


Fig. 9 Intention of RFID adoption

ness case than the average. But when dealing with practical deployment, they have experienced a number of difficulties, which could also explain why only one pilot is being carried out at the moment and none of the respondents was adopting it.

6 Conclusion

This survey focused on the assessment of the awareness and diffusion degree of RFID technology in the UK logistics sector. The questionnaire was circulated to a selection of 302 logistics companies based in the UK and the response rate achieved was 17 %.

The key findings of this investigation were:

- In the UK, as of the end of 2005, 80% of logistics companies are aware of RFID technology, and 17% are already using it.
- The diffusion of RFID technology in the UK logistics sector has just overtaken the inflection point of the S-shaped curve that models the diffusion of an innovation, and is now approaching the stage in which the adoption rate is rapidly increasing.
- Companies are generally interested in the features of RFID technology and in a number of business cases that, in a certain degree, RFID technology might help to effectively deal with, indicating that the diffusion is likely to increase rather than stop and suggesting that soon or later, the curve will progress along the diffusion path.
- Large companies are leading the adoption process.
- The main barrier to adoption was the cost of the system (54%), followed by the conviction of the business case and issues related to the practical application; next, companies are concerned by technological issues such as lack of standards and worries about integration and reliability of RFID technology.

To sum up, in supply chain management it is likely that UHF passive tags will be the most commonly used, and that pallets, cases and cartons tagging, rather than item tagging, will be the most popular application of RFID technology. Moreover, it is predicted that in the early phase big companies will adopt it (as shown by the survey results), paving the way to wider diffusion once return on investment is demonstrated, best practices developed and costs fall.

There were a number of issues revealed by the survey which require further investigation:

- If the few companies that are adopting RFID are gaining the benefits and ROI expected, and how long the testing period lasted before deployment.
- If logistics companies are more attracted by the radio waves identification (multiple and simultaneous reading without opening packaging) or the unique identification and at which level they are interested in tagging (item/case/pallet).

References

Benetton Group (2004) “Benetton: no microchip present in garments on sale. No decision yet taken on industrial use”, Benetton Group press area, 26 Jul. 2005, http://press.benettongroup.com/ben_en/releases/2003-04-04/

Collins, J. (2004) “RFID’s ROI Tops User Concerns”, RFID Journal, 26 Oct. 2004, http://www.rfidjournal.com/article/articleview/1207/1/1/ (accessed 23 Jan. 2006)

De Jonge, P. S. (2004) “Making Waves: RFID Adoption in Returnable Packaging”, LogicaCMG, http://www.logicacmg.com/pdf/RFID_study.pdf (accessed 23 Jan. 2006)

Bureau van Dijk Electronic Publishing (BvDEP), FAME, 2005, http://www.bvdep.com/fame.html

Napolitano, M. I. (2005) “Get ready for RFID”, Logistics Management, vol. 44, no. 8, pp. 83–86

Rogers, E. M. (1995) Diffusion of innovations, 4th ed., New York: Free Press

West, M. A. (1990) Innovation and Creativity at work, Wiley

Zaltman, G. et al. (1973) Innovations and organizations New York: Wiley

“85 per cent of UK companies shun RFID technology”, GS1 UK Media Centre, 2 Feb. 2004, http://www.gs1uk.org/media_story.asp?s=0&fid=75&sid=463 (accessed 23 Jan. 2006)

RFID in UK industry – An executive summary, 2004, http://www.rfidc.com/pdfs_downloads/Benchmark%20Research%20final.pdf (accessed 23 Jan. 2006)

RFID-Based Intelligent Logistics for Distributed Production Networks

Alexander Smirnov, Tatiana Levashova, Nikolay Shilov

St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences 39, 14 Line, 199178, St. Petersburg, Russia

Abstract Improvement of logistics system in production networks can produce a significant advantage for the companies involved. The paper investigates application of modern technologies such as RFID and configuration/optimization possibilities via an example of dynamic transportation problem. The general methodology based on ontology-driven knowledge description is described. It is explained how RFID technology can improve the logistics system and shown that in case of dynamic environment it is preferable to use fast algorithms producing feasible decisions instead of time-consuming optimisation.

1 Introduction

Today manufacturing companies have to restructure in order to respond to their rapidly changing environment. This causes a growing recognition of the need for organizational structures that could be distributed, mobile and flexible and therefore exhibit characteristics of innovation, resilience, and self-management. One of the most widespread forms of organizational structures is the form of distributed production network. Logistics systems play an important role in companies based on the concept of the production network. An intelligent decision making support based on modern technologies (Web services, radiofrequency identification (RFID), etc.) may significantly enhance the logistics system abilities (e.g., reduce costs and times of delivery). The paper proposes an application of knowledge logistics as an intelligent service for creation of efficient transportation plans (as one of the major logistics tasks in distributed production network management) under given constraints and preferences. This application is illustrated via a case study of delivering goods to customers.

An important factor for efficient implementation of decision support systems (DSS) providing intelligent support to logistics is usage of up-to-date technologies supporting automated logistics management systems. One of such technologies is the technology of RFID for automated product tracking. RFID is a way of storing

H.-D. Haasis et al., *Dynamics in Logistics*

DOI: 10.1007/978-3-540-76862-3,

and remote writing/reading data via radio signals to/from small and cheap devices (RFID tags or transponders). This technology enables collecting information about objects (e.g., materials and products), their locations and transportations, performing tracking and getting information concerning operations performed with the object without human intervention and with minimal amount of errors.

The RFID system consists of tags storing the data and readers or sensors (devices for writing and reading the data to and from tags). The antenna of the reader emits a radio signal that is received by the tag and powers its microchip. Using the received energy the tag performs an exchange of radio signals with the reader for self-identification and data transfer. Then, the reader sends the information received into the information system. RFID doesn't need a direct connection or direct visibility; the tags are read quickly and precisely, through dirt, steam, water, paint, plastics, etc.; tags can store relatively large volume of information. This technology enables a wide range of possibilities in the areas of logistics, identification, access control, and other areas of human activities.

2 Context-Driven Methodology

The reference model of the proposed intelligent decision making support is based on the earlier developed idea of knowledge logistics (Smirnov et al. 2004). It correlates with the conceptual integration developed within the Athena project (Ruggaber 2005) and meets requirements to information systems for distributed production networks.

To provide for semantic interoperability the usage of ontologies as one of the most advanced approaches to knowledge mark-up and description is proposed. Ontologies establish the common terminology between members of a community of interest (SemanticWeb 2005). This enables organization of semantic interoperability between various logistics tasks in distributed production networks. Besides, the ontology model is a means to overcome the problem of semantic heterogeneity.

The methodology presented proposes integration of environmental information and domain knowledge in the context of the current situation. It is done through linkage of representation of this knowledge to semantic models of information sources providing information about the environment. The methodology (Fig. 1) considers context as a problem model based on the knowledge extracted from the application domain and formalized within the application ontology by a set of constraints. The set of constraints, additionally to the constraints describing domain knowledge, includes information about the environment and various preferences of the user (user defined constraints) concerning the problem solving. Logistics assumes different user roles that are taken into account by the methodology as different levels of user responsibility. The problem is modelled by two types of contexts: abstract and operational. *Abstract context* is an ontology-based model integrating information and knowledge relevant to the problem. *Operational context* is an instantiation of the abstract context with data provided by the information sources.

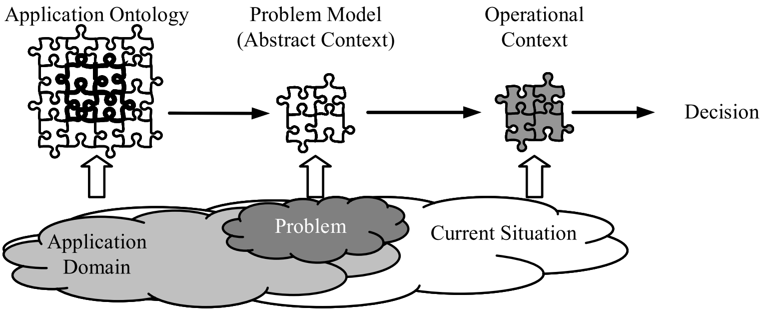


Fig. 1 Context-driven decision support

The methodology proposes a two-level framework (Fig. 2) of context-driven information integration for operational decision making. The first level addresses activities over a pre-starting procedure of the decision support system (DSS) as creation of semantic models for DSS' components; accumulating domain knowledge; linking domain knowledge with the information sources; creation of the application ontology describing the macro-situation; indexing the set of available e-documents against the application ontology. This level is supported, if required, by the subject

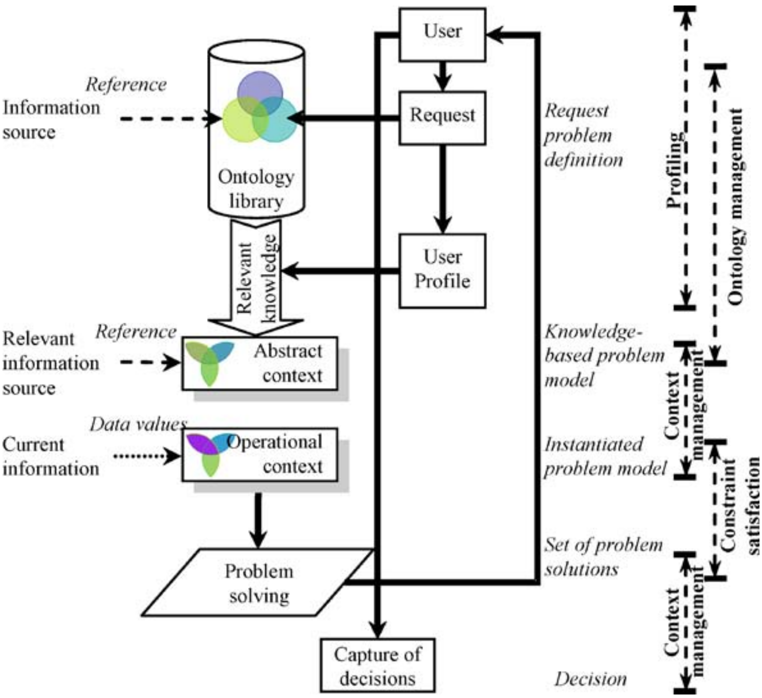


Fig. 2 Conceptual framework of ontology-driven information integration for operational decision making

experts, knowledge and ontology engineers. The second level focuses on decision making supported by DSS. This level addresses request recognition, abstract context creation, operational context producing, identification of relevant e-documents, generation of a set of problem solutions, and making a decision by the user. Both levels of the framework require maintenance of user profile.

A detailed description of the approach can be found in (Smirnov et al. 2005a, 2005b).

Since Web-services (W3C 2006) are currently a de facto standard for inter-system communications, widely supported, and quickly spreading, it is reasonable to use this standard a means of communication. The idea of open services arises from the concept of virtual organization. It can be said that "services are oriented to virtualization of resources" (The Globus Project Tutorial 2002).

3 Case Study

For the purpose of testing the possibilities of the RFID technology a table imitator that allows performing different experiments has been built. The scheme of coupling the table imitator with the presented technological framework is given in Fig. 3. RFID tags installed on trucks and containers are read by readers. A PC with installed software processes this information to define current locations of the trucks and containers (in real life can be combined with GPS or similar systems) and statuses of deliveries. This information is used by DSS to update delivery plans and schedules. For this purpose DSS also acquires information about the road network of the region provided by a Geographical Information System (GIS), traffic situation is acquired form an Intelligent Transportation System, current weather conditions are provided by a weather service, available warehouses and their current capacities are acquired from a special directory. Currently, the table imitator is equipped with two RFID readers ID ISC.M02 and ID ISC.PR101, and RFID tags RI-I11-112A-03. The characteristics of this equipment can be found in (FEIG Electronic 2007; Texas Instruments 2007). Car models are used for movement imitation.

A dynamic transportation problem is considered as a case study. There are several vehicles in different known locations that change in time. There is a central depot with a number of containers to be transported. Also there exist several warehouses with given capacities where the containers (products) are to be transported. It is necessary to take into account the current situation in the area (traffic jams, closed roads, etc.) to find a feasible solution with possibly minimal time of transportation. Though the problem seems to be simple it appears to be more complicated than it looks. For example, it can be more reasonable for one vehicle to make two or more rides instead of using two or more vehicles.

The problem has been formalized as follows.

Main variables of the problem:
b – vehicles $(1 \ldots B)$.
i – iterations $(1 \ldots PQ)$, where PQ is a number of containers to be transported, e.g.

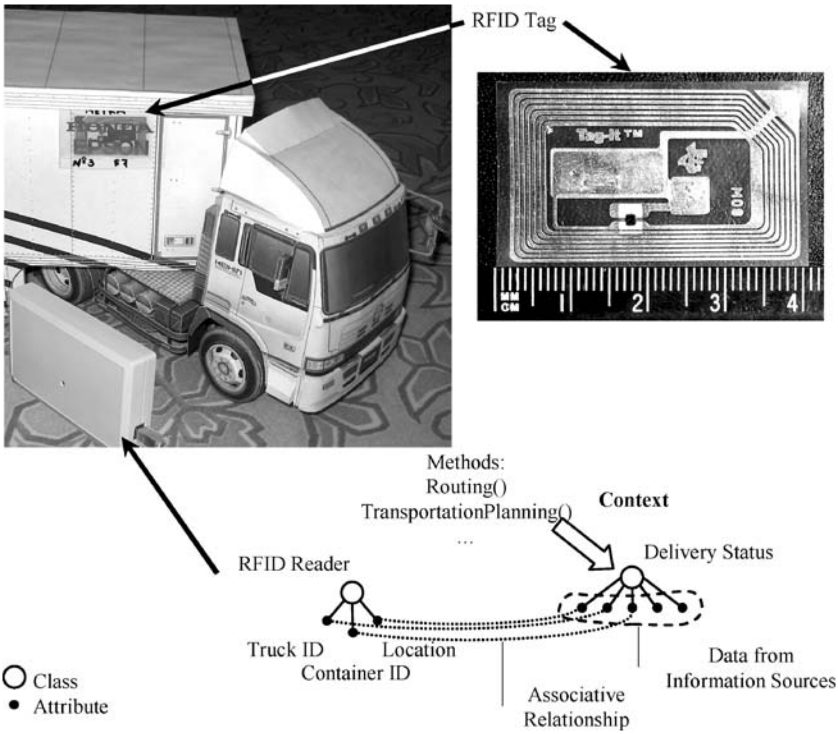


Fig. 3 Coupling RFID table imitation with technological framework

iteration 1: arriving to the depot and transporting a container; iteration 2: coming back to the depot and transporting the next container, ...
h – warehouses $(1 \ldots H)$.

Auxiliary variables:
Set of Boolean variables P_{bih} (whether vehicle b transports a container at iteration i into warehouse h).
Auxiliary Boolean variables $i > 1$: R_{bih} defining if vehicle b has to return to the depot after iteration i – 1.
Auxiliary variable $TE_{\max}$ (maximal time of transportation)

Constraints:
$\forall h$: $\Sigma_{bi} P_{bih} \leq HC_h$ (warehouse capacity)
$PQ \leq \Sigma_{bih} P_{bih}$ (number of transported containers should not be less than the given number of containers)
$\forall (b,i)$: $\Sigma_h P_{bih} \leq 1$ (not more than 1 container can be transported at once)
$\forall (b,i_1,i_2), i_1 < i_2$: $\Sigma_h P_{bi1h} - -\Sigma_h P_{bi2h} \leq 0$ (iteration for a vehicle cannot take place if no containers were transported by this vehicle at the previous iteration)
$\forall (b,i,h), i > 1$: $\Sigma_h P_{bih} + P_{b,i-1,h} - R_{bih} \leq 1$ (constraint between R and P)

$\forall(b,i), i = 1$: $\Sigma_h P_{bih}(AT_b + HT_{bh}) - TE_{\max} \leq 0$, where AT_b – time required for getting to the depot at the first time, HT_{bh} – time required for getting from the depot to warehouse h.
$\forall(b,i), i > 1$: $\Sigma_h P_{b1h}(AT_b + HT_{bh}) + \Sigma_{k=2}^{i} \Sigma_h HT_{bh}(P_{bih} + R_{bih}) - TE_{\max} \leq 0$

Goal: minimize TE_{max} by changing P_{bih} and R_{bih}.

It was found that in the considered case time pressure on problem solving made it reasonable to use algorithms for finding feasible solutions instead of trying to find the optimal solution. As a result it was proposed to create an algorithm to solve the problem in a direct way. Though, the algorithm does not always find the best solution, it provides for a feasible solution in a very short time and doesn't change when the size of the problem increases. This makes it possible to talk about scalability of the developed algorithm.

Inputs: AT_b, HT_{bh}, HC_h, PQ
Result: Transportation plan
Initialize times required for vehicles to get to the depot from their current locations:
For each vehicle b

$TimeToDepot_b = AT_b$

End For
For each container find the closest vehicle. This vehicle goes to the depot, picks up the container and transports it to the closest available warehouse:
For each container

Find the closest vehicle:
Find $b_{\min} \in [1, B]$ such that $TimeToDepot_{b_{\min}}$ is minimal
Find the closest available warehouse for vehicle $b_{\min}$:
Find $h_{\min} \in [1, H]$ such that (1) $HT_{b_{\min}h_{\min}}$ is minimal, (2) $HC_{h_{\min}} > 0$
Update transportation plan:
Add to the transportation plan information that vehicle $b_{\min}$ transports container to warehouse $h_{\min}$
Update times required for the chosen vehicle to get to the depot considering that the return time is the same:
$TimeToDepot_{b_{\min}} = TimeToDepot_{b_{\min}} + HT_{b_{\min}h_{\min}} + HT_{b_{\min}h_{\min}}$
Update capacity of warehouse $h_{\min}$:
$HC_{h_{\min}} = HC_{h_{\min}} - 1$

End For

The algorithm is very straightforward and simple, and in most cases it provides efficient solutions at short time. The algorithm could be improved at the cost of performance but in the project considered it was not needed. The results and working time (about 2 seconds on a Pentium IV computer for the task with 7 vehicles and 9 containers) were sufficient.

The prototype interfaces are Web-based what enables using Web browsers (from PCs, PDA or mobile phones) for working with the system. Fig. 4 illustrates examples

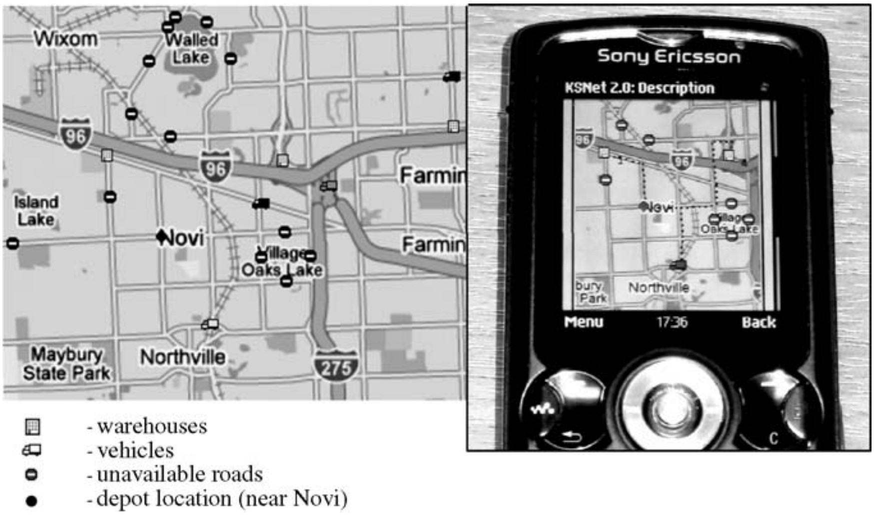


Fig. 4 Current situation for the dynamic transportation problem (a fragment) and vehicle driver assignment on the screen of a mobile phone (bottom right)

of screenshots presenting (a) the current situation (a fragment) for dispatcher and decision maker and (b) vehicle driver assignment on the screen of a mobile phone (*bottom right*).

4 Conclusion

The paper presents a technological framework, case study and a prototype of an approach to intelligent decision support of logistics in distributed production networks. The approach assumes acquisition of information from distributed heterogeneous sources via Web services. RFID is considered as a possible technology for providing actualized information into the decision support system. A dynamic transportation problem is considered as a case study. Planned future work is aimed at introduction of self-organising networks. The advantage of self-organisation is its robustness to possible infrastructure failures (e.g., when the central planning unit is not longer available).

Acknowledgements The paper is due to the research carried out as a part of projects funded by grants # 05-01-00151 and # 06-07-89242 of the Russian Foundation for Basic Research, projects # 16.2.35 of the research program "Mathematical Modelling and Intelligent Systems", # 1.9 of the research program "Fundamental Basics of Information Technologies and Computer Systems" of the Russian Academy of Sciences (RAS), project of the scientific program of St. Petersburg Scientific Center of RAS, as well as project of the Science and High School Committee of St. Petersburg Government.

References

Smirnov, A., Pashkin, M., Chilov, N., Levashova, T. (2004), ‘Knowledge Logistics in Information Grid Environment’, The special issue “Semantic Grid and Knowledge Grid: The Next-Generation Web” (Zhuge H., ed.) of *International Journal on Future Generation Computer Systems*, Vol. 20, No. 1 pp. 61–79.

Ruggaber, R. (2005), ‘Interoperability of Enterprise Systems and Applications’, *Strengthening Competitiveness through Production Networks. A prespective from European ICT research projects in the field of “Enterprise Networking”*, pp. 58–70.

The Globus Project. An Open Grid Services Architecture (2002) ‘Globus Tutorial, Argonne National Laboratory’, http://www.globus.org/ogsa/deliverables/OGSA-January-2002-v3.pdf.

W3C consortium (2006) http://www.w3.org.

Semantic Web. Web site (2005), http://www.SemanticWeb.org.

Smirnov, A., Pashkin, M., Levashova, T., Chilov, N. Fusion-Based Knowledge Logistics for Intelligent Decision Support in Network-Centric Environment. International Journal of General Systems, December 2005, Vol. 34, No. 6, 673–690.

Smirnov, A., Pashkin, M., Chilov, N., Levashova, T. Constraint-driven methodology for context-based decision support. Journal of Decision Systems, 2005, Vol. 14, No. 3, 279–301.

FEIG Electronic. Web site (2007). http://www.feig.de/index.php?option=com_content&task=view&id=71&Itemid=134.

Texas Instruments. Web site (2007). http://www.ti.com/rfid/shtml/prod-trans.shtml.

Methodology for Development and Objective Comparison of Architectures for Networked RFID

Béla Pátkai, Damith Ranasinghe, Mark Harrison, Duncan McFarlane

Distributed Information and Automation Lab, Institute for Manufacturing, University of Cambridge, Mill Lane, Cambridge, CB2 1RX, United Kingdom

Abstract The purpose of this paper is to address the problems presented by a heterogeneous environment of networked RFID implementations. These systems have a growing number of applications and these will have varying requirements that need to be modelled and communicated. The use of an ontology-centred methodology is suggested that can represent the entities in the system and also their relationships in a formalized manner. Such a formal methodology can support the planning, design, development, efficient operation and interconnection of various networked architectures.

1 Introduction

Recent developments in Radio Frequency Identification (RFID), sensor technology, wired and wireless networking of computers and other electronic devices – among others – have established the technological opportunity to build the long anticipated "Internet of things". Even though it is very useful and convenient to use the analogy of the Internet in this area, networking RFID tags, sensors, and various data sources from a network of enterprises is closer to the idea of intranets, extranets and the semantic web. The complexity of design methods necessary to develop industrial strength, secure, trusted and efficient architectures for various industry sectors is a significant challenge. The recent standardization of the EPC Information Services [1] and the concentrated development effort in industry specific architectures in a number of research projects suggests that an abundance of networked architectures will be designed, deployed and will need to interoperate and be compared to each other, especially because the high value and long life cycle of products requires unique identification and Integrated Condition Monitoring far beyond previously known examples. The aim of this paper is to lay down the foundations of a systematic design method that foresees the necessity for the multiplicity and heterogeneity of solutions, interface definitions, design of customized solutions and the definition of performance metrics for objective comparisons.

DOI: 10.1007/978-3-540-76862-3, © Springer 2008

2 Problem Definition

Networked information systems have been developed in the last two decades with increasing speed due to the emergence of numerous standards, including the well-known W3C standards and recommendations – XML, REST, SOAP, WSDL, Corba, WS-*, RPC – the Enterprise Service Bus (ESB), and many other vendor-dependent, but significant, solutions like Microsoft **.Net**. The recent development of EPCglobal standards for networked RFID [1] added useful, problem-specific functionality to these. In this paper we take an agnostic view on whether a single standard is going to address all the challenges of the future in this field, and assume that the proliferation of several standards and vendor-dependent solutions is going to happen.

This brings up a set of problems such as:

- Networked RFID is going to be implemented in a heterogeneous environment
  - Not following a single standard
  - Different vendors will provide solutions not conforming to all standards
- Heterogeneous systems will have to interface
- Ad-hoc solutions will often be built
- Services will be technically easy to implement – difficult to deploy & maintain, hence
  - Solution performance will not foreseeable
    - Computational burden can become problematic
      - E.g. sensor data processing
    - Network load is going to fluctuate
    - Storage requirements will not be known before deployment of service
  - Indirect security issues: *accessing data allows data mining by third parties*
- Systematic interfacing and service development work requires a methodology
- Efficient communication formalism is necessary for system/services development
- Automation of interfacing (or at least part of it) will be very desirable (M2M)

In the realm of logistics these problems will be amplified because:

- Logistic networks are:
  - Naturally cross-organizational
  - Very dependent on each other
  - Use a variety of IT infrastructure
  - At a different point of IT development
  - Have different IT outsourcing policy
  - Their business model may benefit a lot from improved information flow

- Logistic networks require addressing issues in the previous "Problem Definition"
- The concept of "logistics" is being naturally extended, as the lifecycle of products can be very long, their value may be very high and unique identification and condition monitoring throughout their lifecycle requires attention.

The purpose of this paper is to suggest a methodology for architecture designers and operators to develop, deploy and maintain networked RFID-based services in the future.

3 The Design Methodology

Man-made systems expected to grow and interact with other systems need to be designed in a systematic manner incorporating robustness in them to enable them to cope with changes, errors and performance problems. Successful examples are the Internet and the Unix operating system that was built on about a dozen elegant design principles [1] and continues to be the most robust and flexible computing and server platform. The simple and well defined interfaces of network architectures currently in development aim at the same robustness, efficiency and simplicity, but either lack functionality to handle complex filtering, querying or access control, or only address a specific industry sector and don't scale well. The stack of EPCglobal standards relevant in this context address many of the problems foreseen in the future, but it is not reasonable to expect that it will be the only solution in the market. The methodology suggested in Fig. 1 shows an ontology based design method that is meant for a heterogeneous environment.

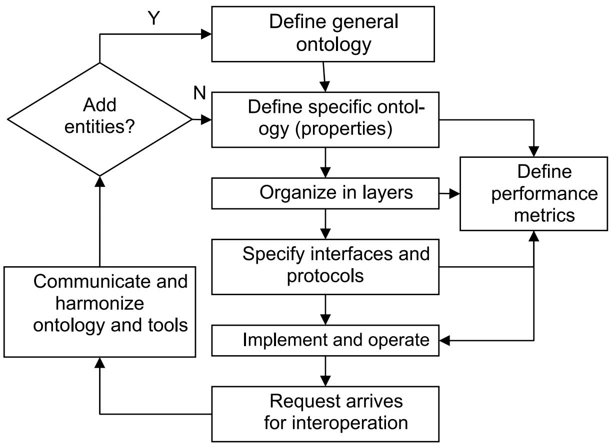


Fig. 1 Ontology-centered system development

The steps shown in the figure are the following:

1. *Define a general ontology* that is essentially just a list of entities, without relationships and generalization. This step is necessary because the following steps are already subjective, but the entity list is not.
2. *Define a specific ontology* including classes (individuals, concepts and roles), attributes and interfaces of the building blocks of networked architectures (RFID tags, sensors, data sinks, computers, databases, network interfaces and protocols, applications).
3. *Organize in layers,* i.e. define a stack of layered protocols (the example in the next section uses the ISO Open Systems Interconnection standard to show this).
4. *Specify interfaces and protocols* – for interoperability of the entities and also different architectures if necessary in a heterogeneous environment.
5. *Implement and operate* – the actual building blocks are implemented and start operation according to current requirements.
6. *Define performance metrics* – on the basis of class attributes. This provides the basis for any M2M (Machine to Machine) communication and automation of related decision making. E.g. when a request is made by another system then the one providing services is able to use these metrics to determine whether it has suitable resources to satisfy the request.
7. *Request arrives for interoperation*: either a person to person or M2M request arrives and is evaluated either manually or automatically.
8. *Communicate and harmonize ontology and tools*: in case both parties use an ontology they can use it for quick and efficient communications and they can also harmonize it.
9. *Add entities*: in case there are new ontological entities in any of the involved systems these entities can be added, their relationships and properties can be defined.

This general methodology using ontologies addresses the issues listed in the problem definition and provides tools for interfacing, integrating and operating networked IT systems.

4 Demonstrative Example

4.1 General Ontology Definition

The general ontology only provides the most basic relationships between the entities and entity classes, because at this level the preservation of generality is necessary. The more specific the ontology becomes in the following steps, the more disagreement and differentiation between them can be expected.

The general hardware ontology for a simple networked RFID system includes the following (list nesting already shows basic hierarchy):

- Computers
  - Functional nodes
    - RFID tags
      - Passive tags
      - Semi-passive tags
      - Active tags
    - Sensor node
      - Wired sensors
      - Wireless sensors
    - Sensor tags
    - Interrogators
      - RFID readers
      - Sensor sink

The same kind of list can be generated for software entities and extended as required with all the necessary details (e.g. aggregations, generalizations, properties, methods and relationships).

4.2 Specific Ontology Definition

Figure 2 shows the specification of the ontology by an RDF (Resource Description Framework) diagram. UML is also a very useful and usually well understood alternative for communicating a system structure, but it has some limitations – e.g.

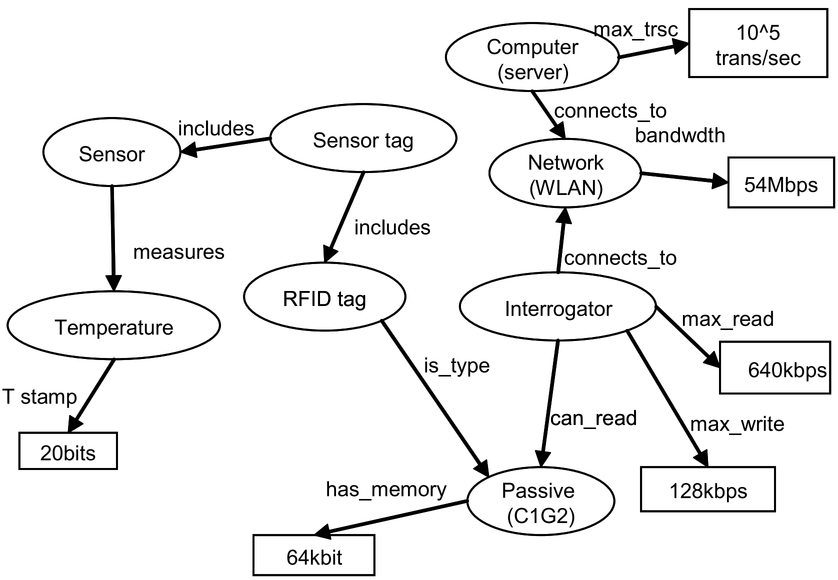


Fig. 2 RDF of Example Networked RFID System

the lack of syntax for an object to specify the ordering of instances [3] – that require an actual ontology language. RDF and RDFS (RDF Schema) uses XML to define entities and relationships. OWL (Web Ontology Language) is its extension and contains further features not necessary in this work.

4.3 Definition of Layers

Figure 3 shows how the entities defined earlier can be mapped on a layered hierarchy. The example of the OSI (Open Systems Interconnection) standard is used in this paper even if it's not the most successful one, mainly because it is well-known in many industries. Such a hierarchic layering assists us in determining the interfaces that need to be defined between interacting systems. This figure provides a different cross section of the system than the RDF model, the first is about the place of entities while the other one is about their functionality and relationships.

4.4 Usage of the Ontology Model

Let's consider the following example of a logistic problem:

- Two firms have a logistic link with each other, and they:
  - Use bar codes and ASN (Advance Shipping Notification) on products
  - Want to introduce RFID on products for track and trace
  - Want to add condition monitoring features to their system
- The systematic process provides the following:
  - They both have an ontology model
  - They harmonize it to have a mutual understanding
  - Derive conclusions from the model
    - Performance of the integrated system
    - Bottlenecks
    - Resource usage and cost

From the RDF model it is straightforward to derive that:

- In case a shipment arrives with 1000 sensor tags every hour, can we share full temperature history with shipping partners?
- Do we need more readers, servers, bandwidth?
- Should we delegate sensor data filtering towards

From a more extensive model we can see whether:

- In case we have full service history of aircraft parts and exchange 500 parts per day with partners, can we sell them information and allow them to query our SQL database?
- Should we sell raw data or invest in processing/mining?

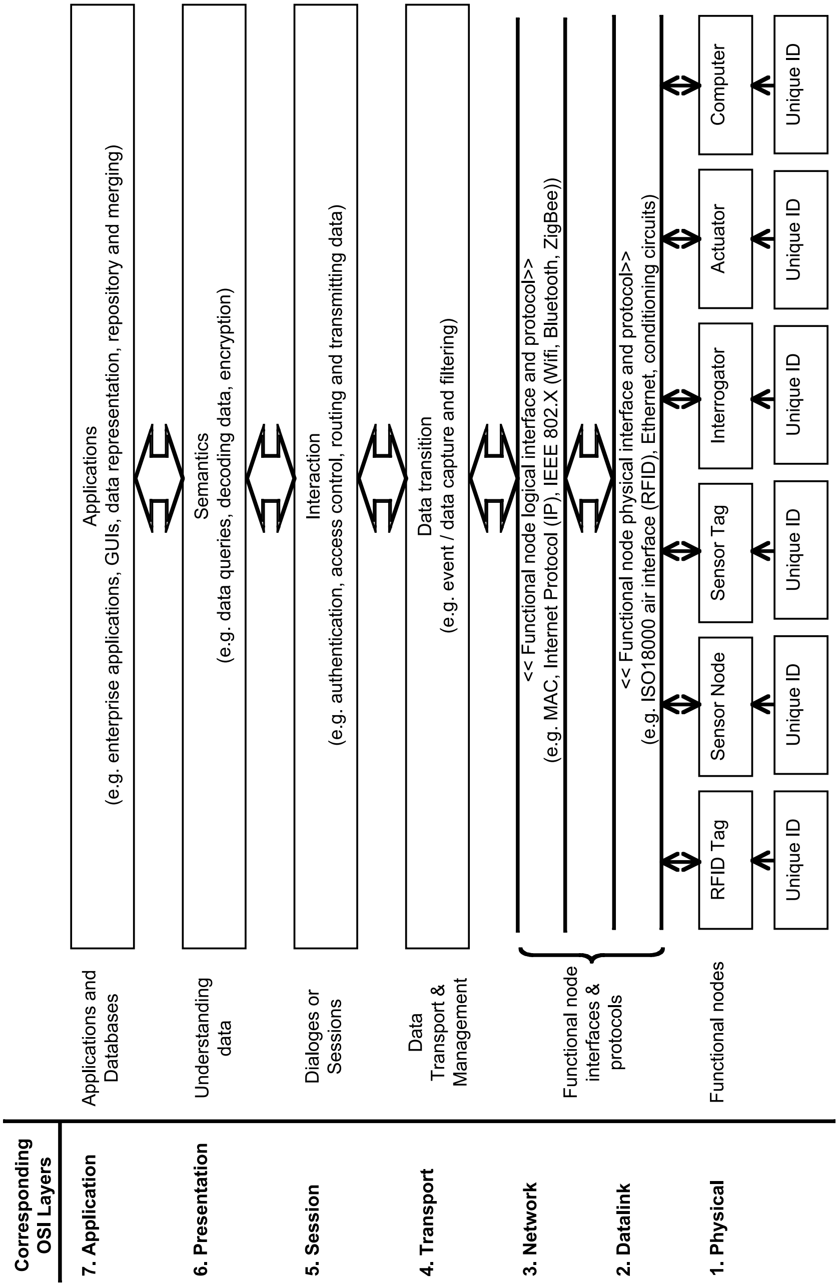


Fig. 3 Correspondence of OSI Layers and Networked RFID

5 Conclusions

In this paper a systematic, ontology based methodology was suggested to overcome the foreseeable challenges in heterogeneous networked RFID system interconnection. The proposed ideas will have to be tested in a laboratory environment and the automation of service capacity planning has to be investigated.

References

1. EPC Information Services (Version 1.0) Specification, Ratified Standard April 12, 2007, www.epcglobalinc.org/standards
2. Eric Steven Raymond, The Art of Unix Programming, 2003, www.faqs.org/docs/artu/
3. Stephen Cranefield and Martin Purvis, UML as an Ontology Modelling Language, IJCAI-99 Workshop on Intelligent Information Integration, 1999.

Supply Chain Control Policies

Determining Optimal Control Policies for Supply Networks Under Uncertainty

Marco Laumanns

ETH Zurich, Institute for Operations Research, 8092 Zurich, Switzerland
email: laumanns@ifor.math.ethz.ch

Abstract This paper shows how to compute optimal control policies for a certain class of supply networks via a combination of stochastic dynamic programming and parametric programming. We consider supply networks where the dynamics of the material and information flows within the entire network can be expressed by a system of first-order difference equations and where some inputs to the system act as external disturbances. Assuming piecewise linear costs on state and control inputs, optimal control policies are computed for a risk-neutral objective function using the expected cost and for a risk-averse objective function using the worst-case cost. The obtained closed-loop control policies are piecewise-affine and continuous functions of the state variables, representing as a generalization of the common order-up-to policies. The optimal value functions are piecewise affine and convex, which is the essential structural property to allow for the solution via a sequence of parametric linear programs. Some numerical results are given on an example network with two suppliers with different costs and lead times.

1 Introduction

In managing supply networks an important task is to organize the flow of goods and information between the different components. This includes both the design of the network structure (building and configuring the links) as well as controlling the material flow through the network over time. A major challenge in this respect is to cope with uncertainty, for instance in the customer demand, availability of materials, variability of production and transportation times and so forth [13]. This immediately raises questions of how to effectively reduce the risk in supply networks associated with these uncertainties, which is often addressed by holding reserves such as excess inventory and capacity or redundant suppliers [3].

There is a large body of research concerned with devising optimal control policies for processes on different scales within the value chain, including control of manufacturing lines as well as inventory management. As a common approach

H.-D. Haasis et al., *Dynamics in Logistics*

DOI: 10.1007/978-3-540-76862-3,

for situations where decision are made in stages, such stochastic control problems are often addressed using Bellman's principle of optimality and dynamic programming [2]. Due to the complexity of the processes involved, an analytical derivation of the optimal control policies is usually only possible for very simple models involving only a few, often only one, state variables. A related, slightly different approach is to assume a family of control policies in parametric form, prove its optimality in a certain setting using the principle of optimality, and afterwards compute the optimal parameters depending on the data of the current problem instance [5]. Yet, this approach often requires restrictive assumptions about the network structure as well [9].

The dynamic programming algorithm can be used to obtain optimal control inputs numerically by tabulating the optimal costs for all possible states in all tail subproblems recursively. Obviously, this is only practical when the set of reachable states is sufficiently small or can be coarsely discretized. In the latter case the obtained solutions represent only an approximation, as is the case when simulation-based or heuristic methods are used.

In this paper, we show how to use a combination of dynamic programming and parametric programming, recently developed for model predictive control [1], to determine explicit control policies for general supply networks algorithmically. Our assumptions are that the system works in discrete time, that there are fixed time delays for both material and information flows between the nodes in the network and that the cost per time period is a piecewise linear convex function of the state and control inputs. We further assume that the disturbances are additive and their distribution has finite support. We present algorithms to compute the optimal control policies with respect to two different objective functions, the expected cost and the worst-case cost, representing one risk-neutral and one risk-averse approach. The approach is demonstrated on an example network involving two suppliers with different lead times and different costs.

2 Optimal Control by Stochastic Dynamic Programming

We consider supply networks as discrete-time controlled dynamical systems with uncertainty

$$\mathbf{x}_{k+1} = f(\mathbf{x}_k, \mathbf{u}_k, \mathbf{d}_k) \tag{1}$$

where $\mathbf{x}_k \in X \subseteq \mathbb{R}^n$ denotes the state of the system, $\mathbf{u}_k \in U(\mathbf{x}_k) \subseteq \mathbb{R}^{n_u}$ the control input and $\mathbf{d}_k \in \mathscr{D} \subseteq \mathbb{R}^{n_d}$ the (uncertain) disturbances, at time k. The map $f : \mathbb{R}^{n+n_u+n_d} \to \mathbb{R}^n$ depends on the structure of the network and on the dynamics of the individual nodes. As the basic network structure for the material flow we assume a directed connected graph $G = (V, E)$ whose vertex set $V = \{v_1, \ldots, v_{n_V}\}$ contains the facilities (nodes) of the supply network. The set of directed edges E is divided into two sets $S = \{s_1, \ldots, s_{n_S}\}$ for the material flows and $R = \{r_1, \ldots, r_{n_R}\}$ for the information flows, where $E = S \cup R \subseteq V \times V$. With each edge $e \in E$ we associate an integer time delay (transportation time) $\tau_e \in \mathbb{N}$. A small example network is displayed in Fig. 1.

A procedure to derive the global dynamics (1) from the network structure and the individual node dynamics was given in [12]. Assuming that the transition function of the individual nodes are linear, the resulting global dynamics can be written as

$$\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k + \mathbf{B}\mathbf{u}_k + \mathbf{d}_k$$

with matrices $\mathbf{A} \in \mathbb{R}^{n \times n}$ and $\mathbf{B} \in \mathbb{R}^{n \times n_u}$. This description will be used in the following.

Our goal is to derive explicit state-feedback control policies, i.e., control laws as functions of the current state of the system. This results in a closed-loop control in the sense that by making the control a function of the current state, the decision at time step k implicitly takes into account all disturbance realizations prior to time k.

Let $\mathbf{u}_k(\mathbf{x}_k)$ denote the control input when system is in state $\mathbf{x}$ at time k. A *control policy* is then defined as a sequence $\pi = (\mathbf{u}_0, \ldots, \mathbf{u}_{N-1})$ of such state-to-control mappings, where we consider $\mathbf{x}_0$ as the initial state and a fixed finite time horizon $N \in \mathbb{N}$. An *admissible* control policy is such that $\mathbf{u}_k(\mathbf{x}_k) \in U(\mathbf{x}_k)$ for all $0 \leq k \leq N-1$. The *expected* cost for an admissible control policy π is then given by

$$\bar{J}(\mathbf{x}_0, \pi) := \mathbb{E}\left\{\mathbf{p}^T\mathbf{x}_N + \sum_{i=0}^{N-1} \mathbf{p}^T\mathbf{x}_i + \mathbf{q}^T\mathbf{u}_i(\mathbf{x}_i)\right\} \tag{2}$$

where $\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k + \mathbf{B}\mathbf{u}_k(\mathbf{x}_k) + \mathbf{d}_k$ and $(\mathbf{p}^T\mathbf{x})$ and $(\mathbf{q}^T\mathbf{u})$ are linear cost functions on state and control inputs, respectively. The expectation is taken with respect to the random sequence $(d_k)_{k=0}^{N-1}$. Similarly, the *worst-case* cost can be written as

$$\hat{J}(\mathbf{x}_0, \pi) := \max_{(\mathbf{d_k})_{k=0}^{N-1}} \mathbf{p}^T\mathbf{x}_N + \sum_{i=0}^{N-1} \mathbf{p}^T\mathbf{x}_i + \mathbf{q}^T\mathbf{u}_i(\mathbf{x}_i). \tag{3}$$

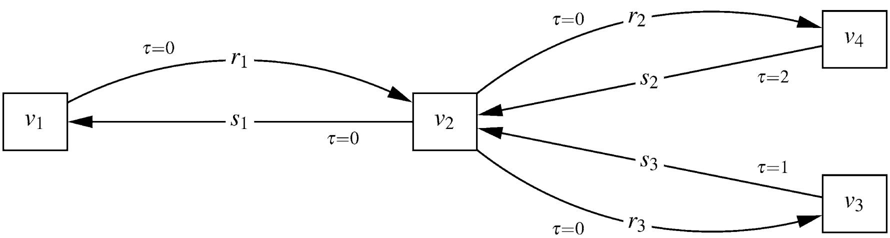


Fig. 1 Example of the basic supply network model, where the vertices v_i represent the facilities and the edges s_i and r_i represent material flows (shipments) and information flows (orders), respectively. The edges are annotated by their time delays τ

To find the optimal value functions $J_k^* : \mathscr{X}_k \to \mathbb{R}$ and corresponding optimal controls u_k^* for both objective functions, we initialize the dynamic programming recursion by setting the terminal costs in both cases as

$$\bar{J}_N^*(\mathbf{x}_N) \equiv \hat{J}_N^*(\mathbf{x}_N) \equiv \mathbf{p}^T \mathbf{x}_N \tag{4}$$

and letting $\mathscr{X}_N = X$ denote the set of feasible states at the time horizon N. The optimal value function can now be computed recursively for the expected cost as

$$\bar{J}_k^*(\mathbf{x}_k) = \min_{\mathbf{u}_k \in \mathscr{U}_k} \left\{ \mathbf{p}^T \mathbf{x} + \mathbf{q}^T \mathbf{u} + \mathbb{E}\{J_{k+1}^*(\mathbf{A}\mathbf{x}_k + \mathbf{B}\mathbf{u}_k + \mathbf{d})\} \right\} \tag{5}$$

and for the worst-case cost as

$$\hat{J}_k^*(\mathbf{x}_k) = \min_{\mathbf{u}_k \in \mathscr{U}_k} \left\{ \mathbf{p}^T \mathbf{x} + \mathbf{q}^T \mathbf{u} + \max_{\mathbf{d}_k \in \mathscr{D}} \{J_{k+1}^*(\mathbf{A}\mathbf{x}_k + \mathbf{B}\mathbf{u}_k + \mathbf{d})\} \right\} \tag{6}$$

where the set of admissible controls in both cases is

$$\mathscr{U}_k = \{u \in U(\mathbf{x}_k) : \mathbf{A}\mathbf{x}_k + \mathbf{B}\mathbf{u}_k(\mathbf{x}_k) + \mathbf{d}_k \in \mathscr{X}_{k+1} \ \forall \, \mathbf{d}_k \in \mathscr{D}\} \tag{7}$$

and the feasible state sets are given by

$$\mathscr{X}_k = \{x_k \in \mathscr{X} : \mathscr{U}_k \neq \emptyset\}. \tag{8}$$

The sets $U(\mathbf{x}_k)$ represent additional joint constraints on controls and states, which we assume to be polyhedral set.

Under the additional assumption that the set of possible disturbances $\mathscr{D}$ is finite or, for the problem of minimizing the worst-case cost, that it is the convex hull of a finite set of points, it can be shown that

- both problems are multiparametric linear programs for all k, where the current state $\mathbf{x}_k$ acts as a parameter of dimension n affecting the right-hand-side of the inequality system of the LP,
- the optimal value functions $J_k^*(\mathbf{x}_k)$ are piecewise affine and convex in $\mathbf{x}_k$,
- there exists a piecewise affine $\mathbf{u}^*(\mathbf{x}_k)$ that is continuous in $\mathbf{x}_k$.

Note that degeneracy, i.e, the existence of multiple solutions for the optimal control, can be avoided using lexicographic perturbation [6], in which case the only resulting optimal control policy will be continuous in $\mathbf{x}_k$.

The optimal value functions and the control policies can therefore be stated explicitly as piecewise affine functions of the state $\mathbf{x}_k$ as

$$J_k^*(\mathbf{x}_k) = \mathbf{V}_k^{(i)} \mathbf{x}_k + \mathbf{W}_k^{(i)} \tag{9}$$

$$\mathbf{u}_k^*(\mathbf{x}_k) = \mathbf{G}_k^{(i)} \mathbf{x}_k + \mathbf{H}_k^{(i)} \tag{10}$$

where $\mathscr{R}_k := \{\mathscr{R}_k^{(i)}\}_{i=1}^{n_R}$ is a polyhedral partition of the feasible state space $\mathscr{X}_k$ (the closure of the $\mathscr{R}_k^{(i)}$ are polyhedra). As a consequence, both optimal value function and optimal control policy for a fixed horizon N can be computed by recursively solving a sequence of N multiparametric linear programs. We can further define $\mathbf{u}^* := \lim_{N\to\infty} \mathbf{u}_0^*$ as the infinite-time optimal control policy, if this limit exists.

3 Numerical Example

In this section we consider the network of Fig. 1 for a numerical example. The goal is to compute the optimal control policy for a fixed, sufficiently large time horizon N for the different objective functions (average and worst-case cost) and different disturbance sets D. This is achieved by iterating the dynamic programming recursion for a sufficiently large number of iterations.

In the example network, the node v_2 represents a retailer facing uncertain demand $d \in [0,8]$ by its customer v_1. The retailer can choose to order $\mathbf{u}_k^{(1)} \in [0,8]$ from supplier v_4 for a cost of $q^{(1)} = 1$ per unit or an amount $\mathbf{u}_k^{(2)} \in [0,6]$ from supplier v_4 for a cost of $q^{(2)} > 1$ per unit. Inventory holding costs are 1 per unit per time step, hence $\mathbf{p} = [1,0,0,0]^T$. The dynamics of the system is given by

$$\mathbf{x}_{k+1} = \underbrace{\begin{bmatrix} 1&1&0&1\\ 0&0&1&0\\ 0&0&0&0\\ 0&0&0&0 \end{bmatrix}}_{\mathbf{A}} \mathbf{x}_k + \underbrace{\begin{bmatrix} 0&0\\ 0&0\\ 1&0\\ 0&1 \end{bmatrix}}_{\mathbf{B}} \mathbf{u}_k - \underbrace{\begin{bmatrix} -d\\ 0\\ 0\\ 0 \end{bmatrix}}_{\mathbf{d}_k}. \tag{11}$$

Enforcing a non-negative feasible state space $X = \mathbb{R}_{\geq 0}$, the sequence of multiparametric linear programs quickly converges to a solution whose resulting feasible region $\mathscr{X}_f$ is given by

$$\begin{aligned} x^{(1)} + x^{(2)} + x^{(4)} &\geq 8 \text{ and} \\ x^{(1)} + x^{(2)} + x^{(3)} + x^{(4)} &\geq 10. \end{aligned}$$

The optimal control law can be stated succinctly as

$$\begin{aligned} u^{(1)*} &= \min\{\max\{20 - x^{(1)} - x^{(2)} - x^{(3)} - x^{(4)}, 0\}, 4\} \\ u^{(2)*} &= \max\{16 - x^{(1)} - x^{(2)} - x^{(3)} - x^{(4)}, 0\}. \end{aligned}$$

This is a so-called dual base-stock policy, which has been proven to be optimal for certain discrete-time inventory models with two sources [15]. The hedging points are 16 for the quick supplier v_3 with lead time 1 and L for the cheaper supplier v_4 with lead time 2.

Of course, the optimal strategy depends on the relation of the different order costs $q^{(1)}$ and $q^{(2)}$ as well as the chosen objective function, and is constrained by the

Table 1 Hedging point L as a function of the cost ratio $q^{(2)}/q^{(1)}$. For the expected cost, the disturbances are assumed to be uniformly distributed on the set of numbers given in the second column

Objective	Disturbance						
worst case cost	$[0,8]$	16	16	17.6	20	22	23
expected cost	$\{0,8\}$	16	16	24	24	24	24
	$\{0,4,8\}$	20	20	20	24	24	24
	$\{0,1,2,3,4,5,6,7,8\}$	19	20	21	22	23	24
Cost ratio $q^{(2)}/q^{(1)}$		1.5	2	2.5	4	8	16

maximum allowed order quantities that have been agreed upon by the retailer and each of its suppliers, as specified by the constraints.

Table 1 displays the resulting hedging points L for a range of different cost ratios $q^{(2)}/q^{(1)}$ for both the worst case and the expected cost as control objectives. For the worst-case cost, the disturbance can take any value in $[0,8]$. For the expected cost, the disturbance is assumed to be distributed uniformly on the finite support listed in the second column of the table.

4 Conclusions

We have presented a method that allows us to compute explicit closed-loop state-feedback control policies for a general class of supply networks with piecewise linear dynamics in discrete time. The method makes use of the fact that the optimal value function can be shown to be piecewise affine and convex so that the dynamic programming recursion can be executed by solving a sequence of multiparametric linear programs. The method works for both the expected cost and the worst-case cost as objective functions under the assumption that the set of possible disturbances is a finite set (for the expected cost) or the convex hull of a finite set of points (for the worst-case cost), respectively. The concept of uncertainty is very general and can be used to model various aspects such as variations in customer demand and transportation times, perishing or other losses of goods and so forth.

The numerical example for a network with two suppliers demonstrates that the method finds the optimal dual base-stock policy without any a priori assumptions about the actual structure of the control policy itself. The actual values of the hedging points depend on the coefficients of the cost functions and constraints. These values enter the algorithms as input values and are therefore not explicit parameters of the resulting control policy, which will only be the state variables.

The actual structure of any optimal control policy of the class of problems considered here as a piecewise affine and continuous function of the state variables can be seen as a generalization of the common base-stock or hedging-point policies that are known to be optimal in many simple settings even beyond the piecewise linear objective function case considered here. The generalization achieved here is rather with respect to a general network scenario, where any control variable could depend on any state variable anywhere within the network.

The latter aspects also highlights the fact that this approach assumes global state information. It is therefore applicable in cases where there is, e.g., one central owner of the networks who can exercise global control. Nevertheless, as the method does not only yield optimal control inputs for a given state, but the control policy as a function of any arbitrary system state, the policy can simply be distributed to the individual nodes in the network. The structure of the policy then indicates which state information needs to be communicated to any given node such that it can determine its optimal control values.

The stability of supply networks is a further topic, where existing studies investigate how the dynamics of the system depends on the chosen control strategy [4, 7, 11, 14] or on the structure of the network [8]. It would be an interesting topic of future research to study the stability of control policies obtained by the approach presented here with respect to disturbances that are outside the range of values considered in the algorithm.

Acknowledgements I am grateful to Erjen Lefeber for earlier joint work on this topic and to Colin Jones, Rico Zenklusen, and Sebastiano Columbano for numerous discussions on parametric linear programming. For the numerical results presented here, the parametric LP solver of the Multi-Parametric Toolbox [10] of the Automatic Control Lab, ETH Zurich, was used.

References

1. Bemporad A, Morari M, Dua V, Pistikopoulos EN (2002) The explicit linear quadratic regulator for constrained systems. Automatica 38(1), 3–20
2. Bertsekas DP (2000) Dynamic programming and optimal control. Volume I. Athena, Belmont
3. Chopra S, Sodhi MMS (2004) Managing risk to avoid supply-chain breakdown. MIT Sloan Manag Rev 46(1):53–61 8
4. Daganzo CW (2003) A Theory of Supply Chains. Springer, Berlin
5. De Kok AG, Graves SC, ed (2003) Supply chain management: design, coordination and operation. Handbooks in Operations Research and Management Science. Elsevier, Amsterdam
6. Fukuda K, Lüthi HJ, Namkiki M (1997) The existence of a short sequence of admissible pivots to an optimal basis in LP and LCP. Int Trans Oper Res 4(4), 273–384
7. Helbing D, Lämmer D, Seidel T, Seba P, Platkowski T (2004) Physics, stability and dynamcis of supply networks. Phys. Rev. E 70(066116)
8. Helbing D, Witt U, Lämmer S, Brenner T (2004) Network-induced oscillatory behavior in material flow networks and irregular business cycles. Phys. Rev. E 70(056118)
9. Graves SC, Willems SP (2000) Optimizing strategic safety stock placement in supply chains. Manuf Serv Oper Manag 2(1):68–83
10. Kvasnica M, Grieder P, Baotić M (2004) Multi-Parametric Toolbox (MPT). Available via http://control.ee.ethz.ch/ mpt/
11. Mosekilde E, Larsen ER (1988) Deterministic chaos in a beer production-distribution model. Syst. Dyn. Rev. 4:131–147
12. Laumanns M, Lefeber E (2006) Robust optimal control of material flows in demand-driven supply networks. Physica A 363(1):24–31
13. Liberopoulos G et al (2004) Stochastic models of production-inventory systems. Ann Oper Res 125:17–19
14. Ouyang Y, Daganzo CF (2006) Counteracting the bullwhip effect with decentralized negotiations and advance demand information. Physica A 363(1):14–23
15. Tan B, Gershwin SB (2004) Production and subcontracting strategies for manufacturers with limited capacity and volatile demand. Ann Oper Res 125:205–232

Adaptive Production and Inventory Control in Supply Chains against Changing Demand Uncertainty

Markus Zschintzsch[1], Amir Sheikh Jabbari[2], Walter Lang[2], Bernd Scholz-Reiter[1]

[1] Department of Planning and Control of Production Systems, BIBA, University of Bremen, Germany
[2] Institute for Microsensors, -actuators and -systems, University of Bremen, Germany

Abstract We applied an adaptive controller to the general model of a replenishment rule. The replenishment rule is the "automatic pipeline, variable inventory and order based production control system". In the current literature the possibility to reduce the bullwhip effect and demand fluctuations by setting the right smoothing parameters to this replenishment rule was shown for independently and identically distributed stationary stochastic demand. In the real world the demand uncertainties are changing (e.g. along the product life cycle). On this account we developed a gain scheduling adaptive control mechanism, which is able to adjust the parameter based on states of the system and demand uncertainties. With the adaptive controller the system can be adapted to uncertainties in demand by holding the production fluctuations stable and adjust smoothly to a new appropriate inventory coverage level by setting the right inventory adjustment parameter, the right target inventory level and the right forecast time.

1 Introduction

The design of the production and inventory control (PIC) policy is an important factor for effective supply chain. This rule determines the dynamics in the supply chain and the costs for production and inventory holding. Inventory costs rises based on inventory holding increase which is created by uncertainties in demand, supply, production and control policy. Production costs increase with ramp-up and phase out of production and capacity holding, caused by changing order and production rate. Furthermore fluctuations in order quantity build up along the supply chain from the point of origin to the point of consumption. This effect called "Bullwhip Effect" (Lee et al. 1997) and was shown in many real-world supply chains (e.g. Holmström 1997).

The main causes for the bullwhip effect are "non-zero lead times", "demand signal forecasting", "order batching", "gaming" and "promotions" (Lee et al. 1997). We focus on "non-zero lead times" and "demand signal forecasting" as core prob-

H.-D. Haasis et al., *Dynamics in Logistics*

DOI: 10.1007/978-3-540-76862-3,

lems of the supply chain and neglect the other effects. The similar bound are used by Chen et al. (2000) and Lalwani (2006). The bullwhip effect is measured by the "variance ratio" based on Chen et al. (2000). The variance ratio is the long term ratio between the variance of orders (OR) over the variance of demand (D). The ratio can be applied for single order decision (e.g. Disney and Towill 2003) or order decision in each echelon over the whole supply chain (e.g. Hosoda 2006). In order to measure the effectiveness of the policy in terms of inventory holding the variance ratio of the inventory is used. The variance inventory ratio is defined as the variance of inventory (I) over the variance of demand.

The aim is the development of a production and inventory control policy under the trade-off of production and order fluctuations on one hand and inventory fluctuations on the other hand. As starting point we focus on the order-up-to policy. Dejonckheere et al. (2003) showed that the order-up-to policy is a special case of the general "automatic pipeline, variable inventory and order based production control" (APVIOBPC) policy. Beyond he pointed out that by setting the right parameters the APVIPBPC policy is able to eliminate the bullwhip effect and even reduce the demand fluctuations. On the same path, Disney et al. (2004) calculated the "golden ratio" between production and inventory fluctuations depending on the weight on inventory and capacity costs. In their research they set restrictive assumptions on the parameters and they ignore changes in demand fluctuations.

In this paper, we present a production and inventory control policy with adaptive parameters considering changing of demand fluctuations. Our approach refers using adaptive control theory, which gives new possibilities to adjust parameters despite uncertainties like changing demand.

The paper is organised as follows: In Sect. 2 the production and inventory control system with policies is described by using block diagram, transfer function and stability analysis. In Sect. 3 the measurement of the bullwhip effect and the inventory holding is characterised. Furthermore we discuss the overall objective function and introduce a new one. The adaptive control methods are compared and the "Gain Scheduling Controller" is chosen and also the proceeding is determined. In Sect. four the production and inventory control policy is expanded by adaptive control variables considering changing uncertainty in demand.

2 The production and inventory control policy

In this paper we studied the APVIOBPC system after John et al. (1994) which is the most parameterised form of inventory and order based production control (IOBPC) system based on Towill (1982). The IOBPC family is the basic for control theoretical approaches and has been used in many studies (Lalwani et al. 2006).

The APVIOBPC policy defines the desired production and the replenishment orders as sum of "demand forecast", "work-in-progress adjustment" and the "inventory adjustment" with "variable inventory target". For the demand forecast different forecast can be used. In our case we used exponential smoothing forecasting with T_a

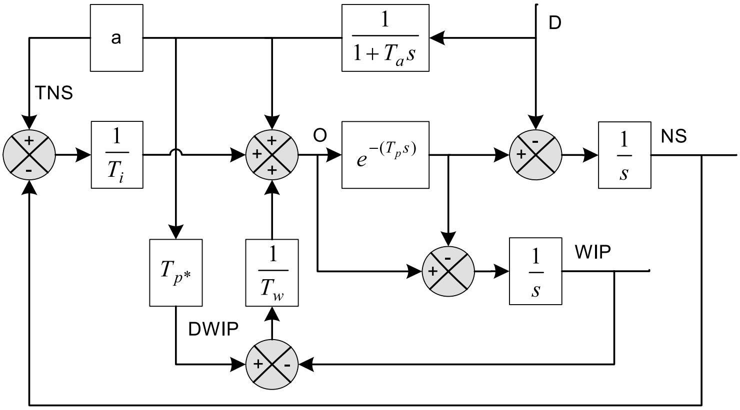


Fig. 1 Block diagram of APVIOBPC

as adjustment parameter. The work-in-progress adjustment is defined as the ratio of the actual work-in-progress minus the desired work-in-progress over the adjustment time (T_{WIP}). The inventory adjustment calculated in the similar way by the ratio of the actual inventory minus the desired inventory over the adjustment time (T_I). Furthermore the desired inventory (a) is the product of inventory target and forecasted demand.

The block diagram shown in Fig. 1 is a continuous time representation of the APVIOBPC policy. Disney et al. (2006) developed two transfer functions in frequency domain of the system using Laplace transformation. One between orders (O) and demand (D) and the other between net stock (NS) and demand (D).

In the transfer functions five parameters describe the system. They are: T_p, production lead-time, is modelled as a pure time delay, $T_p > 1$; T_{p*}, estimation of production lead time, $T_p > 1$; T_a, average age of the exponential forecast, $T_a > 1$; T_{wip}, smoothing parameter for the gap between actual work-in-progress and desired work-in-progress, $T_{wip} > 1$; T_i, smoothing parameter for the gap between actual inventory and desired inventory, $T_i > 1$.

For stability analysis the bounded input, bounded output (BIBO) stability is used. The system is stable when the system response on the bounded input (e.g. step input) with a steady state in infinite time is reached. The drivers for stability of the system are only feedback loops. Riddalls and Bennet (2002) proof the stability of the APVIOBPC policy excluding the variable target inventory feedback loop.

3 Variance ratios and objective function

The variance ratios of the order variance over the demand variance and inventory variance over demand variance are important measurements of the system and the policy respectively. They describe the bullwhip effect and the inventory variance based on the input variance. In order to understand the system better, Disney and Towill (2003) defined another expression of the general variance ratio based on independently and identically distributed stationary stochastic input.

Based on the general variance ratio Disney et al. (2004) calculated the bullwhip and the inventory variance ratio depending on T_a, T_p and T_i. They assumed following relationships $T_{p*} = T_p$ and $T_i = T_{wip}$, because they need to reduce the complexity. For the bullwhip ratio they conclude that the bullwhip effect decreases with a smaller production lead time T_p, increasing forecast smoothing and increasing smoothing parameters for work-in-progress and inventory feedback loop. The inventory ratio decreases with the smaller production lead time. The minimum inventory variance occurs by the smoothing parameters for work-in-progress and inventory feedback loop equals to one. In contrast to the bullwhip ratio the inventory ratio increases with smaller smoothing parameter (T_i).

Therefore the identification of the appropriate smoothing parameter (T_i) is clearly a trade-off between production (bullwhip) and inventory amplifications. In order to solve this trade-off an objective function is introduced. The basic objective function is the minimum of the sum of the production and inventory amplifications. Disney et al. (2004) used this objective function with weighting parameter (w) for the trade of for inventory and capacity costs dominance. For $w = 1$ the production fluctuation costs are very high comparing the inventory holding costs, this is called capacity costs dominance. In opposite for $w = 0$ the production fluctuation costs are negligible compared to the inventory costs. They solved the objective function and detect the "golden ratio".

The "golden ratio" split the demand fluctuations in percentages of production and inventory fluctuations. If uncertainties in demand are constant (modelled as constant distribution of demand), the objective function is appropriate. For changing uncertainties the objective function is not sufficient, because the production fluctuations are bounded by capacity and inventory fluctuations are bounded by inventory. On this account, adaptive controller of the system is developed by considering the energy of demand signal and the target inventory level.

4 Methodology

When use of a linear and fixed controller cannot achieve a satisfactory compromise between stability and desired performance of the process then adaptive control should be considered for controlling the system despite new conditions. Therefore because of adaptation of the controller to nonlinear nature of the system and benefits of stability concepts, nowadays application of adaptive controllers have been considered in process control studies; Adaptive control design deals with adjusting system parameters for obtaining desired performance and global stability.

There are various forms of adaptive controllers which were established for adapting whole system to new changes. The important similarity between the various adaptive control approaches is adjusting parameters of the controller based on evo-

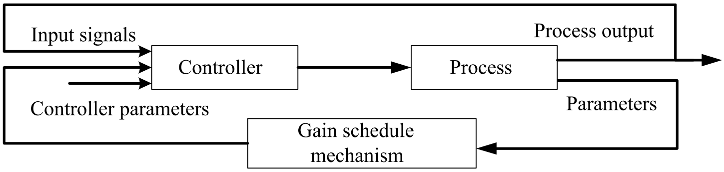


Fig. 2 Schematic of Gain scheduling adaptive controller

lutions of variations in the plant. There are some important adaptive approaches like model reference adaptive control, self tuning, dual control and gain scheduling controller. In gain scheduling technique, controller is tuned for several operation conditions of the process. Based on the process measurement, the schedule is then used to change the tuning despite new conditions. In this project, gain scheduling adaptive controller was applied on the process for achieving better performance and stability.

The idea of gain scheduling adaptive control is to find sensitive process parameters which will correlate with changes in dynamics of the process. By this way, it will be possible to compensate variation of plant variables by proper controller commands based on the events (Fig. 2). In fact, after any changes in sensitive parameters of the system, the controller will adapt the whole process to new circumstances. In operation, the parameters are measured and the controller is scheduled according to the parameter variations. When the process changes despite operating conditions in known and predictable manner, gain scheduling can provide the required performance of the system.

It is necessary to assure stability of the whole system after installation of the controller. The system should be stable against uncertainties, disturbances and new unknown events. Thus for designing the controller, also stability studies should be considered for obtaining desired performance. There are some stability analysis methods for non-linear systems like bounded- input bounded- output (BIBO) and Lyapanov direct and indirect approaches. These approaches are widely used for stability analysis of different systems (Bacciotti et al. 2002).

In BIBO systems, if every bounded input $(x(t))$ results in a bounded output $(y(t))$, then the statement that $x(t)$ is bounded, means that there exists a finite value (N) which $|x(t)| <= N$ for all times, and also for output signal. If all Bounded Inputs produce bounded Outputs, the system will be externally stable and the whole system will be stable against uncertainties. In this paper, BIBO stability analysis was considered for inventory control.

5 Adaptive policy

The development of an adaptive PIC policy required knowledge about environment, the system and the behaviour of the system (e.g. sensitive parts of the system). For the system environment we focus on demand uncertainty. The changing uncertainty in demand has many reasons (e.g. changing customer behaviour and changing competition). The demand uncertainties can be measured as the energy of noise over the demand trend (in rms). In order to define the adaptive policy based on demand uncertainties first we have to find important operating points. The approximation for the operating points is based on the new objective function described in Sect. 3. This function defined as minimum between production and inventory variance considers the boundaries of production and inventory fluctuations. This function depends on the driving feedback loops and related adjustment parameters (see Sect. 2). In detail there are the "demand forecast" (T_a), "work-in-progress adjustment loop" (T_{wip}), the "inventory adjustment loop" (T_i), and the "variable inventory target loop" (a).

Afterwards a sensitivity analysis for changing demand uncertainties and the adjustment parameters are accomplished based on model of the system and the objective function in Matlab. The "work-in-progress adjustment loop" with the adjustment parameter (T_i) doesn't have a big influence on the objective function in the range of $1.5 < T_{wip} < 20$. The reason is the structure of the control loop as damped balancing loop with the relatively short time delay T_p. Damped balancing loops with short time delays are not able to generate large fluctuations. The "inventory adjustment loop" with the parameter T_i in the case of changing demand uncertainties is more sensitive. The causes are again in the structure. The "inventory adjustment loop" is partly target value for the "work-in-progress adjustment loop" and generates thereby additional fluctuations in this loop. Moreover the "inventory adjustment loop" with the adjustment parameter $1.5 < T_i < 20$ is also a damped balancing loop but with a longer variable time delay, approximately T_p plus a. On this account the parameter T_i is an important adjustment variable. The "demand forecast" with the average smoothing time T_a works as a filter for short-waves. The short-waves disturb the system, because the system has long time delays. For increasing uncertainties in demand we need increasing T_a to filter all short waves out. Additionally the parameter is the target for the WIP and inventory adjustment loop, and the target inventory and should be an adaptive parameter. Also in the sensitivity test parameter was highly sensitive. The "target inventory" (a) gives the target for the "inventory adjustment loop" and influences thereby the "WIP adjustment loop". For this reason the parameter is a highly sensitive and should only use as last opportunity to set the system to a new equilibrium.

Based on this relationship between the described components of the system, the following parameter-scheduled mechanism was defined and applied to the system. First the demand uncertainty (measured in rms) is used to define the parameters T_i and T_a in the system. The parameter of T_i and T_a holds the production fluctuation stable in desired boundaries and increases partly the inventory fluctuations. If the rms value of inventory is higher, then a certain percentage of the inventory than the inventory target will be set to a higher value. This adjustments increase the instability of the system, because the inventory target loop changes the target of the whole system. The target adjustment generates high changes in the work-in-progress loop and small changes in the inventory target loop, due to the fact that the adjustment parameter T_{wip} has a low value and T_i a high value. Therefore by installation of gain scheduling controller, the process obtained adaptation features to behaviour of the system, external uncertainties and parametric changes. The applied gain scheduling controller against changing demand uncertainties is shown in Fig. 3, including demand, process, and adaptive control diagram and parameter adjustment mechanism.

After development of the adaptive control policy we simulated the new adaptive APVIPBPC and compared with the APVIPBPC one (see Fig. 4). The demand as input has mean of 150 [parts/day], standard deviation is 10 [parts/day] (1–60 day) and 100 [parts/day] (61–160 day). The behaviour is shown after a transition state. In Fig. 4 two policies are shown, the APVIPBPC policy and the adaptive policy. In the right Figure the production start rate (PSR) and in the left Figure the inventory level are presented. In the APVIPBPC policy the PSR fluctuations are increasing with the increased standard deviation of demand. The inventory fluctuations are increasing in appropriate way. The adaptive policy comparing to the normal policy decreases PSR fluctuations by small increasing in inventory fluctuations. For this reason we introduced the adaptive policy to damp the fluctuations depended on demand uncertainties. This behaviour optimises the objective function for every demand uncertainty levels.

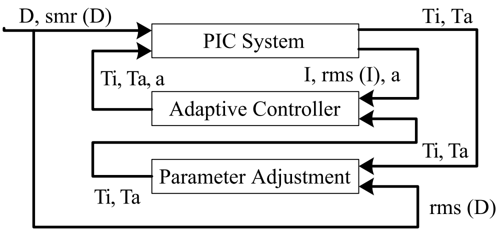


Fig. 3 Application of gain scheduling approach for adaptive inventory control

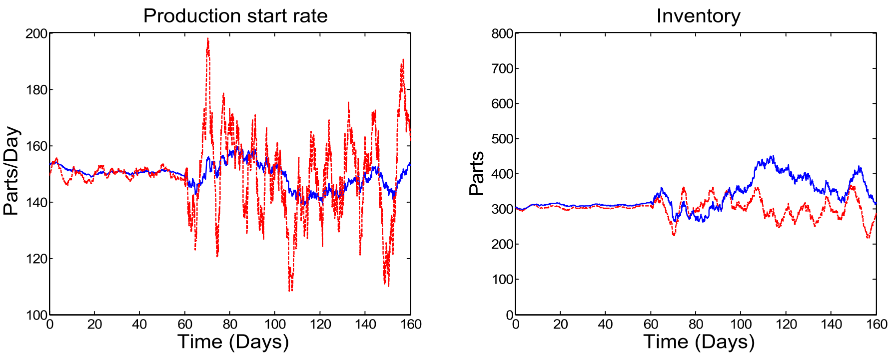


Fig. 4 Comparison between APVIOBPC (Dashed) and the adaptive policy (Line)

6 Summary

In this paper, an adaptive controller has been applied to production inventory control system. The aim of the controller is to balance the trade-off between production and inventory fluctuations under changing demand uncertainties. For this purpose gain scheduling algorithm has been considered in setting sensitive parameters including inventory adjustment, forecast parameters and target inventory for adapting behaviour of system to new circumstances. Bounded-input, bounded-output method is applied for stability analysis despite demand uncertainties and performance of the process is compared before and after using adaptive controller.

References

Bacciotti A, Mazzi A (2000) A Necessary and Sufficient Condition for Bounded-Input Bounded-State Stability of Nonlinear Systems, SIAM Journal on Control and Optimization, 39 (2) 478–491.

Chen FZ, Drezner JK, Simchi-Levi D (2000) Quantifying the Bullwhip Effect in a Simple Supply Chain: The Impact of Forecasting, Lead-times and Information, Management Science, 46, 436–443.

Dejonchheere J, Disney SM, Lambrecht MR, Towill DR (2003) Measuring the Bullwhip Effect: a Control Theoretic Approach to Analyse Forecasting Induced Bullwhip in Order-Up-To Policies, European Journal of Operations Research, 147 (3), 567–590.

Disney SM, Towill DR (2003) On the bullwhip and inventory variance produced by an ordering policy, OMEGA, 31 (3) 157-167.

Disney SM, Towill DR, van de Velde W (2004) Variance Amplification and the Golden Ratio in Production and Inventory Control, Int. J. Production Economics, 90, 295–309.

Disney SM, Towill DR, Warburton DH (2006) On the equivalence of control theoretic, differential, and difference equation approaches to modelling supply chains, Int. J. Production Economics, 101, 194-208.

Holmström J (1997) Product range management: A case study of supply chain operations in the European grocery industry, 2 (3), 107–115.

Hosoda T, Disney SM (2006) On variance amplification in a three-echelon supply chain with minimum mean square effort forecasting, OMEGA, 34, 344–358.

John S, Naim MM, Towill DR (1994) Dynamic analysis of a WIP compensated decision support system, Int. J. of Manufacturing System Design 1 (4), 283-297.

Lalwani CS, Disney SM, Towill DR (2006) Controllable, observable and stable state space representations of a generalized order-up-to policy, Int. J. Prod. Eco. 101, 172–184.

Lee HL, Padmanabhan P, Whang S (1997) The bullwhip effect in supply chains, Sloan Management Review, 38 (3), 93–102.

Riddalls CE, Bennet S (2002) The stability of supply chains, Int. J. of Production Research 40, 459-475.

Towill DR (1982) Optimisation of an inventory and order based production control system. Proceedings of the Second Int. Symposium on Inventories, Budapest, Hungary.

A Framework of Adaptive Control for Complex Production and Logistics Networks

Dmitry Ivanov, Marina Ivanova

Chemnitz University of Technology, Chair of Production Management, Germany

Abstract The collaborative networks (CN) must be configured according to the project goals and reconfigured in dynamics according to the current execution environment. In practice, CN design decisions are rarely focused on "green Field" situation. More typically questions are centred on adaptation and rationalizing the CN in response to permanent changes of CN itself and its environment. That is why we consider the problem of business continuity supporting and continuous adjustment of collaborative networks in accordance with permanent changes of project execution environment as a critical point in the CN research.

Subject of this contribution is to elaborate a methodological basis of the CN adaptation. In the CN, adaptation challenges are caused by self-interested competitive behavior of participated enterprises, unpredictable execution environment, structure dynamics structure dynamics of both CN itself and of the supporting infrastructures like information processes. To answer these challenges, we introduce the framework of CN adaptive control to increase the quality of decision-making about the CN reconfiguration and adjustment based on planning under the terms of uncertainty, execution under the terms of adaptation, and reflections of planning and execution phases of the CN control on the adaptation principles.

We illustrate the methodological framework on an analytical example, which matches the stability and robustness analysis in the planning phase with the CN adaptation in the execution phase based on a systematic approach to derive the adaptation measures according to different deviations. First, the multi-stage adaptation concept is presented. We draw the conclusion that the CN robustness must be considered in connection to the stability analysis and CN adaptation. We show that particularly the user-controlled adaptation plays the key role in supply chains to achieve or to hold the stable or rather robust state. To amplify this idea, the CN adaptation and stability/robustness analysis are considered interrelated to each other. We finish this section with a short numerical example. The article is concluded with a summary of the achieved results and prospects for future developments. The practical applicability of the presented results can be seen in elaborating a comprehensive approach to decision making in complex value-adding partnerships, which allows generating stable plans and guidelines for operative CN adjustment.

H.-D. Haasis et al., *Dynamics in Logistics*
DOI: 10.1007/978-3-540-76862-3,

1 Introduction

In many branches, wide-broaden hierarchical supply chains with pre-determined suppliers structure and product programs evolve into customer-oriented, temporary networking of core competences of small and medium enterprises. These concepts, known as Virtual Enterprise or Agile Networks, become a popular form of competitive organizations. In this paper, the term Collaborative Network (CN) is used as a general definition for such networks (Camarinha-Matos et al. 2005).

The CN must be configured according to the project goals and reconfigured in dynamics according to the *current execution environment*. In practice, CN design decisions are rarely focused on "green Field" situation. More typically questions are centred on adaptation and rationalizing the CN in response to permanent changes of CN itself and its environment. That is why we consider the problem of *business continuity supporting and continuous adjustment of collaborative networks in accordance with permanent changes of project execution environment* as a critical point in the CN research.

Subject of this contribution is to elaborate a methodological basis of the CN adaptation. In the CN, adaptation challenges are caused by self-interested competitive behavior of participated enterprises, unpredictable execution environment, structure dynamics structure dynamics of both CN itself and of the supporting infrastructures like information processes. To answer these challenges, we introduce the framework of CN adaptive control to increase the quality of decision-making about the CN reconfiguration and adjustment based on planning under the terms of uncertainty, execution under the terms of adaptation, and reflections of planning and execution phases of the CN control on the adaptation principles.

2 State-of-the-art

The CN execution is accomplished by permanent changes of internal network properties and external environment. It requires dynamic CN adaptation to the current execution environment and the goals and decisions of the configuration phase. Adaptation is a category, which attracts more and more research attention in the area of enterprise networks. Complex adaptive systems are special cases of complex systems. The term *complex adaptive system was* coined by John H. Holland, Murray Gell-Mann and others (Holland 1995, Gell-Mann 1995). A Complex Adaptive System (CAS) is a dynamic network of many agents acting in parallel, constantly acting and reacting to what the other agents are doing. The control of a CAS tends to be highly dispersed and decentralized. If there is to be any coherent behavior in the system, it has to arise from competition and cooperation among the agents themselves. The overall behavior of the system is the result of a huge number of decisions made every moment by many individual agents. Natural complex systems are modeled using the mathematical techniques of dynamical systems, which include differential equations, difference equations and maps. A view on CN management from the

CAS theory is considered i.e. in (Tielbein 2006). (Kopfer and Schoenberger 2006) consider the online-optimization of problems with several multi-layered objectives, which cannot be integrated in a multi-criteria optimization problem, because they are situated on different decision levels. The authors present an approach for the extension of computer-aided decision systems so that they are capable to make a self-adaptation to a given superordinated objective and apply it to a problem of dynamic vehicle routing and scheduling.

The ideas of adaptive production planning and control are based on the systems planning approach (Bellmann 1972) that interprets planning not as discrete operations, but as a continuous adaptive process. A posteriori, current, and a priori information can be used for plan adaptation (adaptation to the "past", "present", or "future"). The adaptive planning considers the processes of the structure dynamics (including the control of the CN structure reconstruction) in terms of general context and their interrelations with the CN control processes at various stages of the CN lifecycle. Therefore, corresponding groups of structural dynamics control problems may be regarded as the subgroup of control problems of generalised dynamic systems with reconfigurable structure.

The Systems Science and Control theory can be used for the CN adaptation. They can serve as general methodological basics of complex systems (re)-synthesis and analysis. However, the disadvantage of the research grounded in modern *systems and control theories* regarding complex business systems is that the system elements are being controlled from a central decision-maker and cannot change their states and interactions on their own free will (the system elements are passive). In complex business systems, the elements are *active* (their can compete and have conflicting aims, interests, and strategies). However, the classic methods do not allow developing practical comprehensive models taking into account the *goal-oriented (active) behaviour* of enterprises.

3 Research methodology: MARINA

This section presents general principles of the MARINA (Method of Analysis, Reconfiguration and Integrated Network Adjustment). The MARINA approach is developed to implement the comprehensive consideration of uncertainty for production and logistics processes planning and control. The approach aims at developing *comprehensive modelling approach for business continuity supporting and continuous adjustment of collaborative networks in accordance with permanent changes of project execution environment.* The MARINA approach provides interlinked conceptual methods, mathematical models and algorithms for generating CN stable plans and operative decision making for the CN reconfiguration. The main distinguishing features of the proposed approach MARINA are the following:

- application of adaptation principles,
- reflection of planning and execution models, and
- reflection of conceptual and mathematical models.

We developed the following constructive tools to implement CN planning under uncertainty. They are:

- Embedding risk management in planning models (uncertainty factors known to the planning moment i.e., by means of stochastic models),
- Stability analysis (uncertainty factors unknown to the planning moment),
- Matching static and dynamics models by means of functors (interlinking of planning and execution models).

The goal of CN planning is to generate a stable plan. Uncertainty analysis on this stage is comprised of two phases: taking into account factors known to the planning moment (i.e., by means of stochastic models) and factors unknown to the planning moment. For the first phase we apply an elaborated general scheme of CN modeling under the terms of uncertainty (Ivanov 2007). For the second phase, stability analysis is applied. In general case stability analysis consists in investigation of influence of the CN execution parameters deviations on the final CN goals. The stability analysis allows proofing execution plan feasibility, selecting a plan with the sufficient stability degree from a set of alternative plans, determining CN execution bottle-necks and steps for their strengthening. In Sect. 4, we will come back to the stability analysis in the settings of the CN adaptation.

The execution models should ensure operative decision making about the CN changing and provide feedbacks to the planning models. We developed the following constructive tools to implement these principles in order to ensure correspondence of the CN states to the actual execution environment. These principles allows taking into account the third type of uncertainty (uncertainty factors emerging while CN execution). These principles are:

- Information update concept,
- Conceptual model of CN adaptation and supporting mathematical models,
- and a system of CN performance indicators.

The conventional tools of enterprise network operative control (such as APS and SCEM systems) evince considerable deficiencies. Their optimization cycle is slow and does not let appropriate taking into account operative oscillations in demand, material availability, lead times, production charges etc. Besides the parametrical oscillations, the structural and goal oscillations must be taken into consideration. Various structures changing, such as organizational, technological, informational, financial, might let to the situation when initial CN models would be no more representative and adequate. The clients and network participants' goal changing can also cause the necessity of the models changing (adaptation). The execution models should provide feedbacks to the planning models. We developed the following constructive tools to implement these principles. They are:

- Information update concept,
- Conceptual model of CN adaptation and supporting mathematical models and algorithms, and
- System of CN performance indicators.

In the information update concept (Bellmann 1972, Sethi et al. 2005) posterior, current, and a priori information can be used for plan adaptation. The information update concept presumes continuous renewing information about the CN functioning, including changing external environment and internal CN properties. The basis of the information update concept is CN monitoring. In our approach, CN *monitoring* is based on the monitoring of the CN *macro-structural macro states* (Ivanov 2007).

The elaborated concept of the complex CN adaptation is built as a five-level structure. Each level characterizes certain control loop in accordance with the oscillations and deviations appeared (Ivanov et al. 2006). These adaptation levels and management actions are determined based on the stability analysis (Ivanov et al. 2006), which was performed on the planning stage. The mathematical models of the CN adaptation are based on this conceptual model from the DIMA-methodology (Ivanov 2007).

System of CN performance indicators is elaborated to support operative decision making. It provides quantitative indicators which reflect operative states of CN execution regarding their correspondence to the planning parameters.

4 Illustration

This section illustrates the methodical principles described in the Sect. 3. *According to MARINA,* CN were planned under the terms of uncertainty. There were defined possible bottle-necks (risk management concept), various classes of CN execution parameter deviations (stability analysis), and rules of matching CN planning and execution models. While executing, based on information update and performance indicators analysis, the CN can be systematically adapted to the actual execution environment. Figure 1 presents the conceptual model of the CN adaptive control.

According to MARINA, CN is planned under the terms of uncertainty and risk. There are defined possible bottle-necks (risk management concept), various classes of CN execution parameter deviations (stability analysis), and rules of matching CN planning and execution models. While executing, based on information update and performance indicators analysis, the CN can be systematically adapted to the actual execution environment.

The Fig. 2 presents the dynamic of the CN execution in case of deviations and paths of the operability recovery. To take into account particular features of the CN stability analysis (system elements' activity and structure dynamics), the CN execution is described in terms of multi-structural macro states S (Ivanov et al., 2004, 2005) and the decision making about the CN reconfiguration are based on the concept of decision making in the CN based on the combined using of multi-agent and control theory frameworks (Ivanov 2007).

The Fig. 2 depicts various variants of system behaviour changing in case of any disturbances at the system state $S_1^{(u)}$ of the initial execution plan. The perturbation impacts may cause various execution parameters deviations Δp_i and corresponded stability decrease regarding the final goals $C(t = end) = \{c_1 \ldots, c_c\}$ of a supply

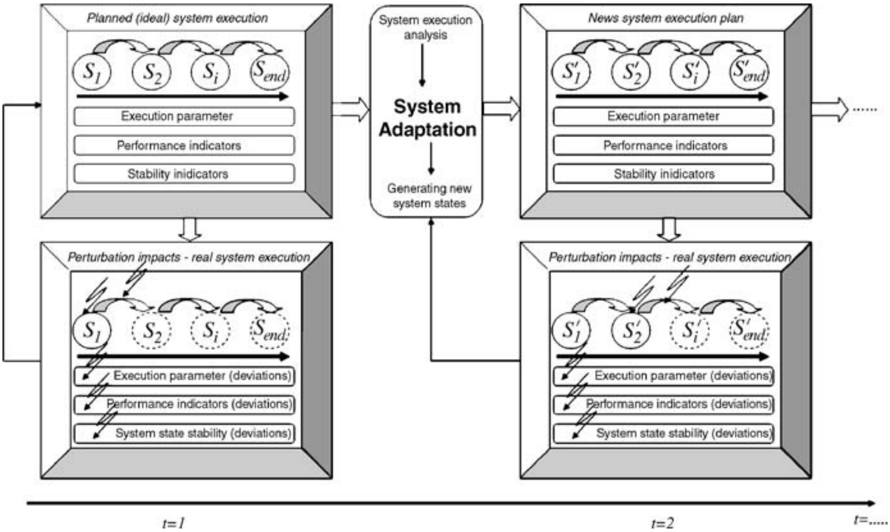


Fig. 1 Conceptual model of the CN adaptive control

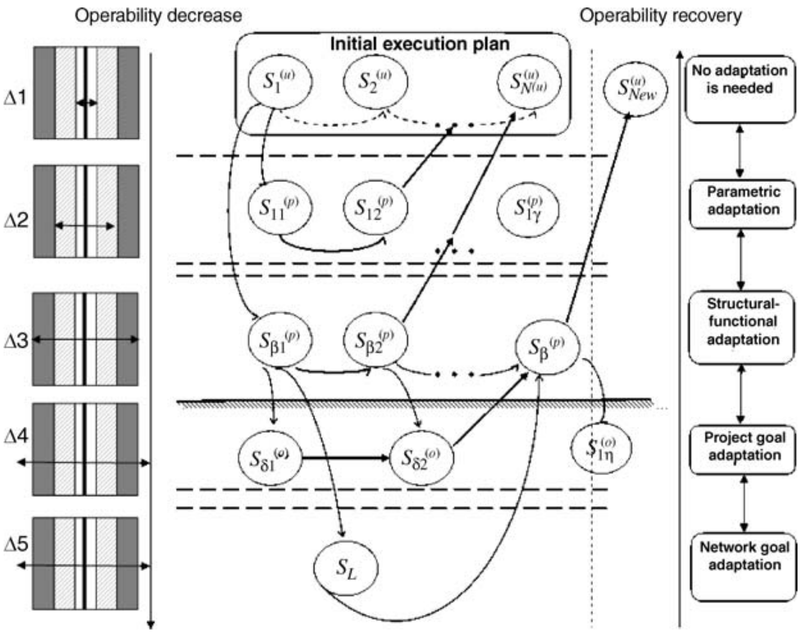


Fig. 2 Dynamics of the CN execution and operability recovery

chain or project. To match the system stability analysis and taking steps to recover the system stability, we elaborated a concept of the complex CN adaptation is built as a five-level structure.

Each level characterizes certain control loop in accordance with the oscillations and deviations appeared. Each adaptation level corresponds to certain management actions. We distinguished parametric adaptation (i.e., rush orders), structural-functional adaptation (i.e., supplier structure changing), project goals adaptation (i.e., delivery delay), CN goals adaptation (i.e., network profit changing) as well as CN models adaptation (Ivanov et al. 2006). It makes it possible to match the results of stability analysis and actual execution analysis. It also provides a decision-maker with a tool what decisions about the CN adjustment must be taken in a current situation. The Fig. 3 depicts this idea.

Various deviations Δp_i from the ideal parameter value $p_i(t)$ correspond to three classes of CN states: CN state is stable and no adjustments are needed, CN state is stable, but adjustments are needed, and CN lost its stability (the final goal can not be achieved any more under the current conditions). The operative decision making about CN reconfiguration and adjustment are based both on the execution information and planning results (i.e., selection of adaptation level in case of certain parameter deviations is related to the stability analysis). In case of deviations (both in future – preventive reconfiguration, and during the current operation – reactive reconfiguration), the decision maker can timely take necessary steps for the CN adjustment. It is ensured by that the system of CN performance indicators is methodically and practically interconnected to the framework of stability analysis (planning stage) and the five-level adaptation framework (execution stage).

Finally, let us present an example of the CN adaptation in case of a certain deviation. We consider the following case. We refer to the Fig. 2 and assume, that a perturbation impact at the system state $S_1^{(u)}$ of the initial execution plan caused a deviation Δp_i (20%) of the execution parameter "Capacity".

According to the results of stability analysis, the decision maker classifies this deviation under the class Δ_2, and derives that the CN will not loose its stability in case of the parametrical adaptation correspondingly (s. also Fig. 3). The following schema of the CN adaptation contains the following steps: interactions between CN

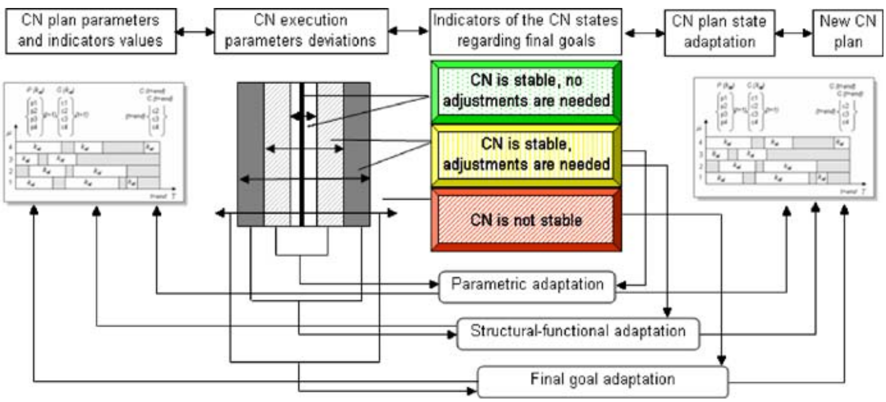


Fig. 3 Matching the CN stability analysis and adaptation

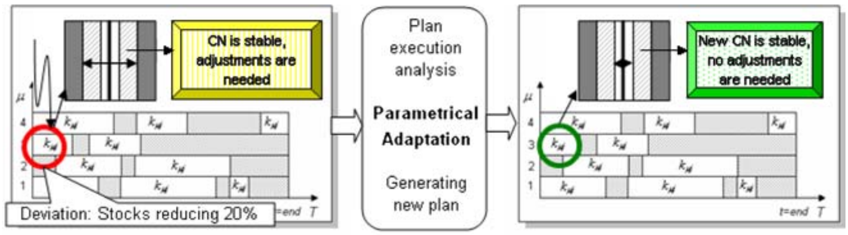


Fig. 4 CN reactive adaptation in case of a deviation

coordinator and agents (new CN execution plans generated by the network coordinator are adjusted and specified by agents' interactions) → decision making about the CN reconfiguration (in this particular case of the parametrical adaptation – a rush order) → generating of a new CN execution plan. In our case – the path $\{S_1^{(u)}, S_2^{(u)}\}$ is replaced with the path $\{S1_1^{(p)}, S_{12}^{(p)}\}$ with the following transition to the initial execution plan trajectory (s. Fig. 2).

