IFIP AICT 553

Luis M. Camarinha-Matos
Ricardo Almeida
José Oliveira
(Eds.)

Technological Innovation for Industry and Service Systems

10th IFIP WG 5.5/SOCOLNET
Advanced Doctoral Conference on Computing,
Electrical and Industrial Systems, DoCEIS 2019
Costa de Caparica, Portugal, May 8–10, 2019
Proceedings

Springer

IFIP Advances in Information and Communication Technology 553

IFIP – The International Federation for Information Processing

IFIP was founded in 1960 under the auspices of UNESCO, following the first World Computer Congress held in Paris the previous year. A federation for societies working in information processing, IFIP's aim is two-fold: to support information processing in the countries of its members and to encourage technology transfer to developing nations. As its mission statement clearly states:

IFIP is the global non-profit federation of societies of ICT professionals that aims at achieving a worldwide professional and socially responsible development and application of information and communication technologies.

IFIP is a non-profit-making organization, run almost solely by 2500 volunteers. It operates through a number of technical committees and working groups, which organize events and publications. IFIP's events range from large international open conferences to working conferences and local seminars.

The flagship event is the IFIP World Computer Congress, at which both invited and contributed papers are presented. Contributed papers are rigorously refereed and the rejection rate is high.

As with the Congress, participation in the open conferences is open to all and papers may be invited or submitted. Again, submitted papers are stringently refereed.

The working conferences are structured differently. They are usually run by a working group and attendance is generally smaller and occasionally by invitation only. Their purpose is to create an atmosphere conducive to innovation and development. Refereeing is also rigorous and papers are subjected to extensive group discussion.

Publications arising from IFIP events vary. The papers presented at the IFIP World Computer Congress and at open conferences are published as conference proceedings, while the results of the working conferences are often published as collections of selected and edited papers.

IFIP distinguishes three types of institutional membership: Country Representative Members, Members at Large, and Associate Members. The type of organization that can apply for membership is a wide variety and includes national or international societies of individual computer scientists/ICT professionals, associations or federations of such societies, government institutions/government related organizations, national or international research institutes or consortia, universities, academies of sciences, companies, national or international associations or federations of companies.

More information about this series at http://www.springer.com/series/6102

Luis M. Camarinha-Matos ·
Ricardo Almeida · José Oliveira (Eds.)

Technological Innovation for Industry and Service Systems

10th IFIP WG 5.5/SOCOLNET
Advanced Doctoral Conference on Computing, Electrical and Industrial Systems, DoCEIS 2019
Costa de Caparica, Portugal, May 8–10, 2019
Proceedings

Springer

Editors
Luis M. Camarinha-Matos
NOVA University of Lisbon
Monte Caparica, Portugal

Ricardo Almeida
NOVA University of Lisbon
Monte Caparica, Portugal

José Oliveira
NOVA University of Lisbon
Monte Caparica, Portugal

ISSN 1868-4238 ISSN 1868-422X (electronic)
IFIP Advances in Information and Communication Technology
ISBN 978-3-030-17770-6 ISBN 978-3-030-17771-3 (eBook)
https://doi.org/10.1007/978-3-030-17771-3

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This proceedings volume, which collects selected results produced in engineering doctoral programs, focuses on research and development in the field of technological innovation for industry and service systems. The industry and service sectors are undergoing profound transformation towards digitalization and integration of new levels of "smartness." The idea of a fourth industrial revolution, represented by terms such as "industry 4.0," "smart manufacturing," and "economy 4.0" are an expression of such transformation. This movement is characterized by an increasing digitalization and interconnection of systems, products, value chains, and business models. The interconnection between the physical and the cyber worlds – cyber-physical systems and Internet of Things – and the integration of the so-called exponential technologies, are central features of this innovation trend.

The DoCEIS series of advanced Doctoral Conferences on Computing, Electrical and Industrial Systems aims at providing a venue for sharing and discussing ideas and results from doctoral research in these inter-related areas of engineering, while promoting a strong multidisciplinary dialog. Furthermore, the conference aims at creating collaborative opportunities for young researchers while also collecting valuable feedback from colleagues in a welcoming environment. As such, participants were challenged to look beyond their specific research question and relate their work to the selected theme of the conference, namely, to identify in which ways their research topics can benefit from or contribute to industry and service systems. Current trends in strategic research programs are confirming the fundamental role of multidisciplinary and interdisciplinary approaches in innovation. More and more funding agencies are including this element as a key requirement in their calls for proposals. In this context, the challenge proposed by DoCEIS is a contribution to the process of acquiring such skills, which are mandatory in the profession of a PhD.

The tenth edition of DoCEIS, which was sponsored by SOCOLNET, IFIP, and IEEE IES, attracted a good number of paper submissions from a large number of PhD students and their supervisors from 14 countries. This book comprises the works selected by the international Program Committee for inclusion in the main program and covers a wide spectrum of application domains. As such, research results and ongoing work are presented, illustrated, and discussed in areas such as:

- Collaborative systems
- Collaboration and resilient systems
- Decision and optimization systems
- Assistive systems
- Smart environments
- Smart manufacturing
- Water monitoring systems
- Communication systems
- Energy systems

We hope this collection of papers will provide readers with an inspiring set of promising ideas and new challenges, presented in a multidisciplinary context, and that by their diversity these results can trigger and motivate richer research and development directions.

We would like to thank all the authors for their contributions. We also appreciate the efforts and dedication of the DoCEIS international Program Committee members, who both helped with the selection of articles and contributed with valuable comments to improve the quality of the papers.

March 2019

Luis M. Camarinha-Matos
Ricardo Almeida
José Oliveira

Organization

10th IFIP/SOCOLNET Advanced Doctoral Conference on Computing, Electrical and Industrial Systems
Costa de Caparica, Portugal, 8–10 May 2019

Conference and Program Chair

Luis M. Camarinha-Matos	NOVA University of Lisbon, Portugal

Organizing Committee Co-chairs

Luis Gomes	NOVA University of Lisbon, Portugal
João Goes	NOVA University of Lisbon, Portugal
João Martins	NOVA University of Lisbon, Portugal