This case illustrated a so called reactive adaptation (reaction of any deviation, appeared at any CN state while operating this state). An other situation is a preventive adaptation or adjustment, which is undertaken in case of any deviations, appeared at any CN states in the future while operating another CN states. The schema of the CN adaptation will be the same with only exception that the backwards principle of adaptation will be used.

5 Conclusions

The elaborated adaptation framework aims at developing *comprehensive modelling approach for business continuity supporting and continuous adjustment of collaborative networks in accordance with permanent changes of project execution environment.* It makes it possible to increase the quality of decision-making about the CN reconfiguration and adjustment based on planning under the terms of uncertainty, execution under the terms of adaptation, and reflections of planning and execution phases of the CN control on the adaptation principles. It also allows comprehensive taking into account three main types of uncertainty regarding a decision maker: factors known to the planning moment (i.e., by means of stochastic models), factors unknown to the planning moment, and uncertainty factors emerging while CN execution.

References

Bellmann R (1972) Adaptive Control Processes: A Guided Tour. Princeton Univ. Press, Princeton, New Jersey

Camarinha-Matos LM, Afsarmanesh H (2005) Collaborative networks: a new scientific discipline, in Journal of Intelligent Manufacturing, 2005, 16, 439–452

Gell-Mann M (1995) Complex Adaptive Systems, in: Morovitz H, Singer JL (eds.), The Mind, The Brain, and Complex Adaptive Systems, Reading: Westview Press, pp. 11–23

Holland JH (1995) Hidden Order: How Adaptation Builds Complexity, Reading: Addison Wesley

Ivanov D, Käschel J, Arkhipov A, Sokolov B (2006) A Conceptual Framework for Modeling Complex Adaptation of Collaborative Networks, In: Network-centric collaboration and supporting frameworks, edited by. L.M. Camarinha-Matos, H. Afsarmanesh, M. Ollus, Springer, pp. 15–22

Ivanov D (2007) DIMA – Decentralized Integrated Modeling Approach. Interdisziplinäre Modellierung von Produktions- und Logistiknetzwerken (2007), Verlag der GUC, Chemnitz

Ivanov D, Arkhipov A, Sokolov B (2007) *Intelligent planning and control of manufacturing supply chains in virtual enterprises*, In: International Journal of Manufacturing Technology and Management, Vol.11 (2), pp. 209–227

Kopfer H, Schoenberger J (2006) Die Lösung von Optimierungsproblemen mit mehrschichtigen Zielsetzungen durch die Adaption von selbststeuernden Planungsagenten, in: Proceedings of the German-Russian Logistics Workshop, 20.–21 April, Saint Petersburg, Russia, pp. 93–102

Sethi SP, Yan H, Zhang H (2005) Inventory and Supply Chain Management with Forecast Updates. Springer

Tilebein M (2006) Decentralized Supply Chain Management: A View form Complexity Theory, In: Complexity Management in Supply Chains, edited by T. Blecker, W. Kersten, Erich Schmidt Verlag, pp. 21-36

Mechanisms of Instability in Small-Scale Manufacturing Networks

Reik Donner[1], **Uwe Hinrichs**[2], **Bernd Scholz-Reiter**[2]

[1] Institute of Transport and Economics, Dresden University of Technology, Andreas Schubert-Str. 23, 01062 Dresden, Germany
[2] Bremen Institute of Industrial Technology and Applied Work Sciences, University of Bremen, Hochschulring 20, 28359 Bremen, Germany

Abstract The intrinsically unstable dynamics of production networks is a serious economic danger not only for small companies. For example, fluctuations of processing times cause an irregular behaviour which can lead to production breakdowns due to a lack of material or time intervals with an intermittent accumulation of material or unprocessed orders. However, event-discrete simulations indicate that such fluctuations may also have a constructive effect on the dynamics: If a push strategy is applied for the production process, long down times of the machines after each processed workpiece may lead to a successive increase of certain stocks, which is decelerated in the presence of fluctuating processing times. In the more realistic case of a pull control, improper lot sizes result in an aperiodic behaviour even for constant processing times. The critical parameters associated with the different effects are identified and further evaluated for linear supply chains and symmetrically interacting manufacturers.

1 Introduction

Present-day companies act on a global market which faces them with numerous challenges. As this situation requires the ability to flexibly and fastly adapt to changing economic conditions, manufacturing systems or, more general, business units are characterised by an increasing complexity and diversity of goods and services. During the last decades, many enterprises have reacted to the corresponding requirements by a concentration on their specific core competences, incorporating concepts like outsourcing of special tasks of production and logistics. As a consequence, large-scale economic networks have successively developed, whose dynamics is additionally influenced by a variety of uncontrollable network-external factors, for example, varying supply and demand conditions [1].

The combination of both, diversification and specialization tendencies, is however known to have severe negative effects on the overall production process. Prominent examples are production breakdowns due to a lack of material as well as large (and usually irregular) oscillations of stocks which amplify along a supply chain.

H.-D. Haasis et al., *Dynamics in Logistics*

DOI: 10.1007/978-3-540-76862-3,

The classical approach to meet this situation is the decoupling of business units by large buffers. This approach allows balancing variations in the delivery of material. However, high inventory levels require large initial investments, store large amounts capital in terms of material inside the buffers and increase the time between the delivery of material and the production of the final goods. As the cycles of development and production are gradually accelerating in many businesses, these features mean a high economic risk, especially for smaller manufacturers. Therefore, instead of a decoupling, real-world companies often tend to closely link to each other by applying strategies like just-in-time logistics [2]. As a consequence, the resulting business networks may become complex, which may result in a very irregular and unpredictable dynamics again [3].

The simulation-based optimisation of production processes in terms of their logistic performance focuses on the minimisation of inventory levels and throughput times under a simultaneous maximisation of the reliability of the networks [4]. In order to map a production system onto a suitable mathematical model, different approaches are possible. For the analysis of general features of production networks, continuous-flow [3,5] or discrete-map models can be used to prescribe the essential structure of interactions and study the main features of the dynamics on such networks. However, in order to map the particular structure of a real-world manufacturing system, there are well-developed software packages which allow to study the dynamics on a designated network structure in more detail by means of continuous-time or discrete-event simulations. In contrast to rather general physical networks, the study of the corresponding models allows the identification of economic potentials and the successively testing of various optimisation and control strategies.

In real-world large-scale manufacturing networks, the complex dynamics and high dimensionality of the system require not only an adaptation of existing, but also a development of new methods for modelling, analysis, and control. Consequently, a fundamental understanding of the dynamics in small-scale systems is necessary at first, which applies to both the intrinsic dynamics and the potential impact of external influences. The corresponding information may be derived by applying nonlinear methods for the analysis of time series of logistic quantities observed in sophisticated models [6–8]. In this study, we use event-discrete simulation to evaluate potential sources of internal instability in small-scale manufacturing networks and the corresponding robustness of these systems. In Sect. 2, we describe the basic features and variants of our simulation model. Different sources of instability are analysed in Sect. 3 by applying nonlinear methods of time series analysis. Finally, we discuss the potential relevance of our findings for the stabilisation of real-world manufacturing networks.

2 Model Description

The systems investigated in this study consist of two to four collaborating manufacturers (nodes) which are firmly connected by pre-defined supply lines. For the topology of the corresponding network, we have distinguished two limiting cases (see Fig. 1): In a system of *N symmetrically interacting manufacturers*, every node delivers one respective component to every other collaborator and one final product to an external costumer. In addition, every node has one external supplier, such that there are *N* buffers for different commodities and *N* production lines. Topologically, the corresponding system forms a circular alignment of nodes with an all-to-all coupling. In addition, we also study *linear supply chains* where each company produces only two products, of which one is supplied to the next company, whereas the other one is delivered to an external costumer. Hence, in this linear alignment, every node has only 2 buffers for semi-finished material and 2 production lines, independent of the total number of nodes, which corresponds to a nearest-neighbour coupling. In our future research, we also plan to extent our investigations to combinations of both settings, which is probably the most realistic case.

Our models have been studied by means of event-discrete simulations based on a commercial simulation software [6–8], which means that all production and delivery events are time-discrete. The dynamics at each node is controlled by a sub-model whose structures are assumed to be identical. All buffers of a manufacturer have an unlimited size and can be assessed by every production line. The manufacturing process of a certain product starts only when the appropriate commodities are present in sufficient quantities. For the realistic illustration of a production procedure, operating times and working-on-intervals have been implemented. This setup allows an explicit computation of the time required for completing a product, and to prohibit or prevent a multiple treatment on the production lines by the definition of working-on-intervals. By prescribing different combination of lot sizes, it is possible to consider different scenarios corresponding to different capacities of the transport containers. Different transport routes with different lengths can be modelled by prescribing certain transportation times. For the accurate control of the supply of semi-finished material from outside the network, all mentioned parameters can also be defined for the external suppliers.

In order to systematically investigate the effects of different production strategies, order policies, network topologies, and process variables, several model variants have been considered for different sets of model parameters:

Fixed and stochastic simulation periods: In the standard case of deterministic (fixed) simulation periods, all time scales of the model have been prescribed to constant values. For the sake of simplicity, operating times, transportation times and working-on-intervals have been fixed multiples of this basic time scale. As an alter-

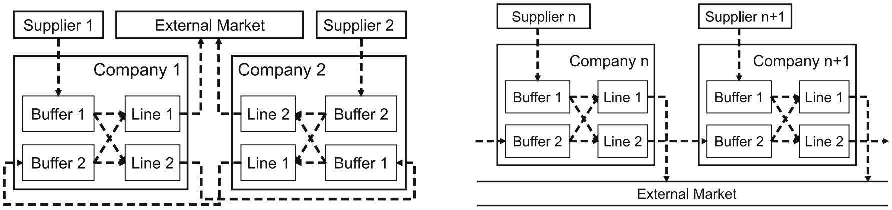


Fig. 1 Visualisation of the symmetrically coupled network with two nodes (*left*) and a linear supply chain (*right*). The direction of material flows is shown by *dashed* lines

native, we have considered the case where all these time scales vary randomly with a Gaussian distribution. The average values have been fixed to the former deterministic clock rate, while the corresponding standard deviation serves as an additional parameter of the model which can be varied over a suitable range.

Production strategies: In the case of a *pull strategy*, each manufacturer produces firstly the product for the external market. If the available amount of semi-finished material is not sufficient, orders of an appropriate size are sent to the corresponding supplier, which are processed in the next period. If the different stock values are sufficiently large, the semi-finished material for the network-internal collaborators is additionally manufactured. Otherwise, corresponding orders are initialised again. As an alternative, a *push strategy* may be used where each manufacturer produces both the final product as well as the semi-finished material in dependence on the respective buffer contents without further orders from the recipients.

Order policies: In addition to the production strategy, all manufacturers may influence the performance of the entire network by their specific order policy. For this purpose, the available stocks of semi-finished materials are compared with a minimum value in pre-defined time intervals. In the case of *provision policy (PP)* and *order point policy (OPP)*, a fixed amount of material is ordered if the corresponding buffer is empty (PP) or has less than a minimum inventory level (OPP). As an alternative, a *periodic order policy (POP)* may be applied where all orders take place periodically, but with a variable size which is adjusted to the actual buffer contents and demand.

3 Classification and Quantification of Instabilities

For the analysis of our simulations, we have focussed our attention to time series of inventory levels in the different buffers, which may also be used to derive different other logistically meaningful quantities. The specific features of this kind of data are a special challenge for time series analysis: Firstly, in microscopic models like the one used in our study, order sizes as well as buffer contents have discrete, non-negative values, whereas most methods of time series analysis have been developed for application to continuous data. Hence, using these concepts for the analysis of the stocks may systematically lead to erroneous results. Secondly, event-discrete data may have an uneven sampling, which causes additional problems.

Concerning the last point, one may interpolate the stock time series to an even sampling, which is easily done due to the event-discrete nature of the model [6]. Alternatively, one may completely neglect the uneven sampling and assume a given basic clock [7]. Comparisons have shown that the corresponding results do practically not differ from each other, such that the second approach has been used for further analysis. In contrast to this, the problem of discrete-valued time series can only be solved by a sophisticated choice of the methods used for their analysis, for example, symbolic time series analysis [6–8] or recurrence quantification analysis [9].

In general, the dynamic vulnerability of material flows in complex networks increases in the presence of feedback loops [10]. Therefore, we will mainly concentrate our investigations to the case of symmetrically interacting manufacturers, as in this case, there is a back-reaction between the different collaborators which is not present in linear supply chains. Our analysis reveals the following factors which significantly increase the dynamic complexity of manufacturing networks:

Fluctuating Production Intervals: If the flow of material in the network is balanced in terms of lot sizes and intrinsic time scales, the system is stable if the basic time intervals are constant. In the case of a push control, the corresponding dynamics of the buffer levels is remarkably stronger than for a pull control. In contrast to this, in a realistic manufacturing system, internal stochastic influences are always present and change the actual length of production and/or transportation times. Such influences are found to have a destructive effect on the logistic performance of a manufacturing network by leading to intermittent time intervals of accumulation of material or unprocessed orders [6,7].

A systematic investigation of the resulting probability distributions (see Fig. 2) and symbolic autocorrelation functions for different buffers shows that this kind of behaviour sets in instantaneously. Moreover, the corresponding effect does not depend on the magnitude of stochasticity over a certain range of moderate standard deviations. Only if these standard deviations are significantly larger than 10% of the mean production intervals, significant differences are observed.

Whereas the accumulation effect is rather moderate and almost equally distributed over all production lines in the case of a pull strategy, under a push control, those lines producing exclusively for external consumers show low and stable stocks, while the lines producing semi-finished material which remains in the network have stocks which increase much stronger than in the case of a pull control.

Imbalance of Material In- and Outflows: In the presence of a push control and large working-on-intervals, the stocks of semi-finished material may increase above

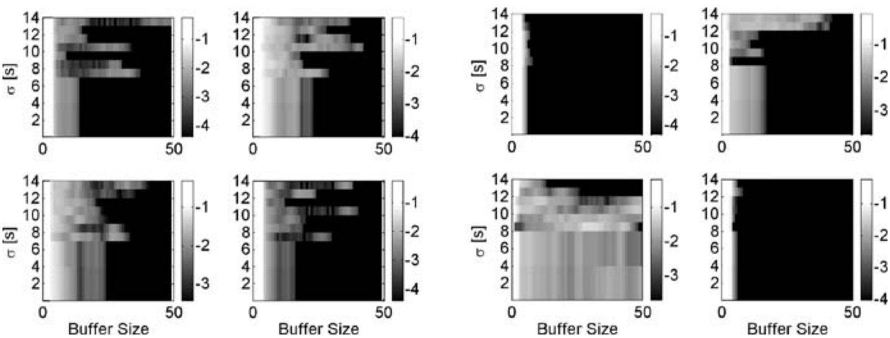


Fig. 2 Estimated probability distribution functions of inventory levels (values in powers of 10) in dependence on the standard deviation of processing times σ (basic clock: 60 s) for the symmetrically coupled two-manufacturer system with all time scales being set to 60 s and all lot sizes to 1, for a pull (*left*) and a push control (*right*). The upper panels correspond to the production lines of the first manufacturer, the lower panels to those of the second one

all limits. In contrast, in the case of a pull control, this effect is absent as the balance between material supply and production is provided here by appropriate order volumes.

In a linear supply chain, the described instabilty can easily occur as there is no back-flow of goods between the respective collaborators. However, it is also found in the case of a symmetric interaction. For example, the two-node system becomes super-critical if for two of the four production lines, these intervals increase the processing times at least by factors of 3 and 2, respectively. Under these conditions, three different cases may be distinguished: (i) If both production lines responsible for network-internal supply are super-critical, both lines producing for the external market become unstable. (ii) If the working-on-intervals of the external line of manufacturer A are larger than those of the internal line of manufacturer B, the internal line of manufacturer A becomes unstable. (iii) If the internal lines are stable at manufacturer A and unstable at manufacturer B, the external line of manufacturer B becomes also unstable.

If in the case of an unstable system, stochasticity is added to the model in terms of distributed production intervals, the rate of material accumulation surprisingly decreases. The strength of this decrease depends on the respective parameters and may lead to an almost complete suppression in case of only weakly unstable systems. Such a stabilisation of the dynamics by application of noise has already been observed in different nonlinear systems, but not yet been found in manufacturing networks. However, the particular amplitude of stochasticity has almost no influence on this process, supporting the results described above.

Incomplete Logistic Synchronisation: In the case of a pull control, an improper choice of lot sizes necessary for production increases the complexity of material flows in the network in a way that the strict periodicity of the production process in the case of fixed production intervals is lost. However, in contrast to the accumulation instability described above, the stock values remain bounded on rather low levels. The results of symbolic time series analysis [6] suggest that the change in the corresponding dynamics is associated with additional frequencies, resembling the transition from fully periodic over quasi-periodic towards chaotic states in nonlinear systems. More detailed investigations show that there are different competing effects contributing to this type of instability. For example, for some settings with three different lot sizes, the corresponding instability is very weak, while for other cases with only two different values within the network, the dynamics is very complex. Interestingly, different odd values of lot sizes (> 1) lead to a much more irregular behaviour (with auto-correlation functions showing hardly any regularity) than even ones of a comparable value.

Order policies: The application of suitable order policies often allows to avoid production breakdowns due to a lack of material. However, this comes on the cost of a logistic desynchronisation as unlike for a usual pull strategy, the sizes and timings of orders are not determined by the current, but the *anticipated* demand. Consequently, the use of order policies results in a strong increase of complexity and a corresponding decrease of the regularity of the dynamics, which leads to a non-periodic behaviour even in the case of equal production and transportation times. In a previous study of the behaviour of the symmetrically coupled four-manufacturer network [8], this has already been demonstrated in some detail. In particular, the behaviour under a periodic order policy shows clearly weaker periodic contributions than for provision and order point policies because in this case, a fixed period of orders is not necessarily present.

4 Conclusions

We have demonstrated that apart from random external influences, at least three different types of internal factors may lead to a very complex dynamics of small-scale manufacturing networks: (i) stochastic internal influences, which occur instantaneously independent of the detailed process parameters and lead to an intermittent accumulation of material or unprocessed orders, (ii) dynamic transitions due to an improper logistic synchronisation (which can be caused by lot sizes or order policies), and (iii) instabilities due to an imbalance of material flows. Type (ii) occurs only for pull or related strategies where the different nodes control *inward* material flows by setting orders of appropriate size. In contrast to this, instabilities of type (iii) occur only for push strategies, i.e., when the nodes control only the *outward* material flows.

Our findings allow to draw some general conclusions about the behaviour of manufacturing systems which may help to develop appropriate control strategies for real-world systems. In particular, the occurrence of different types of instability demonstrates that the active influence of separate "agents" which act such that their individual logistic performance is optimised is not necessarily optimal for the behaviour of the overall network. This observation is especially reflected by the dynamics in the case of improper logistic synchronisation or imbalances of the material flows. As a necessary consequence, the implementation of a fully decentralized control in a manufacturing network may fail as long as the different collaborators do not take care of the states of *all* production lines. Possible solutions of this problem may include the (phase) synchronisation of stock oscillations at the different production lines [11] as well as the application of centralised or external control mechanisms in terms of concepts like the Fourth Party Logistics [12].

A very interesting finding of our study is that in the case of stochastic components in the dynamics, the resulting changes do almost not depend on the magnitude of stochasticity. Moreover, stochasticity is found to strongly decrease the undesired effects of imbalances of material flows. Further studies of time-continuous production models are necessary to further verify that the corresponding results are actually no effects of the discrete-event simulation. In case of their significance, it may be possible to explicitly use stochasticity to improve the logistic performance of generically unstable manufacturing systems, which cannot be stabilised by appropriate control mechanisms or structural changes within the production network.

Acknowledgements This work has been supported by the Volkswagen foundation (project no. I/78218) and the German Research Foundation (project no. He 2789/8-1).

References

1. Sterman JD (2000) Business Dynamics. McGraw-Hill, Boston
2. Hopp WJ, Spearman ML (2000) Factory Physics. McGraw-Hill, Boston
3. Daganzo C (2005) A Theory of Supply Chains. Springer, Berlin
4. Bloomberg DJ, LeMay SB, Hanna JB (2002) Logistics. Prentice Hall, New York
5. Radons G, Neugebauer R (eds) (2004) Nonlinear Dynamics of Production Systems. Wiley, New York
6. Scholz-Reiter B, Hinrichs U, Donner R, Witt A (2006) Modelling of networks of production and logistics and analysis of their nonlinear dynamics. In: Wamkeue R (ed) Modelling and Simulation. IASTED, Montréal, pp 178–183
7. Donner R, Hinrichs U, Scholz-Reiter B, Witt A (2006) Nonlinear dynamics and control of small-scale manufacturing networks. In: Marquié P (ed) Nonlinear Dynamics in Electronic Systems (NDES2006). Dijon, pp 22–25
8. Scholz-Reiter B, Hinrichs U, Donner R (2007) Application of methods of nonlinear dynamics for the control of networks of production and logistics. In: Proceedings of the 2^{nd} International Conference on Changeable, Agile, Reconfigurable and Virtual Production (CARV 2007). Toronto, pp 263–272
9. Marwan N, Romano MC, Thiel M, Kurths J (2007) Recurrence plots for the analysis of complex systems. Phys Rep 438: 237–329
10. Helbing D, Lämmer S, Witt U, Brenner T (2004) Network-induced oscillatory behaviour in material flow networks and irregular business cycles. Phys Rev E 70: 056118
11. Scholz-Reiter B, Tervo JT, Freitag M (2006) Phase-synchronization in continuous flow models of production networks. Physica A 363: 32–38
12. Reindl M, Oberniedermeier G (2002) eLogistics. Addison-Wesley, Boston

Decentralized Decision-making in Supply Chains

Aspects of Agent Based Planning in the Demand Driven Railcab Scenario

Wilhelm Dangelmaier, Benjamin Klöpper, Nando Rüngerer, Mark Aufenanger

Heinz Nixdorf Institute, University of Paderborn,
Fürstenallee 11, 33102 Paderborn, Germany
e-mail: {whd,kloepper,mark.aufenanger,nando.ruengerer}@hni.upb.de

Abstract Multi-agent Systems are a promising approach to implement flexible and effective decentralized control mechanismen for the logistic domain. This paper intends to introduce the key features of a multi-agent system that is under development for several years. The agents autonomously plan, optimize, and control a railway transportation system.
For this purpose the agents interact, communicate and negotiate. This paper presents an overview how different techniques from operations research, artificial intelligence and soft computing are integrated into a multi-agent system to solve this difficult and complex real life problem.

1 Introduction

This paper presents a multi-agent system (MAS) controlling and optimizing the processes in a railway network. It is under development for several years and was constantly extended. The agents are designed to plan, control and optimize the processes of shuttles in the Railcab scenario. Railcab is an innovative logistic concept, which tries to combine the effectiveness of railway traffic with the flexibility and individuality of road traffic. Therefore transport disposition in the Railcab concept is demand driven, no fixed schedule exits. If a passenger wishes to travel from Paderborn to Munich (about 540 km), he advises the customer interaction module of the departing place, the destination station, the wanted time and shuttle type (business, tourist class etc.) and the number of travelling passengers. The customer interaction module provides the customer with various offers from the provided resources (tracks, wagons, station capacity) and a price finding procedure. Finally, the passenger decides on an offer and books bindingly. A shuttle will be available at the given time at the train station Paderborn. With a small number of stops to pick up and disembark other passengers, the shuttle carries the passenger to Munich. A short travel time is not inevitably reached through a higher maximum speed, but through a higher average speed. By reducing the number of stops compared to conventional railway

H.-D. Haasis et al., *Dynamics in Logistics*
DOI: 10.1007/978-3-540-76862-3,

systems, travel times are decreased in the Railcab concept. This and the flexibility of the individual transport units build the foundation for an attractive train system. The lower maximum speed and the therewith related relatively lower energy consumption and few empty drives, lead in total to an economically attractive system. For a detailed description of the logistic concept see [1].

The scheduling of such a railway system is a complex and dynamic problem. The system is subject to frequent changes (e.g. new transport orders and disturbances) and encompasses a large number of objects (e.g. shuttles, tracks, stations, switches, goods). For these reasons centralized planning and coordination algorithms seem not reasonable. In centralized planning for distributed execution an effective and efficient communication infrastructure is of crucial importance [2]. The communication in the Railcab system necessarily relies on wireless communication, which cannot provide a sufficiently reliable and fast communication for a centralized planner. Thus we decided to implement a MAS in order to plan, control, and optimize the Railcab logistic system. The remainder of this paper will outline the development of this MAS in historical order. Each of the subsequent section will describe a step in the evolution of the MAS and tackle with a particular problem. The next section briefly describes the logistic problem solved by the MAS.

2 Problem Description

The planning problem defined by the Railcab scenario is a online version of the *Pick Up And Delivery Problem With Time Windows* (PDPWTW). Informaly this means a number of n passengers or goods have to be pick-up at a pick-up station p within a time window from *earliest departure time* until *latest departure time* and delivered to a delivery station within a time window from *earliest arrival time* until *latest arrival time*. New transport jobs occur constantly. The solution of this problem requires special properties due to the fact that Railcab is a rail bound system. In particular the utilization of tracks and the capacities of stations have to be considered. Furthermore the solution should be optimal or (due to the complexity of the problem) at least good with respect to some cost functions:

$$cost_i = \sum_{t=1}^{n} c_t + e_t \cdot ec_t + \sum_{st=1}^{m} c_{st}\left(departure_{st} - arrival_{st}\right) \tag{1}$$

$$cost_g = \sum_{sh=1}^{o} cost_{sh}\,. \tag{2}$$

Equation (1) shows a simplified version of the individual cost of a shuttle. It sums the cost of travelling on tracks and stopping at stations. c_t denotes the toll for travelling on track and c_{st} the cost per time unit stopping at station. e_t denotes a factor for energy consumption, which can be reduced by the formation of convoys (due to the slip stream effect) and ec_t denotes the normalized energy consumption for travelling on track t. These individual cost functions are subsumed for evaluation issue to a global cost function overall shuttles shown in Eq. (2).

3 Asynchronous Coordination and Synchronous Optimization

This section describes the first version of the MAS. It tried to solve two important sub problems in the Railcab scenario: first the provision of reliable and fast answers to transport orders on the one hand, the requirement of cost effective transport plan on the other hand. The MAS meets these two requirements by an implementation of the cooperative planning paradigm of generalized partial global planning (GPGP) [3]. In GPGP each agent constructs an individual plan in order to achieve it's goals. In order to improve the coordination in the system agents can exchange their individual plans. An agent that has knowledge of more than one individual plan knows a so called partial global plan and can use this information to coordinate the individual plans by altering them. GPGP cannot assure to resolve all conflicts between local plans, e.g. resource conflicts. Resolving this kind of conflict is of crucial importance for safety and reliability in general in railway systems. Franke developed an agent based transport planning system, which performs the coordination and planning process in two phases: the initial planning and the optimization. The main task of the initial planning is to provide a feasible transport plan under consideration of all resources in the railway network, especially the tracks and stations. To solve the coordination problem an agent is assigned to every resource in the network. For example a track agent has to decide if a shuttle can travel on the

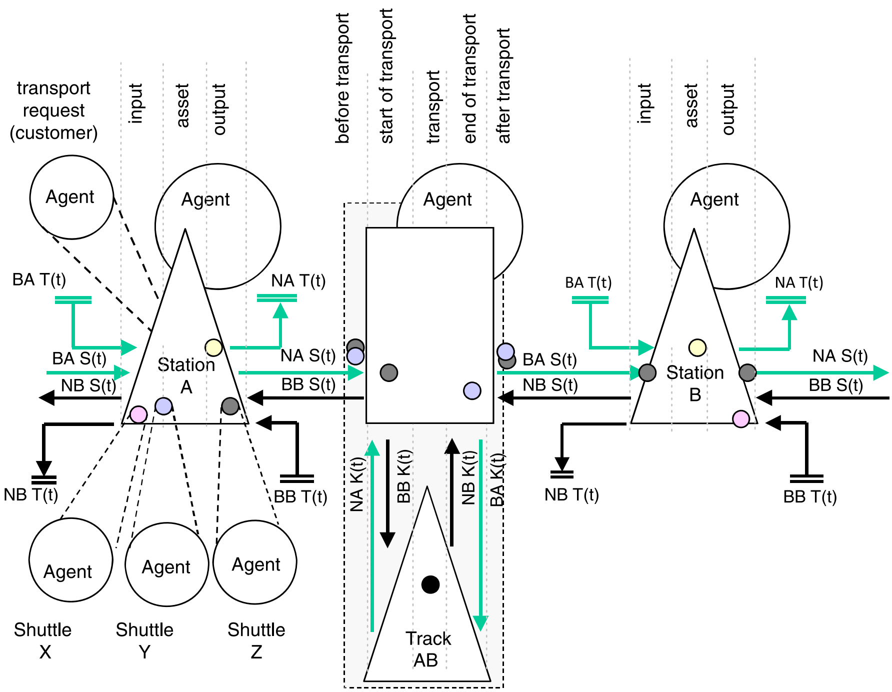


Fig. 1 MFERT Agent Model

corresponding track section at a particular time in a particular direction. The relationships and interactions in the MAS are systematized by a state transition net profile called MFERT (see Fig. 1), which is a domain specific profile for production planning and logistics. In order to avoid long blocking and waiting times in the MAS and to increase the performance of the initial planning, the agents communicate in an asynchronous way. (for details see [4] and [5]).

In the second phase the transport plans from the initial planning are optimized. It starts with an aggregation of all individual plans into a single global plan. In a parallel heuristic procedure (taboo search) the initial plan is optimized. The heuristic procedure is performed by a number of synchronously communicating agents, which synchronize their plan in fixed time steps. The parallel optimization procedure was able to realise a decrease of cost of 30% compared to the initial solution. After synchronization, the currently best plan is used by all agents for further optimization. Both phases are performed alternating. New transport orders can be buffered and are handled by the asynchronous communicating agents after the phase (for details see [6] and [7]).

4 Decentralized Optimization

The approach presented in the prior section has one critical shortcoming: the synchronous optimization procedure is a centralized approach with all disadvantages mentioned in the introduction. Furthermore the complete exchange of information is not possible if the agents perform the planning for different principals.

This is the case when competing companies use the same railway network. Thus we were looking for a more decentralized way of optimization. There are two possible coordination actions between agents which can lead to a cost decrease: formation of convoys and swapping of transport jobs. For both optimizing strategies, in scenarios with self interested agents the actual realisation of a cooperation is subject of negotiations. The negotiation may encompass issues like compensation payments, detours, and timing. An overview of possible negotiation techniques is given in [15].

4.1 Decentralized Swapping of Jobs

The next step in the evolution of the MAS was the implementation of the jobswapping. In a decentralized job swapping each shuttle constantly analyzes it's local plan. In this local analyzes searches for jobs that cause inappropriate costs, e.g. such jobs that cause large detours for picking up and/or delivering a single passenger. This kind of job is advertised for bids of other shuttles. The idea is to find a shuttle the travels along the pick up and deliverance point anyway. The problem of bringing cooperation partners together is considered as the matchmaking problem [12]. There are several methods to solve the matchmaking problem: the contract net protocol (CNP) [11], middle agents and brokers [12] and unstructured matchmaking [10]. In the Railcab MAS we used a combination of the contract net protocol and middle agents to solve the matchmaking problem. The amount of communication required for the CNP is determined by the ability of the manager agent (the agent who wants to delegate a job) to identify good candidates for taking over the job. Thus distributed blackboards (a special kind of middle agents) are used to find shuttles which are interested in taking over an offered job. Blackboards are assigned to station agents, since pick-up and delivery stations are the most promising places to look for shuttles interested in the job. The rest of the agent interaction is organized according to the CNP. The decentralized implementation of the job swapping was able reduce the cost of track utilization by 15% in average during simulation experiments. This result is remarkable compared to the 30% cost reduction by the centralized optimization. Details about job swapping can be found in [8].

4.2 Decentralized Convoy Formation

The implementation of the second coordination action is currently under development. The formation of convoys requires a more detailed coordination of local plans. Actually a nesting of a new (partial) route plan into the local plans of the shuttles is required. Since this nesting is very complex and the number of considered partners has strong impact to the complexity, a more sophisticated method for the identification of partners is used. The matchmaking – again initiated by a shuttle agent that discovers the potential of optimization – encompasses three steps: search for possible partners, filtering of partners and clustering of partners. The search process will be implemented by a distributed depth-first-search along the route of the shuttle. The depth of the search is limited by a parameter d. The search process will be actually performed by the station agents and is only triggered by shuttle agents. The shuttle agents query the station along their routes. Each station, which has a depth smaller than d performs a look-up for passing shuttles in its schedule and propagates a query to its neighbour station. A station with a depth of d performs the look-up but does not propagate the query to its neighbours. Equation 3 shows how the depth of a station is calculated.

$$depth(st) = \begin{cases} 0, & \textit{if st is queried by shuttle agent} \\ d(st') + 1, & \textit{if st is queried by station } st' \end{cases} \quad (3)$$

In the next step the resulting set of shuttles is filtered for improper candidates. Ephrati and Pollack demonstrated that filtering improves the efficiency in multi-agent planning [13]. Filter criteria are arrival and departure times and the direction of travel. After filtering a hierarchical clustering [14] is used to identify groups of shuttles which may constitute a good convoy. This grouping is based on the similarity of shuttles. Shuttles which have similar routes (especially similar pick-up and delivery stations), similar travel speed and similar breaking distances can form an effective convoy.

5 Consideration of Uncertain Travel Times

Like in many approaches, known from operations research, the shuttle agents consider travel times on the tracks in the transportation network to be constant. In fact, travel times are never constant, but depend on a large number of factors. For instance weather (ice or leaves on the tracks) or the amount of traffic may influence the travel times. The paradigm of Bayesian thinking teaches us that we can estimate variables like travel time by partial knowledge of the influences and their conditional probability distribution. Thus we extended the MAS by the ability of reasoning about uncertain travel times. Expert agents were introduced into the system which have knowledge about the probabilistic nature of travel times in the railway network. This knowledge is represented by hybrid Bayesian Networks. A Bayesian Network is a directed, acyclic graph (G,V), where the nodes V represent discrete or continuous variables. A directed edge from node X_i to node X_j denotes that variable X_j depends on X_i. This dependency is quantified by the conditional probability distribution $P(X_i|Parents(X_i))$ for each node in the network [21]. Bayesian Networks with both discrete and continuous variables are called hybrid networks. According to Jensen [17], Gaussian distributions are the only applicable continuous probability distributions. Other distributions cannot be used for the lack of mathematical concepts. Figure 2 shows two examples of such expert networks. For inference in this network a modification version of Pearl's Message Passing [20] algorithm was implemented. The modification enables the inference in complex networks with polynomial runtime.

Shuttle agents use Dynamic Bayesian Networks (DBN) [16] to model their shortest path problem. A DBN is a graphical representation of a first order Markov Process. The DBN of the shortest path problem is constructed by querying the expert agents. Figure 2 illustrates the interaction between shuttle and expert agents.

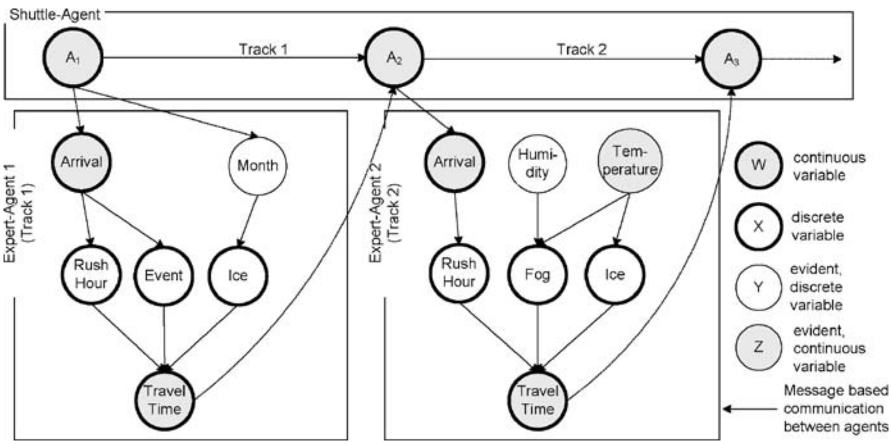


Fig. 2 Interaction of Agents

In simulations we recognized a more accurate estimation of travel times by the probabilistic approach. Furthermore, the new approach resulted in shorter travel times. A detailled discussion of the planning procedure and the results can be found in [18]. In [19] the MAS was extended by an advanced consideration of risk. On basis of the probability distributions provided by the expert agents, the likelihood of delays is included in the shortest path problem. In simulations, the number of delays was reduced by almost 10% compared to the probabilistic approach without consideration of risk.

6 Conclusion

This paper presented a MAS for planning, controlling and optimizing a rail bound transportation systems. Different aspects of demand driven railway concepts are reflected in the development process of the MAS. These different aspects are the safety critical coordination of shuttles, the demand for efficient transport plan and potentially competing forwarding companies in the same transportation network.

The safety critical requirements are met by formal modeling of the resources and interactions in the transportation system. A state-transition net profile is used for implicit synchronization of the local shuttle plans. Furthermore methods for centralized and decentralized optimization of transportation plan were introduced. The decentralized optimization is suitable for scenarios with competing forwarding companies, when a complete share of information is not conceivable. Finally a new concept for path planning in dynamic and probabilistic environment was introduced. During simulation this concept proved to be useful to provide better estimation of travel times and due able to find paths with shorter travel times, e.g. by driving around traffic jams. Further research will focus on this direction. The consideration of risk is subject of further improvement and a new method for the construction of conditional plans on basis of the probabilistic planning is currently under development.

References

1. Dangelmaier, W., Fahrentholz, M., Franke, H., Mueck, B.: A demand-driven logistics concept for the fully automated rail system NBP. In: World Congress on Railroad Research, Köln, 2001
2. Durfee, E.H.: Distributed problem solving and planning. In: Lecture Notes In Computer Science, Vol. 2086, pp. 118–149, Springer Verlag, Berlin, Heidelberg, 2001
3. Decker, K.S., Lesser, V.R.: Designing a family of coordination algorithm. In: Proceedings of the Thirteenth International Workshop on Distributed Artificial Intelligence (DAI-94), pp. 65–84, 1994
4. Dangelmaier, W., Franke, H.: : Decentralized management for transportationlogistics: A multi agent based approach. Integrated Computer-Aided Engineering, 10(2), pp. 203–210, 2003

5. Dangelmaier, W., Franke, H., Klöpper, B., Scheideler, P.: Modelling and controlling decentralized logistics networks. In: APPLIED SIMULATION AND MODELLING (ASM04), pp. 17–26 2004
6. Dangelmaier, W., Franke, H., Scheideler, P.: A Parallel Procedure for a Distributed Logistics Scheduling. In: Proceedings of Industrial Simulation Conference 2003 (ISC-2003), pp. 233–237, 9.–11. Jun. 2003
7. Dangelmaier, W., Franke, H., Klöpper, B., Koesters, C.: Synchronously communicating agents for parallel improvements in transport logistics. In: Artificial Intelligence and Applications AIA, pp. 277–286, Innsbruck, Austria, 2004
8. Danne, C., Dück, V., Klöpper, B.: Selfish Motivated Cooperative Planning in Cross Networked Mechatronic System. International Transactions on Systems Science and Applications, 2(2), pp. 221–225, 2006
9. Sycara, K., Decker, K., Williamson, M.: Middle Agents for the Internet. In: Proceedings of IJCAI-97,pp. 578–583, 1997
10. Ogston, E., Vassiliadis, S.: Unstructured Agent Matchmaking: experiments in timing and fuzzy matching. In: SAC'02: Proceedings of the 2002 ACM Symposium on Applied Computing, pp. 300–305, New York, NY, USA, ACM Press
11. Zhou, P.-C., Hong, B.-R., Wang, Y.-H., Zhou, T.: Multi-Agent Cooperative Pursuit Based on Extended Contract Net Protocol, Proceedings of the Third International Conference on Machine Learning and Cybernetics, pp. 169–173, Shanghai, 2004
12. Jang, M.-Y., Momen, A.A., Agha, G.: ATSpace: A Middle Agent to Support Application Oriented Matchmaking and Brokering Services. In: IEEE/WIC/ACM IAT(Intelligent Agent Technology)-2004, pp. 393–396, 2004
13. Ephrati, E., Pollack, M.E., Ur, S.: Deriving Mulit-Agent Coordination through Filtering Strategies. In: 14th International Joint Conference on Artificial Intelligence, pp. 679–685, 1995
14. Jain, A.K., Murty, M.N., Flynn, P.J.: Data Clustering: A Review. In: ACM Computer Surveys, Vol. 31(3), pp. 264–323, 1999
15. Fatima, S., Wooldridge, M., Jennings, N.R.: Bargaining with Incomplete Information. Annals of Mathematics and Artificial Intelligence, Vol. 44, pp. 207–232, 2005
16. Murphy, K.P.: Dynamic Bayesian Networks: Representation, Inference and Learning. Doctoral Thesis. UMI Order Number: AAI3082340
17. Jensen, F.V.: Bayesian Networks and Decision Graphs. Springer-Verlag, New York, 2001
18. Dangelmaier, W., Klöpper, B., Wienstroer, J., Timm, T.: The Shortest Path Problem in Uncertain Domains - an Agentbased Approach with Bayesian Networks. In: Sixth International Conference Intelligent Systems Design and Applications, Vol. 2, pp. 943–948, IEEE Computer Society Press. 2006
19. Dangelmaier, W., Klöpper, B., Wienstroer, J., Döring, A.: Risk Averse Shortest Path Planning in Uncertain Domains. In: International Conference on Computational Intelligence for Modelling, Control & Automation 2006, IEEE Computer Society Press, pp. 115–121, 2006
20. Pearl, J: Probabilistic Reasoning in Intelligent Systems – Networks of Plausible Inference. Morgan Kaufmann, 1988
21. Charniak, E: Bayesian Networks without Tears. AI MAGAZINE, pp. 50–63, 1991

Merging Time of Random Mobile Agents

Shehla Abbas, Mohamed Mosbah, Akka Zemmari

LaBRI, University of Bordeaux 1,
351 cours de la Libération, 33405 Talence-France
email: {abbas,mosbah,zemmari}@labri.fr

Abstract In this paper, we investigate the following problem: k mobile agents are placed on a graph G. Each agent performs a random walk, and when two or more agents meet at a node, they merge into a single agent. We present a Markov chain modeling the agents behaviour and show how this can be used to upper bound the expected time for all the k agents to merge into a single agent.

1 Introduction

In today's world, immense importance has been given to distributed computing and algorithms to give the maximum throughput with minimum resources. Mobile agents are one of the ways to implement these computations and algorithms. In recent years emphasis has also been shifting from deterministic algorithms to randomized algorithms which are often more simple and more efficient. So we are also interested in randomized algorithms and solutions for the displacements of agents in the network. These algorithms require less energy and few resources to perform any delegated task.

The Model. We consider a *synchronous distributed* network of processors with an arbitrary topology. It is represented as a connected graph where vertices represent processors, and two vertices are connected by an edge if the corresponding processors have a direct communication link.

A *mobile agent* (MA) is an entity which moves in the graph. There are many models to study the movement of the agents depending on the parameters of interest and on the applications. In this paper, we consider MAs with identities whose storage memory can not deal with topological information of the network. This is useful for instance if the network is huge (e.g. Internet). We study in this work agents that do not store information about the visited vertices. One interesting model that we deal with is that of random walks. Hence, our MA moves randomly in the graph. The MA performs a *random walk*: when it is located in a vertex v, if $B(v,1)$ denotes the ball of center v and radius 1, the MA chooses *uniformly at random* one of the

H.-D. Haasis et al., *Dynamics in Logistics*

DOI: 10.1007/978-3-540-76862-3,

vertices in $B(v,1)$, if the chosen vertex is v, the MA does not move, else, it moves to the chosen vertex.

The Problem. In this paper, we are interested in the following problem: a set of $k \geq 2$ mobile agents are placed arbitrarly on the network. These agents can work in collaboration by performing random walk and can give efficient results by the combined effort (for example in applications such as collecting information, spanning tree or gathering problem), however, the agents are reduced each time the agents meet on any node v of the netowrk. Upon meeting the agents can share their information they already acquired by doing some computations. In this case, the agent with the smaller identity disappers itself from the nework. At the end of our algorithm all agents merge into a *single* MA having the biggest identity.

Another objective of this paper is to calculate the expected time for all the agents to meet in a network hence analyzing the complexity of the algorithm. We make another assumption that a MA remains on the vertex where it is located or moves to one of the vertices in the neighbourhood. Indeed, without this hypothesis, even if we use random algorithm to break symmetry, the *rendezvous* can never occur. To see that, consider the case of the complete graph K_2 with two vertices and two MAs one on each vertex. Then if we do not allow the MAs to remain on a vertex, they will never meet.

Related Works. The calculation of the time for two agents to meet in a graph can be considered as the *rendezvous problem*. A lot of work has been done on the rendezvous problem and vast literature can be found in (Alpern and Gal, 2002). Rendezvous on graphs has been presented in (Alpern et al., 1999; Alpern, 1995). Some papers also discuss the probabilistic solutions like in (Alpern, 1995, 2002; Anderson and Weber, 1990; Baston and Gal, 1998). Gathering is another technique (Flocchini et al., 2001; Lim and Alpern, 1996; Thomas, 1992) when more than two agents have to meet at one fixed location. Random walk is a useful tool for many applications such as the construction of a spanning tree by multiple mobile agents (Abbas et al., 2006). It has many other applications and it can also be considered as an Election Problem (Bar-Ilan and Zernick, 1989). In another work of agent rendezvous by (Yu and Yung, 1996), the authors considered their agents (or distributed units as they call it) to be intelligent agents having computing power, control and memory that is having decision capability which helps in understanding the topology of the network. The agents are coordinated of their behaviour in a distributed environment. The meeting time of agents for the complete graph is $O((n/k)\log n)$ with high probability and for hypercube, it is $O((n/k)\log^2 n)$ with high probility, however their model is different from ours in which agents do not have control, intelligence and memory. A survey on mobile agent rendezvous can be found in (Kranakis et al., 2006). In this survey, authors have narrated various results on the mobile agent rendezvous on distributed networks having deterministic and randomized rendezvous among agents; the agents having computing power, control and memory and the number of agents vary from 2 to k. To the best of our knowledge, the present paper is the first to consider the *randomized* meeting of any k agents without memory and the reduction of the number of agents to one.

Our Contribution. In this paper, we first give a Markov chain formulation of the problem, then we investigate some particular cases and analyze the merging time in complete graphs, star graphs and the hypercube.

1.1 An Introductory Example

As a first investigation, consider a graph $G = (V, E)$ with a star topology. Consider two agents in this graph. Depending on the placement of the two agents, if we consider the random variable $(X_t)_{t\geq 0}$, such that X_t is the distance between the two agents at time t, then $\forall t \geq 0$, $X_t \in \{0,1,2\}$ and its behaviour can be modelled by a Markov chain $\mathscr{M} = (\{0,1,2\}, \mathscr{T})$, where $\mathscr{T}$ is defined by the following transition matrix:

$$\mathscr{P} = \left(\begin{array}{c|ccc} & 0 & 1 & 2 \\ \hline 0 & 1 & 0 & 0 \\ 1 & \frac{1}{n} & \frac{1}{2} & \frac{n-2}{2n} \\ 2 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{array}\right).$$

This chain is an absorbing one, the single absorbing state is 0 and its *fundamental matrix* (see (Kemeny and Snell, 1960)), can be written as follows

$$\mathscr{N} = \left(\begin{array}{c|cc} & 1 & 2 \\ \hline 1 & \frac{6n}{n+4} & \frac{4(n-2)}{n+4} \\ 2 & \frac{4n}{(n+4)} & \frac{4n}{n+4} \end{array}\right).$$

Hence, if we denote by T_i the time for the two agents to obtain a rendezvous, given that they are initially placed at distance i ($i \in \{1,2\}$) one from the other, we have

$$\mathbb{E}(T_1) = \frac{10n-8}{n+4}$$

and

$$\mathbb{E}(T_2) = \frac{8n}{n+4}.$$

Hence, if we denote by T the time for the two agents to obtain a rendezvous, we have, $\mathbb{E}(T) = O(1)$.

If we consider $k \geq 3$ MAs in the star graph, the parameter of interest will be the number of MAs in the graph and the behaviour of this number. Then, instead of the distance between MAs, we consider the r.v. $(X_t)_{t\geq 0}$ which counts the number of MAs in the network at time t. We then start by the following lemma:

Lemma 1. *Let X_t be the number of agents on the leaves at time $t \geq 0$. Then*

$$\mathbb{E}\left(X_{t+1} \mid X_t\right) \leq \frac{X_t}{2} + 1.$$

Proof. To prove the lemma, we have to consider two cases:

1. There are 0 MAs at the center. Let $\mathbb{E}_0(X_{t+1} \mid X_t)$ denotes the expected number of agents on the leaves at time $t+1$ knowing that there were X_t agents on the leaves

at time t, then, for any $i \in \{0, 1, \dots, X_t\}$, if i MAs move to the center of the star, then the number of MAs on the leaves decrease by i, hence, a combinatorial and probabilistic reasoning yields to

$$\mathbb{E}_0(X_{t+1} \mid X_t) = \sum_{i=0}^{X_t} \left(\frac{1}{2}\right)^{X_t} \binom{X_t}{i} (X_t - i) = \frac{X_t}{2}.$$

2. There is an MA at the center. Let $\mathbb{E}_1(X_{t+1} \mid X_t)$ denotes the expected number of agents on the leaves, a similar reasoning as in 1) yields to

$$\mathbb{E}_1(X_{t+1} \mid X_t) = \left(1 - \frac{1}{n}\right)\frac{X_t}{2} + \left(1 - \frac{1}{n}\right).$$

The lemma follows. □

Then we have the following theorem.

Theorem 1. *If we denote by T_k the time for k MAs to merge into a single MA, then*

$$\mathbb{E}(T_k) = O(\log(k)).$$

Proof. From the formulae in Lemma 1, we can see that the number of times we can apply the recursion is given by $\log(X_0)$. Since $X_0 \le k$, the theorem follows. □

2 A Genral Markov Chain Formulation

2.1 Configurations Graph

The first formulation of the problem is inspired from (Bshouty et al., 1999). As in (Bshouty et al., 1999), a *configuration* of k agents on the graph G is a multiset of k vertices of G. A configuration $\{x_1, x_2, \cdots, x_k\}$ describes a placement of k agents on the vertices of G, where x_i is the *position* of agent a_i. If $x_i = x_j$ then the agents a_i and a_j are merged into one agent: a_i if $a_i > a_j$ and a_j otherwise. The process starts in an *initial configuration* $v_0 = \{x_1, x_2, \cdots, x_k\}$ with $x_i \ne x_j, \forall i \ne j$ and ends with a *terminal configuration* $\{x_1, x_2, \cdots, x_k\}$ with $x_i = x_j, \forall i \ne j$. If $v_p = \{x_1, x_2, \cdots, x_k\}$ is the configuration of the process at step p then a *next configuration* $v_{p+1} = \{y_1, y_2, \cdots, y_k\}$ of v_p satisfies the following conditions:

1. if $y_i \ne x_i$ then $y_i \in N(x_i)$ where $N(x_i)$ denotes the set of neighbours of x_i,
2. for all $i \ne j$, if $x_i = x_j$ then $y_i = y_j$

An *execution* starting at initial configuration v_0 is a finite sequence of configurations $\{v_0, v_1, \dots, v_m\}$ such that for all $p = 0, 1 \cdots, m-1$, v_{i+1} is a next configuration of v_i. Configuration v_m is called the *current configuration* at time step m. $v_m = \{x_1, x_2, \cdots, x_k\}$ is a *terminal* configuration if $x_i = x_j$ for all $i \ne j$, this configuration is noted *TERM*.

As in (Bshouty et al., 1999), the *configuration graph* $G_k = (V(G_k), E(G_k))$ is the directed graph defined as follows: $V(G_k)$ is the set of all the possible configurations and

$$E(G_k) = \Big\{(v_1, v_2) \mid v_1 = \{x_1, x_2, \cdots, x_k\} \neq TERM \text{ and}$$

$$v_2 = \{y_1, y_2, \cdots, y_k\} \text{ is a next configuration of } v_1\Big\} \bigcup \{(TERM, TERM)\}$$

Since G is connected, it is clear that G_k is connected. The behaviour of the set of agents $\{a_1, a_2, \cdots, a_k\}$ in the undirected graph G is then the same as the behaviour of a random walk on the directed configuration graph G_k. This random walk is a Markov chain with one absorbing state *TERM*, the other states are all transient.

Since G_k is connected, it is clear that, with probabilty 1, the random walk in G_k will reach the state *TERM*, meaning that all the MAs will end by merging on a single MA.

In this study, we are interested on the expected time before all the agents merge into one agent. This can be interpreted as the expected time for a random walk starting in an initial configuration to reach the state *TERM* in G_k. In terms of random walks, this corresponds to the notion of *hitting time* $H(v_0, TERM)$.

2.2 Components Graph

Let G be a connected graph. As we have seen in the previous section, the behaviour of k MAs on G is the same as the bahaviour of a random walk on the configurations graph G_k. Nevertheless, this random walk, even if it is a Markov chain, seems not easy to study and do not provide a simple method to upper bound the merging time of k MAs on G.

Let v_i be any configuration, we define the *size* of v_i, and we denote it $|v_i|$, the number of different vertices in v_i. We define the *components graph* $\mathscr{G}_k = (\mathscr{V}_k, \mathscr{E}_k)$ as follows:

- $\mathscr{V}_k = \{C_1, C_2, \cdots, C_k\}$ such that $\forall i \in \{1, 2, \cdots, k\}$, $C_i = \{v \in V_k \,|\, |v| = i\}$
- $\mathscr{E}_k = \{(C_i, C_j) \mid \forall v_1 \in C_i, \exists v_2 \in C_j$ such that $(v_1, v_2) \in E_k\}$

The main idea of this construction is to gather, on a single component, all the configurations of the same size and there is a transition from a component C to a component C' iff for any configuration v_1 in C, there is at least a configuration v_2 in C' such that $(v_1, v_2) \in E_k$.

Let $(Y_t)_{t \geq 0}$ be a random walk on G_k. For any $(i,j) \in \{1, 2, \cdots, k\} \times \{1, 2, \cdots, k\}$, we define $p_{i,j}$ as follows:

$$p_{i,j} = \begin{cases} \min\{\mathbb{P}r(Y_{t+1} = v_2 \mid Y_t = v_1) \mid (v_1, v_2) \in C_i \times C_j\} & \text{if } i > j \\ 1 - \sum_{l<i} p_{i,l} & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$$

Then, consider the following process: $X_0 = C_k$ and for any $t \geq 0$,

$$\mathbb{P}r(X_{t+1} = C_j \mid X_t = C_i) = \begin{cases} p_{i,j} & \text{if } i > j \\ 1 - \sum_{l<i} p_{i,l} & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$$

We have the following lemma:

Lemma 2. *If $\mathscr{G}_k$ is connected then*

$$H(v_0, TERM) \leq H(C_k, C_1), \textit{for any initial configuration } v_0.$$

Proof. By induction on k. □

2.3 From 2 to k Agents

In this section we show how the study of the particular case of two agents in a graph $G = (V, E)$ allows us to obtain an upper bound for the general case of $k \geq 3$ MAs in G.

Let $G = (V, E)$ be a given graph of size n and $k \geq 3$ be the number of the MAs in G. Let $\mathscr{G}_k$ the corresponding components graph. In this section, we assume that $\mathscr{G}_k$ is connected. Consider a random walk in $\mathscr{G}_k$, and let $(X_t)_{t\geq 1}$ be the r.v. defined as follows: $X_0 = k$ and $\forall t > 0$, X_t is the size of the component at time t.

The behaviour of the r.v. $(X_t)_{t\geq 0}$ can be modelled by a Markov chain $\mathscr{M} = (\mathscr{S}, \mathscr{T})$ such that $\mathscr{S} = \{1, \cdots, k\}$ and

$$\mathbb{P}r(X_{t+1} = 1 \mid X_t = 1) = 1$$
$$\mathbb{P}r(X_{t+1} = j \mid X_t = i) = \begin{cases} p_{i,j} > 0 & \text{if } j \leq i \\ 0 & \text{otherwise.} \end{cases}$$

$\mathscr{M}$ is an absorbing Markov chain with a single absorbing state (the state 1) with the following transition matrix:

$$P = \left(\begin{array}{c|cccccc} & 0 & 1 & 2 & \cdots & k-1 & k \\ \hline 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 1 & p_{2,1} & p_{2,2} & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \cdots & \cdots \\ k & p_{k,1} & p_{k,2} & p_{k,3} & \cdots & p_{k,k-1} & p_{k,k} \end{array} \right).$$

Let $(Y_t)_{t\geq 0}$ be the r.v. defined as follows: $Y_t \in \mathscr{S}$, $\forall t \geq 0$ and

$$\mathbb{P}r(Y_{t+1} = 1 \mid Y_t = 1) = 1$$
$$\mathbb{P}r(Y_{t+1} = j \mid Y_t = i) = \begin{cases} q_{i,j} = p_{i,j} & \text{if } j = i \\ q_{i,j} = \sum_{l=i-1}^{1} p_{i,l} & \text{if } j = i-1 \\ 0 & \text{otherwise.} \end{cases}$$

This corresponds to an absorbing Markov chain $\mathscr{M}' = (\mathscr{S}, \mathscr{T}')$ with the following transition matrix:

$$P' = \left(\begin{array}{c|ccccccc} & 1 & 2 & 3 & 4 & \cdots & k-1 & k \\ \hline 1 & 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 2 & q_{2,1} & q_{2,2} & 0 & 0 & \cdots & 0 & 0 \\ 3 & 0 & q_{3,2} & q_{3,3} & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & \ddots & \ddots & \ddots & \vdots & \vdots \\ k & 0 & 0 & 0 & \cdots & 0 & q_{k,k-1} & q_{k,k} \end{array} \right).$$

$\mathscr{M}'$ is a Markov chain whose set of states is the same as for $\mathscr{M}$ and the only possible transitions are between i and $i-1$. This intuitively means that whereas in $\mathscr{M}$, from a state i it is possible to reach all the states $i-1, i-2, \cdots, 1$, one can reach only the state $i-1$ in $\mathscr{M}'$. It is clear that:

Lemma 3. *For any $i \in \{1, 2, \cdots k\}$, if we denote by T_i (resp. T'_i) the time for the Markov chain $\mathscr{M}$ (resp. $\mathscr{M}'$), starting at the state i to reach the state* 0*, then*

$$\mathbb{E}(T_i) \leq \mathbb{E}(T'_i).$$

Proof. From the transition matrices P and P', we deduce the following recursions: $\forall i \geq 2$, $\mathbb{E}(T_i) = 1 + p_{i,i}\mathbb{E}(T_i) + \sum_{j=1}^{i-1} p_{i,j}\mathbb{E}(T_j)$ and $\mathbb{E}(T'_i) = 1 + p_{i,i}\mathbb{E}(T'_i) + (\sum_{j=1}^{i-1} p_{i,j})\mathbb{E}(T_j)$. Hence, an induction reasoning yields to $\mathbb{E}(T_i) \leq \mathbb{E}(T'_i), \forall i \geq 2$. □

Application: The Complete Graph K_n

Let $K_n = (V, E)$ be the complete graph of size n. We start by studying the case of $k = 2$ MAs. If we denote by D the distance between the two MAs in K_n, it is clear that $D \in \{0, 1\}$. Then, if T is the r.v, which counts the number of steps before the two MAs meet, T is a r.v. with a geometric distribution of the parameter

$$p = \mathbb{P}r(D = 1) = \frac{1}{n},$$

hence, for any $m \geq 1$,

$$\mathbb{P}r(T = m) = p(1-p)^{m-1} = \frac{1}{n}\left(1 - \frac{1}{n}\right)^{m-1}.$$

Yielding to the expected number of steps before the two MAs meet:

$$\mathbb{E}(T) = \frac{1}{p} = n.$$

Now, let $k \geq 3$ be the number of agents in the graph K_n. From Lemma 3, it suffices to study the Markov chain $\mathscr{M}'$. To do so, we have to calculate the transition probabilities $q_{i,i}$ and $q_{i,i-1}$, $\forall i \in \{2, 3, \cdots, k\}$.

Lemma 4. *Let $\mathscr{M}'$ be the Markov chain defined above. Then we have the following for any $i \in \{2, 3, \cdots, k\}$:*

$$q_{i,i} = \frac{(n-1)(n-2)\cdots(n-i)}{(n-1)^i} \tag{1}$$

and $q_{i,i-1} = 1 - q_{i,i}$.

This yields to:

Lemma 5. *If we denote by T the time for all the k MAs to merge in a single agent, then*

$$\mathbb{E}(T) = O(n \log k).$$

Proof. Once $(n - j) \leq (n-1)$, $\forall j \geq 1$, we obtain from Eq. (1) that $q_{i,i} \leq \frac{n-i}{n-1}$ for any $i \in \{2, 3, \cdots, k\}$. On the other side, we have

$$\mathbb{E}(T) = \mathbb{E}(T'_k) = \sum_{i=2}^{k} \frac{1}{q_{i,i-1}}.$$

Hence,

$$\mathbb{E}(T) \leq \sum_{i=2}^{k} \frac{n-1}{n-i} = (n-1)H_k,$$

where $H_k = 1 + \frac{1}{2} + \cdots + \frac{1}{k}$ is the harmonic number. Since $H_k \sim \log(k)$ the lemma follows. □

3 Hypercubes

An hypercube of dimension d is a graph $\mathscr{H}_d = (V, E)$ such that $V = \{\omega \in \{0,1\}^d\}$ and $E = \{(\omega_1, \omega_2) \in V \times V \mid h(\omega_1, \omega_2) = 1\}$ where $\forall(\omega_1, \omega_2) \in V \times V$, $h(\omega_1, \omega_2)$ denotes the *Hamming distance* between ω_1 and ω_2, i.e., the number of bits to change in ω_1 to obtain ω_2. That is, this graph has some nice properties: each vertex has a degree d, the maximum distance between any two vertices in the graph is d and the size of the graph is $n = 2^d$.

We first investigate the case of two MAs in $\mathscr{H}_d$. As for the previous cases, we consider the random variable $(X_t)_{t \geq 0}$ such that $\forall t \geq 0$, X_t corresponds to the distance between the two MAs at time t. The behaviour of this r.v. can be modelled by a Markov chain $\mathscr{M} = (\mathscr{S}, \mathscr{T})$ where $\mathscr{S} = \{0, 1, \cdots, d\}$ and the transitions are defined as follows:

$$Pr(X_{t+1} = 0 \mid X_t = 0) = 1$$

$$Pr(X_{t+1} = j \mid X_t = i) = \begin{cases} \frac{d+2di-2i^2+1}{(d+1)^2} & \text{if } j = i \\ \frac{2i}{(d+1)^2} & \text{if } j = i-1 \\ \frac{2(d-i)}{(d+1)^2} & \text{if } j = i+1 \\ \frac{i(i-1)}{(d+1)^2} & \text{if } j = i-2 \\ \frac{(d-i)(d-i-1)}{(d+1)^2} & \text{if } j = i+2 \\ 0 & \text{otherwise.} \end{cases} \quad \forall i \geq 0, j \geq 1,$$

That is, $\mathscr{M}$ is an absorbing Markov chain with a single absorbing state: the state 0.

Remark 1. If the distance between the two MAs is close to n, there are more chances that the distance between them decreases. If the distance between them is close to 0, which is an absorbing state, there are more chances that the distance increases. Similarly, if the distance is in the middle then there are more chances that it remains the same.

The last remark makes the study of the above chain more complicated than the previous ones. Indeed, in this case, the transition probabilities are not constant but are given by means of d and the actual value of the distance between the two MAs. Hence, we use another method to study this process: we decompose the chain $\mathscr{M}$ into two Markov chains $\mathscr{M}_1$ and $\mathscr{M}_2$, each of them with the following transition matrix:

$$A_1 = A_2 = \left(\begin{array}{c|cccccccc} & 0 & 1 & 2 & 3 & 4 & 5 & \cdots & d \\ \hline 0 & \frac{1}{d+1} & \frac{d}{d+1} & 0 & 0 & 0 & 0 & \dots & 0 \\ 1 & \frac{1}{d+1} & \frac{1}{d+1} & \frac{d-1}{d+1} & 0 & 0 & 0 & \dots & 0 \\ 2 & 0 & \frac{2}{d+1} & \frac{1}{d+1} & \frac{d-2}{d+1} & 0 & 0 & \dots & 0 \\ 3 & 0 & 0 & \frac{3}{d+1} & \frac{1}{d+1} & \frac{d-3}{d+1} & 0 & \dots & 0 \\ \vdots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ d & 0 & 0 & 0 & 0 & 0 & \dots & \frac{d}{d+1} & \frac{1}{d+1} \end{array} \right)$$

If we denote by $Q = (q_{i,j})_{0 \leq i,j \leq d}$ the matrix $Q = A_1 \times A_2$ and by P the transition matrix of the chain $\mathscr{M}$ then it is easy to verify that

$$p_{i,j} = q_{i,j}, \ \forall (i,j) \in \{1, 2, \cdots, d\} \times \{1, 2, \cdots, d\}. \tag{2}$$

The formulae (2) can be interpreted as follows. There is a transition, in the process modelled by $\mathscr{M}$, from a state i to a state j if and only if there is a transition, in the process modelled by $\mathscr{M}_1$, from i to a state $l \in \{0, 1, 2, \cdots, d\}$ and a transition, in the process modelled by $\mathscr{M}_2$, from l to j.

Then, rather than studying the chain $\mathscr{M}$, we first investigate the chain $\mathscr{M}_1$. The parameter of interest will be the expected value of the r.v. which counts the number of transitions for the process modelled by $\mathscr{M}_1$ to reach the state 0. Indeed, when this occurs, this means that the two MAs are at distance at most 1.

$\mathscr{M}_1$ is a regular Markov chain which admits a stationary distribution $\pi = \pi_{i(0\leq i\leq d)}$, that is, a vector of values in $[0,1]$ such that $\pi A_1 = \pi$ and $\sum_{i=1}^{d} \pi_i = 1$. This yields to a system which we can solve to obtain

$$\forall i \in [0,d],\ \pi_i = \frac{\binom{d}{i}}{n}.$$

Let T_i denote the time for the process $\mathscr{M}_1$ being at the state i to reach the state 0, $(T_i)_{1\leq i\leq d}$. The expected value of T_i verifies the following system:

$$\begin{cases} \mathbb{E}(T_0) = \frac{1}{\pi_0} = n \\ \mathbb{E}(T_{i+1}) - \mathbb{E}(T_i) \leq \frac{1}{d-i} n,\ \forall\, 1 \leq i < d\,. \end{cases}$$

Summing the left and the right sides, we obtain

$$\begin{aligned} \mathbb{E}(T_i) - \mathbb{E}(T_0) &\leq \left(\sum_{j=0}^{i} \frac{1}{d-j}\right) n \\ &\leq \left(\sum_{j=0}^{i} \frac{1}{i-j}\right) n \\ &= H_i n, \end{aligned}$$

where for any $i \geq 1$, H_i is the harmonic number. Hence,

$$\mathbb{E}(T_i) \leq H_i n + n.$$

Yielding to

$$\mathbb{E}(T_i) \sim n \log i.$$

Since $i \leq d$, we obtain the fundamental result:

Theorem 2. *With* 2 *MAs in the hypercube* $H_d = (V,E)$*, if we denote by* T *the time for the two MAs to obtain a rendezvous, then*

$$\mathbb{E}(T) = O(n \log\log n).$$

This permits to obtain the general result:

Theorem 3. *With* k *MAs in an hypercube* $\mathscr{H}_d$*, if* T *denotes the time for* all *the* k *MAs to merge in one agent, then*

$$\mathbb{E}(T) = O(kn \log\log n).$$

Proof. To upper bound $\mathbb{E}(T)$, we use the Lemma 3. A probabilistic reasoning yields to the fact that $q_{i,i-1} \geq q_{2,1}, \forall i \geq 2$, which allows us to obtain an upper bound for $\mathbb{E}(T')$:

$$\mathbb{E}(T') \leq \sum_{i=2} k \frac{1}{q_{2,1}},$$

using the result of Theorem 2, we obtain $\mathbb{E}(T) = O(kn \log\log n)$. □

Remark 2. In the introduction we said that this problem models the election problem in agents based system. This can also be used to solve a more general problem: the election problem in distributed system. Indeed, it suffices that each vertex in the graph generates a MA with the vertex identity and then the vertex which has generated the MA which is eventually elected in the n agents is the elected vertex. The complexity of the resulting algorithm is obtained by substituting n to k in the formulae.

4 Conclusion and Perspectives

In our algorithms, we calculated the time it takes for the agents to meet in a network. Starting with k agents, each one of them collect information from the nodes of the network. By working in parallel to visit nodes and by gathering their information, the global algorithm is more efficient than if it was executed by a single agent.

We calculated the complexities for different classes of graphs with two agents and then we extended it to k agents. We start our problem with star topology which takes the constant time $O(1)$ for the two agents to meet. Extending it to k agents gives the result of $O(\log k)$. In the case of a complete graph, the time it takes by two agents is linear with n. Similarly for k agents, the calculation yields $O(n \log k)$. Then we further studied the case of the hypercube; with two agents it gives the time of $O(n \log\log n)$. In case of k agents, we get the time of $O(kn \log\log n)$. This shows that it is better and more efficient than using one agent in the network.

This problem can be further extended to other kind of graphs such as d-regular graphs. We are also working on implementing these algorithms on a software tool to obtain the experimental results.

Acknowledgements This work has been supported by the French project "Agence Nationale de la Recherche ANR-06-SETI-015". Authors are very grateful to Nasser Saheb for helpful discussion.

References

Abbas S, Mosbah M, Zemmari A (2006) Distributed computation of a spanning tree in a dynamic graph by mobile agents. In: Proceedings of IEEE 1st International Conference on Engineering of Intelligent Systems, pp 425–430

Abbas S, Mosbah M, Zemmari A (2007) Collecte d'informations par des agents mobiles. In: Proceedings of New Technologies for Distributed Systems (NOTERE'2007), HERMES

Alpern J, Baston V, Essegaier S (1999) Rendezvous search on a graph. Journal of Applied Probablity 3(1):223–231

Alpern S (1995) The rendezvous search problem. SIAM J on Control and Optimization 33(3):673–683

Alpern S (2002) Rendezvous search on labelled networks. Naval Research Logistics 49:256–274

Alpern S, Gal S (2002) The thoery of search games and rendezvous. International Series in Operations Research and Management Science, number 55. Kluwer Academic Publishers

Anderson E, Weber R (1990) The rendezvous problem on discrete locations. Journal of Applied Probablity 28:839–851

Bar-Ilan J, Zernick D (1989) Random leaders and random spanning trees. In: Proceedings of the 3rd IWDAG

Baston V, Gal S (1998) Rendezvous on the line when the players' initial distance is given by an unkown probability distribution. SIAM J on Control and Optimization 36:1880–1889

Bshouty NH, Higham L, Warpechowska-Gruca J (1999) Meeting times of random walks on graphs. Information Processing Letters 69(5):259–265

Flocchini P, Prencipe G, Santoro N, Widmayer P (2001) Gathering of asynchronous oblivious robots with limited visbility. In: Proc. 18th Annual Symposium on Theoratical Aspects of Computer Science, LNCS 2010, pp 247–258

Kemeny, Snell (1960) Finite Markov Chains. D. Van Nostrand

Kranakis E, Krizanc D, Rajsbaum S (2006) Mobile agent rendezvous: A survey. In: SIROCCO, LNCS 4056, pp 1–9

Lim W, Alpern S (1996) Minimax rendezvous on the line. SIAM J on Control and Optimization 34(5):1650–1665

Thomas L (1992) Finding your kids when they are lost. Journal on Operational Reseach 43(5):637–639

Yu X, Yung M (1996) Agent rendezvous: A dyanmic symmetry-breaking problem. In: 23rd International Colloquium on Automata, Languages and Programming (ICALP), LNCS 1099, pp 610–621

Dynamic Decision Making on Embedded Platforms in Transport Logistics – A Case Study

Reiner Jedermann[1], **Luis Javier Antúnez Congil**[1], **Martin Lorenz**[2], **Jan D. Gehrke**[2], **Walter Lang**[1], **Otthein Herzog**[2]

[1] Institute for Microsensors, -Actuators and -Systems (IMSAS), University of Bremen, Germany, email: {rjedermann,ljacongil,wlang}@imsas.uni-bremen.de
[2] Center for Computing Technologies (TZI), University of Bremen, Germany, email: {mlo,jgehrke,herzog}@tzi.de

Abstract Autonomous logistic processes aim at coping with logistic dynamics and complexity by local decision making to gain flexibility and robustness. This paper discusses resource-bounded logistics decision making using software agents and task decomposition. Simulations show the feasibility of dynamic vehicle routing and quality monitoring on embedded systems for the transport of perishable goods.

1 Introduction

Within the last years logistics has become a key success factor in globally distributed production because of its cross-sectional function. But enhanced product life cycles, rapid changes in company structures and information flows change the requirements for logistic processes. On the one hand the rising complexity of inter-organizational structures and also a relative shortage of logistic infrastructure lead to increasing utilization of the existing logistic processes. On the other hand a specialization of the ways of transportation and the carriers, which are connected to the transported goods can be observed. These factors combined with changing customer market conditions have considerable effects on planning and controlling logistic processes in such a dynamic environment. A possible approach to face these challenges is the dezentralization of logistic control and coordination by entities with the ability for autonomous decision making—in other words: autonomous logistic processes.

Especially the management of supply chains for agricultural products has to cope with multiple dynamic factors. Increasing and varying consumer expectations meet hardly predictable harvest quality and amount, which in turn puts high demands on supply chain management in the face of uncertainty (Scheer, 2006). Large inventories within the chain are generally to be avoided.

The quality of perishable products like fruits, meat, fish, flowers and pharmaceuticals is a dynamic factor by itself, depending on harvest/production conditions, time and stress factors during transport and storage.

DOI: 10.1007/978-3-540-76862-3,

In the future "quality oriented tracking and tracing system[s]" (Scheer, 2006) will offer new approaches to cool chain management. Stock rotation could be based on current quality instead of fixed 'sell by' dates. Warehouse planning will be organized by 'First *expires* first out' (FEFO) instead of 'First in first out' (Emond and Nicometo, 2006). Cool chain management will not only reduce unnecessary buffers in inventory but also in shelf life. The dynamic shelf life expresses the remaining time span until the product quality falls below an acceptance limit. Individual items with high reserves in shelf life will be sent exactly *where* they are needed, e.g., international deliveries or retail shops with lower turn over.

The main methodology used throughout our work is based on autonomous logistic processes modelled by software agents. Timm (Timm, 2006) defines four levels of autonomy and investigates the possibilities of their respective realization in software agents. Whereas "strong regulation" is associated with classical software engineering the autonomy on levels one and two, "operational autonomy" and "tactical autonomy", can be realized with known decision making methods especially the *Belief – Desire – Intention* (BDI) approach (Rao and Georgeff, 1995). A concept for realizing "strategic autonomy" within a software agent is sketched in (Timm, 2006).

In the context of resource-bounded computation another aspect has to be added to the question of autonomy. An agent's autonomy is limited by the computational resources that are available for its deliberation. In other words, while *operational autonomy* can be realized by a relatively simple model-based reflex agent (Russell and Norvig, 2003, chap. 2) *tactical autonomy* already needs planning and deliberation skills for which current resource-bounded environments are not suitable.

Therefore this paper will show the tradeoff between computational needs of autonomous logistic processes on the one hand and capabilities of current embedded systems on the other hand. Furthermore we show the current state of a prototype implementation using off-the-shelf embedded hardware to demonstrate the opportunities of intelligent local decision making for transportation of perishable goods.

2 Autonomous Decision Making in Transport Logistics

Experiences regarding perishable goods show that loss of quality depends on its location inside the container and on the environment in general (Tanner and Amos, 2003a), without forgetting unexpected problems. Therefore a prediction of actual quality losses in this dynamic environment is a hard problem. In order to enable real-time reaction quality tracing must happen locally and individually in the container.

The intelligent container we have presented in (Jedermann et al., 2006) provides a tracing system that consists of four components: wireless sensor network (WSN), RFID system, CPU and communication system. The WSN provides information about the environment, e.g., temperature and humidity. The RFID system is responsible for the identification of goods. The next step is the extension of this tracing system towards an autonomous decision making system, thus goods will be able to change routing decisions of the carrier if necessary for their goal achievement.

This monitoring and autonomous decision making system (MADMS) should support the following requirements. The MADMS must trace the goods in an autonomous way. It has to react to changes in the environment and the goods' shelf life. The MADMS should be able to communicate with a route planner and must be able, considering these aspects, to make autonomous decisions to adapt the route to the current shelf life state.

The MADMS can be modelled as a multiagent system, using the deliberative agent technology, i.e., the BDI architecture. The BDI architecture is based on practical reasoning and consists of two processes: deliberation and means-ends reasoning. A BDI agent has information about its environment, its goals and ways to achieve them (Wooldridge, 1999).

We consider one BDI agent that is responsible for the goods in a shipping container. The agent's desire is to deliver all goods without losses. The agent's beliefs are: the goods' respective shelf-life, destination and delivery time as well as information about the local environment provided by the WSN and other information systems, e.g., the route planner. The intentions determine the agent's planned actions to achieve its current goals.

A BDI agent is able to make autonomous decisions, for example if a truck cannot connect with the route planner system and it does not have information about the next destination. An embedded system has bounded resources, e.g., memory or CPU, compared to desktop PCs. The challenge is the further development of a MADMS for embedded systems using the BDI architecture, where the agents must deliberate with only scarce resources.

3 Implementation in Embedded Systems

Quality-oriented tracking and tracing (t&t) for chilled transport and storage is primarily based on permanent temperature supervision. Literature reports (Rodríguez-Bernejo et al., 2006; Punt and Huysamer, 2005; Tanner and Amos, 2003b; Moureh and Flick, 2004) and preliminary tests (Jedermann and Lang, 2007) revealed spatial deviations of 5 °C or more over the length of a container or temperature zone in trucks with separated compartments. Simulations with shelf life models showed that these temperature differences could lead to variances in the remaining quality of more than 50 % at the end of transport (Jedermann and Lang, 2007). In cool down to deep freezer mode, deviations of up to 10 °C were still present after 5 hours.

Evaluation of sensor data by current t&t systems is centralized. But due to limited communication bandwidth and the cost of mobile or satellite communication it is not feasible to transfer the vast information created by spatial temperature supervision to a remote server for evaluation. Therefore we shift part of the logistic decision system into the means of transport. Equipped with an embedded processing unit, the truck or container can act as an autonomous entity. In Sect. 4.1 we show by example how an embedded unit can also handle parts of the route planning. As test platform we selected a credit card sized ARM XScale processor module with a clock rate of 400 MHz and 32 MBytes SDRAM memory (`www.dilnetpc.com`).

3.1 Representation of Logistical Objects by Software Agents

In the context of our work the implementation paradigm for autonomous processes is based on software agents. Agents communicate by asynchronous messages independently of their current location to achieve their goals (Wooldridge and Jennings, 1995). Currently the most common environment to test and implement agents is the Java Agent DEvelopment framework (JADE) (Bellifemine et al., 2001).

A special version of JADE, the Light Extensible Agent Platform (LEAP), was developed for devices with limited computing power and memory, e.g., mobile phones (Moreno et al., 2007). The programming language Java is required by the agent framework. Although Java is not a common language to operate microcontrollers, it offers useful advantages for our setting like platform independency and the execution of dynamic code. New programming environments like the Jamaica Virtual Machine (`www.aicas.com`) make it feasible to efficiently run complex Java programs on embedded systems (Siebert, 2002).

As example process we investigated the transfer of a software agent representing the freight-specific supervision instructions. After the loading of a freight item to a new means of transport is detected by a RFID reader, the vehicle sends a request for the corresponding agent. After being transferred to the vehicle the agent is restarted on the local platform provided by the truck or container. Finally a confirmation message is sent to the previous location of the freight.

By optimizations of the agent framework the execution time for the above described process has been reduced from 15 to 6 seconds (Jedermann and Lang, 2006). Major bottlenecks were the translation of messages into the FIPA-ACL format and JADE internal services.

3.2 Interpretation of Sensor Data and Quality Assessment

A wireless sensor system for spatial temperature supervision is provided by the means of transport. The complexity of fruit ripening processes or quality decay of other products can not be reduced to a simple temperature threshold checking. Based on the Arrhenius law on reaction kinetics the freight supervision agent calculates the current loss in shelf per time unit, depending on the measured temperature (Jedermann et al., 2006). The feasibility of this concept for automated quality evaluation by mobile freight agents was demonstrated by our reduced scale prototype of an intelligent container (Jedermann et al., 2007).

4 Distributed Solution of Route Planning Problems

Planning for a logistic domain is a real-time problem with relaxed timing constraints of hours or even days rather than seconds. Although deliberating can be done concurrently to acting it is constrained by onboard computational resources, which may be very limited regarding computing power and memory (see Sect. 3).

Classical planning based on theorem proving dates back to the 1960's. The widely known STRIPS system was introduced in (Fikes and Nilsson, 1971). STRIPS-like planners can cope with "fully observable, deterministic, finite, static, and discrete" environments (Pollack, 1992). Even though numerous approaches exist that address one or the other restriction, none of them can remove all of them.

Decision-theoretic approaches, especially Markov Decision Processes, add the ability of handling uncertainty, observations, the concept of utility, and also partial observability (Boutilier et al., 1999; Wagner et al., 2006). Several approaches and systems have been proposed and implemented to address the problem of timing constraints but still include major shortcomings. A discussion of these approaches and a potential solution for the logistics domain can be found in (Lorenz et al., 2007).

For evaluation purposes in this paper the planning problem is reduced to a simple route planning algorithm which is guaranteed to find the optimal solution at the cost of hyperexponential computation time.

4.1 Distributed Planning by Truck Agents

In order to illustrate how distributed problem solving could shift part of the route planning into the local processing platform of the vehicle we use the following example:

> A truck is loaded with a number of perishable products that should be delivered to multiple customers in different destinations. The products have different initial shelf lives, which are reduced by journey times. The task is to optimize the route in a way that product losses are avoided due to zero shelf life at delivery. The sum of the remaining shelf lives at time of delivery should be maximized. In an extended scenario unexpected shelf life losses by temperature changes force re-planning of the route. A solution for this problem has to consider two sources of information: First the current shelf life state of the loaded products; these data are directly available inside the vehicle. And secondly information about travel distances and traffic situation. These data would be provided by a remote traffic data base.

The above example describes a special case of a Traveling Salesman Problem. To solve this kind of problem we apply a heuristic approach by splitting the route search between remote traffic data base and local vehicle. This distributed problem solving results in increased robustness and autonomy. A remote Route Planning Agent (RPA) searches for routes with low total driving time to the delivery sites that have not been visited so far.

The proposals are fetched by the Local Vehicle Agent (LVA), which selects one of the possible routes based on an evaluation of the achievable grade of goal fulfilment, given as the sum of remaining shelf lives at delivery and avoidance of zero shelf life. After arriving at the next customer the LVA requests new route proposals, if the current grade of goal fulfilment is not satisfactory or unexpected quality losses make re-planning necessary.

By this approach the delivery service can quickly react to sudden changes in freight quality. Even if the RPA could not be reached due to communication failure, the vehicle can continue its planning based on the proposals that are already known.

As an additional win this approach also secures information about shelf life and especially about quality problems as internal state because the actual decision parameters remain local to the vehicle.

4.2 Experimental Evaluation

An important requirement regarding the RPA is that the route proposals must be substantially different from each other. If, e.g., only a list of the routes with shortest driving distance is provided, these routes often resample each other and deprive the LVA of the freedom it needs to find optimal paths in terms of shelf life.

The following approach was tested as example: For each step with N customers still to be visited, the LVA requests a proposal for a short, but not necessarily the shortest round trip. The truck could drive from its current position to one of the N destinations as starting point for the round trip and continue in clockwise or counter clockwise direction. The resulting $2 * N$ options are evaluated according to their grade of goal achievement. After arriving at the first customer the procedure is repeated by sending a new request to the RPA. This approach was compared to a full optimization by a software simulation. For a fixed map of twelve destinations, freight items with a shelf life set by random values had to be delivered to the customers.

In two thirds of the simulation runs, the systems finds a route that delivers as much items before expiration as theoretical possible (see Table 1–row A). The higher flexibility of the system and lower planning effort entails a small reduction of the remaining shelf life at delivery to 92% in average of the possible value. To avoid package losses for the remaining third (see Table 1–row B) of experiments, further improvements of the local planning process are necessary. Such a system should detect cases in which the described simple heuristic is not sufficient and switch to a more costly strategy, e.g., requesting additional route proposals.

Table 1 Summary of 600 simulation runs. The points give a measure for the remaining shelf life at delivery. Late deliveries with zero shelf life were punished with -1000 points (row B)

		Runs	Local planning	Optimal	Ratio
A	(no losses)	402	252,73 points	272,02 points	92,62% ± 7,37
B	(with losses)	198	$-813,64$ points	218,93 points	—

5 Conclusion

In this paper we have shown that it is indeed possible to create autonomous logistic decision making systems based on state-of-the-art deliberation methods of distributed artificial intelligence. By skilled division of planning tasks, as shown by our software simulations, it is feasible to implement decision instances onto off-the-shelf embedded hardware. These autonomous systems are capable of on-line monitoring of certain parameters that a logistic service provider would not want to reveal to third parties and base decision on these data by considering information collected from third party providers.

The solution we propose is autonomous on a level we consider clearly above *operational*. Whether it is jusitified to claim *tactical* autonomy has to be shown by further experiments and refinement of the approach. To increase their capabilities the implementation of the JADE agent platform should be extended to a deliberation on a BDI basis.

Acknowledgements This research was supported by the German Research Foundation (DFG) as part of the Collaborative Research Centre 637 "Autonomous Cooperating Logistic Processes".

References

Bellifemine F, Poggi A, Rimassa G (2001) Developing multi-agent systems with a FIPA-compliant agent framework. Software – Practice & Experience 31(2):103–128

Boutilier C, Dean T, Hanks S (1999) Decision-theoretic planning: Structural assumptions and computational leverage. Journal of Artificial Intelligence Research (JAIR) 11:1–94

Emond JP, Nicometo M (2006) Shelf-life prediction and FEFO inventory management with RFID. In: Cool Chain Association Workshop

Fikes RE, Nilsson NJ (1971) STRIPS: A new approach to the application of theorem proving to problem solving. Artificial Intelligence 2(3–4):189–208

Jedermann R, Lang W (2006) Mobile java code for embedded transport monitoring systems. In: Ester R (ed) Proceedings of the Embedded World Conference 2006, Franzis Verlag, pp 771–777

Jedermann R, Lang W (2007) Semi-passive RFID and beyond – steps towards automated quality tracing in the food chain. International Journal of Radio Frequency Identification Technology and Applications (IJRFITA) In press

Jedermann R, Schouten R, Sklorz A, Lang W, van Kooten O (2006) Linking keeping quality models and sensor systems to an autonomous transport supervision system. In: Kreyenschmidt J, Petersen B (eds) Cold Chain Management–2nd international Workshop, Bonn, 8.–9. May 2006, University Bonn, pp 3–18

Jedermann R, Gehrke JD, Becker M, Behrens C, Morales-Kluge E, Herzog O, Lang W (2007) Transport scenario for the intelligent container. In: Hülsmann M, Windt K (eds) Understanding Autonomous Cooperation & Control in Logistics. The impact on Management, Information and Communication and Material Flow, Springer, Berlin, pp 365–404

Lorenz M, Ober-Blöbaum C, Herzog O (2007) Planning for autonomous decision-making in a logistic scenario. In: Proceedings of the 21st European Conference on Modelling and Simulation (ECMS 2007), pp 140–145

Moreno A, Valls A, Viejo A (2007) Using JADE-LEAP to implement agents in mobile devices. http://jade.tilab.com/papers/EXP/02Moreno.pdf, last visit: 2007-03-05

Moureh J, Flick D (2004) Airflow pattern and temperature distribution in a typical refrigerated truck configuration loaded with pallets. International Journal of Refrigeration 27(5):464–474

Pollack ME (1992) The uses of plans. Artificial Intelligence 57(1):43–68

Punt H, Huysamer M (2005) Temperature variances in a 12 m integral reefer container carrying plums under a dual temperature shipping regime. In: International Conference Postharvest Unlimited Downunder 2004, Acta Horticulturae, vol 687, International Society for Horticultural Science, pp 289–296

Rao AS, Georgeff MP (1995) BDI agents: From theory to practice. In: Proceedings of the First Intl. Conference on Multiagent Systems, pp 312–319

Rodríguez-Bernejo J, Barreiro P, Robla JI, Ruíz-García L (2006) Thermal study of a transport container. Journal of Food Enginnering 80(2):517–527

Russell SJ, Norvig P (2003) Artificial Intelligence: A Modern Approach, 2nd edn. Prentice Hall, New Jersey

Scheer P (2006) Optimising supply chains using traceability systems. In: Furness A (ed) Improving traceability in food processing and distribution, Woodhead, Cambridge, UK, pp 52–64

Siebert F (ed) (2002) Hard Realtime Garbage Collection. aicas GmbH

Tanner DJ, Amos ND (2003a) Modelling product quality changes as a result of temperature variability in shipping systems. In: Proceedings of the International Congress of Refrigeration (ICR 2003), International Institute of Refrigeration, Paris, published on CD-ROM

Tanner DJ, Amos ND (2003b) Temperature variability during shipment of fresh produce. In: International Conference: Postharvest Unlimited, Acta Horticulturae, vol 599, International Society for Horticultural Science, pp 193–204

Timm IJ (2006) Strategic management of autonomous software systems. TZI-Bericht 35, Center for Computing Technologies, University of Bremen, Bremen

Wagner T, Raja A, Lesser V (2006) Modeling uncertainty and its implications to sophisticated control in TAEMS agents. Autonomous Agents and Multi-Agent Systems 13(3):235–292, http://mas.cs.umass.edu/paper/425

Wooldridge M (1999) Intelligent agents. In: Weiss G (ed) Multiagent Systems. A Modern Approach to Distributed Artificial Intelligence, The MIT Press, pp 27–77

Wooldridge M, Jennings NR (1995) Intelligent agents: Theory and practice. Knowledge Engineering Review 10(2):115–152

The Global RF Lab Alliance: Research and Applications

The Value of RF Based Information

Dieter Uckelmann

LogDynamics Lab, University of Bremen, Germany

Abstract Today we see a rapid shift of traditional supply chains towards internet-driven supply networks. Radio frequency (RF) based technologies such as RFID and RF based sensors will support this change by combining the physical and virtual world. While the cost for RF infrastructure faces a substantial decrease, the cost of providing and updating product and supply chain relevant data will increase as a consequence of the growing need for information. In many cases the positive contribution of RF based information to this dynamic change will be hindered because of doubtful return on investment (ROI) in internet-driven supply networks.
Based on three scenarios, the paper will show a new concept of billable RF based information, which may be integrated into the concept of the "Internet-of-Things" as it is known today.

1 Introduction

Nowadays logistics is affected by a rapid change of traditional supply chains towards internet-driven supply networks. [1] While traditional supply chains are defined by long-lasting 1-to-1 or n-to-1 relationships the newly formed supply networks are more flexible and are characterized by n-to-n relationships. This trend is accompanied by a similar development within information logistics. Information logistics may be divided into an information flow that is linked to a physical product and the information flow that moves independent of the product through the internet or dedicated networks. Main data carriers in combination with the product are human readable text and barcode, but currently we see more and more use of RFID in this context. RFID-installations are still mostly limited to 1-to-1 or n-to-1 applications, which are mostly pushed through small pilot projects or mandates. Still, we will see more open n-to-n installations, if RFID becomes ubiquitous. Within the second information flow which is not physically linked to the product, fax and paper still are common but there has been a trend towards electronic data interchange (EDI). EDI is used to exchange information across organizational boundaries based on informa-

H.-D. Haasis et al., *Dynamics in Logistics*
DOI: 10.1007/978-3-540-76862-3,

tion technology. The "Internet-of-Things" seems to be the next logical step towards electronic data interchange in n-to-n relationships (Table 1).

The concept of "Internet-of-Things" will be explained in more detail later on. It is based on the idea that every product is uniquely identifiable. Even though identification can be achieved by multiple technologies, the "Internet-of-Things" is always mentioned in relation with RFID. RFID is a highly potential technology, but simultaneous with a few abashments. The technology is on its way to supplement the barcode technology and replace it in certain implementations for pure object identification. The term RFID itself may be misleading and should be replaced by RF based identification, data storage and sensing to encompass the real potential of the technology. Generally, there are two options of providing additional data beyond pure identification. First, data can be stored in databases and forwarded through online messaging infrastructure. The concepts of electronic data interchange (EDI) and the "Internet-of-Things" are well known examples. Second, the memory space of the transponder can be used as a storage device for additional relevant data, although this strategy is not used very often. According to a study performed by a business analyst (IDC), two thirds of all companies planning to implement or already using RFID only focus on identification. This leaves a lot of the potential of RF based technologies unused [2]. RF based data storage may be used as a cheap alternative to cost intensive infrastructure for the electronic data interchange [3]. Still, the costs of buying and maintaining the RF based data storage and exchange infrastructure as well as the cost of providing high-quality and on-time data may add up to a considerable sum. These costs have to be matched by productivity improvements or revenue generation.