International Program Committee

Antonio Abreu, Portugal
Andrew Adamatzky, Poland
Vanja Ambrozic, Slovenia
Amir Assadi, USA
Olga Battaia, France
Luis Bernardo, Portugal
Andrea Bottino, Italy
Erik Bruun, Denmark
Barbora Buhnova, Czech Republic
Giuseppe Buja, Italy
Luis M. Camarinha-Matos, Portugal
Roberto Canonico, Italy
Laura Carnevali, Italy
Wojciech Cellary, Poland
Alok Choudhary, UK
Noelia Correia, Portugal
Luis M. Correia, Portugal
Jose de la Rosa, Spain
Stefano Di Carlo, Italy
Florin G. Filip, Romania
Maria Helena Fino, Portugal
José M. Fonseca, Portugal
Adriana Giret, Spain
João Goes, Portugal
Luis Gomes, Portugal
Antoni Grau, Spain
Paul Grefen, The Netherlands
Michael Huebner, Germany
Ricardo Jardim-Gonçalves, Portugal
Vladimir Katic, Serbia
Asal Kiazadeh, Portugal
Hans-Jörg Kreowski, Germany
Bernadetta Kwintiana, Germany
Zbigniew Leonowicz, Poland
Niels Lohse, UK
Marin Lujak, France
João Martins, Portugal
Rui Melicio, Portugal
Paulo Miyagi, Brazil
Renato Moraes, Brazil
Filipe Moutinho, Portugal
Horacio Neto, Portugal

Rodolfo Oliveira, Portugal
Luis Oliveira, Portugal
Joaquin Ordieres, Spain
Angel Ortiz, Spain
Gordana Ostojic, The Netherlands
Peter Palensky, Austria
Marcin Paprzycki, Poland
Luis Palma, Portugal
Nuno Paulino, Portugal
Pedro Pereira, Portugal
Duc Pham, UK
João Pimentão, Portugal
Paulo Pinto, Portugal
Armando Pires, Portugal
Ricardo J. Rabelo, Brazil
Rita Ribeiro, Portugal
Juan Rodriguez-Andina, Spain
Enrique Romero-Cadaval, Spain
Carlos Roncero, Spain
Thilo Sauter, Austria
Eduard Shevtshenko, Estonia
Thomas Strasser, Austria
Kleanthis Thramboulidis, Greece
Damien Trentesaux, France
Antonios Tsourdos, UK
Manuela Vieira, Portugal
Ramon Vilanova, Spain
Soufi Youcef, France
Tamus Zoltán Ádám, Hungary

Organizing Committee (PhD Students)

Duarte Alemão, Portugal
Ricardo Almeida, Portugal
David Borges, Portugal
João Carvalho, Portugal
Nuno Correia, Portugal
João Gião, Portugal
José Lima d'Oliveira, Portugal

Technical Sponsors

Society of Collaborative Networks

IFIP WG 5.5 COVE
Co-Operation infrastructure for Virtual Enterprises and electronic business

IEEE−Industrial Electronics Society

Organizational Sponsors

Organized by: PhD Program on Electrical and Computer Engineering, FCT - NOVA University of Lisbon

Contents

Optimization Systems

Assistive Systems

Smart Environments

Smart Manufacturing

Water Monitoring Systems

Communication Systems

Energy Markets

Energy Control

Power Systems

Collaborative Systems

A Model to Assess Collaboration Performance in a Collaborative Business Ecosystem

Paula Graça[1,2(✉)], Luís M. Camarinha-Matos[1(✉)], and Filipa Ferrada[1(✉)]

[1] Faculty of Sciences and Technology and Uninova CTS, NOVA University of Lisbon, Campus de Caparica, 2829-516 Caparica, Portugal
{cam, faf}@uninova.pt
[2] Instituto Superior de Engenharia de Lisboa, Instituto Politécnico de Lisboa, Rua Conselheiro Emídio Navarro 1, 1959-007 Lisbon, Portugal
mgraca@deetc.isel.pt

Abstract. In a Collaborative Business Ecosystem, Performance Indicators are a useful mechanism to assess collaboration performance, inducing self-adjustment in organization's profile, thereby improving the sustainability of the ecosystem. Using system dynamics and agent-based modelling, a simulation model is assembled to show organizations' self-adjustment by improving their profile in response to an assessment through a chosen set of performance indicators, such as innovation indicator, contribution indicator and prestige indicator. The natural reaction of organizations (similar to individuals) towards improving their performance according to the way they are evaluated, is modelled considering different enterprise profiles categorized into various classes of responsiveness, to better simulate the diversity in a real collaborative business ecosystem. Preliminary results of this approach are presented and discussed.

Keywords: Collaborative Networks · Business ecosystem · Performance indicators · System dynamics · Agent based modelling

1 Introduction

Business ecosystems are crossing a new era anchored in more interconnected and powerful digital platforms, following the dynamics and trends highlighted in [1]. The term Business Ecosystem was first introduced by Moore as a metaphor inspired by ecological ecosystems [2]. On the other hand, the research area of Collaborative Networks (CN) [3], which has a wider scope, classifies a business ecosystem as a subclass of the CN taxonomy [4]. As such, and in order to emphasize the collaborative dimension, the term Collaborative Business Ecosystem (CBE) has been adopted [5], and a CBE was modelled as an environment of agents (representing the organizations), which collaborate to accomplish business opportunities [6].

In this work it is assumed that the assessment of the collaboration benefits of a CBE, using a set of Performance Indicators (PIs) such as the ones proposed in [6] and [7], influence the behaviour of the agents inducing some degree of self-adjustment of their profile, leading to an improvement of their individual performance and of the CBE

Published by Springer Nature Switzerland AG 2019
L. M. Camarinha-Matos et al. (Eds.): DoCEIS 2019, IFIP AICT 553, pp. 3–13, 2019.
https://doi.org/10.1007/978-3-030-17771-3_1

as a whole, promoting its sustainability. The main purpose of the paper is to present an approach to allow verifying the mentioned assumption.

The remaining sections are organized as follows: section two briefly positions this work in the context of innovation in industry and service systems; section three describes the proposed model to assess the CBE and describes the profile of organizations; section four presents one performance indicator used to illustrate the assessment; section five presents the experimental evaluation of the simulation model using some parametrized scenarios, and including a discussion of results. The last section summarizes the contributions and identifies further research directions.

2 Relationship to Innovation in Industry and Services

Business ecosystems have evolved with the continued increasing in digitalization and interconnection of systems, from traditional industrial sectors to digital business ecosystems supported by computer networks and collaboration platforms. This transition may entail substantial improvements for society and economy, also enabling the shift of business ecosystems towards the age of Industry 4.0, particularly in the Collaborative Industry 4.0 [8].

The essence of the Industry 4.0 concept "*the leap from digital back to physical*", "*a state in which manufacturing systems and the objects they create, communicate, analyse and use that information to drive further intelligent action back in the physical transition*" [9], coupled with the concept of Cyber-Physical Systems (CPS), enable the implementation of more effective mechanisms to measure the performance of a CBE and of its individual organizations. Thus, the real time measurements to determine the performance indicators could facilitate the decision-making processes of organizations regarding selection of partners to collaborate in order to achieve more innovative, productive and value-added solutions, contributing to a more sustainable ecosystem.

3 A Simulation Model to Evaluate the Collaboration of a CBE

The objective of a model to evaluate collaboration within a CBE is to be able to study the evolution of the behaviour of its members, when influenced by the measurement of performance indicators, thus allowing to check how these indicators contribute (or not) to the improvement of performance and collaboration sustainability. As such, a simulation model designated by PAAM (Performance Assessment and Adjustment Model), depicted in Fig. 1, was proposed in [6]. The CBE is modelled as an environment containing a set of agents representing the Organizations (O_i), which collaborate by creating Virtual Organizations (VO_i) in response to market opportunities, designated as Collaboration Opportunities ($CoOp_i$). These $CoOp_i$ are represented in the model by links between agents with weights w_{ij}, meaning the number of times an agent O_i collaborated with another O_j.

Collaboration can be assessed using the set of performance indicators such as those proposed in [6] and [7]: the Innovation Indicator (II), to evaluate the capability of the organizations to create new patents, services or products; the Contribution Indicator (CI), to evaluate the value generated by organizations, creating or accepting collaboration opportunities; and the Prestige Indicator (PI), to evaluate the prominence of a particular organization over others, to participate in collaboration opportunities. It is assumed that such assessment through the given performance indicators will influence the behaviour of organizations causing their self-adjustment, trying to "look better" in face of the used metrics, resulting in an improvement of the ecosystem as a whole.

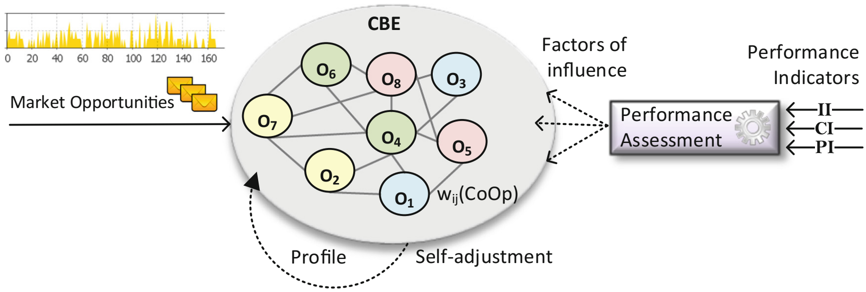


Fig. 1. PAAM (Performance Assessment and Adjustment Model) for a CBE.

The organizations are characterized by different profiles categorized into several classes of responsiveness, to better capture the diversity of a business ecosystem. In the current version of the simulation model, the following classes are considered as described in Table 1: The Social, Selfish, Innovator, and Crook. Each class expresses the way organizations collaborate (or not) in response to market opportunities, using a composition of three parameters (contact rate, acceptance rate and new products rate) with decimal values ranging from 0 to 1, to produce different collaboration behaviour. For instance, social organizations as opposed to selfish, are more likely to contact other organizations to collaborate and are likely to accept more invitations. On the other hand, innovator organizations tend to accept more opportunities involving the creation of new patents or products. Finally, a few crook or opportunistic organizations are also considered to make the simulation model more realistic. The values in Table 1 are only for illustration and can be fine-tuned for each scenario. For each CBE a certain number of agents of each class will be considered.

Table 1. Characterization of the classes of responsiveness of organizations.

Classes of responsiveness of organizations				
Parameters [0..1]	Social	Selfish	Innovator	Crook
Contact rate	0,8	0,1	0,4	0,1
Acceptance rate	0,6	0,2	0,5	0,1
New products rate	0,2	0,2	0,9	0,1

The importance (weight) given to each performance indicator by the CBE manager will result in a set of influence factors that will change the internal behaviour model of agents.

4 The Contribution Indicator to Evaluate the CBE

The Contribution Indicator (CI) is the one chosen among the performance indicators proposed in [6] to illustrate the experimental evaluation, using a simulation model of a CBE. The CI measures the contribution of the organizations to create value in the CBE using the metrics described in Table 2.

Table 2. Description of the metrics used to calculate the Contribution Indicator.

Metrics of the Contribution Indicator (CI)	
Metric	Description
$O_1, \ldots, O_n$	Organizations in the CBE
#O	No. of organizations in the CBE
$\#CoOp_i$ in	No. of collaboration opportunities the organization O_i gained from the CBE
$\#CoOp_i$ out	No. of collaboration opportunities the organization O_i brought in the CBE

The value of CI_i for each individual organization O_i, is calculated by the weighted indegree centrality $C_D(O_i)in$ and the weighted outdegree centrality $C_D(O_i)out$ as expressed in [10] and [11], which correspond, respectively, to the sum of the collaboration opportunities received in and sent from each organization, according to formulas (1) and (2). The resulting $CI_i in$ and $CI_i out$ are normalized values between 0 and 1, because are divided respectively by the maximum weighted indegree centrality, $C_D(O^*)in$, and maximum weighted outdegree centrality, $C_D(O^*)out$.

$$CI_i\, in = \frac{C_D(O_i)in}{C_D(O^*)in} = \frac{\sum_j O_{ij}\#CoOp_{ij}\, in}{C_D(O^*)in} \quad (1)$$

$$CI_i\, out = \frac{C_D(O_i)out}{C_D(O^*)out} = \frac{\sum_j O_{ij}\#CoOp_{ij}\, out}{C_D(O^*)out} \quad (2)$$

The value of contribution for the whole ecosystem, CI_{CBE}, is evaluated by $CI_{CBE}t$ and $CI_{CBE}d$. The $CI_{CBE}t$, calculated by formula (3), is a ratio of the total number of collaboration opportunities created in the CBE by the total number of organizations.

$$CI_{CBE}t = \frac{\sum_i \#CoOp_i}{\#O} \quad (3)$$

The $CI_{CBE}d$, calculated by formula (4), is the centrality degree of the CBE according to [10] and [11], which gives the degree to which the most active organization exceeds the contribution of the others.