Even though we see a lot of interest in RFID, there are still only few productive installations. The following table lists some of the current success factors, which may also be described as RF readiness and distinguishes between todays and tomorrows demands. (Table 2)

Mandating may be possible in markets dominated by few large players. Typical industries are retail, defense, automotive and aerospace. Other industries such as the electronics market are far more distributed, thus mandating is rarely possible. In these competitive environments contractual incentives or cost sharing are possible solutions [4], but these approaches fail to scale in n-to-n relations. Besides mandating, RF installations will most likely be beneficial, if cost and benefit come together or security outweighs the cost.

Traditional supply chains with long-lasting relationships seem to benefit most from installing RF. The necessary investment will eventually pay off over time. As

Table 1 Shift to Digital n-to-n Information Logistics

Type of Information	Analog	Digital 1-to-1, n-to-1	Digital n-to-n
Not physically linked to product	Fax, paper	EDI, XML	"Internet-of-Things"
Physically linked to product	Barcode	RF (-ID), closed supply chains	RF (-ID), open networks

Table 2 RF readiness today and tomorrow

	RF readiness today	RF readiness tomorrow
Cost benefit ratio	Technology investors and beneficiaries are the same (e.g. internal projects)	Technology investor and beneficiaries are separate
	RFID installation is mandatory (e.g. defense, retail)	RFID is not mandatory
	Security outweighs cost (e.g. pharma, homeland security)	Security is a commodity
Supply chain issues	Traditional supply chains	Flexible Supply networks
	Long-lasting supplier customer relationship	Temporary business connections
	Closed loops	Open loops
Technical prerequisites	Existing IT infrastructure (e.g. ERP, SCM, PLM)	IT infrastructure needs to be implemented
	Existing messaging infrastructure (EDI, XML)	Messaging infrastructure needs to be implemented

discussed before, these traditional supply chains will transform into flexible supply networks, where the benefit of installing RF technologies will be harder to achieve. A return on investment (ROI) may also be possible based on cost-sharing in small networks. In open networks other means of balancing cost and benefits need to be applied. The need for an information market allowing to trade information according to supply and demand is eminent.

Most of the time when RF is used today, we see an already existing IT and messaging infrastructure, where linking the real and the virtual world through RF is a small step. If this infrastructure is not in place, further investments beyond RF infrastructure are necessary, but the overall potential of implementing RF based technologies is greater. “The degree to which Auto-ID technologies can enhance operations depends on the sophistication of process automation solutions a company already has in place.” [5] The mentioned technical prerequisites are one decisive factor why RFID is mainly driven by large companies. Similar to the usage of EDI, RFID still is a domain of big companies. A survey by the European Commission showed that only 2–4% of micro and small enterprises are using EDI while 43% of large companies utilize EDI. [6]

In order to gain better acceptance in the market, RF must still prove its benefit to all relevant stakeholders for future RF readiness. New concepts of cost benefit allocation are necessary to provide an incentive for the stakeholders that are not in the position to benefit from RF based technologies today. This article will propose a new concept of billing RF identification and additional RF related data. The proposed concept is based on the EPC network architecture as it has been introduced by the Auto-ID labs (Fig. 1).

The original EPC network architecture consisted of a collection of components, including tags and readers, Savant, the Object Name Service (ONS) and the Physical Markup Language (PML). Since 2004 the concept of the network architecture

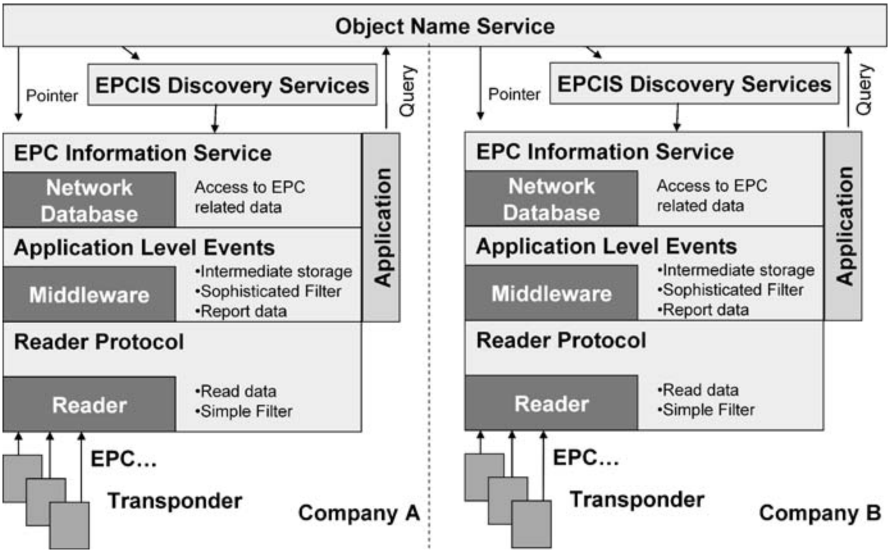


Fig. 1 Revised EPC Network Architecture, based on: [7]

has changed towards defining interfaces that may be implemented by commercial manufacturers. [7]

The first module within the "Internet-of-Things" is the transponder. The EPC Class 1 Generation 2 transponder has gained considerable acceptance in the market and is supported by most UHF reader manufacturers. The original concept is based on having the electronic product code (EPC) as the only identifier scheme with GS1 as the only issuing agency. A more open concept based on multiple issuing agencies as proposed by ISO/IEC 15459 would most likely benefit a lot of industries, where changing the existing identification scheme to the EPC would be similar to the year 2000 problem.

Besides tags and readers, multiple middleware products are offered in the market. The application level events (ALE) are supported by more and more middleware providers to interface to applications and databases. ALE provide raw observation reports from RF networks. These need to be combined with additional context information in order to generate events upon which applications such as ERP or WMS systems can act. [7]

The EPC information service can serve as a repository, which stores low level reports as well as higher level business content. It is designed as a query and update interface to applications and binds multiple databases and information systems. The ONS and the EPCIS discovery services have a coordination role and point to the relevant EPCIS to access the requested information. [7]

Scalability is a key-issue to the success of the network which is supposed to be used in large scale n-to-n relationships. It would be much more flexible than EDI, but still leaves a lot of open questions including security and data semantics.

2 Value of RF based information

As described above, pure identification data is enriched with data from databases and applications. The costs to provide and update these data are often left out of the scope of cost calculations. Instead the cost for tags, readers and antenna dominate within publications and presentations.

Enriched RF related information represents a real value to the stakeholders who can access and use this information to improve their knowledge. The definitions of knowledge and information can be taken from language theory where three dimensions are used to clarify the relationship between characters, messages and information. On the syntactical level, relations between characters are examined. The semantical level goes beyond that. Relations between characters and content are defined, thus representing messages. The pragmatic level looks at the recipient of the messages and characters and their relationship towards these data. Generally, information is existent, if existing knowledge is improved. Messages transporting already existing knowledge are defined as redundancy, not as information.

While there have been numerous studies on cost reduction through EDI and RFID, there has been little research on generating extra revenue with RFID and EDI. There are several reasons for this:

1. Information in B2B product sales is expected to be free of charge
2. The value of information in EDI-messages will most often be in the range of cents or fractions of cents
3. There is no established payment infrastructure to collect and invoice a huge amount of B2B EDI messages

Despite all these obstacles, this article will show that if it is possible to overcome these problems, the diffusion of RFID and EDI usage would be accelerated.

3 Solution Model – The "Billing Integrated Internet-of-Things"

Product related information has always been part of business transactions. With the growing automation of information sharing between business partners through electronic means such as EDI, the gap between virtual and real world has narrowed. The value of information is eminent, even though we see a massive reluctance to pay for product related information. There have been multiple concepts of enforcing the usage of new technologies such as RFID across companies and industries. Mandating may be seen as the main force to spread new means of electronic information sharing such as RF related technologies in markets dominated by large players such as retail, aviation or automotive. Mandates from the Department of Defense, Metro and Walmart are good examples of these strategies. But there are certain limits to mandates in distributed markets without presence of powerful key-players.

Cost and technology sharing is another well known method of enforcing new technologies between business partners. Technology sharing may be supported by technology itself. In case of RFID, a Chip Sharing Approach (CSA) was introduced in 2003 [8]. The CSA is an enhancement to the data on tag (DOT) approach. Several users throughout the supply chain may have their own individual storage area on an RFID chip, which is secured by different authorization. Cost sharing models may be seen as a first step to implement these new technologies in distributed markets, but they fail to have the necessary scalability. Until today, cost and technology sharing models are limited to scenarios with few players involved. The model fails to scale in large environments. The third option to enforce new technologies would be an open and market driven model where each usage of information across the supply chain can be calculated and invoiced. Asymmetries concerning the distribution of costs and benefits between information providers and information users may only be overcome, if standard market rules can be applied. This would include the separation of product and information cost. The supply and demand side would need to negotiate prices for information as well as for products.

The "Internet-of-Things" offers the necessary technology to query RF related product information. If a billing functionality would be included, it would be possible to charge certain information queries. The necessary additional components to the proposed "Billing Integrated Internet-of-Things" (BIIT) would be a billing service, a subscriber identity module (SIM) and the billing awareness of the components of the "Internet-of-Things" (Fig. 2).

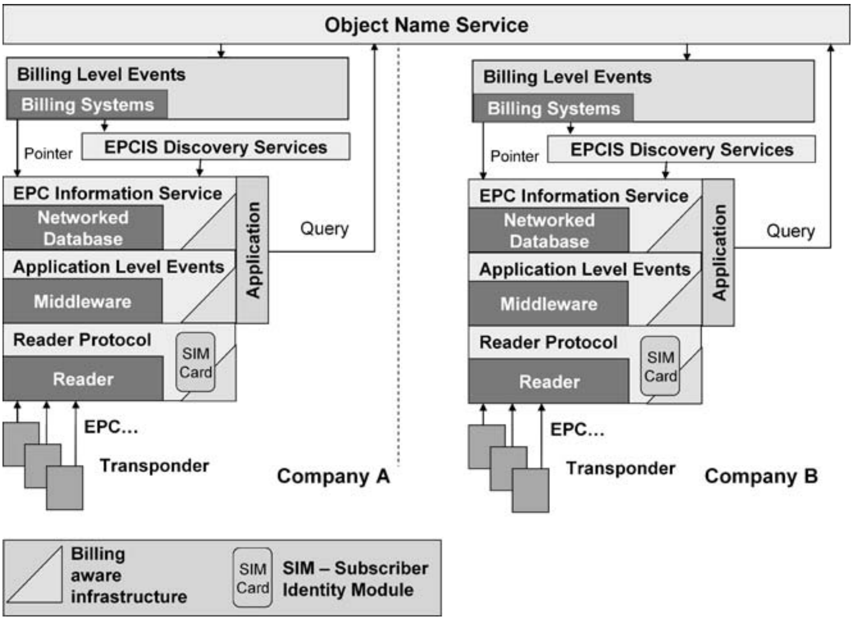


Fig. 2 The "Billing Integrated Internet-of-Things" (BIIT)

The billing service would be between the ONS and the EPCIS and could be integrated with the EPCIS Discovery Services, thus providing information about where the requested data is stored and at what price it can be accessed. It would ensure that billing level events (BLE) are generated, if a query asks for billable information. Billing level events could be standardized in order to forward billing events to other billing systems or to the financial systems of companies. It could be based on the concept of electronic bill presentment and payment. If the model proves to generate sustainable revenue, the entry of billing specialists such as telecom providers will be most likely. They would play an important role in aggregating micro payments to regular electronic bills that can be processed by the companies' ERP systems.

The subscriber identity module (SIM) is well known from mobile phones. Through SIM-cards, users are bound to a certain telecom provider. A similar concept could be introduced for RFID readers. The SIM-card could be obtained from billing system operating companies. In contrast to SIM-cards in mobile phones, there would not be a subscription to the network, because the existing internet connection will be used. The SIM-card would assure the connection to the billing service.

The readers, middleware and EPCIS would need to be enhanced to be "billing aware". This billing awareness could be very basic, e.g. pricing information would need to be forwarded. But it would also be possible to integrate software agents in order to allow autonomous price negotiations.

Today most electronic payment systems are focused on B2C markets. Thus, credit card billing systems such as telebilling systems are mostly used. In our model, we are focusing on information sharing between businesses. Therefore, electronic bill presentment and payment will be the preferred method. Still, the telecom providers would be in a good position, to grab some of the market proposed above.

1-to-1 payments based on EDI would leverage existing infrastructure. If our model proves to generate sustainable revenue, the entry of billing specialists such as telecom providers will be most likely. They will play an important role in aggregating micro payments to regular electronic bills that can be processed by the companies' ERP systems.

The concept of the BIIT would allow cost-sharing and profit generation across supply networks as well as product live cycles. It would allow SMEs to benefit, even if considerable investments are necessary. And it would allow service providers to focus on RF based information trading and offer their services to information providers as well as information users.

4 Business Scenarios

In order to explain our concept in more detail, we will show the relevance of our proposed model on the basis of three typical RFID scenarios. In our first scenario we have adopted a study by Accenture. [8] Accenture had a look at the time savings provided by usage of RFID versus barcode throughout the distribution value chain of DVD-players. While individual time savings did not exceed 188 seconds, the overall time savings within the value chain added up to 514 seconds (Fig. 3).

It is easy to see that the real potential of RFID usage in this case is distributed throughout the supply chain, whereas the cost of providing the necessary technology burdens the manufacturer. If it would be possible to charge the beneficiaries of the atomized identification through RFID, the dissemination of RFID would surely be accelerated. Similar scenarios can be found in most supply chains. This example showed just six process steps, but even supply chains with up to 50 process steps and more are known. [10] Another supply chain example, where the benefit of RF based information is distributed throughout a supply network, is baggage handling. There have been multiple pilots for baggage handling throughout the last years. While technical feasibility of RFID has been proven in most of these pilots, the ROI is hard to find. One airline would need to pay for the RFID tags, multiple airlines would be able to benefit, if the corresponding RF infrastructure is available. If the ROI for implementation and usage of RFID is calculated for a single airline with a good baggage performance, a negative ROI is possible. [11]

In a second scenario, we have a look at rental systems such as providers of returnable transport items (RTI). Collecting real-time value is still an issue in this scenario. The RTI providers would like to know exactly where their products are, who is using them and how long they have been used in order to achieve a time based pricing scheme. RFID could drastically improve visibility in this scenario and an integrated billing system would be well accepted.

The third scenario where our model would prove beneficial would be post key-player information handling. In markets dominated by large, influential players such as retail, aerospace or automotive, we see a problem in data allocation beyond the key-players interest. If for example logistic service providers, recyclers or service companies would like to get accurate and standardized data in combination with RFID, they have little to no influence on these key-players. If an infrastructure would be available at low cost to supply a usage based information billing system, it could be an easy incentive scheme for the OEM.

There will be a lot more scenarios where the concept of a "Billing Integrated Internet-of-Things" would prove to be beneficial.

	Identify Delivery	Check items with ship list	Identify Delivery	Check items with ship list	Check shelf inventory	Check out Customer	Total
Barcode	2 sec	203 sec	2 sec	203 sec	144 sec	10 sec	564 sec
RFID	2 sec	15 sec	2 sec	15 sec	15 sec	1 sec	50 sec

Fig. 3 Case for time savings within the Distribution Value Chain [8]

5 Conclusion and future work

The move away from traditional supply chains towards internet-driven supply networks is accompanied by a similar movement within IT-based messaging. While structured 1-to-1 fitted the traditional supply chains, the "Internet-of-Things" offers a scalable approach towards product related information sharing. RF based technologies may be seen as the combining factor between virtual and real world, but information collection, aggregation, processing and transmission does not come for free. Benefit and cost are not symmetrically distributed along the supply chain, though. Mandating has been successful in some industries, but fails in certain scenarios. Therefore, new scalable concepts of information selling and buying have to be developed.

The proposed concept of a "Billing Integrated Internet-of-Things" offers a scalable solution approach towards billing RF related information. It would offer a possibility to reward investments into the necessary IT infrastructure. Especially SME as well as service providers could benefit from this new approach.

Still there are some problems that need to be solved. Today's billing systems are mostly focussed on B2C scenarios. The usability in B2B scenarios still needs to be proven. The approach "Internet-of-Things" would need to be extended to enable a "billing awareness" of its components. Beyond the technical issue, a paradigm shift towards paying for valuable information is needed. More research will be necessary to overcome these obstacles.

References

1. Laudon K, Laudon P (2002) Essentials of Management Information Systems, Prentice Hall, 5th edition
2. Zwei-Klassen-Gesellschaft bei RFID (in German), Computerwoche, 16.06.2005, http://www.computerwoche.de/knowledge_center/rfid/557481/index.html, Last visited: 25.08.2007
3. Uckelmann, D (2007) Utilizing application- and data fact identifiers for structured, RFID-based data communication, 4th RFID Academic Convocation Brussels, 13./14.03.2007
4. Gaukler G M, Seifert R W, Hausman W H, (2007) Item-Level RFID in the Retail Supply Chain, Production and Operations Management, vol 16 (1), pp. 65-76
5. Chappell G, Ginsburg L, Schmidt P, Smith J, Tobolski J (2003) Auto-ID on the Line: The Value of Auto-ID Technology in Manufacturing, http://www.autoidlabs.org/single-view/dir/article/6/123/page.html, last visited: 22.08.2007
6. ICT Security, e-Invoicing and e-Payment Activities in European Enterprises, European Commission (2005), 09.2005, http://www.ebusiness-watch.org/studies/special_topics/2005/documents/TR_2005_Payments_IV.pdf , last visited: 22.08.2007
7. Harrison M (2004) EPC Information Service, www.m-lab.ch/auto-id/SwissReWorkshop/papers/EPCinformationService.pdf, last visited 22.08.2007

8. Taggruber V (2003) Die Bedeutung des Textilsegments auf dem Weg zu “Ubiquitous computing”, 3. Schönenberger Logistikforum, Landsberg/Lech, www.schoenenberger.de/Deutsch/forum2003/pdf/Vortrag-Taggruber.pdf, last visited: 22.08.2007
9. Establishing Performance Standards and Benchmarks for Pharmaceuticals through Cold Chain Logistics (2004) Department Transport & Urban Planning Transport Planning Agency, Government of South Australia
10. m2p Consulting (2005) Implementing RFID Tagged Baggage, The 3rd Annual Middle East Aviation IT Manager Forum, Dubai 22.11.2005

Reengineering and Simulation of an RFID Manufacturing System

Antonio Rizzi, Roberto Montanari, Andrea Volpi, Massimo Tizzi

Department of Industrial Engineering, University of Parma,
viale G.P.Usberti 181/A, 43100 Parma, Italy

Abstract We present a discrete event simulation model reproducing the adoption of Radio Frequency Identification (RFID) technology for the optimal management of common logistics processes of a Fast Moving Consumer Goods (FMCG) warehouse. In this study, simulation is exploited as a powerful tool to reproduce both the reengineered RFID logistics processes and the flows of Electronic Product Code (EPC) data generated by such processes. Due to the newness of RFID adoption in the logistics context and to the lack of real case applications that can be examined, we believe that the model developed can be very useful both to understand the practical implications of the technology applied in such field, and to provide a "proof of concept" to substantiate the adoption of RFID in the FMCG industry. Moreover, through the simulation tool, we also address the issue of how the flows of EPC data (*Matrix*) generated by RFID technology can be exploited to provide value-added information for optimally managing the logistics processes.

1 Introduction

Recently, Radio Frequency Identification (RFID) technology is experiencing an increasing diffusion for the optimization of many logistics systems (Auto-ID Center 2000, Prater et al. 2005). A main reason for RFID diffusion is the capability of tags to provide more information about products than traditional barcodes (Jones et al. 2004). Manufacturing site, production lot, expiry date, components type are among information that can be stored into the tag chip. Such data are stored into the tag chip in form of an Electronic Product Code (EPC), whose standards have been developed by the Auto-ID Center, a partnership founded in 1999 by five leading research universities and nearly 100 leading retailers, consumer products makers and software companies (Niemeyer et al. 2003). Moreover, tags do not need line-of-sight scanning to be read, since they act as passive tracking devices, broadcasting a radio frequency when they pass within yards of a reader (Karkkainen 2003). RFID tags also solve some of the inefficiencies associated with traditional barcodes, such as

H.-D. Haasis et al., *Dynamics in Logistics*
DOI: 10.1007/978-3-540-76862-3,

manually handling cases to read the codes (Boxall 2000), reducing time consumption and avoiding data capture errors. In some cases, readability of barcodes can also be problematic, due to dirt and bending, reducing accuracy and involving low reading rate (Moore 1999, Ollivier 1995).

In this paper, we present a discrete event simulation model reproducing the adoption of RFID technology for the optimal management of common logistics processes of a Fast Moving Consumer Goods (FMCG) warehouse. Here, simulation is exploited as a powerful tool to reproduce both the reengineered RFID logistics processes and the flows of *Matrix* generated by such processes. Simulation models are nothing new in supply chain literature, where they have been exploited with several different aims (see, for instance, (Wei and Krajewski 2000) and (Sivakumar 2001) for discrete simulation model, or (Homem-de-Mello et al. 1999) and (Tagaras and Vlachos 2001) for continuous ones). Nonetheless, due to the newness of RFID adoption in the logistics context and to the lack of real case applications that can be examined, simulating RFID logistics processes can be very useful to understand the practical implications of the technology applied in such field. Moreover, through the simulation tool, we also address the issue of how the flows of *Matrix* generated by RFID technology can be exploited to provide value-added information for optimally managing the logistics processes.

The study presented stems from a research project, called “RFID Warehouse” (RFIDWH), currently under development at the RFID Lab of the University of Parma, aiming at assessing the technical feasibility of reengineered logistics processes through RFID technology.

The remainder of the paper is organized as follows. In the next section, we briefly describe the RFID Lab of the University of Parma, together with the RFIDWH research project. Then, we detail the logistics processes examined, their reengineering for RFID adoption and the simulation tool developed to analyse them. Business Intelligence Modules (BIMs) developed to generate value-added information starting from *Matrix* are described in Sect. 4. Concluding remarks and future research directions are finally presented.

2 RFID Lab at the University of Parma

RFID Lab spun off from past research projects carried out at the Department of Industrial Engineering of the University of Parma, in the field of RFID-based systems to manage the FMCG and food supply chain. Such projects, launched in 2002, have examined the potentials of RFID for traceability, a mandatory EU requirement in the food industry (Rizzi 2006) as well as its impact on FMCG supply chain processes.

Moreover, specific knowledge and capabilities have been devised from research projects carried out in collaboration with corporations and governmental bodies, such as GS1 Italy (Bottani and Rizzi 2008) and the Italian Ministry for Technological Innovation. This know-how has been embodied into RFID Lab of the University of Parma, a centre for research, training and technology transfer in the RFID area.

The lab, founded and headed by Prof. Antonio Rizzi, opened in May 2006, and is currently located in a 150 m^2 area at the Department of Industrial Engineering of the University of Parma. The lab is the first Italian facility to be cleared by the Italian Ministry of Defense with a temporary site license to operate RFID equipments in the UHF 865–868 MHz spectrum (RFID Journal 2006), according to ETSI 302 208 regulations (ETSI 2006).

RFID Lab also strives to become a national benchmark for technology transfer in the RFID context. In this regards, about 20 technology providers collaborate in RFID Lab activities, together with a steering committee, called "Board of Advisors" (BoA), set up to lead the research activities of the RFID Lab. The board currently encompasses 20 major Italian manufacturers, 3PLs and retailers of FMCG.

Research activities of the lab are very broad-based and reflect different interests in both theoretical and applied areas. Major topics concern Business Process Reengineering (BPR) and the evaluation of RFID impact on logistics and supply chain processes in the FMCG supply chain. RFIDWH is a main research project launched in late 2006 at the lab. As mentioned, the general aim of the project is, on the one hand, to evaluate the technical feasibility of reengineered logistics processes of a representative FMCG warehouse through RFID technology, and, on the other one, to quantitatively assess the improvements brought in by the technology in the FMCG industry, which were analyzed in the above mentioned research. Accordingly, RFIDWH project encompasses 4 subsequent work-packages (WP), namely: (1) "Definition and engineering of RFID logistics processes"; (2) "Business Intelligence Modules (BIMs) definition"; (3) "Software development"; and (4) "Simulation". In this paper, we mainly focus on describing the results of WP2 and WP4. To this extent, the development of BIMs in WP2 represents a relevant achievement of the project, since BIMs strictly related to the simulation of reengineered logistics processes performed in WP4. As a matter of fact, BIMs are software tools which process RFID data coming from the (simulated) field to generate value-added information that could be usefully exploited to optimize logistics processes. Hence, properly structuring BIMs and the simulation model allows both reproducing RFID logistics processes to assess their technical feasibility and addressing how *Matrix* generated by RFID equipments can be exploited as value-added tools for the optimal management of the warehouse.

3 Reengineering and simulation of logistics processes

The processes analysed and reengineered in WP1 were defined together with several companies adhering to the RFID Lab BoA, and operating in the FMCG supply chain; similarly, reengineered models were approved by the board.

Figure 1 provides the general scheme of the processes examined. Reengineered logistics processes for RFID adoption are described in the following.

- Receiving: this process encompasses all activities related to physically acceptance of materials at the warehouse. It includes unloading products from the in-

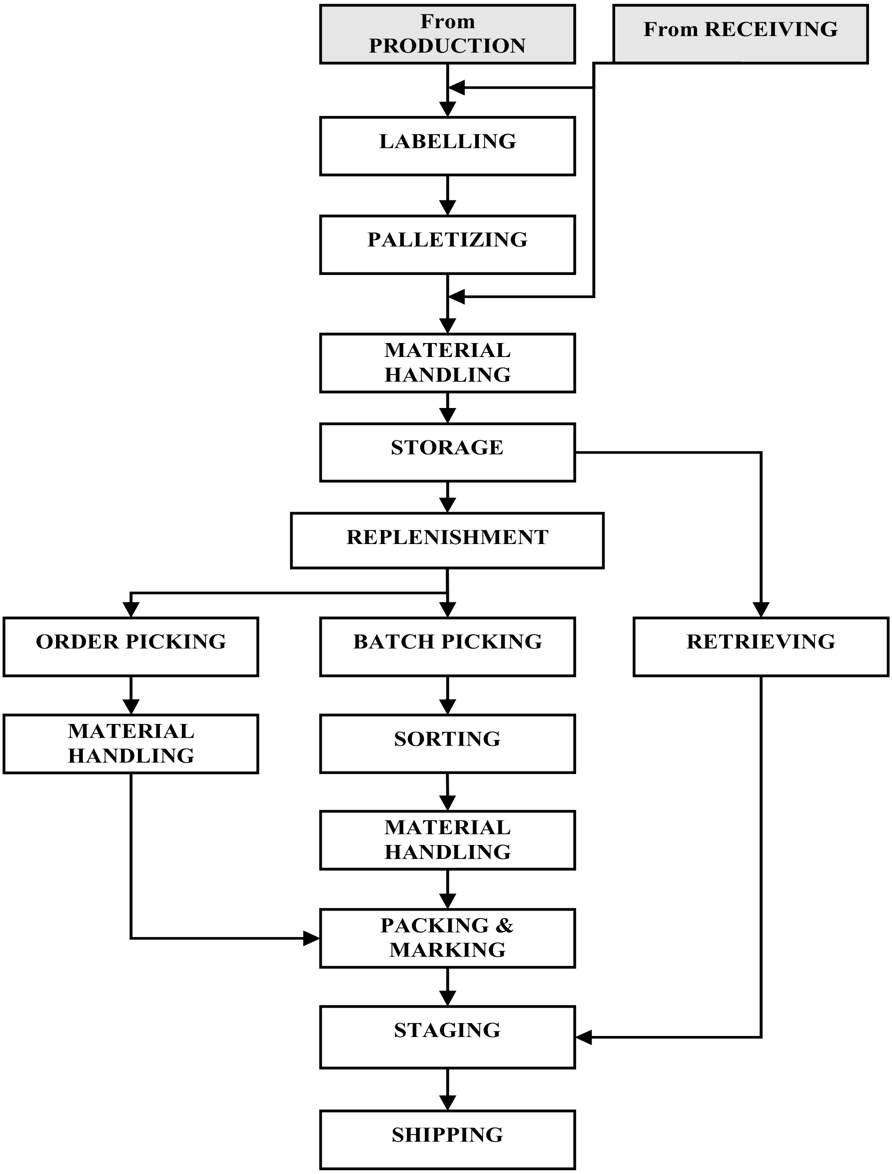


Fig. 1 General scheme of the warehouse processes

bound transportation mode, staging, verifying quantity, conditions and type of the materials and related documents, documenting information and updating the company's Information System (IS).

- Labelling: such a process can be performed either at the end of the production process or after receiving operations at the warehouse. Product cases are the input of the process, and should be positioned on conveyors and labelled with RFID tags (whenever cases are not RFID tagged); then, the company's IS should be updated with information stored in the EPCcases, such as GTIN, LN or traceability information. Checking for labelling errors could also be performed.

- Palletising: labelled cases, whose data have been stored into the company's IS, are first located on a sorter, which acts as a reader of EPC codes and sorts the cases according to a defined palletizing scheme. Single-item (same item, different lots on a pallet), single-lot (same item and same lot on a pallet) and mixed pallets (different items) schemes are considered. As a result, each pallet is identified by a RFID tag, which, through the IS, is linked to information related to input cases.
- Storage: activities involved in this process basically consist in picking up labelled pallets from a temporary storage area, transporting them to the storage rack and placing them in the proper location. This latter is identified by the companies Warehouse Management System (WMS), based on a defined storage policy.
- Material handling: this process encompasses several different activities, whose aim is to change the location of labelled cases or pallets within the warehouse. Material handling operations can be are requested by the WMS as:
  - replenishment: it consists in relocating materials from a storage location to an order picking location and documenting this transfer;
  - staging: in this operation, materials are physically moved from the packing area to the shipping area based on a prescribed set of instructions related to a particular outbound vehicle or delivery route often for shipment consolidation purpose. Staging is performed after Retrieving or Packing & Marking processes, and is required whenever shipping orders are received;
  - material handling: it involves a generic movement of pallets/cases between two different locations of the warehouse.
- Order selection: this activity involves selecting and picking the required quantity of products, according to the shipping orders to be fulfilled, and to move them to the packing area. Both single-item and mixed pallets may results from the process. In the reengineered model, three picking policies are examined, namely:
  - order picking: employees perform picking activities on cases according to a specific customer order list. In this case, all items related to a given order are picked;
  - batch picking: picking is performed on a specific product, which is required in different orders. Orders are pooled, and cases are picked and automatically sorted according to the shipping orders received;
  - retrieving: full pallets which match orders received are picked from the storage area.
- Sorting: sorting follows the batch picking operation, and receives, as input, labelled cases resulting from picking activities, together with information related to the number of cases picked and the required quantity in orders to be fulfilled. Cases should thus be sorted based on orders received. Hence, as a results of the

process, several set of cases are prepared in as many packing positions; through the company's IS, each set is linked to the related order.

- Packing & Marking: this activity encompasses checking, packing and labelling one or more items of an order into a shipping unit (i.e. usually a pallet). Thus, pallets are identified by EPCpallets, containing shipping destination data and additional information that can be required.
- Shipping: materials are moved from the staging area and loaded into an outbound vehicle; documents related to the movement are updated.

Each of the above reengineered models has been implemented into an appropriate discrete event simulation tool. In reproducing the processes, particular attention has been paid to satisfy physical and informative constraints of the warehouse. This has allowed to faithfully reproduce the real logistics processes and to generate a wide amount of consistent data, that could be used to test the correctness of the BIMs developed. Such data have thus been exploited to compute the performance indexes implemented in the BIMs.

To achieve the above aim, the following steps were followed in developing the simulation model:

1. definition of scope and problem in exam;
2. development of a conceptual model;
3. validation of the conceptual model;
4. analysis of input data;
5. mathematical formulation of the model;
6. model calibration;
7. definition of the experimental campaign;
8. analysis of output data.

An overview of the simulation model, in its final form, is provided in Fig. 2. The model was developed using commercial software Simul8^{TM} Professional release 12 (Visual Thinking International Inc.).

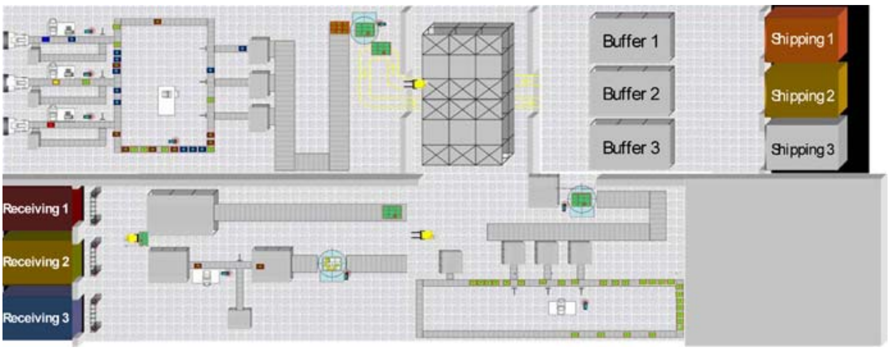


Fig. 2 Overview of the simulation model

4 Development of BIMs and results

Matrix resulting from the simulation runs become the input parameters of the BIMs, which provide, as output, value-added information for the optimal management of the warehouse processes.

To this extent, data are first organised into a properly designed database, consistent with the OLAP (on-line analytical processing) cube theory; starting from these data, three BIMs have been developed and implemented in MS Excel^{TM}. They are described in the following.

- Flow Time Management (FTM): it provides quantitative indications about activities that increase products value and activities that add costs but do not add value to products. As a result, the FTM module provides a chart describing the time required to perform the processes, distinguishing between "value-added time" and "non-value-added time". An example is provided in Fig. 3.
- Key Performance Indicators (KPIs): KPIs provide a quantitative assessment of processes performance; among others, such performance include time management and productivity, assets and systems utilization, distance from defined target values. Since the full list of performance indicators is too long to be fully reported in the paper, we describe, as an example, KPIs related to the storage process. For such process, the BIMs allow determining storage area utilization, flowtime and average stock level. As can be seen from Fig. 4, the BIMs also track which SKUs have been stored in each storage location in a defined time period.
- Track and Trace (T&T): this module provides value-added information that can be used for the optimal management of tracking and tracing of either pallets or cases. For instance, for each case in the warehouse T&T allows identifying manufacturing lot, related pallets, processes performed on it, related time period

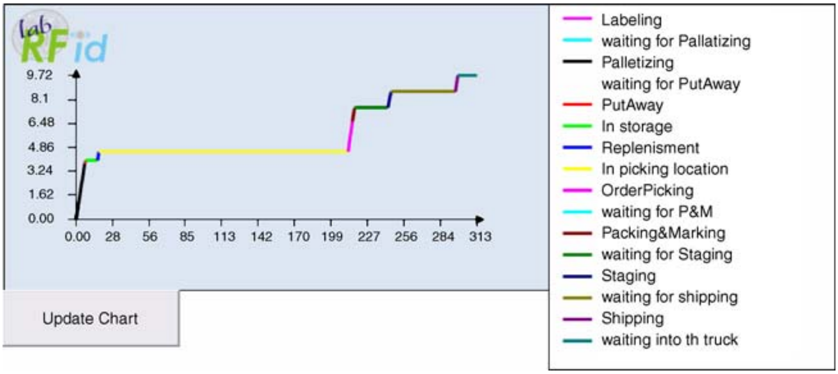


Fig. 3 Example of FTM chart

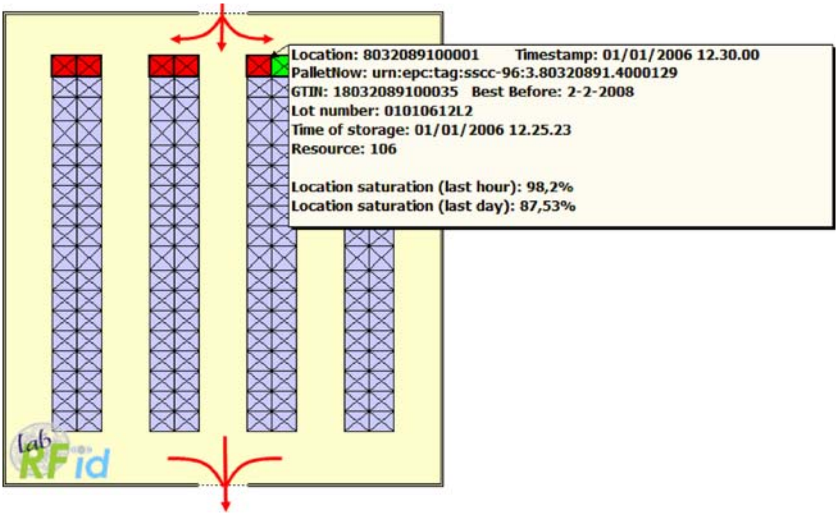


Fig. 4 Example of KPI chart for the storage process

and employee; similarly, for shipped/received products, manufacturing lot, supplier/customer, related order and shipping/receiving bay can be easily traced. Figure 5 provides an example of traceability information that can be managed through the T&T module.

Read Time EPCcase	EPCpallet	process Labelling	Palletising
urn:epc:tag:sgtin-96:3.80320891.10145.107	urn:epc:tag:sscc-96:3.80320891.10000002	1/1/06 9.02.09	1/1/06 9.02.36
urn:epc:tag:sgtin-96:3.80320891.10145.113	urn:epc:tag:sscc-96:3.80320891.10000002 urn:epc:tag:sscc-96:3.80320891.20000010	1/1/06 9.02.19	1/1/06 9.02.48
urn:epc:tag:sgtin-96:3.80320891.10145.115	urn:epc:tag:sscc-96:3.80320891.10000002 urn:epc:tag:sscc-96:3.80320891.20000010	1/1/06 9.02.22	1/1/06 9.02.50
urn:epc:tag:sgtin-96:3.80320891.10145.13	urn:epc:tag:sscc-96:3.80320891.10000002	1/1/06 9.00.19	1/1/06 9.00.36
urn:epc:tag:sgtin-96:3.80320891.10145.15	urn:epc:tag:sscc-96:3.80320891.10000002	1/1/06 9.00.20	1/1/06 9.00.43
urn:epc:tag:sgtin-96:3.80320891.10145.34	urn:epc:tag:sscc-96:3.80320891.10000002	1/1/06 9.00.47	1/1/06 9.01.08

Fig. 5 Example of T&T chart

5 Future research directions and conclusions

In this study, we have shown how simulation can be exploited to test the applicability of the RFID technology in common logistics processes of the FMCG supply chain. To our knowledge, this is the first study attempting to investigate reengineered RFID logistics processes and the flow of data generated by such processes. Our study has two main outcomes. First, by reproducing real logistics processes, the simulation model developed can provide a "proof of concept" of the technical and economical feasibility of RFID implementation in the FMCG industry. Moreover, the BIMs developed also show that *Matrix* can be properly processed to provide value-added information that can be usefully exploited to optimize the warehouse processes.

Our study could be the starting point to develop a pilot project, where the same processes analysed through the simulation tool could be in-field examined. In field measurements could provide a direct assessment of potential savings for each process and for each supply chain partner involved.

References

Auto-ID Center (2002) Technology Guide. Cambridge University, Auto-ID Center (available at http://interval.hu-berlin.de)

Bottani E, Rizzi A (in press) Economical assessment of the impact of RFID technology and EPC system on the Fast Moving Consumer Goods supply chain. Int J Prod Econ

Boxall G (2000) The use of RFID for retail supply chain logistics. Paper presented at Tag 2000, Baltic Conventions, The Commonwealth Conference & Events Centre, London, 24 May

ETSI (2006) Electromagnetic compatibility and Radio spectrum Matters (ERM) - Radio Frequency Identification Equipment operating in the band 865 MHz to 868 MHz with power levels up to 2 W – Part 1: Technical requirements and methods of measurement (available at www.etsi.org)

Homem-de-Mello T, Shapiro A, Spearman ML (1999) Finding optimal material release times using simulation-based optimization. Manage Sci 45(1): 86–102

Jones P, Clarke-Hill C, Shears P, Comfort D, Hillier D (2004) Radio frequency identification in the UK: opportunities and challenges. Int J Retail and Distribution Management 32(3): 164–171

Karkkainen M (2003) Increasing efficiency in the supply chain for short shelf life goods using RFID tagging. Int J Retail and Distribution Management 31(3): 529–536

Moore B (1999) Barcode or RFID: which will win the high speed sortation race? Automatic ID News 15(7): 29–36

Niemeyer A, Pak MH, Ramaswamy SE (2003) Smart tag for your supply chain. The McKinsey Quarterly 4: 6–8

Ollivier M (1995) RFID enhances materials handling. Sensor Review 15(1): 36–39

Prater E, Frazier GV, Reyes PM (2005) Future impacts of RFID on e-supply chains in grocery retailing. Supply Chain Management 10(2): 134–142

RFID Journal (2006) RFID News Roundup (available at http://www.rfidjournal.com)

Rizzi A (2006) La tracciabilità nel settore alimentare, un'analisi critica (in Italian). Logistica Management 166: 85–98

Sivakumar AI (2001) Multi-objective dynamic scheduling using discrete event simulation, Int J Comp Integ M 14(2): 154–167

Tagaras G, Vlachos D (2001) A periodic review inventory system with emergency replenishments. Manage Sci 47(3): 415–429

Wei J, Krajewski L (2000) A model for comparing supply chain schedule integration approaches. Int J Prod Res 38(9): 2099–2123

LIT Middleware: Design and Implementation of RFID Middleware Based on the EPC Network Architecture

Ashad Kabir, Bonghee Hong, Wooseok Ryu, Sungwoo Ahn

Dept. of Computer Engineering, Pusan National University, Republic of Korea
e-mail: {makcse,bhhong,wsryu,swan}@pusan.ac.kr

Abstract Radio Frequency Identification (RFID) technology can be used to significantly improve the efficiency of business processes by providing the capability of automatic identification and data capture. RFID system must implement functions to posses the enormous event data generated quickly by RFID readers. The proliferation of RFID tags and readers will require dedicated middleware solutions that manage readers and process event data. In this paper, we describe our middleware (called LIT Middleware) architecture with key features and also illustrate overall functional components of the middleware software as well as the system framework. Our component based middleware architecture provides extensibility, scalability and abstraction, and is also compatible with EPCglobal standards.

1 Introduction

RFID technology uses radio-frequency waves to transfer data between readers and movable tagged objects, thus it is possible to create a physically linked world in which every object is numbered, identified, cataloged, and tracked. RFID is automatic and fast, and does not require line of sight or contact between readers and tagged objects.

RFID systems have recently begun to find greater use in industrial automation and in supply chain management. In these domains RFID technology holds the promise to eliminate many existing business problems by bridging the gap between the virtual world of IT systems and the real world of products and logistical units. Common benefits include more efficient material handling processes, elimination of manual inventory counts, and the automatic detection of empty shelves and expired products in retail stores [1]. However, not only the business community that can benefit from the use of RFID tags, but also the consumer. The magic medicine cabinet [10], the magic wardrobe, and the often-cited fridge are some of the applications in which the consumer would benefit from these smart products.

The widespread adoption of RFID requires not only low cost tags and readers, but also the appropriate networking infrastructure. To facilitate application development

DOI: 10.1007/978-3-540-76862-3, © Springer 2008

even further, RFID infrastructure can also feature another component that consumes the events delivered by the middleware, combines the RFID data with application logic, and generates application-level events.

The concept of a distributed networking infrastructure for RFID was initially proposed by the Auto-ID Center, an industry sponsored research program to foster RFID adoption, which coined the term Electronic Product Code (EPC) Network. Sun Java System also provides RFID software architecture for enabling the EPC-global Network [2]. Java System RFID Software consists of two major components – The Java System RFID Event Manager and the Java System RFID Information Server. Oracle sensor based services offers a complete sensor based computing infrastructure for capturing, managing and responding to RFID and other sensor data implemented in the physical environment [3].

In this paper, we have analyzed those requirements the RFID middleware component should meet in order to manage large deployments of heterogeneous readers and the amount of data those readers capture. The main contribution of this paper is designing software component for RFID middleware, which addresses both application needs and the constraints of RFID technology. We called our middleware as Logistics Information Technology (LIT) Middleware.

This paper is organized as follows. In Sect. 2, we review the EPCglobal network architecture relevant to our work. In Sect. 3, we describe features of our middleware. In Sect. 4, we illustrate implementation of LIT middleware with software architecture. We conclude in Sect. 5 with avenues for further research.

2 Overview of EPC Network Architecture

The EPC Network, originally developed by the Auto-ID Center with its standards now managed by EPCglobal Inc. (also named shortly EPCglobal) [5], was designed to enable all objects in the world to be linked via the Internet. The EPC Network is built upon many fundamental technologies and standards. Figure 1 shows the structure of typical EPC Network [7] proposed by EPCglobal. In this paper, we focus on Application Level Events (ALE) and EPC Information Services (EPCIS), and consider both of them as RFID Middleware. In the rest of this section, we shortly discuss ALE and EPCIS.

The EPC Network Architecture defines an interface through which clients may obtain filtered and consolidated EPC data from a variety of sources. This interface and functionality within the EPC Network Architecture are called ALE. The processing done at ALE typically involves: (1) *receiving* EPCs from one or more data sources such as readers; (2) *accumulating* data over intervals of time, *filtering* to eliminate duplicate EPCs and EPCs that are not of interest, and *counting* and *grouping* EPCs to reduce the volume of data; and (3) *reporting* in various forms. The role of the ALE interface within the EPC Network Architecture is to provide independence among the infrastructure components that acquire the raw EPC data, the architectural component(s) that filter and count EPC data, and the applications that use the data [8].

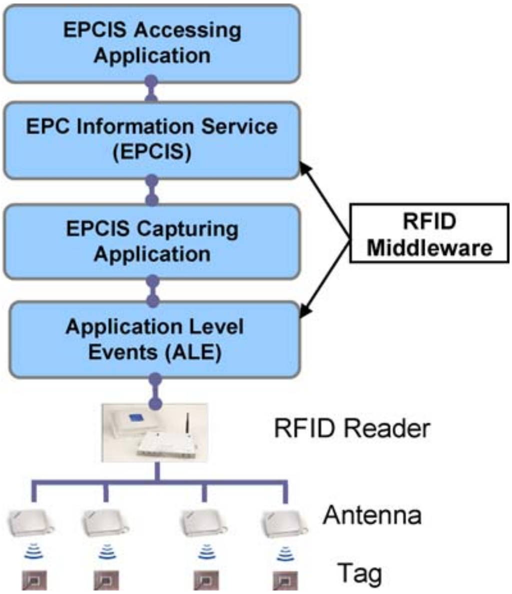


Fig. 1 RFID Middleware on the EPC Network Architecture

The EPCIS provides a uniform programmatic interface to allow various clients to capture, secure and access EPC-related data, and the business transactions with which that data is associated. The goal of EPCIS is to enable disparate applications to leverage EPC data via EPC-related data sharing, both within and across enterprises. *EPCIS Accessing Application* is responsible for carrying out overall enterprise business process, such as warehouse management, shipping and receiving, historical throughput analysis, and so forth, aided by EPC-related data. *EPCIS Capturing Application* supervises the operation of the lower EPC elements, and provides business context by coordinating with other sources of information involved in executing a particular step of a business process [6].

3 Features of LIT Middleware

Our LIT Middleware considers the concept of ALE and EPCIS. In the rest of this section, we have summarized the key features of our middleware on the basis of ALE and EPCIS.

3.1 Features of ALE

High Performance and Scalability

RFID Middleware can have many applications, queries and readers. These criteria increase query processing time and also network traffic. RFID middleware should have ability to handle many applications, queries and readers. Hence, we have implemented *Duplication Removal technique* for eliminating duplicate data and ultimately reducing network traffic. We also implemented *Continuous Query Process* and *State Based Execution* for high performance.

Continuous Query Process RFID data is usually regarded as the streaming data which are huge, because they are gathered continuously by many readers. RFID middleware filters those data to process queries which applications request. The EPCglobal proposed ALE specification which mention Event Cycle Specification (ECSpec) and Event Cycle Report (ECReport) as the standard of interface for the RFID middleware. ECSpec is a specification for filtering and collecting RFID tag data and is treated as a Continuous Query (CQ) process during fixed time interval repeatedly. ECReport is a specification for describing the results of ECSpec. Figure 2 shows the architecture of CQ processing. In CQ processing, first application requests query through ECSpec to CQ Manager. CQ Manager builds *query index* (representing query using spatial data structure that improves the speed of operation) on the basis of query request. Reader interface continuously receives EPC data from RFID readers and evaluate queries using query index. Finally, send the result to user or application through CQ Manager.

Duplication Removal Technique RFID Middleware receives EPC codes from readers. As tag reads are flowed into the middleware, there are a lot of noises and duplicate reads, so filtering operation is used to eliminate that redundant information, called duplication removal technique. During request query, user or application also mention time duration of each query. ALE treats this duration as *Event Cycle*. During each *Event Cycle* RFID readers can read same tag many times. Duplication removal technique removes those duplicate data.

State-based Execution In our continuous query processor, we use *state-based execution* model instead of *thread-based execution* model [4]. When the number of thread is large, the thread-based execution model incurs significant overhead due to cache misses, lock contention, and switching. On the other hand, *state-based execution* model is a single scheduler thread that tracks system state and maintains the execution queue. The execution queue is shared among a small number of worker

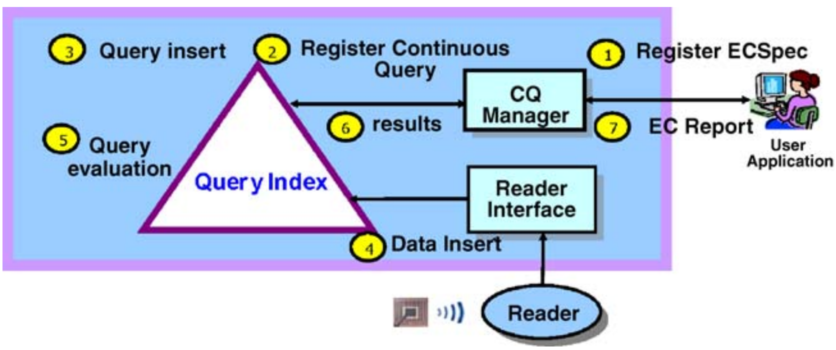


Fig. 2 Continuous Query Processing in LIT Middleware

threads responsible for executing the queue entries. This state-based model avoids the limitations of the thread-based model, enabling fine grained allocation of resources according to application specific targets.

Abstraction and Extensibility

Reader Abstraction and Extensibility There are many vendors of RFID Readers. Our system is independent from types of RFID Readers and vendors and also provides extensible architecture for new type of RFID reader. We used Common *Reader Management Interface* to manage heterogeneous RFID reader and also used UNIX device driver and Jini architecture concept, which can add new adapter at run time. We also used logical reader mapping which provide independence from number and arrangement of RFID readers.

Application Abstraction There are many types of applications those use RFID middleware. Our system provides independence from different types of application using *Common Application Interface* and *Common Application Query*. We used common application interface provided by EPCglobal, which are Tag Memory Specification API, ALE Reading API, ALE Writing API, ALE Logical Reader API and Access Control API [8]. We also used common application query (ECSpec) provided by EPCglobal and standard format (e.g., XML) and standard protocols (HTTP, FILE and TCP).

3.2 Features of EPCIS

High Performance

The question of performance has always to be seen in relation to the questions of response time. In case of only few requests or queries to be processed, too much care about performance is not necessarily required. Having thousands of queries to be processed (in a short period of time) requires a well designed system, being able to bear the burden. We used CQ Index Technique to process enormous query efficiently. For track and trace query, we have implemented Fixed Interval R-tree [9]. Fixed Interval R-tree proposed new indexing, insert and split policy to process queries efficiently. It used to trace trajectories as well as monitor present locations of the tags. A tag trajectory is represented as a line by connecting two spatiotemporal locations captured when the tag enters and leaves the vicinity of a reader.

Abstraction and Extensibility

Our system is persistent and compatible to store different types of EPCIS event and also support commercial legacy database (Oracle, MSSQL, MYSQL etc.). Hence,

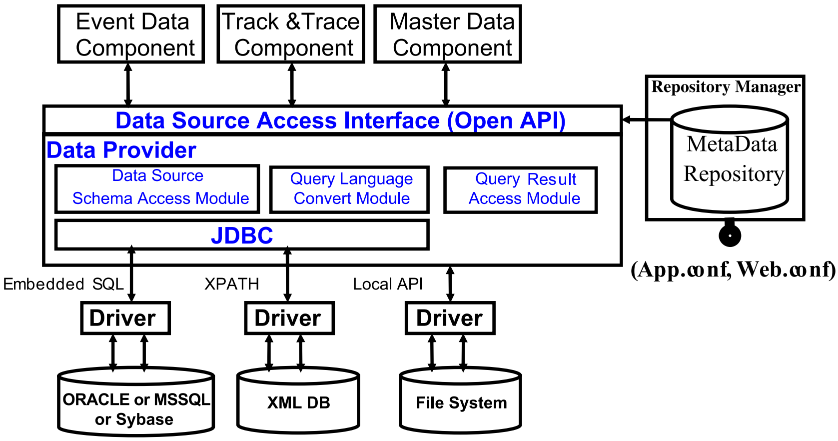


Fig. 3 Data Source Access Component

we designed *Data Source Access Component* as shown in Fig. 3. It consists of *Data Source Access Interface* and *Data Provider*. Using interface different components of EPCIS can interact with legacy database. Using data provider, we can access various types of data sources.

Different types of applications use EPCIS. EPCIS provide independence from different types of applications and platforms by supporting standard query interface (EPCIS query interface) proposed by EPCglobal [6]. It also uses standard data format (e.g., XML) and standard communication protocol (HTTP, TCP and WSDL). EPCIS also provide extensible architecture. That is, it support new EPCIS event (e.g., Sensor Event, e-Seal Event) by extending XSD and WSDL.

4 Design and Implementation of LIT Middleware

Figure 4 shows the ALE Software architecture. It consists of four layers: Application Abstraction Layer, State-based Execution Layer, Continuous Query Layer and Reader Abstraction Layer.

Application Abstraction Layer provides RFID data access via web-based ALE API by using JAX-WS. Our application also uses different types of logger (TCP, HTTP and FILE) to send result (ECReport) to subscriber. *State-based Execution Layer* is overall controller of middleware. It consists of controller, scheduler, query manager, reader manager and Thread Pool. Controller initiates different modules and manages whole system. Scheduler is thread based class used to manage schedule of middleware. Query manager stores and manages the query information. Reader manager keeps information of logical and physical reader mapping and makes the upper layer independent from the change of the physical reader configurations. The primary role of *Reader Abstraction Layer* is to support seamless link between the middleware software and various types of RFID readers. This component

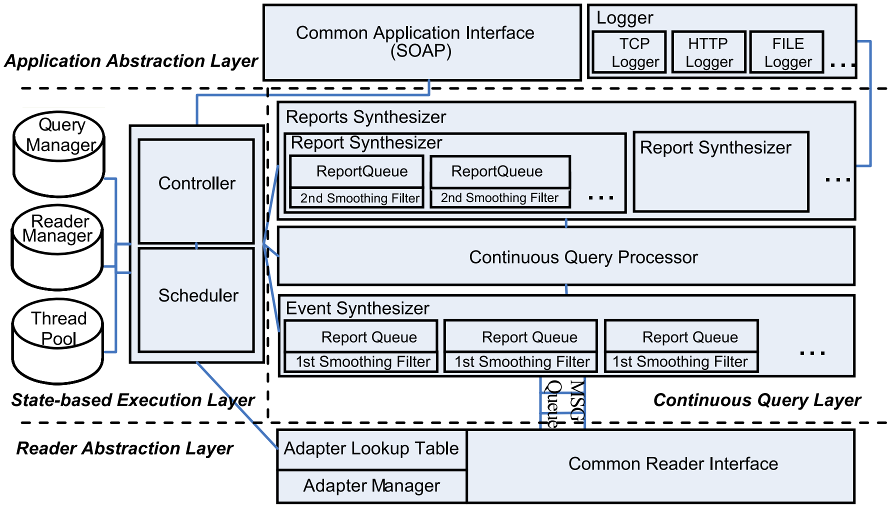


Fig. 4 ALE Software Architecture

abstracts the tag read protocols, supports various types of reader devices and allows users to configure, monitor and control all these devices. It supports XML-based reader control methodology to control reader specific features. The functions of *Continuous Query Layer* are collecting, filtering and removing duplicate EPC data.

Figure 5 shows the software architecture of EPCIS. It consists of three layers. *Query Service Layer* consists of query control component and standard query inter-

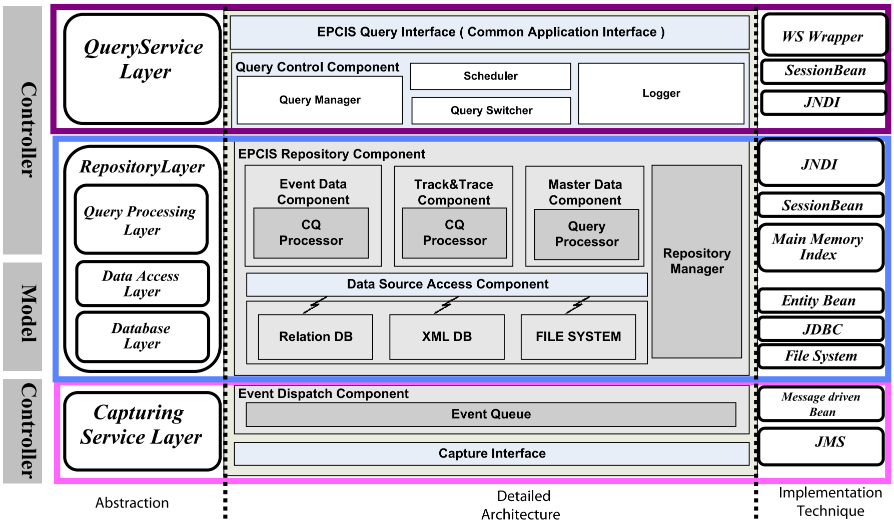
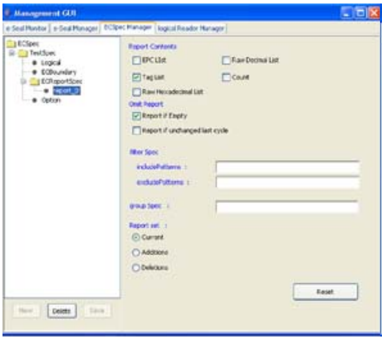


Fig. 5 EPCIS Software Architecture



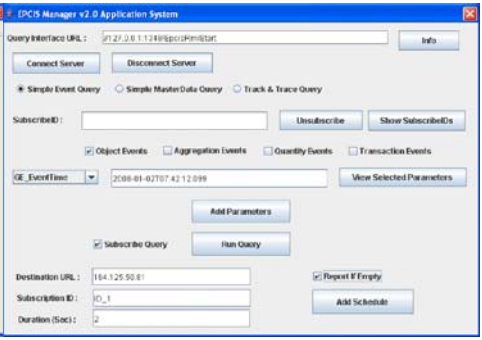


Fig. 6 Management GUI of LIT Middleware

Fig. 7 Client GUI of LIT Middleware

face that is independent on application program. *Repository Layer* consists of Query Processing Layer (here we implemented CQ Technique and Fixed Interval R-Tree), Data Access Layer (here we implemented Data Source Access Component) and Database Layer (consists of Legacy DB). *Capturing Service Layer* consists of event queue and capturing interface that gather information from ALE.

We have implemented our system using several software tools: JAVA (JDK 5.0), Tomcat Web-server (version 6.0), and Smart UML (version 5.0.2). We have tested our system using RFID devices from several vendors such as Alien and Intermec. Figure 6 is the Management GUI of LIT Middleware, used to manage ECSpec and logical reader. Figure 7 is the Client GUI of LIT Middleware, used to query in EPCIS Repository.

5 Conclusions

In this paper, RFID-based system called 'RFID Middleware' is examined and recently ratified EPCglobal de-facto standard interface for RFID middleware, ALE and EPCIS are also discussed. Then, the RFID-based middleware system which is compatible with the ALE and EPCIS standard is introduced with internal architecture. The nature of RFID system, requiring distributed network architecture, drives the requirements for common interface to let various applications have access to RFID tag data in a standardized way and such requirements bear ALE and EPCIS specification. In this paper, we have analyzed the features of our middleware which is compatible with ALE and EPCIS specification, and also summarize its architecture and functional operations. Future research includes requirements analysis, and implementation of Middleware for RTLS and Sensor Network.

Acknowledgements "This work was supported by the Korea Research Foundation Grant funded by the Korean Government (MOEHRD – Ministry of Education and Human Resources Development)".

References

Alexander K, Gilliam T, Gramling K, Kindy M, Moogimane D, Schultz M, Woods M (2002) Focus on the supply chain: Applying auto-id within the distribution center. Technical Report IBM-AUTOID-BC-002, Auto-ID Center

A Technical White Paper (2005) The Sun Java™ System RFID Software Architecture. www.sun.com

An Oracle White Paper (2006) Enterprise Information Architecture for RFID and Sensor-Based Services, www.oracle.com

Carney D, Çetintemel U, Rasin A, Zdonik S, Cherniack M, Stonebraker M (2003) Operator Scheduling in a Data Stream Manager. 29th VLDB Conference, Berlin, Germany, pp 838–849

EPCglobal Inc. http://www.epcglobalinc.org

EPCglobal Inc. EPC Information Services (EPCIS) Specification. Working Draft Version 1.0, June 8, 2005

EPCglobal Inc. EPCglobal Architecture Framework. Final Version July 1, 2005

EPCglobal Inc. The Application Level Events Specification. Working Draft Version 1.1, July 13, 2006

Ahn S, Hong B, Ban C, Lee K (2006) Design and Implementation of an Index Structure Using Fixed Intervals for Tracing of RFID Tags. ICCSA, LNCS 3981, pp175–185

Wan D (1999) Magic medicine cabinet: A situated portal for consumer healthcare. In Proceedings of the International Symposium on Handheld and Ubiquitous Computing, Karlsruhe, Germany

Shelf Life Prediction by Intelligent RFID – Technical Limits of Model Accuracy

Reiner Jedermann[1], Jean-Pierre Emond[2], Walter Lang[1]

[1] Institute for Microsensors, -actuators and -systems
e-mail: {rjedermann,wlang}@imsas.uni-bremen.de
[2] UF/IFAS Center for Food Distribution and Retailing, e-mail: jpemond@gmail.com

Abstract Information about current quality state enables large improvements in the dynamic planning for the logistics of fresh products. The main reason for quality losses are temperature deviations. This article investigates how a prediction system that delivers a shelf life prediction at each transhipment point could be integrated into a semi-passive RFID label. Limiting factors for the possible predictions accuracy like model tolerances and the bound processing resources of low-power microcontrollers will be evaluated.

1 Introduction

Fresh perishables are a great challenge for retailers. These products must move as fast as possible in order to provide the best quality to the consumers. For some produce only few hours delayed in cooling or in the distribution chain can be sufficient to prevent their marketability. Today, new knowledge and technologies can allow better control of the distribution chain from field/processing to store.

For many retailers the best way to keep or gain market share is by presenting an attractive produce section. A positive image of the produce department gives the customers a better perception of the quality of the business. In the minds of many customers, if the store provides a high quality fresh produce, it probably maintains the same high quality for the other products in the store. The visible attributes of the produce, that is the appearance of the produce is maybe the most important factor that determines the market value of fresh fruit and vegetables. However, the price to pay to keep the "image of freshness" is to have an inventory turnover of almost 50% per day. This is the highest percentage in a retail store followed by the meat and fish sections. One of the reasons is the perishable aspect of the produce. Most of the time, conditions found in the store are far from the recommended temperature and humidity needed to maintaining an optimal produce quality. Often, a short exposure of 1 or 2 hours to inadequate conditions (i.e. too high or too low temperatures, too dry or too moist) is enough to cause a dramatically drop in the visual quality of a produce.

H.-D. Haasis et al., *Dynamics in Logistics*

DOI: 10.1007/978-3-540-76862-3,

Nunes et al. [1] reported that temperature is the characteristic of the distribution environment that has the greatest impact on the storage life and safety of fresh perishables. Effective temperature management is in fact, the most important and simplest procedure for delaying product deterioration [1]. For example, optimum temperature storage retards aging of fruit and vegetables, softening, textural and colour changes, as well as slowing undesirable metabolic changes, moisture loss, and loss due to pathogen invasion. Temperature is also the factor that can be easily and promptly controlled. Preservation of perishables quality and safety can be only achieved when the product is maintained under its optimum temperature as soon as possible after harvest or production.

Understanding more about the whole distribution chain is important. Knowing in advance what it is going to happen is critical. Any systems that can communicate additional information regarding shipment quality and safety aspects before it gets to the distribution centre are essential from a quality assurance perspective. Currently, most digital temperature loggers have to somehow be connected to a host device to download data, and because of this, they have limited 'real-time' data interactivity, and result in after-the-fact analysis for claims, quality analysis and related issues. RFID temperature loggers add wireless communication to read the temperature logger in real-time. The RFID tag, with associated hardware and software will add the benefit of having a pallet/tag scanned on receipt, so that if an alert (alert trigger programmable prior to shipping) is active, the receiver knows immediately (not after-the-fact) that there is a potential problem with the shipment and can spend the time required on specific shipments rather to use random inspections.

This new approach can enable retailers and suppliers to plan the movement of perishable products more efficiently and thus reducing waste and providing better quality products to consumers. A study done by Pelletier et al. [2] demonstrated that knowing the full temperature history from the field to the distribution centre can make a difference of as much as $48 000 of loss of revenue per trailer of strawberries for a retailer. Further study done by Emond et al. [3] demonstrated that for fresh strawberries the full visibility of the temperature of the shipment combined with a shelf life prediction model can bring $13,076 in profit per trailer compared to a potential lost of $2,303 with the actual process.

This full visibility combined with prediction models can allow retailers, food service and restaurant chains to manage their inventories by First to *expire* – first out (FEFO) rather than first in – first out (FIFO). Prediction shelf life models like the ones developed by Nunes et al. [4] can integrate the quality attributes selected by the users and predict which loads should be sent to specific stores or restaurants in order to maximize inventories and quality.

2 Intelligent RFID as enabling technology

One technical challenge for the implementation of the described quality and temperature tracking system is the management of the huge data amount. Especially the

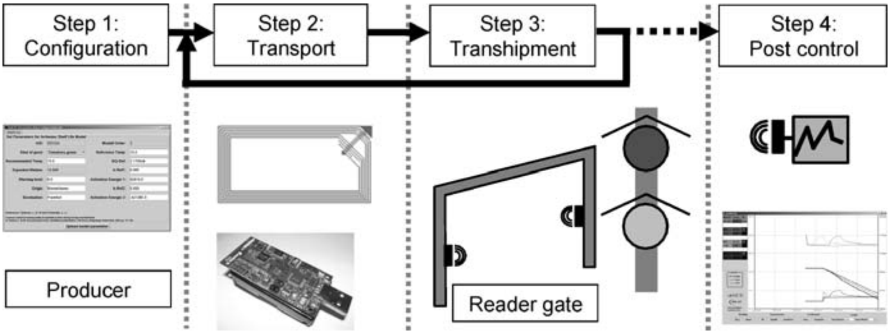


Fig. 1 Transport supervision by intelligent RFID

communication between reader and passive RFID device turned out to be a major bottleneck. Reading full temperature charts from hundreds of RFID data loggers at each transhipment point is hardly possible. On the other hand, for quality based inventory and transport management it is rather sufficient to have an indicator for current quality and pending losses than knowing the detailed temperature history.

The approach of the 'intelligent RFID' circumvents the communications bottleneck and provides the necessary information for management by integration of the shelf life model into the hardware of an RFID temperature logger. After being configured by the producer (Step 1), the intelligent RFID measures the temperature during transport and calculated the change in the product properties per time unit (Step 2). When passing a reader gate at transhipment the RFID tag only transmits the calculated quality state of the attached transport good (Step 3). A traffic light signals whether the product is 'ok' (green) or needs additional checking ('red'). The temperature history of products in a crucial state can be read out by a handheld reader (Step 4) for detailed analyses [5]. The process of transport supervision by intelligent RFID is depicted in Fig. 1. An alternate approach that allows online access to temperature and quality during transport on the costs of higher hardware expenses, the intelligent container (www.intelligentcontainer.com) is described in [6].

3 Modelling approaches

Beside the technological basis, the precise modelling of shelf-life changes is the other key challenge. One very common approach is to describe the quality decay according to the law of Arrhenius for reaction kinetics. The change in shelf life or loss per day is calculated as a function of temperature [6].

These shelf life or keeping quality models have the advantage that only three model parameters have to be estimated by analysing experimental data. Tijskens [7] showed that the sensitivity of some fruits against chilling injuries caused by too low temperatures can be modelled by adding a second Arrhenius type function with negative activation energy, resulting in five parameters in total.

For many food products, a chain of chemical or enzymatic processes contribute to a loss of quality. The approximation by one or two Arrhenius function does not always deliver the required accuracy. For a precise prediction the time function of several quality factors like colour, firmness, mould and vitamin C content, have to be estimated. Because it is not possible to know in advance which of the curves will be the most critical one, all of them have to be continuously calculated and monitored.

Descriptions of the underlying chemical reactions by a set of differential equations are only known for a minority of products, Bobelyn [8] for example describes the colour loss of Mushrooms by a set of 3 or 6 equations.

In many cases only tables for different constant temperatures are available, listing the change of physical properties over time. Examples can be found in [9] and [4]. To extract a set of differential equations from these measurements requires expert chemical knowledge that is in general not available.

The suggested table-shift approach provides a prediction for dynamic temperature condition based on curves recorded at constant temperatures. The value of each physical property for the next step is read out from the pre-measured table. If the temperature changes, the focus in the table is shifted to the row that matches the current quality inside the column for the new temperature. In the mathematical description the change per physical property is calculated as two-dimensional interpolation over the parameters temperature and current property state.

4 Software simulation for the table-shift approach

This table-shift approach was verified by a MATLAB software simulation. The prediction might be distorted by effects of unknown or hidden states that also affect the speed of change. Full accuracy can only be guaranteed, if all underlying reactions have the same relation to temperature or activation energy. In this case a change of temperature can be expressed as a stretch of the time scale.

If the activation energies are close to each other, the table-shift approach will give a good approximation. As example the table-shift model was compared with a 3 differential equation sigmoid model for the browning of mushrooms described by Bobelyn [8]. The reactions that contribute to the oxidation of the colour pigment have activation energies of 74,600 and 65,500 $\mathrm{J\,mol^{-1}}$. In the first step a table for colour change over time was created for four different temperatures by simulation of the differential equations. Figure 2 shows the colour change for an arbitrary temperature function calculated by the differential equations and by interpolation of the tables. The effect of the differences in the activation energy turned out to be minimal.

A more crucial problem are tolerances of the measured table values. Measurements of physical properties like firmness or nutrition content have only a very limited resolution. Taste and aroma are evaluated on a subjective scale. The effects of a measurement resolution limited to 1% or 5% of the maximum value were also tested by the MATLAB simulation in Fig. 2.

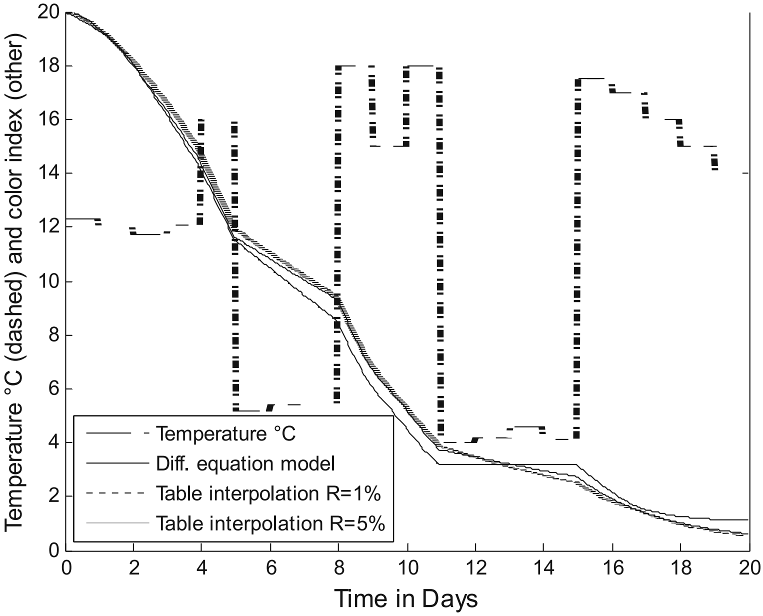


Fig. 2 Matlab simulation of predicted colour loss by differential equation model and table shift model with table resolution of 1% and 5% for an arbitrary temperature input. Colour index scaled to 20 as initial value

To avoid misleading results, the accuracy of the table-shift model has to be verified for each product by simulation for a dynamic temperature curve. Possible deviations caused by different harvest conditions should also be tested.

5 Implementation

Several developments can be found that point into the direction of a low-cost shelf life tag. The Turbo-Tag data logger from Sealed Air/KSW Microtec (www.turbo-tag.com) uses a semi-passive RFID tag equipped with an additional battery to record up to 700 temperature values. A number of manufactures announced UHF RFID temperature recorders with extended reading range.

Some data loggers already include shelf life estimation, but are restricted to Arrhenius type models. The electronic time temperature integrator from CliniSense [10] can be ordered in small quantities, but it does not offer RF communication. The announced Freshtime tag from Infratab is still not available yet.

Wireless sensor networks (WSN) provide the second platform for a shelf life monitoring device beside passive RFID. Ranges of 50 meters or more are reached by active communication. The Tmote Sky from Moteiv (www.moteiv.com) is equipped with a free programmable MSP430 ultra-low-power microcontroller running at

8 MHz clock rate with 16 bit integer arithmetic and hardware multiplier. Drawbacks are the higher price, energy consumption and resulting battery size.

Because currently only WSN offer additional on-board calculation facilities, first tests of different model types were carried out on this platform. Our demonstration system measures the freight temperature and sends a calculated quality index over a 2.4 GHz link to a PC. Three different model types can be uploaded to the sensor notes: The first model calculated losses directly by two Arrhenius functions. The second micro controller application is based on a pre-calculated look up table for the loss per day as function of temperature as proposed in [11]. The tables were generated by evaluation of the same equations that were used in the first model. The table-shift approach with two-dimensional interpolation was tested as third application.

5.1 Required resources

To transfer the model calculation to a semi-passive RFID tag, a new hardware development will be necessary. As preliminary step for construction of an intelligent RFID it should be carefully considered, which model types and level of accuracy are feasible inside a credit card sized label in regard to limited processing power and energy resources.

Quality modelling can be seen as add-on to existing data logger technology, which provides required system components as RFID interface, sensor query, accurate timer and data storage, powered by a paper thin battery. So the question is to determine the additional costs for quality modelling.

To answer this question the required resources for the implementation on the wireless sensor platform were analysed. The required energy per model step was calculated based on the measured execution time of the modelling algorithm. The necessary memory resources for the program in ROM as well as variables and constant tables in RAM were extracted from the compiler log file. The results for the both considered models are summarized in Table 1.

The required energy has to be compared against the battery capacity and the stand-by current of the microcontroller.

The table-shift model consumes about 10 times more energy than the interpolation of a one-dimensional look up table. But the consumption of the table-shift model is still negligible compared to the capacity of small size batteries. If the shelf life of quality state is updated every 15 minutes, 20 mJ are required per modelled

Table 1 Required resources per model step

Type of Resource	Calculation of Arrhenius equations	Look up table for Arrhenius model	Table-Shift Approach
Processing time	1.02 ms	0.14 ms	1.2 ms
Program memory	868 bytes	408 bytes	1098 bytes
RAM memory	58 bytes	122 bytes	428 bytes
Energy	6 μJoule	0.8 μJoule	7 μJoule

physical property and month. A typical button cell battery has a capacity between 300 J and 3000 J. The Zink oxide battery of the credit card sized turbo tag data loggers has a capacity of 80 J.

Design issues for the RFID implementation

A more crucial issue is the stand-by current of the microcontroller. Although the microcontroller does nothing 99.99% of the time, it still consumes energy for the clock oscillator, wake-up circuit and RAM retention. The typical stand-by current of the MSP430 of 1 μA at 2.2 V results in a consumption of 5.7 J per month.

From the point of required resources precise quality prediction including modelling the temperature dependency of several physical properties will be feasible. The challenge is rather to add integer arithmetic capabilities to a semi-passive RFID tag without increasing the stand-by current to much.

6 Summary and outlook

Losses of perishable good can be avoided by managing transport and warehouse planning based on the current quality state. To install this process two elements are necessary: Accurate shelf life models and a way for efficiently measure the freight conditions and deliver this information.

In most cases it is hard to develop a bio-chemical model to describe the change of physical product properties as a function of temperature and time. The table-shift approach provides a method for a direct calculation of the shelf life based on a set of measured reference curves. The accuracy of this approach has to be validated for new product types by comparison of simulated curves for dynamic temperature conditions with measured data.

Intelligent RFID is the most promising concept to capture temperature and forward quality information without trapping into communication bottlenecks. This approach will probably not provide multi-factor quality parameters details like as a main frame system but still can provide rapid information for immediate response at any steps of the distribution chain. Although no hardware solution is currently available, the measurements of the required resources demonstrate that it would be possible to integrate the calculation of a shelf life model into an RFID temperature logger. Limiting factors for the model accuracy are measurement tolerances of the pre-recorded reference curves. Related experiments to connect a low-power microcontroller with an UHF-RFID communication interface are currently carried out by the IMSAS.

Acknowledgements This research was supported by the German Research Foundation (DFG) as part of the Collaborative Research Centre 637 “Autonomous Cooperating Logistic Processes”.

References

1. M.C.N. Nunes, J.P. Emond, K.V. Chau, M. Rauth, S. Dea, and W. Pelletier. Effects of in-store conditions on the quality of fresh fruits and vegetables. Research report to Publix Super Markets. Food Science and Human Nutrition: University of Florida 262 pages, 2006
2. W. Pelletier, J.P. Emond and K.V. Chau. Effects of post harvest temperature regimes on quality of strawberries. Report 1 (2006). University of Florida. 48 pages 2006.
3. J.P. Emond, B. Hardgrave. Quantifying RFID's Cold Chain Benefits. In: Fifth RFID Academic Convocation. Orlando, Florida, April 30th, 2007
4. M.C.N. Nunes, J.P. Emond, J.K. Brecht. Quality curves for highbush blueberries as a function of the storage temperature. *Small Fruits* 3:423–440, 2004
5. R. Jedermann and W. Lang. Semi-passive RFID and beyond – steps towards automated quality tracing in the food chain. *International Journal of Radio Frequency Identification Technology and Applications IJRFITA* in Press
6. R. Jedermann, R. Schouten, A. Sklorz, W. Lang, and O. van Kooten. Linking keeping quality models and sensor systems to an autonomous transport supervision system. In J. Kreyenschmidt and B. Petersen, editors, Cold Chain Management – 2nd international Workshop, Bonn, 8.–9. May 2006, University Bonn, 3–18, 2006
7. L.M.M., Tijskens, J.J. Polderdijk. A generic model for keeping quality of vegetable produce during storage and distribution. *Agricultural Systems*, 51(4):431–452, 2006
8. E. Bobelyn, M.L.A.T.M. Hertog, and B.M. Nicolaï. Applicability of an enzymatic time temperature integrator as a quality indicator for mushrooms in the distribution chain. *Postharvest Biology and Technology*, 42(1):104–114, 2006
9. M.C.N. Nunes, J.P. Emond, J.K. Brecht. Quality of strawberries as affected by temperature abuse during ground, in-flight and retail handling operations. *Acta Hort*. 604:239–246, 2003
10. S.E. Zweig. Advances in vaccine stability monitoring technology. *Vaccine*, 24(33–34): 5977–5985, 2006
11. W. Roberts and J. Cox. Proposal for Standardized Core Functionality in Digital Time-Temperature Monitoring SAL Devices. White Paper by the Temperature Tracking Work Group of the SAL Consortium (Sealed Air Corporation, Cox Technologies), 2003

Sustainable Collaboration

Effects of Autonomous Cooperation on the Robustness of International Supply Networks – Contributions and Limitations for the Management of External Dynamics in Complex Systems

Michael Hülsmann[1], **Bernd Scholz-Reiter**[2], **Christoph de Beer**[2], **Linda Austerschulte**[1]

[1] Management of Sustainable System Development,
Institute for Strategic Competence-Management,
Faculty of Business Studies and Economics, University of Bremen, Germany
[2] Department of Planning and Control of Production Systems, BIBA,
University of Bremen, Germany

Abstract This paper intends to show possible contributions of the concept of autonomous cooperation (AC) to enable logistics management of International Supply Networks (ISN) to improve dealing with external dynamics caused by environmental complexity and dynamics. The concept of AC as one possible approach to cope with external dynamics of ISN will be analysed either from a theoretical and an empirical point of view.

1 Risks of External Dynamics for the Robustness of Complex International Supply Networks

The phenomenon of real-time economy could be described by timely convergence of order, production and distribution of products and services (Tapscott 1999; Siegele 2002). It evolved from the requirement to fulfil consumers' demands marked by diversified preference structures to a shorter time between order and distribution (Herzog et al. 2003). If an organisation wants to meet these demands it has to adapt its structures and processes constantly in order to meet the real-time-changing consumers' requirements and thereby to ensure its existence and survival on the long run (Hülsmann et al. 2006). A possible way to cope with characteristics of a real-time economy might be to intensify business relationships in order to provide the organisation with a sufficient amount of addressable resources, capabilities and ways of distribution (Geoffrion and Powers, 1995). In turn, this implies an understanding of companies which are interwoven in supply chain networks on global markets. They are legally separate, however economically to a greater or lesser extent dependent on each other (Sydow 2002). Hülsmann and Grapp characterized these global networks of the world wide connected flow of management decisions, information and ma-

H.-D. Haasis et al., *Dynamics in Logistics*
DOI: 10.1007/978-3-540-76862-3,

terials as International Supply Networks (ISN) (Hülsmann and Grapp 2005). The shown kind of diversity and changeability in customer demands and corresponding variation in organisational structures are assumed to lead to higher complexity and dynamics for the involved management (Hülsmann and Grapp 2005) that has to deal with these phenomenon (Hülsmann and Berry 2004). According to Dörner the existence of multiple interrelations between the elements of a system as well as interrelations between the system's elements and environmental elements, is a characteristic attribute of a complex system (Dörner 2001, p. 60; Malik 2000, p. 186). As ISN are characterised by multiple internal and external relations an ISN can be understood as a complex system. Another factor that characterises a system is "dynamics" that describes the variation of a systems status over time (Kieser 1974). For ISN-Management especially variations in the environment as to say external dynamics of the system are a critical factor because of their number of relations to their environment and the amount of changes in these relations. External complexity and dynamics could also be reasoned by phenomena like hyper-linking, hypercompetition and hyper-complexity (D'Aveni 1995, p. 46; Tapscott 1999; Siegele, 2002; Hülsmann and Berry 2004). Consequently, robustness in ISN might be endangered by a higher amount of information resulting from complexity and dynamics. Because the capacity to handle information is steady (e.g. due to rigidities of organizational competences (Schreyögg et. al. 2003)) the proportion between the amount of information and the capacity to handle information might deteriorate. That could lead to instability of the system.

Organisational flexibility is needed to enable an organisation to respond to changes in environmental conditions like technological progress (e.g. acquiring new technologies for efficient logistic processes in ISN). It is maintained by opening the borders of the system to absorb external complexity (Sanchez 1993). It is necessary for sustaining the integration of the system in its environment. Organisational stability is needed to compensate the absorbed complexity within the system. It is realized by the closure of the respective system. This leads to a definition of the system's identity which is necessary for the internal integration of all system's elements (Luhmann 1973). For adaptivity and therefore to maintain the existence of the ISN both processes (flexibilisation and stabilisation) are needed and have to be balanced (Hülsmann and Grapp 2006; Hülsmann et al. 2006).

From the perspective of logistics the achievement of goals of logistics-management e.g. time, quantity, quality, price etc. (Plowman 1964) is endangered if the level of information availability decreases because a lack of information leads to sub optimal decision-making (e.g. if not every possible offer can be taken into account the products and services might be to expensive).

To enable an ISN to cope with complexity and dynamics and therefore to achieve the goals of logistics-management the adaptivity of the system has to be ensured by its ability to balance flexibility and stability (Hülsmann and Grapp 2006).

One management approach that is currently discussed to balance the adaptivity in ISN is the concept of autonomous cooperation (e.g. Hülsmann and Grapp 2006). The focus of the concept of autonomous cooperation is the autonomous evolution of ordered structures in complex systems (Hülsmann and Wycisk 2005). Therefore, to apply autonomous cooperation in ISN might increase the ability of logistics management to cope with external risks

According to this, the research question of this paper will be: To what extend might the concept of autonomous cooperation contribute to the management of external risks in ISN? To answer this research question the concept of autonomous cooperation will be introduced and theoretically applied to strategic logistics-management in ISN, in order to deduce theoretical findings about the ability of autonomous cooperation to contribute to the management of external dynamics in ISN. The next research step comprises a simulation based empirical analysis in order to proof the theoretical findings. Finally, conclusions will be drawn which lead to implications for strategic logistics-management in ISN.

2 Autonomous Cooperation as an Approach to Increase the Robustness of ISN

Autonomous cooperation is based on the concept of self-organisation and is currently discussed in different fields of research. The concept of self-organisation has its scientific roots in multiple academic fields like physics, biology or chemistry and belongs to the research field of complexity science in the broadest sense (Hülsmann and Wycisk 2005). The different concepts of self-organisation, for example synergetics (Haken 1973), cybernetics (von Foerster 1960), chaos theory (Peitgen and Richter 1986), autopoiesis (Maturana and Varela 1980), and dissipative structures (Prigogine and Glansdorff 1971) have been points of origin for the concept of autonomous cooperation (Hülsmann et al. 2007). In the 70's a basis for an enfolding interdisciplinary theory could be established, due to the fact that the different concepts of self-organisation have a common background in complexity and order (Hülsmann and Wycisk 2005). The aim of this interdisciplinary field of research is to explain and identify how complex system create ordered structures autonomously (Hülsmann and Wycisk 2005). To elucidate the correlations between self-organisation and autonomous cooperation the concepts could be categorized in the following way: Self-organisation as a part of management describes the way in which complex systems autonomously create emergent order in structures and processes (Bea and Göbel, 1999; Probst 1987). Autonomous cooperation has a more narrow perspective than self-organisation. It describes processes of decentralized decision making in heterachical structures. To create a common understanding of autonomous cooperation for this research paper a general definition for autonomous cooperation, which has been formed by Windt and Hülsmann, shall be used:

> "Autonomous Control describes processes of decentralized decision-making in heterarchical structures. It presumes interacting elements in non-deterministic systems, which possess the capability and possibility to render decisions. The objective of Autonomous Control is the achievement of increased robustness and positive emergence of the total system due to distributed and flexible coping with dynamics and complexity." (Windt and Hülsmann 2007, p. 8).

According to this definition main characteristics of autonomously cooperating systems are for example autonomy, interaction and non-determinism. To answer the question whether autonomous cooperation is able to contribute to strategic logistic-management of ISN the different attributes can be analysed

A system or an individual is **autonomous** if it forms, guides and develops itself. It is operationally closed, meaning that its decisions, relations and interactions are not dependent on external instances (Probst 1987). As there is no system completely operational closed it is spoken about relative autonomy of a system or an individual in relation to certain criteria (Varela 1979, Malik 2000, Probst 1987). According to theory, an organisation has to have a suitable degree of autonomy to be able to form and develop itself on the one hand and to maintain its identity on the other hand. Therefore a suitable degree of autonomy might preserve the ISN from information overload because it is able to balance between stability and flexibility (e.g. the ability of a load carrier to decide for itself what to transport).

The second attribute is **interaction**. Haken (1987, p. 132) states that open dynamic and complex systems develop a self-organized order through various interactions of the systems individual elements. Due to this processes the system develops new qualitative characteristics through emergence (Haken 1993). The emergent structures might contribute to the capacity for information handling because new structures and processes are developed (e.g. new ways of decision-making by intelligent objects). If the capacity to handle external dynamics rises, the risk of sub optimal decision-making due to information overload falls.

Non-determinism is another characteristic of autonomous cooperation. Non-determinism implies that the behaviour of a system can not be predicted over a longer period of time. With the characteristic of non-determinism, autonomous cooperation aims at higher efficiency for dealing with complexity and uncertainty within processes (Windt and Hülsmann 2007, p. 10). For the management of ISN non-determinism might imply that the processing of information can be handled more flexible. It enables the system to react to changes in the structure of ISN and the resulting problems. For the ISN-Management this could mean that its capacity to cope with complexity and dynamics is increasing. Therefore autonomous cooperation might contribute to the robustness of an ISN because the capacity to handle external dynamics might be increased.

3 Empirical Analysis

To measure the effects of autonomous cooperation on the robustness of ISN a simulation and measurement system has been developed that enables to compare the external risks in ISNs and SCs. In earlier work of the authors, a shop floor scenario has been analysed (Hülsmann et al. 2006). A similar approach will be used to simulate the behaviour of an ISN. Figure 1 shows the different characteristics of ISNs and SCs in order to show the different levels of complexity and resulting external dynamics. The scenario shows a simple supply chain that consist of one source (e.g.

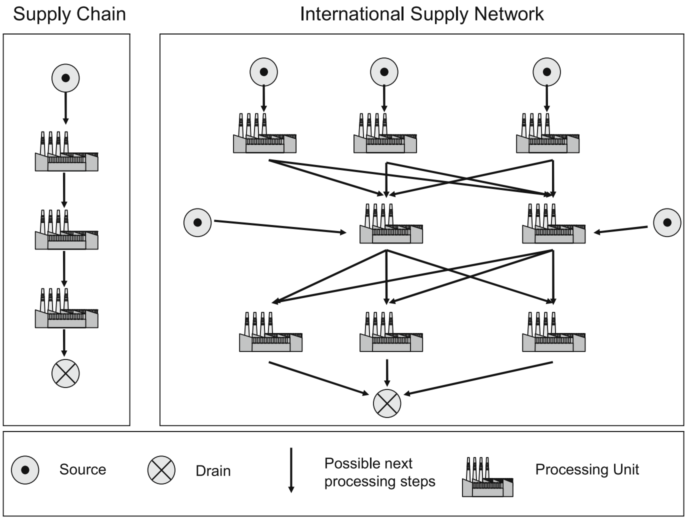


Fig. 1 Different organisational level of complexity

information, resources, orders) and a multi-stage production process. On the other hand, an ISN is depicted that in comparison to the SC consists of diverse sources and interlinked and substitutable production stages. In this model, the supply chain has the lowest level of complexity.

This model inherits the opportunity to evaluate the systems ability to cope with different amounts of external dynamics using different levels of autonomous cooperation. The orders enter the system at the sources and leave it finalised at the drain. Each order has a specific processing plan i.e. a list of processing steps that have to be undertaken to produce goods. In the model, the orders are not directed by a centralised control entity but have the ability to render decisions on their next processing step autonomously by using different concepts of autonomous cooperation. Depending on the different autonomous control methods, the overall system shows altered behaviour and dynamics.

In the following, the applied autonomous control methods will be described. The first method called **queue length estimator** compares the actual buffer level at all parallel processing units that are able to perform the next production step. Therefore, the buffer content is not counted in number of parts but the parts are rated in estimated processing time and the actual buffer levels are calculated as the sum of the estimated processing time on the respective machine. When a part has to render the decision about its next processing step it compares the current buffer level i.e. the estimated waiting time until processing and chooses the buffer with the shortest waiting time (Scholz-Reiter, et al., 2005). The **pheromone method** does not use information about estimated waiting time, i.e. information about future events but uses data from past events. This method is inspired by the behaviour of foraging ants that leave a pheromone trail on their way to the food. Following ants use the pheromone trail with the highest concentration of pheromone to find the shortest path to food. In the simulation this behaviour is imitated in a way that whenever a good leaves a processing unit, i.e. after a processing step is accomplished, the good leaves information about the duration of processing and waiting time at the respective processing unit. The following parts use the data stored at the machine to render the decision about the next production step. The parts compare the mean throughput times from parts of the same type and choose the machine with the lowest mean duration of waiting and processing. The amount of data sets that are stored define the up-to-datedness of the information. This number of data sets can be used to tune the pheromone method. The replacement of older data sets resembles to the evaporation of the pheromone in reality (Scholz-Reiter, et al., 2006). The **due date method** is a two-step method. When the parts leave a processing unit they use the queue length estimator to choose the subsequent processing unit with the lowest buffer level. The second step is performed by the processing units. The due dates of the parts within the buffer are compared and the part with the most urgent due date is chosen to be the next product to be processed (Scholz-Reiter et al. 2007).

The following simulation analyses the overall systems ability to cope with rising structural complexity and rising external dynamics using different autonomous control methods. At each source the arrival rate is set as a periodically fluctuating function. Therefore, the external dynamics rises with increasing complexity of the ISN. The logistical goal achievement is measured using the key figure throughput time for different levels of complexity and different autonomous control methods.

Figure 2 shows the results i.e. the mean throughput times for the three different autonomous control methods in dependence of the systems complexity. The left side corresponds to the supply chain consisting of three processing units. To the right of

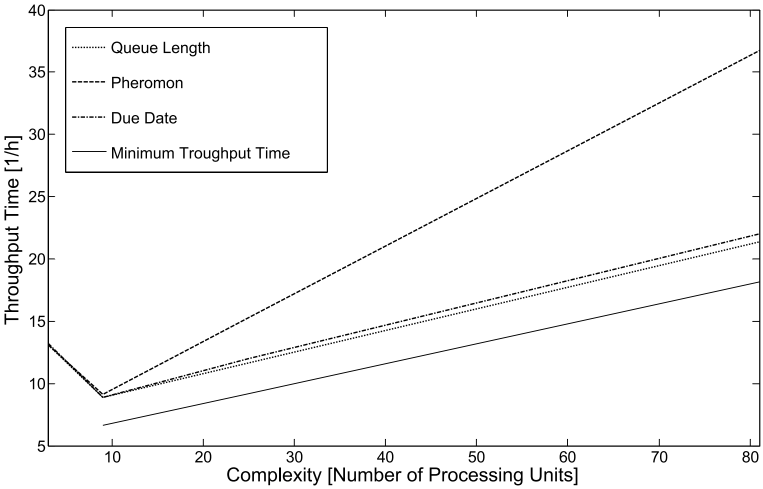


Fig. 2 Logistical goal achievement for different organisational level of complexity and multiple autonomous control methods

the figure the systems complexity is increased by enlarging the amount of processing units as well as the number of sources. Furthermore, the minimal throughput time, which is rising with increasing complexity level, is shown. In the first part of the figure, the throughput time declines for all three curves. This is caused by the fact that parallel processing units can be used in situations of overload i.e. if the arrival rate is higher than the capacity of the processing unit. In the second part it is shown that the curves for the due date method and the queue length estimator show almost the same results, the due date method shows a slightly worse performance because of sequence reordering, while the pheromone method shows inferior goal achievement. The first two curves are almost parallel to the minimal throughput time. This means that a constant logistical goal achievement is achieved during rising complexity and rising external dynamics. The pheromone method shows an inferior behaviour. In this scenario, dynamics is too high and the boundary conditions change faster than the pheromones are updated.

In a second simulation, external dynamics is varied to determine the system's robustness i.e. the system's ability to cope with external dynamics without beeing unstable and to become a locked-organisation. In this simulation the system is called unstable if one of the systems parameters increases without restraint. To determine this boundary of stability, the mean arrival rate at all sources has been increased and the highest possible arrival rate before the system starts to be unstable is measured.

Figure 3 shows the results for the three different autonomous control methods. It has been found that the pheromone method shows declining robustness with rising complexity and dynamics while the other two control methods lead to increasing robustness, with a faintly better robustness of the Queue length estimator. The reason for this rising robustness is the higher amount of parallel processing units that are available for the autonomous objects to use in case of overload.

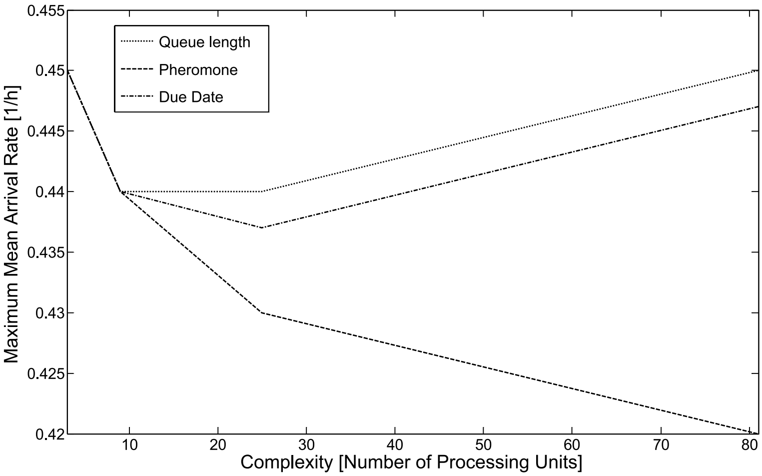


Fig. 3 Robustness measured as the maximum mean arrival rate for different organisational level of complexity and multiple autonomous control methods

The due date method and the queue length estimator show a small difference because only few changes in the sequence occurred and the mean values were not highly affected.

Both simulations have shown that for this kind of scenario, representing different levels of external dynamics, the queue length estimator is the appropriate autonomous control method. Because of the fast changing dynamics, the pheromone method has not been able to adapt to the changed boundary conditions. For the ability of autonomous cooperation to cope with rising external dynamics this means that the adequate autonomous control method has to be chosen depending on the scenarios parameters like complexity, dynamics and reliability of information. If this selection is done properly autonomous cooperation is a possible approach to cope with external dynamics.

4 Conclusions

The aim of this paper was to examine the contributions and limitations of autonomous cooperation on the robustness of ISN especially for the management of external dynamics. From a theoretical perspective it has been outlined, that autonomous cooperation might enable an ISN to enhance its capacity of information handling. For ISN-Management this can lead to a larger capacity to cope with external dynamic that results from changing organisational structures. This implicates that the employment of methods of autonomous cooperation might increase the robustness of the organization. Empirically, it has been shown that autonomous cooperation could be one alternative to cope with rising external risks in ISN but that it depends on the organisational structure and the scenario which level of autonomous cooperation is adequate. In the shown case, the queue length estimator has been the best degree of autonomous cooperation to cope with rising complexity and dynamics. For achieving progress in research other concepts of autonomous cooperation could be analysed. For logistics-management practice in ISN the findings imply that there are concepts of autonomous cooperation like the queue length estimator that might increase the quantity of external dynamics the organisation can cope with but that there is no general method of autonomous cooperation that can be applied in every scenario.

Acknowledgements This research was supported by the German Research Foundation (DFG) as part of the Collaborative Research Centre 637 “Autonomous Cooperating Logistic Processes – A Paradigm Shift and its Limitation”.

References

Bea FX, Göbel E (1999) Organisation – Theorie und Gestaltung. Lucius & Lucius, Stuttgart

D’Aveni RA (1995) Coping with Hypercompetition: Utilizing the New 7S’s Framework. Academy of Management Executive, Vol. 9, No. 3: 45–57

Dörner D (2001) Die Logik des Misslingens: Strategisches Denken in komplexen Situationen. Rowohlt, Hamburg

von Foerster H (1960) On Self-Organizing Systems and their Environment. In: Yovits MC & Cameron S (eds.) Self-Organizing Systems. Pergamon Press, London

Haken H (1973) Synergetics: cooperative phenomena in multi-component systems. In: Proc. of the Symposium on Synergetics. Teubner, Stuttgart

Haken H (1993) Synergetik: Eine Zauberformel für das Management?, In: Rehm W (ed.): Synergetik: Selbstorganisation als Erfolgsrezept für Unternehmen; ein Symposium der IBM, Stuttgart. pp. 15–43

Herzog O et al. (2003) Antrag auf Finanzierung des Sonderforschungsbereichs 1980 "Selbststeuerung logistischer Prozesse – ein Paradigmenwechsel und seine Grenzen", Bremen

Hülsmann M, Berry A (2004) Strategic Management Dilemma – Its Neccessity in a World of Diversity and Change. In Wolff R et al. (eds) Conference Proceedings of the SAM/IFSAM VII World Congress: Management in a World of Diversity and Change. 5–7 Juli 2004. Göteborg, Schweden

Hülsmann M, Grapp J (2005) Autonomous Cooperation in International-Supply-Networks – The Need for a Shift from Centralized Planning to Decentralized Decision Making in Logistic Processes. In Pawar KS et al. Conference Proceedings of the 10th International Symposium on Logistics (10th ISL) – Innovations in global Supply Chain Networks. 3–5 Juli 2005. Lissabon, Portugal, pp. 243–249

Hülsmann M, Grapp J (2006) Monitoring of Autonomous Cooperating Logistic Processes in International Supply Networks. In: Pawar, K.S et al (eds) Conference Proceedings of 11th International Symposium on Logistics (11th ISL). Loughborough, United Kingdom, pp. 113-120

Hülsmann M et al. (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) Proceedings of the International Conference on Complex Systems (ICCS 2006). Boston, MA, USA, 8 pages

Hülsmann M, Wycisk C (2005) Unlocking Organizations through Autonomous Co-Operation – Applied and Evaluated Principles of Self-Organization in Business Structures. The 21th EGOS Colloquium: Unlocking Organizations. June 30 – July 02 2005, Berlin (Web-publication)

Hülsmann M, Wycisk C (2005b) Contributions of the concept of self-organization for a strategic competence-management. In: Sanchez R et al. (eds) The Proceedings of the 7th International Conference on Competence-Based Management (ICCBM): Value Creation through Competence-Building and Leveraging. June 2 – 4, 2005. Antwerp, Belgium

Kieser A (1974) Organisationsstruktur. In: Schmalenbachs Zeitschrift für betriebswirtschaftliche Forschung, Vol. 25, No. 9: 569–590

Lambert DM et al. (1998) Supply Chain Management: Implementation Issues and Research Opportunities. The International Journal of Logistics Management, Vol. 9, No. 2: pp. 1–19

Luhmann N (1973) Zweckbegriff und Systemrationalität. In Series Suhrkamp-Taschenbuch Wissenschaft, Vol. 12, Frankfurt am Main

Malik F (2000) Strategie des Managements komplexer Systeme: Ein Beitrag zur Management-Kybernetik evolutionärer Systeme. Haupt, Bern

Maturana HR, Varela F (1980) Autopoiesis and cognition: the realization of living. Reidel, Dordrecht

Peitgen HO, Richter RH (1986) The Beauty of Fractals: Images on Complex Dynamical Systems. Berlin, Springer

Plowman EG (1964) Lectures of Elements of Business Logistics. Stanford University, Graduate School of Business

Porter ME (1995) Wettbewerbsstrategie – Methoden zur Analyse von Branchen und Konkurrenten. Campus, Frankfurt et al.

Porter ME (1999) Wettbewerbsvorteile – Spitzenleistungen erreichenen und behaupten. Campus, Frankfurt et al.

Prigogine I, Glansdorff P (1971) Thermodynamic Theory of Structure, Stability and Fluctuation. Wiley, London et al.

Probst GJB (1987) Selbstorganisation, Ordnungsprozesse in sozialen Systemen aus ganzheitlicher Sicht. Parey, Berlin et al.

Reimann B. (1995) Rationalität. In: Fuchs-Heitz, W.: Lexikon zur Soziologie, Westdeutscher Verlag, Opladen

Sanchez R (1993) Strategic Flexibility, firm organization, and managerial work in dynamic markets – A strategic options perspective. In: Advances in Strategic Management 9, pp. 251–291

Scholz-Reiter B et al. (2005) Modelling Dynamics of Autonomous Logistic Processes: Discrete-event versus Continuous Approaches. In: Annals of the CIRP, Vol. 55, No. 1: pp. 413–417

Scholz-Reiter B et al. (2006) Modelling and Simulation of a Pheromon based Shop Floor Control. In: Cunha, P, Maropoulos, P (eds.): Proceedings of the 3rd International CIRP Sponsored Conference on Digital Enterprise Technology - DET2006. University of Setubal, Setubal

Scholz-Reiter B et al. (2007) Analysing the Dynamics caused by Autonomously Controlled Logistic Objects. In: Proceedings of the 2nd international conference on changeable, agile reconfigurable and virtual production – CARV 2007. To appear

Schreyögg G et al. (2003) Organisatorische Pfade – Von der Pfadabhängigkeit zur Pfadkreation? In: Schreyögg G, Sydow J (eds.) Managementforschung 13; 257–294

Seebauer P (2003) Logistikentwicklungen – Nach-denken lohnt sich. In: Logistik Heute Vol. 4: pp. 62–63

Siegele L (2002) How about now? A survey of the real-time economy. In: The Economist, Vol. 362: pp. 18–24

Sydow J (2002) Zum Management von Logistiknetzwerken. In: Logistik Management, Vol.4

Tapscott D (1999) Creating value in the network economy. In: Harvard Business School Press, Boston

Varela FJ (1979) Principles of biological autonomy. North Holland, New York

Windt K, Hülsmann M (2007) Changing Paradigms in Logistics – Understanding the Shift from Conventional Control to Autonomous Co-operation and Control. In: Hülsmann M, Windt K (eds.) Understanding Autonomous Cooperation & Control – The Impact of Autonomy on Management, Information, Communication, and Material Flow. Springer, Berlin

Sustainability and Effectiveness in Global Supply Chains: Toward an Approach Based on a Long-term Learning Process

Bernd Scholz-Reiter, Enzo Morosini Frazzon

Bremen Institute of Industrial Technology and Applied Work Science (BIBA), Hochschulring 20, 28359 Bremen, Germany

Abstract The industry has undergone significant changes in recent decades. Today, it is no longer home-based and operates in a global market. The integration of demand and supply has widened the strategic relevance of logistic networks, processes and systems. This integration along global supply chains depends on trustful partnership, international and hence cross-cultural cooperation. The design, plan and management of logistics are particularly complex because of the diversity of contexts, organisations and individuals involved. Therefore, suitable long-term learning process should taken into account agents' role and background, the context in which it takes place as well as intra- and inter-organisational culture. In fact, enhancing competitive advantage on the basis of both overlap and diversity of knowledge is an important capability within today's supply chains. On this foundation, the present paper aims to address business scenario and logistic systems from a viewpoint that approaches competitive advantage in terms of sustainability and effectiveness. Some aspects impacting on dynamics in logistics will be circumscribed in agreement with a holistic perspective embracing context, organisation and individual levels. A descriptive model and further steps toward an approach based on a long-term learning process will be then introduced.

1 Introduction

The manufacturing and service industry have changed in recent decades. A relevant side of this evolutionary and dynamical process has been the pursuit for new and attractive markets as well as the search for high quality and low cost sources of raw materials, parts and products. Both fronts have ceaselessly boosted the globalisation of commerce and production, and made material and information flows more dynamic – due to vibrant demand and supply – and structurally complex. Therefore, the enhancement of processes and systems that fit and connect these fronts should be aspired. Nevertheless, this connection must be consistent with firms' strategy which delineates how activities and processes are planned, coordinated and per-

H.-D. Haasis et al., *Dynamics in Logistics*
DOI: 10.1007/978-3-540-76862-3, © Springer 2008

formed. Business strategy also deals with the allocation of finite resources (Porter 1996) which should be, in order to generate unique competitive advantage: valuable, in the sense that they exploit opportunities and/or neutralises threats in a firm's environment; rare among current and potential competition; imperfectly imitable – either through specific historical conditions, causal ambiguity, or social complexity; and without strategically substitutes (Barney 1991). Moreover, the adoption of business strategies that meet the needs of the enterprise and its stakeholders today while sustaining the resources that will be needed in the future (IISD 1992) is a concurrent business obligation. In fact, business networks have to be effective and, at the same time, to sustain their economic, ecological and social resource base (Ehnert et al. 2006).

The present paper aims to address business scenario and logistic systems from a viewpoint that approaches competitive advantage in terms of sustainability and effectiveness. Some aspects impacting on dynamics in logistics will be circumscribed in agreement with a holistic perspective embracing context, organisation and individual levels. A descriptive model and further steps toward an approach based on a long-term learning process will be then introduced.

2 Logistic Systems

The connection between markets and sources, demand and supply has increased the strategic relevance of logistics. In this broad context, the strategic integration of logistic systems' diverse components is a requirement to achieve effectiveness and sustainability in global supply chains. Customer expectations, the pressure of competition on turbulent global markets and virtualisation of logistics companies result in complex and dynamic logistics systems, structures and networks (Scholz-Reiter et al. 2004). Moreover, logistic management includes the configuration and coordination of structures and processes of logistic networks focused on strategic, tactical and operational goals (Kopfer 2000).

In fact, the increasing dynamic and structural complexity within logistics networks – caused by changing conditions in the markets and the increasing importance of customer orientation and individualisation – challenges logistic processes and systems (Arndt and Müller-Christ 2005). Furthermore, contributing to such challenging scenario, activities, processes and structures of a number of organisations are interwoven worldwide, where the management has to deal with multiple interrelations between actors situated in distinctive cultural, economic, political and social environments (Hülsmann et al. 2006). On this scenario, people and companies' singularities (e.g. social and cognitive aspects) may represent barriers or boosters and therefore need to be better understood. In fact, collaboration and the environment in which it takes place is a key area in which processes and systems meet head on with people of all shapes, sizes, backgrounds and intellectual capability (EU 2006).

Each firm might be understood as a system derived and evolved from contextual, organisational and individual influences (Fig. 1). For instance, agents in global sup-

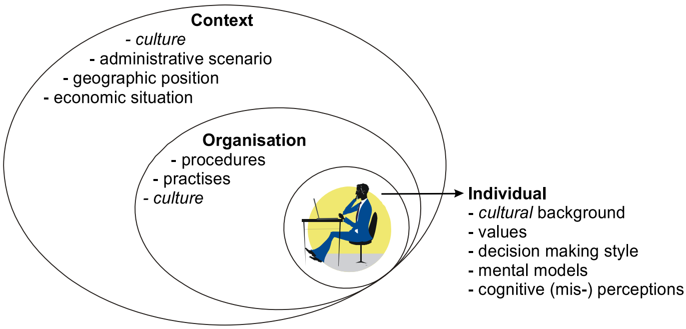


Fig. 1 Systemic perspective

ply chains have varied cultural background as well as values, decision making style, mental models and cognitive perceptions.

In fact, people at various organisational levels, with diversified culture and background, are responsible for the design, plan, management and execution of logistic activities and process. The variety of existing mental models, deeply ingrained assumptions, generalizations that influence how people take actions (Senge 1990) is closely related to cultural environment and background. So, the strategic integration along global supply chains depends on trustful partnership, and is closely dependent on international and hence cross-cultural cooperation.

Suitable long-term (i.e. double-loop) learning process should taken into account agents' role and background, the context in which it takes place as well as intra- and inter-organisational culture. Dense, informal networks and strongly embedded gatekeepers increased the effective flow of information and recombinant potential of the organisation (Fleming 2001). Furthermore, internationalisation creates a need to know how managers in different parts of the world make decisions and how differences in their socialization as well as business environments may affect both their decision-making processes and the choice that they make (Martinsons and Davison 2006).

Commenting current business scenario, Forrester (1998) vigorously argue that attention must be placed on systems and structures design as well as on policy planning, instead of purely on managing contingencial day-to-day decision making. This translates perfectly the holistic view fostered here. System thinking is an iterative learning process in which a reductionist, narrow, short-run, static view of the world is replaced with a holistic, broad, long-term, dynamic view, reinventing our policies and institutions accordingly (Sterman 2006). On this basis, systems thinking can help understand this system by revealing which underlying structures exist, how complex problems are generated and which/how factors influence them over time (Senge 1990) and space. To subsidise the following analysis, a logistic system and its distinctive yet interlaced levels will be illustrated as an interactive system that extend themselves to the contextual level of a national economy (Forrester 1958; Peck 2006). Logistic systems (Fig. 2) could be represented as linked levels and inter-

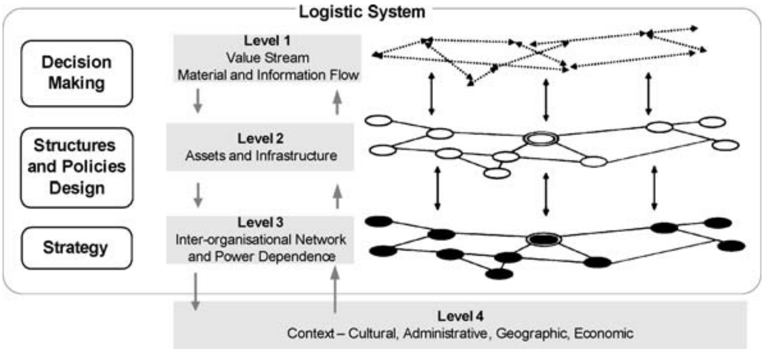


Fig. 2 Logistic system (adapted from Peck 2005)

dependent universe of actions, i.e. strategy, structures and policies design as well as decision making. In fact, the discrete levels of analysis are inextricably attached elements of the logistic system. Together they cover elements of a logistic system and the context within they are embedded, though each level reflects quite different perspectives (Peck 2005).

Specifically regarding to a logistic network, competitive advantage and a long-term evolution can be created by developing knowledge that is embedded in the cooperation context and thus hard to imitate (Hülsmann et al. 2005). In line with paper's aims, logistic systems integration in current fast-moving scenario as well as the impact of diversity on competitive advantage will be tentatively circumscribed through a descriptive model.

3 Logistic Systems' Potential Absorptive Capacity

The goal of knowledge management is to support the agent in improving its decisions in the presence of incomplete, imprecise, or debatable information as well as the inherent uncertainty that results from the dynamic of the domain (Langer et al. 2006). It is also important to note that knowledge management is restricted by various sociological and technological boundaries (Langer et al. 2006). Usually it is claimed that organisations need to develop cultures where their members are encouraged to share knowledge in order to gain a strategic advantage. Nevertheless, the impact of contextual and organisational culture (Gray and Densten 2005) in knowledge management is strategic and deserves further research.

The absorptive capacity construct represents the firm's ability to acquire, assimilate, transform and exploit knowledge to create value through innovation (Cohen and Levinthal 1990; Zahra and George 2002). The authors studied how knowledge sharing and diversity across individuals affects the development of organisational absorptive capacity, concluding that both overlap (to enable internal communication) and diversity (to potentialize the comprehension of outside information) of knowledge structures are beneficial and needed. The concept emphasises that successful knowledge identification and acquisition require the receiver to possess suitable knowledge base and skills to assimilate new information.

Logistic systems integration in current fast-moving scenario will be circumscribed through a descriptive model based on the potential absorptive capacity construct. Potential absorptive capacity (adapted from Zahra and George 2002) for logistic system's (Fig. 3) refers specifically to system's ability to recognize the value of new information, assimilate and integrate it in order to be applied to commercial ends. It is segmented in identification and obtainment (influenced by awareness), and integration (double-loop/long-term learning).

External knowledge comprises a broad range of elements, mainly external data, information, tacit and explicit knowledge as well as experience. Currently, it has on one hand gained amplitude and scale; and, on the other hand, evolved in a structurally complex and dynamic fashion. Internal and external activation triggers are events that motivate and compel the system to respond. Moreover, individuals have central role in influencing search patterns and awareness, activation of triggers as well as knowledge integration.

Furthermore, innovation and the concomitant evolution of social and physical technologies coordinated under evolving business plans (Beinhocker 2006) play a central role in today's economy and can be viewed as complex processes arising out of systemic interaction between actors and stakeholders involved in the production, diffusion and use of new and economically-useful knowledge (Nelson 1993).

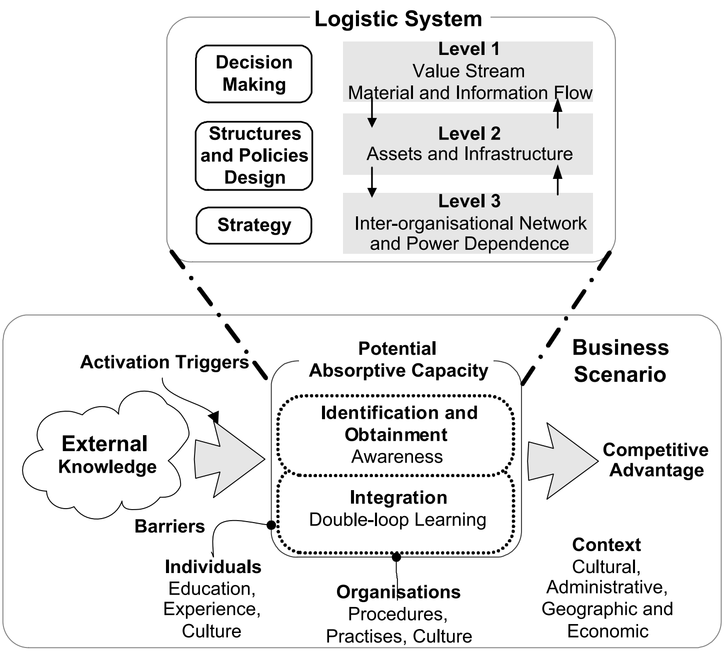


Fig. 3 Logistic Systems' Potential Absorptive Capacity – Descriptive model

This approach sees innovation as an iterative matching of technical possibilities to market opportunities, through both market and non-market interactions, feedbacks and learning processes throughout research and development (R&D), production and logistic networks, rather than as a one-way, linear flow from R&D to new products (Foxon et al. 2005). In fact, innovation in an interwoven world impregnated with external knowledge, characterised by evolving and cyclical opportunities and threats, is distinctly challenging.

On a congruent fashion, it has been empirically demonstrated that the element of individual diversity is a decisive driver for innovation (Fleming 2004). In a case study in a manufacture firm, Fleming (2001) approached innovation breakthrough as recombinant and boundedly rational search process, involving many inventive trials, mix and juxtapose of diverse technologies, professions and experience. Moreover socio-psychological research (Nemeth 1986) demonstrated that exposure to a diversity of opinions encourages divergent thinking and generates more creative solutions. The recombinant search concept can also be applied to the creation of novelty in other contexts (Nelson and Winter 1982) and/or marketplaces.

One of today's most important sources of diversity, culture is a collective programming of the mind which distinguishes one collectivity of people from another (Hofstede 1991). Culture as shared meaning, understanding, values, belief systems, or knowledge depends upon both community and diversity (Hatch 1997). Hofstede (1984) argue that:

> In fact, management deals with a reality that is man-made, people build organisations according to their values, and societies are composed of institutions and organisations that reflect the dominant values within their culture. Management within a society is very much constrained by its cultural diversity and context, because it is not effective to coordinate the actions of people without understanding their values, beliefs, social norms and expressions.

The existence of diverse perceptions of reality shaped by different context and organisational cultures (Hülsmann et al. 2005) may be a reason for the existence of cultural, inter-organisational learning barriers. Analogically, the interaction between agents and stakeholders in the challenging logistic domain may represent an interesting opportunity to address systems' structures and policies. Finally, it is argued that the proposed descriptive model would help to enhance the understanding about the referred logistic system.

4 Preliminary Conclusions and Prospective Research

Through a holistic perspective embracing context, organisation and individual levels, the present paper aimed to address business scenario, logistic systems and correlated evolving competitive advantage. In fact, manufacturing and logistics within global supply chains are facing new challenges and, to cope with them, they have to consider and deal with influences coming from those levels. Furthermore, the strategic and systemic integration of logistics diverse agents might influence systems' competitive advantage.

On this broad context, a model was employed to describe the argued integration perspective. Logistic systems' potential absorptive capacity represents one relevant characteristic in current dynamic marketplaces where abundance of information does not necessarily leads to strategic and valuable knowledge. It could be also concluded that diversity (e.g. cultural) apart from the potential positive influence on awareness also brings new challenges to logistic systems with impacts on complexity and dynamics. In fact, enhancing competitive advantage on the basis of both overlap and diversity of knowledge is an important capability within today's supply chains. Finally, it is here acknowledged the empirical limitations of current research, based mostly on exploratory literature review and on personal experiences in cross-cultural environments.

Further research steps toward an approach based on a suitable long-term learning process, aiming to enhance knowledge and information transferring between supply chain partners in cross-cultural contexts, will be constituted by: (i) deepening of the descriptive model in order to approach specific causal interrelations within logistic systems; (ii) proposal of an evaluation framework to assess potential impacts on sustainability, effectiveness and innovativeness; and (iii) development of a business application prototype.

References

Arndt, L., Müller-Christ, G.: Robustness in the Context of Autonomous Cooperating Logistic Processes: A Sustainability Perspective. In: Haasis, H.D., Kopfer, H., Schönberger, J. (eds.): Operations Research Proceedings 2005. Selected Papers of the Annual International Conference of the German Operations Research Society (GOR), Springer, Berlin, pp. 67–72

Barney, J. (1991) Firm Resources and Sustained Competitive Advantage, Journal of Management, 17, 99–120

Beinhocker, E.D. (2006), The origin of wealth: evolution, complexity, and the radical remaking of economics, Harvard Business School Press, Boston

Bowersox, D.J., Daugherty, P.J. (1995) Logistics paradigms: The impact of information technology. Journal of Business Logistics

Camarinha-Matos, L.M.; Afsarmanesh, H. (2006): Collaborative Networks: Value creation in a knowledge society. In Proceedings of PROLAMAT'06 (Springer) – Shanghai, China, 2006

Cohen, W.M.; Levinthal, D.A. (1990). Absorptive capacity: A new perspective on learning and innovation. Administrative Science Quarterly 35: 128-152

Ehnert, I., Arndt, L., Mueller-Christ, G. (2006). A sustainable management framework for dilemma and boundaries in autonomous cooperating transport logistics processes, Int. J. Environment and Sustainable Management, Vol. 5, No. 4, pp.355–371

EUROPEAN COMMISSION (2006): ICT – Information and Communication Technologies: A Theme for research and development under the specific programme "Cooperation" implementing the Seventh Framework Programme (2007-2013) of the European Community for research, technological development and demonstration activities. Brussels, 2006

Fleming, L. (2001). Finding the organizational sources of technological breakthroughs: the story of Hewlett-Packard's thermal ink-jet. Industrial and Corporate Change, Vol. 11, Number 5, pp. 1059-1084

Fleming, L., 2004, Perfecting Cross-pollination. Harvard Business review, September 2004

Forrester, J. W. (1998), Designing the Future. Presentation by Jay W. Forrester at Universidad de Seville, Spain, December 15, 1998

Forrester, J.W. (1958), Industrial dynamics: a major breakthrough for decision makers. Harvard Business Rev., 38, 37–66

Foxon, T.J., Pearson, P., Makuch, Z. and Mata, M. (2005), Transforming policy processes to promote sustainable innovation: some guiding principles – Report for policy-makers, ISBN 1 903144 02 7, Imperial College London

Gray, J.H. and Densten, I.L. (2005). Towards an Integrative Model of Organisational Culture and Knowledge Management. International Journal of Organisational Behaviour, Volume 9(2), 594-603

Hatch, M. J. (1997), Organization Theory, Oxford University Press

Hofstede, G. (1984), Cultural Dimensions. In: Management and Planning, Asia Pacific Journal of Management, 81–99

Hofstede, Geert. Culture's Consequences. Newbury Park: Sage, 1991

Hülsmann, M., Lohmann, J., Wycisk, C. (2005): The Role of Inter-Organisational Learning and Self-Organising Systems in Building a Sustainable Network Culture. In: International Journal of Knowledge, Culture & Change Management, 5, 2, pp. 21-30

Hülsmann, M.; Grapp, J.; Li, Y.: Strategic Flexibility in Global Supply Chains – Competitive Advantage by Autonomous Cooperation. In: Pawar, K.S. et al. (eds.): Conference Proceedings of 11th International Symposium on Logistics (11th ISL). Loughborough, United Kingdom, 2006, pp. 494-502

INTERNATIONAL INSTITUDE FOR SUSTAINABLE DEVELOPMENT (1992): Business Strategy for Sustainable Development, 19 pages, Copyright: IISD, ISBN 1-895536-00-6, Winnipeg, Canada, 1992

Kopfer, H.; Giesa, F. (2000): Management logistischer Dienstleistungen der Kontraktlogistik. Logistik Management, 1/2000, pp. 43-53

Langer, H.; Gehrke, J.D.; Hammer, J.; Lorenz, M.; Timm, I.J.; Herzog, O.: A Framework for Distributed Knowledge Management in Autonomous Logistic Processes. In: International Journal of Knowledge-Based & Intelligent Engineering Systems, 10(2006)4, pp. 277-290

Martinsons, M.G. and Davison, R.M. (2006). Strategic decision making and support systems: Comparing American, Japanese and Chinese management, Decision Support Systems (2006), doi:10.1016/j.dss.2006.10.005

Nelson, R. and Winter S. (1982), An Evolutionary Theory of Economic Change. Belknap Press: Cambridge, MA

Nelson, R. (1993), National Innovation Systems: A comparative analysis, Oxford University Press, New York

Nemeth, C. (1986), Differential contributions of majority and minority influence, Psychological Review, 93, 23-32

Peck, H. (2005), Drivers of supply chain vulnerability: an integrated framework, International Journal of Physical Distribution Logistics Management, 35, 210–232

Peck, H. (2006), Reconciling supply chain vulnerability, risk and supply chain management. International Journal of Logistics: Research and Applications Vol. 9, No. 2, June 2006, 127–142

Porter, M. E. (1996), What is Strategy? Harvard Business Review, Nov/Dec 1996, Vol. 74 Issue 6.

Scholz-Reiter, B., Windt, K. and Freitag, M. (2004) 'Autonomous logistic processes: new demands and first approaches', in L. Monostri (ed). Proceedings of the 37th CIRP International Seminar on Manufacturing Systems, Budapest, Hungaria, pp. 357–362

Senge, P. (1990). The fifth discipline: the art & practice of the learning organization. New York: Currency Doubleday.

Simon, H.A. (1955), 'A behavioural model of rational choice', Quarterly Journal of Economics 69, 1-18

Sterman, J.D. (2006). Learning from Evidence in a Complex World. American Journal of Public Health, March 2006, vol. 96, No. 3

Zahra, S.A., George, G. (2002): Absorptive Capacity: A Review, Reconceptualization and Extension. Academy of Management Review, 2002 Vol. 27, No. 2. pg. 185-203

Risk Management in Dynamic Logistic Systems by Agent Based Autonomous Objects

Boris Bemeleit[1], **Martin Lorenz**[2], **Jens Schumacher**[3], **Otthein Herzog**[2]

[1] Bremen Institute for Industrial Technology and Applied Work Science—BIBA, email: bem@biba.uni-bremen.de
[2] Center for Computing Technologies—TZI, email: {mlo,herzog}@tzi.de
[3] University of Applied Science Vorarlberg, email: jens.schumacher@fhv.at

Abstract Intelligent logistic objects with the ability of autonomous control are a possibility to face new challenges in dynamic logistic systems. On the other hand autonomy of logistic entities poses new questions to the underlying logistic system. In this paper we will focus on the possibilities of reducing the objective and subjective drawbacks of autonomy in logistics by an autonomous risk management for the logistic entity. A technical solution based on intelligent agent technology will be presented.

1 Introduction

Within the last years logistics has developed to a key success factor in globally distributed production because of its cross-sectional function. Enhanced product life cycles, rapid changes in company structures and information flows alter the requirements for logistic processes. With the reduction of the vertical range of manufacturing and the tendency to globally distributed production, logistics and the design of logistic processes gain significance (Eschenbächer and Gühring, 1999). As a result the importance of logistics grows and new concepts for planning and control of the logistic processes are needed.

On the one hand the rising complexity of inter-organizational structures and also a relative shortage of logistic infrastructure lead to increasing utilization of the existing logistic processes. On the other hand a specialization and intermodalisation of the ways of transportation and respective carriers can be observed (Chang, 2006). These factors combined with changing customer market conditions have considerable effects on planning and controlling of logistic processes in such a dynamic environment.

The resulting dynamic and structural complexity of logistics networks makes it very difficult to provide all information necessary for a central planning and control instance in time during the planning phase and react on incoming information during the phase of execution. A possible approach to face these challenges is the devel-

H.-D. Haasis et al., *Dynamics in Logistics*

DOI: 10.1007/978-3-540-76862-3,

opment of autonomous logistic processes which have the ability and capabilities for decentralised coordination and decision making.

The autonomous entities are self contained and follow an individually rational goal. In the basic assumption they are individually rational decision-makers in the sense of game theory, each aiming at maximizing their individual utility function.

In the remainder we will further investigate the issue of complexity and dynamics in modern logistic systems and the possibilities that arise from introducing autonomy (Sect. 2). Section 3 will show that risk management on the level of individuals in an autonomous logistic system can answer questions of dependability and trust that could arise when making logistic entities autonomous. We will furthermore investigate the requirements for a technical system that can address those answers in Sect. 4 and sketch it's implementation in Sect. 5. An Outlook in Sect. 6 will wrap up over results. An approach to face the challenges on existing and upcoming problems in logistics is the concept of autonomous logistic processes represented by autonomous logistic objects. Autonomy in logistic processes is defined by "Autonomous control in logistics systems is characterized by the ability of logistic objects to process information, to render and to execute decisions on their own" (Windt et al., 2007). The autonomous control of logistic processes can be realized through decentralized control systems, which select alternatives autonomously or logic based semi-autonomously and decide within a given framework of goals.

2 Complexity and Dynamic in Logistic Systems

Complexity can be understood as interaction between complicatedness and dynamics (Schuh, 2005). Due to the fact that the described approaches of complexity only refer to single aspects of complexity, as for instance the structure of a considered system, they seem insufficient for an entire understanding of the term complexity in the context of logistic systems, in particular production systems (Philipp et al., 2007).

Different approaches aim at explaining the complexity of systems that also apply to logistic systems. We subsume those under two main categories, the *element-* or *relation complexity* approaches like Bar-Yam (Bar-Yam, 2003), who names (1) elements (and their number), (2) interactions (and their strength), (3) formation/operation (and their time scales), (4) diversity/variability, (5) environment (and its demands), and (6) activity(ies) (and its[their] objective[s]) as the characterizing properties of a complex system, whereas Ottino (Ottino, 2004) as a representant of the class that defines complexity by its property of *emergence* states: "A complex system is a system with a large number of elements, building blocks or agents, [...] that [...] display organisation without any external organizing principle being applied."

The dynamics of a logistic system is characterised by its temporal behaviour. A dynamic system is subject to permanent changes on micro and meso level however it can take a constant perceivable state on macro level. The number of possible states resulting from events influencing the system and the interaction between the embedded entities of the system is a representative factor for measuring the dynamic of the logistic system and is as important for tasks of planning and realising logistic processes as the complexity of the logistic system.

Philipp et al. (Philipp et al., 2006) believe that it is essential to define different categories of complexity and to refer them to each other to obtain a comprehensive description of complexity. They divide complexity into three categories *Organisational complexity*, which resembles element- and relation complexity, *Time-related complexity*, thus redefining dynamics as a form of complexity, and *Systemic complexity*, which introduces the border between system and environment. The emergence aspect of complexity is not present in this approach.

However, the complexity of a logistic system and its dynamic as well as the overall behaviour of the system still allow no conclusion regarding the sensitivity of the system in relation to the malfunction of individual entities or their relations as well as regarding non-deliberate or malevolent disturbances from outside the system.

Following Luhmanns theory that only complexity (of the system) can reduce complexity (the system is exposed to) (Luhmann, 1984) we enable the (logistic) system to deal with increased complexity in it's environment by increasing the complexity of the system itself and managing this complexity (by means of technology) (Baecker, 1999). This approach is hereby advanced by the development of novel information and communication technology (ICT).

Fast and ongoing development of modern ICT, e.g., telematics, mobile data transfer, and transponder technology opens new possibilities for the development and emergence of intelligent logistic systems which can fulfil the requirements of autonomous logistic processes. However to maintain a controllable dynamic logistic system, technological development must not only provide short-run autonomous replacements for standard logistic operations but also take into account that introducing autonomy will impact the operational and strategic management of logistic services.

Because the dynamic and structural complexity of logistics networks makes it very difficult to provide all management information, which would be necessary for a central planning and control instance the autonomy of the logistic entities is a promising approach.

This autonomy can be realised by the development of adaptive logistic processes including autonomous capabilities for the decentralised coordination of autonomous logistic entities in a heterarchical structure. Autonomy describes processes of decentralised decision making in heterarchical structures. Autonomy assumes that interacting entities have the ability and possibility to decide autonomously in non deterministic systems (Windt et al., 2007).

The autonomy permits and requires new control strategies and autonomous decentralised control systems for logistic processes. In this setting, aspects like flexibility, adaptivity and reactivity to dynamically changing external influences while maintaining goals are of central interest.

The integration of strategic and tactical planning combined with an amount of actual data and possible communication between the systems entities enables the system to act autonomously and maybe compensate a temporary or unlimited malfunction of an entity or a relation between two or more entities. A consequence of the autonomous acting of the involved entities is a shift of the responsibility for the realisation of the decisions from a central decision system—be it technical or human—to the single logistic entity.

This has to be regarded by developing a management concept of autonomous logistic objects and the complexity of the total system.

3 Control of a Dynamic System by Online Risk Management

The increased complexity of logistic systems is followed by a more complicated planning and control of logistic systems and of the related processes in combination with an increased sensitivity of the total system to disturbances and malfunctions. The hazard of delayed delivery in transportation, latency in manufacturing and reduced adherence to delivery dates are results of complex system structures and increased customer requirements. All these numerated disturbances and changed conditions clarify that logistic systems and the related logistic processes are very fragile and the contained hazards and chances have to be managed to ensure the success of the logistic processes. These circumstances show that the development of a management system for risks is essential for a successful realization of autonomous logistic objects.

Avoiding, reducing and partly compensating are selected strategies for risk by a proactive autonomous risk management system relies on a a-priory identification and analysis of events, which could be dangerous for the fulfillment of goals given to an autonomous logistic object. Thereby the aim should be to model risks on an abstract level and to integrate operational risk detection into the autonomous system. Such a risk management system will be developed in our sub project.

The consequence of the proposed shift of responsibility from a central instance to an autonomous logistic object is a fundamentally different situation in the face of events or situations, which could endanger the success in terms of reaching the goals of a logistic process.

In classic logistic systems a malfunction of the centralized deciding instance is the main danger for the success of the whole of logistic processes involved. Other problems are, that central systems are suitable to only a limited extent in reacting on changing local conditions and that a local lack of information affects the total system. With the application of autonomous logistic objects this disadvantage can be compensated but for an autonomous logistic object it has to be kept in mind that there are additional risks. These risks result from the required communication between the involved objects and that the interaction between them which leads to non calculable states on local and global level. It is also important to consider that contradictory information generated from different objects is another source of risk for the logistic processes in relation to their specific goals and that an optimization on object level can compromise the goal of the total system.

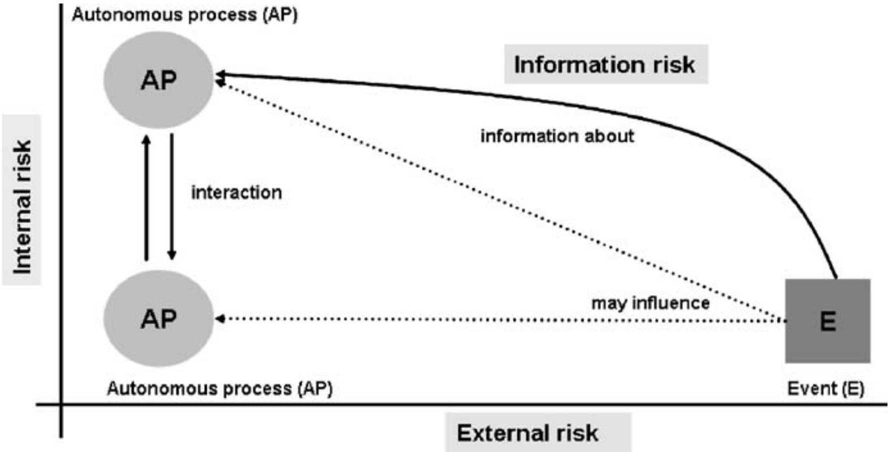


Fig. 1 Risks induced by events

Direct disturbances of the processes caused by risks which exist impartial from the logistic objects and risks which result from the interaction of the logistic processes.

These flexible characteristics of disturbances can be categorized in 3 types of risk: *External risk*, which is caused by an event, that exists independently from the autonomous processes and may affect them, *Internal risk*, which is a result of the interaction between autonomous processes, the reasoning within an autonomous process and *Information risk*, which is related to the information which are available but may be inconsistent, contradictory fuzzy, incomplete or unreliable. An overview about the different characteristics of risk which could influence the logistic objects is given in Fig. 1.

In order to managing different types of risk it is essential to understand the meaning of risk for autonomous logistic objects and their environment. To handle existing and new risks for autonomous processes and autonomous objects a proactive risk management has to be established as a part of the whole system, because it helps to develop logistic processes which are robust and insusceptible to existing and occurring risks: A risk management system supports the autonomous objects in decision making and realizing these decisions considering the risk which is related to the whole logistic processes. For this reason the development of a proactive risk management system can be considered as a relevant success factor for autonomous logistic processes.

4 Risk Management of Autonomous Objects

Goal fulfillment is the defining characteristics of a risk concept for autonomous logistic entities. In the logistics domain this goal might be to reach a given destination in shortest possible time, at lowest cost or with lowest possible fuel consumption. The autonomous entities aim to maximize its local utility will usually subsume primary goal fulfillment but aspects like system continuance or contribution to a global utility of the enterprise the entity belongs to induce different risks. The autonomous system therefore needs to acquire and maintain an internal model of its environment and the processes therein. Using a "foretelling" mechanism can than enable the assessment of situations that will be occurring. Such a mechanism has of course to be of technical nature and thus needs to calculate future states of the world based on probabilities.

In a technical autonomous system one can either employ classical—brainstorming based—methods of risk assessment in advance (the "design time") or find a computer implementable method to assess risks. The former is simply a matter of completeness of the design process. The disadvantage of design-time assessment is obviously that new situations in which risks occur cannot be handled by the autonomous system. In conventional control tasks a human operator will be responsible and able to intervene. In the autonomous decision-making case this task is delegated to the system itself. Therefore enabling autonomous risk assessment is the only remaining alternative.

Thus for a risk management within an autonomous logistic entity we need five technically implementable components. (1) An internal local model of the environment, which will contain static elements that are common to all entities and inherently subjective parts originating from local perception and communication with other entities. To fulfill a given goal it will (2) need to make plans using the knowledge it has and (3) generate hypotheses about future states of the environment. The subjective part of the knowledge needs (4) a mechanism to assign a certainty value to each item and evaluate its contribution to hypotheses, triggering the acquisition of additional information as necessary. Finally it will need to (5) evaluate plans it made and predicted states of the environment for their potential of risk.

5 Technical Risk Aware Decision-Making

In conventional research on multiagent systems, it is claimed, that the local interaction of autonomous systems (microscopic behavior) should lead to an optimized behavior on the global level (macroscopic behavior) (Langer et al., 2005). However current agent architectures are not designed to model this complex decision-making process which requires agents to process knowledge about internal structures and organizations, show awareness of other agents and communicate or even cooperate with them, and perceive changes in their environment.

An important challenge for this project is to augment the agent's deliberation cycle with the ability to identify and assess the underlying risks that are associated with the options that determine the next course of action. If necessary, the agent must be able to augment its knowledge base with missing or updated knowledge, for example, from other agents, to be able to properly assess and evaluate the feasible options. In (Langer et al., 2005) we proposed a framework for an enhanced agent deliberation process. This framework is being developed as a common basis for risk- and knowledge-management in agent decision-making (Langer et al., 2006). Generally speaking, we use risk management to identify and assess the risks associated with one or more options.

The first step is the identification of potential risks associated with each option. Each identified risk must be evaluated to assess the magnitude of the risk and its probability of occurrence. In the ideal case, the agent has sufficient knowledge to arrive at a meaningful risk assessment. Upon completion, the result of the assessment is returned to the deliberation process which uses the information to aid in the selection of the best possible option.

Risk is thereby represented as a set of patterns with an attached "severeness" attribute, which are matched against the current or a predicted state of the world. For example assume the agent knows from a weather forecast, that it might rain within an hour. Together with the knowledge that it carries a paper roll this world state wold match a risk pattern, which states, that water harms paper with a severity of 0.8—where 1 would mean certain total loss.

Due to incomplete or uncertain knowledge, risk management may be unable to decide on risk, e.g., the weather report might be ten hours old. This triggers knowledge management to acquire the missing information or detailed information on the current situation. Knowledge acquisition may retrieve knowledge from other agents or directly from external sources/sensors. In our example a local weather service may offer up-to-date reports on an hourly basis. The agent could then decide to alter its plan from using the direct route across open space to a roofed but much longer way. The tradeof between the risk of loosing the load and choosing the longer, hence more expensive route will make the differnce of success or failure.

6 Conclusion

In the present paper we related different aspects of complexity and dynamics in modern logistic systems to the possibilities that arise from introducing autonomy. We showed that autonomous processes, that can serve as a building block for managing the growing dynamics and complexity, will need a mechanism for dealing with risks.

The concept of risk, as it is known from economics and project management as well as plant and machinery safety, has been re-introduced to fit for assessment and management by an autonomous process, implemented as an autonomous software agent. Risk management on the individual level in an autonomous agent has been introduced.

This risk management system supports the autonomous logistic objects in decision making and realizing these decisions considering the risk which is related to the whole logistic processes.

Acknowledgements This research was supported by the German Research Foundation (DFG) as part of the Collaborative Research Centre 637 "Autonomous Logistic Processes".

References

Baecker D (1999) Organisation als System, suhrkamp taschenbuch wissenschaft, vol 1434. Suhrkamp, Frankfurt/M

Bar-Yam Y (2003) Dynamics of Complex Systems (Studies in Nonlinearity). Westwing Press

Chang TS (2006) An international intermodal routing. In: Operations and Technology Management 2: Complexity Management in Supply Chains, Erich Schmidt Verlag, Berlin, pp 87–103

Eschenbächer J, Gühring T (1999) Prozeßbenchmarking in der Materialversorgung. In: Logistik Management, Springer, Berlin, pp 31–40

Langer H, Gehrke JD, Hammer J, Lorenz M, Timm IJ, Herzog O (2005) A framework for distributed knowledge management in autonomous logistic processes. International Journal of Knowledge-Based & Intelligent Engineering Systems 10(4):277–290, http://www.tzi.de/ mlo/download/kes84.pdf

Langer H, Gehrke JD, Herzog O (2006) Distributed knowledge management in dynamic environments. In: Hülsmann M, Windt K (eds) Understanding Autonomous Cooperation & Control in Logistics, Springer, Berlin, chap B4, pp 215–230

Luhmann N (1984) Soziale Systeme. Grundriss einer allgemeinen Theorie. Suhrkamp, Frankfurt/M

Ottino JM (2004) Complex systems. AIChE Journal 49(2):292–299 DOI 10.1002/aic.690490202

Philipp T, Böse F, Windt K (2006) Autonomously controlled processes characterisation of complex production systems. In: Cunha P, Maropoulos P (eds) Proceedings of 3rd CIRP Conference in Digital Enterprise Technology, Setubal, Portugal

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 M, Windt K (eds) Understanding Autonomous Cooperation & Control in Logistics The Impact on Management, Information and Communication and Material Flow, Springer, Berlin, to appear

Schuh G (2005) Produktkomplexität managen. Carl Hanser Verlag, München

Windt K, Böse F, Philipp T (2007) Autonomy in logistics – identification, characterisation and application. International Journal of Computer Integrated Manufacturing Forthcoming

Knowledge Management and Service Models in Logistics

Knowledge Management in Intermodal Logistics Networks

Hans-Dietrich Haasis

University of Bremen and ISL – Institute of Shipping Economics and Logistics, Universitätsallee GW1A, 28359 Bremen, Germany, e-mail: haasis@isl.org

Abstract Product differentiation in transport and logistics mainly can be realized by quality improvements of related services. In this context, amongst others, demands for intermodal solutions arise due to organisational and infra-structural induced bottlenecks in mobility. Therefore knowledge management approaches rank among the best design options of a modern and innovative management in these companies. This is also emphasized by results of the initiative "Fit for the knowledge competition" of the German Ministry of Economics and Technology. Using intermodal logistics networks as an example, in this paper it will be demonstrated and outlined on how knowledge management yield benefits, and which trends can be expected in line with the application of knowledge management tools in this dynamically characterized decision environment.

1 Intermodal logistics networks

Realizing the vision of sustainable economics on an international level, logistics can take over a leading force. Already today, sustainable supply chains can be designed, at least for selected needs (Haasis 2004). For this, amongst others, intermodal logistics networks play an essential role.

As an example, the extension of the EU results in new freight transportation, increased cargo quantities and changed cargo qualities. Therewith it comes to new cooperative solutions and new allocations of responsibilities in value adding logistics networks and to new intermodal logistics products.

As known, freight related intermodal logistics networks are defined as integrated transportation systems consisting of two or more modes. Modes on intermodal networks are connected through terminal facilities, which allow freight to shift from one mode to another during a trip from an origin location to a destination. These networks have to handle network complexity – for example expressed by the amount of interrelated process decisions and the number of interfaces within the network –, time challenges, and individuality in customer requirements, bottlenecks

H.-D. Haasis et al., *Dynamics in Logistics*

DOI: 10.1007/978-3-540-76862-3,

in resources and capacities, safety and security directives (Chandra and Kamrani 2004; Haasis 2005; Haasis and Szafera 2005).

Moreover, both the economic and the flexible arrangement of cooperative systems within logistics networks, taking into account all main supply chain partners and their various development options, play the leading role especially for international cooperations in logistics (Baumgarten 2004). By this, the fair interaction between production, logistics and services is essential. Hence, supply chain design and controlling has to be extended. In line with this, innovation factors like communication, systems integration, logistics product planning and skills are of vital importance (Haasis 2006).

Thus, logistics today and even more in the future is concerned with an interface-related systems design. For example, interfaces arise between enterprises, business processes, information technologies, decision makers, transport technologies and regulations. Of course, net effects and net externalities have to be analysed and evaluated taking into account a combined view on macro-logistics and micro-logistics. All these are arguments for the realization of a knowledge management approach for the design and the operation of intermodal logistics networks on a national and an international level.

2 Challenges in Knowledge Management

In view of an increasing dynamic environment, characterized by customer oriented ambitious demand behaviour and the creation of partly worldwide cooperative solutions, dealing with knowledge results in essential competitive advantages. The use of the so-called production factor "knowledge" will be decisive for the fluently meshing of business processes and the innovative consolidation or better strengthening of the own position in the international competition. Innovative enterprises recognize the challenges to achieve competitive advantages not only by technical innovations but also by improving the creation, the documentation, the qualification, and the utilization of knowledge in line with the assignment of human resources within the process chain.

The business processes related to integrated logistical service products are much more complex and knowledge intensive than traditional single services. This requires the implementation of new management systems and the use of know-ledge management techniques not only within production companies but also within logistics service providers. Business-related knowledge management represents a structured procedure for dealing with the production factor "knowledge". Knowledge management techniques show a high potential especially for the use in small and medium sized companies (BMWi and KPMG 2001). By this, know-ledge management can serve as engine for strengthening the innovation power, the cooperation capability, and the process flexibility of logistics orientated small and medium sized companies. The implementation and the use of knowledge management hold the potential for the development of sustainable competitive advantages (Haasis 2001).

Also for the optimisation of cross-border transports in international logistics besides technical innovations above all organisational innovations as well as the willingness and the ability for cooperation are required. Future-oriented concepts with a logistics product portfolio meeting customer requirements have to use synergetic effects along the whole supply chain as well as to balance in a holistic way pros and cons of the partners. Conflicts of interest require flexible solutions as well as a minimisation of transaction costs for all partners involved. Thereby knowledge is more than data and information. This knowledge has to be identified, developed, qualified, distributed, used and conserved. Cross-linking in a specified decision situation only can develop knowledge, and that by an organisational learning of individuals and teams. These tasks constitute the modules of a modern knowledge management in enterprises (Ilskenmeier 2001; Nonaka and Takeuchi 1997; Probst et al. 1997; Wilke 1998).

Since shipping companies and in general international supply chain activities are shaped by a dynamic and information intensive environment, for companies involved knowledge management is essential. The improvement of the collaboration between partners and decision makers along the international supply chain, e.g. shipping companies, ship-owners, terminal operators, port operators, forwarders, retailers, and production companies, requires knowledge-based solutions. The increasing defragmentation of the supply chain and the extension of the global production and service networks lead to more and more partners within the network. In parallel the reaction times necessary for success in competition decrease. Due to this, both a business approach for knowledge management as well as a transfer of knows along the mainly intermodal logistics network is necessary.

Moreover production and logistics locations, qualified for cooperation in line with the realization of knowledge regions, have to be positioned as international partners in logistics networks. As an example, international port cooperations as well as cooperations between logistics centres can be mentioned. This implies, that innovative companies in knowledge regions will realize compatible measures, which strengthen their potentials, their efficiency, their attractiveness and their productivity. Besides customer orientation after all cooperation and knowledge management will develop logistics in the next years.

Especially the competitiveness of small and medium sized enterprises will be formed by their capability to initiate, to design and to operate cooperations both in a region and in a supply chain. In line with this, cooperations not only are restricted on suppliers and customers, but also on other partners, stakeholders and, from case to case, competitors (Haasis and Fischer 2007). Due to their regional integration and their regional engagement for them a regional cooperation management like "think global, act local" is essential. This interplay of forces also is expressed within the European policy of regions.

Especially the cross functional and companies trespassing transfer of know-ledge moves in the management focus. The fair collaboration asks for a solid information transparency and mutual trust. A sustainable knowledge management along the whole intermodal logistics process chain only can realize this.

For a related implementation of knowledge management first of all an identification of relevant sources of knowledge and information is required, for the enterprise and for the supply chain. For example, this can be realized and operated by selected Internet portal solutions and so-called knowledge maps. Within a next step towards a living knowledge management in intermodal logistics networks the organisation of knowledge transfer and qualification between the partners in the companies and in the cooperation network is essential. From a technical point of view this can be supported by information platforms. However, the transfer always has to be flanked by personnel measures, e.g. coaching, mentoring, inter-divisional teams, self-improvement, training measures, and temporally restricted exchanges of staff members. Furthermore the conservation and updating of know-ledge as well as a continuously realized controlling are of vital importance for the success of knowledge management activities.

3 Selected aspects of applied knowledge management in intermodal logistics

In line with the realization of process-related logistics chains and intermodal networks barriers may occur when organizing the necessary collaboration between different decision makers and supply chain partners. As an example, operation requirements of the shippers have to be confronted with capacity related transportation restrictions. Nevertheless, an efficient sequencing of tasks along the intermodal logistics chain enhances in defined decision situations the necessity to use distributed and interdisciplinary knowledge. The implementation opportunities in many cases lie within the practical knowledge of the decision makers. Hence, for an efficient design and operation of intermodal logistics organisational, personnel and technical requirements and options have to be balanced.

The complexity of knowledge problems in view of intermodal logistics networks can be characterized as follows:

- Personnel problems:
  - Less state of knowledge on supply chain relations
  - Missing exchange of informations
  - No documentation of practical knowledge of key actors
  - No reward system motivating to think on new solutions
  - Problems in handling and use of information masses
- Organisational problems:
  - No solid business processes for the transfer of information and knowledge
  - Missing understanding of the sphere of action of the other decision makers in the chain
  - Barriers in line with the implementation of electronic business solutions
  - Insufficient process accompanying information logistics

- Technical problems
  - Problems of interfaces in a heterogeneous IT-infrastructure
  - Time consuming consolidation of informations

These general knowledge problems can be demonstrated especially in line with two typical sceneries:

- **Knowledge documentation and conservation due to exchange of staff members**

The exchange of staff members can result in various problems for knowledge conservation and for knowledge qualification of a team, a division and of course a company. However the problems are in general quite different, e.g. according to the possibility of fresh engagements or a cancellation of the position. In the ideal case the position should be staffed with tandems, the old member and the new member in parallel. By this, the old member can transfer his experience and knowledge to the new member. At least vacant positions should be avoided. Besides this, methods and tools of knowledge documentation can be applied, e.g. address lists, documented manuals, project reports, knowledge maps.

- **Improvement of cooperation within the supply chain**

For the configuration and operation of an intermodal supply chain an interface-orientated management of the overall process chain is necessary. Capabilities, capacities and skills have to be allocated. The objective is to improve the performance of the whole supply chain instead of only one part. The provision and the use of knowledge management can result in a more economic and reliable sequencing of tasks in and between companies. Due to complementary or conflicting objectives of the partners in the intermodal logistics chain an interdisciplinary and system-oriented approach is needed, i.e. system integration.

Obviously, the collaboration of decision makers along the logistics chain requires a knowledge-based path. Hence, a business oriented knowledge management is just as necessary as a transfer of know how between decision makers. More information on knowledge management, its tools and benchmarks can be found on www.wissenmanagen.net. This is the official platform of the initiative “Fit for the knowledge competition” of the German Ministry of Economics and Technology. Benefits of applying knowledge management are amongst others an increase in flexibility, in reliability of logistics services, and in a joint innovative learning.

In line with the design and operation of international intermodal logistics chains more and more the improvement of cooperations imply the realization of logistics regions as knowledge regions. These knowledge regions allow to compose individual logistics functions and to bundle effects in transportation and services, for example in view on the organisation of hinterland relations. A knowledge region means a better regional communication and knowledge management. The region as a whole offers his capabilities and skills as an increasingly powerful partner in international logistics networks. Knowledge management is used to improve the personnel, process-oriented, informational, and technological interfaces between the network partners (Haasis 2005).

4 Conclusions

Intermodal solutions in freight transportation more and more play an important role as innovative paths for increasing operations and transportation efficiency. Mainly, demands for intermodality arise due to organisational and infra-structural induced bottlenecks in mobility. Besides technical aspects the integrative approach has to take into account amongst others the comprehension of the different views of related decision makers and their objectives as well as the linkage of corporate decision ranges. By this, the dynamic and structural complexity of these logistics networks imply design and control principles which make complexity controllable by means of intelligence.

The complexity is mainly influenced by the dynamic environment, the related configuration of interfaces and the number of partners in the network. The collaboration of partners or decision makers along the process chain calls for a knowledge-based path. Hence, a business-oriented knowledge management is just as necessary as a transfer of know how between decision makers. Therefore, a customer oriented supply means an excellent control and design of interfaces as well as knowledge based governing of the underlying complexity in the logic of processes and structure.

On an international level also intercultural affected decision processes have to be taken into consideration (Cui et al. 2004). This imply the supply availability of transport infrastructure and quality as well as the skill potentials of the enlisted participants in line with governing of complexity in combination with flexibility advantages. A regional culture of business innovations and social learning supports this trend in logistics and supply chain development (Haasis 2001).

In line with this, benefits of applying knowledge management in intermodal logistics are for example an increase in process efficiency, in flexibility, in reliability of logistics services, in a joint innovative learning and in a related product marketing. All these benefits result in a stronger position in international competition. The joint interdisciplinary study and use of knowledge management measures in the corresponding logistics network is mainly based on a dialogue, on communication and on cooperation. Finally, the knowledge management approach assists to come to a common understanding of sustainability in innovative international intermodal logistics.

References

Baumgarten H (2004) Trends in der Logistik. In: Baumgarten H, Darkow IL, Zadek H (eds) Supply Chain Steuerung und Services. Springer, Berlin Heidelberg New York, pp 1–12

BMWi, KPMG (2001) Studie zur Bedeutung und Entwicklung des multimediabasierten Wissensmanagements in der mittelständischen Wirtschaft. KPMG Schlussbericht 20. August 2001

Chandra C, Kamrani AK (2004) Mass Customization: A Supply Chain Approach. Springer, Berlin Heidelberg New York

Cui Z, Deng A, Haasis HD (2004) The Simple Analysing of Developing of Lean Logistics in China. In: Sun L, Notteboom T (eds) Proceedings of the First International Conference on Logistics Strategy for Ports. Dalian, pp 457-460

Haasis HD (2001) Wissensmanagement und dessen Bedeutung für ein Nachhaltiges Wirtschaften. In: Griesche D, Meyer H, Dörrenberg F (eds) Innovative Managementaufgaben in der nationalen und internationalen Praxis. Wiesbaden, pp 141-156

Haasis HD (2004) Nachhaltige logistische Systeme. In: Schwarz E (ed) Nachhaltiges Innovationsmanagement. Wiesbaden, pp 105-118

Haasis HD (2005) Mass Customization in International Logistics. In: Blecker T, Friedrich G (eds) Mass Customization. Concepts – Tools – Realization. Berlin, pp 189-193

Haasis HD (2006) Logistik der Zukunft. In: Jahrbuch der Hafentechnischen Gesellschaft 2006. Hamburg, pp 63-66

Haasis HD, Fischer H (eds) (2007) Kooperationsmanagement. AWV-Verlag, Eschborn

Haasis HD, Szafera S (2005) Supply Chain Management und intermodale logistische Korridore. In: Ehrig D, Staroske U (eds) Dimensionen angewandter Wirtschaftsforschung: Methoden, Regionen, Sektoren. Hamburg, pp 311-319

Ilskenmeier D (2001) Der Ansatz eines ganzheitlichen Wissensmanagements für Klein- und Mittelbetriebe. Beiträge aus der Forschung der Sozialforschungsstelle Dortmund, Nr. 124. Dortmund

Nonaka I, Takeuchi H (1997) Die Organisation des Wissens. Wie japanische Unternehmen eine brachliegende Ressource nutzbar machen. Frankfurt New York

Probst G, Raub S, Romhardt K (1997) Wissen managen. Wie Unternehmen ihre wertvollste Ressource optimal nutzen. Wiesbaden

Willke H (1998) Systemisches Wissensmanagement. Stuttgart

Knowledge Management in Food Supply Chains

Bernd Scholz-Reiter, Salima Delhoum, César Stoll

University of Bremen, Planning and Control of Production Systems
Hochschulring 20, 28359 Bremen, Germany

Abstract In a food supply chain, companies exchange information and products to create value. The competitive advantage of the company relies on that of the entire chain. This paper addresses knowledge management in a food supply chain especially the promotion of knowledge sharing and transfer. This is a substantial issue when, for example, a company comes to leave the supply chain voluntarily or disappears because of natural or man-made catastrophes. In this case, the question remains whether the substituting organization can acquire the tacit knowledge needed to fulfill its mission. Consequently, the operation of the supply chain becomes an acute issue too. The paper proposes a methodology based on organizational and interorganizational learning and develops a learning laboratory that facilitates knowledge sharing and transfer. This lab supports (a) knowledge sharing through the mental map's knowledge elicitation of every company's leader, and (b) knowledge transfer via dedicated learning-based environments.

1 Introduction

A food supply chain is constituted of a number of organizations responsible for the production and distribution of vegetable and/or animal-based products. Bourlakis and Weightman (2004) defined the food supply chain as controlled processes combining knowledge, skills and technology in addition to a wide range of other disciplines. Products are supplied by farmers and delivered to potential customers. The orientation on customers and the underlying social responsibility lead to a change in the supply chain by the development of strategic management. Partly, this is achieved by globalization which offers the opportunity to develop new businesses and improve the satisfaction of customer demand. Moreover, globalization promotes the applications of information technology, and hence; it supports the movement of products, and the dissemination of product information through the food supply chain (Bourlakis and Weightman 2004). However, globalization can also complicate the performance of a food supply chain when competitiveness is increased and

H.-D. Haasis et al., *Dynamics in Logistics*

DOI: 10.1007/978-3-540-76862-3,

international reallocation of suppliers and consumers is altered. An informational channel between the different sectors of the food supply chain (suppliers, farmers, manufacturers, clients) is needed in order to share and create knowledge. This facilitates the collaboration within the food supply chain. Indeed, success depends on a program of mutual benefits based on trust, communication, production continuity; and time and visibility to make changes (Seyed-Mahmoud 2004).

High levels of information exchange and knowledge transfer represent a common characteristic of lead companies such as Wal-Mart and Rewe. In the food retail industry, the Wal-Mart model is usually applied. This model is dependent on the continuous replenishment of inventories. Accordingly, Wal-Mart shares information with customers and manufacturers electronically, and therefore, it can replenish the inventories in its stores (Seyed-Mahmoud 2004). This electronic data interchange (EDI) system not only performs an easy, quick as well as responsible supply of shelves in supermarkets but also improves the traceability of food with the resulting cost reduction in the food safety program (Wijnands and Ondersteijn 2006). Rewe and other independent retail dealers began a project about the application of radio-frequency identification (RFID) along the meat supply chain. The project aims to examine and optimize information transparence about the handling, transport and storage conditions. This will enhance the traceability and security throughout the food supply chain.

After a review of knowledge management processes, the paper describes how a food company can share and transfer knowledge about its operations in the supply chain (with suppliers, farmers, etc.) based on the foundations of organizational behavior. As an approach, both organizational and interorganizational learning are used in order to create and facilitate knowledge in a food supply chain. Organizational and interorganizational learning suffer from defensive routines and "people barriers" featured by diverse fears and aversion to knowledge sharing (Argyris 1985; Barson et al. 2000). Therefore, it is valuable to have an effective strategy to learning and use dedicated tools such as the learning laboratory discussed in this paper.

2 Knowledge Management Processes

Knowledge management can be described as practices used by companies to identify, create, share, transfer and turn knowledge into a competitive advantage. Davenport et al. (1998) deemed knowledge to be information combined with experience, context, interpretation and reflection. Thus, knowledge exploitation plays an important role within each organization and between organizations integrating the supply chain, especially in the business relations between suppliers and customers. However, extending internal knowledge management programs to external sources is a demanding task because of some typical barriers. Barson et al. (2000) classified these barriers in four categories: technology barriers, e.g., available technology and legacy systems; organisation barriers, e.g., distance, poor targeting of knowledge, costs of managing knowledge transfers, protection of proprietary knowledge; people barriers, e.g., internal resistance, self interest and lack of trust; and finally cross-category barriers which are featured by technology, organisation and/or people, e.g., existing resources, the need for rewards and culture.

A short description of the processes forming knowledge management is presented to clarify the most important components within this concept. Some of them are consequential in the development of the learning laboratory. Alavi and Leidner (2001) identified knowledge management as a salient factor in the accrual of a company's competitiveness and proposed the following processes:

- Knowledge creation: explicit through systematic language, scientific formula; and tacit, i.e. ideals, values, emotions, etc. (Nonaka 1991). Both are strongly interconnected and cannot be easily differentiated
- Knowledge storing: the capacity to retrieve knowledge
- Knowledge transfer: identification, assimilation and exploitation of knowledge
- Knowledge application: use of knowledge in order to increase or maintain the competitive advantage

These processes take place in today's globalized world which is replete with information. Companies exchange information resulting in a potential for knowledge sharing. It is essential that companies recognize and promote it because knowledge sharing prevents reinventing the wheel (Bender and Fish 2000), crates shared understanding (Nickerson and Zenger 2004); reduces uncertainty (Thusma 1978), and turns individual learning into organizational learning (Nonaka 1994). Berends (2005) distinguished between direct and indirect contributions of knowledge sharing. The former has an immediate effect on problem solving with either (a) a contribution to the solution of a problem, (b) a change in the problem statement; or (c) activation of new actions. The latter possesses a long-term plus-value, in that; it helps the development of background knowledge.

3 Organizational Approach to Knowledge Management

Lasting learning involves double-loop learning (Argyris 1985) which questions the underlying goals, assumptions and programs of the organization's shared mental model instead of just adjusting its behavior to achieve a desired outcome as it is the case of single-loop learning. Double-loop learning provides the organization with a shared, tacit learned behavior that survives the people who leave it in a voluntary or forced manner. Organizational learning offers knowledge elicitation methods that facilitate double-loop learning. Examples of these methods are the left-hand column method (LHC), originally introduced by Argyris & Schoen (1974) and the ladder of inference portrayed in Senge et al. (1994). Both methods allow knowledge sharing since they unearth that tacit knowledge held in mental models and render it visible and explicit. Learning-based environments (management flight simulators, microworlds, etc.) are learning processes, tools or environments that help activate double-loop learning. These tools can transfer individual learning to organizational learning since the knowledge source is the individual who interacts with the computer model to find out sustainable strategies which will be later applied to the organization (knowledge recipient). The combination of (a) knowledge elicitation and (b) learning-based environment yields a dedicated framework for learning, often called a learning laboratory. It is an effective environment to share knowledge through mental model elicitation and to transfer knowledge using learning-based environments.

3.1 Learning Lab

First of all, it is important that technology evolves rapidly in the supply chain in order to safeguard that every company has little domination over the others. From one side, the concentration of power is detained by supermarkets in the food supply chain. The retailers apply sometimes pressure on farmers with respect to the buying price, thus, farmers find themselves obliged to sell their products at a price close to the cost of production. From the other side, technology is of standard use to preserve fresh and quality products. Second of all, the member companies should be willing to learn and perhaps have already a history in organizational learning. The initiation of a learning effort is pertinent only when the participating food companies have a true interest in learning. The learning lab sustains knowledge sharing and transfer by the use of the left-hand column method and learning-based environment respectively. The learning laboratory is offered within a workshop and it is led by a facilitator. Its duration depends on the material studied by the participants and the pace of the steps. Overall, the workshop is feasible in three hours.

Knowledge Sharing

One representative member of each company in the supply chain experienced knowledge sharing through the left-hand column (LHC) method. Naturally, a skilled facilitator is required to achieve a rewarding interactional environment. Furthermore, for the prevention of conflict outbursts during the elicitation of decision makers' mental maps, a systematic record of discords is thought to help avoid the cause of the conflict in the subsequent sessions of the learning lab. For learning in groups suffers from defensive routines. Hence, the LHC method detailed in the next section takes a fictive story between two decision makers X and Y as prevention against interpersonal conflicts. The elicitation method adopted follows the methodology described by Argyris and Schoen (1974) summarized in Fig. 1. Y commented on the poor performance of X with regard to the inventory control of food products:

> "Your orders were too much fluctuating.
> Your ordering policy was immoderate.
> The forecast of customer demand you gave me was exaggerated.
> As a result, the whole inventories oscillate.
> You provoked dramatic inventory costs."

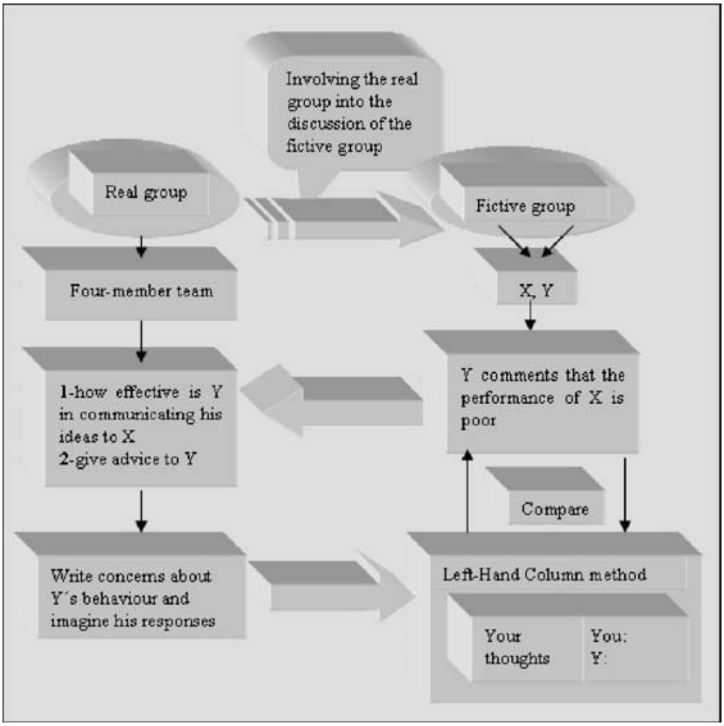


Fig. 1 Procedure for LHC (adapted from Argyris and Schoen 1974)

The team read these comments (of Y towards X) and based on that, it decided whether (1) Y had been effective in communicating with X about those issues and (2) gave Y recommendations in respect to the comments on or reproaches to X. In order to test whether the participants observed the advice provided to Y, they were asked to write, according to the format of the left-hand column method, their concerns about Y's behavior and imagine the perspective responses of Y in a conversation with this decision maker (Fig. 2).

Your thoughts and feelings	What you and Y said
(Tell here what you thought, but what you did not disclose to Y)	I: Y: I: Y:

Fig. 2 Left-hand column in a fictive conversation with Y (Argyris 1985)

The method shows if the team members generate the same problems, issues, and comments as Y. Moreover, it enables them to discover if their theory-in-use (what they actually do as demonstrated by the left-hand column) converges to or diverges from their espoused theories (what they say they would do such as the advice given to Y). The facilitator should look at the responses of the participants and point out if they are making use of Model I and help them develop Model II for the next step of knowledge transfer (see Argyris 1985 for more details on Model I and II). Negative or control attitudes, determining the inconsistencies of the participants as depicted by the left-hand column method, should be stressed.

Knowledge Transfer

The individuals underwent knowledge transfer by running the supply net game (Scholz-Reiter et al. 2006). This simulation game is designed to be played in a team of four persons. Players control the four inventories included in their respective factories, and place orders for the replenishment of the stocks. While regulating their inventories, players should pay attention not to let them drop too much because of the out-of-stock penalty of €1.0 per item per minute. Meanwhile, they should impede the build up of stocks because of the €0.5 cost for each product on hold. The game's objective is the minimization of the total inventory costs. Debriefing sessions are planned to evaluate the thoughts of the subjects and discuss the dynamics of the system, feedback loops, and behavioral patterns of variables such as customers' orders.

This structural setting not only favors organizational learning but also advances interorganizational learning. In other words, learning takes place between the team members of the food supply chain. When players quit the session for whatever reason, it is still possible to retrieve what they experienced since they transferred and shared their knowledge with the rest of the team. Hence, the player who replaces the one who left will immediately benefit from the knowledge of the latter through the team. In this way, people relate to one another's experiences in the food supply chain.

4 Conclusion

Sharing and transferring knowledge in the food supply chain increases the competitive advantage of the whole chain and guarantees the continuity of its operations. The globalized world with the associated technological innovations affords a prospective environment for the development of knowledge management. In spite of some commonplace barriers such as the lack of trust capable of generating delays in practice, the moral and social responsibility of retailers in the food supply chain empowers them to play a significant role for gradually accelerating the change processes, and hence, eliminating the aformentioned barriers. Therefore, it is important that retailers recognize their major role and learn to share knowledge efficiently.

The methodology elaborated around the learning laboratory showed how to share and transfer knowledge in a food supply chain. This approach can be utilized in the case of the networked or extended supply chain. First, knowledge is shared through the knowledge elicitation of the team's mental models. Second, knowledge is transferred via a learning-based environment such as a gaming simulation tool. Few knowledge management initiatives have addressed extended supply chains. This is why the paper certainly bridges a gap by its contribution to this research arena. Even though this methodology focuses on the food industry, it can be applied to other industries, with a few adjustments, to deal with the real problem at issue. Knowledge elicitation via the left-hand column method should always be part of the learning laboratory. This lab represents an effective strategy to knowledge sharing and transfer in food supply chains considered as vulnerable to team rotation or change in today's highly dynamic environments. All in all, the methodology proposed is highly flexible and can be practically applied. Finally, the novel adaptation of the left-hand column method as introduced, with the reliance on a fictive group, is thought to lessen the barriers to knowledge sharing in food supply chains.

The learning lab was successfully employed for the instance of a production network manufacturing items whose demand is quite variable and oscillating within the network. This phenomenon is referred to as the bullwhip effect in a linear supply chain. The participants of the workshps were asked to remedy for this situation when replenishing the inventories of their manufacturing units by minimizing the inventory costs of either the products available in stocks or those backlogged. Preliminary experimental results show that the best performing team is one that experienced a mental model elicitation (Scholz-Reiter and Delhoum 2007). Consequently, the learning lab is assumed to have influenced this team to achieve this quite positive outcome.

References

Alavi M, Leidner D (2001) Review: knowledge management and knowledge management systems: conceptual foundations and research issues. MIS Quarterly 25(1): 107–136

Argyris C (1985) Strategy, change and defensive routines. Pitman Publishing Inc., Marshfield, Massachusetts

Argyris C, Schoen D (1974) Theory in Practice. Jossey-Bass, San Francisco

Barson R, Foster G, Struck T, Ratchev S, Pawar K, Weber F, Wunram M (2000) Inter- and intra-Organisational barriers to sharing knowledge in the extended supply-chain. In: Standford-Smith B, Kidd PT (eds) E-business – Key Issues, Applications, Technologies, IOS Press

Bender S, Fish A (2000) The transfer of knowledge and the retention of expertise. Journal of Knowledge Management 4(2): 125-137

Berends H (2005) Exploring knowledge sharing: moves, problem, solving and justification. Knowledge Management Research & Practices 3: 97-105

Bourlakis M, Weightman P (2004) The food supply chain management environment. In: Bourlakis M, Weightman P (eds) Food supply chain management, Blackwell Publications, Oxford, pp. 1-10

Davenport T, de Long D, Beers M (1998) Successful knowledge management projects. Sloan Management Review 39(2): 43-57

Nickerson J, Zenger T (2004) A knowledge-based theory of the firm: the problem solving perspective. Organization Science 15(6): 617-632

Nonaka I (1991) The knowledge creating company. Harvard Business Review 69: 96-104

Nonaka I (1994) A dynamic theory of organizational knowledge creation. Organization Science 5(1): 14-47

Scholz-Reiter B, Delhoum S (2007) The effect of enhanced collaboration on the performance of subjects for the task of inventory control. In: Thoben KD, Baalsrud Hauge J, Smeds R, Riis JO (eds) Multidisciplinary research on new methods for learning and innovation in enterprise networks. Mai 20–22, BIBA, University of Bremen, Verlagsgruppe Mainz GmbH Aachen: 149-158

Scholz-Reiter B, Delhoum S, Rekersbrink H (2006) Supply Net Game: Ein Planspiel für die Ingenieurausbildung in verteilten Produktionssystemen. Industrie Management 22 (5): 15-18

Senge P, Kleiner A, Roberts C, Ross RB, Smith RJ (1994) The Fifth Discipline Fieldbook, Strategies and Tools for Building a Learning Organization. Doubleday, New York, 1994

Seyed-Mahmoud A (2004) Improving logistics operations across the food industry supply chain. International Journal of Contemporary Hospitality Management 16(4): 263-268

Thusma M (1978) Technical communication in R&D laboratories: The impact of project work characteristics. Academy of Management Journal 21(4): 624-645

Wijnands JHM, Ondersteijn CJM (2006) Quantifying the agri-food supply chain. In: Ondersteijn CJ, Wijnands JH, Huirne RB, van Kooten O (eds) Quantifying the agri-food supply chain. Springer, The Netherlands, pp 1-10

Service Models for a Small-sized Logistics Service Provider – A Case Study from Finland

Jukka Hemilä

VTT Technical Research Centre of Finland, P.O. Box 1000, 02044 VTT, Finland

Abstract Manufacturing companies today are cutting operational costs considerably in Europe. Many globally operating companies have moved their production to lower labour cost countries, like China or India. However, some companies have decided to stay in Europe, but they then must focus on cost reduction. One solution for decreasing operational costs is outsourcing. This paper is a case example from the Finnish manufacturing industry, where a small sized manufacturing company decided to focus on its core competencies. Core competencies were product development and production, but not the logistics process. The case company decided to outsource its logistics process. The challenge being, where to find a small sized logistics service provider that would be interested in the needs of the case company? The case company's requirements for services included in-bound, in-house and out-bound logistics. In our project, the other case company was the service provider. This paper describes the service provider's service offerings development process. During that project, VTT Technical Research Centre of Finland, was with the aid of the service provider, developing new logistics service offerings for industry needs.

1 Introduction

Logistics outsourcing has lately been an emerging business not only in Finland, but also on a global level. Logistics Service Providers (LSP) supply a wide range of services from express deliveries to warehousing and supply chain solutions. This kind of service provider utilisation has been common in the electronics industry now for over a decade (e.g. Jarimo and Hemilä, 2005; Sink and Langley, 1997). Global electronics manufacturers use LSPs to operate the logistic hubs, where all material flows from suppliers are integrated and the LSP directs the materials straight to the production line, but also manages the out-bound flow from production to the customers (Hemilä and Jarimo, 2005).

Recent studies have indicated that customers are requiring more value-adding services from logistics service providers, and not only warehousing and transporta-

H.-D. Haasis et al., *Dynamics in Logistics*
DOI: 10.1007/978-3-540-76862-3, © Springer 2008

tion (Naula et al. 2006; Stone 2006). Globalisation has driven companies to utilise globally-operating service providers that are able to provide the required value-added services (Naula et al. 2006). Rising requirements for services have driven LSPs to standardise their service offerings. Many globally-acting LPS have standardised their services all over the world, so every kind of tailor-made logistics service is likely to be expensive. Sink and Langley (1997) prove that the scope of service offerings may range from a relatively limited set of services to a comprehensive, fully integrated set of logistics activities.

Berglund et al. (1999) have presented a framework for the segmentation of logistics service providers, and there are two main categories for LSPs:

1. LSPs that offer a specific service, for example the distribution of spare parts
2. LSPs that cover a complete range of services and offer their customers a logistics solution

Service providers focus on a few standard services, maybe add some features to attract extra customers and use scale economies to increase profits. Solution providers focus on a few industries, take over complete, well-defined processes and customise their services (Berglund et al. 1999). There are two main categories for LSPs, ones that carry out traditional transportation and warehousing activities only, versus providers that offer additional activities, such as value-added services (Berglund et al. 1999). In Fig. 1, these two categories are indicated as basic logistics and value-added logistics (Berglund et al. 1999).

The demand for logistics services in the manufacturing industry is rising on a global level, but the challenge is the same for locally operating manufacturers.

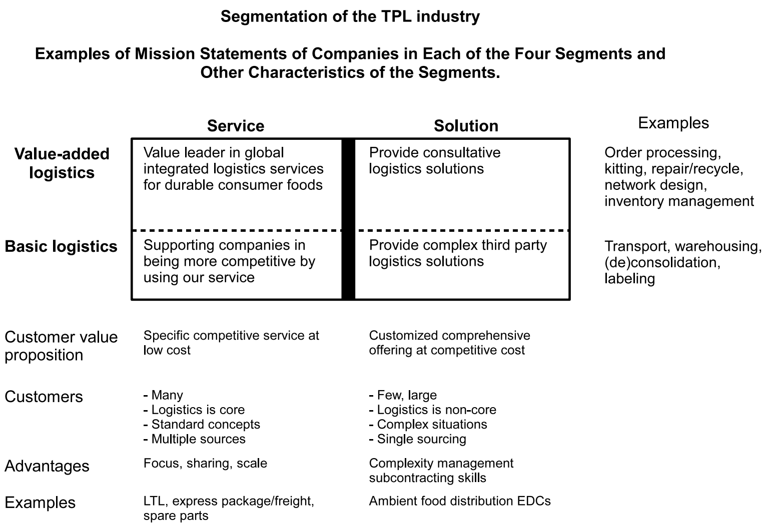


Fig. 1 Framework for segmenting service providers (Berglund et al. 1999)

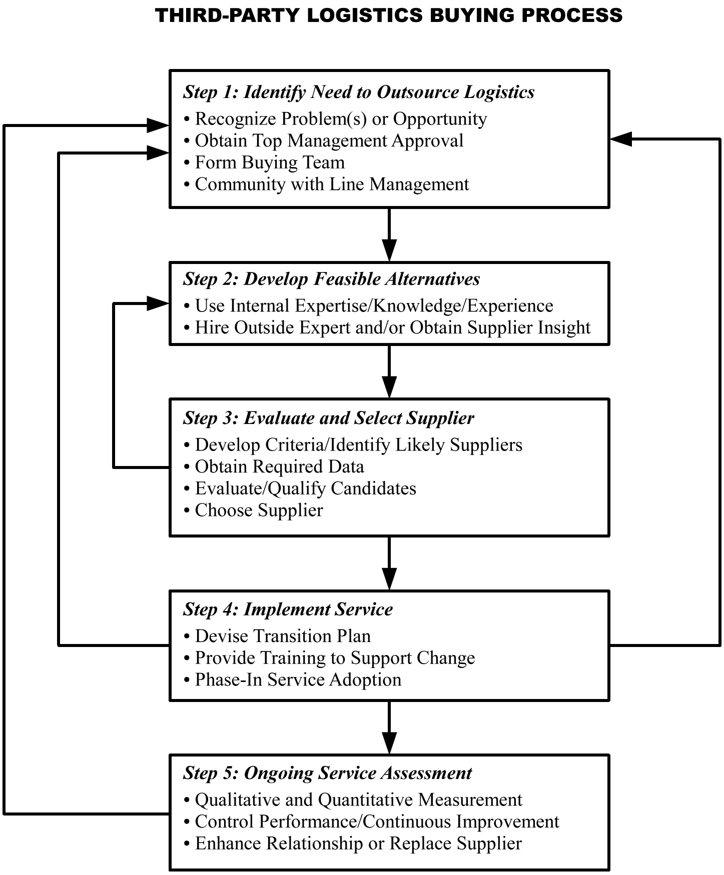


Fig. 2 Logistics service buying process (Sink and Langley, 1997)

We have noticed that many customer companies are willing to buy logistics services rather from smaller providers and not as global services.

Finnish manufacturing companies operating only in national market today are decreasing their operational costs, so one solution could be outsourcing. Logistics is seen to be one function suited to outsourcing, especially when there is a need to focus on the core competencies of the company (Sink and Langley, 1997).

Then a new challenge arises: where can they find a service provider for flexible services according to their needs? LSP are mostly large-sized global companies, which have standard outsourcing solutions. SMEs have faced difficulties in creating business relationships with large LSPs. The problem is that SME manufacturers are like peanuts to global LSPs, so LSPs are not interested in wasting energy on these small customers. Holter et al. (2007) have analysed the purchasing of logistics services in SMEs. The purchase of logistics services is central to an SME's operational needs, but very difficult for them to organise (Holter et al. 2007). Due to their size and lack of "purchasing power" SMEs may be treated as "order takers" by LSPs rather than being "order makers" who tend to have some control over the purchasing process parameters (Holter et al. 2007).

Sink and Langley (1997) have also analysed the logistics purchasing and they have developed a conceptual model for the logistics service buying process (Fig. 2). In our project, the manufacturing company managed their outsourcing process as Sink and Langley defined on their conceptual model.

The development phase of new logistics services in our project was carried out during Steps 4 and 5 of Sink and Langley's (1997) conceptual model. In our project, the service provider developed their services during the implementation phase (Step 4). We had a loop from step 5 to step 4, where through continuous improvement they implemented new services for use.

2 Service development steps

The starting point of service development was process analysis: what kind of internal processes the LSP had. The service process included nine steps:

1. Delivery request
2. Transport service (optional)
3. Material receiving and shelving
4. Material stocking
5. Material picking
6. Pre-assembly (optional)
7. Delivery to customer's site
8. Packing and delivery
9. Distribution

The service process begins with the delivery request or an order made by the customer or in some case the LSP. Then material is delivered from the supplier to the customer or to the LSP-operated warehouse. Transport could be arranged by the LSP as an optional process phase, but it could be managed by the transporting company. Then materials should be received and shelved, also information should be added to the warehouse system. Process phase four is stocking where there is usually nothing

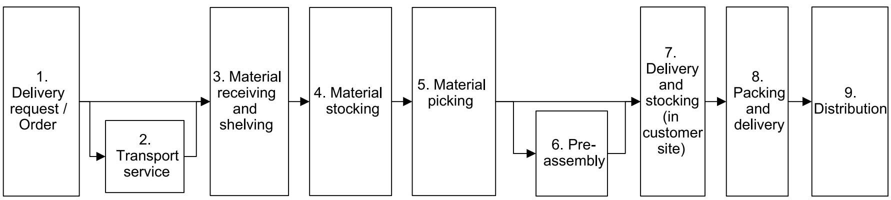


Fig. 3 Internal process for Logistics Service Provider

to do, but there could also be some follow-up of conditions (temperature, pressure, etc.). Material picking is performed when the LSP receives the information from the customer about the required items. Pre-assembly (phase 6) is also optional, it is dependent on the customer. Phase seven is delivery of the items to the place the customer wants them. After phase seven, the customer has its own manufacturing process which is followed by the LPS managing the packaging and delivery of the final products. The final process phase is distribution to the markets.

The internal process was the basis for service development. The challenge was how to make service offerings to the customer from service process phases.

3 Services for small sized LSP

The customer analysis was carried out after an internal process analysis of the LSP. According to the customer analysis results, we developed the new service offerings. The customer value chain included four phases, from the logistics point of view: 1) In-bound logistics; 2) In-house logistics; 3) Out-bound logistics; and 4) Planning and management. These four phases included many processes that the customer could outsource to the LSP.

The in-bound phase of the customer process includes in-bound and return logistics. There the LSP services are material receiving, acceptance, quality control and shelving. Another major service is the VMI (Vendor Managed Inventory) service, where the LSP is responsible for all material management straight to the customer's production line. In that case, the supplier supplies through the VMI model to the LSP-operated warehouse. Return logistics is the last of three in-bound services. Return logistics on the in-bound side means the return of disqualified, non-marketable or broken items back to supplier.

The in-house part is concerned mainly with manufacturing logistics. Our definitions of the customer's in-house processes are: procurement, manufacturing, final product sales and delivery. The LSP service offerings are then: procurement service, manufacturing logistics, sales promotion services and return logistics.

The out-bound part is simply out-bound logistics and return logistics, from the customer's point of view. The services for out-bound are then: maintenance/service logistics, delivery/distribution and return logistics.

The fourth main process in the customer value chain was planning and management. That process covers the other three processes. LSPs could offer services for the customer's planning and management. We identified that those services could include information logistics and knowledge services. Information logistics could be data about warehouse levels, lead-times, delivery accuracy, etc., which could be shared with suppliers and customers. The knowledge services could then be, for example, consultation about how to improve the logistic process through outsourcing.

Figure 4 shows our framework for the customer logistics process and how LSP service offerings meet the customer's needs at every process phase.

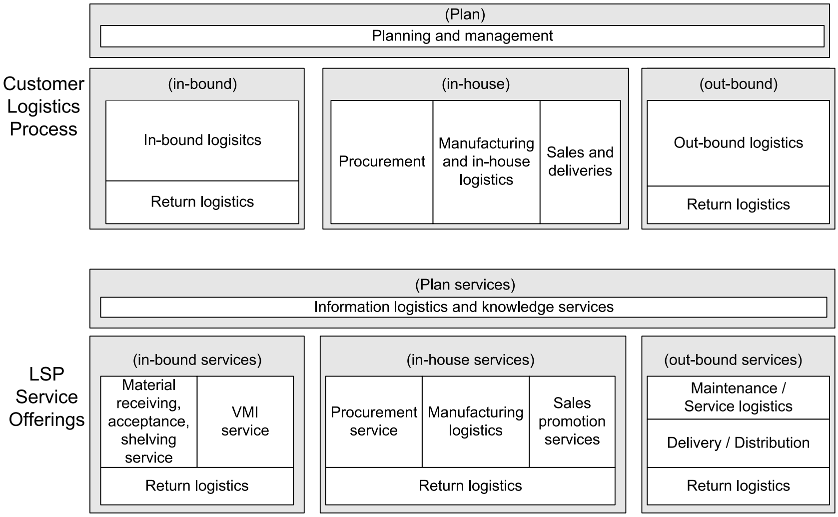


Fig. 4 Model for LSP services for the Customer's value chain

The service offerings were developed and implemented during our project. Still continuous improvement and other changes, particularly in cases of new customers, must continue.

4 Conclusions

We have noticed that there is a huge need for an industrial logistics service, particularly focused on small- and medium-sized manufacturers. There are not so many "total outsourcing" service providers in Finland, particularly not ones which could manage all of the logistics processes. This paper presented one example of service development.