$$CI_{CBE}d = \frac{\sum_i [C_D(O^*) - C_D(O_i)]}{\max \sum_i [C_D(O^*) - C_D(O_i)]} \tag{4}$$

where $C_D(O_i)$ is the weighted indegree centrality of the organization O_i, $C_D(O^*)$ is the largest value of $C_D(O_i)$ for any organization in the CBE and in the denominator is the maximum possible sum of differences in organization centrality for the CBE. According to [10], the maximum sum of differences is achieved for a star or a wheel graph (the most centralized graphs), which in an unweighted undirect graph with n nodes, is calculated by (n − 2) pairs of nodes multiplied by (n − 1) maximum links. In weighted direct graphs, the maximum sum of differences is given by the maximum weighted input (maximum indegree of all organizations) multiplied by ($\#O - 1$) maximum links, allowing the reformulation of (4) resulting in formula (5).

$$CI_{CBE}d = \frac{\sum_i [C_D(O^*)in - C_D(O_i)in]}{C_D(O^*)in \ * \ (\#O - 1)} \tag{5}$$

In formula (5), the indegree is used instead of outdegree, because it better represents the importance of the organizations in the CBE. Degree-based measures of graph centrality reflects the relative dominance of a single point [10]. Thus, an organization with a high indegree, means that it was invited to collaborate by many others, which is more relevant than a high outdegree (organizations invited). Similarly, a scientific paper that is cited by many others has high relevance. The resulting $CI_{CBE}d$ is a normalized value between 0 and 1, meaning that if $CI_{CBE}d = 0$, all organizations have equal relevance, and $CI_{CBE}d = 1$ if there is an organization, O^*, that completely dominates the CBE with respect to centrality.

5 Experimental Evaluation of the CBE

The proposed PAAM simulation model for the experimental evaluation of a CBE, was developed with the AnyLogic Multimethod Simulation Software [12], using agent-based modelling (ABM), system dynamics (SD), and discrete elements (state-charts, events, timers, etc.).

The CBE is composed of an environment of agents (representing the organizations), which have a certain behaviour, simulated through a state-chart. Organizations are characterized by a profile classified according to the mentioned classes of responsiveness (Social, Selfish, Innovator and Crook) that define how they respond to market opportunities.

A market opportunity is also modelled as an agent including a task description, a number of resources needed to accomplish the task, a duration expressed in days and a flag indicating if it is a new product. For the current experiment, and for illustration purposes, new opportunities are generated by following discrete uniform distributions suitable to better simulate the randomness and diversity of market demand. When a new market opportunity is created, the number of resources follows a discrete uniform distribution, which generates values with equal probability between 1 and a parametrized maximum number of resources. The duration also follows a discrete uniform distribution between 1 and a parametrized maximum number of days. Finally, the flag indicating if it is a new product, follows a Bernoulli distribution [13] with probability p = 10% of total generated market opportunities in the time window of the simulation model. This is a discrete probability distribution of any single experiment that asks a yes–no question, whose result is "yes" with probability p and "no" with probability (1 – p). The Bernoulli distribution is adequate since it allows to randomly generate in the CBE, a certain percentage (p) of new patents or products from the total number of market opportunities.

Responding to the incoming market opportunities, organizations interact creating collaboration opportunities by inviting and accepting invitations from other organizations in the CBE. The interactions are registered in the model by weighted links that represent the number of collaboration opportunities each organization received from or sent to others.

Figure 2 illustrates the PAAM simulation model, showing a CBE environment after a period of one year, parametrized with 6 social organizations, 5 selfish, 3 innovative and 1 crook. In this example, market opportunities arrive at a rate of 1 thousand per year using the Poisson distribution [14], a discrete probability distribution quite popular for modelling the number of times an event occurs in an interval of time or space. This distribution allows to simulate a parameterized number of market opportunities, arriving randomly to the CBE during the time of the experiment.

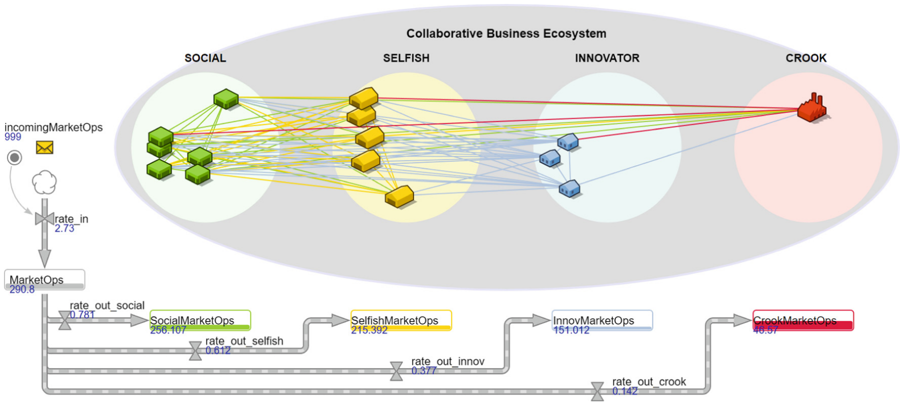


Fig. 2. PAAM simulation model, showing the CBE environment.

As mentioned, the organizations in the CBE, are autonomous agents characterized by a profile (contact rate, acceptance rate and new products rate) according to their classes of responsiveness. Figure 3 shows the example of a social organization whose behaviour is modelled by a state-chart and by two stocks of resources to perform the tasks associated with the incoming market opportunities. Stocks manage the flow of available and allocated resources, which after being allocated and after the execution time of tasks, are released to accept new ones. When an organization receives a new market opportunity, based on its class and needed resources for the task, it distributes business units inviting other organizations to collaborate, using formula (6).

$$units_{to_{distribute}} = binomial(contactRate, nResources) \tag{6}$$

The binomial distribution used in (6), is a discrete probability distribution with parameters p = *contactRate* and n = *nResources*, which gives the number (bounded by [0, n]) of successes in a sequence of n independent experiments, each asking a yes-no question. High values of p give greater probability to get higher values of n. Thus, this distribution allows the simulation of the behaviour of the agents depending on their class of responsiveness, that is, the higher the contact rate of an organization, the greater the probability of distributing more business units.

The organizations that receive the invitation, accept it or not based on the organization's class and available resources, using formula (7).

$$\begin{aligned} & accept_{to_{collaborate}} \\ & = [bernoulli(acceptanceRate) || bernoulli(newProductsRate)] \\ & \qquad \&\&\ availableResources \end{aligned} \tag{7}$$

The Bernoulli distribution [13] used in (7), is a discrete probability distribution of any single experiment that asks a yes–no question, whose result is "yes" with probability p = *acceptanceRate* or p = *newProductsRate*, and "no" with probability (1 – p).