For small- and medium-sized manufacturing companies, the logistics management could require a lot of resources, which could be utilised better during the manufacturing process. In Finland, manufacturing labour is a hard to find these days. As a result, SMEs are willing to focus on manufacturing, use labour effectively during the production process and outsource all logistics operations. It is then hard to find such a service provider providing in-bound, in-house and out-bound logistics. Anther important issue in outsourcing is the measurement of costs. As our case company managed the logistics process earlier by themselves, the costs were made up of the manufacturing costs. The LSP could manage logistics much more effectively, which also offers a cost advantage for the customer. Operational costs could then be analysed better and also business performance could be measured. Measurement has a direct effect on business management.

The logistics service providing business is currently consolidating, big player are becoming even bigger. Still there are some small companies coming to the LSP markets, and those companies are the most innovative and flexible and are usually more focused on the needs of small- and medium-sized customers. There is still room for the new entries to the markets, and it is good to have local competitors for global players. There are many reasons why LSP markets are growing, but the need to control costs and improve services are the main drivers. For small sized LSPs, the cost advantage is realised through economies of scale.

The big challenge in outsourcing is information technology utilisation. There is a need for effective information flow, but it is quite expensive to create a transparent information flow through suppliers to the LSPs and to the customer.

References

Berglund, M., van Laarhoven, P., Sharman, G., Wandel, S. (1999): Third-party logistics: Is there a future? International Journal of Logistics Management, 1999, 10, 1, p. 59–70.

Hemilä, J. and Jarimo, T. (2005) Hub-Model Utilisation in Inbound Logistics – A Case Study from the Finnish Electronics Industry. Conference Proceedings of the ICIL2005, Montevideo, Uruguay, Feb. 14–18, 2005.

Holter, A.; Grant, D. and Ritchie, J. (2007): Purchasing Transport And Logistics Services In Small And Medium Size Enterprises. The conference proceedings of The 19th Annual NOFOMA Conference. 7-8 June 2007 Reykjavik Iceland.

Jarimo, T.; Hemilä, J. (2005): An Inbound Logistics Hub Model for Supply Chain Management. EurOMA International Conference on Operations and Global Competitiveness, Budapest, Hungary, June 19-22, 2005, Demeter, Krisztina (Ed.). Diamond Congress Ltd. (2005), 925–932

Naula, T., Ojala, L. and Solakivi T. (2006): Finland State of Logistics 2006. Publications of the Ministry of Transport and Communications 45/2006. Edita Publishing Ltd, Finland.

Sink, H. and Langley, J. (1997): A managerial framework for the acquisition of third-party logistics services. Journal of Business Logistics; 1997; 18, 2; p. 163-189.

Stone, M. (2006): Effect of globalization on growth strategies for European logistics service providers. Conference Proceedings of the Logistics Research Network 2006. 427-432.

Container Logistics

A Framework for Integrating Planning Activities in Container Terminals

Seung Hwan Won, Kap Hwan Kim

Dept. of Industrial Engineering, Pusan National University, South Korea

Abstract This paper introduces various planning activities in container terminals. And their functions, planning horizons, and required resources are described. This paper also proposes a new framework for various planning activities in container terminals. The new framework considers capacities of various resources related to each planning activity, simultaneously. The resources include not only the storage space but also the handling equipment. For considering the handling capacity explicitly, a concept of the resource profile for each planning activity is proposed. It is shown how the resource profile can be utilized for checking the feasibility of an operational plan in container terminals.

1 Introduction

Port container terminals are located in shores and places where containers are transshipped from one transportation mode to another. Their main function is to provide transfer facilities for containers between vessels and land transportation modes such as trucks and trains.

Operations in container terminals can be classified into the ship operation, during which containers are discharged from and loaded onto a vessel, and the receiving and delivery operation during which containers are transferred from and to external trucks. During the operations, the assignment of resources to tasks and the scheduling of these tasks are done and thus decision-making methods for the assignment have been major planning issues in container terminals.

Many papers reviewed the planning problems for operations in container terminals (Ramani 1996; Meersmans and Dekker 2001; Vis and de Koster 2003; Steenken et al. 2004; Murty et al. 2005; Günther and Kim 2006; Crainic and Kim 2007). This study provides a new template for various planning activities in container terminals. This template is a general procedure that can be applied to various planning problems in container terminals. First, we identify planning problems and describe decision variables for each problem. Second, we investigate the input parameters and

H.-D. Haasis et al., *Dynamics in Logistics*
DOI: 10.1007/978-3-540-76862-3,

the decision process for the decision activity. Note that the input parameter has to include all related resources for each decision activity. We also define the length of each planning time bucket and the planning horizon for the decision activity. Third, we suggest a method to construct the resource profile for each planning activity.

Section 2 provides a framework for a planning procedure in container terminals. Section 3 discusses resource profiles and their usages for various operational plans. Finally, Sect. 4 presents a conclusion.

2 The framework for a planning procedure

Container terminals are carrying out various handling operations by utilizing such facilities as a quay or berth, quay cranes (QCs), a storage yard (SY), yard cranes (YCs), straddle carriers (SCs), and yard trucks (TRs). Efficiency of the operations depends on the capacities of the resources and how to utilize the capacities of the resources.

The capacities of these recourses are represented as follows. The capacities of berths are the product of one dimensional space (usually length) and time. The capacities of QCs, TRs, and YCs are measured in QC times, TR times, and YC times, respectively. If all QCs have the same productivity, we can evaluate the capacity of QCs by multiplying the number of QCs by their available time. In a similar way, the capacities of the other resources can be evaluated. Table 1 represents the units of each resource.

Every planning activity for operations in container terminals requires related resources. The amount of resources required by an activity can be estimated as follows. Berthing a vessel requires the resource of the berth by the amount of the length of the vessel multiplied by the occupation time. Unloading or loading containers consumes the resource of QCs by the amount of the number of containers multiplied by the standard handling time per container. The workloads of the unloading and the loading operation on TRs can be evaluated by multiplying the number of containers with the average transportation time per container including empty travels. The workloads of unloading and loading operations on YCs are calculated by multiplying the number of containers with the standard handling time per container. The workload on TAs is added by the expected future occupation of TAs

Table 1 Resources and their units

Resources	Units
Berth (H)	Length of the berth × Time
QC (C)	Number of QCs × Time
TR (R)	Number of TRs × Time
YC (Y)	Number of YCs × Time
TA (A)	Number of TRs passing the TA × Time
SY (S)	Number of slots (in TEU) × Time

by TRs. The storage spaces of the SY are consumed by the reservation before the storage of containers and the actual occupation by the containers.

Every planning process must consider the availability of the related resources. Fig. 1 shows several key planning processes and their related resources to be checked.

The berth planning is a decision on the berthing location and time for the ship. A certain zone of the berth is assigned to a ship for loading and unloading containers for a certain period of time. QC work scheduling (split) is a decision on the service sequence of bays in a ship by each QC and the time schedule for the services. Several QCs are usually assigned to one ship. Yard planning is a decision on the storing locations for containers during unloading operation, receiving operation, and remarshalling operation. Unload scheduling is a decision on the work sequence and time for the containers which are discharged from the ship to the storage location in the SY. Load scheduling is a decision on the work sequence and time for the containers that are loaded from the SY into the stowage location in the ship. Remarshal planning is a decision on the changed storing position (from one storage block to another) and time of the containers in the SY.

Decisions to be made by each planning process are summarized in Table 2. The first line of each planning process represents the activity or the activity unit on which a decision will be made. The second line represents the reference moment of the activity based on which the time-phased additional amount of resource consumption resulting from the decision will be estimated. The third line shows contents of the decision to be made by the corresponding planning process.

When the same resource is used in different decisions, the aggregation level for the resource may be different among different types of decisions. The capacity of berths is represented by the total length of them multiplied by the length

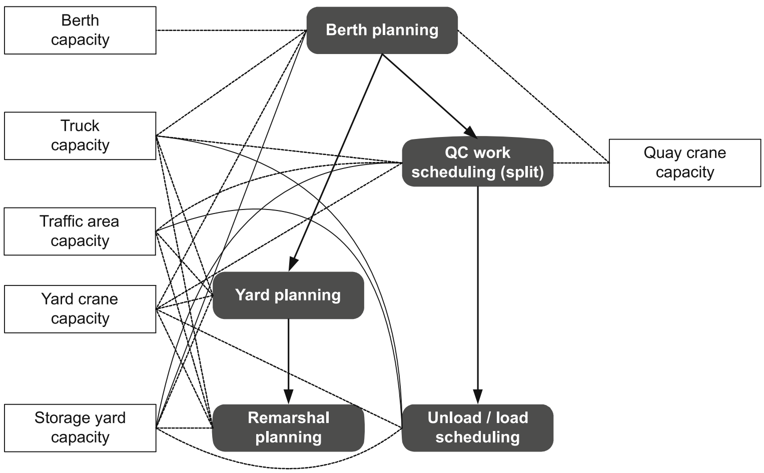


Fig. 1 Various operational plans and their related resources

Table 2 Contents of the decision for various operational plans

Berth planning (B)	
Activity to be planned	Berthing of each vessel
Reference moment of the activity	The beginning of the berthing
Contents of the decision to be made	Berthing position and time of each vessel
QC work scheduling (Q)	
Activity to be planned	Loading or unloading task on deck or in hold of a bay by a QC
Reference moment of the activity	The beginning of each task
Contents of the decision to be made	A schedule for a QC to discharge and load containers on a vessel
Yard planning (P)	
Activity to be planned	Receiving outbound containers of a group for a vessel or unloading of inbound containers by a QC for a vessel for a period
Reference moment of the activity	The start of arrivals of outbound containers at the gate or the unloading of inbound containers from the vessel
Contents of the decision to be made	Storage positions for receiving or unloading containers
Remarshal planning (M)	
Activity to be planned	Moving a set of containers from one block to another for a period
Reference moment of the activity	The start of the moving
Contents of the decision to be made	The containers to be moved and their source and destination positions for a period

of a period. The capacities of entire QCs, TRs, and YCs are represented by the number of available pieces of equipment multiplied by the length of available time for a period. The capacity of TAs is represented by the total number of TRs that can pass the TA for a period. The capacity of the entire SY is represented by the total number of available slots multiplied by the length of a period. Moreover, the capacities of QCs, YCs, TAs, and the SY can be divided into a more detail level.

3 Resource profiles for various activities

We consider six types of resources (berths, QCs, TRs, YCs, TAs, and the SY). There are three types of routings for container flows (inbound, outbound, and transshipment flows). The operation time of the resources are the berthing time, the handling time by QCs, the transportation time by TRs, the handling time by YCs, the transfer time in the TAs, and the storage time in the SY, respectively.

For describing the resource profiles, the following notations are used:

Indices

r The index for the type of resources where $r = H$ (*berth*), C (*QC*), R (*TR*), Y (*YC*), A (*TA*), and S (*SY*).
t The index for periods where $t = 1, 2, \ldots, m$.
a The index for activities where $a = 1, 2, \ldots, n$.

Problem data

c_r^t The entire capacity of resource r at period t. An activity can be scheduled only when the total workloads on the resources required for the activity do not exceed the capacities of the resources.
r_r^t The remaining capacity of resource r at period t. This can be evaluated by the capacity of a resource minus the total workload on the resource.
s_{ar}^t The amount of resource r which must be consumed with the time offset of t for carrying out activity a. In berth planning, for example, if a vessel is decided to berth at a quay at period p (considering this as activity a), the resource r at period $(p+t)$ will be consumed by the amount ofs_{ar}^t. This is the data of the resource profile.

Sets

P_{ar} The set of time offsets in which resource r is consumed by activity a.
B The set of activities related to berth plans. Each activity corresponds to a decision on the berthing of a vessel.
Q The set of activities related to QC work schedules. Each activity corresponds to a decision on the work for a vessel by a QC.
P The set of activities related to yard plans.

In the following, resource profiles for the berth planning are illustrated. The berth planning determines the berthing position and time of a vessel. Table 3 shows the

Table 3 The definition of the berth planning problem

Input parameters	Calling schedule of vessels
	Favorable berthing location of each vessel
	Length of each vessel
	Draft required for each vessel
	Number of unloading and loading containers of each vessel
	Resource profiles
Decision variables	Berthing position of each vessel
	Berth time of each vessel
Objectives	To minimize the delays in the departure of vessels
	To minimize the travel distances between shore and yard for all containers of vessels
Constraints	Depth of water for berths
	Due time for the departure of vessels
	Availability of resources

input parameters, the decision variables, the objectives, and the constraints for the berth planning. For constructing a berth plan, the calling schedule, the favorable berthing location, the lengths of vessels, the required draft of each vessel, and the required unloading and loading time of each vessel are necessary. For a berth plan to be a good one, vessels must berth at the most favorable positions – which will reduce the container delivery time between the marshaling yard and QCs – and also must be able to depart the port before their committed due times, which is an important service criterion of the berth planning process.

The requirements of resources depend on the number of unloading and loading containers and are evaluated by using the load profiles of the resources for the berth planning. The requirements of each resource can be time-phased with respect to the berth time of the vessel. The decision variables of the berth planning consist of the berthing position and the berth time of each vessel.

By adding the schedule of a vessel to the berth plan, related resources will be required as shown in Figs. 2 a) and b). Figures 2 a) and b) show the resource profile for outbound flows and inbound flows, respectively.

The resource profile can be evaluated as follows: s^0_{BH} indicates the amount of berths required for a berth plan B. It can be evaluated by using (the length of the

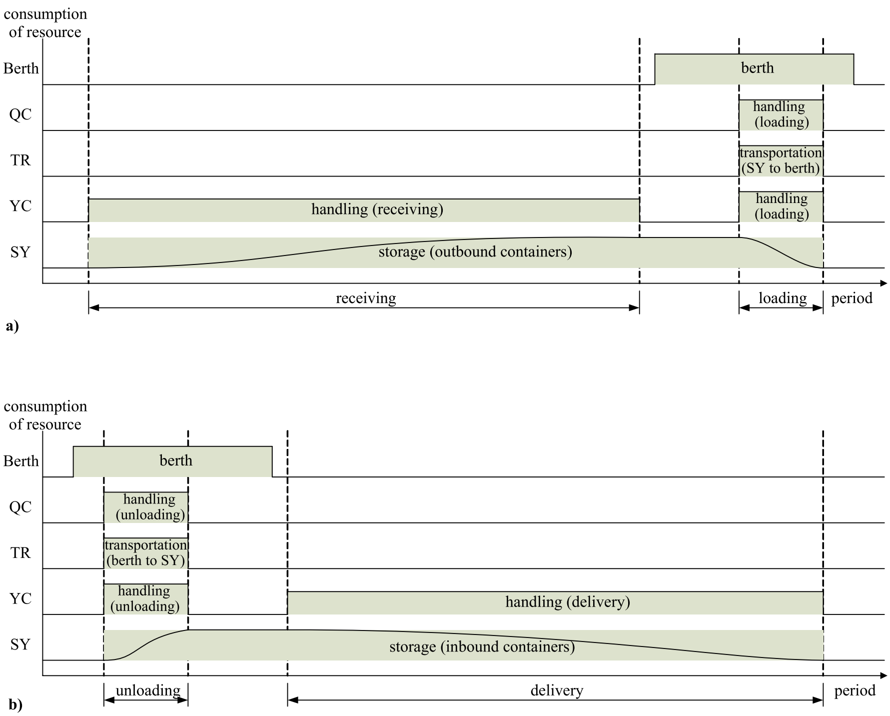


Fig. 2 Resource profile for berthing of a vessel; **a)** Resource Profile for outbound containers; **b)** Resource profile for inbound containers

vessel plus the allowance between adjacent vessels) × (the berthing duration of the vessel). s^0_{BC} can be evaluated by using {(the time for a QC to transfer an outbound container to a slot of the vessel) × (the percentage of loading containers among all containers for the vessel) + (the time for a QC to transfer an inbound container to a TR) × (the percentage of unloading containers among all containers for the vessel)}. The transfer operation by QCs can be done while the vessel stays at the berth. The standard time, which includes not only the pure operation time but also the unavoidable delay and the personal and fatigue allowance, must be used.

Because TRs are required for the loading and the unloading operations, s^0_{BR} can be evaluated in the similar way as s^0_{BC}. s^t_{BY} for $t < 0$ can be evaluated by using (the time for a YC to receive an outbound container from an external truck) × (the percentage of containers, among all the outbound containers, arriving at the SY on the tth day before loading). s^t_{BY} for $t > 0$ can be evaluated by using (the time for a YC to transfer an inbound container to an external truck) × (the percentage of containers, among all the inbound containers, leaving the terminal on the tth day after unloading). s^t_{BY} for $t = 0$ can be evaluated by using {(the time for a YC to transfer an outbound container to a TR) × (the percentage of loading containers among all containers for the vessel) + (the time for a YC to receive an inbound container from a TR) × (the percentage of unloading containers among all containers for the vessel)}. s^t_{BS} for $t \leq 0$ can be evaluated by using (the cumulative percentage of containers, among all the outbound containers, having arrived at the SY until the tth day before loading) × (the length of a period). s^t_{BS} for $t \geq 0$ can be evaluated by {(1 – the cumulative percentage of containers, among all the inbound containers, having left the terminal until the $(t-1)$th day after unloading) × (the length of a period)}.

As an example, we illustrate hows^t_{BY} for $t < 0$ can be evaluated in the following. In order to evaluates^t_{BY} for $t < 0$, we have to find two things, the time for a YC to receive an outbound container from an external truck and the percentage of containers, among all the outbound containers, arriving at the SY on the t^{th} day before loading. Lee and Kim (2007) proposed formulas for estimating the cycle time of various types of yard crane in container terminals. They also analytically formulate expectations and variances of the cycle time for the receiving, loading, unloading, and delivery operations. We can use their results to estimate handling times for receiving operations of a YC. Table 4 represents an example of s^t_{BY}, for $t < 0$. The expected handling time for a receiving operation of a YC used in the example is 1.521 minutes.

Table 4 An example of s^t_{BY}, for $t < 0$

t	−6	−5	−4	−3	−2	−1
Percentage of containers	2%	5%	6%	10%	12%	65%
s^t_{BY}	0.030	0.076	0.091	0.152	0.183	0.989

Table 5 Data used for calculating the resource profile in the berth planning

The length of the vessel plus the allowance between adjacent vessels	300 meters
The berthing duration of the vessel	18 hrs or 1,080 mins
The number of loading containers for the vessel	540
The number of unloading containers for the vessel	560
The time for a QC to transfer an outbound container to a slot of the vessel	1.9 mins
The time for a QC to transfer an inbound container to a TR	1.9 mins
The turnaround time for a TR to travel between shore and yard	10 mins
The time for a YC to receive an outbound container from an external truck	1.521 mins
The time for a YC to transfer an inbound container to an external truck	2.242 mins
The time for a YC to transfer an outbound container to a TR	1.134 mins
The time for a YC to receive an inbound container from a TR	1.114 mins
The storage duration of the container for a period	24 hrs or 1,440 mins

Table 6 Percentage of containers left the terminal on the *n*th day after unloading

n	1	2	3	4	5	6
Percentage	34%	22%	15%	14%	11%	4%

Table 7 A resource profile for berthing of a vessel (s_{Br}^{t}) (unit: minute)

t / *r*	−6	−5	−4	−3	−2	−1	0	1	2	3	4	5	6
H	-	-	-	-	-	-	324,000	-	-	-	-	-	-
C	-	-	-	-	-	-	1.9	-	-	-	-	-	-
R	-	-	-	-	-	-	10	-	-	-	-	-	-
Y	0.030	0.076	0.091	0.152	0.183	0.989	1.124	0.762	0.493	0.336	0.314	0.247	0.090
S	28.8	100.8	187.2	331.2	504	1,440	1,440	1,440	950.4	633.6	417.6	216	57.6

Table 5 summarizes data used for calculating the resource profile. Tables 4 and 6 show the percentage of containers arrived at the SY on the days before loading and the percentage of containers left the terminal on the days after unloading, respectively. Table 7 illustrates the resource profile for berthing of a vessel.

4 Conclusion

This paper proposes a new framework for various planning activities in container terminals. In previous planning processes, handling capacities have not been considered explicitly. Instead, they have been considered as constraints or rules for decision-making. This paper showed how the handling capacity can be considered explicitly during the planning process. For that purpose, a concept of the resource profile for each planning activity is proposed. The resource profile for the berth planning was illustrated.

Acknowledgements This research was supported by the MIC(Ministry of Information and Communication), Korea, under the ITRC (Information Technology Research Center) support program supervised by the IITA (Institute of Information Technology Advancement) (IITA-2006-C1090-0602-0013).

References

Crainic TG, Kim KH (2007) Intermodal transportation. In: Barnhart C, Laporte G (eds) Handbooks in Operations Research and Management Science: vol 14, Transportation, North-Holland, Amsterdam, chap 8

Günther H-O, Kim KH (2006) Container terminals and terminal operations. OR Spectrum 28: 437–445

Lee BK, Kim KH (2007) Cycle time models for yard cranes considering block layouts in container terminals (in Korean). Journal of Korean Institute of Industrial Engineering 33: 110-125

Meersmans PJM, Dekker R (2001) Operations research supports container handling. Econometric Institute Report EI 2001-22, Erasmus University Rotterdam, Rotterdam

Murty KG, Liu J, Wan Y-w, Linn R (2005) A decision support system for operations in a container terminals. Decision Support Systems 39: 309-322

Ramani KV (1996) An interactive simulation model for the logistics planning of container operations in seaports. Simulation 66: 291-300

Steenken D, Voß S, Stahlbock R (2004) Container terminal operation and operations research – a classification and literature review. OR Spectrum 26: 3-49

Vis IFA, de Koster R (2003) Transshipment of containers at a container terminal: an overview. European Journal of Operational Research 147: 1-16

Vollmann TE, Berry WL, Whybark DC (1997) Manufacturing planning and control systems, 4th edn. McGraw-Hill, New York

Won SH, Kim KH (2006) Allocating storage spaces for temporary inventories considering handling, transportation, and storage capacities (in Korean). Journal of Korean Operations Research and Management Science Society 31: 11-25

Electronic Seals for Efficient Container Logistics

Kateryna Daschkovska[1], **Bernd Scholz-Reiter**[2]

[1] International Graduate School for Dynamics in Logistics, University of Bremen
Hochschulring 20, 28359, Bremen, Germany
[2] BIBA IPS, University of Bremen
Hochschulring 20, 28359, Bremen, Germany

Abstract New security requirements for international container trade are forcing changes in global supply chain processes. Simultaneously, shippers look to advanced technology to enhance security and efficiency of their container logistics. Customs inspections at international borders, especially in maritime ports, are one of the biggest inefficiency factors, leading to delays and additional logistics costs. An electronic seal as an important part of a multi-layered security system has the ability to enhance the security as well as to improve the efficiency of container logistics. The paper is focused on the problem of reducing customs inspections variation at the ports, and evaluating investments in electronic seals. As a number of world security initiatives require implementing electronic seals on containers, the paper tries to measure the influence of different types of eSeals on the performance of container logistics and with special consideration of customs inspections in the ports.

1 Introduction

After the terrorist incident in 2001 the cost associated with the closing of U.S. ports for a few days surpassed $1 trillion [6]. Consequently the maritime industry needed to solve thousands of problems connected with new high security requirements that impact the efficiency of ever-increasing "just-in-time" container movements. Companies that use maritime transportation as a part of "just-in-time" logistic process become more vulnerable to changes and disruptions in their supply chains. The key point affecting the efficiency of "end-to-end" shipping, especially in maritime business, is customs inspections of containers in the ports. Processes of container transportation are vulnerable to security threats and uncertainties concerned with them. The biggest problem points for shippers in international container supply chains are uncertainties and lead time in customs and in transit and visibility of their cargo en route from the point of origin to the destination point. Existing capabilities to reduce the customs lead time for container processes and increase visibility in supply chains are still quite limited.

H.-D. Haasis et al., *Dynamics in Logistics*
DOI: 10.1007/978-3-540-76862-3,

As a result of increased security requirements on international supply chains, technology can be an important bridge between efficiency and security of supply chain process. Electronic seal technology has such potential. ESeals serve both commercial and security interests by tracking commercial container shipments from their point of origin, while en route, through customs, and to their final destination. Companies, that voluntary take part in security programmes and are certified by U.S. Customs and Boarder Protection (CBP) can expect reduced container inspections and related importing costs. The set of requirements for participation includes application with a written copy of the company's supply chain security plan, that approved by CBP and then the intensive process of business validation. In addition, firms that implement additional security measures such as "smart" containers equipped with electronic seals may apply for "third-tier" status, the highest level of security validation. Third-tier companies will benefit from use of "green lane" advantage for containers, meaning that their containers will clear customs without undergoing inspection [2].

Besides fewer delays at border controls, the implementation of high-level security in container supply chain processes produces positive effects on efficiency on the whole process such as reducing logistics costs, decreasing inventory stocks and increasing customer service levels.

Thereby, the article addresses the following research questions:

1. How can shippers benefit from investments in electronic seals?
2. What monetary benefit results from avoiding customs inspections in the ports?
3. What type of electronic seal is the most profitable over a 5-year investment period?

2 Container Electronic Seals

Electronic Seals (eSeals) present a very simple and at the same time very strong defence against different weaknesses in container safety of containers in worldwide trade. By implementing various types of eSeals it is possible to enhance container security as well as improve efficiency of container logistics processes throughout the whole supply chain.

According to current ISO 18185 [4] definition an eSeal is a "Read-only, non-reusable freight container seal conforming to the high security seal defined in ISO/PAS 17712 [5] and conforming to ISO 18185 or revision thereof that electronically evidences tampering or intrusion through the container door". At the current time there are no global standards for frequencies and technical specification for electronic seals. The International Standards Organization's (ISO) Technical Committee 104 is trying to specify data protection technology, and as a result, ISO 18185-4 Gen 1 was released in August 31, 2005. However, the ISO 18185-4 Gen 1 did not satisfy requirements for the data protection and device authentication for eSeals.

In this paper we introduce potential opportunities of eSeals to enhance the efficiency of lead times of containers going through customs inspection processes at a port. An eSeal provides not just physical security but also can contain a set of useful information such as seal number, container number, user data, security, battery and environment statuses and some different data useful for supply chain management. An electronic seal has the great advantage of maintaining visibility en route and allows for real-time event reports using satellite or cellular communications. With high-end capability it becomes more attractive for security and business applications.

Finally, the adaptation of eSeals is connected with additional expenses for container business providers and for government in particular. The problems of investments in such security devices on containers are evaluated in the paper.

3 Cost-Effective Investments and Returns on ESeals

Heightened container security requirements, especially for import operations, influence the container customs inspection process in ports. Extensive inspections become a bottleneck for container supply chains that slow it down and decrease operational performance. Tightened security control traditionally means increasing the inspection rate for container content rather than enhancing the effectiveness of it [3]. The effect of increased random sampling checks of containers leads to extended dwell time of containers in a port through the increasing number of containers waiting for inspection and additional logistic costs.

The greatest part of world container volume is repetitive: the same shippers make the same shipments for the same consignees with a large number of containers. These routine container flows are quite easy to "secure" and can be placed under less precise customs scrutiny [1]. Participation in the C-TPAT program and third-tier status will give companies an opportunity to avoid the additional expenses on containers waiting for inspection, to save costs of manual customs inspections itself and to accelerate the container turnover through the whole supply chain. Electronic seals, as a necessary part of "green lane" advantage, lead to additional investments in "secure" container logistic processes.

We present the evaluation of cost-benefit influence of three different eSeals on efficiency of container supply chain by using the "green lane" opportunity. Cost-Benefit Analysis (CBA) allows us to weigh the total expected costs against the total expected benefits of one or more alternatives in order to choose the best or most profitable option. The process involves monetary calculations of expected expenses versus expected returns on investments.

The cost-benefit calculations of eSeals investments involve, at first, using time value formulas for Present Value (PV) and Net Present Value (NPV) calculations as well as Profitability Index (PI); and secondly, formulas based on record appraisals like Payback Period (PP) and Return on Investments (ROI).

To estimate additional logistic expenses caused by investments in electronic seals we compare in this paper three types of seals:

1. "Container Security Device" (CSD), containing seal ID number, container ID number, additional sensors to indicate environmental status of container content, alarm function to inform in real-time and satellite communication via GPS/INMARSAT systems. With the ability to provide real-time global visibility for container supply chain, CSD has the highest level of costs for its use. We assume that one CSD costs 4,000 € [7] and the possible rent is equal to 10% from the original price, i.e. 400 € per trip. We assume additional operating costs for container sealing (2.5 € per trip) and costs for information transaction (3 € per trip), when the minimum number of transactions is equal to 6.
2. Reusable or permanent active RFID eSeals also include a seal ID number, a container ID number; can initiate alarm calls and record time/date of container tampering. The current problem of RFID seals is that, for its implementation, they need worldwide RFID reading infrastructure. The question of who will invest to build a global RFID net is still under discussion among many parties involved in container logistics. We assume that one permanent eSeal costs 500 € [7] and the possible rent is equal to 10% from the original price, i.e. 50 € per trip. We assume both additional operating costs for container sealing (1 € per trip) and costs for information transaction (3 € per trip). The minimum number of transactions is equal to 6.
3. Read-only, non-reusable eSeal defined in ISO 18185. This type of electronic seals contains only a seal ID number and is attractively priced in comparison to the two previous eSeals. The most common technology for such eSeal is passive RFID, which requires an appropriate reading device and software. We assume that one-used eSeal costs 5 € per trip. We assume as well that operating costs for container sealing is 5 € per trip and costs for information transaction is 3 € per trip, the minimum number of transactions is equal to 6.

We assume that necessary global reading infrastructure already exists for RFID transponders or will be a part of investments from the government. The RFID reading infrastructure is the most extensive share of all investments in eSeals and the issue of its responsibility is still open. The annual repair costs for eSeals are assumed as 1% from original price.

We assume that 1,000 containers move through the ports and each container makes 10 trips per year.

We consider two scenarios for containers shipping through the ports:

1. **Scenario without eSeals.** When containers enter the port of loading A, as shown in Fig. 1, two alternatives of physical container flows exist: the container goes directly to the waiting area to be loaded on the ship, or it is moved to the customs checking territory to be manually inspected by customs. The latter alternative brings additional costs for container logistics as well as extends container dwell time at the port, which in turn creates additional expense. We assume that the volume of such expenses is equal to 1% of lost sales. We assume inspection costs to be 100 € per container. The realistic rate of customs inspections is

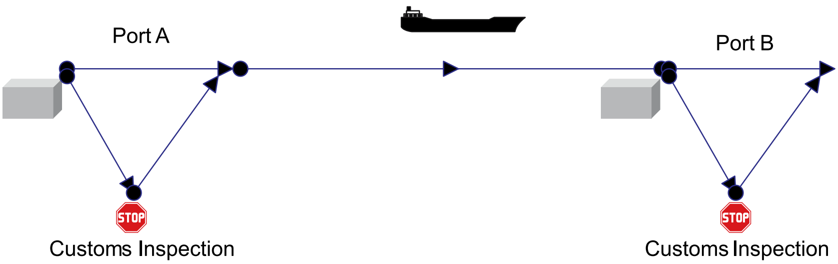


Fig. 1 1st Scenario for container processing through the ports under customs inspections

2% of total number of containers. We assume as well that inspections cause additional storage costs for a container at a port. It is assumed that storage costs 1 € per day. The number of storage days due to inspection control is 7 days in the port of loading and 4 days in the port of discharge. Following inspection container a will be loaded on a ship and directed to a port of discharge B. We consider 3 possible variants of physical inspections: first, a container is checked once at the port of loading; second, a container is inspected once at the port of discharge, and; third, a container is checked once at the port of loading and once at the port of discharge. Customs inspections are random checks, so no shipper can be fully confident that his containers will not be opened to scrutiny.

2. **Scenario with eSeals.** The second scenario describes the perspective of using "green lane" advantage for "secure" containers equipped with eSeals. A container moves from port A to port B without any stops for physical inspection, (see Fig. 2). We assume that it brings one benefit for shippers like the possibility to increase container turnover at least by 1%.We consider two acceptable situations: in the first case, customs checks randomly 0.5% of total number of shipped containers for the first 2 years following implementation of eSeals; in the second case, "green lane" advantage is fully available to use from the first year of eSeals implementation for "secure" containers.

The 1st scenario presumes that no investment costs (IC) are realized in logistic processes. The authors look only at the change in logistic expenses from the point of view of customs inspections and investments in electronic seals. Thus the assumption is that all other logistic costs for container transportation through the ports do not change. Consequently, from this scenario we can first calculate the customs costs in the cases when the inspection takes place at one or at two ports for the same container, and secondly a possible loss in 1% of annual sales of container cargo (estimated value of 1 container is equal 5,000 €).

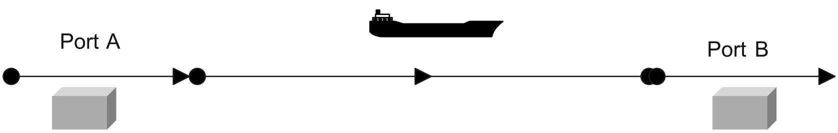


Fig. 2 2nd Scenario for container processing through the ports without customs inspections

In the 2nd scenario to calculate investments in eSeals we distinguish two variants of eSeal use: renting or buying them. The project time period is 5 years. We assume an increase in container turnover by 1% after the implementation of eSeals in container logistic process as one of others possible benefits. Its monetary effect is assumed as 1% annual growing in sales of carrying by container merchandise.

Our investigation distinguishes investment projects when CSD or reusable eSeal are rented by shipper for each container trip each year (I and II projects) and considers the situation where eSeals are purchased once (III-V projects) for whole project period:

1. "Rent CSD"
2. "Rent reusable eSeal"
3. "Buy CSD"
4. "Buy disposable eSeal"
5. "Buy reusable eSeal"

Table 1 presents projected positive and negative cash flows for the investment projects and customs scenarios. From the above-mentioned table we can see the difference between projects with customs control (A, B and AB) and potential investment projects (I–V). The last column summarizes negative CFs for projects with 2% customs inspections and positive incomes in the cases of investments in project I, II, IV and V.

Thus, the analysis of annual incomes denotes possible efficiencies from eSeals versus customs inspections expenses.

Nevertheless, it is required to further analyze every potential investment project to consider the rent prices on eSeals, costs to buy each type of eSeal and additional operational expenses that incur with their use, e.g. transaction costs, sealing and

Table 1 Potential Cash Flows

Scenarios/ Projects	CF [10^6 €					Total CFs [10^6 €]
	1 year	2 year	3 year	4 year	5 year	
A	−0.52	−0.52	−0.52	−0.52	−0.52	−2.6
B	−0.52	−0.52	−0.52	−0.52	−0.52	−2.6
AB	−0.542	−0.542	−0.542	−0.542	−0.542	−2.71
I	0.49465[1]	0.49465[1]	0.5[1]	0.5[1]	0.5[1]	2.5[1]
	0.5[2]	0.5[2]	0.5[2]	0.5[2]	0.5[2]	2.49[2]
II	0.49465[1]	0.49465[1]	0.5[1]	0.5[1]	0.5[1]	2.525[1]
	0.5[2]	0.5[2]	0.5[2]	0.5[2]	0.5[2]	2.49[2]
III	−0.11035[1]	−0.11035[1]	−0.105[1]	−0.105[1]	−0.105[1]	−0.525[1]
	−0.105[2]	−0.105[2]	−0.105[2]	−0.105[2]	−0.105[2]	−0.5357[2]
IV	0.25965[1]	0.25965[1]	0.265[1]	0.265[1]	0.265[1]	1.325[1]
	0.265[2]	0.265[2]	0.265[2]	0.265[2]	0.265[2]	1.3143[2]
V	0.25965[1]	0.25965[1]	0.265[1]	0.265[1]	0.265[1]	1.30[1]
	0.26[2]	0.26[2]	0.26[2]	0.26[2]	0.26[2]	1.3143[2]

[1] Customs check randomly 0.5% of containers 2 years after implementation of eSeals;
[2] "Green lane" advantages are from the first year.

Table 2 Parameters of the investment values for different types of eSeals

Projects	IC [10^6 €]	NPV [10^6 €]	PP [years]	PI [%]	ROI [%]
I. "Rent CSD"	20.00	−18.11[1]	40.40[1]	0.09[1]	0.02[1]
		−18.10[2]	40.00[2]	0.09[2]	0.03[2]
II. "Rent reusable eSeal"	2.50	−0.61[1]	5.00[1]	0.75[1]	0.20[1]
		−0.60[2]	5.00[2]	0.76[2]	0.20[2]
III. "Buy CSD"	4.00	−4.41[1]	−38.1[1]	−0.10[1]	−0.03[1]
		−4.39[2]	−38.1[1]	−0.10[2]	−0.03[2]
IV. "Buy disposable eSeal"	0.25	0.75[1]	0.90[1]	3.98[1]	1.05[1]
		0.75[2]	0.90[2]	4.02[2]	1.06[2]
V. "Buy reusable eSeal"	0.50	0.50[1]	1.90[1]	1.99[1]	0.53[1]
		0.48[2]	1.91[2]	1.97[2]	0.52[2]

[1] Customs check randomly 0.5% of containers 2 years after implementation of eSeals;
[2] "Green lane" advantages are from the first year.

repair costs. We computed the values of NPV, PI, PP and ROI parameters for each project of potential investment by using Excel.

The analysis of the obtained results (Table 2) noted that variants II, IV and V are acceptable for realization over a reasonable 5-year project period. Project IV is the most profitable. It has the greatest NPV value and the highest ROI. The payback period is also the lowest for project IV. At the same time, reusable eSeals have a lower NPV and PP than project IV for the variant when a shipper buys the seals (Project V). It shows that this project is also acceptable for future investment. Furthermore, considering that the NPV of single-use seals is not much larger then that of reusable eSeals, these two variants of investment should be analyzed as alternative projects taking into account investments in reading infrastructure, development of device prices in the future and the benefits for potential eSeal users.

Project III has all negative parameters as seen from the Table 2. So we can conclude that, with current prices of CSD, to buy and exploit them the shipper has to get more monetary benefits. These benefits should be included in the investment analysis to get more realistic results.

4 Conclusions

In the paper we have analyzed two scenarios for container shipping through the ports. The scenario without eSeals presents the current inspection costs for containers. For the scenario with eSeals we have analyzed how effective investments in container "security" by means of eSeals will be. The direct impact of "secure" container trade comes from avoiding customs inspections. Already this advantage can bring monetary benefits for shippers in less then one year. Considering the obtained results the paper shows the positive tendency in influence of investing in security devices on efficiency of container business.

Our further research is required more detailed investing analysis for reusable RFID eSeals. They provide more possibilities to be used in logistic purposes and to get new benefits for improving container shipping processes. We also plan to analyze the advantages of investing in CSD due to more visibility benefits for the container supply chains.

References

1. Crist P, Crass M, Miyake M (2005) Container transport security across modes. (OECD and ECMT Final report), ISBN 92-821-0331-5, 2005
2. Eyerdam R (2005) Bonner announces "green lane" shipping (C-TPAT). Florida Shipper, 2005.02.05
3. Hau LL, Wolfe M (2003) Supply Chain Security without Tiers. Supply Chain Management Review January/February 2003: 12–20
4. ISO/DIS 18185-1 (2005) Freight containers – Electronic seals – Part 1: Communication protocol, 2005.04.28
5. ISO/PAS 17712 (2003) Freight containers – Mechanical seals, 2003.10.01
6. U.S. Government Accountability Office. http://www.gao.gov/ Accessed 2005.11.03
7. Wolfe M (2002) Electronic Cargo Seals: Context, Technologies, and Marketplace. FHWA, North River Consulting Group

Towards Autonomous Logistics: Conceptual, Spatial and Temporal Criteria for Container Cooperation

Arne Schuldt, Sven Werner

Centre for Computing Technologies (TZI)
University of Bremen, Am Fallturm 1, D-28359 Bremen

Abstract Shipping containers handle most of today's intercontinental transport of packaged goods. Managing them in terms of planning and scheduling is a challenging task due to the complexity and dynamics of the involved processes. Hence, recent developments show an increasing trend towards autonomous control with software agents acting on behalf of the logistic objects. Despite of the high degree of autonomy it is still necessary to cooperate in order to achieve certain goals. This paper argues in favour of conceptual, spatial, and temporal properties on which shipping containers can form groups in order to jointly achieve certain goals. A distributed clustering method based on concept, location, and time is introduced. Subsequently, a case study demonstrates its applicability to a problem in the shipping container domain.

1 Introduction

Shipping containers play an important role in intercontinental transport. They have been established since 1956 when Malcom Purcell McLean started applying intermodal transport on a large scale (Levinson, 2006). The first containers in Germany were unloaded in Bremen in 1966 (Schwerdtfeger et al., 1991). Today, shipping containers handle most of the intercontinental transport of packaged goods (Günther and Kim, 2005).

Planning and scheduling the route of a container from its source to its sink is a complex task. This is due to the high number of containers in service as well as the task of finding an optimal utilisation of the available transport and storage facilities. Furthermore, containers are sometimes delayed or even lost. This prevents static planning and leads to highly dynamic processes. Centralised approaches, such as (Dantzig, 1951), exhibit only a limited degree of efficiency in addressing this is-

H.-D. Haasis et al., *Dynamics in Logistics*

DOI: 10.1007/978-3-540-76862-3, © Springer 2008

sue. By contrast, there is an increasing trend towards autonomous logistics. In this paradigm logistic objects are represented by autonomous software agents which act on their behalf in a multiagent system (Weiss, 1999). In terms of shipping containers this massively reduces the complexity since only the parameters of each single container have to be taken into consideration. Furthermore, replanning (e. g., due to delays) can be done locally without rescheduling the whole network.

According to the above discussion it is desirable to achieve a high degree of autonomy in order to cope with the underlying complexity and dynamics. Nevertheless, it still makes sense to cooperate in achieving certain goals (Wooldridge and Jennings, 1999). For instance, transporting containers by train is much cheaper than transporting them by truck. This, however, is only true if several containers choose to share the same train. By contrast, it is not profitable to employ a train in order to transport a single container. As another example, it is preferable to receive containers comprising the same content in the same warehouse. This allows distribution centres to be centrally supplied; the probability for empty space on the employed trucks is reduced. Regarding autonomous logistics containers have to coordinate and group themselves accordingly to their content.

Forming clusters is a possibility for shipping containers to succeed in these goals. Previous clustering methods like k-means (MacQueen, 1967) are not applicable for this task as they conflict the autonomy by their centralised perspective on the data to be clustered. Approaches from the wireless sensor network domain (Akyildiz et al., 2002; Al-Karaki and Kamal, 2004) seem to be more promising as they work in a distributed way. However, they are generally limited to quantitative spatial distances. The problems outlined above can therefore not be covered as they require to take semantic knowledge into consideration. In particular, conceptual, spatial, and temporal aspects can be deduced as clustering criteria from the above examples.

The remainder is structured as follows: Section 2 introduces the clustering criteria identified for cooperating containers and gives a formal definition. Subsequently, a case study with an example application is presented in Sect. 3. Finally, Sect. 4 discusses the work presented.

2 Criteria for Cooperation

The abstract model for cooperation (Wooldridge and Jennings, 1999) divides cooperative problem solving into four steps: recognition of potential for cooperation, team formation, plan formation, and team action. The focus of this paper is a concrete approach for the second step, i.e., forming a distinct team that can reach a certain goal. In order to address this task it is necessary to compare the properties of the respective containers. In particular, the presented approach takes concept, location, and time constraints into account. These properties are considered relevant to containers that may take advantage of forming clusters. Having conducted the team formation step the formed cluster is supposed to be adequate for realising the desired goal.

2.1 Conceptual and Spatial Constraints

The knowledge concerning the content of shipping containers is represented by an ontological approach based on description logics (Baader et al., 2003). Description logic is a decidable fragment of first-order logic. The modelling consists of concepts representing sets of objects and roles as relationships between these objects. Figure 1 shows an example ontology that consists of three major parts: article, property, and location. The first part arranges the transported goods by their type into a taxonomical hierarchy. The taxonomy allows the recognition of more general classes of goods; for instance, it can be concluded that T-shirts are textiles. A second part of the ontology comprises properties that goods may have. By introducing further roles it is possible to model that, say, fruits are perishable or jewellery is valuable. In contrast to the type of goods, which generally never changes, the properties may change (although rather seldom). As an example, goods might be damaged during transport. Furthermore, the location of goods can be integrated as an ontological concept, which is the third part. Thus, further restrictions can be realised, e.g., a class comprising all containers with valuable jewellery that are currently at a container terminal. Changes in this part of the description are most likely as goods are transported between different locations.

As discussed above, the ontological descriptions of logistic objects change over time. The time-span in which an ontological concept holds can be characterised with the help of the following definition of temporal intervals:

Definition 1 (Temporal Interval). A temporal interval τ is defined by a start point t_s and an end point t_e, where $t_s < t_e$:

$$\tau = (t_s, t_e);\; t_s, t_e \in \mathbb{N}, t_s < t_e\,.$$

Based on this definition, the ontological concept of a container agent during a given time-span can be defined as a subclass of the universal concept $\top$:

Definition 2 (Agent Concept). Let α be an agent. The ontological concept of α during τ is represented by

$$\mathsf{c}(\alpha, \tau) \sqsubseteq \top\,.$$

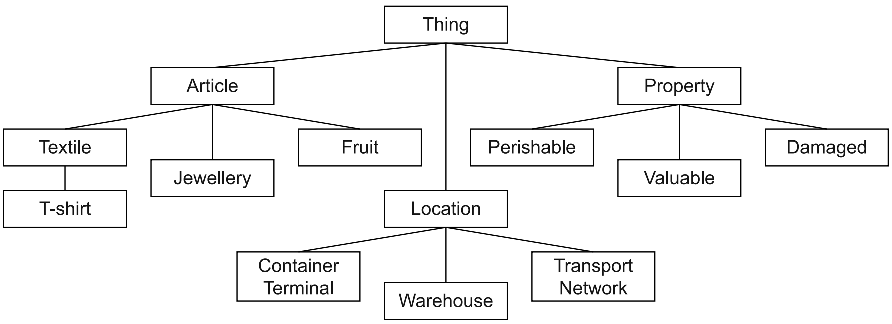


Fig. 1 An example ontology for goods with three major parts: article, property, and location. The *lines* connecting concepts indicate taxonomical relations

2.2 Agent Clusters

In order to achieve a common goal, container agents are capable of clustering. The lifetime of the whole cluster can thereby exceed the membership of a single agent. Furthermore, each agent can leave a cluster and rejoin later. Therefore, the membership of an agent cannot be characterised by a single temporal interval. By contrast, a set of intervals has to be applied (which might also be empty):

Definition 3 (Cluster Membership). Let α be an agent, let γ be a cluster. The membership of α in γ is defined by a set of temporal intervals:

$$\mathsf{m}(\alpha,\gamma) = \{\tau_1,\ldots,\tau_n\}, n \in \mathbb{N}\,.$$

Before joining a cluster, a container has to determine its similarity to the cluster. This procedure is a special case of the so-called matchmaking problem (Shvaiko and Euzenat, 2005). It denotes the decision whether advertisements offered by business entities or agents match requests issued by other agents. An approach proposed by (Li and Horrocks, 2004) addresses this task by considering the formal semantics of ontology-based knowledge representation techniques. They propose five different degrees of match between two concepts c_1 and c_2: exact $c_1 \equiv c_2$, plug-in $c_1 \sqsupseteq c_2$, subsume $c_1 \sqsubseteq c_2$, intersection $\neg(c_1 \sqcap c_2 \sqsubseteq \bot)$, and disjoint $c_1 \sqcap c_2 \sqsubseteq \bot$.

Like agents, clusters are also characterised by concepts that hold within given temporal intervals. During its membership, the concept of each agent is supposed to be subsumed by the cluster concept:

Definition 4 (Cluster Concept). Let γ be a cluster. The ontological concept of γ during τ is represented by

$$\forall_{\alpha\in\gamma}\forall_{\tau_i\in\mathsf{int}(\tau,\mathsf{m}(\alpha,\gamma))}\,\mathsf{c}(\alpha,\tau_i) \sqsubseteq \mathsf{c}(\gamma,\tau) \sqsubseteq \top\,.$$

Apart from subsumption, Definition 4 also uses the auxiliary function int, that computes for a given set of temporal intervals their intersection with a reference interval. This ensures that the concept of the cluster members is only compared for those times that are covered by τ:

$$\mathsf{int}(\tau,\{\tau_1,\ldots,\tau_n\}) = \{(\tau\cap\tau_i) \mid i \in \{1,\ldots,n\} \wedge (\tau\cap\tau_i) \neq \emptyset\}\,. \tag{1}$$

Besides matching containers and clusters, the approach can also be applied in order to find appropriate warehouses. As an example, it can be deduced from the ontology that a cluster of containers carrying damaged T-shirts can be received in a warehouse that is capable of receiving damaged textiles.

2.3 Temporal Constraints

As discussed in the previous section, the concept (including location) of container agents plays an important role during cluster formation. So far, the application of time has been limited to checking whether agents meet a cluster concept during a given temporal interval. Besides, some applications also demand certain temporal relationships between the agents within a cluster. Figure 2 illustrates three shipping container examples. First, if containers plan to be transported by the same truck, one after another, their temporal intervals have to be disconnected (i.e., without overlap). Second, if one container conducts the planning for the whole group, its membership has to subsume the memberships of all other agents. Finally, if containers plan to share a train, their expected dates of arrival at the station should correspond, i. e., the respective intervals must share a common end.

A classical approach for temporal reasoning is formed by Allen's qualitative relational system (Allen, 1983) (qualitative in the sense that it abstracts quantitative data to a semantically defined representation) which is depicted in Fig. 3:

Definition 5 (Temporal Relation). Let τ_i, τ_j be temporal intervals. The position $\tau_{i\,\tau_j}$ of τ_j w. r. t. τ_i is then characterised as

$$\tau_{i\,\tau_j} \in \{<, >, \mathrm{m}, \mathrm{mi}, \mathrm{o}, \mathrm{oi}, \mathrm{s}, \mathrm{si}, \mathrm{d}, \mathrm{di}, \mathrm{f}, \mathrm{fi}, =\}.$$

However, these 13 relations characterise pairs of temporal intervals. Hence, this paper proposes a generalised approach of predicates about arbitrary sets of temporal intervals in order to meet the above requirements.

In order to reason about temporal intervals within a cluster, it is first necessary to obtain all these intervals. This is achieved by the following auxiliary function:

$$\mathsf{allIntervals}(\gamma) = \bigcup_{\alpha \in \gamma} \mathsf{m}(\alpha, \gamma). \tag{2}$$

Since the above set is not necessarily sorted (accordingly to start and end points), it is transferred into an ordered list by a further auxiliary function:

$$\mathsf{orderedIntervals}(\gamma) = \langle \tau_1, \ldots, \tau_n \rangle, \tau_i \in \mathsf{allIntervals}(\gamma),$$
$$\forall_{\tau_i = (t_{s_i}, t_{e_i})} \begin{cases} t_{e_i} \leq t_{e_{i+1}} & \text{iff } t_{s_i} = t_{s_{i+1}} \\ t_{s_i} < t_{s_{i+1}} & \text{else.} \end{cases} \tag{3}$$

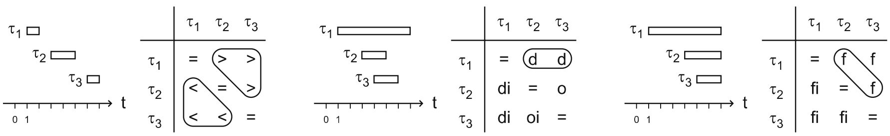


Fig. 2 Three examples for relations between sets of temporal intervals within an agent cluster and the respective temporal matrices

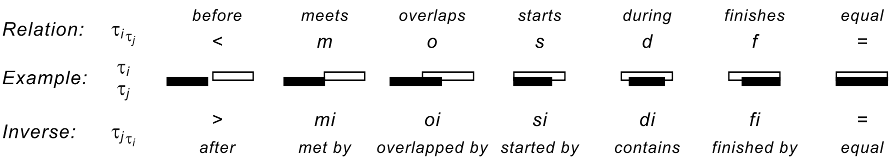


Fig. 3 The 13 relations that have been proposed by Allen (Allen, 1983) in order to characterise the qualitative relationship between two temporal intervals

Based on these auxiliary functions, it is then possible to define the matrix M_γ that comprises the relations for all pairs of temporal intervals of a cluster:

Definition 6 (Temporal Matrix). Let γ be an agent cluster. The matrix

$$M_\gamma = \langle \tau_1, \ldots, \tau_n \rangle \times \langle \tau_1, \ldots, \tau_n \rangle, \tau_i \in \mathsf{orderedIntervals}(\gamma)$$

characterises the relations $\tau_{i\,\tau_j}$ between all temporal intervals in γ with $\tau_{i\,\tau_j}$ being the entry in row τ_i and column τ_j.

Figure 2 gives these matrices for the depicted configurations. Restrictions like those discussed in the introductory paragraph of this section can be defined as predicates on the temporal matrices. For instance, in the first example the only relations that are allowed to occur in the whole matrix are $\{<,>\}$. Thereby, the matrix' main diagonal is excluded as it relates each interval to itself, which always results in $=$. For the second example, it has to be ensured that there exists one row comprising at most the relations $\{\mathrm{s}, \mathrm{d}, \mathrm{f}, =\}$. In order to determine the third example it is sufficient to examine pairs of subsequent intervals (which is a diagonal of the matrix). The occurring relations are restricted to $\{\mathrm{f}, =\}$. The general definition for these three types of predicates is as follows:

Definition 7 (Temporal Matrix Restriction). Let γ be an agent cluster. Restrictions on its temporal matrix M_γ can be described by a set ρ of temporal relations and one of the following predicates:

$$\begin{aligned} \mathsf{wholeMatrix}(M_\gamma, \rho) &= \forall_i \, \forall_j \, (\tau_{i\,\tau_j} \in M_\gamma \wedge i \neq j) \rightarrow \tau_{i\,\tau_j} \in \rho \\ \mathsf{matrixRow}(M_\gamma, \rho) &= \exists_i \, \forall_j \, (\tau_{i\,\tau_j} \in M_\gamma \wedge i \neq j) \rightarrow \tau_{i\,\tau_j} \in \rho \\ \mathsf{matrixDiagonal}(M_\gamma, \rho) &= \forall_i \;\; \tau_{i\,\tau_{i+1}} \in M_\gamma \rightarrow \tau_{i\,\tau_{i+1}} \in \rho \,. \end{aligned}$$

3 Case Study

In order to test the clustering approach it is implemented within the Java-based agent-framework JADE (Bellifemine et al., 2007). The clustering is thereby conducted as follows: as soon as a container joins the network it queries a catalogue service for existing clusters. Subsequently, it communicates its properties to all cluster-heads. Clusters matching the properties of the container send positive answers. If the requesting agent receives a positive answer it joins the respective cluster. Otherwise, the agent chooses to register as a cluster-head itself.

As a case study the approach is applied to data of about 2,400 containers which were in operation during three months in 2006. The goal is to demonstrate that the approach overcomes the limitations of other approaches regarding distribution and semantic knowledge. Furthermore, its applicability to real-world problems has to be examined. As an example application similar containers located anywhere are expected to form clusters in order to be jointly received in a common warehouse

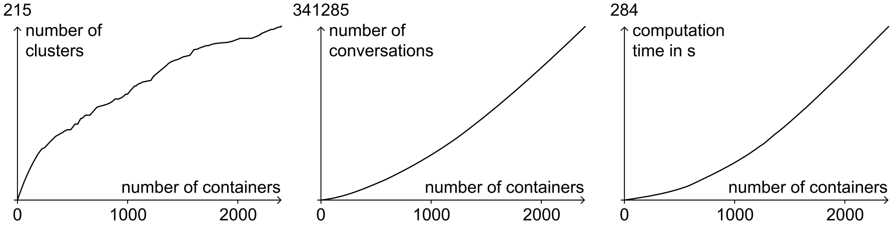


Fig. 4 The number of clusters (*left*) and agent conversations (*centre*) and the computation time (*right*) in relation to the number of containers

(see Sect. 1). The lifetimes of the agents are thereby expected to be overlap pairwise without a gap in between.

The containers join the network distributed over this time-span. A manual data inspection reveals that there exist 215 clusters; so this is the expected outcome of the experiment. Since each of the n containers contacts all m clusters at most once it is expected that the asymptotic development of agent conversations and computation time is $O(nm) = O(n^2)$.

The average results of 50 test runs are given in Fig. 4. The number of clusters in relation to the total number of containers is depicted on the left hand side. The exact shape of the curve strongly depends on the underlying data. As a result it can, however, be discovered that the final number of clusters is 215 for all test runs, i. e., the algorithm is capable of solving the addressed problem. The total number of conversations (Fig. 4 centre) is 341,285, which is below the expected complexity. That is because not all clusters exist right from the start. Nevertheless, as expected the plotted part of the curve indicates an increase that is faster than linear. This observation also corresponds with the computation time that is depicted on the right hand side of Fig. 4. On a computer with Windows XP and an Intel Centrino Duo processor with 2.16 GHz the total time for clustering is only 284 seconds.

4 Discussion

In autonomous logistics, it is a challenging task for the participating logistic objects to cooperate in achieving their goals. The framework introduced allows similar agents representing autonomous shipping containers to form clusters by concept, location, and time. The conceptual, spatial, and temporal properties are thereby represented as semantic knowledge and qualitative relations. The case study conducted demonstrates the applicability to problems in the shipping container domain. The method presented does not make special assumptions on the represented objects. Hence, it is not limited to containers and can also be applied in order to realise clustering of other logistic objects (e. g., autonomous packages). As concept, location, and time are general properties, it is also possible to apply the approach in multiagent systems that model domains not related to logistics.

References

Akyildiz IF, Su W, Sankarasubramaniam Y, Cayirci E (2002) Wireless Sensor Networks: A Survey. Computer Networks 38:393–422

Al-Karaki JN, Kamal AE (2004) Routing Techniques in Wireless Sensor Networks: A Survey. IEEE Wireless Communications 11(6):6–28

Allen JF (1983) Maintaining Knowledge about Temporal Intervals. Communications of the ACM 26(11):832–843

Baader F, Calvanese D, McGuinness D, Nardi D, Patel-Schneider PF (eds) (2003) The Description Logic Handbook. Cambridge University Press, Cambridge, UK

Bellifemine F, Caire G, Greenwood D (2007) Developing Multi-Agent Systems with JADE. John Wiley & Sons, Chichester, UK

Dantzig GB (1951) Application of the Simplex Method to a Transportation Problem. In: Koopmans TC (ed) Activity Analysis of Production and Allocation, John Wiley & Sons, New York, NY, USA, pp 359–373

Günther HO, Kim KH (eds) (2005) Container Terminals and Automated Transport Systems. Springer-Verlag, Berlin, Germany

Levinson M (2006) The Box: How the Shipping Container Made the World Smaller and the World Economy Bigger. Princeton University Press, Princeton, NJ, USA

Li L, Horrocks I (2004) A Software Framework for Matchmaking Based on Semantic Web Technology. International Jounal of Electronic Commerce 8(4):39–60

MacQueen J (1967) Some Methods for Classification and Analysis of Multivariate Observations. In: Fifth Berkeley Symposium on Mathematical Statistics and Probability, pp 281–296

Schwerdtfeger H, von Zabern T, Pölking-Eiken HJ (1991) Bremen/Bremerhaven Container Story. BLG, Bremen, Germany

Shvaiko P, Euzenat J (2005) A Survey of Schema-Based Matching Approaches. Journal on Data Semantics 4:146–171

Weiss G (ed) (1999) Multiagent Systems. A Modern Approach to Distributed Artificial Intelligence. MIT Press, Cambridge, MA, USA

Wooldridge M, Jennings NR (1999) The Cooperative Problem Solving Process. Journal of Logic & Computation 9(4):563–592

Distributed Process Control by Smart Containers

Türk Kiziltoprak[1], René Schumann[2], Axel Hahn[3], Jan Behrens[1]

[1] OFFIS e.V., Escherweg 2, 26121 Oldenburg, Germany
[2] Universität Frankfurt, FB 12, 60325 Frankfurt, Germany
[3] Universität Oldenburg, Dept. für Informatik, 26111 Oldenburg, Germany

Abstract Most of the actual sea freight is transported in ISO standard containers today. The number of containers shipped per year has been increasing continuously over the last years and is expected to further grow, see e.g. [1]. Dealing with these increasing numbers and the data generated along their way is a great challenge for today's logistic and IT systems. We believe that these problems can best be faced through encapsulation and decentralization. Our goal is to increase the automation within the port by using emerging technologies to enable a distributed container handling approach with proactive containers. These smart containers monitor their environment, provide status information and interact with others. All these capabilities should be used to control the process flow of the container's handling.
Our research vision is focussed on the port. We are aware that additional requirements arise, for example from supply chain event monitoring, but nevertheless believe this to be an acceptable constraint.
We assume a smart container that is capable of wirelessly registering itself, when entering the port. The further processing of the container can then be controlled by the container itself through its state. The container would know its contents (pack list) along with its routing. Additionally the container would be equipped with sensors to monitor its interior and exterior environment, and change its state proactively. Technologically we want to use the ZigBee and LEACH routing protocol to build up the network and develop a software platform as a service-oriented architecture based on the UPC UA Architecture and using OSGi as the framework.

1 Introduction

Currently various information systems exist that are involved in a container's handling, for example port information management systems (e.g. BHT or Dakosy), customs information systems (e.g. ATLAS), dangerous goods information systems or carrier's systems. All these have to have different information and views on the state of each container.

H.-D. Haasis et al., *Dynamics in Logistics*
DOI: 10.1007/978-3-540-76862-3,

Our vision implies both challenges for the sensor network technology and the information processing entities. There are also many challenges in integration through the paradigm shift towards decentralisation. Chiefly those, arising from the need to assure the interoperability of software systems and processes. While the container communication today is generally unidirectional and focused on centralised systems, the proposed communication will be bidirectional and will have to make use of encapsulation through decentralisation. A direct integration is not feasible because of the many different stakeholders and their varying information needs. It would also prevent bidirectional communication as the container would not know who and how to talk to in the current ports software infrastructure. For example a reefer container with temperature sensors could ask for assistance in case of failure, or ask for a change in routing if its contents would otherwise perish. The information would also enable a better monitoring of dangerous cargo and potential safety threats by any container. Because the containers amongst themselves would establish an ad hoc network, they could also share sensor data or communication infrastructure and react on changes of their neighbours. Through these autonomous state changes, and the possibility to communicate them, the container can directly influence its handling. Furthermore automatic creation of manifests would become possible, leading to a considerable advantage in time and expenses for both port operators and carriers. We plan to develop a platform as an abstraction layer that allows the bidirectional interaction between the container network and the various software systems in the port through standardized means. This platform will provide a SOA, using web services, that facilitates the communication with independent containers and analysis on groups of containers – and enable containers to use web services provided by the software systems. The container would only need to know the role of the stakeholder he would like to communicate with (e.g. carrier, terminal operator), the platform would then relay to the correct system in the current ports software infrastructure.

2 Problem

We assume two major problems in the area of container handling within the port area. First of all there are various information systems integrated in the process. These information systems are very heterogeneous. There are systems for the port management, systems for the ship management and stakeholder information systems. All these systems need to have different information and states of the containers for their special usage, and nowadays the interoperability of these systems is small.

Beside the heterogeneity of these systems the collection of data is only little automated. The little automation of data collecting leads to cost intensive data collecting processes and only low data quality. Nevertheless new regulatory frameworks, coming from the United States or the European Union in the nearer future, make it necessary to collect even more data.

3 Solution ideas

To face these challenges we identify three topics. First of all the automation of data collection must be increased. Nevertheless an overflow of information generated from the data must be avoided. Therefore well known concepts, like Planning, Data Mining or Decision Support, might be useful. Furthermore the communication with the goods which should be established has to be a semantic enriched standardised communication, by using concepts like Web Services and Ontology.

4 Technical aspects

A networked microsystem node consists of a small computer with wireless communication and integrated peripheral devices such as sensors. We would also see each node equipped with an (ISO 10374) RFID transponder and reader to scan container contents. We identify three major challenges regarding these sensor nodes; the establishment of ad hoc networks of containers, the location of containers within such a network, and especially the composition and utilisation of emergent properties of the network. We want to establish an ad hoc network based on the LEACH (Low Energy Adaptive Clustering Hierarchy) [2] routing protocol, using 802.15.4 (ZigBee base) on the physical layer allowing for a medium range of 500m. Location services would be facilitated through signal propagation delays – thus enabling a continuous location in the port with a maximum offset of 2m at 100m reach. We furthermore plan to build prototypic sensor nodes that allow the usage of embedded OSGi to enable dynamic loading of software components. This will enable each node to react not only on its own perception but also on the perception of those nodes surrounding it, by loading different behaviours based on the state of the network.

For implementing a smart container approach various technologies are available to be aligned. In addition to the already mentioned communication standards the idea of smart containers needs platform operation and sensor integration concepts. Some of them are presented below.

OSGi defines a framework and predefined services structures in 5 layers. The framework forms the core of the OSGi Service Platform Specifications [3]. It provides a general-purpose, secure, and managed Java framework that supports the deployment of extensible and downloadable applications known as bundles.

The Security Layer is based on Java 2 security but adds a number of constraints and fills in some of the blanks that standard Java leaves open. Actual Services are the application specific components.

OPC Unified Architecture is the most recent OPC specification from the OPC Foundation and differs significantly from its predecessors [4]. The OPC UA architecture is a Service-Oriented Architecture (SOA) and is based on different logical levels.

All of the Base Services defined by OPC are abstract method descriptions which are protocol independent and found the basis for the whole OPC UA functionality. The transport layer puts these methods into a protocol, which means it serializes/de-

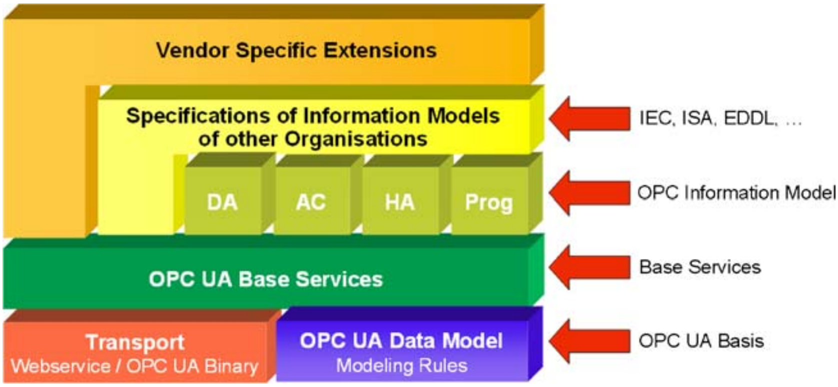


Fig. 1 The OPC UA Architecture [5]

serializes the data and transmits it over the network. Currently there are two protocols specified for this purpose. One is a binary, towards high performance optimized TCP protocol, the second is Webservice-based. The OPC information model isn't just a hierarchy based on folders, items and properties anymore, but a so-called Full Mesh Network based on nodes instead. These nodes can additionally transmit all varieties of Meta information. The easiest comparison of a node would be an object, as known in object oriented programming (OOP). It can own attributes for read access (DA, HAD), methods which can be called (Commands) and triggered events which can be transmitted (AE, DA DataChange). The nodes are used for process data as well as all other types of meta data. The used OPC namespace also contains the type model used for description of all possible data types. Based on the above, other organizations e.g. like EDDL are specifying their own information source. Client software has the ability to verify which of the so-called Profiles a server supports. This is necessary to obtain information if a server only supports the DA functionality or additionally the AE, HAD, etc. functions. Additionally, information can be obtained whether a server supports e.g. the EDDL profile and therefore the client knows that there are also EDDL-specific device descriptions available.

For the usage of these standards an adoption into the logistics domain is required. An open issue is the definition of the semantics of container operation. Here an ontological approach is proposed by the authors.

5 Communicational aspects

To establish a standardized way for the communication between the different systems we developed an Ontology. Ontologies meet several demands in computer science. Based on description logics, they offer a rich set of constructs for defining a shared understanding of the entities and their relationships of a certain domain, modelled as classes and their relationships [6]. The modelled world can be checked for consistency so that a sound definition of the entities of a real-world subset can be established [7]. Using this definition, it is possible to establish communication between the actors of the domain using instances of the classes. Each class represents certain entities of the real world and can be persisted, e.g. through an XML [8] representation. Those representations are then used by the actors to interchange information. This approach offers two main advantages: On the one hand, every actor has complete freedom in the design of his hardware and software. He only needs to offer and accept his communication needs via the agreed ontology.

On the other hand, every participant can be ensured that no communication problems on the semantics of the messages occur. The latter is ensured through a mutual understanding of the modelled semantics by all stakeholders during the development process.

There are several process models for the development of an ontology (see [9] for a comparison). Without going into further detail there must be a shared understanding by the stakeholders of the entities relevant in the scenario and a correct modeling of those entities through the modeling techniques provided by an adequate ontology representation language, e.g. OWL [10]. Using our previous example one can identify four main stakeholders: Containers, ships, shippers, and terminal operators. We will show an extract of an ontology that contains relevant classes for all stakeholders. The use case that is covered by the scenario is a reefer container which's content heated up to a critical temperature and calls for assistance by a local service team. Additionally the shipper gets informed that the cargo might be spoiled before it arrives at the intended location so that a rerouting might be considered in order to get the cargo sold earlier on its originally intended way. As already mentioned we will not present a complete ontology for the given scenario but rather show a use case driven beginning of developing an ontology.

Figure 2 shows an extract of the proposed ontology of the autonomous container handling scenario suitable for the use case described above.

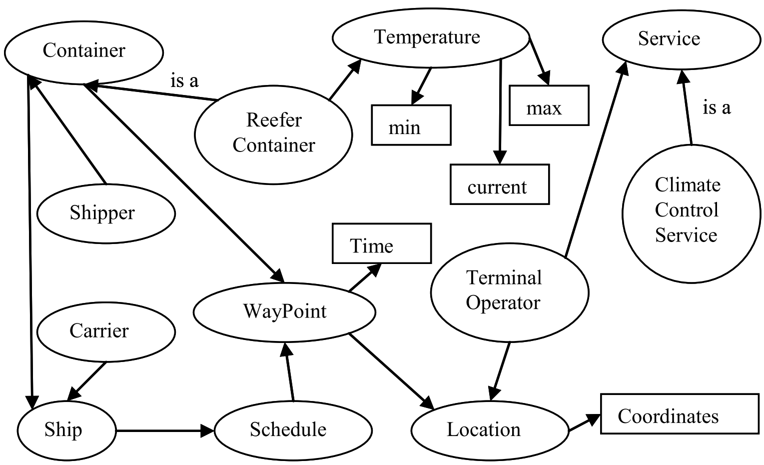


Fig. 2 Ontology for proposed scenario (simple types in rectangles, only edges describing inheritance are depicted)

A '*reefer container*' is a specialization of an ordinary '*container*' that has an uncritical temperature span. *'Containers'* belong to *'Shippers'*. Carriers have several '*Ships*' underway that consist of a set of '*Containers*'. Each *'Ship'* has a '*Schedule*' that consists of a set of '*WayPoints*' sortable through their '*Times*' that are associated with them. The '*WayPoints*', additionally to their '*Times*', also have a '*Location*' that is defined through (e.g. GPS-) '*Coordinates*'. Each '*Container*' can reach those '*WayPoints*' where also the '*TerminalOperators*' are located at. Those '*TerminalOperators*' offer '*Services*' such as '*ClimateControlServices*' that can be requested by the '*Containers*' through the highlighted connections in the ontology.

The presented ontology shown in Fig. 2 is not described formally correct but shows the interrelationships of the entities relevant in the proposed scenario. The container having a malfunctioning climate control would ask the terminal operator for assistance by requesting a list of all instances of the '*ClimateControlService*' class, choosing one of them (probably after inviting several offers) and finally asking for assistance. It then informs its shipper of the possible delay who can, if he thinks that it is necessary, reassign the *'Container'* to another *'Ship'* with another *'Schedule'* so that the cargo might get sold elsewhere.

6 Modern Information processing

As the process of information gathering can be completely automated, it can be expected that the amount of collected data grows significantly. But on the other hand the number of participating information systems and their abilities will not grow that fast. The new challenge rises from the availability of information. It was to be clarified which information is needed in which system and in what granularity. The amount of data will not only grow by the emerging technologies for data collection but also as a consequence of the growing number of container transports in the next years [1]. The number of containers an operator has to supervise will grow while the information available will grow for each container. In consequence not only the data collection has to be automated, it is necessary that the information processing has to be automated as well. This automated information processing is subject of research in the domain of autonomous computing. The current port information systems were developed for centralized computing running on mainframe computers. These systems are not capable of processing all generated information, for two reasons. First the collected data have to be transported to the computing entity processing this data, so communication will become a bottleneck. Second is the relation of the increased computing power of the mainframe systems actual software works on and the number of data and in consequence the computing power needed to process this data, so performance will become a bottleneck. To handle this challenge the information processing and hence decision taking has to be decentralized. Information processing has to be done locally to avoid the aforementioned computing and communication bottleneck. Normal container processing should be handled by the autonomous systems without operator interaction. In case of unforeseen events that make new decision taking necessary the autonomous systems should generate a suggestion.

This support by autonomous working entities could make it possible that operators can handle their work even if the number of containers and collected data will grow.

To implement this vision the technical aspects and the semantic enriched communication are supporting steps that have to be done, if intelligent information processing shall be applied successfully. Intelligent information processing can comprise techniques from the domains of automated inference, planning (predictive and even more reactive), data mining and machine learning. The more scientific term for the discussed autonomous systems is actually agent.

To detail our vision we will give an example for the routing of containers. In this example we assume that a malfunctioning reefer container at a port is waiting for its transhipment that can not keep its goods at the needed temperature. Using the port communication infrastructure it can

1. Ask the local administrator to send a repairman.
2. Inform interested entities like the terminal operator, the carrier and the owner of the goods within. As it may take time until the cooling is fixed the transport of the container might get delayed or miss the next vessel. This can lead to a situation that the goods might be unusable when the goods arrive at the port of destination. The container can detect that situation, as it knows about the state of its contained goods and the approximate time needed to its port of destination. If it detects such a situation it can propose different alternative solutions.
3. It can propose to change the destination, to sell the contained goods earlier. Then it might become necessary that other containers have to be replanned to fulfill the original order.
4. It can propose the change of alternate routing or ask for the transport of goods by another medium, e.g. by plane.

The carrier operator can be informed that a transport exists that needs more attention and can evaluate these suggestions. Therefore using additional information or experience he can make a decision on actual and precise data using local aggregated information and projections. With this decision support he can take the decision and inform the container. The detailing and realization of this decision can then be done by the container, again.

In essence the current centralized information systems evolve into decision support systems, where the actual state is presented to operators that can formulate goal states for certain transports. The realization of the goal state, the planning and control to reach these states is in the scope of the smart containers.

7 Related work

The research work being done by the SFB 637 in Bremen is similar to our work [11]. But while we are focusing on the port as our research area the SFB sets its focus on the whole transport chain. Beside this we are first of all interested in propagating the state of the containers and manage their way through the port area, as a bottleneck within the transportation chain, more efficiently.

8 Future work

At the moment the technological parts, like the sensor nodes, are being developed prototypically. The conceptual parts are being defined and all components were built up as a model to test the usability of the components.

As a next step a testbed and a reference implementation of the aforementioned architecture must be created. This testbed will be more realistic than the developed test model. In the testbed all components (the hardware as well as the software) will work together as realistic as possible so a prove of concept can be done. Further work on the development of a prototype capable of working in a real world environment has to be done in cooperation with partners to evaluate the behaviour in a real world surrounding.

References

1. ISL; *ISL Monthly Container Port Monitor*, Retrieved at 27.03.2007, http://www.isl.org/products_services/publications/pdf/MCPM_3rd_quarter_2006_sum.pdf
2. Heinzelman W, Chandrakasan A, Balakrishnan H: *Energy-efficient Comunication Protocols for Wireless Microsensor Networks,* Proc. Hawaiian Int'l Conf. on Systems Science, January 2000
3. About the OSGi Service Platform (Technical Whitepaper), Revision 4.1 7. June 2007
4. http://www.opcfoundation.org/ Retrieved at 27.04.2007
5. http://www.ascolab.com/index.php?file=ua&lang=de Retrieved at 27.04.2007
6. Gruber TR: *A translation approach to portable ontologies.* In: Knowledge Acquisition, 5(2):199-220, 1993.
7. Antoniou G, van Harmelen F: *A Semantic Web Primer.* The MIT Press, Cambridge, Massachusetts, London, England, 2004
8. *Extensible Markup Language (XML) 1.0* (Fourth Edition) W3C Recommendation 16, 2006 http://www.w3.org/TR/2006/REC-xml-20060816/ (2007-04-25)
9. López, M: *Overview of the methodologies for building ontologies.* In Benjamins VR, Chandrasekaran B., Gmez-Prez A, Guarino N, and Uschold M (Eds): Proceedings of the IJCAI-99 Workshop on Ontologies and Problem-Solving Methods (KRR5), Stockholm, Sweden, 1999.
10. *OWL Web Ontology Language Semantics and Abstract Syntax* W3C Recommendation, 2004. http://www.w3.org/TR/owl-semantics/ (2007-04-25)
11. SFB 637 Collaborative Research Centre Homepage. http://www.sfb637.uni-bremen.de (2007-04-25)

Autonomous Control in Logistics

Autonomous Units for Communication-based Dynamic Scheduling

Karsten Hölscher, Peter Knirsch, Melanie Luderer

Department of Computer Science, University of Bremen, Bremen, Germany

Abstract Logistics has to deal with dynamics and uncertainties. In order to cope with such problems we introduce a communication-based approach built on distributed autonomous systems. In this work graph transformation with autonomous units is used as a rule-based instantiation of multi-agent systems to model logistic systems. The approach will be presented using a scenario taken from transport logistics. Here loads which have to be transported are queued in order of their arrival but scheduled for further transportation according to their own constraints. In the paper we propose a negotiation between loads and the respective truck based on payment of transportation rates.

1 Introduction

Logistics is a field characterized by mainly two types of dynamics. First, the organized supply with goods and information demands for controlled material and information flows. In addition, these flows are from the operational point of view dynamic changes. To manage this kind of dynamics is the paramount challenge of logistics. Second, the whole setting is equipped with uncertainties and risks. Not all circumstances can be planed in advance; not all information needed is accessible, correct, or consistent at the point of scheduling. Occurring changes in the prerequisites can force to restart the whole scheduling process repeatedly. Especially this type of dynamics is hard to cope with using the classical approaches. However, these difficulties exist in real world scenarios and being able to react to them can result in a great competitive advantage.

The Bremen Collaborative Research Centre CRC 637 "Autonomous Cooperating Logistic Processes – A Paradigm Shift and its Limitations" tries to overcome these kinds of problems by avoiding a strictly centralistic view and by passing control capabilities to the logistic objects in order to make them smarter.

In this work, a communication-based approach will be introduced that incorporates autonomous units as a formal framework for modeling autonomous behaviors.

H.-D. Haasis et al., *Dynamics in Logistics*
DOI: 10.1007/978-3-540-76862-3,

A scenario taken from the transport logistics will demonstrate the usefulness of the chosen approach: Unit loads arriving at consolidation points for further transportation are queued (see, e.g., (Cooper 1981) for an introduction to queueing theory) according to their arrival time in a *first come, first served* manner but are scheduled according to their own constraints. Often the order of arrivals is rather arbitrary and does not reflect the real priorities of the transports to be accomplished. In addition, transport orders can arrive at any time, so early scheduling cannot react accordingly. Additionally, due to limited resources, the simultaneous transport of unit loads can be impossible and can force loads to pause. Imagine a queue of 100 unit loads where only one load at position 98 is having a hard time constraint to be met. Wouldn't it be fair to give this unit load the chance to be served first? In this case the load would have to pay a higher transportation rate while the remaining loads' disadvantages can afterwards be compensated monetarily.

In this work, a negotiation- and market-based approach is chosen to reorganize the processing of the queue in a decentralized way. For the sake of simplicity, this approach does not take into account the transhipment times.

2 Autonomous Units

For the structuring of the logistic system, we consider autonomous units as a basic entity having all means for modeling autonomous behavior of distributed entities. Autonomous units are a generalization of the concept of transformation units as studied in (Kreowski and Kuske 1999) and (Kreowski et al. 1997). They are a rule-based instantiation of the idea of multi-agent systems (see, e.g. (Weiss 1999)) as first introduced in (Knirsch and Kreowski 2000). A first discussion on the relation of these concepts can be found in (Timm et al. 2007).

An *autonomous unit* consists of a *goal* g (formulated in a proper logic or language), a set of identifiers U naming used autonomous units (that are imported in order to use their capabilities), a set P of rules (also called productions) specifying the operational capabilities, and some control condition c, which restricts the possible orders of actions.

The goal g describes the unit's intent. An autonomous unit acts in a specific environment, which it may change by applying a suitable rule from its productions P. It may also use help from other units in U. A unit is considered autonomous in the sense that the next action is rather based on a non-deterministic selection of a rule or of an imported unit than on control from outside the unit. A simple unit would always randomly decide on the next action. The control condition c provides the means for realizing a more sophisticated form of autonomy. A control condition may be very restrictive and thus eliminating the non-determinism completely or it may leave some room for non-deterministic choice.

An autonomous unit is typically chosen to represent an autonomous entity of the overall system to model. For this reason, a community of autonomous units is defined as a system comprising an overall goal *Goal*, an initial environment *Init* (both formulated in a proper logic or language), and the set *Aut* of the autonomous units that belong to the community.

Since the underlying rule-based concept of autonomous units is graph transformation (see e.g. (Rozenberg 1997)), the productions are graph transformation rules and the environment is specified as a graph. The changes of the environment happen in a well-defined way as application of graph transformation rules on the environment graph, yielding a rigorous formal operational semantics not only for a single autonomous unit but also for the community as a whole. The sequential semantics as discussed in (Hölscher et al. 2006b) is used if exactly one unit is performing an action at any given point of time. The parallel semantics, where a number of actions take place in parallel at the same time is investigated in (Kreowski and Kuske 2006). The concurrent semantics (see (Hölscher et al. 2007) for a short introduction) is used if there are no chronological relations between the acting units except for causal dependencies. In the context of this work, we consider the sequential semantics, i.e. the respective units act one at a time in sequential order.

3 Communication-based Dynamic Scheduling

3.1 Transport Networks

The sample scenario in this work is based on a simplified transport network. We assume a number of consolidation points in German cities and unit loads (ULDs) to be transported by trucks along road connections from one consolidation point to another one. The main relations of each truck are fixed, i.e. the routing has been arranged in advance as a regular service with timetables for each truck. Based on the knowledge of these timetables, the routes of the ULDs have also been planned in advance. What remains to be scheduled in our scenario is which of the waiting ULDs are actually transported by the respective truck. We propose a negotiation between the truck and the ULDs based on the payment of transportation rates.

In this scenario, trucks and ULDs are realized as autonomous units, called truck unit and load unit, respectively. Thus, the underlying environment comprises the consolidation points and their connections. This is modeled as a graph in a straightforward way. Now each truck unit and each load unit becomes part of the environment as a special truck resp. load node. Both kinds of autonomous units utilize tour nodes for a representation of their planned tours (as introduced in (Hölscher et al. 2006a). A truck tour is divided into tour sections, which are represented by tour nodes that are connected to the source and target consolidation points of the respective section. These tour nodes are labeled with the estimated time of departure, estimated time of arrival and the capacity for that tour section. A planned load unit tour is represented by tour nodes in a similar way, except for the labels. Load unit tour nodes are labeled with the weight of the ULD and a desired time of arrival (which is ‘n.def’ in the case that it does not matter when the ULD arrives). Figure 1

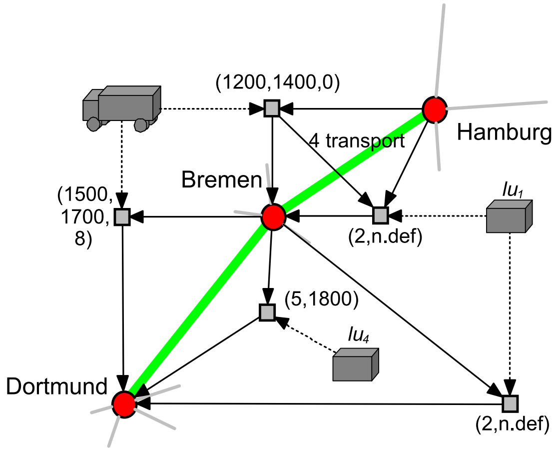


Fig. 1 Excerpt of a sample transport network

shows an excerpt of a sample environment, with a truck that has started in Hamburg at noon and drives via Bremen to Dortmund, estimated to arrive at 5pm. There is no capacity left on the first tour section, since the truck transports the load unit lu_1 and others (which are not depicted for clarity). The capacity on the second tour section is eight, since it is not yet negotiated which load units are transported on that section. The load unit lu_1 with a weight 2 has planned a tour from Hamburg to Dortmund via Bremen, with no fixed arrival time. It is currently being transported from Hamburg to Bremen for a transportation rate of 4 (represented by an accordingly labeled edge between the corresponding tour nodes). Another load unit lu_4 with a weight of five has planned a tour from Bremen to Dortmund with an arrival time not later than 6pm.

3.2 Sample Negotiation

Now consider the following concrete situation. A truck with a load capacity of eight has recently started its journey from Hamburg to Dortmund via Bremen, transporting different ULDs. One of these is lu_1, which has a weight of two and is scheduled for further transport to Dortmund, while all the others are unloaded in Bremen. Additionally three load units lu_2, lu_3, and lu_4 with the respective weights 4, 2, and 5, are queued for pickup in Bremen for transportation to Dortmund (here the position in the waiting queue is represented by the indices, in the environment the queue is represented by corresponding queue edges between the load unit nodes). ULDs are queued in the order of their arrival at the consolidation point, resp. the arrival of the corresponding transport order, preferring those ULDs that are already loaded on a truck.

Now each load unit may make an offer for transportation to the desired truck unit. The standard transportation rate in our simplified scenario is calculated to be the weight of the load unit multiplied by the transport time for the tour section in hours given by the timetable (so lu_2 would offer a rate of $4 * 2 = 8$). The offer is inserted into the environment by a graph transformation rule, which inserts an edge from the respective tour node of the load unit to the corresponding tour node of the truck unit, labeled with the actual offer and a question mark. These offers have to be made until the truck arrives at the consolidation point where the ULDs are queued. The graph in the left-hand side of Fig. 2 shows an excerpt of the sample environment after all load units that desire transportation from Bremen to Dortmund have made their offers.

When the truck arrives at the consolidation point in Bremen, the corresponding truck unit scans the offers of the load units respecting the queue order. In the concrete example, it will accept the ULDs lu_1, lu_2, and lu_3, which will pay the overall rate of 16 and moreover result in a full truckload. Technically this is achieved by a truck rule, which considers the offer of every load unit in the order in which they are queued. This rule is applicable, if the weight of a load unit is below or equal to the remaining capacity of the tour section under consideration (skipping every load unit with a weight that exceeds the currently remaining capacity for the tour section). When applied, the rule replaces the edge labeled with the offer by a reversely directed edge labeled with the same offer and an exclamation mark. If the rule is not applicable anymore then there is not sufficient capacity left for additional load units. The rejection of offers for those load units is handled by a truck rule, which replaces all remaining offer edges with reversely directed edges labeled with the same offer and '#'. This is depicted in the graph of the right-hand side of Fig. 2.

In the next step each load unit answers to the accepted or rejected offer. If a unit accepts the truck's decision it labels the corresponding edge with an additional 'OK'. If it does not accept the decision of the truck, it labels the corresponding edge with an additional 'Not OK'. In our example, lu_4 would not accept the decision, as it has to be transported by the considered truck in order to arrive in Dortmund on time. The left-hand side of Fig. 3 shows the situation in the environment after each load unit has commented on the decision of the truck unit.

The truck unit may now fix the transportation schedule if all edges are labeled with 'OK'. If an edge is labeled as 'Not OK', then the truck unit may calculate a new

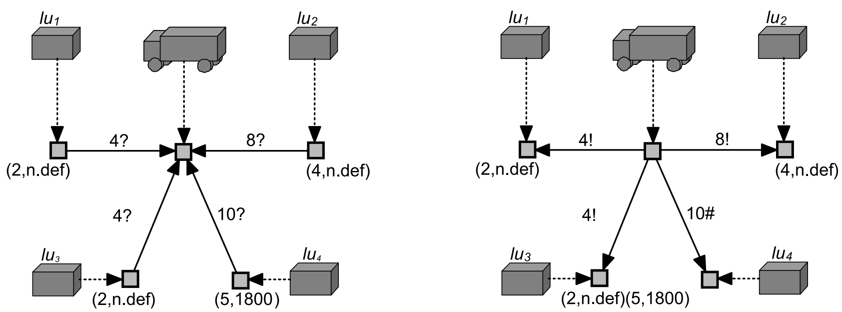


Fig. 2 Offers of load units and the truck unit

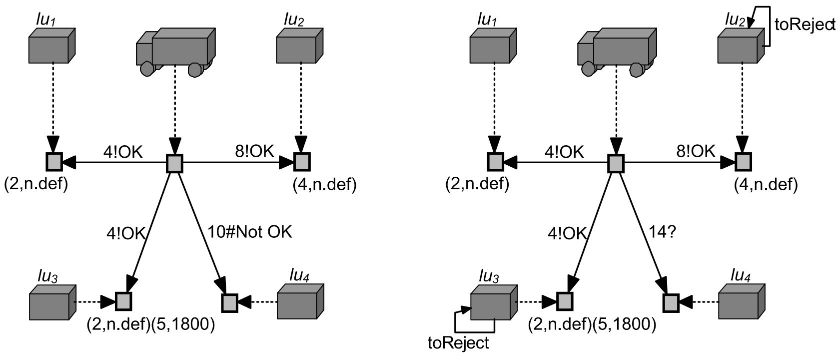


Fig. 3 Comments of the load units and new offer of the truck unit

rate for the transportation of the corresponding load unit based on the potential rejection of other load units. The necessary amount to pay is calculated by the truck as the missing amount compared to the full truckload plus one for every previously accepted load unit that now has to be rejected. This additional amount is then paid to the now rejected load units, so that they gain a small advantage for further negotiations. In the concrete example, the transportation of lu_4 with a weight of five would only be possible by likewise rejecting lu_2 and lu_3, resulting in an overall payment loss of two and the usually desired full truckload. In the given scenario, there is no chance to transport lu_4 and to get a full truckload. The necessary amount to pay for lu_4 is now calculated by the truck in the following way. In order to transport lu_4, the load unit lu_2 definitely has to be rejected due to its weight. Because the load unit lu_1 is considered first in the queue (since it is already loaded onto the truck), the load unit lu_3 will also be rejected. This yields enough capacity in the truck to transport lu_4. The payment rate of lu_1 is 4, a full truckload would amount to an overall transportation rate of 16 (the capacity of 8 multiplied by 2 hours). Therefore, the load unit lu_4 will have to pay a rate of 12 for compensating the difference compared to the full truckload, and additionally one for every previously accepted and now rejected load unit, in this case 2. Therefore, the overall sum in this example would be 14. The much higher cost of 14 compared to 10 can be justified by the fact, that this load unit has a higher urgency and is waiting for a much shorter time (as can be seen by its position in the waiting queue). Technically the rule of the truck unit handling this situation would replace the label of the corresponding edge with the calculated amount and a question mark (the direction of the edge makes it distinguishable from a load unit's offer). It would also mark the potential load units to reject with a loop labeled 'toReject'. The right-hand side of Fig. 3 depicts this situation.

If the load unit accepts the suggested payment, it can apply a rule which adds 'OK' to it. This in turn makes a rule of the truck unit applicable, which replaces the corresponding question mark by an exclamation mark. It also marks the previously flagged load units as rejected by replacing '!OK' with '#' in their corresponding edge labels and removing the loop edges labeled 'toReject'. The situation of the concrete scenario is depicted in the left-hand side of Fig. 4.

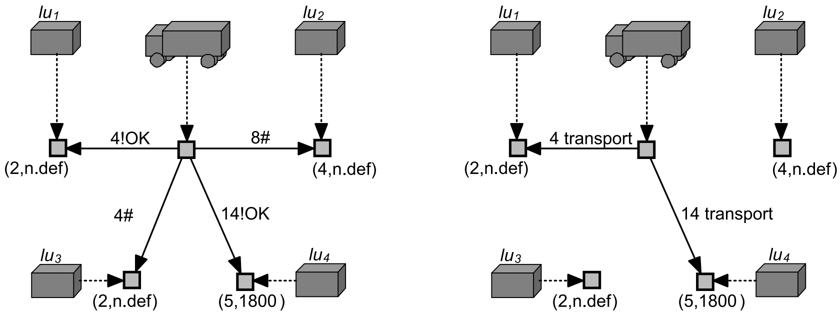


Fig. 4 Accepting a recalculated offer and finishing the negotiation

Now every rejected load unit again comments on the new decision by adding 'OK' or 'Not OK' to the corresponding edge labels. In the case of the concrete scenario, lu_2 and lu_3 would accept the truck's new decision, because they have no fixed arrival time and thus can wait for a later truck. In this case, the negotiation is considered finished which is handled by a rule of the truck unit. The application of this rule labels the edges connecting the truck tour node and the tour nodes of the transported load units with the negotiated transport rate and additionally 'transport'. The edges connecting the truck unit's tour node and the tour nodes of the rejected load units' tour nodes are removed. The right-hand side of Fig. 4 shows this situation.

However, if a previously accepted and now rejected load unit would disagree with the new situation, it would again add 'Not OK' to the respective edge label. Then the truck would in turn recalculate the now necessary amount to overrule the current proposition. This recalculation is done as before plus one for every negotiation round in order to avoid repeated transportation rates. An endless negotiation is not possible due to the fact, that the transportation rates are increasing with every negotiation round and every load unit has only a fixed amount at its disposal (and is of course not allowed to accept a transport rate which exceeds its own budget). An alternative could be the restriction of a maximum number of offer and accept/reject steps.

4 Conclusion

In this paper, we presented an approach to the intelligent scheduling of transports using communicating autonomous units. All logistic entities are represented by autonomous units having their own goals, their own capabilities, and their own control devices. It turned out that compared to traditional scheduling a sophisticated support of decentralized decision-making can significantly contribute to the performance of the overall system.

Autonomous units are a formal framework still under development. One aim is to use the formal framework to prove properties of the system and its components in general. An interesting aspect to prove in the context of the mentioned situations is the fact that every load unit will be transported to its final destination in time (provided that the overall constraints like capacities and timetables permit this).

Future work will regard extensions of the rather simple negotiation presented in this paper. Aspects like customer retention, transhipment costs, and competition between different logistic companies have to be considered for the negotiation to become more realistic. The possible effects on stability aspects, as investigated in e.g. (Scholz-Reiter et al. 2005), have to be considered. Alternative queueing techniques, as proposed in e.g. (Scholz-Reiter et al. 2006) should also be investigated.

A simulation tool that realizes executable autonomous units is currently being implemented. Once this tool is available, different negotiation concepts will be simulated and compared.

Although we presented a merely theoretical modeling concept, the prerequisites for a practical realization are already met (see, e.g., (Jedermann et al. 2006)).

Acknowledgements This research was supported by the German Research Foundation (DFG) as part of the Collaborative Research Centre 637 "Autonomous Logistic Processes".

References

Cooper RB (1981) Introduction to Queueing Theory, 2nd Edition, Elsevier North Holland

Jedermann R, Behrens C, Westphal D, Lang W (2006) Applying autonomous sensor systems in logistics; Combining Sensor Networks, RFIDs and Software Agents. Sensors and Actuators A (Physical), 132:370–375

Hölscher K, Knirsch P, Kreowski H-J (2006a) Modelling Transport Networks by Means of Autonomous Units. In: Haasis H-D, Kopfer H, Schönberger J (eds) Operations Research Proceedings 2005, Springer, Berlin Heidelberg New York, pp 399–404

Hölscher K, Kreowski H-J, Kuske S (2006b) Autonomous Units and their Semantics – the Sequential Case. In: Corradini A, Ehrig H, Montanari U, Ribeiro L, Rozenberg G (eds) Proc. 3rd International Conference on Graph Transformations (ICGT 2006), Lecture Notes in Computer Science vol 4178, Springer, Berlin Heidelberg New York, pp 245–259

Kreowski H-J, Kuske S (2007) Autonomous Units and Their Semantics – The Parallel Case. In: Fiadeiro JL, Schobbens PY (eds) Recent Trends in Algebraic Development Techniques, 18th International Workshop, WADT 2006, La Roche en Ardenne, Belgium, June 1–3, 2006, Revised Selected Papers. Lecture Notes in Computer Science Vol. 4409, Springer, Berlin Heidelberg New York, in print

Kreowski H-J, Kuske S (1999) Graph Transformation Units with Interleaving Semantics. Formal Aspects of Computing 11(6):690–723

Kreowski H-J, Kuske S, Schürr A (1997) Nested Graph Transformation Units. International Journal on Software Engineering and Knowledge Engineering, 7(4):479–502

Knirsch P, Kreowski H-J (2000) A Note on Modeling Agent Systems by Graph Transformation. In: Nagl M, Schürr A, Münch M (eds) International Workshop of Graph Transformation with Industrial Relevance (AGTIVE 1999), Lecture Notes in Computer Science Vol. 1779, Springer, Berlin Heidelberg New York, pp 79–86

Rozenberg G (ed) (1997) Handbook of Graph Grammars and Computing by Graph Transformation, Vol. 1: Foundations. World Scientific, Singapore

Scholz-Reiter B, Freitag M, de Beer C, Jagalski T (2006) Modelling and Simulation of a Pheromon based Shop Floor Control. In: Cunha P, Maropoulos P (eds) Proceedings of the 3rd Interna-

tional CIRP Sponsored Conference on Digital Enterprise Technology – DET2006. University of Setubal, Setubal
Scholz-Reiter B, Wirth F, Freitag M, Dashkovskiy S, Jagalski T, de Beer C, Rüffer, B (2005) Some remarks on the stability of manufacturing logistic networks. Stability margins. In: Proceedings of the International Scienific Annual Conference on Operations Research. Springer, Berlin Heidelberg New York, pp 91–96
Timm IJ, Knirsch P, Kreowski H-J, Timm-Giel A (2007) Autonomy in Software Systems. In: Hülsmann M, Windt K. (eds) Understanding Autonomous Cooperation & Control in Logistics – The Impact on Management, Information and Communication and Material Flow, Springer, Berlin Heidelberg New York, pp 255–274
Weiss G (ed) (1999) Multiagent Systems – A Modern Approach to Distributed Artificial Intelligence. The MIT Press, Cambridge, Massachusetts

Autonomously Controlled Adaptation of Formal Decision Models – Comparison of Generic Approaches

Jörn Schönberger, Herbert Kopfer

Chair of Logistics, University of Bremen, Wilhelm-Herbst-Straße 5,
28359 Bremen, Germany, e-mail: {jsb,kopfer}@uni-bremen.de

Abstract We consider a deployment problem from transportation logistics that requires a repeated update of an existing transportation plan. The system load varies significantly and unpredictable over time. In order to consider modifications of the decision premises we propose to adjust reactively the used decision model for a transport planning problem. Since the transportation plan update instances are represented in terms of an optimization problem we try to adapt the objective function as well as the constraint set in order to inject up-to-date problem knowledge into the formal problem representation. Furthermore, we provide a comprehensive report and comparison about the numerical results observed in simulation experiments and describe the different impacts of adapting the objective function and the constraint set of the model.

1 Introduction

This article is about different strategies to adapt automatically a decision model to a changed problem situation. By means of an example from transportation logistics, we assess two strategies for implementing recent knowledge about a volatile decision problem into an optimisation model representing the problem at hand: Adaptation of the constraint set and adjustment of the objective function.

Section 2 introduces the considered decision problem. Section 3 describes the proposed model adaptation strategies. Section 4 describes the experimental setup and reports the achieved numerical results.

2 Vehicle Scheduling Problem

The problem we are investigating in this contribution generalises the common vehicle routing problem with time windows (Solomon, 1987) in three aspects.

DOI: 10.1007/978-3-540-76862-3, © Springer 2008

a) Soft Time Windows. Lateness at a customer site is possible but causes penalty costs. Although a particular request is allowed to be late, the portion p^{target} of the f_t requests completed in $[t^-;\ t]$ and expected to be completed in $[t;\ t^+]$ must be in time. Let f_t^{comp} be the number of the requests completed timely within the last t^- time units and let f_t^{expec} be the number of punctually scheduled requests within the next t^+ time units, then $p_t := (f_t^{\text{comp}} + f_t^{\text{expec}})/(f_t) \geq p^{\text{target}}$ has to be achieved.

b) Subcontracting. Logistics service providers (LSP) are paid for the reliable fulfilment of selected requests. An LSP receives a certain amount of payment for this service and ensures that the request is fulfilled within the specified time window. A subcontracted request remain unconsidered while constructing the routes for the own vehicles. If a request has been subcontracted then this decision cannot be revised later on.

c) Uncertain Demand. Only a subset of all requests is known to the planning authority at the time when the decision concerning subcontracting is made and the routes for the own vehicles are generated. The planning authority decides about subcontracting or self-fulfilment of a request as soon as it becomes known.

Decision model. A transportation plan describes how the known requests are fulfilled. Subsequently arriving requests are accepted and handled by updating the existing transportation plan. A sequence of transportation plans TP_0, TP_1, TP_2, ... is generated reactively at the ex ante unknown update times t_0, t_1, t_2, ... and each single transportation plan is executed until it is updated. It is aimed at keeping the costs for the execution of the additional requests as low as possible but, on the other hand, to provide a sufficiently high reliability within the request fulfilment.

A single request r attains consecutively different states. Initially, r is known but not yet scheduled (F). Then, r is assigned to an own vehicle (I) or subcontracted (E). If the operation at the corresponding customer site has already been started but not yet been finished the state (S) is assigned to r. The set $R^+(t_i)$ is composed of additional requests released at time t_i. Requests completed after the last transportation plan update at time t_{i-1} are stored in the set $R^C(t_{i-1}, t_i)$. The new request stock $R(t_i)$ is determined by $R(t_i) := R(t_{i-1}) \cup R^+(t_i) \setminus R^C(t_{i-1}, t_i)$ and the set $R^X(t_i)$ contains all currently available requests belonging to the state $X \in \{F, E, I, S\}$.

The transportation plan update problem at time t_i is as follows. Let V denote set of all own vehicles, $P_v(t_i)$ the set of all routes executable for vehicle v and let $P(t_i)$ denote the union of the sets $P_v(t_i)(v \in V)$. If the request r is served in path p then the binary parameter a_{rp} is set to 1, otherwise it is set to 0. A request r, already known at time t_{i-1} that is not subcontracted in TP_{i-1} is served by vehicle $v(r)$ according to TP_{i-1}. The travel costs associated with route p are denoted as $C^1(p)$ and $C^2(p)$ refers to the penalties associated with p. Finally, $C^3(r)$ gives the subcontracting costs of request r.

We deploy two families of binary decision variables. Let $x_{pv} = 1$ if and only if path $p \in P(t_i)$ is selected for vehicle $v \in V$ and let $y_r = 1$ if and only if request r is subcontracted.

$$\sum_{p \in P(t_i)} \sum_{v \in V} \left(C^1(p) + C^2(p)\right) x_{pv} + \sum_{r \in R(t_i)} C^3(r) y_r \to \min \tag{1}$$

$$\sum_{p \in P_v(t_i)} x_{pv} = 1 \quad \forall v \in V \tag{2}$$

$$x_{pv} = 0 \quad \forall p \notin P_v(t_i), v \in V \tag{3}$$

$$y_r + \sum_{p \in P(t_i)} \sum_{v \in V} a_{rp} x_{pv} = 1 \quad \forall r \in R(t_i) \tag{4}$$

$$y_r = 1 \quad r \in R^E(t_i) \tag{5}$$

$$\sum_{p \in P_{v(r)}} a_{rp} x_{pv(r)} = 1 \quad \forall r \in R^S(t_i) \tag{6}$$

The overall costs for TP_i are minimised (1). One route is selected for each vehicle (2) and the selected path p is realisable by vehicle v (3). Each single request known at time t_i is served by a selected vehicle or given away to an LSP (4) but a once subcontracted request cannot be re-inserted into the paths of the own vehicles (5). An (S)-labelled request cannot be re-assigned to another vehicle or LSP (6).

Test cases. The construction of artificial test cases from the Solomon instances (Solomon, 1987) is described by Schönberger and Kopfer (2007). In these scenarios, demand peaks that represent significant changes in the decision situations interrupt balanced streams of incoming requests. Two fulfilment modes are available for each request: self-fulfilment and subcontracting. The costs for the first mode are normalized to 1 monetary unit for each travelled distance unit. For each subcontracted distance unit α monetary units have to be paid to the LSP. Each single request r causes overall costs of $F_r := C^3(r)$ monetary units calculated by multiplying the distance between the LSP depot and the customer site location of r with α.

3 Online Decision Strategies

We use the Memetic Algorithm (MA) scheduling framework introduced in Schönberger (2005) for the repeated update of the transportation plans. Whenever new requests are released then the MA is re-called in order to produce an updated transportation plan. This approach is a realisation of the generic online optimisation concept (Krumke, 2001). Is referred to as NONE because neither the search space nor the objective function are adapted to current situations.

Search Space Adaptation. The severeness of the consecutively solved optimization models changes due to the variation of the system load. In order to consider load modifications, we enforce the subcontraction if the currently observed punctuality p_t falls below the desired least punctuality p^{target}.

A continuous piece-wise linear control function h that is equal to 1 if $p_t \leq p^{\text{target}} - 0.05, 0$ if $p_t \geq p^{\text{target}} + 0.05$ and that falls proportionally from 1 down to 0 if p_t increases from $p^{\text{target}} - 0.05$ up to $p^{\text{target}} + 0.05$ is used in order to determine the intensity of the enforcement of the fulfillment mode “subcontraction”. Therefore, at first we select randomly the portion $h(p_t)$ of $R^+(t)$ and the fulfillment mode of the selected requests (collected in $R^{\text{pre}}(t)$) is pre-select as “subcontraction”. It is not allowed to change the selected mode anymore so that we have to replace the constraint (5) by the new constraint (7).

$$y_r = 1 \quad r \in R^E(t_i) \cup R^{\text{pre}}(t_i) \tag{7}$$

Since the constraint (7) is adapted with respect to the current punctuality, this online decision strategy is referred to as Constraint Set Adaptation (CSAD).

Search Direction Adaptation. An adaptation of the cost coefficient in the objective function (1) leads to a re-valuation of the costs. This adjustment corresponds to the adaptation of the search direction of the used solver. In order to enable such a search direction adaptation (SDAD), the objective function (1) is replaced by the function (8) and the coefficient f_i is set to $f_0 = 1$ and $f_i = 1 + \alpha h(p_{t_t})$.

$$\sum_{p \in P(t_i)} \sum_{v \in V} f_i \left(C^1(p) + C^2(p)\right) x_{pv} + \sum_{r \in R(t_i)} C^3(r) y_r \rightarrow \min \tag{8}$$

As long as the punctuality is sufficiently large $(p_{t_i} \geq p^{\text{target}} + 0.05)$ both fulfilment modes are weighted only by the original costs. As soon as the punctuality decreases f_i increases and if p_{t_i} has reached an unsatisfactory level $(p_{t_i} \geq p^{\text{target}} - 0.05)$, then $f_i = 1 + \alpha$, which means that in (8) the costs for subcontraction are weighted less than the self-entry costs. After the coefficient has been adjusted, the updated model is solved and is used for the generation of a transportation plan update.

4 Numerical Experiments

Experimental setup. We deploy a piece-wise linear penalty function, which is 0 for delays shorter than 10 time units and which increases proportionally up to a maximal value of 25 money units for delays longer than 100 time units. The target punctuality is set to $p^{\text{target}} = 0.8$.

We perform experiments for different tariff levels $\alpha \in \{1, 1.25, 1.5, 1.75, 2, \text{and} 3\}$ ($\alpha = 1$ represents a fair and comparable LSP tariff which leads to the same costs as in the case of self-fulfillment). Requests are taken from the Solomon instances R103, R104, R107, R108. Each of the $|\{$R103, R104, R107, R108$\}| \cdot |\{$1, 1.25, 1.5, 1.75, 2,3 $\}| \cdot |\{$NONE, CSAD, SDAD$\}| = 72$ scenarios is simulated three times leading to overall 216 performed simulation experiments. Here, we report the average results observed for each scenario. For analysing the impacts of the objective function adaptation and the constraint set adaptation with respect to the tariff level α and the decision policy $\varepsilon \in \{$NONE,CSAD,SDAD$\}$, we first calculate p_t's maximal decrease $\delta(\varepsilon, \alpha) := \min_{t \geq 1500}(p_t(\varepsilon, \alpha)/p_{1000}(\varepsilon, \alpha))$ after the demand peak's start.

Let $T^{\text{below}}_{\varepsilon,\alpha}$ denote the first time in which p target is not achieved and $T^{\text{heal}}_{\varepsilon,\alpha}$ refers to the time in which p^{target} is finally re-achieved. We define $\pi(\varepsilon, \alpha) := (T^{\text{heal}}_{\varepsilon,\alpha} - T^{\text{below}}_{\varepsilon,\alpha})/(4000)$ as the percentage of low quality situations $(p_t < p^{\text{target}})$ within the observation interval $[1000, 5000]$.

Beside the effects on the process reliability, we have recorded the resulting process costs. Let $C_{\varepsilon,\alpha}(t)$ denote the cumulated overall costs realized up to time t. In order to quantify the impacts of tariff level rising, we calculate the relative growth $c(\varepsilon,\alpha) := (C_{\varepsilon,\alpha}(5000))/(C_{\varepsilon,1}(5000)) - 1$ caused by increasing freight tariffs. To compare the impacts of the cost criteria in the mode decision, we calculate the cost increase $r(\varepsilon,\alpha) := c(\varepsilon,\alpha)/c(NONE,\alpha) - 1$ caused by switching from NONE to CSAD and SDAD. Similarly, we calculate the relative growth of the travel costs $c^{\text{int}}(\varepsilon,\alpha)$, of the subcontracting costs $c^{\text{ext}}(\varepsilon,\alpha)$ as well as of the penalty costs $c^{\text{pen}}(\varepsilon,\alpha)$. The contribution of the travel costs to the overall costs is defined as $m^{\text{int}}(\varepsilon,\alpha) := C^{\text{int}}_{\varepsilon,\alpha}(5000)/C_{\varepsilon,\alpha}(5000)$ (the portion $m^{\text{ext}}(\varepsilon,\alpha)$ of the subcontracting costs as well as the portion $m^{pen}(\varepsilon,\alpha)$ of the penalties are determined in the same way).

Simulation Results. Independently from the used strategy ε, the reliability decreases if the freight level α is increased. A severe decrease of $\delta(\varepsilon,\alpha)$ is observed for increasing α. The application of CSAD as well as SDAD leads to values of $\delta(\varepsilon,\alpha)$ above 91 %. However, SDAD seems to be able to keep $\delta(\varepsilon,\alpha)$ on a slightly higher level than CSAD (Table 1).

The percentage of situations with a punctuality below p^{target} increases if the tariff level is lifted by the tariff level increase but the application of a model adaptation defers the occurrence to higher tariff levels (Table 2).

The increase in the service level is mainly based by an extension of the LSP incorporation. The results in Table 3 show that the application of CSAD as well as SDAD leads to an increase of $\sigma(\varepsilon,\alpha)$.

CSAD and SDAD overrule the cost criteria in the mode selection for α>1 so that additional costs occur. Table 4 shows that SDAD leads to significantly less additional costs (10.9%) than the application of CSAD (49.8%). However, the severeness of the cost increase after an increase of the freight tariff level is quite different in the three strategies (Table 5).

Table 1 Minimal punctuality $\delta(\varepsilon,\alpha)$

ε	α					
	1	1.25	1.5	1.75	2	3
NONE	98.0%	97.0%	93.6%	80.7%	81.5%	72.0%
CSAD	92.9%	96.4%	93.6%	93.5%	93.3%	91.1%
SDAD	99.0%	97.0%	95.6%	96.1%	93.1%	92.2%

Table 2 Percentage $\pi(\varepsilon,\alpha)$ of replanning situations with a punctuality below p^{target}

ε	α					
	1	1.25	1.5	1.75	2	3
NONE	–	5.0%	15.0%	42.5%	42.5%	47.5%
CSAD	–	–	5.6%	8.3%	13.9%	19.4%
SDAD	–	–	–	–	5.0%	7.5%

Table 3 Maximal quote of subcontracted requests $\sigma(\varepsilon,\alpha)$

ε	α					
	1	1.25	1.5	1.75	2	3
NONE	18.8%	13.5%	9.6%	10.9%	5.7%	9.3%
CSAD	21.2%	17.1%	20.0%	20.2%	20.3%	20.9%
SDAD	21.9%	17.1%	16.0%	15.2%	15.5%	16.5%

The overall costs are split into three cost drivers: travel costs (int), subcontracting costs (ext) and penalties (pen) as shown in Table 6. Independently from the applied adaptation strategy, the highest cost increase is caused by additional penalty costs. The second highest cost driver is the travel costs. However, the highest discrepancies are observed for the subcontracting costs. In the NONE experiment, the pure cost-based decision strategy disqualifies the subcontracting mode as more as the tariff level is increased. At the end ($\alpha = 3$), the external costs have been reduced by 91.4 %. If the search direction is adapted in the SDAD experiment

Table 4 Cost increase $r(\varepsilon,\alpha)$ after switching from NONE to CSAD or SDAD

ε	α					
	1	1.25	1.5	1.75	2	3
CSAD	2.8%	11.4%	17.3%	23.1%	25.5%	49.8%
SDAD	4.6%	8.2%	7.5%	6.8%	5.9%	10.9%

Table 5 Increase $c(\varepsilon,\alpha)$ of the cumulated overall costs

ε	α					
	1	1.25	1.5	1.75	2	3
NONE	0.0%	4.9%	13.6%	20.4%	27.6%	43.0%
CSAD	0.0%	13.8%	29.6%	44.3%	55.7%	108.5%
SDAD	0.0%	8.6%	16.7%	23.0%	29.2%	51.7%

Table 6 Cost increase

ε		α					
		1	1.25	1.5	1.75	2	3
NONE	$c^{\text{int}}(\varepsilon,\alpha)$	0.0%	70.0%	119.5%	148.7%	167.1%	206.4%
	$c^{\text{ext}}(\varepsilon,\alpha)$	0.0%	–45.1%	–67.0%	–77.6%	–80.8%	–91.4%
	$c^{\text{pen}}(\varepsilon,\alpha)$	0.0%	93.3%	141.4%	182.6%	237.3%	430.0%
CSAD	$c^{\text{int}}(\varepsilon,\alpha)$	0.0%	50.3%	87.8%	115.9%	132.6%	154.1%
	$c^{\text{ext}}(\varepsilon,\alpha)$	0.0%	–10.7%	–8.5%	–1.7%	6.7%	78.3%
	$c^{\text{pen}}(\varepsilon,\alpha)$	0.0%	78.1%	113.7%	129.2%	140.4%	182.9%
SDAD	$c^{\text{int}}(\varepsilon,\alpha)$	0.0%	39.0%	66.1%	90.5%	106.8%	137.1%
	$c^{\text{ext}}(\varepsilon,\alpha)$	0.0%	–7.8%	–9.9%	–13.1%	–12.3%	5.8%
	$c^{\text{pen}}(\varepsilon,\alpha)$	0.0%	56.7%	92.4%	119.2%	137.9%	182.1%

then the subcontracting cost increase is rather small but if the constraint set is adapted then the promotion of this fulfilment mode leads to additional subcontracting costs of 78.3 %. As long as the freight tariffs are comparable ($\alpha = 1, 1.25$), the LSP costs contribute mostly to the overall costs independently from the applied strategy ε. For higher freight tariffs, travel costs of the own vehicles become the most important cost driver in the NONE-experiment but if the decision model is adapted then LSP charges as well as travel costs contribute similarly to the overall costs although CSAD leads to a higher contribution of LSP charges (Table 7).

In order to understand the different costs, we have analysed the recorded control function values $h(t)$ in the CSAD as well as in the SDAD experiment. Figure 1 shows the intervention intensities for CSAD (dashed bold line) and SDAD (continuous bold line). Although the CSAD control function produces a lower average value

Table 7 Split of costs

ε		α					
		1	1.25	1.5	1.75	2	3
NONE	$m^{\text{int}}(\varepsilon,\alpha)$	39.8%	64.5%	76.9%	82.2%	83.3%	85.2%
	$m^{\text{ext}}(\varepsilon,\alpha)$	57.2%	29.9%	16.6%	10.6%	8.6%	3.4%
	$m^{\text{pen}}(\varepsilon,\alpha)$	3.1%	5.6%	6.5%	7.2%	8.1%	11.4%
CSAD	$m^{\text{int}}(\varepsilon,\alpha)$	36.0%	47.6%	52.1%	53.9%	53.8%	43.9%
	$m^{\text{ext}}(\varepsilon,\alpha)$	61.2%	48.1%	43.2%	41.7%	41.9%	52.3%
	$m^{\text{pen}}(\varepsilon,\alpha)$	2.8%	4.4%	4.6%	4.4%	4.3%	3.8%
SDAD	$m^{\text{int}}(\varepsilon,\alpha)$	32.3%	41.3%	45.9%	50.0%	51.7%	50.5%
	$m^{\text{ext}}(\varepsilon,\alpha)$	65.7%	55.8%	50.7%	46.4%	44.6%	45.8%
	$m^{\text{pen}}(\varepsilon,\alpha)$	2.0%	2.9%	3.3%	3.6%	3.7%	3.7%

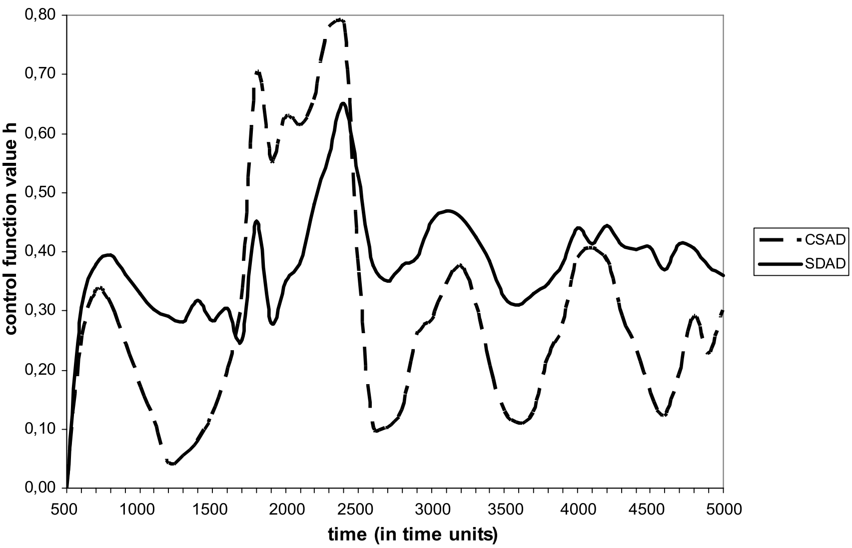


Fig. 1 Control function values and self entry quotes observed in the $\alpha = 3$ experiment

than the SDAD control function, the amplitude of the CSAD control function oscillation is quite larger than the amplitude of the SDAD control function oscillation. This might be a reason, why CSAD leads to more intensive subcontracting usages with a lower number of requests served by own vehicles.

With respect to the costs, SDAD dominates CSAD (cf. Tables 4 and 5), however, CSAD is able to enforce the subcontracting of requests to a larger extend (cf. Table 3). This might be an idea to combine SDAD and CSAD into a common strategy that uses SDAD mechanisms to heal small punctuality deficiencies and that employs CSAD capabilities to correct severe punctuality decreases.

5 Conclusions

We have analysed an evolving decision problem from transportation logistics. In order to implement ex ante unknown problem knowledge into the formal model decision, we have proposed extensions to the general online optimization framework that allow the contextual formulation of the instances in an online optimization problem. Within comprehensive numerical experiments, we have proven the general applicability of the adaptation strategies and a comparison has been carried out. Future research will address the combination of the so far separately used adaptation strategies.

Acknowledgements This research was supported by the German Research Foundation (DFG) as part of the Collaborative Research Center 637 "Autonomous Cooperating Logistic Processes" (Subproject B7)

References

Krumke S (2001) Online-Optimization – Competitive Analysis and Beyond. Habilitation Thesis, Technical University Berlin, Germany.

Schönberger J, Kopfer H (2007) On Decision Model Adaptation in Online Optimization of a Transport System, in: Günther H-O, Mattfeld DC, Suhl L: Management logistischer Netzwerke, Springer 2007, Berlin, Heidelberg, S. 361–381

Schönberger, J (2005) Operational Freight Carrier Planning. Springer.

Solomon, M (1987) Algorithms for the vehicle routing and scheduling problem with time window constraints. Operations Research 35 (2), pp. 254–265.

Clustering in Autonomous Cooperating Logistic Processes

Gulshanara Singh, Bernd-Ludwig Wenning, Amanpreet Singh, Carmelita Görg

Communication Networks, University of Bremen, Bremen, Germany

Abstract Due to the growing complexity of logistic processes, "Autonomous Cooperating Logistic Processes" are considered as a way to handle this complexity growth (Scholz-Reiter et al. 2004). In this concept, knowledge and decisions are distributed among the participants of the logistic process. Vehicles and goods become intelligent, interactive, and capable of deciding about how to achieve their aims.
Logistic components may have common aims, e.g., several goods that are at the same location and have the same destination. In such a case, it can be sensible to form communities of those components and determine a community leader that acts on behalf of all members. It is expected that thereby, the required communication among the logistic components can be optimized.
This paper identifies challenges in the area of communication that arise from the distributed decision process and the interacting components. An approach to form clusters among the goods is proposed to address these challenges.

1 Introduction

The increasing complexity in logistic processes together with rising dynamics has resulted in embedding new technologies or paradigms in the logistic domain. Autonomous Cooperation (AC) is one of the growing paradigms that has witnessed tremendous upsurge in recent years for handling this dynamism and complexity. This paradigm shift is facilitated by the availability of a wide range of information and communication technologies that can be utilised to devolve decision making down to the level of a vehicle and indeed the individual item in the logistic chain [1],[2]. The implementation of AC aims at a flexible self-organizing system structure that is able to cope with the dynamics and complexity while maintaining a stable status [3]. Thus, logistic entities can be defined as items that have the ability to interact and communicate with other entities in the logistic network. Software

DOI: 10.1007/978-3-540-76862-3,

agent technology is said to provide the means of creating autonomous, intelligent and social software entities capable of supporting autonomous decision making by sharing and interacting with each other. The distribution of planning and decision-making to autonomous components is a widely accepted promising solution to handle complex problems [4]. The motivation comes from the fact that the agent-based systems reflect the distributed nature and are able to deal with the dynamics of execution and planning on the real-time settings [5]. In case of large networks with large number of logistic entities it may result in substantially large communication. So the key challenge is the reduction in the amount of communication to bring about autonomous cooperation in a logistic network. So an approach of clustering the packages first and then routing to finally reach their destination has been introduced in this paper and a result on amount of communication traffic is presented with analytical formulations.

2 Routing and Clustering Approach

In case of transport logistics, AC means that the vehicles and goods are able to act independently according to their objective such as vehicles try to achieve aims such as cost efficiency while the goods aim to find a fast path to their destination. As the goods need to be transported by vehicles, there is an interdependence between the goods' and the vehicles' decisions. In order to deal with this interdependence, the "Distributed Logistic Routing Protocol" (DLRP) was proposed in [6]. In this protocol, information retrieval is done using route request and route reply messages. In large networks, the route request and route reply messages cause a significant amount of communication traffic. If goods with similar aims can be combined into a community with one "speaker", a reduction of this traffic is expected.

Logistic entities are represented by software agents, and the solution proposed to reduce the communication traffic is the approach of clustering these logistics entities (software agents). The objective is to cluster the logistic entities which have common goals like packages having the same destination, type of packages etc shown in Fig. 1. Organizing logistical entities into clusters can reduce the communication overhead associated with each of the individual entities and provide scalable solutions. The data associated with the package agent (like the destination ID) can be used as a reference to form clusters by performing intra-communication (communication within packages of same destination) and inter-communication (communication between various cluster-heads). Once the cluster is formed, a cluster-head can be selected to act on behalf of the cluster members instead of the individual entity communicating with each other. Thus, the cluster-head will be the information pool which can negotiate with other entities, for example vehicles. Hence, once the vehicle gets the knowledge of a particular cluster via a cluster-head, it can decide which route is appropriate to reach the destination.

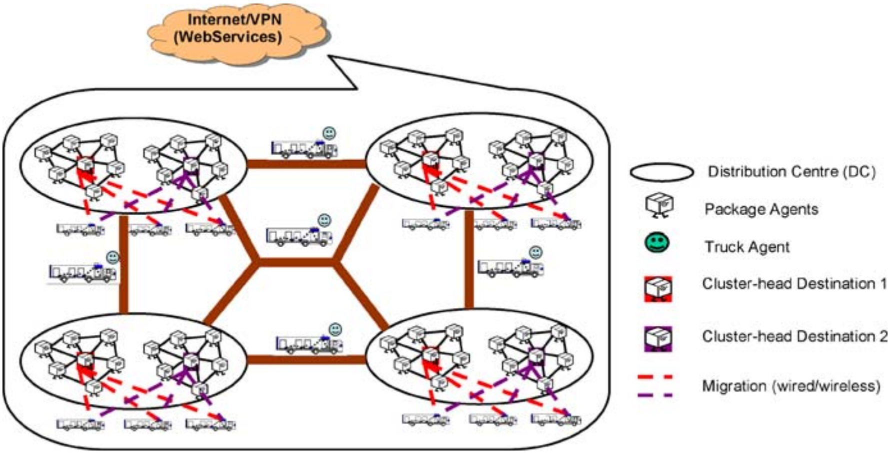


Fig. 1 Transportation Logistics: Clustering of Logistic Entities

This paper addresses the measure of communication traffic analytically with the approach of clustering and routing of logistic entities. Clustering has been a research topic in various fields of communication networks like sensor networks, ad-hoc networks where all deal with the problem of energy saving (battery life) by reducing communication [7]. The underlying problem in all domains is the same; given a number of objects, a cluster should be created, such that items in the same cluster are more similar to each other than they are to items in other clusters. Since mobile agents are autonomous and intelligent they can be used for creating dynamic and adaptive clusters in a logistical network. So the algorithms applied in the previously mentioned fields can be mapped to logistic networks. For mapping of these algorithms from communication networks, it is necessary to identify where these networks are similar and where they are different.

Clustering can be applied at various levels in a logistical network. Clustering of the packages can be done based on common destinations or type of objects like food, clothing etc. Clustering can also be done by the cities depending on the route taken by the vehicle, or the vehicles itself based on the material they are carrying or destination to where they are travelling via a certain route.

3 Scenario Description

In this paper, a semi-autonomous scenario is assumed first concentrating on the cluster formation and then routing of the logistic entities. The packages are clustered based on their common destinations. Once the clusters of packages are formed and the cluster-heads nominated (selection of cluster-heads based on different criteria like large lifetime, delivery date etc.), the cluster-heads will initiate the routing procedure instead of the individual packages. The amount of communication traffic for the routing procedure with and without-clustering of the packages is measured and compared analytically based on the number of messages exchanged.

3.1 Messages Sent during Clustering

In the clustering procedure, the logistical entities involved are the package agents, the associated vertex (Distribution Centre), and the cluster-heads (after the selection procedure). The messages exchanged between the packages, vertices and the cluster-heads are presented in the Fig. 2. In the initial stage, we assume that, once the package (agent) arrives at its associated vertex, it sends a registration request (RegReq) and gets back a registration acknowledgement (RegAck). The associated vertex on sending acknowledgement looks out for the cluster which has the same destination as for the presently arrived package. Once it finds that the newly arrived package has no clusters based on this destination, it forms a new cluster based on that destination and selects the present package as the cluster-head sending it a message (CHAnn). In case it has a cluster with the same destination, its sends the cluster-head information (CHInfo) of that cluster to the package so that it can register itself with that cluster (CRegReq). In addition, the cluster-head will send back the acknowledgement (CRegAck). The message cluster completion (CComplete) is considered in this scenario based on the cluster size assumed.

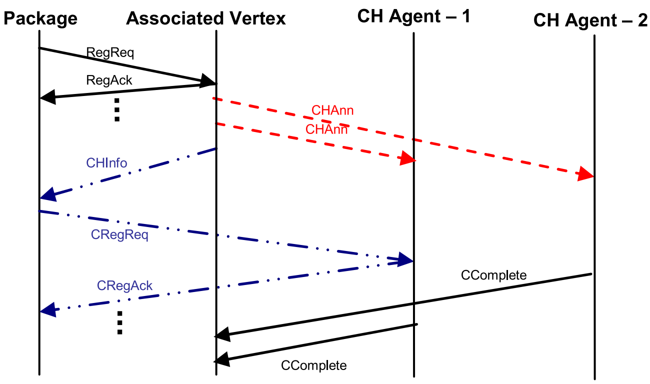


Fig. 2 Messages for Clustering

4 Communication Traffic for Clustering

The total amount of communication traffic for clustering taking into account some assumptions of the message exchange, the number of packages, the number of destinations, the number of clusters and cluster size are presented below.

4.1 Representation and Assumption

Number of packages stored in a DC = N_packs
Number of destinations = N_dests
Number of Clusters = N_Clusters
Cluster size = Cl_size
Total number of Register Request (RegReq) = N_packs
Total number of Register Acknowledge (RegAck) = N_packs

Cluster-head Announcement/Information:
Total number of Cluster-head Announcements (CH_Ann) = N_dests
Total number of Cluster-head Information (CH_Info) = N_packs – N_dests

Clustering Process:
Total number of Cluster Register Request (CRegReq) = N_packs – N_dests
Total number of Cluster Register Acknowledge (CRegAck) = N_packs – N_dests

$Total\ Clustering\ Volume = 5 * N\_packs - 2 * N\_clusters$
where $N\_clusters = N\_dests * \lceil N\_packs/(N\_dests * Cl\_size) \rceil$

4.2 Messages Sent during Routing

During the routing process in DLRP, the entity (package or vehicle) that performs the routing generates a high amount of data traffic. As there are usually more packages than vehicles and clustering should be applied to the packages, only the package routing is covered in the following. The routing process is illustrated by the Fig. 3.

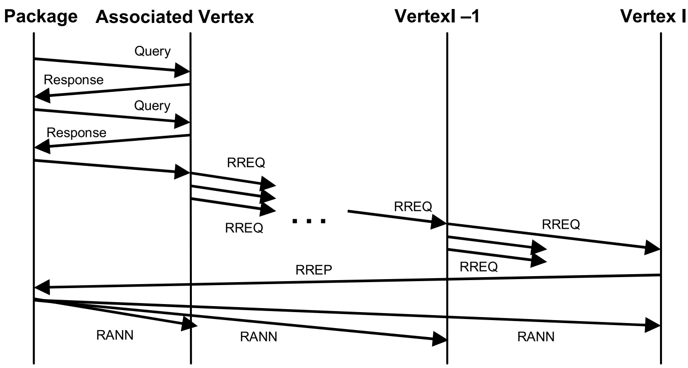


Fig. 3 Messages for Routing

The routing starts with two queries to the associated vertex and the corresponding responses. These queries inform the package about some initial parameters such as destination, associated vertex etc that it needs for the routing. Then the package sends a route request (RREQ, exactly one) to its associated vertex, which in turn adds some data to the request and forwards it to all neighbour vertices, thus it is multiplied by the vertex's branching factor, which states how many neighbours are available as recipients of the forwarded route request. This is continued until the request reaches the destination or its hop limit. A route reply (RREP) is then sent back for each request that reaches the destination.

Assuming an average branching factor b and an average route length of l hops, the amount of route replies is b^{l-1}, while the total number of route requests sent in the network is $\sum_{i=0}^{l-1} b^i$.

After having received the route replies, the package decides the route and announces it to the affected vertices, which leads to other l route announcements (RANN) per route, so if the package announces n alternative routes, the total number of route announcements is nl.

For a small logistic network with a branching factor of 3 and packages choosing 3 alternative routes, an average route length of 4 would correspond to 40 route requests and 27 route replies for a single routing process, as depicted in Fig. 5. If for a network the average route length is just one more i.e. 5 then this would lead to huge increase in routing traffic (121 route requests and 81 route replies).

In the case where each package routes individually, each package generates this amount of messages. In contrast, if the routing is only done by one cluster head instead, only the cluster head generates the messages. Therefore, clustering brings

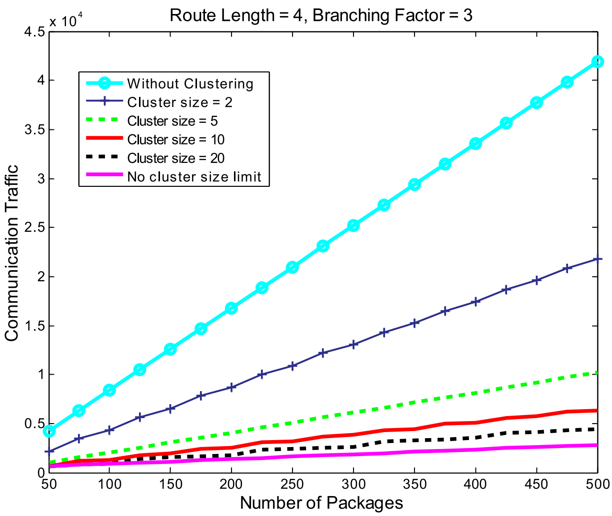


Fig. 4 Communication traffic vs. Number of packages

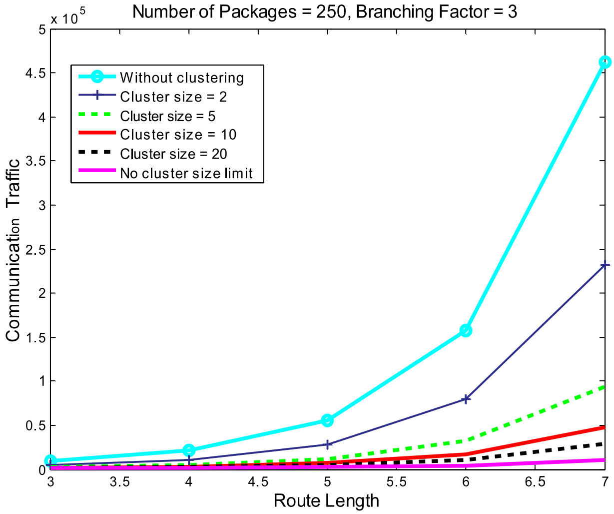


Fig. 5 Communication traffic vs. Route length

a reduction of routing messages by a factor of Cl_size when compared to the non-clustered case.

The assumptions with respect to number of packages, number of vehicles is presented in the Table 1.

Table 1 Parameters for Clustering and Routing

Parameters	Representation	Value
Number of Packages	N_packs	Min 50 Max 500
Number of Destinations	N_dests	5
Route Length	l	Min 3 Max 7
Branching Factor	b	Min 1 Max 8
Number of alternate routes	n	3
Cluster size	Cl_size	Min 2 Max 20

5 Results

The results depicted in Figs. 4, 5 and 6 give the total amount of communication traffic taking into account clustering and routing messages for varying number of packages, route length, and branching factor, respectively. The set of curves are

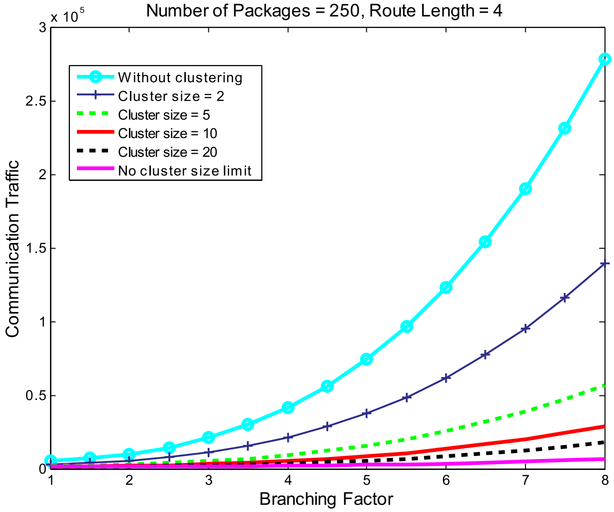


Fig. 6 Communication traffic vs. Branching factor

presented for the parameters such as *Without clustering,* varying *Cluster size*, and *No cluster size limit*. As seen in Fig. 4 the communication traffic goes on increasing linearly with increasing number of packages and is of maximum value in the case *Without Clustering*, while the communication volume decreases gradually with the increase in the cluster size. Figure 5 depicts the results for varying route length, and it is observed that the communication traffic has its maximum value in case of *Without clustering* and the result is similar for the case of varying branching factor as presented in Fig. 6. This implies that as the route length and branching factor increases the amount of communication traffic increases substantially but not as high as compared to the increase in the number of packages.

Hence, it can be concluded that clustering of the entities helps in reducing the communication volume depending on the size of the clusters. The larger the cluster size, the less the communication between the logistical entities will be as only one of the members needs to communicate on behalf of all others whereas there will be a slight increase in the communication within the cluster members to form the clusters.

6 Summary and Outlook

In this paper, the communication-related challenges of autonomous cooperating logistic processes were briefly introduced, and a cluster-based approach to reduce the communication volume was proposed. Some results are presented with respect to the communication volume with varying number of packages, route length, and branching factor. This approach will be investigated by simulation in the near future, expecting a significant improvement in communication efficiency.

References

1. Scholz-Reiter B, Windt K, Freitag M (2004) "*Autonomous logistic processes: New demands and first approaches*." Monostori L (ed) Proc 37th CIRP International Seminar on Manufacturing Systems. Hungarian Academy of Science, Budapest, pp. 357–362
2. Tarantilis CD, Kiranoudis CT (2002) "*Using a Spatial Decision Support System for Solving the Vehicle Routing Problem*." Information and Management, 39, pp. 359-375
3. Windt B and Hülsmann M (2007) "*Perspective on Initial Ideas and Conceptual Components of Autonomous Cooperation and Control*." Understanding Autonomous Cooperation and Control in Logistics, Springer, Berlin, pp. 23-38
4. Jennings NR (2001) "*An Agent-based Approach for Building Complex Software Systems*." Comms. Of the ACM 44(4), pp. 35-41
5. Dorer K and Calisti M (2005) "*An Adaptive Solution to Dynamic Transport Optimization*", Proc 4th AAMAS, pp. 45-51
6. Scholz-Reiter B, Rekersbrink H, Freitag M (2006) *"Kooperierende Routingprotokolle zur Selbststeuerung von Transportnetzen."* Industrie Management 3/2006, pp. 7-10
7. Bandyopadhyay S, Coyle Edward J (2003) "*An Energy-Efficient Hierarchical Clustering Algorithm for Wireless Sensor Networks*", Proc IEEE INFOCOM 2003, pp. 1713-1723

Application of Small Gain Type Theorems in Logistics of Autonomous Processes

Sergey Dashkovskiy, Björn Rüffer, Fabian Wirth

Zentrum für Technomathematik, Universität Bremen, Germany

Abstract In this paper we consider stability of logistic networks. We give a stability criterion for a general situation and show how it can be applied in special cases. For this purpose two examples are considered.

1 Introduction

The control of one processing machine or a small plant of several machines can be performed by one (central) control unite. A control can be designed in such a way that a production process is stable in the sense that the number of parts waiting in the buffer of the machine to be processed remains bounded. In case of a plant we can speak of the number of orders to be complete. Such number describes the state of a system. In case of large systems a centralized control becomes unflexible and may be even impossible. One of the possible solutions is the introduction of autonomous control (Scholz-Reiter et al., 2007a), i.e., to allow single parts or machines to make decisions. Modelling of some simple scenarios with autonomous control by means of differential equations have been given in (Dachkovski et al., 2005), (Scholz-Reiter et al., 2007b). Due to globalization many enterprizes begin to influence each other, to cooperate or merge together. Large logistic networks appear in this way. The question arises, under which conditions such networks are stable. Since the behavior of logistic processes is often nonlinear a suitable stability notion is the input-to-state stability (ISS) defined below. In this paper we will give a stability criterion for such logistic processes modelled by differential equations. Complex logistic processes are often described and modelled with help of graphs. A nodes of a graph may represent a processing machine or a plant in case of production logistics as well as a traffic junctions or a warehouse in case of transport logistics. The edges of the graph describe the connections or relations between the nodes. The stability property is very important for the design of logistic networks. This property ensures that the state of the whole system remains bounded for bounded external inputs. There are several stability criteria of a small gain type in the literature in

H.-D. Haasis et al., *Dynamics in Logistics*

DOI: 10.1007/978-3-540-76862-3,

the last decade (Jiang et al., 1994), (Jiang et al., 1996), (Dashkovskiy et al., 2007a). Here we going to show how such theorems can be applied for the investigation of stability properties of logistic networks with autonomous control. For this purpose we consider two examples to show the essence and the method of usage of the small gain criteria. Some simulation results will be also presented.

2 Motivating example

Let us consider a network of car production. The first node Σ_1 receives the orders of customers. Depending on the customer wishes the order is then forwarded to a plant Σ_4 where the car will be completed or the preparatory steps in Σ_2 and/or Σ_3 are needed. The car is then supplied to one of the distribution centers Σ_5 or Σ_6. The corresponding graph of the network is given in Fig. 1. The state of a plant is often model by differential equations (Armbruster et al., 2006), (Helbing et al., 2004). Let the following system on describe the evolution of states of the nodes.

$$\dot{x}_1 = u - \frac{ax_1 + b\sqrt{x_1}}{1 + x_2 + x_3} \tag{1}$$

$$\dot{x}_2 = \frac{1}{3}\frac{ax_1 + b\sqrt{x_1}}{1 + x_2 + x_3} + \frac{1}{2}\min\{b_3, c_3x_3\} - \min\{b_2, c_2x_2\} \tag{2}$$

$$\dot{x}_3 = \frac{1}{3}\frac{ax_1 + b\sqrt{x_1}}{1 + x_2 + x_3} + \frac{1}{2}\min\{b_2, c_2x_2\} - \min\{b_3, c_3x_3\} \tag{3}$$

$$\dot{x}_4 = \frac{1}{3}\frac{ax_1 + b\sqrt{x_1}}{1 + x_2 + x_3} + \frac{1}{2}\min\{b_2, c_2x_2\} + \min\{b_3, c_3x_3\} - \min\{b_4, c_4x_4\} \tag{4}$$

$$\dot{x}_5 = \frac{1}{2}\min\{b_4, c_4x_4\} - c_5x_5 \tag{5}$$

$$\dot{x}_6 = \frac{1}{2}\min\{b_4, c_4x_4\} - c_6x_6 \tag{6}$$

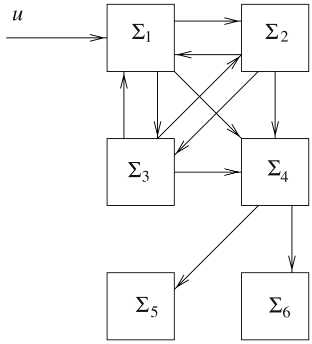


Fig. 1 Graph structure of the logistic network

One can check that each subsystem is ISS using ISS Lyapunov functions $V_i(x_i) = x_i$, see (Dashkovskiy et al., 2006) and (Dashkovskiy et al., 2007b), where the gains are also found. However the interconnection of these systems is not always stable. The small gain condition can be applied for these systems. It imposes certain restrictions on the constants a, b_i, c_i in the equations. See (Dashkovskiy et al., 2007b) for details. A numerical procedure was developed there to check the small gain condition locally. The procedure was applied for the stability investigation of the system (1–6). The motivation for influx and outflux terms is similar as in the example with two nodes, which we consider in detail in the next section.

3 Feedback loop as a two nodes network

We consider two stations processing parts coming from outside to the first station and forwarded to the second one after being processed there. After the processing on the second station the parts leave the network. The arrival process of parts is described in terms of the time varying arrival rate function $u = u(t)$ which can be considered as a result of another autonomous process. There are queues of length $x_1(t)$ and $x_2(t)$ in front of each station. The stations are considered to be self-controlled, i.e., able to increase or decrease their processing rates depending on certain circumstances, as for example current arrival rate, own queue length or the queue length of the neighbor. Such a system can be modelled with help of differential equations. Let be $b_1 = b_1(t, u, x_1, x_2)$ and $b_2 = b_2(t, u, x_1, x_2)$ the processing rate of the corresponding station. Then the current state of the system, i.e., the queue lengths, is then given as the solution of the following differential equations

$$\dot{x}_1 = u - b_1, \tag{7}$$

$$\dot{x}_2 = b_1 - b_2. \tag{8}$$

For our limited purpose it is enough to consider such a feedback loop. However it can be considered as a part of a more complex network.

3.1 Interpretations

Here we give some possible interpretations of the scenario. Consider a container ship being offloaded at a container terminal. The autonomous containers on the ship arrive to the offloading facility (the first station) in a certain rate u determined by their internal rules. The offloading facility consists of several cranes, each of them has several operating regimes. For example if the queue is short not all of them need to be active. If the queue becomes longer they are able to switch themselves to a faster regime. A desired processing rate can be achieved in this way. After offloading a container arrives to the customs clearance (the second station) and waits there to be processed. Since the space for this second queue is limited the offloading facility takes into account the length of the second queue and its processing rate can be taken in the form

$$b_1 = \frac{ax_1 + b\sqrt{x_1}}{1+x_2}, \tag{9}$$

where a and b are positive constants. For vanishing x_2, i.e., for empty queue at customs, the processing rate is essentially proportional to x_1 if $x_1 \gg 1$ and to $\sqrt{x_1}$ if $x_1 \ll 1$. With growth of the second queue x_2 this processing rate is reduced by the factor $(1+x_2)$. The customs has several gates for container clearance with varying number of service personal. It uses only one gate if its queue is short and opens further gates when the queue grows. Let the processing rate of customs be proportional to the length of its own queue, i.e.,

$$b_2 = cx_2, \tag{10}$$

with a positive constant c. In this case the system (7–8) becomes

$$\dot{x}_1 = u - \frac{ax_1 + b\sqrt{x_1}}{1+x_2} \tag{11}$$

$$\dot{x}_2 = \frac{ax_1 + b\sqrt{x_1}}{1+x_2} - cx_2\,. \tag{12}$$

3.2 State equation and stability of the queues

The state equations (11–12) describe the evolution of the queues in dependence of the inflow u and the initial conditions. We assume that each subsystem is ISS. This means that there exists an ISS-Lyapunov function V_1 for each of them, i.e., for some $\psi_{11}, \psi_{12}, \psi_{21}, \psi_{22}, \gamma$ of class $\mathcal{K}_\infty, \chi_1, \chi_2$ of class $\mathcal{K}$ and α_1, α_2 positive-definite functions it holds

$$\psi_{1i}(|x_1|) \le V_i(x_1) \le \psi_{2i}(|x_1|), \quad \forall x_1 \in \mathbb{R}^n,\, i = 1,2, \tag{13}$$

$$|x_1| \ge \max\{\chi_1(|x_2|), \gamma(u)\} \Rightarrow \frac{dV_1}{dx_1}(x_1)\left(u - \frac{ax_1 + \sqrt{x_1}}{1+x_2}\right) \le -\alpha_1(|x_1|) \tag{14}$$

$$|x_2| \ge \chi_2(|x_1|) \Rightarrow \frac{dV_2}{dx_2}(x_2)\left(\frac{ax_1 + \sqrt{x_1}}{1+x_2} - cx_2\right) \le -\alpha_2(|x_2|) \tag{15}$$

Remark 1. Note than the solution of (11) is nonnegative for any initial condition $x_1(0) = x_1^0 \ge 0$. This follows from $u \ge 0$ and any time when x_1 reaches zero $\dot{x}_1 = u \ge 0$ holds true. In the same way the solution of (12) is also nonnegative for any initial condition $x_2(0) = x_2^0 \ge 0$. This is in agrement with the notion of queue length, which cannot be negative. Hence in the following we will write x_1 and x_2 instead of $|x_1|$ and $|x_2|$ respectively.

The small gain theorem in this case reads as follows

Theorem 1. *Let* V_1 *and* V_2 *satisfy* (13)–(15) *and assume that*

$$\chi_1 \circ \chi_2(s) < s, \tag{16}$$

then the interconnection (11)–(12) *is ISS with an ISS-Lyapunov function given by*

$$V(x_1, x_2) = \max\{\sigma(V_1(x_1)), V_2(x_2)\}, \tag{17}$$

where σ *is a smooth class* $\mathcal{K}_\infty$ *function with*

$$\chi_2 < \sigma < \chi_1^{-1}.$$

We consider $V_1(x_1) = |x_1|$ and $V_2(x_2) = |x_2|$. These functions are not smooth in zero, however one can change both in an arbitrary small neighborhood of zero to be smooth. We will not go onto this technical details and will work with them as with smooth functions. For nonnegative arguments we have then

$$\frac{dV_1(x_1)}{dx_1} = \frac{dV_2(x_2)}{dx_2} = 1.$$

Let $\gamma(u) := u^2/c_u^2$ and $\chi_1(x_2) := x_2^2/c_1^2$, then from $x_1 > \chi_1(x_2)$ and $x_1 > \gamma(u)$ follows $u < \gamma^{-1}(x_1) = c_u\sqrt{x_1}$ and $x_2 < \chi_1^{-1}(x_1) = c_1\sqrt{x_1}$, i.e.,

$$\begin{aligned} u - \frac{ax_1 + b\sqrt{x_1}}{1 + x_2} &< \gamma^{-1}(x_1) - \frac{ax_1 + b\sqrt{x_1}}{1 + \chi_1^{-1}(x_1)} \\ &= \frac{c_u\sqrt{x_1} + c_u\sqrt{x_1}c_1\sqrt{x_1} - ax_1 - b\sqrt{x_1}}{1 + c_1\sqrt{x_1}} \\ &= -\frac{(a - c_u c_1)x_1 + (b - c_u)\sqrt{x_1}}{1 + c_1\sqrt{x_1}} \end{aligned}$$

and $\frac{(a-c_u c_1)x_1+(b-c_u)\sqrt{x_1}}{1+c_1\sqrt{x_1}} =: \alpha_1(x_1)$ is a positive-definite function if $a > c_u c_1$ and $b > c_u$. This shows that (13) and (14) holds true, i.e., V_1 is an ISS-Lyapunov function and the Eq. (11) is ISS. Now consider Eq. (12) and let $\chi_2(x_2) := \sqrt{x_2}/\sqrt{c_2}$, then from $x_2 > \chi_2(x_1)$ follows $x_1 < \chi^{-1}(x_2) = c_2 x_2^2$ and

$$\begin{aligned} \frac{ax_1 + b\sqrt{x_1}}{1 + x_2} - cx_2 &< \frac{ac_2x_2^2 + b\sqrt{c_2}x_2 - cx_2 - cx_2^2}{1 + x_2} \\ &= -\frac{(c - ac_2)x_2^2 + (c - b\sqrt{c_2})x_2}{1 + x_2} \end{aligned}$$

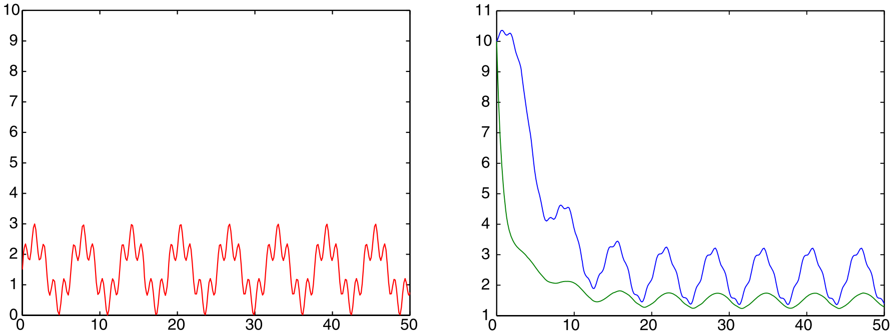


Fig. 2 Input u and the queues x_1, x_2

where $\frac{(c-ac_2)x_2^2+(c-b\sqrt{c_2})x_2}{1+x_2} =: \alpha_2(x_2)$ is a positive-definite function if $c > ac_2$ and $c > b\sqrt{c_2}$. Now we see that the small gain condition reads as

$$\chi_1 \circ \chi_2(s) < s \quad \Leftrightarrow \quad c_2 c_1^2 > 1$$

and the network is stable if for the given parameters a, b, c of the system there exist positive c_u, c_1, c_2 such that $c_u < b$, $c_1 < a/c_u$, $c_2 < c/a$, $\sqrt{c_2} < c/b$ and $c_2 c_1^2 > 1$. We can always take $c_2 < \min\{c/a, c^2/a^2\}$, $c_1 > 1/\sqrt{c_2}$ and $c_u < \min\{b, a/c_1\}$. Hence this network is input to state stable. The simulated solution of (11-12) with the input $u(t) = 1.5 + \sin(t) + \sin(5t)/2$ and the initial conditions $x_1(0) = 10, x_2(0) = 10$ is presented on the Fig. 2. We see that the solution remains bounded in this case.

4 Conclusions

The most important for the logistics is the knowledge of the long term behavior of the state of a system. Hence stability of a logistic networks is an important property which guarantees that the state remains bounded. A general criterion for the input-to-state stability was proposed. We have discussed how it can be applied in some special cases. In general it yields some restriction on the parameters of a system and can be checked numerically.

Acknowledgements This research was supported by the German Research Foundation (DFG) as part of the Collaborative Research Center 637 "Autonomous Cooperating Logistic Processes".

References

Armbruster D, Marthaler DE, Ringhofer C, Kempf K, Jo T (2006) A continuum model for a re-entrant factory. Operations Research 54(5):933–950

Dachkovski 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, NL

Dashkovskiy S, Rüffer B, Wirth F (2005) A small-gain type stability criterion for large scale networks of ISS systems. Proc of 44th IEEE Conference on Decision and Control and European Control Conference ECC 2005, pp 5633–5638

Dashkovskiy S, Rüffer BS, Wirth FR (2006) An ISS Lyapunov function for networks of ISS systems. In: Proceedings of the 17th International Symposium on Mathematical Theory of Networks and Systems (MTNS), Kyoto, Japan, pp 77–82

Dashkovskiy S, Rüffer B, Wirth F (2007a) An ISS small-gain theorem for general networks. Mathematics of Control, Signals, and Systems 19:93–122

Dashkovskiy S, Rüffer B, Wirth F (2007b) Numerical verification of local input-to-state stability for large networks. In: Proc. 46th IEEE Conf. Decision and Control, CDC 2007, IEEE, New Orleans, LA, submitted

Helbing D, Lämmer S, Witt U, Brenner T (2004) Network-induced oscillatory behavior in material flow networks and irregular business cycles. Physical Review E 70(056118):6 pages

Jiang ZP, Teel AR, Praly L (1994) Small-gain theorem for ISS systems and applications. Math Control Signals Systems 7(2):95–120

Jiang ZP, Mareels IMY, Wang Y (1996) A Lyapunov formulation of the nonlinear small-gain theorem for interconnected ISS systems. Automatica J IFAC 32(8):1211–1215

Scholz-Reiter B, Jagalski T, de Beer C (2007a) Selbststeuerung logistischer Prozesse in Produktionsnetzen. Industrie Management 23(1):19–22

Scholz-Reiter B, Wirth F, Freitag M, Dashkovskiy S, Jagalski T, de Beer C, Rüffer B (2007b) Mathematical models of autonomous logistic processes. In: Hülsmann M, Windt K (eds) Understanding Autonomous Cooperation & Control in Logistics: The Impact on Management, Information and Communication and Material Flow, Springer, Berlin

Sontag ED (1989) Smooth stabilization implies coprime factorization. IEEE Trans Automat Control 34(4):435–443

Appendix: Definitions and known results

Let a logistic network be given. Consider equation describing dynamics of the state $x_i \in \mathbb{R}^{N_i}$ of the i-th node depending on the input $u_i \in \mathbb{R}^{M_i}$ and states of other nodes x_j, $j \neq i$

$$\dot{x}_i = f_i(x_1, \ldots, x_n, u_i), \quad i = 1, \ldots, n \tag{18}$$

with $f_i : \mathbb{R}^{\sum_j N_j + M_i} \to \mathbb{R}^{N_i}$, $i = 1, \ldots, n$. The system of these equations can be written in the form

$$\dot{x} = f(x,u), \tag{19}$$

$x^T = (x_1^T, \ldots, x_n^T) \in \mathbb{R}^N$, $N = \sum_{i=1}^n N_i$, $u^T = (u_1^T, \ldots, u_n^T)$, $f(x,u)^T = (f_1(x,u_1)^T, \ldots, f_n(x,u_n)^T)$. Let $\mathbb{R}_+$ denote the interval $[0,\infty)$ and $\mathbb{R}_+^n$ be the positive orthant in $\mathbb{R}^n$. For any $a,b \in \mathbb{R}_+^n$ let $a < b$ and $a \leq b$ mean $a_i < b_i$ and $a_i \leq b_i$ for all $i = 1, \ldots, n$ respectively. Recall the definition of comparison functions.

Definition 1. (i) A function $\gamma : \mathbb{R}_+ \to \mathbb{R}_+$ is said to be of class $\mathcal{K}$ if it is continuous, increasing and $\gamma(0) = 0$. It is of class $\mathcal{K}_\infty$ if, in addition, it is unbounded. (ii) A function $\beta : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$ is said to be of class $\mathcal{KL}$ if, for each fixed t, the function $\beta(\cdot, t)$ is of class $\mathcal{K}$ and, for each fixed s, the function $\beta(s, \cdot)$ is non-increasing and tends to zero at infinity.

The concept of input-to-state stability (ISS) has been first introduced in (Sontag, 1989).

Definition 2. System (19) is *input-to-state stable*, if there exists a $\gamma \in \mathcal{K}_\infty$, and a $\beta \in \mathcal{KL}$, such that for all $\xi \in \mathbb{R}^n_+$, $u \in L_\infty$

$$\|x(t,\xi,u)\| \leq \beta(\|\xi\|,t) + \gamma(\|u\|_\infty) \quad \forall\, t \leq 0, \tag{20}$$

in this case γ is called gain.

It is known that ISS defined in this way is equivalent to the existence of an ISS-Lyapunov function.

Definition 3. A smooth function $V : \mathbb{R}^N \to \mathbb{R}_+$ is an *ISS-Lyapunov function* of (19) if there exist $\psi_1, \psi_2 \in \mathcal{K}_\infty$, $\chi \in \mathcal{K}_\infty$, and a positive definite function α such that

$$\psi_1(|x|) \leq V(x) \leq \psi_2(|x|), \quad \forall x \in \mathbb{R}^N, \tag{21}$$

$$V(x) \geq \chi(|u|) \Longrightarrow \nabla V(x) \cdot f(x,u) \leq -\alpha(V(x)), \tag{22}$$

for all $\xi \in \mathbb{R}^n_+$, $u \in L_\infty$. Function χ in then called Lyapunov-gain.

Subsystem (18) is ISS, provided there exist $\gamma_{ij}, \gamma_i \in \mathcal{K}_\infty$, and a $\beta_i \in \mathcal{KL}$, such that for all $\xi_i \in \mathbb{R}^n_+$, $u_i \in L_\infty, \forall\, t \geq 0$,

$$\|x_i(t,\xi_i,x_j : j \neq i, u_i)\| \leq \max\{\beta_i(\|\xi_i\|,t), \max_{j \neq i} \gamma_{ij}(\|x_j\|_\infty),\, \gamma_i(\|u_i\|_\infty)\} \tag{23}$$

If all n subsystems are ISS then these estimates give rise to a *gain matrix*

$$\Gamma = (\gamma_{ij})^n_{i,j=1}, \quad \text{with} \quad \gamma_{ij} \in \mathcal{K}_\infty \quad \text{or} \quad \gamma_{ij} \equiv 0, \tag{24}$$

where we use the convention $\gamma_{ii} \equiv 0$ for $i = 1, \ldots, n$. The gain matrix Γ defines a monotone operator $\Gamma : \mathbb{R}^n_+ \to \mathbb{R}^n_+$ by $\Gamma(s)_i := \max_j \gamma_{ij}(s_j)$ for $s \in \mathbb{R}^n_+$. The global small gain condition assuring the ISS property for an interconnection of ISS subsystems was derived in (Dashkovskiy et al., 2005). An alternative proof has been given in (Dashkovskiy et al., 2007a). We quote the following result from these papers.

Theorem 2 (global small-gain theorem for networks). *Consider system* (19) *and suppose that each subsystem* (18) *is ISS, i.e., condition* (23) *holds for all* $\xi_i \in \mathbb{R}^n_+$, $u_i \in L_\infty$, $i = 1, \ldots, n$. *Let* Γ *be given by* (24). *If there exists an* $\alpha \in \mathcal{K}_\infty$, *such that*

$$(\Gamma \circ D)(s) \not\geq s, \qquad \forall s \in \mathbb{R}^n_+ \setminus \{0\}, \tag{25}$$

with $D = \operatorname{diag}_n(\mathrm{id} + \alpha)$ *then the system* (19) *is ISS from* u *to* x.

Next Generation Supply Chain Concepts

Web-service Based Integration of Multi-organizational Logistic Process

Hyerim Bae

Pusan National University, San 30, Jangjeon Dong, Guemjoeng Gu, Pusan, Korea
e-mail: hrbae@pusan.ac.kr

Abstract Recent business environments of keen competition require companies to communicate frequently with each other. This makes a company deal with more complex processes. Business Process Management (BPM) system plays an important role for the management of such complex processes. Recently, it moves their focus onto multi-organizational collaboration, and a platform independent architecture for the communication is required. Web services are platform-independent and service-oriented architectures for easy communication among heterogeneous systems. In this paper, we propose a new method of integrating BPM and web services to manage multi-organizational processes. A standard, Business Process Execution Language (BPEL) plays a key role in our approach. We consider several scenarios for integrating web service and BPM using the BPEL when a company performs a complex process with other partners. We also provide technical issues of the integrated system. A BPM system is implemented to present the result of our research. Our approach enables each business partner to carry out business collaboration seamlessly.

1 Introduction

In recent business environments, business processes get increasingly complex, and the competition among companies gets keener. Companies in such environments adopt various kinds of information systems such as ERP (Enterprise Resource Management), CRM (Customer Relationship Management), and SCM (Supply Chain Management). Though the systems are able to resolve immediate problem, introduction of the heterogeneous systems generates a new problem of disharmony. The problem forms a severe obstacle for a company to interact with other companies in collaborative environments. The necessity of collaboration has doubled with improvement of network technology, and interoperation of heterogeneous systems becomes one of the most critical issues in the area of information system. As internet based information systems are more developed, new paradigms such as e-Business,

H.-D. Haasis et al., *Dynamics in Logistics*

DOI: 10.1007/978-3-540-76862-3,

electronic commerce, and collaborative commerce are emerging. Since they bring about frequent interaction among companies, the ‘integration’ is a key word for effective communications.

For the interoperability among enterprise information systems, XML (eXtensible Markup Language) technology is considered the practical solution. XML is not only a standard of internet documentation but also a standard means of data exchange and communication among systems. XML-based interoperation among different companies is extending its areas, and gives rise to XML-based standard framework such as ebXML and RosettaNet. Web services, which have been hailed as a new standard for distributed computing, are also based on XML. They provide service-oriented architectures which enable universal interoperability. BPEL4WS (Business Process Execution Language for Web Services) [1], which is a standard for process modeling language of web services, provides an executable process model by composing a set of existing services.

In this paper, we provide methods of integrating multi-organizational workflows by using BPEL4WS. The method enables collaborative commerce in an e-business environment by integrating Business Process Management (BPM) and web services. We develop a BPM system support the method, which shows effectiveness of our approach.

2 Backgrounds

2.1 Workflow interoperability

Interoperability among different information systems is considered an important issue in an environment of keen competence and high speed. Although the necessity of collaboration is increasing, heterogeneity of information systems adapted can be a kind of restriction for the interoperability. When interoperation with rare intervention of persons is needed, it is common for multiple companies to carry out a global process. A workflow is a computer-readable representation of a business process, which can be useful for the interoperability in such an environment. Some research has dealt with workflow interoperability among different companies and most of them focus on software architecture [7],[8]. For better interoperability, an agreement on process definition and exchange is required and we, in this paper, provide the method.

WfMC (Workflow Management coalition), which is an international standard organization on workflow, defines workflow interoperability as the ability of two or more workflow management systems to communicate and interoperate in order to coordinate and execute workflow process instances across those systems [9]. WfMC also provides definition of an interface that specifies a set of API’s and methods of message exchange. An interoperable architecture between different workflows by WfMC is presented in Fig. 1. Each company deals with its own inner process. At

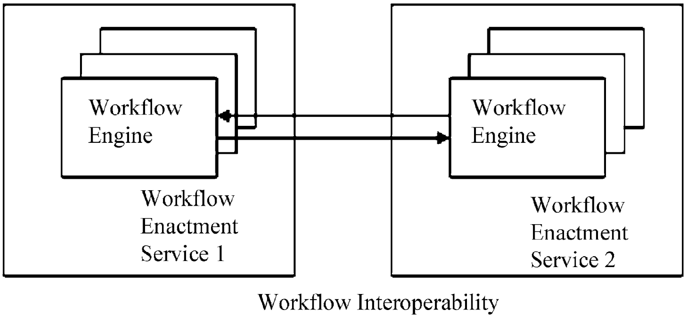


Fig. 1 Architecture of workflow interoperability

the same time, it participates in a higher level process. That is, a number of inner processes compose an inter-organizational process.

2.2 XML and interoperability

We use XML technology as a message exchanging format. XML is recently getting attention as a new standard of data exchange since it can be used independent of computing platforms. XML is able to not only overcome SGML's excessive com-

Table 1 XML based interoperability standards

Standard	Brief description
Wf-XML	Wf-XML is WfMC's message encoding standard specification based on XML. The specification is described on the basis of workflow system architecture called Workflow Reference Model (WRM), and does not provide binding SOAP and WSDL.
WSFL	IBM's WSFL specifies two types of Web services composition. One is an executable business process known as flowModel, and the other is business collaboration known as globalModel. WSFL is interoperable with SOAP, UDDI, and WSDL.
XLANG	Microsoft's XLANG is a business process modeling language for BizTalk, and support Enterprise Application Integration (EAI). It provides BizTalk Orchestration and BizTalk Orchestration Designer which are a workflow engine and XLANG based visual process modeling tool respectively.
BPEL4WS	BPEL4WS is a merge of WSFL and XLANG for Web services' orchestration and workflow composition.
ebXML BPSS	BPSS of eBusiness Transition Working Group defines workflow conversation and orchestration
WSCI	WSCI by a consortium of Sun, BEA, Intalio and SAP is an XML based interface description language, which specifies message flows by Web services.
WSCL	W3C's WSCL is proposed by Hewlett-Packard, which defines abstract interface, XML document exchange, and sequence of documents of Web services.
PIPs	RosettaNet's Partner Interface Process (PIP) defines business processes among trade partners using XML-based communications.

plexity and HTML's restriction of expressive power but also construct a document effectively. Due to the merits, XML is used to acquire interoperability among heterogeneous information systems going far beyond merely to browse data. We can conclude that XML is the only and the most optimal solution for workflow interoperability.

The usage of XML in exchanging data can be viewed in two ways. The first one is standard e-commerce frameworks such as ebXML and RessetaNet. The frameworks define standard specifications related to processes of establishing partnership between different organizations and carrying out their business collaboratively. Data exchange during the process is entirely based on XML. The other is web service. Web services provide interface that enables interaction based on the XML related standards such as XML, SOAP (Simple Object Access Protocol), WSDL (Web Service Description Language), and UDDI (Universal Description, Discovery and Integration).

Up to date, interoperability standards using XML are introduced increasingly, and major ones are summarized in Table 1.

3 Web service and BPEL4WS

3.1 Web service

Web services, in a broad concept, are services provided on the Web. Since this broad concept even include simple browsing, it cannot be considered proper concept in this paper. Web services, in this paper, are communication frameworks where applications are serviced via internet based on XML technologies. Web services settle down as a new paradigm of information system because of it merits such as automatic search and discovery, existence of process description standards, XML based protocol, and plentiful support of vendors. Internet users accessing the Web using diverse channels require receiving services beyond simple search. Internet banking, shopping, and dealing in stocks are typical services that can be provided on the Web.

3.2 Process-oriented web service integration and BPEL4WS

Integration of information systems is a problem that has been presented since the advent of interests on interaction among companies. One of conspicuous trends is the adoption of process-driven application integration, which is developed from data-driven integration. Companies participating in an e-Business with their own inner process collaboratively pursue seamless execution of a global process. In this sense, a standard language that describes a complex process by composition of web services is required. Such standards are BPEL4WS and WSCI. In this paper, we consider inter organizational workflow based the BPEL4WS.

BPEL4WS is proposed to describe a business process based on Web service. It is suggested by BEA Systems, IBM, and Microsoft in August, 2002 to increase the utilization of Web services. A process specified by BPEL4WS uses only Web service interface to import and export its functions. It incorporates and extends the functions of both IBM's WSFL and Microsoft's XLANG, which is expected to replace the two.

A business process description language, BPEL4WS with the help of Web service communication protocols, WS-Coordination, WS-Transaction enables reliable integration of inter organizational workflows and business tractions.

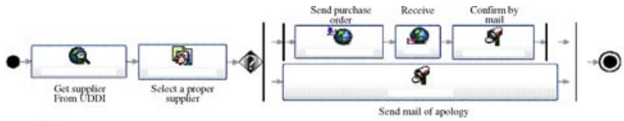
4 Workflow integration using BPEL4WS

4.1 Using BPEL4WS as a process definition language

Since BPEL4WS basically supports an executable mode of business processes, we can say that it provides modeling semantics required to describe enterprise business processes. The first and simplest way of utilizing BPEL4WS is to use it as a process definition language. Wohed shows that BPEL4WS is able to model a wide variety of workflow patterns [9]. In Table 2, several process definition languages are compared with BPEL4WS in their expressive power. We can conclude that BPEL4WS has as

Table 2 Comparison of BPEL4WS with process definition languages [10]

Pattern	BPEL	XLANG	WSFL	Staffw.	MQS.
Sequence	+	+	+	+	+
Parallel Split	+	+	+	+	+
Synchronization	+	+	+	+	+
Exclusive Choice	+	+	+	+	+
Simple Merge	+	+	+	+	+
Multiple Choice	+	−	+	−	+
Synchronizing Merge	+	−	+	−	+
Multi Merge	−	−	−	−	−
Discriminator	−	−	−	−	−
Arbitrary Cycles	−	−	−	+	
Implicit Termination	+	−	+	+	+
MI without Synchronization	+	+	+		+
MI with a Priori Design Time Knowledge	+	+	+	+	+
MI with a Priori Run Time Knowledge	−	−	−	−	−
MI without a Priori Run Time Knowledge	−	−	−	−	−
Deferred Choice	+	+	−	−	−
Interleaved Parallel Routing	+/−	−	−	−	−
Milestone	−	−	−	−	−
Cancel Activity	+	+	+	+	−
Cancel Case	+	+	+	−	−
Request/Reply	+	+	+	−	−
One−way	+	+	+	−	−
Synchronous Polling	+	+	+	−	−
Message Passing	+	+	+	−	−
Publish/Subscribe	−	−	−	−	−
Broadcasting	−	−	−	−	−



a) a process model example


```
<?xml version="1.0"?>
<process xmlns:prefix_0="http://schemas.xmlsoap.org/ws/2003/03/business-
process/"
xmlns:prefix_1="namespace"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
name="purchase_proc_v4" targetNamespace=
"urn:purchase_porc_v4">
<partnerLinks>
<partnerLink name="mailServer" partnerLinkType="prefix:plnktype" part-
nerRole="partnerRole">
</partnerLink>
<partnerLink name="broker" partnerLinkType="prefix:plnktype" part-
nerRole="partnerRole"></partnerLink>
<partnerLink name="suppliler" partnerLinkType="prefix:plnktype" part-
nerRole="partnerRole"></partnerLink>
<partnerLink name="procurement_officier" partnerLinkType="prefix:plnktype"
partnerRole="partnerRole"></partnerLink>
<partnerLink name="client" partnerLinkType="prefix:plnktype" partnerRo-
le="partnerRole"></partnerLink>
<partnerLink name="messengerServer" partnerLinkType="prefix:plnktype"
partnerRole="partnerRole"></partnerLink>
<partnerLink name="Worklist_Service" partnerLinkType="prefix:plnktype" par-
tnerRole="partnerRole"></
partnerLink>
</partnerLinks>
<variables>
<variable name="purchaseOrder" type="prefix:com.supplier.PurchaseOrder">
</variable><variable name="catalogs"
type="prefix:com.supplier.ArrayOfCatalog"></variable>
</variables>
<sequence name="_main_sequence">
<receive partnerLink="self" portType="prefix:onHumanActivityResult_1" op-
eration="self" variable="purchaseOrder"
name="Select_proper_supplier"></receive>
<switch_name="Switch">
<case_condition="boolean exp">
<sequence name="seq">
<invoke partnerLink="supplier" portType="prefix:SupplierService" opera-
tion="order" outputVariable="_dummyVariable"
name="Send_Purchase_Order"></invoke>
<receive partnerLink="self" portType="prefix:Receive_order_result" opera-
tion="self" name="Receive"></receive>
<invoke partnerLink="mailServer" portType="prefix:EmailService" opera-
tion="sendMail" outputVariable="dummyVariable"
name="Confirmation_by_mail"></invoke>
</sequence>
</case_>
<case_condition="Boolean exp">
<invoke partnerLink="mailServer" portType="prefix:EmailService" opera-
tion="sendMail" outputVariable="_dummyVariable"
name="send_mail_of_apology"></invoke>
</case_>
</switch_>
</sequence>
</process>
```


b) BPEL presentation

Fig. 2a,b Example BPEL4WS model for a process much expressive power as commercial definition languages, and thus it is enough to use the standard as a process definition. Figure 2 presents an example BPEL4WS model of a workflow process example. After a user designs a process like Fig. 2a), our system generates BPEL4WS codes shown in Fig. 2b)

4.2 Using BPEL4WS as a process exchange format

BPEL4WS can be utilized as a standard format of business process exchange. One of problems raised by traditional methods of integrating business processes is heterogeneity of adopted systems. Each system has its own computing platform and defines its business processes in a different way. Consequently, it has been difficult to exchange, share and execute inter-enterprise processes.

Since BPEL4WS, which is based on XML, describes business processes platform independently, it can be a method of exchanging business processes. In order for a workflow system to support the process exchange, it needs to develop and install a transformation tool that can export its own processes into BPEL4WS.

To use BPEL4WS for business process exchanging, we propose architecture of inter-organizational business interaction in Fig. 3. In the figure, two companies carry out a common process and each of the two is in charge of parts of the process. They use their own BPM system to automate inner workflow process. During one of the two execute its process; if it needs to hand over a process model, the model has to be transformed to another format. The model is exported in the format of BPEL4WS language.

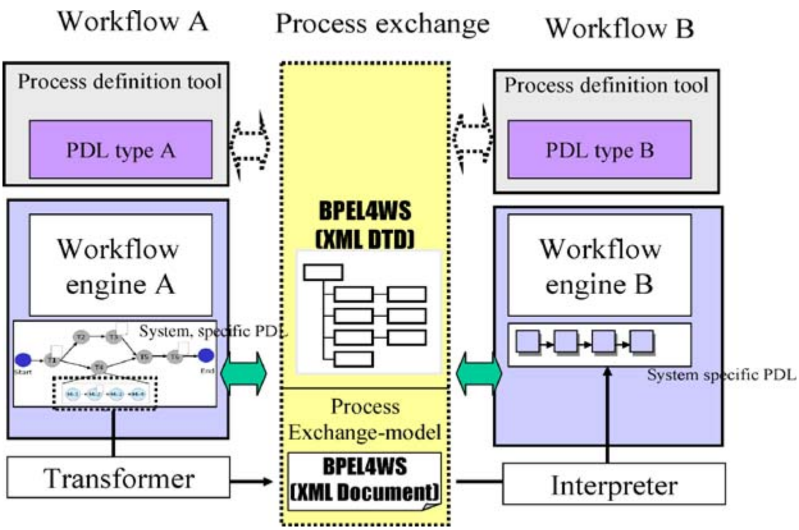


Fig. 3 Process exchange using BPEL4WS

4.3 Workflow as Web service

According to WfMC's standard, a workflow means computerized presentation of a business process that can be automated. On the other hand, Web services aims at providing new services by composition of automated enterprise services. It is not extraordinary that business processes executed in a company are considered services by outer companies. Then, outer companies use APIs predefined and provided by the company. Therefore, workflow processes in a company can be serviced as Web services based on the workflow interoperability standards. Figure 4 shows the architecture of workflow interfaces proposed by WfMC. To make workflow processes into Web services, the interface 4 in the above needs to be used.

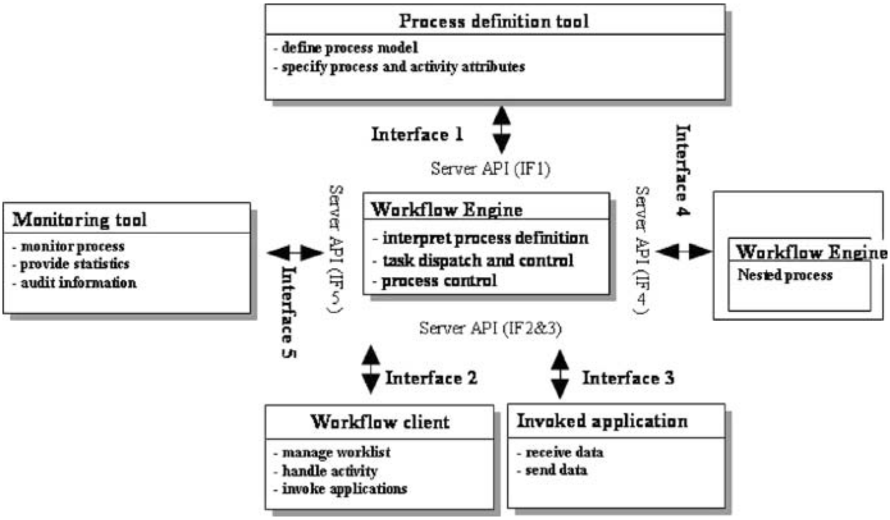


Fig. 4 Workflow components [5]

4.4 Workflow as Web services' coordinator

Recent BPM is an extended concept of workflow, and support collaborative processes helping SCM (Supply Chain Management). A collaborative process is composed of several nested sub-processes, and the execution of the process needs to be managed as whole [2],[3],[4]. Thus, each company executes its process, and at the same time, the nested process can be considered nested in a global process. Sub-processes can be served via Web services. Then the execution of the global process requires some coordinator in charge of executing. We utilize BPM as the coordinator, which should be independent of platforms adopted by companies. For BPM to manage global processes, it has to describe processes in the manners enabling flexible integration of heterogeneous systems. BPM system using BPEL4WS can be a good coordinator of Web services.

5 System implementation: uEngine

In this paper, we propose an approach to integrate multi-organizational workflows based on BPEL4WS. We show that the approach can achieve the goal of the integration using BPEL4WS based on the four scenarios described in Sect. 3. In Fig. 5, architecture of uEngine [6], which is a BPM system to support the scenarios and integrate inter-organizational workflows, is presented.

Our system provides GUI tool for process definition, by which users can model process and transform to BPEL4WS models. The transformed process model can be exported to a partner, and BPEL4WS is used as process exchange format. The system allows Web services to be registered as workflow activities, and orchestrate the execution of the Web services. On the other hand, it also allows inner processes to be served as Web services.

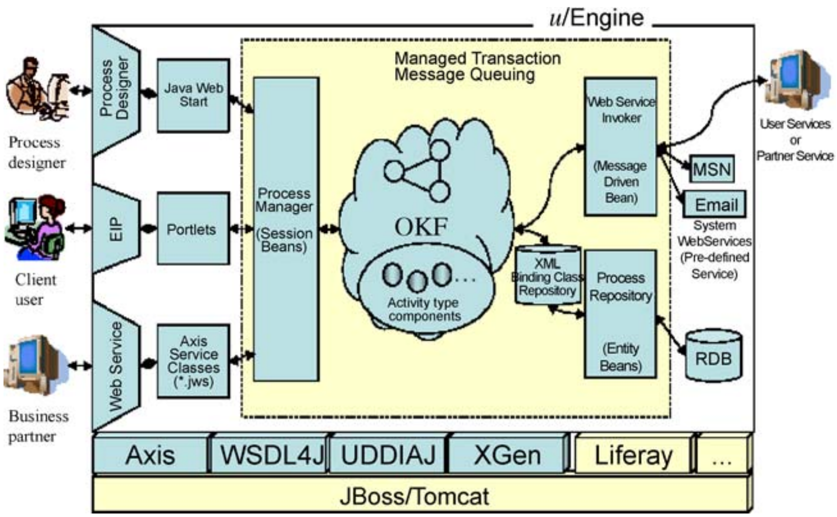
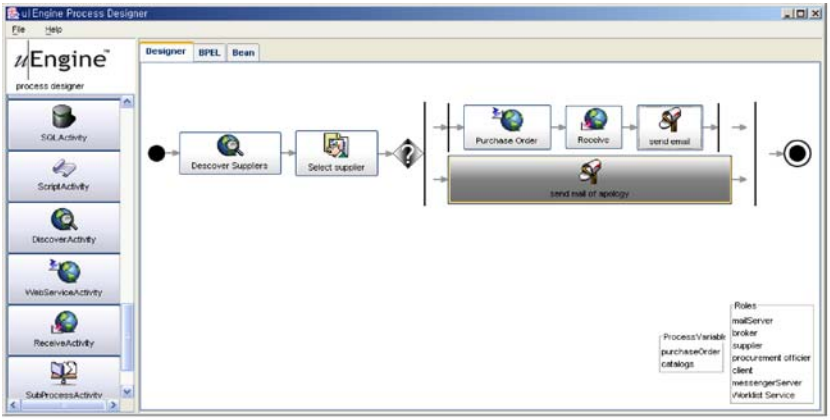
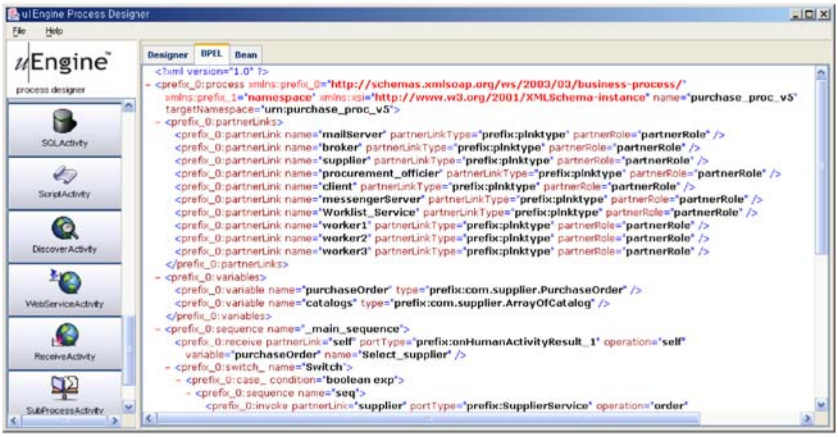


Fig. 5 Architecture of uEngine



a) Example process model



b) Transformed BPEL4WS

Fig. 6 a), b) Screen capture of uEngine

In Fig. 6, a) is a screen capture of an example process model, where a Web service is modeled as a workflow activity, and b) shows the transformed BPEL4WS for the model of a).

6 Conclusions

In this paper, we provide four ways that enable to integrate inter-organizational workflows using BPEL4WS, a standard process execution language. First way is to use BPEL4WS as a process definition language. Second, while each workflow in a global process has its own process definition language, BPEL4WS is utilized as exchanging standard format. Third, a workflow process described in a format of BPEL4WS is served as a Web service. Finally, BPM system plays a role of coordinator of Web services, and manages the execution of a global process consisting of Web services. We expect that the proposed approach contribute to solve a problem of interoperability among heterogeneous information systems. Companies participating in a global collaborative process can carry out their tasks with easy interactions.

References

1. T. Andrews et al., Business Process Execution Language for Web Service, BPEL4WS v. 1.1, 2003.
2. H. Bae, "e-Business Document Management: Change Management of Business Documents for Workflow Processes," Ph. D. dissertation of Seoul National University, 2002.
3. H. Bae and Y. Kim, "A document-process association model for workflow management", Computers in Industry, vol. 47, no. 2, pp 139–154, 2002.
4. H. Bae, W. Hur, W.S. Yoo, B. Kawk, Y. Kim, and Y. Park, "Document Configuration Control Processes Captured in a Workflow", Computers in Industry, vol. 53, No. 2. pp. 117–131, 2004.
5. D. Hollingsworth, "Workflow management coalition specification: the workflow reference model", WfMC specification, 1994.
6. J. Jang, "uEngine user and developer's guide (in Korean)," http://www.uEngine.org, 2004.
7. M. Kwak, D. Han, and J. Shim, "A framework supporting dynamic workflow interoperation and enterprise application integration," Proceedings of the 35th Annual Hawaii International Conference on System Sciences, IEEE Press, 2002, pp. 3766–3775.
8. G. Shegalov, M. Gilmann and G. Weikum, "XML-enabled workflow management for e-services across heterogeneous platforms," The International Journal on Very Large Data Bases, vol. 10, no. 1, pp. 91–103, 2001.
9. WfMC-TC-1012, Workflow Standard – Interoperability Abstract Specification, Workflow Management Coalition, Winchster, United Kingdom, 1999.
10. P. Wohed, W.M.P. van der Aalst, M. Dumas, and A. H. M. ter Hofstede, Pattern Based Analysis of BPEL4WS, Technical Report FIT-TR-2002-4, QUT.

An Approach for the Integration of Data Within Complex Logistics Systems

Carl Hans, Karl A. Hribernik, Klaus-Dieter Thoben

BIBA, University of Bremen
e-mail: {han,hri,tho}@biba.uni-bremen.de

Abstract Due to a global competition today's business world changed dramatically whereas the ongoing trend towards global enterprise networks has lead to increasing customer expectations regarding the performance of logistics systems which have to be simultaneously reliable, robust and cost-effective. One the one hand reliability, robustness and cost-efficiency mainly depend on the availability of suitable ICT-solutions as well as their proper integration into the business processes. As a consequence various systems were developed in the past and are currently in use. On the other hand the seamless access to the data kept by all these different systems is a key issue in order to meet the requirements of logistics systems of today and tomorrow. In this context data integration represents a major challenge which is still unsolved.
This paper describes a proposal for a service-oriented approach towards the seamless integration of logistics data which aims at combining existing systems and standards and thus overcoming today's data and information barriers. By accepting existing standards and interfaces this requires innovative approaches addressing the semantic description of data, their transparent transformation between different representations as well as a common interface which can be provided to the various systems which are currently applied in the field of logistics.

1 Motivation

The evolution of today's business towards enterprise networks requires logistics systems which are simultaneously reliable, robust and cost-effective. In order to fulfil these needs the application of new technologies (e.g. RFID) is an absolute must. Latest initiatives in this field go beyond that by evaluating and developing a new paradigm of autonomous, cooperating logistics processes (http://www.sfb637.uni-bremen.de). Alongside the increased integration of ICT into logistics processes the

H.-D. Haasis et al., *Dynamics in Logistics*
DOI: 10.1007/978-3-540-76862-3,

seamless access to accompanying logistics data became a major issue. A general problem in this field results from the existence of various data sources and accompanying data management and support systems as well as a diversity of underlying standards and protocols.

For competitive reasons, there is usually no interest in supporting a general interface for data access from the perspective of solution providers. Additionally, as end users, logistics service providers are interested in keeping their data safe, secure and protected in order to eliminate the potential risk that their data can be accessed, interpreted and applied by their competitors to improve their process performance. Thus the huge variety of data sources is treated in an isolated manner and the potential for a further improvement of logistics processes remains unused.