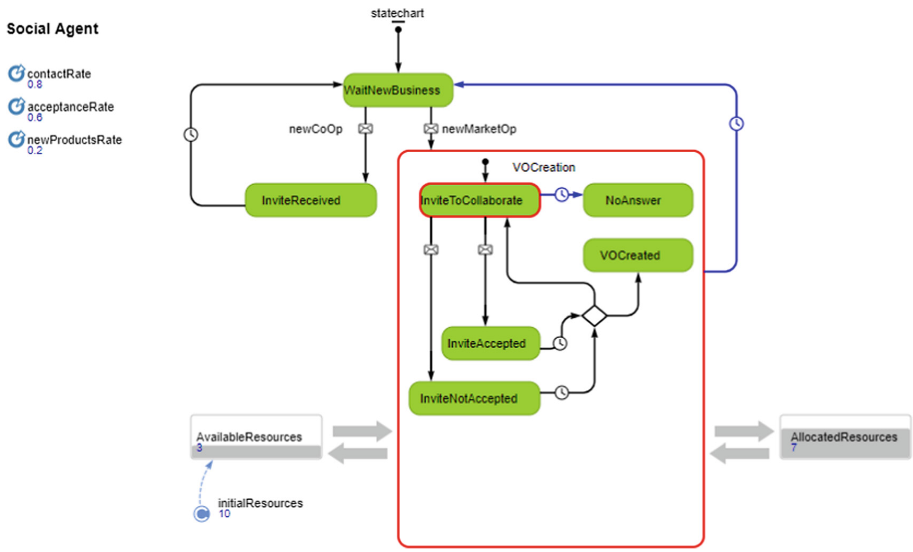


Fig. 3. Zoom-in of a social agent, showing its behaviour modelled by a state-chart and system dynamics.

Thus, this distribution can be used to simulate the behaviour of the agents according to their class of responsiveness. The higher the acceptance rate or the new product rate of an organization, the greater the likelihood of accepting the invitation to collaborate if the required resources are available.

After running the model, the CI is calculated for each organization and for the whole ecosystem, resulting in the values displayed in Table 3. The column $CI_i in$ shows the normalized value of collaboration opportunities that each organization received from others, accepting invitations to collaborate. On the other hand, the column $CI_i out$ shows the normalized value of collaboration opportunities that each organization created in the CBE, inviting other organizations to collaborate.

As expected, it can be verified in the results of Table 3, that the social organizations have the highest values of $CI_i out$ due to their high contact rate. Social organizations, along with innovative organizations, also have the highest values of $CI_i in$, as both have relatively high acceptance rates. In the case of innovative organizations, the acceptance of collaboration is reinforced by a high new products rate.

Finally, $CI_{CBE}t = 19{,}5$ is the ratio of the total number of collaborations opportunities generated in the CBE by the total number of organizations, and $CI_{CBE}d = 0{,}42$, the degree to which the most active organization in the CBE exceeds the number of collaboration opportunities of the others. This last value of centrality indicates that there is some polarization of the distribution of collaboration (as expected in social and innovative organizations) in the CBE, since this value is normalized between 0 and 1, where zero indicates an equal distribution of collaboration among all organizations.

Table 3. Calculated values of the CI for each individual organization and for the whole CBE.

Contribution Indicator (CI_i and CI_{CBE})				
Class of Resp.	**Profile of O_i**		**CI_i in**	**CI_i out**
Social	Contact rate Acceptance rate New products rate	0,8 0,6 0,2	1,00 0,81 0,91 0,69 0,84 0,72	0,63 1,00 0,84 0,68 0,82 0,68
Selfish	Contact rate Acceptance rate New products rate	0,1 0,2 0,2	0,34 0,28 0,25 0,19 0,28	0,26 0,42 0,29 0,29 0,26
Innovator	Contact rate Acceptance rate New products rate	0,4 0,5 0,9	0,94 0,97 0,81	0,42 0,34 0,44
Crook	Contact rate Acceptance rate New products rate	0,1 0,1 0,1	0,09	0,26
		$CI_{CBE}t$	**19,5**	
		$CI_{CBE}d$	**0,42**	

After running the simulation model again, varying the profile of social organizations, increasing their acceptance rates to 1,0, it can be seen from the results obtained in Table 4 that collaboration increased in social organizations, also causing an increase in the ratio of collaboration $CI_{CBE}t$ = 27,1 as well as in the centrality $CI_{CBE}d$ = 0,48.

Table 4. Calculated values of the CI after incrementing the contact and acceptance rates of social organizations.

Contribution Indicator (CI_i and CI_{CBE})				
Class of Resp.	**Profile of O_i**		**CI_i in**	**CI_i out**
Social	Contact rate Acceptance rate New products rate	0,8 1,0 0,2	0,90 0,80 0,92 0,80 0,82 1,00	0,76 0,69 1,00 0,73 0,63 0,76
Selfish	Contact rate Acceptance rate New products rate	0,1 0,2 0,2	0,16 0,31 0,22 0,22 0,22	0,35 0,12 0,24 0,35 0,49
Innovator	Contact rate Acceptance rate New products rate	0,4 0,5 0,9	0,53 0,76 0,59	0,39 0,63 0,57
Crook	Contact rate Acceptance rate New products rate	0,1 0,1 0,1	0,06	0,27
		$CI_{CBE}t$	**27,1**	
		$CI_{CBE}d$	**0,48**	

The results achieved in this work, based on the presented simulation model and scenarios, are a preliminary basis for the ongoing research. The next step is the verification of the evolution of the behaviour of a CBE, varying the weights of the PIs. It is expected that by using the PIs as a factor of influence in the profile of the organizations, they can self-adjust their behaviour, improving the performance of the CBE.

6 Conclusions and Further Work

The experimental evaluation in the previous section showed that a simulation model of a CBE can be defined and evaluated through the PIs proposed, more particularly, the CI to measure the collaboration of individual organizations and the CBE as a whole. It also showed that several scenarios could be modelled, varying the profile of the organizations, resulting in different measures of collaboration.

The ongoing work encompasses the calculation of the measures for the remaining PIs (II and PI), using all the indicators as a factor of influence of the behaviour of the organizations, to be possible to analyse the changes in the CBE by varying the weights of the PIs.

For future work, a more complete and dynamic simulation model can be set considering the integrating of elements of VO Performance Measurement (VOPM) proposed in the ECOLEAD project [15]. The characteristics and requirements of VOs establish specific performance perspectives of collaboration of their members, whose classes of responsiveness (Table 1) and behaviour (Fig. 3), used in this experimental evaluation, can be improved considering characteristics of reliability, flexibility, commitment and communication. On the other hand, the influence on the organizations of the calculated PIs, for the selection of the collaboration partners, as well as their dynamic shift from one class to another in different periods of time, are also possibilities for future work.

Finally, more diversified simulation scenarios should be defined, varying the number of organizations and tested boundary conditions.

Acknowledgments. This work benefited from the ongoing research within the CoDIS (Collaborative Networks and Distributed Industrial Systems Group) which is part of both the New University of Lisbon (UNL) - Faculty of Sciences and Technology, and the UNINOVA - CTS (Center of Technology and Systems). Partial support also comes from Fundação para a Ciência e Tecnologia through the PEST program UID/EEA/00066/2013.

References

1. Canning, M., Kelly, E.: Business Ecosystems Come of Age (2015). https://www2.deloitte.com/content/dam/insights/us/articles/platform-strategy-new-level-business-trends/DUP_1048-Business-ecosystems-come-of-age_MASTER_FINAL.pdf. Accessed 7 Jan 2019
2. Moore, J.F.: Predators and prey: a new ecology of competition. Harvard Bus. Rev. **71**(3), 75–86 (1993)
3. Camarinha-Matos, L.M., Afsarmanesh, H.: Collaborative networks: a new scientific discipline. J. Intell. Manuf. **16**(4–5), 439–452 (2005)
4. Camarinha-Matos, L.M., Afsarmanesh, H.: Collaborative Networks: Reference Modeling. Springer, Boston (2008). https://doi.org/10.1007/978-0-387-79426-6
5. Graça, P., Camarinha-Matos, L.M.: The need of performance indicators for collaborative business ecosystems. In: Camarinha-Matos, L.M., Baldissera, T.A., Di Orio, G., Marques, F. (eds.) DoCEIS 2015. IAICT, vol. 450, pp. 22–30. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-16766-4_3
6. Graça, P., Camarinha-Matos, L.M.: Evolution of a collaborative business ecosystem in response to performance indicators. In: Camarinha-Matos, L.M., Afsarmanesh, H., Fornasiero, R. (eds.) PRO-VE 2017. IAICT, vol. 506, pp. 629–640. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-65151-4_55
7. Graça, P., Camarinha-Matos, L.M.: A proposal of performance indicators for collaborative business ecosystems. In: Afsarmanesh, H., Camarinha-Matos, L.M., Lucas Soares, A. (eds.) PRO-VE 2016. IAICT, vol. 480, pp. 253–264. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-45390-3_22

8. Camarinha-Matos, L.M., Fornasiero, R., Afsarmanesh, H.: Collaborative networks as a core enabler of Industry 4.0. In: Camarinha-Matos, L.M., Afsarmanesh, H., Fornasiero, R. (eds.) PRO-VE 2017. IAICT, vol. 506, pp. 3–17. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-65151-4_1
9. Sniderman, B., Mahto, M., Cotteleer, M.J.: Industry 4.0 and Manufacturing Ecosystems: Exploring the world of Connected Enterprises. Deloitte Consulting (2016) https://dupress.deloitte.com/content/dam/dup-us-en/articles/manufacturing-ecosystems-exploring-world-connected-enterprises/DUP_2898_Industry4.0ManufacturingEcosystems.pdf. Accessed 7 Jan 2019
10. Freeman, L.C.: Centrality in social networks conceptual clarification. Soc. Netw. **1**(3), 215–239 (1978)
11. Opsahl, T., Agneessens, F., Skvoretz, J.: Node centrality in weighted networks: generalizing degree and shortest paths. Soc. Netw. **32**(3), 245–251 (2010)
12. Borshchev, A.: The big book of simulation modeling: multimethod modeling with AnyLogic 6. AnyLogic North America, Lisle (2013)
13. J. H. C. W.: Introduction to Mathematical Probability. Science Progress (1933-), vol. 33, no. 130, pp. 350–350 (1938)
14. Haight, F.A.: Handbook of the Poisson distribution. Wiley, New York (1967)
15. Camarinha-Matos, L., Afsarmanesh, H., Ollus, M.: Methods and Tools for Collaborative Networked Organizations. Springer, New York (2008). https://doi.org/10.1007/978-0-387-79424-2

Organizational Structure for Mass Collaboration and Learning

Majid Zamiri(✉) and Luis M. Camarinha-Matos

Faculty of Sciences and Technology and UNINOVA - CTS, NOVA University of Lisbon, 2829-518 Caparica, Portugal
zamiri_majid@yahoo.com, cam@uninova.pt

Abstract. The rapid emergence of collaborative communities supported by Internet has led to unprecedented waves of novelty in the ways people create and share knowledge. In this framework, the notion of mass collaboration has opened new windows of opportunity for collective learning. Mass collaborative learning is a new paradigm, through which large numbers of people engage in collaborative initiatives to learn from each other and alter the nature of formal education. Even though mass collaboration opens up an apparently limitless field for promoting social inclusion in effective learning, not all aspects, features, and characteristics of this phenomenon such as the organizational structures are quite clear at present. Therefore, this study is conducted to review the organizational structures of 14 real examples of mass collaboration. Through the analysis of the most suitable features of those structures we expect to be able to propose a general organizational structure for mass collaborative learning purpose. It is expected that such organizational structure could help developing a better insight into this field of study.

Keywords: Mass collaboration and learning · Knowledge creation and sharing · Community · Innovation

1 Introduction

Advances in information and communication technologies (ICT) and the emergence of IoT (Internet of Thing) have enabled people to redefine the boundaries of collaboration. This trend provides the possibility for large scale groups of distributed humans to join into mass collective projects and harness their potential joint power to deal with multi-faceted problems in social, economic, and environmental contexts. The emergence of the mass collaboration paradigm and its application to different domains is now reshaping the landscape of a wide variety of tasks, both locally and globally. Evidences clearly show that mass collaboration, by exploiting the capabilities of thousands of people, can create a kind of agile problem-solving system which is almost superior to any type of intelligent artefact that is made to serve similar purpose [1].

Mass collaboration brings together multitudes of individuals that may have not had the opportunity to work together before and may remain anonymous. It brings the opportunity to utilize the brainpower of participants in a collective effort and orchestrate their attempts in order to reach a common goal. In this context, Internet and ICT

Published by Springer Nature Switzerland AG 2019
L. M. Camarinha-Matos et al. (Eds.): DoCEIS 2019, IFIP AICT 553, pp. 14–23, 2019.
https://doi.org/10.1007/978-3-030-17771-3_2

have a facilitating role to play. In such collective action participants can efficiently and quickly contribute in developing an idea, plan, action, process, project, or artefact, to help solving a grand challenge [2].