As shown in Fig. 1 significant effort was and still is spent in order to achieve at least a coupling between systems of certain business partners by bridging the technological islands through specific ICT solutions. However most of these solutions will become obsolete sooner or later due to the ongoing development of the islands which is reflected by a continuously decreasing sea level as well as the highly dynamic partnerships within today's enterprise networks. Instead of developing solutions for 1:1 relationships a general solution must be found which allows a unique access to all relevant logistics data while accepting the diversity of existing systems and standards.

In addition to the physical access to data such an approach has to address their semantics towards a general, system-independent and seamless integration of logistics data in order to allow the interpretation and exchange of data between the various systems of the different partners involved in logistics.

A real integration therefore requires a general approach in order to describe data semantics as well as transformation approaches which do the conversion of data into a format the recipient will understand.

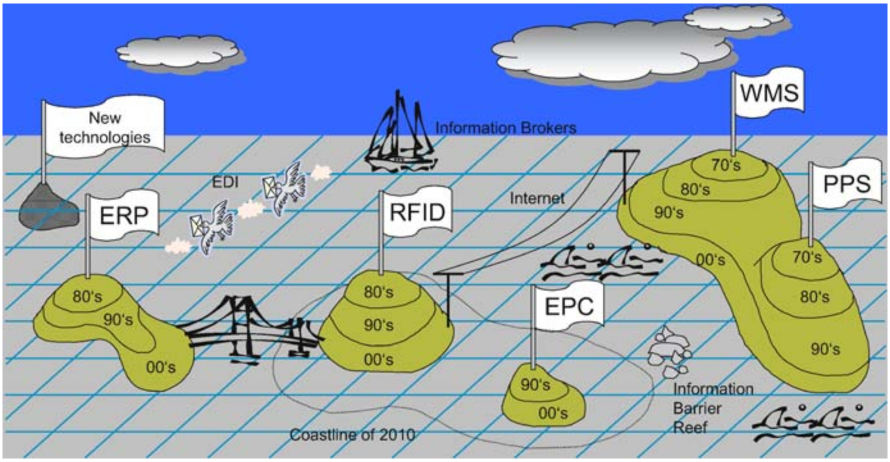


Fig. 1 Bridging Islands (based on Hannus 1996)

2 Challenge

The requirements set by complex logistics systems towards the integration of data regarding the individual entities within them prove immensely challenging. In order to implement complex behaviour with regards to autonomous control, dynamism, reactivity and mobility, these entities, including objects such as cargo, transit equipment and transportation systems but also software systems such as disposition, Enterprise Resource Planning (ERP) or Warehouse Management Systems (WMS) require the development of innovative concepts for the description of and access to data.

The availability of context and environmental data regarding logistic entities is especially relevant to the field of autonomous control. Here, real-time, local decision-making takes the place of central planning and control. It is thus imperative for the individual logistic entity to be able to access, understand, process and thus integrate context data stemming from a host of diverse, mobile and sometimes unreliable data sources, such as RFID readers, sensor networks, embedded systems, etc. This scenario sets the problem at hand apart from traditional data integration questions which generally deal with static, predictable and reliable data sources such as enterprise or database systems.

In this context, heterogeneous environments such as those found in complex logistics systems demand a unique perspective be taken upon the representation and meaning of data. Here, the heterogeneity and distribution of the syntactic and semantic representations and descriptions of the individual entities within such sys-

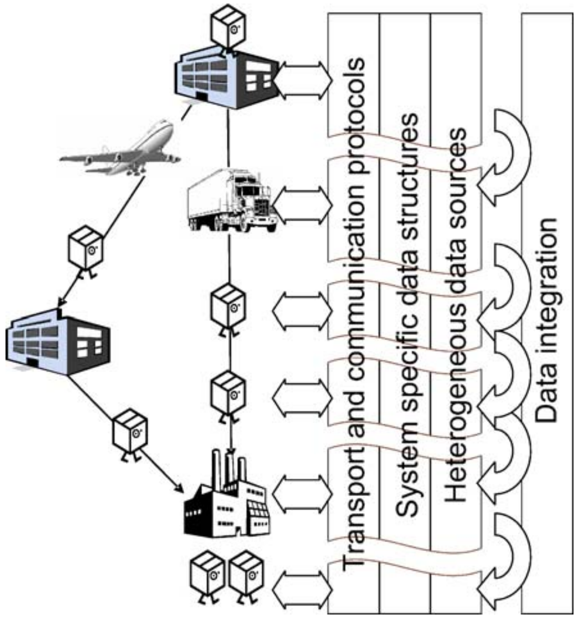


Fig. 2 Levels of Required Data Integration

tems pose the greatest challenge. However, in order for such a complex logistics system to operate, the individual system components need to be able to access and understand data regarding arbitrary entities in a predictable and reliable manner, regardless of the formal, structural and physical properties of the underlying, individual systems.

Figure 2 summarises the challenge towards data integration for complex logistic systems, especially those facilitating autonomous control. The data integration approach must be capable of bridging the semantic and syntactic gaps in the digital representation of arbitrary logistics entities caused by the storage of data across distributed heterogeneous data sources and system specific data structures. Furthermore, it must take into account the restrictions and requirements imposed by the necessity for communication across different transport and communication protocols in complex logistics systems.

3 State of the Art

Currently, a plethora of systems, data representation models and exchange formats are in use in logistics systems in industry. The logistics system landscape is diverse and complex. Solutions such as Global Connect, UPS Online Tools, PAS and EasySped are employed by logistics service providers and are each based on specific proprietary data models. The same can be said for the ERP systems in use in the field of logistics, which themselves strive to implement the flexibility to represent processes covering a wide range of sectors.

The main standardising factors to be found in this landscape are predominantly common data exchange formats such as specific EDIFACT subsets (Ballnus, 2000) such as FORTRAS or the interface languages of popular ERP systems. Furthermore, XML-based formats play an increasing role due to their ability to represent data in a platform independent, structured way (Yergau et al., 2004).

For example, the EPC (Electronic Product Code) Network offers the EPCIS (Electronic Product Code Information Services) specification, which concentrates on defining an interface to views upon item-specific product data within supply chains. EPCIS can successfully be employed to model lifecycle dynamics and other related areas, and can readily be extended in scope, syntax and semantics by means of "user vocabularies". However, the strong focus on item-specific tracking and tracing remains dominant in the scope of application of the EPCIS standard.

The Integrated Project PROMISE (Product Lifecycle Management and Information Tracking Using Smart Embedded Systems) is currently developing a standardised approach to exchange instance-specific product information in an XML messaging format, PMI (Promise Messaging Interface). As the name implies, this format focuses on providing access to product information and concentrates on adequately representing the product lifecycle. Whilst significant areas of lifecycle management do overlap with those of logistics, PMI doesn't explicitly support item-specific logistics information exchange.

Solutions such as Seeburger Business Integration Server do indeed provide the technology to integrate heterogeneous enterprise systems across all stakeholders of a supply chain. However, the challenges of data integration in complex logistics systems go beyond the focus of such solutions, which mainly concentrate on the optimisation of trade relationships by integrating large-scale enterprise systems such as CRM, SCM and ERP. For one, the type of item-specific granularity with regards to logistics entities are outside of the scope of such solutions, for another, the integration of real-time, dynamic data sources such as sensor network or embedded systems are generally not part of this class of integration server.

4 Approach to Data Integration

The diversity of systems involved in complex logistic processes defines challenging requirements towards an approach to data integration in this field. The variety of different applications, data sources, exchange formats and transport protocols demand a high level of flexibility and scalability.

To address these challenges, the literature in the field of traditional data integration needs to be studied bearing in mind the specific requirements complex logistics systems exhibit. Here, a number of different, traditional solutions to data integration may be taken into account, foremost amongst these tightly coupled, loosely coupled and object-oriented approaches (Wache 2003). Whilst a tightly coupled approach can quickly be dismissed on grounds of its inflexibility, loosely coupled and object-oriented approaches cannot be adopted without critical analysis.

An object oriented approach generally provides good mechanisms for avoiding integration conflicts. However, when considering this approach, one must take into account that a single canonical model is required to describe the entire data model, which clearly restricts its flexibility and scalability. Each time a new stakeholder or data source enters the logistics system, the model needs to be extended. Depending on the dynamics of the logistics system, this may or may not be a disqualifying factor with regards to this approach. As the fluctuation of data sources, stake holders and systems in a complex logistics system with any degree of autonomous control can be assumed to be high, an object-oriented approach to data integration is likely to be unsuitable.

A loosely coupled approach requires detailed knowledge of each of the heterogeneous data sources to be able to be successfully employed. With regards to complex logistics systems, further analysis is required to determine whether this is feasible or not. The possibility of requiring highly flexible, and thus possibly not always predeterminable, context data, for example from sensor networks, may prove to be an argument against this approach.

Besides the aforementioned traditional approaches to data integration, a number of predominantly semantic approaches remain to be taken into account. Here, the main concepts constituting architecture of such data integration systems are mediators (Ullmann 1997). In this approach, both syntactic and semantic descriptions of the data to be integrated are applied. The semantic mediator is capable of extracting knowledge regarding the data structures of the underlying data sources and subsequently transforming, decomposing and recomposing data requests according to that knowledge. The mediator relies on semantic descriptions of the data sources. In the case of item-specific product information management, this implies a wholly semantic modelling of product information and data across the distributed, heterogeneous sources, for which a number of approaches, such as ontologies, may be chosen. Here, extensive research is required to determine whether such semantic product models are feasible and adequate to address the requirements of item-specific product data management. However, the application of the semantic mediator to the problem area currently offers the most promising solution candidate—the following paragraphs attempt to sketch a technical solution to the problem on that basis.

Figure 3 introduces a possible solution architecture combining mediator technology with a service-oriented design pattern. The goal of this architecture is to provide a standard interface towards freely definable logical views of the data describing logistics entities which may be stored in arbitrary data repositories, including mobile, dynamic sources such as sensor networks, embedded systems etc.

The core of the suggestion is the provision of transparent mechanisms for the transformation between different standards and data structures. This is represented in Fig. 3 below in the upper section by "data transformation and integration", which can technically be realised by means of a suitable implementation of the mediator concept. To the left and right of that component, the semantic descriptions of the data sources are shown along with the necessary transformation rules by means of

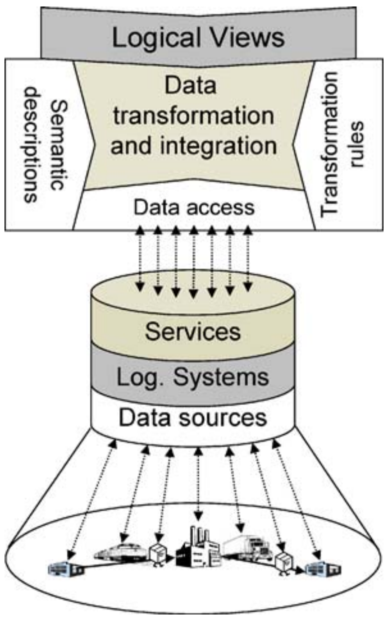


Fig. 3 Suggested Approach to Data Integration

which the mediator is able to understand and decompose queries, forward them to the data sources in question and recompose a logical view of the returned data. Obviously, the nature of complex logistics systems demands a close analysis of system requirements with regards to the synchronisation of data access and maintenance of data persistence across the highly varied and dynamic data sources involved.

Access to both the logical views by the querying entities as well as to the individual data sources by the mediator is proposed by means of a service-oriented architecture. An architecture is described as service-oriented (SOA) when it uses loosely coupled software services to provide functionality (Stojanovic et al., 2004). Here, logic is not packaged as individual programmes, but is distributed across an amount of independent services. The actual implementation details of these services by their provider are completely transparent to the consumer. Currently, the most popular implementations of SOA are carried out using Web Services (Thoben et al 2003) which are built using the combination of the XML standards SOAP (Simple Object Access Protocol, cf. Gudgin et al, 2003), WSDL (Web Service Definition Language, cf. Christensen et al, 2001) und UDDI (Universal Description, Discovery and Integration, cf. Clement et al, 2004). These standards, in combination with the Hypertext Transfer Protocol (http), provide a system independent approach to the discovery, identification, provision and consumption of services according to SOA. However, a SOA may also be built using other technology such as CORBA (Common Object Request Broker Architecture), DCOM (Distributed Component Object Model) or Enterprise Java Beans (Blanke et al, 2004).

5 Approach and Methodology

Figure 4 shows a matrix of methods and processes, the objectives to be reached by employing them, the standards and norms that can be utilised to support the methods and concrete tools that can be used. Three methods and processes have been identified which represent major areas of work towards the realisation of the suggested approach to data integration—the evolutionary development of the overall system, the application of ontologies for the semantic descriptions of data sources, and finally, methods and processes of software engineering to drive forward the technical implementation of the system.

The overall approach to the development of the proposed solution should be one of an iterative nature. Several standardised development processes support this type of approach, foremost among these being the IBM Rational Unified Process (RUP). RUP is not a single concrete prescriptive process, but rather an adaptable process framework, from which the appropriate elements of the process can be selected. Several de-facto standard software tools exist to support project management in this context, one example of which is Microsoft Project.

Technologies and methods for semantic descriptions play a major role in the development of the proposed data integration approach. On the one hand, they can

Methods and processes	Evolutionary system development	Ontologies	Software Engineering
Objectives	Iterative approach	Service modelling, Description of semantics	Requirements definition, system conception, architecture
Standards and Norms employed	RUP	OWL, RDF, DAML+OIL, WSMO, usw.	UML 2.0
Supporting tools	MS Project	Protege	Poseidon

Fig. 4 Matrix of Methods, Processes, Standards and Tools to be Employed

be applied to service-oriented architectures, with the aim of modelling them and the individual services from a semantic perspective. According to (de Bruijn 2005), Web Services, for example, can be modelled by the Web Ontology Language (OWL) based Web Service Modelling Ontology (WSMO) in combination with the Web Service Modelling Language (WSML).

The second major area in which ontologies can be applied in the context of this work is the semantic description of the individual data sources as well as to describe transformation rules. A number of formal languages are available for the definition of ontologies adequate for these purposes. The majority of these are based on mark-up languages such as XML. DAML+OIL (Darpa Agent Markup Language + Ontology Inference Layer, cf. Harmelen et al. 2001) is an agent mark-up language for the semantic description of ontologies, which is based on XML and RDF (Resource Description Framework). The further development of DAML+OIL was taken up by the Web Ontology Language (OWL), which is the specification of a formal language for the description of terms and domains. OWL was developed to be machine-readable, that is, that software, such as agents, is able to understand the semantic descriptions formalised in OWL (Smith et al 2004). These properties make OWL a logical candidate for use in the development of the semantic descriptions of data sources in the proposed approach to data integration for complex logistics systems.

6 Conclusion

Currently today's logistics systems neither exploit their full potential nor do they fulfil the requirements of tomorrows dynamic logistics systems. Although powerful technologies and ICT-systems are available only local optima regarding efficiency, robustness etc. can be reached. The lack of relevant logistics data and information must be seen as the main reason for this situation. Business decisions are taken based on more or less local data and information. This lack of relevant data and information can be overcome by an approach for seamless data integration. It grants access to relevant data and information covering the whole value chain and therefore supports an intelligent decision making.

While this would mean a significant improvement of today's logistics systems it has to be considered as an absolute must for future organisational approaches like autonomously cooperating logistics environments.

Due to the diversity of existing ICT-solutions in the field of logistics a suitable data integration approach requires a unique access to different data sources while furthermore dealing with various still changing and evolving data formats and standards. Thus there is a need for a common interface which is generic enough as well as for structural knowledge supporting the identification and interpretation of data within the variety of different sources. The combination of a service oriented interface for data access with an intelligent mediator-based approach for the interpretation and transformation of logistics data as it is proposed here offers the flexibility which is required in the context of logistics systems.

References

Hannus, M.: Islands of automation in construction. In: Žiga Turk, ed., Construction on the information highway, number 198 in CIB publication, University of Ljubljana, 1996.

Ballnus, R.: Erfolg mit EDI und E-commerce—Handlungsempfehlungen für die Abstimmung und Organisation interorganisationaler Netzwerke. Tectum Verlag, Deutschland, 2000.

Blanke, K., Krafzig, D., Slama, D.: Enterprise Soa: Service Oriented Architecture Best Practices, Prentice Hall International, 2004.

De Bruijn, J. et al: Web Service Modeling Ontology (WSMO), W3C Member Submission, World Wide Web Consortium, 2005.

Christensen, E., Curbera, F. Meredith, G., Weerawarana, S.: Web Services Description Language (WSDL) 1.1. W3C Recommendation, World Wide Web Consortium, http://www.w3.org/TR/wsdl.html, 2001.

Clement, L., Hatley, A., von Riegen, C., Rogers, T. (Hrsg.): UDDI Version 3.0.2, http://uddi.org/pubs/uddi_v3.htm, OASIS Open, 2004.

Gudgin, M., Hadley, M., Mendelsohn, N., Moreau, J.-J., Nielsen, H.F.: SOAP Version 1.2 Part 1: Messaging Framework, W3C Recommendation, World Wide Web Consortium, http://www.w3.org/TR/soap12-part1/, 2003.

Harmelen, F., Patel-Schneider, P.F., Horrocks, I. (Hrsg.): Reference Description of the DAML+OIL Ontology Markup Language. Joint United States/ European Union ad hoc Agent Markup Language Committee 2001.

Smith, M.K., Welty, C., McGuinness , D.L. (Hrsg.): OWL Web Ontology Language Guide, World Wide Web Consortium, 2004.

Stojanovic, Z., Dahanayake, A.N.W., Sol, H.G.: Modeling and Design of Service-Oriented Architecture. In Proceedings of the 2004 IEEE International Conference on Systems, Man and Cybernetics: Impacts of Emergence Cybernetics and Human-Machine Systems. IEEE Computer Society, 2004.

Ullman, J.D.: Information integration using logical views. In Afrati, F.N., Kolaitis, P., editors, Proceedings of the 6th International Conference on Database Theory (ICDT'97), 1997.

Wache, H.: Semantische Mediation für heterogene Informationsquellen. Akademische Verlagsgesellschaft, Dissertationen zur Künstlichen Intelligenz, Berlin, 2003.

Yergeau, F., Bray, T., Paoli, J., Sperberg-McQueen, C.M., Maler, E.: Extensible Markup Language (XML) 1.0 (Third Edition)—W3C Recommendation. World Wide Web Consortium, (Massachusetts Institute of Technology, European Research Consortium for Informatics and Mathematics, Keio University), 2004.

Developing a Measurement Instrument for Supply Chain Event Management-Adoption

Rebekka Sputtek, Joerg S. Hofstetter, Wolfgang Stölzle, Phillip Kirst

University of St. Gallen, Chair of Logistics Management
Dufourstrasse 40a, CH-9000 St. Gallen, Switzerland

Abstract Supply Chain Event Management (SCEM) is an emerging topic in both business practice and academia. It receives increasing attention as more companies implement SCEM-systems. However, despite its practical relevance, no general definition of SCEM exists in literature or within companies until today. Also, so far SCEM has not been operationalized for quantitative studies. Thus, the goal of this paper is to contribute to an understanding of SCEM by defining a measurement instrument to measure the level of SCEM-adoption in companies. This has been approached on the one hand by pursuing a literature review of both practitioners and academic literature and on the other hand by conducting case studies in companies using SCEM-systems. First, a common definition of SCEM is generated from a combination of findings from literature and expert interviews, and second, a measurement instrument for SCEM-adoption is developed. Implications include advice on how to test the SCEM-adoption measurement instrument statistically in order to asses its ability to measure SCEM-adoption with high validity.

1 Introduction

During the last two decades, the logistics function has experienced an increased demand for high-quality logistics while logistics costs were to decrease simultaneously. This circumstance forced companies to drive down costs by reducing or even eliminating time- and stock buffers within the supply chain or sourcing out activities. This resulted in an increasingly segmented chain of value creation. The consequence of an increasingly segmented chain of value creation is a high degree of interorganizational integration with numerous intersections between companies involved.

When a value chain is very segmented and time- as well as stock buffers are low, these circumstances lead to an increasing probability of unexpected events and at the same time to a larger scope of these events. The success of the supply chain partners heavily depends on their ability to quickly detect disturbances within the supply chain and introduce counteractions (2003).

H.-D. Haasis et al., *Dynamics in Logistics*
DOI: 10.1007/978-3-540-76862-3,

Within the context of these developments, the concept of Supply Chain Management (SCEM) has received increased attention. While usual approaches focus on the optimization of intraorganizational processes, SCEM takes an interorganizational perspective and aims at increasing end-customer benefit as well as the benefit of the companies involved (Cooper et al. 1997). SCEM aims at building a bridge between medium- and long term planning and the operative execution of these plans (Wieser and Lauterbach 2001). This is done by increasing visibility along the supply chain which is supposed to enable the timely detection of disturbances within material-, goods-, and information-flows of a company and counteract proactively in order to fulfill plans on time (Nissen 2002).

Despite its relevance, no general definition of SCEM exists in literature or within companies until today. Also, so far SCEM has not been operationalized for quantitative studies. Thus, the goal of this paper is to contribute to an understanding of SCEM by defining a measurement instrument for SCEM-adoption in companies. This is approached by first developing a common definition of SCEM and then proposing an instrument to measure SCEM-adoption by a company.

2 Methodology

Statistical analysis follows a rather standardized approach in management research (Bollen 1989; Homburg and Giering 1996). In general, first the measurement model is developed and then the structural model is tested. This can either be done simultaneously with full-information estimation models or in a two-step approach recommended by Anderson and Gerbing (1988). The authors name four main advantages of using the two-step approach. First, the two-step approach enables a test of significance for all pattern coefficients. Second, within the two-step approach, the fit of any structural model possibly being build with the available items is assessed. Third, the measurement model and the structural model can be tested independently. And fourth, the two-step approach enables formal comparisons between the model of interest and other likely alternatives. Consequently, we follow the two-step approach within this research. In this paper, we describe the initiation of the first step, namely the development of the measurement instrument.

According to Bollen (1989), the measurement process begins with a concept. A concept can be defined as an “idea that unites phenomena under a single term”. In order to represent concepts in a measurement model, a construct based on the concept developed needs to be established. The aim of this paper is to develop a valid measurement instrument for the concept “SCEM-adoption”.

In order to develop a measurement instrument, according to Homburg and Giering (1996) the first step is to develop a broad conceptualization of the phenomenon under investigation. The aim is to perform a broad qualitative analysis in order to approach the concept from various directions and shed light upon its different facets. It includes the evaluation of existing literature and relevant documents, expert interviews, focus groups and the critical incident technique (Homburg and Giering

1996). The result of this first step is a rough conceptualization of the concept and an initial amount of indicators ready to be refined by quantitative analysis.

The quality of a measurement instrument is determined by analyzing its validity (Homburg and Giering 1996). Validity is the basis for any empirical research since it ensures that what "...we conclude from a research study can be shared with confidence" (Mentzer and Flint 1997, p. 65). In order to ensure validity of results once a measurement instrument is used for causal analysis, the measurement instrument needs to fulfill certain validity criteria (Dunn et al. 1994), to which construct validity is central. In order to ensure construct validity, its sub-dimensions content validity, substantive validity, unidimensionality, reliability, convergent validity, discriminant validity, and predictive validity need to be assessed.

2.1 Development of a measurement instrument for SCEM-adoption

We followed the approach described above to measurement instrument development. Thus, we first started with a literature review of existing SCEM definitions, objectives, and applications.

Second, in order to further investigate these primary notions from literature, case studies in twelve SCEM-adopting companies were pursued. The expert interviews conducted in this context were supported by a semi-structured interview guide (Fontana et al. 1994) and were structured to cover all areas of interest. The interview-guide was not changed to ensure comparability to the utmost extent.

Third, after the initial items were developed, we presented them to a round-table of academics for verification and further sophistication.

Fourth, we discussed the measurement instrument developed with a round-table of ten business practitioners responsible for SECM in their organization using the critical incident technique. This led to a further refinement of the items.

Ultimately, we derived the measurement instrument for SCEM-adoption by combining the results from the literature review, the exploratory case study research, the academic round table and the workshop with SCEM-business practitioners.

A review of supply chain event management definitions

To initially approach SCEM, we have reviewed the academic literature on SCEM-definitions. Otto (2003, p 1) declares that: "SCEM attempts to identify, as early as possible, the [...] deviations between the plan and its execution across the multitude of processes and actors in the supply chain to trigger corrective actions according to predefined rules." Christopher and Lee (2004, p 391) find that: "The idea behind event management is that partners in a supply chain collaborate to identify the critical nodes and links through which material flows across the network. At these nodes and links, control limits are agreed within which fluctuations in levels of activities are acceptable [...]. If for whatever reason the level of activity goes outside the control limit, then an alert is automatically generated to enable corrective action to be taken." According to the academic point of view, the focus of SCEM is on an increase in interorganizational visibility of logistics processes by providing timely event-related information. This increased visibility allows to discover disturbances within material-, product-, and information flows in advance and counteract proactively in order to fulfill plans on time (Nissen 2002; AberdeenGroup 2006; Zimmermann et al. 2006). Generally, supply chain event management can be understood as providing a connection between the medium and long term planning and the operative execution of these plans (Wieser and Lauterbach 2001; Heusler et al. 2006).

In business literature, supply chain event management has also received attention. IDS Sheer (2007) defines supply chain event management as a tool with which companies can: "... control events and processes over the entire logistics chain and in doing so create an adaptive logistics network for a real time enterprise. In case of plan deviations predefined decision rules automatically activate the dispatch of messages or system-reactions. This methodology enables implementation of the comprehensive supply chain idea in practice ..." SAP (2001, p 1) constitutes that by supply chain event management companies can "... perform supply chain event management, watching over the processes, inventories, and other 'events' that take place across the supply chain. They manage information over heterogeneous systems and multiple tiers, at every stage of the supply chain process; detect, evaluate, and solve problems in real time; and create a community of businesses that can collaborate in a responsive, adaptable supply chain." Despite these numerous notions, until today no standardized definition for Supply Chain Event Management (SCEM) exists (Teuteburg and Schreber 2003).

Additionally, in order to gain an understanding of definitions of SCEM currently used in companies, we analyzed the results from the twelve case studies in SCEM-adopting companies as well as the round-table discussion with ten SCEM-business experts. Summarizing these sources it can be concluded that in companies there is often only a common understanding but no standardized, explicit definition of SCEM. Furthermore, the understanding of SCEM depends on the role and position within the supply chain (industry, services, trade) and correlates with the anticipated objectives of SCEM to improve transparency of the supply chain and differentiate towards the customer. Also, the management component within the understanding of SCEM is at present rather weakly developed in companies. Controlling processes is most prevalent, while the controlled processes can contain far more than supply activities.

Definition of SCEM

Generating a definition for SCEM from theory and practice implies combining the elements discovered in both fields. The first common element in academic literature, business literature and business practice is the objective of SCEM to increase visibility along the supply chain and the according reaction of the SCEM-system towards critical deviations from plans (SAP 2001; Nissen 2002; Otto 2003; Christopher and Lee 2004; AberdeenGroup 2006; Heusler et al. 2006; Zimmermann et al. 2006; IDSSheer 2007). While the academic literature concentrates more on the proactive character of the SCEM-idea (Bodendorf and Zimmermann 2005; Heusler et al. 2006) and stresses the planning activity possible with SCEM, practically oriented literature focuses more on controlling processes along the supply chain (SAP 2001; IDSSheer 2007).

Consequently, the following definition of SCEM has been developed from a merger between theory and practice reviewed above and is used throughout this paper: Supply Chain Event Management (SCEM) is a concept to establish maximum transparency in supply chains by planning and actively controlling inter- and intra-organizational material, information and financial flows, allowing to react as early as possible to relevant deviations with predefined escalation procedures.

Characteristics of an SCEM-System

The most important context factor in SCEM-adoption is the strategic meaning of supply chain visibility within a company. As the meaning of an improved visibility along the supply chain increases, the strategic meaning of SCEM increases accordingly. Vice versa, with an increase in strategic meaning of supply chain visibility, also the adoption level of SCEM is likely to accelerate.

The adoption level of SCEM can generally be evaluated by determining the implemented key elements and functionalities of an SCEM-system. This adoption level can be basic, only including unit tracking elements, or more advanced, such as in the adoption of true supply chain disruption management or most advanced comprising supply chain optimization. Commonly, by increasing the adoption level of an SCEM-system, companies cannot only manage their supply chains better, but also optimize them.

The literature review and the case studies revealed that the SCEM-adoption level can overall be described along five different dimensions which are deducted from the fact that SCEM builds upon tracking and tracing applications and additionally incorporates an active, event-related supply chain management component (Bretzke et al. 2002; Heusler et al. 2006).

Adoption width

We start by assessing the elements of SCEM which refer to active supply chain management. Supply chain management means the inter- and intraorganizational integration of material, goods and information flows across numerous stages of value creation.

Nissen (2002) points out, that the intraorganizational visibility of logistics processes is central to SCEM. Thus, first of all, the supported business processes which depend on the companies industry and its position in the supply chain are a dimension of SCEM-adoption.

According to the case studies pursued, the spectrum of processes captured reaches from supply, over internal logistical processes to order fulfillment, and distribution. In general all business processes could be supported with an SCEM-system but a focus on a few core-processes is recommended. Most of the case-study companies were concentrating on capturing logistical processes like transport, terminal operations, and warehousing with their SCEM-systems. Thus, in order to asses the adoption level of SCEM, the amount and kinds of processes controlled by the system along the supply chain is a valuable indicator. We will name this aspect the adoption-width of SCEM. On an item-level, we formulated the following questions in order to find out which processes in companies are controlled by an SCEM-system:

Table 1 Adoption width

Dimension	Items[a]	Likert-Scale	Inspired by
Adoption width	The following business processes are captured by our SCEM-system ...transport/storage/handling/ order fulfillment	1–7	Pfohl (2004a,b), Supply Chain Council (2007), Case studies

[a] The Items were originally developed in German, this translation is not necessarily valid

Adoption depth

Second, since SCEM is a concept based upon the general idea of supply chain management, the extension of the system towards the whole supply chain is a relevant indicator for SCEM-adoption. Comparing an SCEM-system to an ERP-system leads to the notion that an ERP system focuses mainly on the linkage between a focal company, first tier suppliers on the one hand and customers on the other hand. SCEM in turn ideally links all supply chain partners, possibly until the furthest tier supplier. Most of the SCEM systems available today have been developed by adding event management capabilities (e.g. rules for event handling) to existing centralized ERP systems (Zimmermann and Paschke 2003). This implies that SCEM differs from ERP with respect to depth of implementation. From the case studies could be found that some companies do not only monitor their own processes but the processes of their suppliers and service providers as well. Besides the monitoring of processes of external organizations, the distribution of SCEM-generated information to external organizations is a differentiating characteristic as well. Commonly, logistics service providers and customers will be provided with SCEM-information. However, it can be summarized that the more supply chain partners are involved in the SCEM-system, the deeper its adoption has progressed. In order to reflect this we developed the following items measuring SCEM-adoption depth:

Table 2 Adoption depth

Dimension	Items[a]	Likert-Scale	Inspired by
Adoption width	All divisions/suppliers/service providers/customers we intend to connect to our SCEM-systems have been connected to them	1–7	Olhager & Sedin (2003), Case studies

[a] The items were originally developed in German, this translation is not necessarily valid

Management elements
Third, considering the information within supply chains provided by tracking and tracing (T&T) applications it is reported that these systems often lead to an information overload, since they do not differentiate between (relevant) deviations from plans and regular "events". Additionally, the data provided often is of too low quality and actuality. In contrast, the SCEM system works as a filter providing the process owner with selective information only if differing significantly from normality. The system "activates" data by actively and selectively sending information to the decision maker and as such changing demand towards supply. This enables a reduction of information to the necessary level. With other words, SCEM enables management by exception (Bretzke et al. 2002; Heusler et al. 2006). However, T&T systems are not obsolete but can serve as input (delivery of data) to the SCEM system. Nevertheless, an SCEM systems needs to have more capabilities than current T&T systems. An SCEM system extends the functionality of T&T systems and as such is able to diminish its weaknesses. In addition to providing data more selectively to the process owner than T&T systems do, an SECM-system includes a planning and management function which goes beyond notifying decision makers of events, since it generate input into planning and execution systems (Heusler et al. 2006).

From the case studies we learned that first the relevant processes of a company are mapped, and then specific measurement capabilities (e.g. RFID, temperature surveillance) have to be installed which enable the company to track the current status of the planned processes. This functionality is the fundament of all SCEM-systems and thus is implemented by each company within the case-study sample. The monitoring-function refers to the SCEM-system capability to monitor real workflows in order to be able to detect relevant deviations from plans. Most of the case-study companies do not use a continuous monitoring e.g. via GPS. They are more likely to rely on a discrete monitoring of specific events, since continuous measurement always is associated with higher costs and increasing complexity. The notifying-function refers to an active and automated information provision of the system. If an event or alert involves the necessity for management action, the systems notify employees from the lowest-possible level in order to relieve upper management from time-consuming trouble-shooting. This functionality is offered by every company's SCEM-systems in the case-study sample. However, different degrees of sophistication of those notifying-functions can be identified. Some companies use simple single-staged email-escalation to either a predefined set of responsible employees or email notification to a group of employees of which the responsibility is not further differentiated. Other companies use multi-stage escalation schemes. Those systems firstly sent out emails (or SMS etc.) to a predefined set of responsible persons. If these persons do not initiate a reaction within a predefined time-frame, other (superior) employees are informed. None of the companies in the whole sample uses any kind of simulation tool, which could support managers by the evaluation of possible process-alternatives. SCEM-systems often fail to provide such a function due to the enormous data-complexity, which would be required to provide meaningful and valid alternatives (e.g. a system would need to beequipped with all shipping schedules from a given port A to a given port B). Finally, none of the SCEM-systems adopted by companies in the case-study sample automatically pursues alternative actions. Thus, always human decision-making is necessary. Consequently, the SCEM-control-function is not adopted by anyone of the companies within the case-study sample.

Summarizing the above discussion, SCEM is a concept which goes beyond the mere monitor function of most T&T applications by including the conceptual elements (event-related) notification, simulation, control and measurement. In order to measure this dimension of an SCEM-system, we developed the following items representing the SCEM-adoption level with respect to the management elements adopted:

Table 3 Management elements

Dimension	Items[a]	Likert-Scale	Inspired by
Management elements		1–7	Bretzke (2002), Heusler et al. (2006), Knickle and Kemmeter (2002), Case studies
Monitor	Our SCEM-systems automatically generates an aimed throughput for incoming orders		
	The planned workflow is equipped with …Aimed times/…Aimed places/…Aimed conditions		
	…Criteria to detect relevant deviations		
	The SCEM-systems support us in controlling …Aimed times/…Aimed places/…Aimed conditions		
Notify	Deviations of the real workflow from the planned workflow are automatically identified		
	If real workflows deviate beyond tolerance from planned workflows, alerts are send automatically to the responsible unit		
Simulate	The SCEM-systems support us in simulating corrective actions if the planned workflow cannot be pursued anymore		
	The SCEM-systems automatically propose alternative actions in case of relevant deviations from plans		
	The SCEM-systems control if there is a reaction to a warning		
Control	The SCEM-systems automatically pursue alternative actions if the real workflow deviates relevantly from the planned workflow		
	The SCEM-systems escalate their alerts independently to another predefined responsible if no reaction is taken to such an alert within a defined period of time		
	We regularly analyze our real workflow with the help of our SCEM-systems in order to identify potential improvements		

Table 3 (continued)

Dimension	Items[a]	Likert-Scale	Inspired by
Measure	The results of the real workflow are regularly used in order to optimize the SCEM-system (hardware and configuration) We regularly use the results of the real workflow in order to picture our business processes more realistically		

[a] The items were originally developed in German, this translation is not necessarily valid

Degree of aggregation and degree of differentiation

Fourth and fifth, referring to the fact that SCEM is based upon old tracking and tracing applications (Heusler et al. 2006), the units of the objects controlled by the SCEM-system as well as the aspects concerning the condition of the SCEM-objects controlled are indicators for SCEM-adoption (Stefansson and Tilanus 2000).

Also within the case studies, these two dimensions can be identified. The SCEM-objects controlled by the SCEM-systems range from relatively small pieces like packages to containers and whole shipments. The objects can be either tangible such as containers, pallets and boxes or intangible like contracts, orders and bills. No matter if the SCEM-object is tangible or not, they are differentiated in terms of their degree of aggregation. The term "aggregation" refers to the number of dividable units, which are monitored as a whole. For example packages might be not dividable anymore whereas a container can comprise several pallets which are bear a specific number of packages. The second SCEM-measurement characteristics identified within the case studies is the degree of differentiation. This is related to the scope of status information, which includes data about timeliness, physical condition (e.g. temperature), and completeness. The more status information is collected by the SCEM-system, the more transparent is the supply chain process in terms of the process condition.

In order to represent these two aspects of SCEM-adoption, we developed the following items to measure the elements degree of aggregation and degree of differentiation:

Table 4 Degree of aggregation & Degree of differentiation

Dimension	Items[a]	Likert-Scale	Inspired by
Degree of aggregation	We entirely control with our SCEM-system the following SCEM-objects: ... customer orders (for LDLs = transport orders) ... order positions ... production orders		Olhager and Sedin; Case studies

Table 4 (continued)

Dimension	Items[a]	Likert-Scale	Inspired by
	... transport orders ... transport positions (transport objects/packaging pieces) ... loading equipment (containers, pallets, boxes etc.) ...orders from suppliers/orders to service providers (for LDLs e.g. storage SPs)		
Degree of differentiation	With our SCEM-system we control the ... ...on-schedule clearing of SCEM objects ...correct location of SCEM objects ...intactness of SCEM objects ...completeness of SCEM objects ... physical condition of loading equipment ... on-schedule reaction towards reported alerts of the system		Bretzke et al. (2002), Heusler et al. (2006), Knickle and Kemmeter (2002), Case studies

[a] The items were originally developed in German, this translation is not necessarily valid

The above discussion leaves us with five dimensions comprising the measurement instrument for SCEM-adoption:

- Adoption width, namely the question concerning which of the financial, material and information flows are controlled by the SCEM-system
- Adoption depth, referring to the degree of extension of the system towards the whole supply chain
- Management elements, namely which elements of the functions monitor, notify, simulate and control are implemented
- Degree of aggregation, measuring which units of the SCEM-objects are controlled and the
- Degree of differentiation, determining which aspects concerning the condition of the SCEM-objects are controlled.

3 Implications

In this paper we have developed the measurement instrument SCEM-adoption. Following the methodology described above, we propose to now test it statistically in order to asses its ability to in deed represent the concept SCEM-adoption. This is done by ensuring construct validity by assessing its sub-dimensions content validity, substantive validity, unidimensionality, reliability, convergent validity, dis-criminant validity, and predictive validity.

Content validity

Content validity and substantive validity are subjective components of construct validity which need to be ensured qualitatively. Content validity refers to the question whether a construct is adequately represented in its various dimensions by the items developed (Straub et al. 2003). In order to prepare the quantitative analysis of a measurement instrument, content validity of the constructs needs to be ensured. Content validity implies substantive validity, since the first one refers to the correlation between the construct and its items and the second one to the relation between individual items and the construct (Garver and Mentzer 1999). However, there is no rigorous way to measure content validity or substantive validity (Dunn et al. 1994; Garver and Mentzer 1999). The only way to ensure content and sub-stantive validity is to base the measurement instrument upon thorough literature research and case studies and use at least three latent variables per manifest variable. This has been done in the research at hand.

Unidimensionality

Unidimensionality requires a single construct underlying a combination of items (Anderson et al. 1987; Kumar and Dillon 1990; Steenkamp and van Trijp 1991). Recent research suggests that confirmatory factor analysis is a valid test for unidimensionality (Anderson et al. 1987; Anderson and Gerbing 1988, Iacobucci 2001). In order to assess the fit between the estimated model and the data, overall measurement model fit is evaluated by performing the chi-square-test (X2), the Tucker-Lewis-Index (TLI), the Goodness if Fit Index (GFI), the adjusted Goodness of Fit Index (AGFI) and the Root Mean Squared Error of Approximation (RMSEA).

Concerning the X2-test, it is suggested to accept models whose ratio between the chi-square and the degrees of freedom (X2/df-ratio) is below 3 (Wheaton et al. 1977). For the TLI, an acceptable threshold is .80, while .90 or more is recommended (Baumgartner and Homburg 1996; Hulland et al. 1996). The CFI should be greater than .80 for an acceptable fit and .90 or greater for a preferred fit (Baumgartner and Homburg 1996; Hulland et al. 1996). The GFI and the AGFI should be above .90 for acceptable model fit (Medsker et al. 1994; Baumgartner and Homburg 1996). Finally, the RMSEA should be below .05 for good, below .08 reasonable and below .10 acceptable fit (Medsker et al. 1994; Baumgartner and Homburg 1996; Hulland et al. 1996; MacCallum et al. 1996).

In addition to evaluating overall model fit, also components of the measurement model need to be evaluated upon unidimensionality by assessing standardized residuals, modification indices and parameter estimates (Jöreskog and Sörbom 1993).

Reliability

Reliability is to ensure that measures which should be related to each other within the same construct indeed are related (Straub et al. 2003). Reliability builds upon the assumption of unidimensionality, which necessitates to ensure the last before assessing the latter (Churchill and Peter 1984; Anderson et al. 1987).

Scale reliability demonstrates the internal consistency of a scale when measuring a construct (Peter 1979). It used to be assessed mainly by Cronbach's Alpha. However, more recent approaches which overcome the weaknesses of Cronbach's Alpha (see e.g. Bollen 1989) are to assess composite reliability and variance extracted. Generally, both Cronbach's Alpha and composite reliability should be above .70 for a reliable scale (Nunally 1978). Variance extracted, which measures the total amount of variance within the respective indicators accounted for by the construct, should be above .50 for a reliable construct (Bagozzi and Yi 1988).

Convergent validity
Convergent validity is given if items are significantly correlated with measures from the same scale, especially as compared to other measures (Straub et al. 2003). It can be assessed by ensuring that the latent variables load significantly on the manifest variables (Dunn et al. 1994). Thus, convergent validity is an aspect of unidimensionality and as such is tested alike the latter.

Discriminant validity
Discriminant validity ensures the existence of a construct by determining which manifest variables do measure a latent variable uniquely (Straub et al. 2003). In order to test for discriminant validity it needs to be ensured that correlations between the construct are below 1. Further tests include the Fornell and Larcker's (1981) procedure and to compare the measurement with a theoretical model (Jöreskog 1971).

Predictive validity
Predictive validity assesses whether the construct of interest predicts or covariates with constructs which it is to predict or vary with (Dunn et al. 1994). Thus, predictive validity can be tested by determining whether the measurement model encloses the constructs of interest and the constructs it is to predict.

Finally, within measure development, a basic distinction needs to be made between formative versus reflective indicators. Since we assume the construct SCEM-adoption to be unobservable, but giving rise to the observable indicators adoption width, adoption depth, management elements, degree of aggregation and degree of differentiation defined above, we consider SCEM-adoption a reflective construct (Diamantopoulos and Siguaw 2006).

4 Outlook

The next step within the development of the measurement instrument SCEM-adoption involves a quantitative pre-test. This enables to test the measurement instrument developed according to the various dimensions of validity described above. After the pre-test, the measurement instrument can be used for causal analysis.

References

AberdeenGroup (2006) The Supply Chain Visibility Roadmap

Anderson J, Gerbing DW (1988) Structural Equation Modeling in Practice: A Review and Recommended Two-Step Approach. Psychological Bulletin 103(3): 411–423

Anderson JC, Gerbing DW, Hunter JE (1987) On the assessment of unidimensional measurement: Internal and external consistency, and overall consistency criteria. Journal of Marketing Research 24(4): 432–437

Bagozzi RP, Yi Y (1988) On the evaluation of structural equation models. Journal of the Academy of Marketing Science 16(1): 74–94

Baumgartner H, Homburg C (1996) Applications of structural equation modeling in marketing and consumer research: A review. International Journal of Research in Marketing 13(2): 139–161

Bodendorf F, Zimmermann R (2005) Proactive Supply-Chain Event Management with Agent Technology. International Journal of Electronic Commerce 9(4): 57–89

Bollen KA (1989) Structural equations with latent variables. New York, NY

Bretzke W-R, Stölzle W, Karrer M, Ploenes P (2002) Vom Tracking&Tracing zum Supply Chain Event Management. Aktueller Stand und Trends. Studie der KPMG Consulting AG, Düsseldorf

Christopher M, Lee H (2004) Mitigating supply chain risk through improved confidence. International Journal of Physical Distribution & Logistics Management 34(5): 388–396

Churchill GAJ, Peter JP (1984) Research design effects on the reliability of rating scales: A meta-analysis. Journal of Marketing Research 21(4): 360–375

Cooper MC, Lambert DM, Pagh JD (1997) Supply Chain Management. More than a new name for logistics. The international Journal of Logistics Management 8: 1–14

Diamantopoulos A, Siguaw JA (2006) Formative Versus Reflective Indicators in Organizational Measure Development: A Comparison and Empirical Illustration. British Journal of Management 17: 263–282

Dunn SC, Seaker RF, Waller MA (1994) Latent variables in business logistics research: Scale development and validation. Journal of Business Logistics 15(2): 145–172

Fontana A, Frey JH, Denzin NK, Lincoln YS (1994) Interviewing: The art of science. Thousand Oaks, CA, US: Sage Publications, Inc. 361-376

Fornell C, Larcker DF (1981) Evaluating structural equation models with unobservable variables and measurement error. Journal of Marketing Research: 39–50

Garver MS, Mentzer JT (1999) Logistics research methods: Employing structural equation modeling to test for construct validity. Journal of Business Logistics 20(1): 33–57

Heusler KF, Stölzle W, Bachmann H (2006) Supply Chain Event Management: Grundlagen, Funktionen und potenzielle Akteure. WiSt -Wirtschaftswissenschaftliches Studium 36(1)

Homburg C, Giering A (1996) Konzeptualisierung und Operationalisierung komplexer Konstrukte. Marketing ZFP 1(1): 5–24

Hulland J, Chow YH, Lam S (1996) Use of causal models in marketing research: A review. International Journal of Research in Marketing 13(2): 181–197

IDSSheer (2007) Supply Chain Event Management: Geschwindigkeit, Transparenz & Kontrolle in der Geschäftsprozesskette In: Sheer I, editor

Jöreskog KG (1971) Statistical analyses of sets of congeneric tests. Psychometrica 26(2): 109–133

Jöreskog KG, Sörbom D (1993) LISREL 8: Structural equation modeling with the SIMPLIS command language. Hillsdale, NJ: Lawrence Erlbaum

Knickle K, Kemmeter J (2002) Supply Chain Event Management in the Field: Success With Visibility. Boston

Kumar A, Dillon WR (1990) On the use of confirmatory measurement models in the analysis of multiple-informant reports. Journal of Marketing Research, 27(2): 102–111

MacCallum RC, Browne MW, Sugawara HM (1996) Power analysis and determination of sample size for covariance structure modeling. Psychological Methods 1(2): 130–149

Medsker GJ, Williams LJ, Holahan PJ (1994) A review of current practices for evaluating causal models in organizational behavior and human resources management research. Journal of Management 20(2): 439–464

Mentzer JT, Flint DJ (1997) Validity in logistics research. Journal of Business Logistics 18(2): 199–216

Nissen V (2002) Supply Chain Event management. Wirtschaftsinformatik 44(5): 477–480

Nunally JD (1978) Psychometric theory. New York, NY: McGraw-Hill

Olhager J, Selldin E (2003) Enterprise resource planning survey of Swedish manufacturing firms. European Journal of Operational Research 146(2): 365

Otto A (2003) Supply Chain Event Management: Three Perspectives. International Journal of Logistics Management 12(2): 1–13

Peter JP (1979) Reliability: A review of psychometric basics and recent marketing practices. Journal of Marketing Research 16(2): 6–17

Pfohl HC (2004a) Logistiksysteme

Pfohl HC (2004b) Logistikmanagement: Konzepte und Funktionen. Berlin

SAP (2001) Supply Chain Event Management. SAP Solution Brief. pp. 1–4

SCC SC (2007) Supply-Chain Operations Reference-model

Steenkamp JBEM, van Trijp HCM (1991) The use of LISREL in validating marketing constructs. International Journal of Research in Marketing 8(4): 283–299

Stefansson G, Tilanus B (2000) Tracking and tracing: Principles and practice. International Journal of Technology Management 20(3/4): 252

Straub D, Bodreau MC, Gefen D (2003) Validation guidelines for IS positivist research: Gorgia State University, University of Georgia, Drexel University. pp. 1–70

Teuteburg F, Schreber D (2003) Mobile computing and auto-ID technologies in supply chain event management – an agent-based approach

Wheaton B, Muthén B, Alwin DF, Summers GF (1977) Assessing reliability and stability in panel models. In: Heise DR, editor. Sociological methodology. San Francisco, CA: Jossey-Bass

Wieser O, Lauterbach B (2001) Supply Chain Event Management mit mySAP (Supply Chain Management). HMD Praxis der Wirtschaftsinformatik n.n. (219): 65-71

Zimmermann R, Paschke A. PAMAS – An Agent-based Supply Chain Event Management System; 2003

Zimmermann R, Winkler S, Bodendorf F. (2006) Agent-based Supply Chain Event Management – Concept and Assessment, Hawaii

Developing a Security Event Management System for Intermodal Transport

Rainer Müller

Institute of Shipping Economics and Logistics, Information Logistics
Universitätsallee GW 1 Block A, 28359 Bremen, e-mail: mueller@isl.org

Abstract Bremen, Federal State of Germany, is planning to set up a headquarter for GMES - Global Monitoring of Environment and Security. One goal of the Institute of Shipping Economics and Logistics (ISL) is to develop services for GMES to improve the security in intermodal container transport with emphasis on the maritime sector.
ISL is developing a Security Event Management System which takes care of all security related events in the intermodal chain. The security system registers the schedule and transport information of a container and assigns a corridor for its transport.
During the transport the system receives events which will be used for computing a security risk factor for each container, by considering restrictions like position with respect to the assigned corridor, duration of standstill and others. According to the value of the security risk factor the user will be informed using web services, EDI or email.

1 Introduction

Nowadays, road transport is by far the preferred transport mode for passenger and freight transport (European Commission 2001). As an example, in 2005, about 80% of all transported goods were transported on the road in Germany (Bundesministerium für Verkehr 2005). Road transport benefits from the road network which can be used to transport goods almost everywhere. Thus, regarding organisation and performance, road transport is much more simple than e.g. rail transport. In road transport, the goods can be delivered directly on-site at the receiver's premises. In contrast to this, as only few consignors or receivers own a private railway siding, rail transport implies pre-carriage from the consignor to the starting railway station and post-carriage from the destination railway station to the receiver of the goods, which

H.-D. Haasis et al., *Dynamics in Logistics*

DOI: 10.1007/978-3-540-76862-3,

are performed by truck. This example already shows the additional effort regarding organisation and coordination which is necessary in rail transport. Furthermore, the transfer of the load from one transport mode to another causes delays and additional costs, which additionally affect the competitiveness in a negative way.

On the other hand, road transport is clearly connected to disadvantages. Throughout Europe, efforts are made in order to improve the road infrastructure. Despite this fact, the road network is congested frequently. As there is a rising demand of transport, the congestion of motorways and roads will even increase in the future.

Looking at short distances, there is no alternative to road transport by truck. Apart from that, looking at medium and long distances, it is desirable to implement intermodal concepts combining truck, rail, and barge in order to relieve motorways and roads. Thus, effective intermodal networks have to be implemented. The higher expenses concerning organisation and coordination implied by intermodal transport with respect to road transport must be compensated by appropriate measures in order to achieve competitiveness.

As far as security is concerned, the optimisation of container transportation relies primarily on a system of trust, where sealed container moves unimpeded through the supply chain, noted in (APL Logistics 2003). Most of the containers will be transported by vessel which are endangered for terrorists acts.

This paper describes a project of ISL to develop a demonstrator to increase the security of supply chains of container transport to decrease the risk for vessels and for hinterland transportation.

First the result of a former project of ISL will be introduced, which had the goal to develop a Logistics Event Manager which has a lot of similarities to the new Security Event Manager. Afterwards the new concept of this software and the needed security related events will be described.

After introducing several ways how to generate those events and their restrictions, the calculation of a security factor will be described.

2 The SCEM approach

The Supply Chain Event Management (SCEM) methodology for planning and monitoring supply chains is well known in the production industry. The intermodal transport of a container is accompanied by so-called events, which divide into two categories: On the one hand there are expected events such as loading and unloading messages. These and the sequence of their occurrence are clearly defined and can easily be monitored. The second class are events which occur unexpectedly and which may allude to problems such as delay messages or a notice indicating a technical defect.

While traditional systems for planning, tracking and tracing are passive, i.e. the user gets updated information only if he looks into the systems, the idea of SCEM systems is to play an active role. To this end, the planned logistics chains (e.g. in door-to-door transports) with expected events at expected times will be defined and matched against incoming event messages. Depending on the configuration, certain triggers will be fired

- either if expected events do not happen
- or other events occur which were not expected.

The SCEM system will compare the expected events with the actual events and decide on appropriate actions, e.g. inform the user, here the manager of the intermodal transport chain, in case (and only in case) of problems. The chain manager is enabled to react in time on exceptions. Problems can be coped with soon after their occurrence and before they cause a severe impact to the transport process. Thus, an optimisation of the transport chains will become feasible.

Status events can be obtained via EDI (e.g. discharge messages from container terminals), via mobile devices (e.g. handheld computers using GPRS communication) or using web interfaces. Another method to obtain automated events is the introduction of RFID technology.

In the course of the R&D project MaTIB (Management of transports and incidents in rail traffic) (MaTIB) funded by the German Federal Ministry of Research and Education, a concept and software tool have been developed to support this approach. Events, decision rules, and corresponding actions can be defined; links to operational systems ensure that the planning data are always up to date. It is a goal within the MaTIB project to implement a system which is able to receive and process these events automatically respectively to react on the absence of events. In case of deviations with respect to the planned transport process, the manager of the transport chain will be informed such that he can intervene as soon as possible, for example by re-scheduling a delayed container to a later train.

3 Logistics Event Manager

As already mentioned, the intermodal transport of a container is accompanied by so-called events, which fall into two categories: On the one hand there are expected events such as loading and unloading messages. These and the sequence of their occurrence are clearly defined and can easily be monitored. On the other hand there are events which occur unexpectedly and which may allude to problems such as delay messages or a notice indicating a technical defect.

Nowadays there are Track & Trace systems where a controller is able to get informed about the current events of a shipped container. The disadvantages of such systems are:

- The user gets updated information only on request based on manual interactions
- Users have to contact several sources (e.g. terminals, shipping lines, railways) to get information about the complete chain
- A controller who is responsible for a huge amount of containers has to manually review all these data although he is just interested in transports where something is going wrong.

A better solution instead of those passive Track & Trace systems are SCEM (Supply Chain Event Management) systems, which are well known in the production industry.

ISL has developed such a SCEM system for logistic purposes to evaluate how SCEM concepts can be used for intermodal container transports covering several sub-transports performed by different transport modes (e.g. imports using ocean shipping to one of the big North Sea ports, rail transport into the hinterland, final distribution to the receiver by truck).

This SCEM system, called Logistics Event Manger, compares the expected events with the current events and decides on appropriate pre-configured actions, e.g. to inform the user in case (and only in case) of problems. Problems can be coped with soon after their occurrence and before they cause a severe impact to the transport process. Thus, an optimisation of the transport chains will become feasible.

For each event there are decision rules which examine its occurrence on time, delay or total absence. Depending on the result of these examinations, the SCEM system is able to initiate appropriate actions in a flexible way. It can send emails or SMS which can notify their receivers about the occurrence of a specific event. In addition, the user's computer system can be affected such that containers originally associated to a cancelled voyage are marked so they can easily be re-scheduled to another voyage.

4 Security Event Manager

The Institute of Shipping Economics and Logistics (ISL) is now about to use the SCEM concept for security purposes as well. Of course there are logistic events which could be also interesting for security issues, e.g. a delayed arrival of the container at a terminal could also be a fact of a disruption. But most events which are important for security of a transport are not bound directly to logistic processes.

The main idea is to adapt the SCEM concept to security related events. First of all there was an analysis which security related information and events are interesting for an intermodal container supply chain.

4.1 Security related data and events

For a security system we have to differ between static information and events which occur dynamically.

Static information which are interesting for a security system are:

- Value of goods: The higher the value the higher is the risk of theft.
- Type of cargo: dangerous goods could be used for acts of terrorism.

- Destination: There are locations which are more interesting for terrorism acts like others, e.g. terminals.
- Companies: The kind of involved companies during the transportation has a great impact on the security, e.g. if they are screened and certified. The more companies are involved in a intermodal transportation the higher the risk of theft or terrorism acts.

Beside this static data there are events occuring during a transport and could be messaged by different systems and persons.

4.2 Automatic messaging of events

By using smart units or smart containers events could be messaged automatically. These devices are able to detect breaking the door by light and contact sensors. Furthermore by using shock, fire, smoke and temperature sensors they can notify changes inside the container which could be a disruption. Some units are able to send a message by GSM/GPRS to a security system.

Some of those devices are also using RFID technology to track the containers position at important locations. There are also smart units and smart containers which are equipped with an GPS Module which could be used for permanent position tracking.

Due the high initial and maintenance costs of this devices it seems to be useful only for containers with high-value goods or dangerous cargo.

Instead of such expensive units which are applied on each container, ISL has developed a low-price solution which can be used on the vehicle, like train, truck or barge. A standard mobile phone and a GPS receiver (or future Galileo) is used to track the container while it is transported by the equipped carrier. A small Java application on the mobile phone is receiving the GPS position by the GPS receiver and sends it by using web services to a server. Users are able to track the current position in a Google Maps application.

The current position of the container can be used for generating events, which is one functionality of the Logistics Event Manager and will also implemented in the new Security Event Manager.

In the Logistics Event Manager an event will be generated, when a transported container enter or leaves a defined area, e.g. a terminal. The user is able to define the position and the radius of locations (like terminals or also motorway junctions) as well as the type of event which should be fired by entering or leaving the defined circle.

A server is receiving the GPS position and calculating if the transported container has entered or left a designed location and fires the defined event (see Fig. 1). The Logistic Event Manager receives the event and decide by using the rule tree to send an e-mail or to use EDI to inform the user. The system is also able to inform the user about standstills of the transport.

Fig. 1 Generating events by entering an area

This functionality will be also used in the Security Event Manager to generate security related events.

On the one hand the user of the system could assign a detailed description of the container's route, which will be used as a security corridor. By comparing the current position and the security corridor the system will be able to identify a deviation and to create a event if necessary, e.g. a truck is leaving the motorway before it was planned (see Fig. 2).

Beside the comparison of the current position and the planned security corridor the algorithm also has to evaluate the probability of a possible threat. This probability is lower if, e.g. the driver of the truck is leaving the motorway in the case of a congestion. There are several countries where TMC (Traffic Message Channel) is available and can be used for retrieving information about congestions. The Security Event Manager should be able to use this information to calculate the chance of a threat.

On the other hand the system will be able to recognize standstills when the carrier is not changing its position for a longer time. There are several restrictions concerning such standstills:

- A standstill somewhere in a forest is more dangerous than on a motorway service area
- The longer the time the higher is the chance of a disruption
- A standstill during the night could be more dangerous than in the daytime

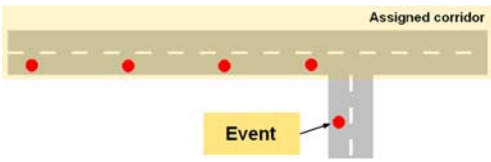


Fig. 2 Detecting events by GPS

The system also has to take care of events which has occurred in the past, because e.g. the chance of a disruption is high if the container has left the assigned security corridor one hour ago and it is not moving anymore.

4.3 Generating events manually

It is handy to generate events in an automatically manner, but there are events which could only be detected by humans. Furthermore not every container will be equipped with a smart unit.

By using software and hardware clients, especially mobile clients, workers could send messages to the Security Event Manager which has to evaluate if an event should be created. Security related events could be:

- Changed goods, which could be detected by opening the container during a customs check
- False, missing or broken seal
- The carrier, chassis or driver is a different one than the planned one
- High security level of a terminal (in cause of ISPS)
- Nervousness of the driver

4.4 Security factor

The Security Event Manager has to calculate a security factor assessing the current threat level of a transport. This factor will last from 0 to 100, where 100 means the highest threat level.

The band of the factor will be divided into three categories: OK, low risk and high risk; which could be used for visualisation purposes.

Following the SCEM idea the security factor itself will be used for generating messages, e.g. Mail or EDI messages, to inform the user proactive.

The calculation of the security factor will base mainly upon the security relevant data (e.g. dangerous cargo, value of goods) and the occurred events.

The user will have the option to declare a container transport to be endangered itself, as a result of circumstances, e.g. a very important container. This would change the threshold in the algorithm which will increase the security factor accordingly.

The algorithm will not use any rule tree, because there are too many variables which lead to huge and complex rule trees. Instead of using a precise algorithm, ISL is about to use Fuzzy Logic to calculate a approximate value.

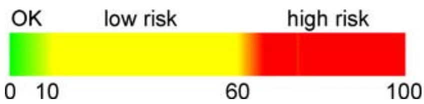


Fig. 3 Security factor and categories

5 Conclusions

It was shown that a Supply Chain Event Management (SCEM) system can support the workflow in intermodal container transport and essentially assist the chain manager. A SCEM system was built using the tool Visual Rules for the modelling of the complex decision rule structure. The system receives events which occur during the transport and examine them concerning punctuality, delay, or absence. Dependent on the results, appropriate actions are performed notifying the transport chain manager about problems by email or SMS or directly affecting other computer systems. Within the MaTIB project, a SCEM system will be set up for a specific railway service in the north of Germany. Results of the test operation will be available in 2007.

Security Event Management is a good application for SCEM, but it is necessary to identify the ways how to generate the security related events.

Using synergies of the Logistics Event Manger and identifying the security related events can help to create a new application to increase the security of intermodal transport.

Instead of a small number of logistic events there are a lot of security related events having several restrictions. These complex restrictions are not manageable by an algorithm based on a rule tree. Therefore, the application of a Fuzzy Logic concept instead offers the opportunity to solve this issue.

References

APL Logistics (2003) Adding Security and Value to the Supply chain, http://www.apl.com/news/documents/security_white_paper.pdf

Bundesministerium für Verkehr, Bau und Wohnungswesen (Hrsg.) (2005) Verkehr in Zahlen 2005/2006, Deutscher Verkehrs-Verlag, Hamburg

Collins, J. (2005) IBM, Maersk Developing Cargo Tracker, RFID Journal Sept. 22, 2005, http://www.rfidjournal.com/article/articleview/1884/1/1/

European Commission (2001), White Paper – European transport policy for 2010: time to decide, Office for Official Publications of the European Communities, Luxembourg

International Maritime Organisation (2003), The International Ship and Port Security Code, 2003 Edition

MaTIB, http://www.matib.de

Savi, http://www.savi.com

Logistic Processes Modelling

Autonomous Control of a Shop Floor Based on Bee's Foraging Behaviour

Bernd Scholz-Reiter, Thomas Jagalski, Julia C. Bendul

Bremen Institute of Industrial Technology and Applied Work Science (BIBA),
Hochschulring 20, 28359 Bremen, Germany

Abstract This paper focuses on the application of a bee-like autonomous control method to a matrix-like shop floor model with setup times. Autonomous control means decentralized coordination of intelligent logistic objects in a dynamically changing environment. By the aid of a continuous flow simulation the system's performance will be analyzed in regard to the application effect on throughput times and inventory levels.

1 Introduction

The ability to cope with increasing internal and external dynamics and complexity grows more and more to the crucial factor of company's going concern. Shorter delivery times in connection with increased schedule performance in the field of multi-stage delivery chains are only few new market conditions present companies have to deal with. These changes require continuous improvement of production responsiveness and call for more flexibility of production and logistic systems.

Apparently present production planning and control systems are unable to cope with this kind of complex dynamics [1]. This leads to the introduction of autonomous control strategies. In accordance to the definition of autonomous control [2], this term can be understood as the decentralized coordination of intelligent logistic objects and their autonomous routing through a production system. For the application it is necessary to develop local decision rules that allow the autonomous decision making while the global objectives are reached through interaction and emergent behaviour [3, 4].

Different autonomous control strategies have been developed for different scenarios with varying processing times, with and without the consideration of setup-times. The queue length estimator [5, 6] is based on local information of buffer

H.-D. Haasis et al., *Dynamics in Logistics*
DOI: 10.1007/978-3-540-76862-3, © Springer 2008

levels and the resulting expected waiting times. In addition a pheromone-like control approach was introduced, which uses data from past events and is inspired by ant's foraging behaviour [7]. This paper aims on the introduction of a further biologically inspired, but non-pheromone-based strategy that utilizes the mechanisms of bee's foraging behaviour and communication to design new rules to improve the ability of production systems to deal with increasing dynamics.

2 Autonomy in production logistics

The core of the concept of autonomous control is the development of decentralized and heterarchical control methods to react contemporary and efficient to changes in the complex and dynamical environment. The decision making process is transferred to the single logistic objects (e.g. machines, parts etc.) that control themselves autonomously by the application of recent information and communication technologies such as radio frequency identification (RFID), sensor networks or wireless communication networks. These technologies can be understood as enablers for intelligent products and parts to process information, to render decisions and to communicate with other logistic objects [8].

The goal of application of autonomous control strategies is the achievement of increased robustness and positive emergence of the total system by allocated accomplishment of dynamics and complexity of non-deterministic systems by higher flexibility and autonomy of decision making. To develop and analyze such autonomous control strategies Scholz-Reiter et al. [9] introduced a matrix-like shop floor scenario, which is modified to proof the capability of the bee-like approach to deal with problems like setup times and unexpected disturbances. The scenario is modelled with the System Dynamics methodology in a continuous flow model as this allows implementing feedback loops according to current data during the simulation easily.

3 Shop floor scenario

The considered shop floor consists of m parallel production lines with each n machine M_{ij}. Each machine has an input-buffer B_{ij} in front of it. K different products can be produced within this system (cf. Fig. 1).

At the source the raw materials for each product type enter the system. We assume that the different product lines are more suitable for certain products: in other words each machine at each stage has different processing times for each product; and each product type has a preferred line For simplicity the priority rule for the different products is set to first-come-first-served (FCFS).The production lines are coupled at every stage. That means every line is able to process every kind of product and the parts can switch to a different line at each production stage. They can

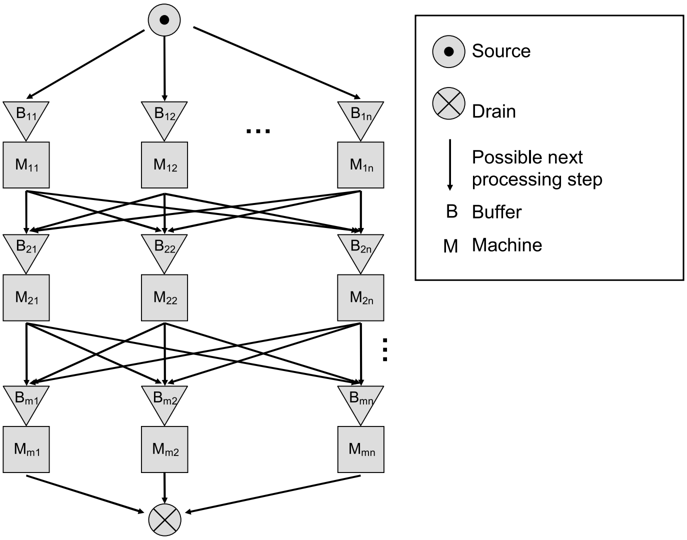


Fig. 1 Schematic illustration of the mxn shop floor scenario

decide autonomously to change their basic process plan and to use a parallel machine instead in times of overload.

4 Autonomous control based on bee's foraging behaviour

Nature often acts as model for the development of new ways to deal with uncertainties in the production environment. Colonies of insects like ants and bees show an impressive behaviour, which has been classed as Swarm-Intelligence [10]. The individuals follow simple rules that allow solving complex problems beyond the capability of the single group members. These colonies are characterized by adaptiveness, robustness and self-organization [11]. It has been suggested that these adaptive properties can lend themselves to distributed optimization problems i.e. in the field of telecommunication, transportation and manufacturing.

Searching for new ideas for autonomous shop floor control the bee's foraging behaviour can act as a proper model. There are several interesting control mechanisms based on simple rules that can be transferred to these specific problems.

4.1 Choosing the best feeding place in a honey bee colony

A colony of honey bees usually has different food sources to choose from. Bees that are aware of a food source can either advertise the source by performing a 'waggle dance', they can continue to forage at the food source without recruiting nest mates, or they can abandon and go back to a pool of unemployed bees [11]. If the bee decides to start dancing it conveys information about the known food source to the 'onlooking' bees, i.e. its general direction, distance, and quality [12].

It has been experimentally shown that if a colony is offered two identical food sources with the same distance from the hive, the bees exploit the sources equally. Nevertheless, if one source is better the bees are able to exploit or switch to it, even if it was discovered later. The probability of recruiting an onlooker bee for a particular flower patch is directly proportional to the number of dances performed for that source. The length of those dances in turn is proportional to source quality [13, 14]. Thus more bees will be recruited to better food sources usually.

Each homecoming collecting bee evaluates the food source by means of the ratio of energy consumption to the energy conveyed to the hive in form of sugar concentration. The better the individual evaluation of the food source quality is the more dance runs the bee will perform. Experiments have shown that onlooker bees watch only one single dance while meeting a dancing bee by chance. Usually the onlookers leave the dancing bee before it completes its waggle dance. Thus the more runs the dance has the longer it takes and the more unemployed bees can watch it. And, the more collecting bees are attracted the more dances are accomplished. A positive feedback loop emerges [15]. However, if the environment changes this food source can be displaced. Thus the decisions making is reversible [13, 14, 15].

4.2 Transfer of best feeding place choice to the best machining program problem

Searching for new approaches for production planning and control parallels occur between the decision making process of finding the best feeding place and finding the best way through a production line. In the same way the bees have to decide, which flower patch to travel to, in an autonomous production the part itself has to decide, which machine to go to in the next processing step.

One can imagine that even if a machine offers a very good processing time, the transportation time and waiting time should also be included into the decision making process. The throughput time varies dynamically over the time i.e. in consequence of different machining performance for different products, setup times and queue lengths.

The equation for the source quality can be found at Seeley [14, 16]. This ratio is the basis for the decision how many runs the next waggle dance will have. It expresses the ratio of the gain of a food source in Joule reduced by the costs and divided by the costs in Joule

$$\text{source quality} = \frac{\text{gain [J]} - \text{costs [J]}}{\text{costs [J]}}. \qquad (1)$$

Transferring this mechanism to the machine scenario this means that a part leaving the machine after its processing has to decide about the duration of a signal given back to the following part. The signal strength is set to the value of 1 according to the one single bee performing a dance. According to the bee mechanism a part's choice of the next machine is depending on the sum of signals performed by the preceding parts (depending on the number of signals and the length of the signals). In the production scenario the evaluation equation expresses the ratio of the value added at one device reduced by the costs (caused by time spent for processing, transportation and waiting for each machine and product type) divided by these costs in monetary units [MU]

$$\text{machine quality} = \frac{\text{value added [MU]} - \text{costs [MU]}}{\text{costs [MU]}}. \tag{2}$$

For the application to the simulation model the machine quality MQ_{mnk}(t) of each machine n on each production stage m and product type k is generated by the reduction of value added VA_{mk} to the part k on production stage m [MU] by the throughput time $TPT_{mnk}(t)$ of a certain product type in a machine n on stage m [TU] multiplied by the cost rate R in monetary units per time unit divided by these costs as a time dependent function:

$$MQ_{mnk}(t) = \frac{VA_{mk} - TPT_{mnk}(t) * R}{TPT_{mnk}(t) * R}. \tag{3}$$

5 Simulation Results

5.1 Scenario without setup times

The simulation model is reduced to 3×3 machines producing 3 different products in order to handle the complexity. To model a highly dynamic market situation the demand for the different products is set as an oscillating curve with situations of over and under load. Therefore the arrival functions is for all product types sinusoidal with an amplitude $a_1 = 0.25$ and identical except for a phase shift of $1/3$ period. Every 2:24 h a new part of every type arrives to the system. The processing times for the products are cyclic: 2 h, 2.5 h and 3 h respectively for the first, second and third best machine respectively. After a phase of another 30 days to avoid transient effects the second month is chosen to analyze the throughput time and input buffer level behaviour.

Figure 2 shows the buffer levels of the three machines on the first production stage. The buffer levels of the first production step are only considered because the following stages act qualitatively in the same way but with an influx, which is smoothed by the processing times of the machines. The left side shows the system's behaviour in the case of conventional control. Conventional control means centralized pre-planned policy that schedules the parts to the line with the lowest processing time.

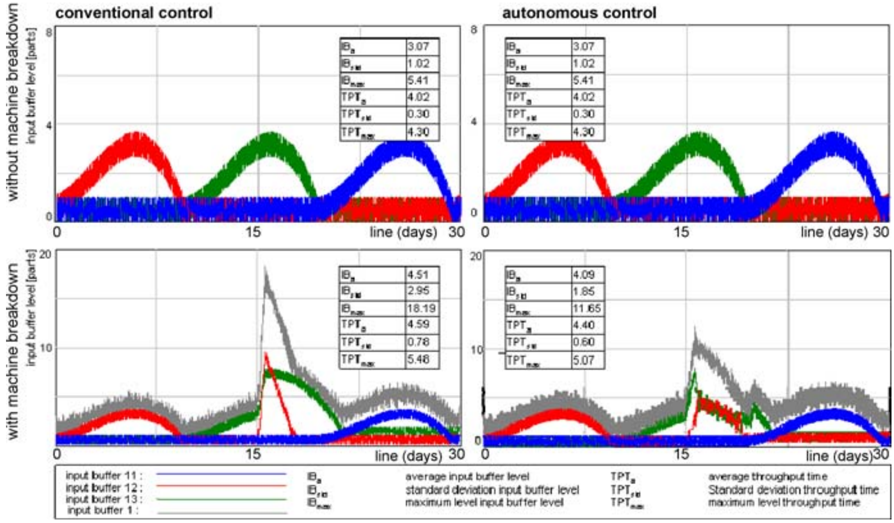


Fig. 2 Input buffer level behaviour comparison for a conventionally controlled and an autonomous controlled production scenario

Because of the identical arrival functions (except for the phase shift) the time series of the buffer levels have the same shape. The buffer levels illustrate the oscillations of the given sinusoidal arrival functions. Within one simulation period we can observe the three maxima IB_{max} at 5.41 pieces and the mean buffer level IB_a of 3.07 pieces with a standard deviation IB_{std} of 1.02 pieces. One can observe that the curve shapes and key figures are identical for the autonomous control approach without a machine break down (top).

To analyze the robustness of the bee concept a 12-hours machine breakdown was modelled. The comparison in the lower part of Fig. 2 shows that while the conventionally controlled system piles up the parts in the respective buffer until the machine starts to work again, the autonomous controlled system allocates the workload more efficient. The maximum buffer level is dramatically reduced from 18.19 pieces to 11.65 pieces. Thus the standard deviation of input buffer levels is decreased to nearly two thirds of the former value, from 2.95 pieces to 1.85 pieces and the mean buffer level is reduced from 4.51 pieces to 4.09 pieces.

The effects on the buffer levels are reflected by the throughput time behaviour. The mean throughput time TPT_a is increased compared to the systems without disturbances from 4.02 h to 4.59 h for the conventionally controlled system and only 4.40 h in the autonomous system. The maximum throughput time TPT_{max} reaches 5.48 hours for the conventionally control and 5.07 h for autonomous control compared to 4.30 h in case of no disturbance. However, the most interesting key figure may be the standard deviation of throughput time TPT_{std}. For the conventionally controlled system it is increased to 0.78 h compared to 0.30 h without a machine breakdown while the autonomous controlled system is only increased to 0.60 h.

5.2 Scenario with setup times

To design a more realistic scenario the meaning of setup times are added. For this purpose we assume that each machine n at each stage m has the same processing time of 2:00 h for each product. A setup time is added if the product type changes (Table 1). Thus, the machines' priority rule for the choice of one part out of the buffer has to be adjusted to this new situation. As a simple approach a priority rule based on the queue length estimator is chosen [5, 6]. This rule leads to the complete depletion of a certain type of parts in the buffer before a change to another product type is accomplished.

The core of the idea is the communication of the actual setup status to the parts entering the production system. These parts take the additional information into account for their decision making process, which buffer to go to.

Although there is nothing comparable in a beehive system to setup times in a machine system, it is possible to do justice to the meaning of setup times in the existing idea. In the previous section the signal strength was set to 1. By changing the signal strength to a higher value for the duration of only one time step a special advertisement can be given only to the direct successor part.

Table 2 shows that the enhanced control method improves all key figures for the input buffer levels and the throughput time. The simulation study has shown that the positive effects on the value are higher the more dynamical changes of the varying workloads are. Table 2 shows the results for a sinusoidal workload with an amplitude of $a_2 = 0.3$ and $a_3 = 0.5$: The standard deviation of throughput time TPT_{std} is

Table 1 Setup times for product changes on the different machines

Setup times [min] product type x → y	machine Mm1	machine Mm2	machine Mm3
A→ B	30	10	60
A→ C	60	30	10
B→ A	10	60	30
B→ C	60	30	10
C→ A	10	60	30
C→ B	30	10	60

Table 2 Comparison of input buffer levels and throughput time key figures for different workloads and without and with transfer of setup status information

amplitude		0.3		0.5	
Setup status advertisement (yes/no)		no	yes	no	yes
INPUT	IB_a	5.88	5.74	6.89	6.44
BUFFER	IB_{std}	1.78	1.02	1.76	1.50
	IB_{max}	11.87	9.07	13.38	11.45
THROUGH-	TPT_a	6.68	6.18	7.37	6.91
PUT TIME	TPT_{std}	0.46	0.20	0.38	0.25
	TPT_{max}	7.83	6.79	8.50	7.43

decreased for more than the half amount for amplitude a_2 (from 0.46 h to 0.20 h) and for one third for amplitude a_3 (from 0.38 h to 0.25 h). In both cases the maximum buffer level IB_{max} can be reduced for approx. 2 pieces from 11.87 to 9.07 pieces and from 13.38 to 11.45 pieces. Moreover the mean buffer level IB_a is reduced from 5.88 to 5.74 for a_2 and from 6.89 to 6.44 pieces for a_3.

6 Conclusion

It was shown that a non-pheromone-based autonomous control strategy can be applied successfully to shop floor scenarios with set-up times. In comparison with a conventionally controlled line production the autonomous controlled system based on the bee's foraging behaviour shows promising results. The simulation proves that new control method can cope with unexpected disturbances like machine breakdowns much better than a conventionally controlled system. Key figures representing the production logistic objectives like throughput time and buffer levels are improved significantly. Furthermore the additional implementation of setup times pointed out the quality of the new control method, especially in dealing with situations of highly dynamical environment, is not only possible but even improves the results of the proposed idea.

References

1. Kim J-H, Duffie NA (2004) Backlog Control for a Closed Loop PPC System. Annals of the CIRP Annals 53/1:357-360
2. Hülsmann M, Windt K (eds) (2007) Understanding Autonomous Cooperation & Control in Logistics – The Impact on Management, Information and Communication and Material Flow. Springer, Berlin Heidelberg New York
3. Ueda K, Markus A, Monostori L, Kals HJJ, Arai T (2001) Emergent synthesis methodologies for manufacturing. CIRP Annals 50/2:535-551
4. Parunak VDH (1997) Go to the Ant: Engineering Principles from Natural Multi Agent Systems. Annals of Operations Research 75: 69-101
5. Scholz-Reiter B, Freitag M, de Beer C, Jagalski T (2005) Modelling dynamics of autonomous logistic processes: Discrete-event versus continuous approaches. CIRP Annals 55/1: 413-416
6. Scholz-Reiter B, Freitag M, de Beer C, Jagalski T (2005) Modelling and analysis of autonomous shop floor control. In: Proceedings of the 38th CIRP International Seminar on Manufacturing Systems, on CD-ROM
7. Scholz-Reiter B, Freitag M, de Beer C, Jagalski T (2006) Modelling and Simulation of a Pheromone-based Shop Floor Control. In: Cunha PF, Maropoulos P (eds) Proceedings of the 3rd CIRP Sponsored Conference on Digital Enterprise Technology
8. Fleisch E, Kickuth M, Dierks M (2003) Ubiquitous Computing: Auswirkungen auf die Industrie. Industrie und Management 19/6:29-31
9. Scholz-Reiter B, Hamann T, Höhns H, Middelberg G (2004) Decentral closed loop control of production systems by means of artificial neural networks. In: Proceedings of the 37th CIRP International Seminar on Manufacturing Systems, pp 199-203

10. Camazine SD, Deneubourg JL, Franks NR, Sneyd J, Theraulaz G, Bonabeau E (2001) Self-Organization in Biological Systems. Princeton University Press, New Jersey
11. Bonabeau E, Dorigo M, Theraulaz G (1999) Swarm Intelligence. From Natural to Artificial Systems. Oxford University Press, New York Oxford
12. Camazine S, Sneyd J (1991) A model of collective nectar source selection by honey bees: self-organization through simple rules. Journal of Theoretical Biology 149:547-571
13. Seeley T, Camazine S, Sneyd J (1991) Collective decision-making in honey bees: how colonies choose among nectar sources. Behavioral Ecology and Sociobiology 28:277-290
14. Seeley TD (1994) Honey bee foragers as sensory units of their colonies. Behavioral Ecology and Sociobiology 34:51-62
15. Schmickl T (2003) Sammeln, Verteilen und Bewerten von Informationen: Verteilte Intelligenz in einem Bienenvolk. In: F-BD Informationsgesellschaft (eds) Böhlau Verlag, Wien
16. Schmickl T, Crailsheim K (2003) Costs of environmental fluctuations and benefits of dynamic decentralized foraging decisions in honey bees. In: Proceedings of the 2nd International Workshop on the Mathematics and Algorithms of Social Insect
17. Scholz-Reiter B, Jagalski T, de Beer C, Freitag M (2007) Autonomous Shop Floor Control Considering Set-up Times. In: In: Proceedings of the 40th CIRP International Seminar on Manufacturing Systems, on CD-ROM

Proof Principles of CSP – CSP-Prover in Practice

Yoshinao Isobe[1], **Markus Roggenbach**[2]

[1] National Institute of Advanced Industrial Science and Technology, Japan,
email: y-isobe@aist.go.jp
[2] University of Wales Swansea, United Kingdom,
email: M.Roggenbach@Swan.ac.uk

Abstract The process algebra CSP provides a well-established formalism for the modelling, analysis, and verification of concurrent systems. Besides being a specification language, CSP provides a valuable set of proof principles. We show in tutorial style, how these proof principles are made available in our tool CSP-Prover. Overall, CSP-Prover turns out to be an off-the-shelf proof tool ready for use in applications.

1 Introduction

The process algebra CSP (Hoare, 1985; Roscoe, 1998) provides a well-established, theoretically thoroughly studied, and in industry often applied formalism for the modelling and verification of concurrent systems. CSP has been successfully applied in areas as varied as distributed databases, parallel algorithms, train control systems (Buth and Schrönen, 1999), fault-tolerant systems (Buth et al., 1998), and security protocols (Ryan et al., 2001).

Within the spectrum of models of concurrency, semantically CSP is based on the so-called *Hoare-Languages*. In the classification of (Sassone et al., 1996), these are of type 'behaviour' (during a system's execution, a state can only appear once), 'interleaving' (parallel execution is expressed via non-determinism), and 'linear time' (CSP abstracts from the point of branching in non-deterministic choices). For such models of concurrency, CSP provides a rich variety of equivalences and refinements.

Fixing one syntax, CSP offers different semantical models, each of which is dedicated to special verification tasks. The traces model $\mathscr{T}$, e.g., covers *safety properties*. *Lifeness properties* can be studied in more elaborate models. *Deadlock analysis*, e.g., is best carried out in the stable-failures model $\mathscr{F}$, the failures-divergences model $\mathscr{N}$ allows for *lifelock analysis*, the stable-revivals model $\mathscr{R}$ (Roscoe, 2005a) has recently been designed to study *responsiveness*. The analysis of *fairness properties* requires models based on infinite traces, see (Roscoe, 1998, 2005b) for further details.

The literature on process algebra, see, e.g., (Bergstra et al., 2001), has suggested a variety of proof principles for process algebra in general, which, naturally, are also

H.-D. Haasis et al., *Dynamics in Logistics*

DOI: 10.1007/978-3-540-76862-3,

available in the context of CSP. These proof principles provide a valuable tool set in order to analyze, validate and verify concurrent systems. To a certain extend, the success of process algebra in general, and especially of CSP, is based on these very principles. The overall purpose of these techniques is to lift *semantical proofs* to *proofs on the syntactic level*: Process algebra is famous for its rich set of algebraic laws, which allow for the manipulation of specifications on the syntactic level.

CSP-Prover (Isobe and Roggenbach, 2007, 2005; Isobe et al., 2005; Isobe and Roggenbach, 2006) is an interactive proof tool for CSP based on the theorem prover Isabelle/HOL (Nipkow et al., 2002). With its theorem proving approach CSP-Prover complements the established model checker FDR (Limited, 2007) as a proof tool for CSP. CSP-Prover is generic in the various CSP models. Currently, CSP-Prover fully supports the traces model $\mathscr{T}$ and the stable-failures model $\mathscr{F}$. Implementations of the failures-divergences model $\mathscr{N}$ and the stable-revivals model $\mathscr{R}$ are well under way.

In this paper we review proof principles of process algebra and show how they can be applied in CSP-Prover. It turns out that CSP-Prover provides a suitable platform to study and apply proof principles of CSP. However, sometimes subtle but important changes are required in order to obtain an 'automatized' version of an established proof technique, which so-far was carried out manually only.

The various CSP models are defined in terms of denotational semantics. Consequently, all syntactic proof principles such as algebraic laws, fixed point induction, or the existence and uniqueness of normal forms have to be proven to be correct with respect to a given denotational semantics. CSP-Prover provides an encoding of these denotational semantics as well as of a wide range of already proven to be correct syntactic proof principles. For the stable-failures model $\mathscr{F}$ the implemented principles actually are complete (Isobe and Roggenbach, 2006). Overall, CSP-Prover can be used as an off-the-shelf proof tool, which easily can be extended by domain specific theorems.

Tools similar to CSP-Prover have been suggested by Tej/Wolff (Tej and Wolff, 1997; Tej, 2003) and Dutertre/Schneider (Dutertre and Schneider, 1997; Schneider, 2002). Here, CSP-Prover continues and extends the work by Tej/Wolf. Their tool HOL-CSP is focused on the failure-divergences model $\mathscr{N}$. While CSP-Prover works with a deep encoding of the language, HOL-CSP provides a shallow encoding of CSP in Isabelle/HOL. Consequently, in HOL-CSP proofs principles related to the CSP syntax are not easy to realise. Schneider/Dutertre base their tool on the theorem prover PVS. Their work is tailored to the verification of security protocols. Their prover implements a restricted version of traces model $\mathscr{T}$.

The paper is organized as follows: First we give an overview on the process algebra CSP as it is available in CSP-Prover. Then we show how to prove and how to apply *algebraic laws* in CSP-Prover. Section 4 discusses recursively described systems, for which *fixed-point induction* provides a powerful and often adequate syntactical proof principle. Finally, we discuss *deadlock analysis*: In the context of CSP, deadlock-freedom is often proved by means of abstraction. An alternative is to look into specialized theorems that cover deadlock-free networks.

2 The process algebra CSP in CSP-Prover

Figure 1 shows the *basic* CSP *processes* available in CSP-Prover. These processes are defined relatively to an alphabet of actions Σ and a set of process names. A process can be a (parametrized) process name or one of the primitive processes *Skip*, *Stop*, and *Div*. Actions can prefix a process. The choice between processes can be external or internal. Both choice operators are also available in a generalized version: the prefix choice operator allows the environment to choose between actions in a subset A of the alphabet Σ; in the replicated internal choice, the process chooses an index s from a set S of selectors, where this set of selectors S is either a set of natural numbers or a set of subsets of the alphabet Σ. The further progress of a process can depend on a condition. Various parallel operators are available, differing in their synchronization condition: with interleaving the processes run independent of each other; synchronous parallel forces the processes to run in lock-step; the general parallel operator synchronizes the processes in the subset X of the alphabet Σ. There is hiding and renaming. Finally, it is possible to restrict a process to its first n actions. Replicated internal choice is described in (Isobe and Roggenbach, 2006), otherwise see (Roscoe, 1998) for a detailed discussion of these operators.

In the setting of CSP-Prover, basic processes are terms of the type

```
('a,'b) proc
```

where `'a` is an Isabelle type variable to be instantiated with the type of process names, and, similarly, `'b` is an Isabelle type variable to be instantiated with the

P ::=	$p(x_1,\ldots,x_n)$	%% (parametrized) process name
|	*Skip*	%% successfully terminating process
|	*Stop*	%% deadlock process
|	*Div*	%% divergence
|	$a \to P$	%% action prefix
|	$P \,\Box\, P$	%% external choice
|	$P \sqcap P$	%% internal choice
|	$?x : A \to P(x)$	%% prefix choice
|	$!s : S \bullet P(s)$	%% replicated internal choice
|	*if b then P else P*	%% conditional
|	$P \mid\mid\mid P$	%% interleaving
|	$P \mid\mid P$	%% synchronous parallel
|	$P \,[\![X]\!]\, P$	%% generalized parallel
|	$P \,;\, P$	%% sequential composition
|	$P \setminus X$	%% hiding
|	$P[[r]]$	%% relational renaming
|	$P \downarrow n$	%% depth restriction

Fig. 1 Syntax of basic CSP processes in CSP-Prover

alphabet of actions. For example, the basic processes $a \rightarrow Skip$ and $?x : \{a,b\} \rightarrow P(x)$ can be typed-checked in Isabelle by declaring them as a `term`:

```
term "a -> SKIP"
term "? x:{a,b} -> P(x)"
```

Isabelle gives both terms the type `('a,'b) proc`.

A CSP *specification* consists of a set of *equations* of the form

$$p(x_1,\ldots,x_k) = P$$

where $p(x_1,\ldots,x_k)$ is a process name with parameters $x_1,\ldots,x_k$ and P is a basic process. A typical example is the following CSP specification of a buffer with capacity one:

$$\begin{aligned} Empty &= left?x \rightarrow Full(x) \\ Full(x) &= right!x \rightarrow Empty\,. \end{aligned}$$

Such a system of equations can be seen as a mapping from process names to basic processes. This is the way, how CSP-Prover encodes these equations, see Fig. 2. Our CSP specification `myBuffer` is declared as a function of type

```
myNames => (myNames, myAlphabet) proc
```

(line 8), the equations are encoded using the Isabelle command `primrec` (line 10 & line 11). The operator `$` is a type constructor which converts process names into basic processes.

Note that the specification in Fig. 2 describes a *generic* buffer. The type `data` is not further specified and can be instantiated with respect to the particular needs of an application. Semantically, `data` stands for an arbitrary set of values. Observe further that the equation in line 11 takes a variable `x` of type `data` as a parameter. This means that in the case that `data` is instantiated with an infinite set, the buffer specification essentially consists of infinitely many equations. CSP-Prover provides the means to analyze both generic CSP specifications as well as CSP specifications consisting of infinitely many equations.

The CSP-Prover User Guide, available at (Isobe and Roggenbach, 2007) discusses the full concrete syntax available.

```
1  (* data *)
2   typedecl data
3   datatype myAlphabet = left data | right data
4   datatype myNames = Empty | Full data
5
6  (* process *)
7   consts
8    myBuffer :: "myNames => (myNames, myAlphabet) proc"
9   primrec
10   "myBuffer (Empty) = left ? x -> $(Full x)"
11   "myBuffer (Full x) = right ! x -> $(Empty)"
```

Fig. 2 A script for the process *Buffer* in CSP-Prover

3 Algebraic Laws

Algebraic laws are at the core of process algebra. They form the basic building block for all advanced techniques of process algebraic reasoning over concurrent systems. In this section we show how to prove algebraic laws and how to apply them in CSP-Prover.

In the context of CSP, algebraic laws come in two flavours, namely as laws on equality and as laws on refinement. Here, we consider laws on equality only. Laws on refinement can be treated in the same way as laws on equality. This has its reason in the following equivalence, which in general holds in the various CSP models:

$$P \sqsubseteq Q \Leftrightarrow P = P \sqcap Q\,.$$

I.e., refinement can be expressed in terms of equality. In this sense, refinement can be considered as syntactic sugar, although in practice refinement surely is a notion of its own right. For this reason, CSP-Prover also offers specific proof procedures for refinement.

The manual verification of process algebraic laws with respect to a given non-algebraic semantics is an important but error prone and complex task. Errors found in algebraic laws years after their publication demonstrates the need for mechanised theorem proving. The latest example in the CSP context is the correction of two step laws, which, published in 1998, were shown to be incorrect in 2006, see (Isobe and Roggenbach, 2006). With the exception of the stable failures model $\mathscr{F}$, for most CSP models it is an open question if they have a complete axiomatic semantics. Thus, the proof practice might require the proof of new, yet not considered algebraic laws. Section 3.1 presents – in the stable failures model $\mathscr{F}$ – a typical proof of a process algebraic law in CSP-Prover and explains the basic techniques involved.

Many tasks in analyzing concurrent systems with process algebra can be carried out solely by applying algebraic laws. CSP-Prover offers a rich collection of such laws, which have already proven to be correct. In Sect. 3.2, we demonstrate – this time in the traces model $\mathscr{T}$ – how CSP-Prover allows one to algebraically reason about systems in a natural and intuitive way. It turns out that after structuring the proof into reasonably small subproblems, CSP-Prover's tactics can automatically discharge the arising proof obligations. Due to the sheer number – about 80 laws are required to completely capture the stable failures model (Isobe and Roggenbach, 2006) – as well as of the complexity of these algebraic laws – the $\Box$–step law studied in Sect. 3.1 below is of 'medium' complexity only – CSP and, we believe, process algebra in general require an automated reasoning approach.