There are many interesting applications of mass collaboration. For instance, its application in social learning occurs at a wider scale than the individual or group learning. In this case a large number of interested people capitalize on one another's resources, skills, and knowledge aiming to learn something new, and create lasting impact together. Mass collaborative learning, indeed, refers to a method of learning that can take place at community level where thousands of participants collectively and proactively engage in the process of knowledge acquisition, building, sharing, and developing, and where they can add their own contributions or even revise others' contributions. As opposed to traditional and formal learning methods delivered by instructors and utilizing systematic learning approaches within educational settings, mass collaborative learning stands upon the contribution of decentralized and self-directed participants who produce knowledge in an informal way [3].

In this case, knowledge creation and sharing can be considered as the core of learning that relies on the participation of a variety of people in learning networks helping to reach the community objectives. In addition, the ability to manage such knowledge is key to community success, which secures its competitive advantage and capability to achieve a sustainable superior performance. In this regard, it is significant to promote knowledge building and sharing that drives communities to create and/or add more value, thus engaging in effective innovation [4].

There are a number of factors that differentiate small organizations and communities from large entities, such as the type of organizational structure. An organizational structure determines how power, roles, and duties can be defined, controlled, and coordinated toward reaching community goals. It also specifies the way in which knowledge, information or data flow across different layers of the organization. Every organization or community certainly needs a structure (even if self-organizing) in order to survive, take actions, and grow [5]. Every community should select its structure based on its requirements and priorities. The type of organizational structure implicitly indicates in which ways internal works can be carried out.

In the past, the structures of communities were mostly designed for effectiveness and efficiency although they are nowadays designed for agility, speed, and adaptability to be able to compete and win in today's global competitive environments. As organizations or communities are becoming more and more digital-based and there is a transformation towards performing projects collaboratively, they are also facing with an imperative to redesign their structures in order to learn more rapidly, quickly respond to demands, and adapt to the characteristics of new workforces and workplaces. While the business environment, customer needs, technology capabilities and the nature of work in organizations and communities are likely to change, the organizational structure needs to reshape as well in a deliberate and strategic way. As such, the design of structures for adaptability is a shift away from traditional organizational structures like the hierarchical, centralized and bureaucratic models, towards unconventional models where projects are fulfilled collectively by network participants [6].

However, thus far there have been very few attempts to report on the role of organizational structures in the context of mass collaboration and learning. Furthermore, there are no clear evidences in the literature that show how mass collaborative projects can define, design, implement, and develop appropriate structures. Therefore, gaining some insight on what kinds of organizational structures have more chances for being adopted in mass collaborative learning projects is the foremost motivation for conducting this research work. Thus, a key research question that emerges is:

What kind of organizational structure within a community should be established to help developing learning through mass collaboration?

The proposed hypothesis to address this research question is:

Community learning through mass collaboration could be helped if existing models of organizational structures for long-term strategic networks are extended to allow more fluid borders and new roles, incentives and internal subgroups are defined to focus on learning and knowledge generation.

For this study, in order to search, choose, and review relevant papers, databases such as, SCOPUS, IEEE Xplore, Web of Science and Google scholar were used, being the goal to identify relevant examples and evaluate their organizational structures.

2 Relationship to Innovation in Industrial and Service Systems

From an organizational perspective, knowledge communities can speed knowledge creation, transfer, and utilization on an ongoing basis, as well as facilitate knowledge mobilization (for example, through providing suitable spaces for discussion in order to narrow the gaps between research and practice) [7].

Knowledge communities are often found to introduce changes to a system, and promote the culture of innovation. Such kind of communities can be called communities of innovation which are dedicated to support innovation. Communities of innovation are creative and dynamic entities that pursue innovative solutions to societal challenges. Communities of innovation are not only responsible for a growing number of innovations, but can also provide a common ground for learning. In this subject, they can freely and efficiently impart information and knowledge to the wider public. The literature shows that some examples of communities of innovation [8, 9] have successfully influenced the learning process.

Furthermore, "*online innovation communities have an ability to learn in a dispersed setting without any formal involvement, their learning capability is actually very remarkable, making it even more striking that we lack academic insight how these learning competences come about*" [9]. Mass collaboration through a large online community can be applied in various domains and fields of study. For instance, it can foster learning and optimization of the innovation portfolio through:

- Increasing the flow of new ideas, knowledge, or information generation,
- Boosting the chance of association between ideas, knowledge, or information,
- Improving the quality of ideas, knowledge, or information,

- Speeding up the collaborative feedbacks,
- Developing connection between members,
- Reaping the power of collective intelligence,
- Etc.

However, it is vital in this new context to be able to evaluate how much trustworthy the acquired knowledge is, because knowledge is power, and it serves as a basis for making choices and decisions in communities. One fundamental step in the process of learning, particularly in an online environment, is to ensure that the created and shared knowledge or information is reliable, as well as the accuracy and credibility of the materials that people encounter with, are high.

It is largely evidenced that in World Wide Web, and specifically in networked-collaborative activities neither all delivered materials are reliable nor will all stay stable. In addition, on one hand, the quality and value of various types of Internet sources (that are available in different formats) are not all high, and on the other hand not all Internet users are able to accurately evaluate the appropriateness of all types of online sources [10]. Thus, along with informal learning in such communities, it is essential to assess the quality and reliability of knowledge or information created in whatever format, particularly in mass collaborative projects. To cope with this challenge, approaches such as machine learning [11], digital audio and video output [12], BS detector [13], and linguistic and network-based approaches [14] can play a relevant role.