Comparing the complexity of the two proof approaches demonstrated, namely first carrying out a proof on the denotational semantics of CSP and then arguing on a system using process algebraic laws only, the proof on the syntactic level turns out to be much simpler although it deals with the more involved equation.

3.1 Correctness proofs of algebraic laws

As an example of a semantical proof we consider the law '$\Box$−step' concerning the external choice operator:

$$(?x : A \to P(x)) \,\Box\, (?x : B \to Q(x)) \;= ?x : (A \cup B) \to$$
$$(\mathit{if}\ (x \in A \cap B)\ \mathit{then}\ P(x) \sqcap Q(x)\ \mathit{else\,if}\ (x \in A)\ \mathit{then}\ P(x)\ \mathit{else}\ Q(x)).$$

This law studies the behaviour of a process, which is obtained by combining the two processes $(?x : A \to P(x))$ and $(?x : B \to Q(x))$ with the external choice operator. The rhs of $\Box$−step captures which actions are possible in the first step, and – depending on this first step – how the combined process behaves further. First, the combined process makes all actions $x \in (A \cup B)$ available to the environment. After the environment has chosen which x the combined process should perform, the behaviour of the combined process depends on the set to which this x belongs: if x is in the intersection of A and B, then the environment has actually made no choice between the two possible branches, and the originally external choice becomes an internal choice between $P(x)$ and $Q(x)$; if x belongs to A but not to B, then $P(x)$ is executed; if x belongs to B but not to A, then $Q(x)$ is executed. In CSP, laws such as $\Box$−step are known as 'step laws'. These step laws capture essential algebraic properties of the various operators: There is a step law for every CSP operator; the step laws hold in all (main) CSP models.

In the following, we will use CSP-Prover to prove that $\Box$−step holds in the stable failures model $\mathscr{F}$. The denotational semantics of $\mathscr{F}$ maps process expressions via functions *traces* and *failures* to pairs

$$(T, F), \quad \text{where } T \subseteq \Sigma^{*\checkmark} \text{ and } F \subseteq \Sigma^{*\checkmark} \times \mathscr{P}(\Sigma^{\checkmark}).$$

Here, Σ is the alphabet of actions from the syntax definition of CSP[1].

To prove the law $\Box$−step, we need to show: for all choices of Σ, for all choices of $A \subseteq \Sigma$ and $B \subseteq \Sigma$, and for all interpretations of the basic processes $P(x)$ and $Q(x)$ in the semantic domain of $\mathscr{F}$ we have:

$$\mathit{traces}(\mathit{Ext}_{(A,P,B,Q)}) = \mathit{traces}(\mathit{Step}_{(A,P,B,Q)}) \qquad (\#1)$$
$$\mathit{failures}(\mathit{Ext}_{(A,P,B,Q)}) = \mathit{failures}(\mathit{Step}_{(A,P,B,Q)}) \qquad (\#2)$$

where $\mathit{Ext}_{(A,P,B,Q)}$ ($\mathit{Step}_{(A,P,B,Q)}$) denotes the lhs (rhs) of the equation $\Box$−step. In the following, we consider the proof of (#2) only:

In a *manual proof*, we first have to compute the sets $\mathit{failures}(\mathit{Ext}_{(A,P,B,Q)})$ and $\mathit{failures}(\mathit{Step}_{(A,P,B,Q)})$ by applying the semantic clauses of the model $\mathcal{F}$. After some simplifications we obtain:

$$\begin{aligned}&\mathit{failures}(\mathit{Ext}_{(A,P,B,Q)}) = \\ &\quad \{(\langle\rangle, X) \mid A \cap X = \emptyset\} \cup \{(\langle\rangle, X) \mid B \cap X = \emptyset\} \cup \\ &\quad \{(\langle a \rangle ^\frown t, X) \mid a \in A \wedge (t, X) \in \mathit{failures}(P(a))\} \cup \\ &\quad \{(\langle a \rangle ^\frown t, X) \mid a \in B \wedge (t, X) \in \mathit{failures}(Q(a))\} \qquad (\#3)\end{aligned}$$

[1] $\Sigma^{\checkmark} = \Sigma \cup \{\checkmark\}$, $\Sigma^{*\checkmark} = \Sigma^* \cup \Sigma^* {}^\frown \langle\checkmark\rangle$, where $\checkmark$ is a special symbol denoting successful termination, see (Roscoe, 1998) for a detailed discussion of $\mathscr{F}$.

```
1   lemma cspF_Ext_choice_step:
2     "(? x:A -> P(x)) [+] (? x:B -> Q(x)) =F[M,M]
3     ? x:(A Un B) -> (IF (x : A Int B) THEN P(x) |~| Q(x)
4                  ELSE IF (x : A) THEN P(x) ELSE Q(x))"
5   apply (simp add: cspF_cspT_semantics)
6   apply (simp add: cspT_Ext_choice_step)
7   apply (rule order_antisym, auto)
8   (* ⊆ *)
9     apply (simp add: in_traces in_failures)
10    apply (auto)
11  (* ⊇ *)
12    apply (simp add: in_traces in_failures)
13    apply (elim disjE conjE exE, force+)
14    apply (case_tac "a : B", simp_all add: in_failures)
15    apply (case_tac "a : A", simp_all add: in_failures)
16  done
```

Fig. 3 A proof script for step law of the external choice

$$\begin{aligned}
&\mathit{failures}(\mathit{Step}_{(A,P,B,Q)}) = \\
&\quad \{(\langle\rangle, X) \mid (A \cup B) \cap X = \emptyset\} \cup \\
&\quad \{(\langle a \rangle ^\frown t, X) \mid a \in A \cup B \wedge \\
&\qquad\qquad \mathit{if}\ (a \in A \cap B)\ \mathit{then}\ (t,X) \in \mathit{failures}(P(a)) \cup \mathit{failures}(Q(a)) \\
&\qquad\qquad \mathit{else\,if}\ (a \in A)\ \mathit{then}\ (t,X) \in \mathit{failures}(P(a)) \\
&\qquad\qquad \mathit{else}\ (t,X) \in \mathit{failures}(Q(a))\} \qquad (\#4)
\end{aligned}$$

Using standard arguments on sets, one can show that the sets (#3) and (#4) are indeed equal and that therefore the step law $\Box$–step holds in $\mathscr{F}$.

In contrast to this approach, Fig. 3 shows a *proof-script* in CSP-Prover for this step-law. First, we state $\Box$–step to be our proof goal (line 1). The command in line 5 splits the main goal into two subgoals corresponding to the Eqs. (#1) and (#2) above. The first subgoal (#1) is discharged at the line 6. Here, we use the lemma `cspT_Ext_choice_step` which belongs to the traces model $\mathscr{T}$. Now we deal with the second subgoal, the Eq. (#2). We prove this equality of sets by two subset relations, which we obtain as new subgoals at the line 7:

$$\forall (t,X) \in \mathit{failures}(\mathit{Ext}_{(A,P,B,Q)}).\ (t,X) \in \mathit{failures}(\mathit{Step}_{(A,P,B,Q)}) \qquad (\#5)$$
$$\forall (t,X) \in \mathit{failures}(\mathit{Step}_{(A,P,B,Q)}).\ (t,X) \in \mathit{failures}(\mathit{Ext}_{(A,P,B,Q)})\,. \qquad (\#6)$$

The subgoal (#5) is proved in the lines 9 and 10, the subgoal (#6) is proved in lines 12–15. For both subgoals, the sets of failures[2] of the both processes, i.e. (#3) and (#4), are automatically derived in CSP-Prover, see the lines 9 and 12. This is a powerful technique. Deriving the denotations of processes according to the semantical clauses of a CSP model is a tedious but error prone and complex task – note that the sets (#3) and (#4) are simplified versions of the sets derived from the semantical clauses. The commands in line 10 and in the lines 13–15 finally check that (#3) and (#4) are equal.

[2] The model $\mathscr{F}$ requires the trace sets in order to derive the failures of the external choice between processes.

3.2 Proofs based on algebraic laws

Given a set of algebraic laws such as $\Box-$step, how can we use these laws in CSP-Prover in order to reason about a CSP specification? Let us consider an example out of the context of testing from formal specifications, see (Kahsai et al., 2007). Given the CSP specification of a system, and given a test in the form of a sequence of actions, shall we expect an implementation of this specification, the so-called system under test, to engage in this sequence of actions? More concrete, let the system be a simple calculator which reads values x and y from a keyboard and then shows on its display the sum $x+y$ of these values:

$$Calculator\ = ?x : Button \rightarrow ?y : Button \rightarrow\ Display!(x+y) \rightarrow Skip\,.$$

Consider as a test the process which represents the sequence of pressing the button for zero, pressing the button for one, and then expecting the display to show a one:

$$Test = Button!0 \rightarrow\ Button!1 \rightarrow\ Display!1 \rightarrow Stop\,.$$

In order to prove that *Test* is a sequence that any correct implementation of *Calculator* has to execute, we need to discharge several proof obligations, see (Kahsai et al., 2007) for the details. Here, we consider just one of the equations that need to be established in this context:

$$(((Calculator \parallel Test)\ [[MyRenaming]])\ [\![\{a\}]\!]\ Count3) \setminus \{a\} =_{\mathcal{T}} OK \rightarrow Stop\,.$$

Here, $MyRenaming = \{(x,a) \mid x \in \Sigma\}$ and $Count3 = a\ \rightarrow a\ \rightarrow a\ \rightarrow OK\ \rightarrow Stop$. This equation can be shown to be correct using the following lemmas:

1. $Calculator \parallel Test =_{\mathcal{T}} Test$ `(* parallel_one *)`
   The *Calculator* and the *Test* agree on the three actions prescribed by *Test*.
2. $Test[[MyRenaming]] =_{\mathcal{T}} a \rightarrow a \rightarrow a \rightarrow STOP$ `(* renaming *)`
   All these actions are renamed into a.
3. $a \rightarrow a\ \rightarrow a\ \rightarrow STOP\ [\![\{a\}]\!]\ Count3 =_{\mathcal{T}} Count3$ `(* parallel_two *)`
   *Count*3 verifies that there are three agreed actions and communicates *OK*.
4. $Count3 \setminus \{a\} =_{\mathcal{T}} OK \rightarrow Stop$ `(* hiding *)`
   After hiding all the a's, the action *OK* remains the only one visible.

Figure 4 shows how this proof idea can be formalized in CSP-Prover, where we assume – for the moment – that the above equations have already been proven. The

```
theorem
"(((Calculator||Test)[[MyRenaming]])|[{a}]|Count3)-{a}=T Ok->STOP"
  apply(tactic { * cspT_simp_with_tac "parallel_one" 1 * } )
  apply(tactic { * cspT_simp_with_tac "renaming" 1 * } )
  apply(tactic { * cspT_simp_with_tac "parallel_two" 1 * } )
  apply(tactic { * cspT_simp_with_tac "hiding" 1 * } )
done
```

Fig. 4 Discharging a proof obligation from formal testing

```
lemma hide_a: "(a -> P) - {a} =T P - {a}"
  apply (tactic { * cspT_hsf_tac 1* } )+
done

lemma hide_OK: "(OK -> P) - {a} =T OK -> (P - {a} )"
  apply (tactic { * cspT_hsf_tac 1 * } )
done

lemma hide_STOP: "STOP - {a} =T STOP"
  apply (tactic { * cspT_hsf_tac 1 * } )
done

lemma hiding: "(a -> a -> a -> OK -> STOP) - {a} =T OK -> STOP"
  apply(tactic { * cspT_simp_with_tac "hide_a" 1 * } )+
  apply(tactic { * cspT_simp_with_tac "hide_OK" 1 * } )
  apply(auto)
  apply(tactic { * cspT_simp_with_tac "hide_STOP" 1 * } )
done
```

Fig. 5 Some typical lemmas

tactic `cspT_simp_with_tac` takes care of the rewriting process. This tactic has as parameters the name of the lemma to be applied, e.g., "`parallel_one`", and the number of the proof goal to which the lemma shall be applied – in our case always the first goal. Using the four lemmas stated above, the proof script rewrites the lhs of the original equation till it is syntactically identical to the rhs, at which point the equation trivially holds. Note how this proof script corresponds directly to the natural line of argument shown above.

It remains to prove the four lemmas used in our theorem. These four lemmas can be proven following a systematic approach, which we demonstrate for the hiding lemma, see Fig. 5. CSP-Prover's tactic `cspT_hsf_tac` is usually able to prove simple equations. Adding a $+$ to a proof command triggers its repeated execution till it fails. In order to prove the hiding lemma, we consider three basic cases: the communication is a, the communication is OK, and hiding is to be applied on the process *Stop*. Note that in `hide_a` we use the process name P as a variable, which later can be instantiated with an arbitrary process.

For the situation that the available tactics fail to prove an equation CSP-Prover also offers proof commands that allow the user to take more detailed control of the rewriting process: Equations can be decomposed into their left and right hand side; equations can be glued together again; basic processes can be decomposed into their parts; specific algebraic laws can be chosen for the rewriting process. These techniques allows the user to choose exactly the point at which an algebraic law shall be applied.

4 Fixed point analysis

In general, CSP specifications are recursive definitions. Take for instance the CSP specification

$$X = (a \to X) \,\Box\, X$$

which shall define the behaviour of a process X. Possible interpretations of X are, for instance, the processes Q_a and Q_{ab}: [3]

$$Q_a = (a \rightarrow Q_a)$$
$$Q_{ab} = (a \rightarrow Q_{ab}) \,\Box\, (b \rightarrow Q_{ab})$$

Thus, the two standard questions on recursively defined objects apply to CSP specifications: Does there exist a solution? If there exists a solution, is this solution uniquely determined? In the context of analysing systems, also a third question arises: What proof principles can be applied to the solution?

For CSP, these questions are answered either in terms of order theoretic techniques based upon Tarski's fixed point theorem or in terms of metric techniques based upon Banach's fixed point theorem. The following tabular summarizes the structures available for the main CSP models, see (Roscoe, 1998, 2005a):

Model	order theoretic structure	metric structure
$\mathcal{T}$	complete lattice	complete metric space
$\mathcal{N}$	for finite alphabet Σ : complete partial order	complete metric space
$\mathcal{F}$	complete lattice	complete metric space
$\mathcal{R}$	complete lattice	no published results

With Tarski's fixed point theorem one can show the existence of a solution, however, there can be several ones. Depending on the model under consideration, CSP selects the smallest (respectively the largest) fixed point under the chosen ordering relation and achieves in this way uniqueness. In contrast to this, Banach's fixed point theorem yields the existence of a unique solution. Consequently, only order-theoretic reasoning can be applied to the specification above. The metric approach fails.

The applicability of Tarski's and Banach's fixed point theorem depends on the semantical properties of the given CSP specification. For Tarski's theorem, on the semantical level all functions are required to be continuous with respect to the chosen ordering relation. In most CSP models, all the standard CSP operators and their composition yield continuous functions. Banach's theorem, however, requires the functions on the semantical level to be contracting. Here, the process algebraic literature has introduced the concept of 'guardedness' of a process expression. In CSP, a guarded process expression yields a contracting function on the semantical level. Thus, both semantical requirements, namely continuity and contractiveness, can be captured on a purely syntactical level. The process expression $(a \rightarrow X) \,\Box\, X$ fails to be guarded, the expressions $a \rightarrow Q_a$ and $(a \rightarrow Q_{ab}) \,\Box\, (b \rightarrow Q_{ab})$ are guarded.

[3] Replacing X on the rhs by Q_a yields with help of the idempotence law $P \,\Box\, P = P$: $(a \rightarrow Q_a) \,\Box\, Q_a = (a \rightarrow Q_a) \,\Box\, (a \rightarrow Q_a) = a \rightarrow Q_a = Q_a$. For Q_{ab} we obtain: $(a \rightarrow Q_{ab}) \,\Box\, Q_{ab} = (a \rightarrow Q_{ab}) \,\Box\, (a \rightarrow Q_{ab}) \,\Box\, (b \rightarrow Q_{ab}) = (a \rightarrow Q_{ab}) \,\Box\, (b \rightarrow Q_{ab}) = Q_{ab}$.

Concerning the analysis of recursively defined objects, fixed-point-induction offers a powerful proof method. In order to prove, e.g., in the traces model $\mathscr{T}$

$$X =_{\mathcal{T}} Q_a$$

one needs to show

$$Q_a \sqsubseteq_{\mathcal{T}} (a \to Q_a) \,\Box\, Q_a \quad \wedge \quad X \sqsubseteq_{\mathcal{T}} a \to X \qquad (*)$$

i.e., Q_a needs to be refined by the process which is obtained by substituting X by Q_a on the rhs of the defining equation of X, and X needs to be refined by the process which is obtained by substituting Q_a by X on the rhs of the defining equation of Q_a. Q_{ab} is not equivalent to X, as $X \not\sqsubseteq_{\mathcal{T}} (a \to X) \,\Box\, (b \to X)$. The two refinements $(*)$ can be proven as follows:

$$Q_a =_{\mathcal{T}} Q_a \,\Box\, Q_a =_{\mathcal{T}} (a \to Q_a) \,\Box\, Q_a, \qquad (*_1)$$

$$X =_{\mathcal{T}} (a \to X) \,\Box\, X \sqsubseteq_{\mathcal{T}} (a \to X) \,\Box\, \mathit{Stop} =_{\mathcal{T}} (a \to X) \quad (*_2)$$

Here, we use the facts that *Stop* is the largest process with respect to the order $\sqsubseteq_{\mathcal{T}}$, and that $P = Q \Leftrightarrow P \sqsubseteq Q \wedge Q \sqsubseteq P$.

4.1 Basic fixed point analysis techniques in CSP*-Prover*

The above examples are complex enough to demonstrate basic techniques for fixed point analysis in CSP-Prover. The next section on deadlock analysis will provide more elaborate examples. Figure 6 shows the encoding of the CSP specifications for X and Q_a. `PNfun` is a *reserved word* of CSP-Prover. It is used to specify that for each process name A the equation $A =$ `PNfun`(A) holds. This function `PNfun` can be used for each type of process names by the option *overloaded*, as shown in lines 8 and 14.

```
1  theory xq = CSP_T:
2    datatype Event = a | b
3
4  (* unguarded process *)
5    datatype ungPN = X
6    consts ungfun :: "ungPN => (ungPN, Event) proc"
7    primrec "ungfun X = a -> $X [+] $X"
8    defs (overloaded) ungfun_def [simp]: "PNfun == ungfun"
9
10 (* guarded process *)
11   datatype gPN = Qa
12   consts gfun :: "gPN => (gPN, Event) proc"
13   primrec "gfun Qa = a -> $Qa"
14   defs (overloaded) gfun_def [simp]: "PNfun == gfun"
15 end
```

Fig. 6 Encoding of recursive CSP specifications

Figure 7 shows the proof script for $Q_a \sqsubseteq_{\mathcal{T}} X$ and $X \sqsubseteq_{\mathcal{T}} Q_a$ (i.e. $X =_{\mathcal{T}} Q_a$). First, we declare that the cpo approach is to be used in this proof (line 4)[4]. The refinement lemmas at lines 10 and 19 give the goals $(*)$ to CSP-Prover. It is necessary to explicitly specify the type of `$X` – otherwise `$X` has no information on the type `Event`. If this typing is left out, the goal becomes more general and is not valid any more. The proof strategy used in lines 11–13 is exactly the same as the one in $(*_1)$ explained above. First, fixed point induction is applied to the rhs (line 11). The function `fx` explicitly states for each process name A on the rhs, which process on the lhs is expected to be refined by A. The command in line 12 instantiates the process name Q_a. After this command, the subgoal `$Qa <=T a -> $Qa [+] $Qa` is displayed. This can be proven by the unwinding (lines 13) as stated above $(*_1)$. The reverse direction $X \sqsubseteq_{\mathcal{T}} Q_a$ can be proven in a similar way (lines 15–22).

While, for mathematical reasons, fixed point induction of type cpo can be applied only for unfolding the rhs as demonstrated in Fig. 7, metric fixed point induction can also be applied on the lhs. In the proof practice this is of major advantage. In a refinement goal *spec* $\sqsubseteq$ *impl*, the process *impl* often has a concurrent structure like $(P1 \,[\![X]\!]\, P2) \setminus Y$, which cannot be unfolded.

The price to pay for metric fixed point induction, however, is that it requires process expression to be guarded. To this end, CSP-Prover provides a predicate `guardedfun` to check if a process expression `fun` is guarded. In our example, the application `guardedfun ungfun` leads to `False`, where `ungfun` is the function defined at line 7 in Fig. 6. This demonstrates how CSP-Prover prevents the use of metric fixed point induction for the proof goal $X =_{\mathcal{T}} Q_a$.

```
1  theory xq_refine_cpo = xq:
2
3  (* CPO apporach *)
4   defs FPmode_def [simp]: "FPmode == CPOmode"
5
6  (* Qa ⊑T X *)
7   consts fx :: "ungPN => (gPN, Event) proc"
8   primrec "fx (X) = $Qa"
9
10  lemma x_refine_qa_cpo: "$Qa <=T ($X::(ungPN, Event) proc)"
11  apply (rule cspT_fp_induct_right[of _ _ "fx"], auto)
12  apply (induct_tac p, auto)
13  by (tactic {* cspT_unwind_tac 1 *})
14
15 (* X ⊑T Qa *)
16  consts fa :: "gPN => (ungPN, Event) proc"
17  primrec "fa (Qa) = $X"
18
19  lemma qa_refine_x_cpo: "$X <=T ($Qa::(gPN, Event) proc)"
20  apply (rule cspT_fp_induct_right[of _ _ "fa"], auto)
21  apply (induct_tac p, auto)
22  by (tactic {* cspT_unwind_tac 1 *})
23
24 end
```

Fig. 7 A proof script for refinement over recursive process

[4] cpo – complete partial order. In contrast to this, it is possible to use to set `FPmode` to the value `CMSmode`; cms – complete metric space, i.e., one works with Banach's theorem.

5 Deadlock analysis

Deadlocks are certainly the best known and also most feared failures exhibited by concurrent systems. In the analysis of concurrent systems, proof of deadlock-freedom is as fundamental as termination proofs are in the context of sequential systems.

In CSP, deadlock is represented by the process *Stop*. A process is deadlock-free, if it never reaches a state equivalent to *Stop*. On the semantical level, this intuition is captured as follows: A process P is defined to be *deadlock-free* if and only if

$$\forall s \in \Sigma^* \bullet (s, \Sigma^{\checkmark}) \notin \mathit{failures}(P).$$

After executing a trace s, the process P is not allowed to refuse the set of possible communications $\Sigma^{\checkmark} = \Sigma \cup \{\checkmark\}$, namely those from the alphabet Σ and the termination symbol $\checkmark$.

This definition on the semantical level has a purely syntactical counterpart: Roscoe (Roscoe, 1998) presents the theorem that P is deadlock-free if and only if

$$DF^{\checkmark} \sqsubseteq_{\mathcal{F}} P,$$

where the process $DF^{\checkmark}$ is defined as $DF^{\checkmark} = (!x : \Sigma \rightarrow DF^{\checkmark}) \sqcap SKIP$ and $\sqsubseteq_{\mathcal{F}}$ stands for refinement in the stable failures model $\mathscr{F}$. As the stable failures model $\mathscr{F}$ has a complete axiomatic semantics (Isobe and Roggenbach, 2006), the above theorem yields that CSP has a complete calculus for checking deadlock-freedom. In Sect. 5.1 we demonstrate how to apply this calculus in CSP-Prover with an example taken from an industrial case study.

While the refinement approach completely abstracts from the structure of the concurrent system to be analyzed, other techniques have been suggested which are tailored specifically to a certain class of systems. Roscoe and Dathi (Roscoe and Dathi, 1987), for instance, present the following theorem on a specific class of networks V: Let V consist of a finite number of processes $\{P_1, \ldots, P_n\}$; if for all pairs of processes P_i and P_j with $i \neq j$ and for all states σ of the network V holds that

$$P_i \xrightarrow{\sigma}\bullet P_j \Rightarrow f_i(\sigma) > f_j(\sigma) \qquad (*)$$

then the network V is deadlock free – see (Roscoe and Dathi, 1987) for the full details of this theorem. The above formula reads: whenever the network is in a state σ such that P_i has an ungranted request to P_j (written $P_i \xrightarrow{\sigma}\bullet P_j$), i.e., the system is in a potential deadlock situation, then the progress $f_i(\sigma)$ which P_i has made in state σ is bigger than the progress $f_j(\sigma)$ which P_j has made in state σ. Essentially, the theorem applies an argument on the exclusion of circular waiting to the network. The functions f_k, $1 \leq k \leq n$, are a so-called *variant* of the network V, which needs to be established for each network individually.

This technique is of purely semantical nature: the state to be considered in condition $(*)$ includes as one of its components the failures of the network to be analyzed. In this sense, the application of this theorem is more involved than the abstraction technique. Its advantage, however, lies in its capability to analyze whole families of systems. In (Isobe et al., 2005) we give a proof of deadlock freedom for a whole family of systolic arrays. It is an open research question, if syntactic methods such as the refinement approach discussed above are capable of proving deadlock freedom for whole families of systems.

5.1 *Proofs by abstraction*

In the following, we give an example of how to prove deadlock-freedom in CSP-Prover via abstraction. To this end, we will show for a process `AC` that it is a refinement of a more abstract process `Abs` in the stable failures model $\mathscr{F}$. For this process `Abs` we prove `$DFtick` $\sqsubseteq$ `Abs`, where `$DFtick` is the CSP-Prover's representation of the process $DF^{\checkmark}$ defined above. As refinement is transitive, this establishes deadlock freedom for `AC`. It should be noted that CSP-Prover also includes the proof of the above syntactical characterization of deadlock freedom.

Our example for this section is taken from an industrial case study on the verification of EP2 (Consortium, 2002), which is a new standard of electronic payment systems. In (Gimblett et al., 2005), major parts of the EP2 system have been formalised in CSP-CASL (Roggenbach, 2006). For a system as complex as EP2, tool support is required in order to prove deadlock freedom for the interaction between the various components.

Translating the data part of the specifications given in (Gimblett et al., 2005) into adequate Isabelle code, we obtain specifications in the input format of CSP-Prover. Figure 8 shows the nucleus[5] of the initialisation procedure of the EP2 `Terminal`. The `Terminal` starts the initialisation (line 12), where `c!?x : X` $\rightarrow$ `P` represents that one of data in X is nondeterministically sent on `c`, and waits then for data sent by the `Acquirer`. If this data is of type `Request`, the `Terminal` answers with a value of type `Response` (line 15). Another possibility is that the `Acquirer` wants to exit the initialisation (line 16). Any other type of communication sent by the `Acquirer` will lead to a deadlock represented by the process `STOP` (line 16). On the other end of the communication, after receiving an initialisation request (line 17) the `Acquirer` internally decides if it wants to exit the process (line 19) or interact with the `Terminal` by sending a request followed by a response of the `Terminal` (line 20). The system `AC` to be analysed here consists of the parallel composition of the `Terminal` and the `Acquirer` synchronised on the channel `c` (line 25).

[5] For the purpose of this paper, the specification text has been simplified. The complete formalisation and proof can be found in (Isobe and Roggenbach, 2007).

```
(* data part *)
typedecl init_d   typedecl request_d   typedecl response_d   typedecl exit_d
datatype Data = Init init_d | Exit exit_d
              | Request request_d | Response response_d
datatype Event = c Data

(* process part *)
datatype ACName = Terminal | TerminalConfigManagement
                | Acquirer | AcConfigManagement
consts ACfun :: "(ACName, Event) proc"
primrec
 "ACfun (Terminal) = c !? init:(range Init) -> $TerminalConfigManagement"
 "ACfun (TerminalConfigManagement) =
   c ? x -> IF (x:range Request)
       THEN c !? response:(range Response) -> $TerminalConfigManagement
       ELSE IF (x:range Exit) THEN SKIP ELSE STOP"
 "ACfun (Acquirer) = c ? x:(range Init) -> $AcConfigManagement"
 "ACfun (AcConfigManagement) =
   c !? exit:(range Exit) -> SKIP |~|
   c !? request:(range Request)
       -> c ? response:(range Response) -> $AcConfigManagement"
defs (overloaded) Set_ACfun_def [simp]: "PNfun == ACfun"

constdefs AC :: "(ACName, Event) proc"
 "AC == ($Acquirer |[range c]| $Terminal)"
```

Fig. 8 EP2 Specification at the Abstract Component Description Level

Using CSP-Prover, we can show in the stable-failure model $\mathcal{F}$ that the above described process AC is a refinement of the process Abs of Fig. 9. This proof gives another nice example on how to apply proof procedures to recursively defined processes in CSP-Prover. Note that Abs is a purely sequential process which can successfully terminate after any even number of communications through the channel c. Which specific data is sent on c can be ignored here.

Figure 10 shows the complete script to prove the refinement relation Abs $\sqsubseteq_{\mathcal{F}}$ AC (line 16) in CSP-Prover. First, it is declared that the CMS approach is applied for the analysis of fixed points in this verification (line 2). Then we check that all processes are guarded. This check is fully automated routine (lines 5–6). Next, a mapping is defined from the process-names of Abs to process expressions in AC (line 9-12)[6]. After these preparations, Abs $\sqsubseteq_{\mathcal{F}}$ AC is given as a goal (line 16).

```
datatype AbsName = Abstract | Loop
consts Absfun :: "(AbsName, Event) procDF"
primrec
 "Absfun (Abstract) = c !? x -> $Loop"
 "Absfun (Loop) = c !? x -> (SKIP |~| c !? x -> $Loop)"
defs (overloaded) Set_Absfun_def [simp]: "PNfun == Absfun"

constdefs Abs :: "(AbsName, Event) proc"
 "Abs == $Abstract"
```

Fig. 9 An abstraction of the process shown in Fig. 8

[6] It is hard to automatically derive such correspondences. However, CSP-Prover can assist users to derive them.

```
1  (* the CMS approach is used for analysing Fixed Points *)
2  defs FPmode_def [simp]: "FPmode == CMSmode"
3
4  (* check the guards *)
5  lemma guardedfun_AC_Abd[simp]: "guardedfun ACfun" "guardedfun Absfun"
6  by (simp add: guardedfun_def, rule allI, induct_tac p, simp_all)+
7
8  (* expected correspondence of process-names in Abs to AC *)
9  consts Abs_to_AC :: "AbsName => (ACName, Event) proc"
10 primrec
11  "Abs_to_AC (Abstract) = ($Acquirer |[range c]| $Terminal)"
12  "Abs_to_AC (Loop) = ($AcConfigManagement |[range c]
13               | $TerminalConfigManagement)"
14
15 (* the main theorem *)
16 theorem ep2_Abs_AC: "Abs <=F AC"
17 apply (unfold Abs_def AC_def)
18 apply (rule cspF_fp_induct_left[of _ "Abs_to_AC"], auto)
19 apply (induct_tac p, auto)
20 apply ((tactic * cspF_hsf_tac 1 *)+,
21        (rule | rule cspF_decompo_ref |
22         rule cspF_Int_choice_left1, auto |
23         rule cspF_Int_choice_left2, rule cspF_decompo_ref, auto)?,
24        (auto simp add: image_iff inj_on_def)?)+
25 done
```

Fig. 10 The complete proof script for `Abs` $\sqsubseteq_{\mathcal{F}}$ `AC`

Using the above mapping, the proof obligations on the recursive process `Abs` are unfolded into a base case and step cases by `cspF_fp_induct_left`. This is the fixed point induction rule for unfolding the left-hand side (line 18). Since a step case is produced for each of the process names of `Abs`, the step cases are instantiated by induction on `AbsName` (line 19). Finally, the theorem is proven by CSP-Prover's tactic `csp_hsf_tac` (line 20) which sequentialises any expression, `csp_decompo_ref` which decomposes CSP-operators (line 21) by considering the refinement, for example,

$$Y \subseteq X \wedge (\forall x \in Y. P(s) \sqsubseteq_{\mathcal{F}} Q(s)) \implies\ !x : X \bullet P(s) \ \sqsubseteq_{\mathcal{F}}\ !x : Y \bullet Q(s),$$

`csp_Int_choice_left1` which selects the first P from $P \sqcap Q$ in the left-hand side (line 22), and Isabelle's tactic `auto` (line 24).

It remains to show that `AC` is deadlock-free. Figure 11 shows the complete script to prove the refinement relation `$DFtick` $\sqsubseteq_{\mathcal{F}}$ `Abs` (line 8) in CSP-Prover. In this proof, `cspF_fp_induct_right`, which is the fixed point induction rule for unfolding the right-hand side, is applied (line 10). Since we use the CMS approach in this example, we can apply the fixed point induction to the both sides. Note that in the CPO approach `cspF_fp_induct_right` is available only for unfolding right-hand sides. The proof strategy used in Fig. 11 is similar to the case of `Abs` $\sqsubseteq$ `AC`.

Finally, we give a script for showing that `AC` is deadlock-free in Fig. 12. The main goal is rewritten to `$DFtick` $\sqsubseteq$ `AC` by `DeadlockFree_DFtick_ref`, which is the encoded syntactical characterization of deadlock-freedom (line 2). We finally

```
(* expected correspondence of process-names in Abs to DF *)
consts Abs_to_DF :: "AbsName => (DFtickName, Event) proc"
primrec
 "Abs_to_DF (Abstract) = ($DFtick)"
 "Abs_to_DF (Loop) = ($DFtick)"

(* the main theorem *)
theorem ep2_DF_Abs: "$DFtick <=F Abs"
apply (unfold Abs_def)
apply (rule cspF_fp_induct_right[of _ _ "Abs_to_DF"], auto)
apply (induct_tac p, auto)
apply ((tactic {* cspF_hsf_tac 1 *}, rule cspF_Int_choice_left1,
                                rule cspF_decompo_ref, auto) |
       (tactic {* cspF_hsf_tac 1 *}, rule cspF_Int_choice_left2, simp))+
done
```

Fig. 11 The complete proof script for $DFtick $\sqsubseteq_{\mathcal{F}}$ Abs

```
theorem AC_isDeadlockFree: "AC isDeadlockFree"
apply (simp add: DeadlockFree_DFtick_ref)
apply (rule cspF_trans[of _ _ _ Abs])
apply (rule ep2_DF_Abs, rule ep2_Abs_AC)
done
```

Fig. 12 The complete proof script for the deadlock-freedom of AC

establish our proof goal by combining $DFtick $\sqsubseteq$ Aba and Abs $\sqsubseteq$ AC, which have been proved in Figs. 10 and 11, respectively.

6 Summary and Future work

In this paper, we have reviewed a number of proof principles of the process algebra CSP, have discussed their merits and limitations, and have shown by the means of practical examples how these techniques can be applied in the analysis of concurrent systems using CSP-Prover. It turned out that mechanized theorem proving is needed for both, the verification and the application of proof principles in CSP and process algebra in general.

Future work will include the implementation of further CSP models in CSP-Prover, the integration of FDR and CSP-Prover, as well as the application of CSP-Prover in further case studies.

Acknowledgements The authors would like to thank Liam O'Reilly and Temesghen Kahsai for help with some of the proof scripts, and Erwin R. Catesbeiana Jr for contributing deep insights into the nature of fixed point induction. This cooperation was supported by the EPSRC Project EP/D037212/1.

References

Bergstra JA, Ponse A, Smolka SA (eds) (2001) Handbook of Process Algebra. Elsevier Science Inc., New York, NY, USA

Buth B, Schrönen M (1999) Model-checking the architectural design of a fail-safe communication system for railway interlocking systems. In: FM'99, Springer, LNCS 1709

Buth B, Peleska J, Shi H (1998) Combining methods for the livelock analysis of a fault-tolerant system. In: AMAST'98, Springer, LNCS 1548, pp 124–139

Consortium E (ed) (2002) eft/pos 2000 Specification, version 1.0.1. EP2 Consortium

Dutertre B, Schneider S (1997) Using a PVS embedding of CSP to verify authentication protocols. In: TPHOL 1997, Springer, LNCS 1275, pp 121–136

Gimblett A, Roggenbach M, Schlingloff H (2005) Towards a formal specification of an electronic payment system in CSP-CASL. In: WADT 2004, Springer, LNCS 3423, pp 61–78

Hoare CAR (1985) Communicating Sequential Processes. Prentice Hall

Isobe Y, Roggenbach M (2005) A generic theorem prover of CSP refinement. In: TACAS 2005, Springer, LNCS 3440

Isobe Y, Roggenbach M (2006) A complete axiomatic semantics for the CSP stable failures model. In: CONCUR 2006, Springer, LNCS

Isobe Y, Roggenbach M (2007) Webpage on CSP-Prover. http://staff.aist.go.jp/y-isobe/CSP-Prover/CSP-Prover.html

Isobe Y, Roggenbach M, Gruner S (2005) Extending CSP-Prover by deadlock-analysis: Towards the verification of systolic arrays. In: FOSE 2005, Kindai-kagaku-sha, Japanese Lecture Notes Series 31

Kahsai T, Roggenbach M, Schlingloff BH (2007) Specification-based testing for refinement. In: Proceedings of SEFM 2007, IEEE Computer Society

Limited FSE (2007) Failures-divergence refinement: FDR2. http://www.fsel.com/

Nipkow T, Paulon LC, Wenzel M (2002) Isabelle/HOL. LNCS 2283, Springer

Roggenbach M (2006) CSP-CASL: a new integration of process algebra and algebraic specification. Theor Comput Sci 354(1):42–71,

Roscoe A (1998) The theory and practice of concurrency. Prentice Hall

Roscoe AW, Dathi N (1987) The pursuit of deadlock freedom. Information and Computation 75(3):289–327

Roscoe B (2005a) Revivals, stuckness and responsiveness. Unpublished draft

Roscoe B (2005b) Seeing beyond divergence. In: Communicating Sequential Processes, the first 25 years, Springer, LNCS 3525

Ryan P, Schneider S, Goldsmith M, Lowe G, Roscoe B (2001) The Modelling and Analysis of Security Protocols: the CSP Approach. Addison-Wesley

Sassone V, Nielsen M, Winskel G (1996) Models for concurrency. Theoretical Computer Science 170(1–2):297–348

Schneider S (2002) Verifying authentication protocol implementations. In: FMOODS 2002, Kluwer, IFIP Conference Proceedings, vol 209, pp 5–24

Tej H (2003) HOL-CSP: Mechanised Formal Development of Concurrent Processes. BISS Monograph Vol. 19, Logos Verlag Berlin

Tej H, Wolff B (1997) A corrected failure-divergence model for CSP in Isabelle/HOL. In: FME'97, Springer, LNCS 1313, pp 318–337

Application of Markov Drift Processes to Logistical Systems Modeling

Mikhail Postan

Department of Economics and Management, Odessa National Maritime University, 34 Mechnikov Str., 65029, Odessa, Ukraine, e-mail: postan@ukr.net

Abstract The activity of logistical systems occurs often under uncertainty (demand fluctuation, irregularity of transport units movement, unreliable equipment, etc.). Therefore, it is naturally to apply for their modeling different types of stochastic processes, in particular the Markov drift processes (with continuous sample paths). The phase space of Markov drift processes is direct product $D \times C$, where D is a finite or denumerable set, C is a continuous set, $C \subseteq R_+^n$. In our paper, a number of examples of Markov drift processes application to modeling the different kinds of logistical systems (production lines, transportation systems, supply chains, etc.) is given.

1 Definition of Markov Drift Process and its Properties

The Markov drift process is a subclass of the Markovian processes with the phase space $\Omega = D \times C$, where D is a finite or denumerable set, C is a continuous set, $C \subseteq R_+^n$ (R_+^n is the non-negative orthant of n-dimensional Eucledian space). More exactly, it is the pair $(Z(t), \xi(t)$, where $Z(t)$ is a discrete component (chain) and $\xi(t)$ is the vector with continuous components (random walk) which satisfies the differential equation:

$$\frac{d}{dt}\xi(t) = v_{Z(t)}(\xi(t)), \quad \text{a.s.} \tag{1}$$

In Eq. (1) $v_{ki}(\mathbf{x}), k = 1, 2, \ldots, n;\ i \in D, \mathbf{x} \in C$, are the given bounded functions describing the "drift" or fluctuations of vector $\xi(t)$.

For the first time this subclass of Markov processes was applied to modeling some industrial processes and data communication systems by Sevast'yanov [14], Kosten [3,4], Wijngaard [15], Anick, Mitra, and Sondhi [1], Mitra [5], Mitra and Mitrani [6], Prabhu [12] and other authors. In the article by Devis [2] and in the work by Schmidli [13] the general theory of wide class of non-diffusion stochastic processes (including the Markov drift processes) was developed but in rather

H.-D. Haasis et al., *Dynamics in Logistics*
DOI: 10.1007/978-3-540-76862-3, © Springer 2008

abstract form. In our works [7–10] the constructive definitions of some types of Markov drift processes were given which have been very suitable for mathematical and/or numerical analysis of different real systems in transport, manufacturing, and data communication.

Let $C = [0,E_1] \times [0,E_2] \times \ldots \times [0,E_n]$ and $\| \lambda_{ij}(\mathbf{x}) \|$ be the infinitesimal matrix (generator) of process $(Z(t), \xi(t))$. We assume that $\lambda_{ij}(\mathbf{x})$, $i,j \in D$, are continuous in the interval $0 < x_k < E_k$, left-continuous at the points $x_k = 0$, right-continuous at the points $x_k = E_k$, $k = 1,2,\ldots,n$, and uniformly bounded. Besides, it is assumed that the following condition is valid:

$$\sum_{i \in D} \lambda_{ij}(\mathbf{x}) = 0, \quad \mathbf{x} \in C.$$

Firstly, consider more simple case $n = 1$. Let us assume that

$$\begin{aligned} v_{Z(t)} = \sum_{i \in D} v_i \mathbf{I}(Z(t) = i) - \sum_{i \in D^-} v_i \mathbf{I}(Z(t) i, \xi(t) = 0) \\ - \sum_{i \in D^+} v_i \mathbf{I}(Z(t) = i, \xi(t) = E), \end{aligned} \tag{2}$$

where $\{v_i, i \in D\}$ is the set of given values of the chain $Z(t), \sup_i |v_i| < l < \infty$;

$$D^- = \{i : v_i < 0, \quad i \in D\} \neq \emptyset, \quad D^+ = \{i : v_i > 0, \quad i \in D\} \neq \emptyset$$

$\mathbf{I}(A)$ is the indicator of an event A.

According to Eqs. (1), (2) 0 and E are the sticky bounds for the random walk $\xi(t)$ on the closed interval $[0,E]$.

Let us introduce the following notations:

$$\begin{aligned} \Pr\{Z(t) = i, x < \xi(t) < x + dx\} &= q_i(x,t)\mathrm{d}x, \qquad && i \in D, \quad 0 \le x \le E, \\ \Pr\{Z(t) = i, \xi(t) = 0\} &= p_i^-(t), && i \in D^- \bigcup D^0, \\ \Pr\{Z(t) = i, \xi(t) = E\} &= p_i^+(t), && i \in D^+ \bigcup D^0, \end{aligned} \tag{3}$$

where $\quad D^0 = \{i : v_i = 0, \quad i \in D\}$.

We suppose that probability densities $q_i(x,t)$ are continuously differentiable for $0 < x < E, t > 0$, and there exist the limits

$$\lim_{x \to 0+} q_i(x,t) = q_i(0,t), \qquad \lim_{x \to E-} q_i(x,t) = q_i(E,t).$$

As it was shown in [8] the functions (3) may be found by solving the following system of PDE with the corresponding boundary conditions:

$$\left(\frac{\partial}{\partial t} + v_i \frac{\partial}{\partial x} \right) q_i(x,t) = -\lambda_i(x) q_i(x,t) + \sum_{j \neq i} \lambda_{ji}(x) q_i(x,t), \quad 0 < x < E, \quad i \in D \tag{4}$$

$$v_i q_i(0,t) = \sum_{j \in D^- \bigcup D^0} \lambda_{ji}(-0) p_j^-(t), \quad i \in D^+, \tag{5}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}p_i^-(t) + v_i q_i(0,t) = -\lambda_i(-0)p_i^-(t) + \sum_{j\in\mathbf{D}^-\bigcup\mathbf{D}^0;\, j\neq i} \lambda_{ji}(-0)p_j^-(t), i\in\mathbf{D}^-\bigcup\mathbf{D}^0,$$

$$-v_i q_i(E,t) = \sum_{j\in\mathbf{D}^+\bigcup\mathbf{D}^0} \lambda_{ji}(E+)p_j^+(t), \quad i\in\mathbf{D}^-, \tag{6}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}p_i^+(t) - v_i q_i(E,t) = -\lambda_i(E+)p_i^+(t) + \sum_{j\in\mathbf{D}^+\bigcup\mathbf{D}^0;\, j\neq i} \lambda_{ji}(E+)p_j^+(t), \quad i\in\mathbf{D}^+\bigcup\mathbf{D}^0,$$

$$\sum_{i\in\mathbf{D}^-\bigcup\mathbf{D}^0} p_i^-(t) + \sum_{i\in\mathbf{D}^+\bigcup\mathbf{D}^0} p_i^+(t) + \sum_{i\in\mathbf{D}} \int_0^E q_i(x,t)\mathrm{d}x = 1\,, \tag{7}$$

where

$$\lambda_i(x) = \sum_{j\neq i}\lambda_{ij}(x), \qquad 0\le x\le E, i\in D,$$

$$\lambda_i(-0) = \sum_{j\neq i}\lambda_{ij}(-0), \qquad i\in D^-\bigcup D^0,$$

$$\lambda_i(E+) = \sum_{j\neq i}\lambda_{ij}(E+), \qquad i\in D^+\bigcup D^0.$$

As $t\to\infty$ from (4)–(7) it follows two-point boundary-value problem for finding the stationary distribution of process $(Z(t),\xi(t))$.

The Eqs. (4)–(7) can be generalized for an arbitrary $n>1$ but the corresponding system of equations is too bulky. Visible results have been obtained only for $n=2$ and in the case when Eq. (1) takes the form:

$$\xi_k'(t) = \sum_{i\in D} v_{ki}\mathbf{I}(Z(t)=i) - \sum_{i\in D_\mathbf{k}^-} v_{ki}\mathbf{I}(Z(t)=i,\xi_k(t)=0)$$
$$-\sum_{i\in D_\mathbf{k}^+} v_{ki}\mathbf{I}(Z(t)=i,\xi_k(t)=E_k), \quad k=1,2, \quad \text{a.s.},$$

where

$$D_k^- = \{i: v_{ki}<0, i\in D\}\neq\emptyset, \quad D_k^+ = \{i: v_{ki}>0, i\in D\}\neq\emptyset.$$

Some generalizations of process $(Z(t),\xi(t))$ may be done by inclusion of one or several additional components, e.g. semi-Markov processes [7, 9].

The Eqs. (4)–(7) and their generalizations are very convenient for modeling the various logistical systems functioning under uncertainty. This may be caused by fluctuation of market's demand, irregularity of transport units movement, restricted reliability of equipment, etc.

One of the specific features of logistical systems modeling is necessity to take into account the combination of continuous production process and discrete nature of the transport units work. The components of random vector $\xi(t)$ may be interpreted as inventory levels of materials at warehouses or financial resources of a firm. As for chain $Z(t)$, it is very suitable for description of transport units movement and control of their distribution among points of destinations.

Some examples of Eqs. (4)–(7) application for modeling the logistical systems will be given in the next Sections.

2 Production Line with Unreliable Units

Consider a production line without intermediate buffers, consisting of N units (machines, equipment). Any unit may be subjected to breakdowns and repairs. If the all units are in good repairs, then the production rate of lone is W (we assume that arbitrary unit has the same production rate). With the rate W a final product comes to warehouse, which capacity is E, if total amount of product inside of it is less than E. From the warehouse product removes with the constant rate $U < W$ (if it isn't empty). If total amount of product at warehouse equals to E, then the production rate equals to U (see Fig. 1).

Note that the parameter U may be interpreted as given demand.

Let us introduce the following designations:

$\nu(t)$ is process which indicates the current state of production line, namely, $\nu(t) = 0$ if no one unit is in failure state at moment t, $\nu(t) = n$ if n-th unit is in failure state at moment t;

$\xi(t)$ is amount of a final product at warehouse at moment t.

It is assumed that in small interval $(t, t + \Delta t)$ with probability $a_n \Delta t + o(\Delta t)$ the n-th unit fails independently from other units and with probability $1 - a_n \Delta t + o(\Delta t)$ a failure doesn't occur. After failure of n-th unit its repair begins immediately and in small interval $(t, t + \Delta t)$ this repair is finished with probability $b_n(x)\Delta t + o(\Delta t)$ under condition $\xi(t) = x$ and with probability $1 - b_n(x)\Delta t + o(\Delta t)$ it isn't finished.

Under these assumptions the stochastic process $(\nu(t), \xi(t))$ will be the Markov drift process, where

$$D = \{0, 1, \ldots, N\}, \quad D^- = \{1, 2, \ldots, N\}, \quad D^+ = \{0\}, \quad D^0 = \emptyset,$$
$$v_i = W - U,\ i \in D^+;\ \ v_i = -U,\ i \in D^-.$$

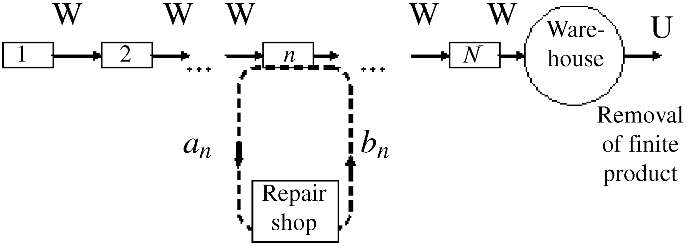


Fig. 1 Scheme of product line with unreliable units

Therefore from (4)–(7) as $t \to \infty$ we get the following system of equations for determination of functions $q_i(x) = \lim_{t\to\infty} q_i(x,t), i \in D$,
and constants $p_i^- = \lim_{t\to\infty} p_i^-(t), i \in D^-$:

$$Vq_0'(x) = -aq_0(x) + \sum_{n=1}^{N} b_n(x),$$

$$-Uq_i'(x) = -b_i(x)q_i(x) + a_i q_0(x), \quad 0 < x < E, \quad i \in D^-, \tag{8}$$

$$Vq_0(0) = \sum_{n=1}^{N} b_n(-0)p_n^-, \quad Uq_i(0) = b_i(-0)p_i^-, \quad i \in D^-, \tag{9}$$

$$Uq_i(E) = a_i p_0^+, \quad i \in D^-; \quad Vq_0(E) = ap_0^+, \tag{10}$$

$$p_0^+ + \sum_{n=1}^{N} p_n^- + \sum_{n=0}^{N} \int_0^E q_n(x)\mathrm{d}x = 1, \tag{11}$$

where

$$V = W - U; \quad a = \sum_{n=1}^{N} a_n.$$

Summerizing Eqs. (8) and taking into account the relations (9), (10) we obtain

$$Vq_0(x) = U \sum_{n=1}^{N} q_n(x), \quad 0 \le x \le E. \tag{12}$$

Substituting this equality to the right-hand side of Eq. (8) we get the system of differential equations

$$-Uq_i'(x) = -b_i(x)q_i(x) + \frac{a_i U}{V} \sum_{n=1}^{N} q_n(x), \quad i \in D^-, \quad 0 < x < E. \tag{13}$$

If $b_1(x) = b_2(x) = \ldots = b_N(x) = b(x)$, then solution to Eq. (13) is given by

$$q_i(x) = e^{\omega_1(x)} \left[q_i(0) - \frac{Va_i}{Ua} q_0(0) \right] + \frac{Va_i}{Ua} q_0(0) e^{\omega_2(x)}, \tag{14}$$

$$\text{where} \quad \omega_1(x) = \frac{1}{U} \int_0^x b(y)dy, \quad \omega_2(x) = \omega_1(x) - \frac{ax}{V}.$$

From (14) and Eqs. (9), (10) we find

$$q_i(0) = \frac{b_0}{U} p_i^-, \quad p_i^- = \frac{a_i}{b_0} e^{-\omega_2(E)} p_0^+, \quad i \in D^-; \quad q_0(0) = \frac{b_0}{V} \sum_{n=1}^{N} p_n^-, \tag{15}$$

where $b_0 = b(-0)$.

Probability p_0^+ is given by

$$p_0^+ = \left[1 + \frac{a}{b_0} e^{-\omega_2(E)} + \left(1 + \frac{U}{V}\right) \frac{a}{U} e^{-\omega_2(E)} \int_0^E e^{\omega_2(x)} \mathrm{d}x\right]^{-1}. \qquad (16)$$

For the arbitrary functions $b_i(x), i = 1, 2, \ldots, N$, the Eqs. (13) may be solved, for example, with the help of matrix calculus.

Note that it is naturally to suppose that $b_n(x)$ are the decreasing functions of $x$taking into account the possible feedbacks between enterprise and market.

Using the solution (12), (14)–(16) it is possible to calculate some important measures of production line work:

a) mean amount of final product at warehouse

$$\mathbf{E}\xi = \int_0^E x \sum_{n=0}^N q_n(x)\mathrm{d}x + Ep_0^+ = p_0^+ \left[E + \frac{a}{V}\left(1 + \frac{V}{U}\right) e^{-\omega_2(E)} \int_0^E x e^{\omega_2(x)} \mathrm{d}x\right];$$

b) probability of market losses caused by standing idle of production line

$$\sum_{n=1}^N p_n^- = \frac{a}{b_0} e^{-\omega_2(E)} p_0^+;$$

c) mean intensity of final product's output flow

$$U\left(1 - \sum_{n=1}^N p_n^-\right) = U\left(1 - \frac{a}{b_0} e^{-\omega_2(E)} p_0^+\right).$$

3 Interaction of Two Transport Units Via Warehouse

Transportation of cargo mostly occurs with the help of several kinds of transport, e.g. railways and sea or river-going ships. That is why it is very important to organize effective interaction of transport flows at points of transshipment (PT). Such interaction is carried out usually by means of warehouses located at PT (e.g. ports).

Let us it is needed the homogeneous cargo to be delivered from point **A** to point **B** through the PT **C**. For cargo delivery at point **C** from point **A** a single transport unit (TU) has been using with the carrying capacity γ_1. For cargo transportation from point **C** to point **B** a single TU has been working with the carrying capacity γ_2. At point **C** cargo has been transshipping from one TU to another via warehouse. The durations of voyages for both TU are T_1 and T_2 respectively (see Fig. 2). In the number of practical situations parameters γ_1, γ_2, T_1, T_2 are the random variables. We shall assume that they are mutually independent random variables.

Denote $\Pr\{T_k \le t\} = A_k(t), \Pr\{\gamma_k \le x\} = G_k(x), k = 1, 2.$

It is assumed that discharging of loaded TU begins immediately after its arrival at PT with the processing rate W1 if another TU is absent from PT and all cargo

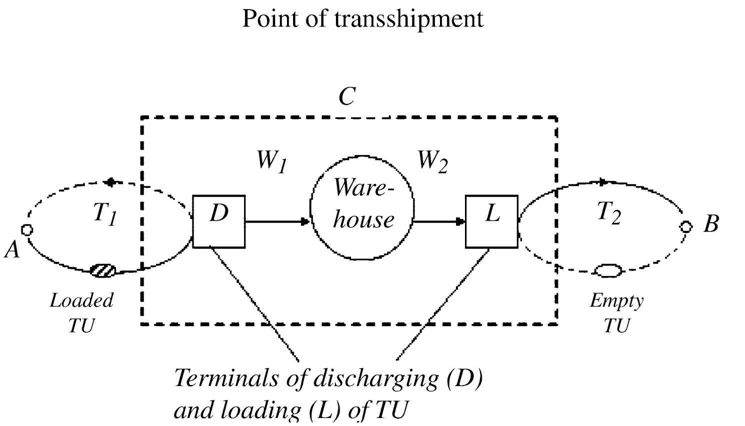


Fig. 2 Scheme of interaction of transport units at point of transshipment

comes at warehouse. Cargo from warehouse comes into TU under loading with the processing rate W2 $<$ W1 (if warehouse isn't empty). Warehouse's capacity is E. At moment when all its capacity is filled up the unloading processing rate becomes equal to W2 if there is the TU under loading. Otherwise the unloading is interrupted.

The current state of the two-modal transportation system described above we shall indicate by the random vector $(\nu_1(t), \nu_2(t), \xi(t))$, where $\nu_1(t)$ and $\nu_2(t)$ are the numbers of TU staying under unloading and loading respectively at moment t, i.e. $\nu_k(t) = 0$ or $1, k = 1,2$; $\xi(t)$ is amount of cargo at warehouse at moment t. Under suppositions

$$A_k(t) = 1 - e^{-a_k t}, G_k(x) = 1 - e^{-x/g_k}, \quad k = 1,2,$$

the process $(\nu_1(t), \nu_2(t), \xi(t))$ will be Markov drift process, where

$$D = \{(0,0),(0,1),(1,0),(1,1)\}, D^- = \{(0,1)\}, D^+ = \{(1,0),(1,1)\},$$
$$D^0 = \{(0,0)\},$$
$$v_{10} = W_1, v_{11} = W_1 - W_2, v_{01} = -W_2, v_{00} = 0.$$

Consequently, the stationary densities and probabilities

$$q_{i_1 i_2}(x) = \lim_{t\to\infty} q_{i_1 i_2}(x,t), \quad i_1, i_2 = 0,1; \quad p_{01}^- = \lim_{t\to\infty} p_{01}^-(t),$$
$$p_{10}^+ = \lim_{t\to\infty} p_{10}^+(t), \quad p_{11}^+ = \lim_{t\to\infty} p_{11}^+(t)$$

satisfy the following two-point boundary-value problem (see(4)–(7)):

$$\begin{aligned} 0 &= -(a_1 + a_2) q_{00}(x) + b_1 q_{10}(x) + b_2 q_{01}(x), \\ W_1 q'_{10}(x) &= -(a_2 + b_1) q_{10}(x) + b_2 q_{11}(x) + a_1 q_{00}(x), \\ -W_2 q'_{01}(x) &= -(a_1 + b_2) q_{01}(x) + b_1 q_{11}(x) + a_2 q_{00}(x), \\ V q'_{11}(x) &= -(b_1 + b_2) q_{11}(x) + a_1 q_{01}(x) + a_2 q_{10}(x), \quad 0 < x < E \end{aligned} \tag{17}$$

$$W_2 q_{01}(0) = a_1 p_{01}^-, \quad V q_{11}(0) = a_1 p_{01}^-, \quad q_{10}(0) = 0, \tag{18}$$

$$\begin{gathered} W_2 q_{01}(E) = b_3 p_{11}^+, \quad -W_1 q_{10}(E) = -a_2 p_{10}^+ + b_2 p_{11}^+, \\ -V q_{11}(E) = -(b_2 + b_3) p_{11}^+ + a_2 p_{10}^+ \end{gathered} \tag{19}$$

$$p_{01}^- + p_{10}^+ + p_{11}^+ + \int_0^E \left[q_{00}(x) + q_{10}(x) + q_{01}(x) + q_{11}(x)\right] \mathrm{d}x = 1, \tag{20}$$

where $V = W_1 - W_2$; $b_1 = W_1/g_1$, $b_2 = W_2/g_2$, $b_3 = W_2/g_1$.

The problem (17)–(20) may be solved by the method of Laplace transform. For the particular case $E = \infty$ it was solved in the book [11]. Its solution allows us to calculate the useful indices of interaction efficiency, for example:

1. probabilities of TU demurrage because of absence if another TU (p_{10}^+) or because of absence cargo at warehouse (p_{01}^+);
2. mean amount of cargo at warehouse

$$\mathbf{E}\xi = \int_0^E x[q_{00}(x) + q_{10}(x) + q_{01}(x) + q_{11}(x)]dx + E(p_{10}^+ + p_{11}^+).$$

For example, if $E = \infty$, then stationary state-probabilities of vector $(\nu_1(t), \nu_2(t))$ are given by the formulas

$$\begin{aligned} p_{01}^- &= 1 - \frac{a_1(a_2+b_2)}{a_2(a_1+b_1)}, \quad p_{10} = \frac{a_1 b_2 (a_1 W_1 + a_2 W_2)}{\Delta} p_{01}^-, \\ p_{11} &= \frac{a_1[a_2(a_1+a_2+b_1)W_2 - a_1 b_2 W_1]}{\Delta} p_{01}^-, \\ p_{01} &= \frac{W_1}{W_2} p_{10} + (\frac{W_1}{W_2} - 1) p_{11} + p_{01}^-, \\ p_{00} &= \frac{b_1}{a_1+a_2} p_{10} + \frac{b_2}{a_1+a_2} (p_{01} - p_{01}^-), \end{aligned}$$

where $\Delta = [(a_1+b_1)a_2W_2 - (a_2+b_2)a_1W_1](a_1+a_2+b_1+b_2)$.This results are valid only under condition (condition of steady-state regime existence)

$$\frac{a_2 W_2}{a_2 + b_2} > \frac{a_1 W_1}{a_1 + b_1}.$$

If random variables T_1 and T_2 have Erlang's distributions, then it is possible on the basis of Eqs. (4)–(7) to investigate the influence of TU movement irregularity on the indices of interaction efficiency.

4 Optimal Cargo-Flows Distribution among a Set of Transshipment Points

The last example concerns the application of results given in Section 1 to optimization of cargo-flows distribution among a set of transport-warehousing systems.

Let us an industrial enterprise sells an homogeneous product to N organizations-mediators. From enterprise the product comes to n-th mediator by pipeline with the constant processing rate W_n, $n = 1, 2, \ldots, N$. It may be, for example, any oil-product or liquefied gas. Every mediator is PT with its own warehouse for storage of product and terminal for loading on TU the product sold to the customers. Removal of sold product from any PT is carried out with help of TU. The input flow of TU arriving at n-th PT is Poisson with the rate a_n, and carrying capacities of TU are mutually independent random variables with distribution function $G_n(x)$. Any TU at n-th PT has been loading with the processing rate $U_n > W_n$ if warehouse isn't empty, and with the processing rate W_n otherwise. The limit length of TU queue at n-th PT is R_n and warehouse's capacity is E_n. Every PT has only one front of TU loading (channel) (see Fig. 3). Thus, the random fluctuation of demand for product is taken into account by irregularity of TU arrival and variation of their carrying capacities.

Let $\nu_n(t)$ denote the total number of TU and $\xi_n(t)$ is the amount of product at warehouse of n-th PT at moment t. The current state of n-th PT we shall describe by

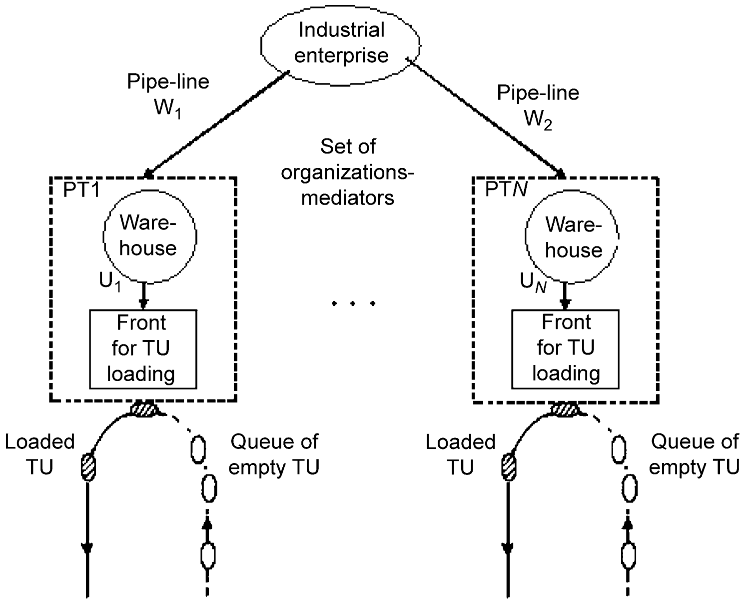


Fig. 3 Scheme of logistical system of liquid product distribution among points of transshipments

the random vector $v_n(t), \xi_n(t))$. For arbitrary distribution function $G_n(x)$ this process will not be Markovian but it becomes Markov drift process if

$$G_n(x) = 1 - e^{-x/g_n}.$$

In this case we have (index n is omitted)

$$D = \{0,1,\ldots,R+1\}, quad D^- = \{1,2,\ldots,R+1\}, \quad D^+ = \{0\}, \quad D^0 = \emptyset,$$
$$v_i = W - U, i \in D^-;\ v_0 = U.$$

For determination of stationary densities and probabilities

$$q_i(x) = \lim_{t\to\infty} q_i(x,t), i \in D,$$
$$p_i^- = \lim_{t\to\infty} p_i^-(t), i \in D^-; \quad p_0^+ = \lim_{t\to\infty} p_0^+(t)$$

from Eqs. (4)–(7) we obtain the following system of differential equations and boundary conditions:

$$\begin{gathered} Wq_0'(x) = -aq_0(x) + bq_1(x), \\ -V_1 q_i'(x) = -(a+b)q_i(x) + aq_{i-1}(x) + bq_{i+1}(x), \quad i = 1,2,\ldots,R \\ -V_1 q_{R+1}'(x) = -bq_{R+1}(x) + aq_R(x), \quad 0 < x < E, \end{gathered} \tag{21}$$

$$\begin{gathered} -b_0 p_1^- + Wq_0(0) = 0, \\ -(a+b_0)p_1^- + b_0 p_2^- + V_1 q_i(0) = 0, \\ -(a+b_0)p_i^- + b_0 p_{i+1}^- + ap_{i-1}^- + V_1 q_{i+1}(0), \quad i = 2,3,\ldots,R, \\ -b_0 p_{R+1}^- + ae(R)p_R^- + V_1 q_{R+1}(0) = 0, \end{gathered} \tag{22}$$

$$\begin{gathered} -ap_0^+ + Wq_0(E) = 0, \\ -ap_0^+ + V_1 q_1(E) = 0, \\ q_i(E) = 0, \quad i = 2,3,\ldots,R+1, \end{gathered} \tag{23}$$

$$p_0^+ + \sum_{n=1}^{N} p_n^- + \sum_{i=0}^{N} \int_0^E q_i(x)\mathrm{d}x = 1, \tag{24}$$

where $V_1 = U - W$; $b = U/g$, $b_0 = W/g$; $e(R) = 1$ if $R > 0$, $e(0) = 0$.

Put $R = 0$. Then the boundary-value problem (21)–(24) may easily be solved:

$$\begin{gathered} q_0(x) = \frac{b_0}{W} e^{\beta x} p_1^-, \quad q_1(x) = \frac{W}{V_1} q_0(x), \quad 0 \le x \le E, \\ p_0^+ = \frac{b_0}{a} e^{\beta E} p_1^-, \quad p_0^- = \left[1 + \frac{b_0}{a} e^{\beta E} + \frac{b_0}{\beta W}\left(1 + \frac{W}{V_1}\right)(e^{\beta E} - 1)\right]^{-1} \end{gathered} \tag{25}$$

where $\beta = b/V_1 - a/W$.

Taking into account (25) we can calculate the mean amount of product at warehouse:

$$\mathbf{E}\xi = Ep_0^+ + \frac{b_0}{aW}\left(1+\frac{W}{V_1}\right)\left(Ee^{\beta E} - \frac{e^{\beta E}-1}{\beta}\right)p_1^-. \tag{26}$$

If $E=\infty$, then from (25), (26) it follows

$$\begin{aligned} p_1^- &= 1-\frac{bW}{aV_1}, \\ \mathbf{E}\xi &= \frac{bW^2}{a[aU-(a+b)W]}. \end{aligned} \tag{27}$$

These formulas are valid only if $aU > (a+b)W$ (condition of steady-state regime existence).

Let us evaluate now the mean current profit ($\bar{P}$) of all mediators (assuming that $E=\infty$):

$$\bar{P} = \bar{I} - \bar{C},$$

where $\bar{I}$ is mean current incomes and $\bar{C}$ is mean current costs of PT (in steady-state regime of their work). The mean current costs may be expressed by the formula

$$\bar{C} = \sum_{n=1}^{N}(p_n+c_n)W_n + \sum_{n=1}^{N} c_n^{st}\mathbf{E}\xi_n, \tag{28}$$

where p_n is the price for $1t$ of product for n-th mediator; c_n is transportation costs for delivery of $1t$ of product to n-th PT; c_n^{st} is storage expenses per time unit for $1t$ of product at n-th PT.

The mean current income for realization of product at n-th PT equals to

$$\bar{I} = \sum_{n=1}^{N} \Pi_n \bar{U}_n, \tag{29}$$

where Π_n is the price for $1t$ of product for customer; $\bar{U}_n$ is mean processing rate of product's removal from n-th PT.

In steady-state regime according to conservation law for material flows the equality $\bar{U}_n = W_n$ takes place. Hence, from (28) and (29) we get

$$\bar{P} = \sum_{n=1}^{N}\left(\pi_n W_n - c_n^{st}\mathbf{E}\xi_n\right), \tag{30}$$

where $\pi_n = \Pi_n - p_n - c_n > 0$.

The production capacity of enterprise is finite value W_0, therefore

$$\sum_{n=1}^{N} W_n \le W_0. \tag{31}$$

Thus, we arrive at the following one-stage stochastic optimization problem: it is required to find out the positive parameters $W_n, n = 1,2,\ldots,N$, maximizing the objective function (30) under constraint (31).

To solve this problem it is necessary to know dependence of $\mathbf{E}\xi_n$ on control parameter W_n in evident form. In more simple case it is given by the formula (27), i.e.

$$\mathbf{E}\xi_n = \frac{b_n W_n^2}{a_n[a_n U_n - (a_n + b_n)W_n]}, \tag{32}$$

which is valid only under condition

$$(a_n + b_n)W_n < a_n U_n. \tag{33}$$

Substituting (32) in the right-hand side of formula (30) we get

$$\bar{P} = \sum_{n=1}^{N} \left\{ \pi_n W_n - c_n^{st} \frac{b_n W_n^2}{a_n[a_n U_n - (a_n + b_n)W_n]} \right\}. \tag{34}$$

Since

$$\frac{\partial^2 \bar{P}}{\partial W_n^2} = -\frac{2c_n^{st} b_n a_n U_n^2}{[a_n U_n - (a_n + b_n)W_n]^3} < 0$$

(see inequality (33)), the function $\bar{P}$ is concave.

The problem of maximization (34) under condition (31) nay be solved by the Lagrange multipliers method. The Lagrangian function for this problem will be denoted thus:

$$L = \bar{+}\varphi \left(\sum_{n=1}^{N} W_n - W_0 \right),$$

where φ is unknown multiplier. From condition $\partial L/\partial W_n = 0$ we obtain the following equation:

$$(a_n + b_n)W_n^2 - 2a_n W_n U_n + \frac{a_n^3}{b_n} U_n^2 (\pi_n + \varphi) \left[\frac{a_n}{b_n} (\pi_n + \varphi)(a_n + b_n) + c_n^{st} \right]^{-1} = 0.$$

The root of this equation which satisfy the inequality (33) is

$$W_n = \frac{a_n U_n}{a_n + b_n} \left[1 - \sqrt{\frac{b_n c_n^{st}}{a_n(a_n + b_n)(\pi_n + \varphi) + b_n c_n^{st}}} \right], \quad n = 1,2,\ldots,N. \tag{35}$$

The multiplier φ is determined as the unique root of equation

$$\sum_{n=1}^{N} \frac{a_n U_n}{a_n + b_n} \sqrt{\frac{b_n c_n^{st}}{b_n c_n^{st} + a_n(a_n + b_n)(\pi_n + \varphi)}} = \sum_{n=1}^{N} \frac{a_{n_n} U_n}{a_n + b_n} - W_0 \tag{36}$$

where it is assumed that

$$\sum_{n=1}^{N} \frac{a_n U_n}{a_n + b_n} > W_0.$$

Thus, the formulas (35), (36) give the solution to optimization problem.

5 Conclusion

The examples given in the Sects. 2–4 show that apparatus of Markov drift processes is sufficiently convenient and flexible for the purpose of modeling the big variety of industrial and logistical systems. Within the formal frameworks of Markov drift processes it is possible to model the informational and material flows simultaneously and study their interaction. In the future investigations, to our opinion, it would be more effective to use the combination of analytical methods, numerical algorithms, and simulation.

References

1. Anick D, Mitra D, and Sondhi MM (1982) Stochastic theory of data-hadling system with multiple sources. Bell System Techn J 61: 1871-1894
2. Devis MHA (1984) Piecwise-deterministic Markov process: a general class of non-diffusion stochastic models. J Royal Statist Soc 3: 353-388
3. Kosten L (1974) Stochastic theory of multi-entry buffer (II). Delft Progress Report 1: 10-18
4. Kosten L (1986) Liquid models for a type of information storage problems. Delft Progress Report 11:71-86
5. Mitra D (1988) Stochastic theory of a fluid model of producers and consumers coupled by a buffer. Adv Appl Prob 20: 646-676
6. Mitra D and Mitrani I (1991) Communication systems. Queueing Systems Theory Appl 9: Nos 1&2
7. Postan MY (1992) A class of mixed Markov processes and their application to teletraffic theory (in Russian). Probl Peredachi Inf 3:40-53
8. Postan MY (1989) Semi-Markov processes of stock accumulation and their usage (in Russian). Preprint No 89-59, Glushkov Inst Cybern, Kiev
9. Postan MY (2006) Modeling complex systems on the basis of semi-Markov drift processes (in Russian). In: Proc. of the V Intl Conference "System Identification and Control Problems" SICPRO'06, Moscow: pp. 604-617
10. Postan MY (2006) On theory of Markov drift processes under semi-Markov control and its applications. In: Abstract of Communications of 9th Intl Vilnius Conf on probability theory and mathematical statistics, Vilnius: pp. 266-267
11. Postan MY (2006) Economic-mathematical models of multimodal ttransport (in Russian). Astroprint, Odessa
12. Prabhu NU (1998) Stochastic storage processes: queues, insurance risk, dams, and data communications, 2nd Edn. Springer, Berlin Heidelberg New York
13. Schmidli HP (1992) A general insurance risk model. Ph.D. thesis, ETH, Zürich
14. Sevast'yanov BA (1962) Influence of storage bin capacity on the average standstill time of a production line (in Russian). Theory Prob Appl 7: 429-438
15. Wijngaard J (1979) The effect of inter-stage buffer storage on the output of two unreliable production units in series, with different production rates. AIIE Trans 11:42-47

Analysis of Decentral Order-picking Control Concepts

Thorsten Schmidt[1], **Guido Follert**[2]

[1] Fraunhofer Institut für Materialfluss und Logistik, Dortmund
[2] Lehrstuhl für Förder- und Lagerwesen, Universität Dortmund

Abstract A paradigm change in the control logic of material flows is propagated towards self-controlled objects that take decisions decentrally as a reaction to cope with the increasing complexity. However, the technical feasibility has yet to be proven, esp. with regard to in-house material handling. For a successful implementation in future planning rules, the overall system performance as well as adapted system hardware is to be investigated. The paper focuses onto part-to-picker order-picking systems (OPS) with the use of AS/RS, which are typically operated by a central control system. The challenge for decentral control concepts is the formulation of optimal control strategies without information of the current system status. By means of a reference model and simulations runs KPIs and various operating strategies are compared between the two different control philosophies. The analyses yield that decentral control strategies are suitable for the operation of part-to-picker OPS, where they control bin supply to a series of order picking stations that even operate according to the inverse picking principle, i.e. parallel picking at the order picking stations. By means of adapted operating strategies a comparable system performance, measured in picks/hr or picker utilization, can be achieved.

1 Introduction

Decentral control concepts offer large potentials for an improvement of the material flow. This includes shorter setup times for complex systems thanks to the integration of elementary functions into system modules as well as shorter reaction times to load variations. Furthermore, they are a basic component of a self-controlled system where the decision is not only made decentrally but also based on information stored in the goods to be transported (e.g. routing information or local intelligence) [Freitag et al. 2004]. This approach uses simple local actuators which reduce the communication expenditure compared to classical central systems and can be considered as a key method to tackle the increasing dynamics in technical material flow systems.

DOI: 10.1007/978-3-540-76862-3,

Of paramount importance however is the question of optimality of the widespread decision taking of decentral control concepts [Bullinger and ten Hompel 2007].

The use of RFID and the promising developments made in this field will propagate the application of decentral control concepts also for internal material flows [ten Hompel and Lange 2004]. For this reason it will be necessary to study the effects that these control concepts have on order-picking systems (OPS), which in the material flow of distribution centers are, for example, a central link to the customer. In addition to this, order-picking systems are often very complex and, due to the large manpower requirements, represent a central expense factor (cp. [Petersen 2000] and [Tompkins et al. 2003]).

2 Application

This paper focuses on the inverse order-picking. In this highly demanding special case of order-related picking the variety of orders, which are handled in parallel at several picking stations, require the exactly synchronised provision of article bins. Earlier research showed promising results for suitable strategies [Schmidt and Follert 2006].

This study analyses the functionality of central control philosophies in a model-based system and compares it with decentral control logics. Opposing influences are analysed by means of a reference model, which studies the part-to-picker picking with a parallel order execution at a series of picking stations arranged in parallel (see Fig. 1) by means of simulation for reasons of complexity [Roidl and Follert 2007].

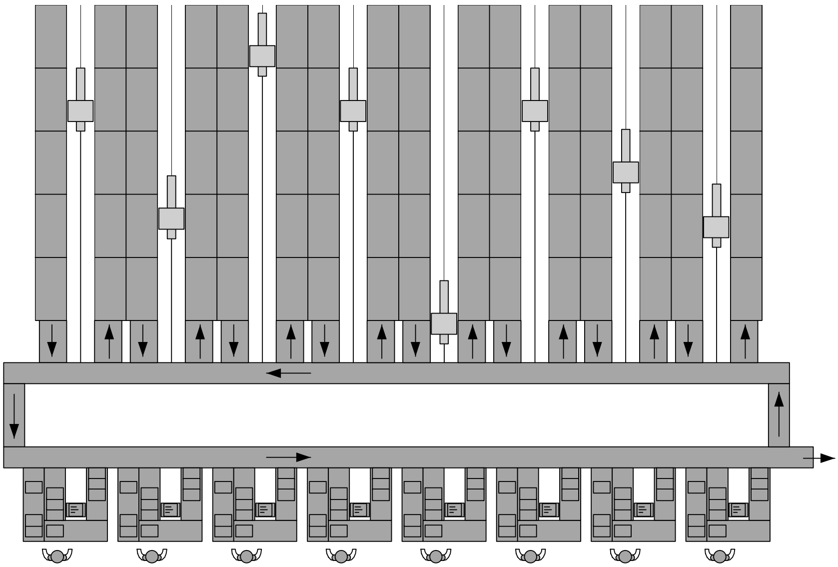


Fig. 1 Layout of the simulation model

Table 1 Primary influential factors and system loads

Influences	system load
Assortment	500/2.000/10.000 articles
Size of orders	
# order lines per order	Exponentially distributed, μ = 5, max. 20
# picks per order line	Uniformly distributed, 1–19; ∅ = 10
# parallel picked orders	1/6/12
Capacity article bins	100 parts/bin
Article structure	20 % A-articles ∼ 80 % OLs, 10 % B ∼ 15 % OLs and 70 % C ∼ 5 % OLs
# order bins per order	Exponentially, $\mu = 3$ bins/order, max. 1 bin/OLs
Performance of order pickers	6 s/pick
Capacity of branches	infeed track: 10 bins and output track: 5 bins

The presented system consists of 8 picking stations and 8 AS/RS, which supply the picking stations with article bins via a conveyor loop (see Fig. 1). In the model this loop has a length of 104 m and a window control with a partition of 4 m for one bin each. Assuming at least one empty bin, the maximum load is 25 bins. The conveying speed is 1 m/s and determines the reaction speed of the conveyor system. There are several functions ensuring the supply and return of order bins of/from the loop, which are not represented in detail to maintain the experimenting flexibility. Each picking station has a bin infeed and bin output tracks for completed bins. At each station a maximum of 12 bins can be handled at the same time whereas an eventual increase of local way times is not considered because it is of no interest for this quantity.

The relevant influential factors and system loads are shown in Table 1. Some parameters, like the picking performance, are no system loads in the classical sense but influence the system output and are thus included in the list of parameters. All values represent typical quantities as they may occur in practice.

3 Control strategies

During the realisation of central as well as decentral control systems a large variety of optimisation processes and rules may apply. To compare the efficiency of central and decentral logics at first the basic structure of the models is tested without a special optimisation algorithm.

The central control acts based on an awareness of the facility status. It mainly has to decide upon the allocation of an article bin to one of the eight picking stations. This decision is made at the central identification point preceding the first picking station, always according to an established strategy. Possible strategies are an allocation according to the order of picking stations within the loop (RoundRobin), the remaining capacity of the infeed (RemainingCapacity) or the minimum quantity of open order lines or picks (MinimumWork). In their study the authors identified the strategy “RemainingCapacity” as the most suitable one and it will set the basis for the description of a central control without further validations.

The decentral control, on the other hand, checks in front of each single picking station and for each arriving bin if

- the article is required at the picking station and
- at least one buffer slot is available at the infeed track.

If the result is positive the bin will be fed in. Thus, the central control can make its decision only at the decision point. In the worst case (demand at several stations) the bin may cycle repeatedly through the loop. In a decentral control, on the other hand, in the best case a bin is moved from the output track of picking station n to the infeed track of station $n+1$.

Irrespective thereof, both controls check if a required article bin is already circulating and contains a sufficient number of articles (quantity reservation). Only if this is not the case articles are taken from storage. In both cases the orders are managed centrally.

4 Experiments

Each parameter combination of the rough results of the discrete-event simulation model was determined during 10 independent simulation runs of 12 hours with varying random numbers. The first two hours of a test represent the start-up time of the simulated system and are not included in the results. The results are very accurate as is demonstrated by the fact that the width of the confidence interval on a 95% confidence level is at maximum 3.82% of the mean of the performance indicator “picker utilisation”.

4.1 Evaluation

The main factors for an evaluation of the procedure efficiency are the Workload of the pickers, which in this case is defined by the sufficient availability of work, i.e. of article bins, and the Number of finished orders, which is closely related to the workload of pickers with just different stochastic influences.

The order lead time of both control principles increases analogously to the number of executed orders in parallel and is defined by the average picking workload. No distinctive features occurred during the studies. The results of a direct comparison of the tests are shown in Fig. 2. The performance of both control principles can easily be compared. The decentral control is slightly ahead in all combinations (+1.9 %). However, a uniform tendency towards certain configurations cannot be seen.

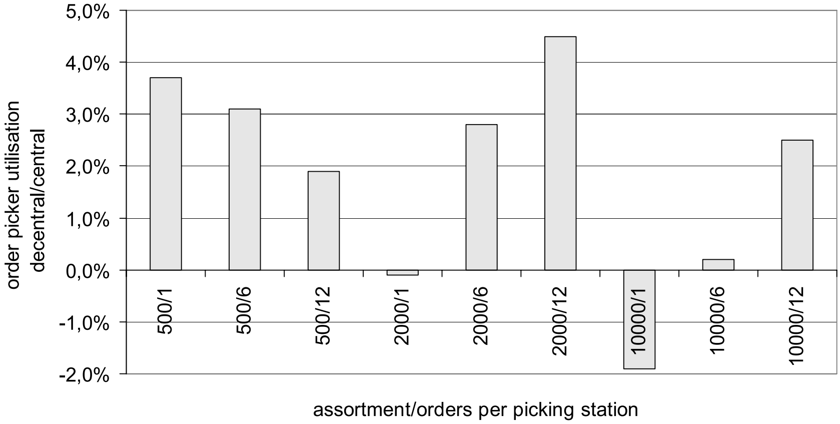


Fig. 2 Performance comparison decentral/central with basic tests

When the non-optimised basic strategy is used, the performance values vary considerably and fall in line with a growing assortment (see Fig. 3). At first, this effect is quite surprising because in case of large assortments the picking performance is principally almost independent on the availability of single article bins. According to an analysis during the parallel execution of orders a high system load correlates with a high frequency of the infeed tracks to the picking stations. As further analyses have shown the low load of the infeed tracks is marked by short peaks at a low average load (see Fig. 5, upper graphic). Since for both assortments fewer grips are made onto the moved inventory and thus the share of bins supplied from store increases, the efficiency of the replenishment control is the critical factor and in the next section will be influenced by adjusted strategies.

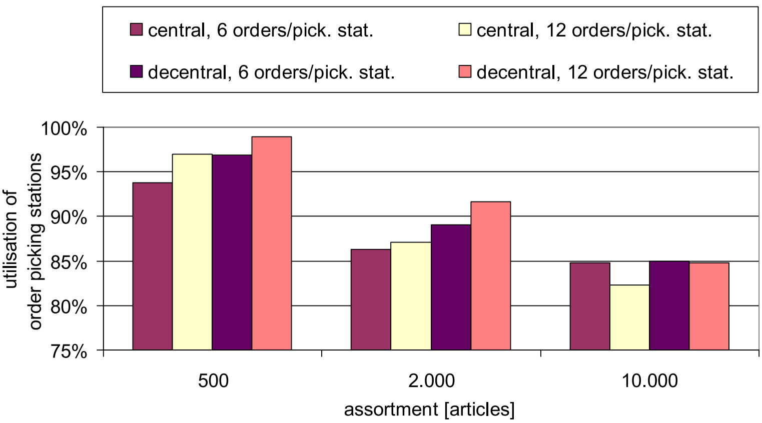


Fig. 3 Basic strategies; performance in case of different assortment volumes

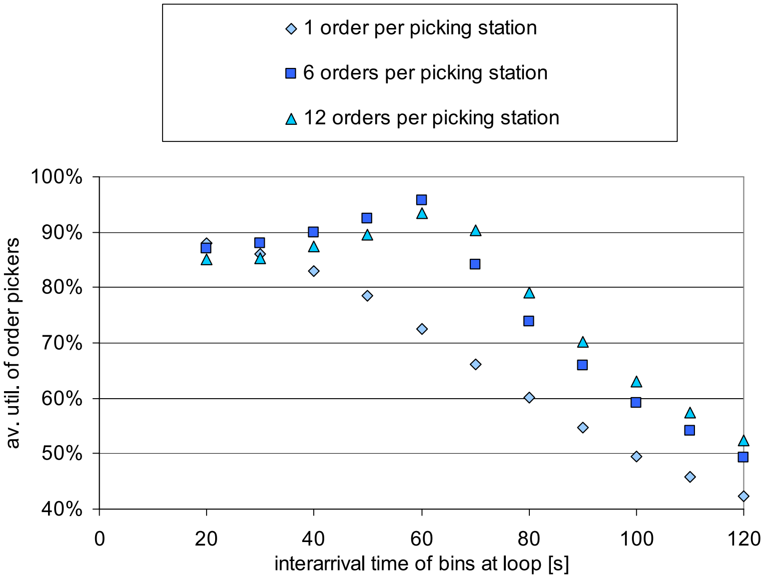


Fig. 4 System performance with controlled bin infeed

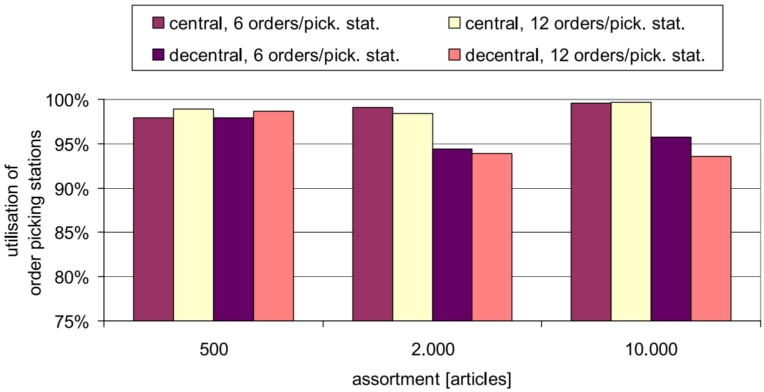


Fig. 5 System performance using controlled bin infeed

Furthermore, the performance of a series execution (1 order/station) is very low (average 84.2%) because the start of a new order and thus the request for a new order bin is directly linked to the termination of the preceding order. This leads to idle times because the new article bins have to be replenished with goods from store via the loop.

The critical factors for both controls are thus the feeding of the loop when new orders are entered and the serial and parallel execution of orders.

4.2 Improvement by strategies

According to the results of the analyses described above the following improvement strategies aim at the smooth and even utilisation of the loop and the avoidance of congestions of the infeed lines of the order picking stations.

The central control strategy uses the RemainingCapacity (RemCap) assignment rule and thus already employs a strategy for the even distribution of bins to the stations. In case of a decentral control such a strategy does not exist. The assignment (I) is activated according to the conditions: “Demand exists” ($Y > 0$) and “buffer available” ($Q \geq 1$), i.e. $Y > 0 \wedge Q \geq 1 = I$. Due to the fact that the stations are arranged in sequence, the early stations are better serviced than the later ones. Thus an advanced assignment rule is defined, which controls the required amount and the queue at the infeed line. It is tested, whether the bin amount (A) is equal or less than the required amount of a station (Y) (i.e. $A \leq Y$) and whether or not the queue capacity (Q) exceeds threshold Γ (i.e. $Q > \Gamma$). In addition, repeated loop circulations (R) are prevented by checking $R > 1$. Thus, a bin is diverted if the condition $Y > 0 \wedge Q \geq 1 \wedge (A \leq Y \vee R > 1) \vee Q > \Gamma, \Gamma = 1,3,5$ is true.

A control of the bin infeed stream from storage to the loop can easily be realised in case of a central control by definition of a maximum number of bins that are released to the system when invoking a new customer order. As a rule, fewer bins are released than the available buffer capacity of the infeed lines (in this model: 2 units difference). After that, new bins are released only after successful divert of a bin at the station.

In case of a decentral control such an algorithm is by definition not possible, since there is no central booking of the performed diverts in the system. As an alternative, a time-based control is established. The infeed of bins from storage to the loop is not done in batch sizes, but one-by-one with a fixed arrival rate in between. The adjustment of this arrival rate is of paramount importance to the system behaviour. It has an immediate influence to the system start and may reduce the system performance, especially for stations with little system load (serial order picking with only one order/station).

The adjustment of the arrival rate follows the rules of the order picking process. It is assumed that the ideal arrival rate depends on the average process time of bins in the picking station. In the given case, assuming an average picking time of 6 s and an average picking amount of 10 parts per order line, the arrival rate is set to 60 s. For validation reasons arrival rates between 20 s and 120 s have been tested. The results confirm this hypothesis and yield to a maximal picker utilization in case of the parallel order execution. However, any arrival rate will reduce the picking performance for serial order execution (as above: serial order picking with only one order/station).

The advanced assignment rule for decentral control based on order amount and buffer capacity has varying effects on the system performance. Varying threshold levels Γ cause irregular performance increases and decreases of picker utilisation

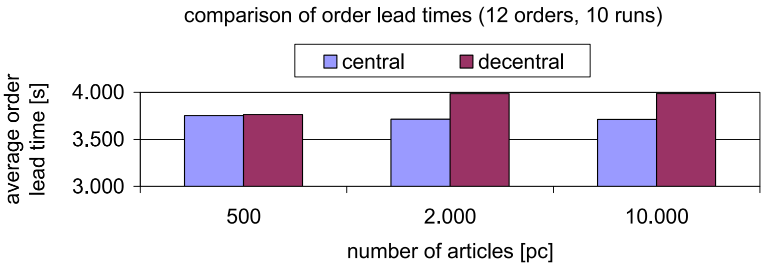


Fig. 6 Occupancy of infeed track no.1; 10 000 articles; 12 orders/station

between +2.3 % and −1.9 %. A positive impact is generated only for small article ranges (500 articles) and serial order execution. Due to this, the rule is not employed for further comprehensive analyses.

Both infeed control strategies (counter and time-based) show positive utilisation of the infeed lines and thus an increase of picker utilisation (see Fig. 5) as well for central and decentral control. Last but not least analysis on the average order lead time shows on the opposite an increasing variance in average lead time especially for large article ranges (see Fig. 6).

5 Conclusion

The experiments have proven that decentral control strategies are suitable for the operation of part-to-picker OPS, where they control bin supply to a series of order picking stations that even operate according to the inverse picking principle, i.e. parallel picking at the order picking stations. The generated output in terms of system performance, measured in picks/hr or picker utilisation, is close to a central control strategy, subject to the condition that suitable improvement strategies are being utilised. Due to the similar performance results, in the future the decision of choosing a central or decentral control system for a certain application can be decided on an investment comparison only. However, the details of the technical availabilities are not within the scope of this analysis.

For future research, the suitability of decentral control concepts shall be tested for large scale order picking systems, which operate according to the inverse picking principle. A high number of picking stations shall be connected with complex transport nets that include short-cuts and refeeds for the minimisation of travelling paths.

References

Bullinger, H.-J.; ten Hompel, M. [eds.]: Internet der Dinge. Springer, Berlin, 2007

Freitag, M.; Herzog, O.; Scholz-Reiter, B.: Selbststeuerung logistischer Prozesse – Ein Paradigmenwechsel und seine Grenzen. In: Industriemanagement 20 (2004), Nr. 1, pp. 23–27

Petersen, C.G.: An evaluation of order picking policies for mail order companies. Production and Operations Management 9 (2000), Nr. 4, pp. 319–335

Roidl, M.; Follert, G.: Simulation von multiagentenbasierten Materialflusssteuerungen. In: INFORMATIK 2007 - Informatik trifft Logistik, Beiträge der 37. Jahrestagung der Gesellschaft für Informatik e.V. (GI), 24.–27. September 2007 in Bremen, GI Proceedings 109, Band 1

Schmidt, T.; Follert, G.: Effects of a decentral control on the performance of part-to-picker order-picking systems. In: Progress in Material Handling Research, The Material Handling Institute, 2006

ten Hompel, M. (editor); Lange, V. (editor): RFID ¬ Logistiktrends für Industrie und Handel. Praxiswissen, Dortmund 2004

Tompkins, J.A, White, J.A., Bozer Y.A., Tanchoco, J.M.A.: Facilities Planning. John, Wiley & Sons, New York, 2003