3 Analysis of Selected Cases

In this study, in order to gain a clear understanding of the organizational structure of mass collaboration, and propose an appropriate structure for mass collaborative learning projects, 14 relevant examples of mass collaboration in different domains were selected from the literature including, (1) *Wikipedia* (a well-known case of a web-based encyclopedia that is written collaboratively by its users), (2) *Digg* (a social networking website that aggregates interesting online news, pictures, and videos), (3) *Yahoo! Answer* (a community-driven question-and-answer website that allows users to ask questions and answer questions), (4) S*ETI@home* (a computing project and scientific experiment that benefits of Internet-connected computers in the search for signs of Extraterrestrial Intelligence), (5) *Scratch* (an online community that enables children to program and share interactive media with other people), (6) *Galaxyzoo* (a crowd-sourced and on-line astronomy project which classifies the morphology of galaxies, and then analyze their pictures and rate them), (7) *Foldit* (an online puzzle video game that uses the power of distributed computing to create and design the primary structure of chosen proteins), (8) *Applications of the Delphi method* (a structured communication technique based on the results of several rounds of questionnaires sent to a panel of experts), (9) *Climate Colab* (an open problem-solving platform where a community of experts on climate change evaluate plans to reach global climate change goals), (10) *Assignment Zero* (an experiment in crowd-sourced journalism, allowing collaboration between lots of people to work on a publishable story, with many parts), (11) *DonationCoder* (a community of programmers who develop and finance their own free software), (12) *Experts Exchange* (a trusted global online community offering millions of verified solutions from industry experts), (13) *Waze* (a navigation app that runs on smartphones and tablet computers, through which users help each other to find directions and avoid traffic jams), and (14) *Makerspaces* (which are physical or digital spaces for open collaboration, where people have access to resources for developing projects with the aim of creating products or services).

In Tables 1 and 2, the organizational structure and main characteristics of two of the above-mentioned examples are summarized as an instance.

Table 1. Organizational structure of Wikipedia

<table>
<tr><td colspan="3">What is Wikipedia: it is a web-based, free and open content encyclopedia based on a model of openly editable content - a wiki. As a general reference, it has been the most popular on the World Wide Web. It is written collaboratively by the people that use it. Many people from around the world are willingly contributing in Wikipedia development.</td></tr>
<tr><td colspan="3">Wikipedia Organizational Structure</td></tr>
<tr><td colspan="3">Wikipedia Membership Method</td></tr>
<tr><td colspan="3">• Wikipedia is open to public to use, read, write, edit, and make changes in articles.
– People from any age group, gender, race, culture, and background can participate in.
• Wikipedia community includes all anonymous editors, supporters, current and potential readers.
• Wikipedia contributors consist of two main groups, those who create an identity, and those who communicate with other contributors.</td></tr>
<tr><td colspan="3">Wikipedia Roles</td></tr>
<tr><td colspan="3">Wikipedians (or editors) are the volunteer who write and edit Wikipedia’s articles. Anyone can become a Wikipedian. Wikipedians do a wide variety of tasks, being free to pick and complete their tasks anytime anywhere.
Wikipedians through collaboration and discussion can gain a sense of collective purpose and connectedness. While there are disagreements about an issue, a consensus can be reached through open and friendly discussion. While unresolved disputes or conflict arise, no matter the reason, it can be pursued in related talk page, either through comprehensive dispute resolution process of Wikipedia or, requesting comments from other Wikipedians.</td></tr>
<tr><td>No</td><td>Roles</td><td>Descriptions</td></tr>
<tr><td>1</td><td>Account creator</td><td>Gets access to a tool that allows trusted Wikipedians to create a high number of accounts for other people who request them.</td></tr>
<tr><td>2</td><td>Editor (from expert scholar to casual reader)</td><td>A volunteer who writes and edits Wikipedia’s articles.</td></tr>
<tr><td>3</td><td>Bureaucrat</td><td>Permitted to perform particular actions on Wikipedians’ accounts.</td></tr>
<tr><td>4</td><td>Volunteer Response Team</td><td>Group of volunteers who answer most email sent to Wikipedia.</td></tr>
<tr><td>5</td><td>Event coordinator</td><td>For a short time, can add new created accounts to verified user groups, hence such accounts could write new articles.</td></tr>
</table>

(*continued*)

Table 1. (*continued*)

6	Edit filter manager	Can make, change, enable, disable, and omit edit filters. Moreover, can check private filters and also their related logs.
7	Arbitration Committee	Handles those conflicts which remain unresolved while all the efforts in dispute resolution have failed.
8	Steward	Can give and revoke any authority to or from any Wikipedian on any wiki operated by the Foundation of Wikimedia that permits creating open account.
9	Oversight	Authorized to delete pages and revisions, and block function pages that it makes possible to hide logs or modify pages from any form of usual access by other Wikipedians.
10	CheckUser	Permitted to check the list of all IP addresses, the list of all edits, and all user accounts.
11	Administrator	A volunteer editor who is granted some technical authority to make particular decision and actions about blocking and unblocking user accounts and IP addresses; editing, modifying, and removing the protected pages, deleting and undeleting pages, protecting and unprotecting pages from editing; and etc.
12	Interface administrator	The only local user group who authorized to edit Cascading Style Sheets, JavaScript, and Javascript Object Notation pages.
Relationship Between Roles in Wikipedia		
• The community of Wikipedia is divided into large number of “spheres” which categories members based on their area of interest, expertise, background, age, etc. • The conversations and debates among members will be facilitated by means of Discussion Pages. • There is possibility for members to nominate each other for awards of Wikipedia. • Neither the quantity or frequency of contributions can be controlled, nor will members be fired. • There are additional administrative responsibilities that can be taken to serve the community better. • It has been attempted to Wikipedians be treated equally with no “power structure” although a hierarchy of positions and permissions is there (from simple editor to Jimmy Wales, the founder of Wikipedia).		
Wikipedia Content Management		
• Wikipedia is a free online encyclopedia and its entire contents are written by unpaid volunteers. • Wikipedia is continually developed and updated. Its articles are intended to be realistic, reliable, and verifiable with cited external sources. • Wikipedia has variety of procedures (e.g., peer review, good article assessment, and the featured article process) for continual article review and improvement. • Feedback about articles, in the first instance, raise on the related discussion pages. • Quality constantly improves via removing or repairing the misinformation and other errors.		

Table 2. Organizational structure of Digg

What is Digg: it is a website that aggregates interesting online news, pictures, and videos through compiling links to the relevant webpages. Users submit stories for promotion, and they are subsequently either voted for (digged) or against (buried). The most popular contents are posted to the front page for making it easier to identify and consume the most important stories, videos, and trends of the day.		
Digg Organizational Structure		
Digg Membership Method		
• Digg is a social networking and user-driven website that anybody can participate. • In Digg, nothing is written by paid editors. Contents are made by hundreds of thousands of users. • All users need to create a Digg user account to access the features of the website. They should play active role in both presenting and Digging stories. Log in is mandatory to get to the website. It provides kind of security assurance for each visit. • After registration, it is possible for users to give comment and vote on others contribution. • All users' information (e.g., past Diggs, friends, feedbacks) will be stored in their Digg profile. • Users can submit their stories and also benefit of all provided features in the website. Bad stories will be ignored, and good ones promoted. Thus, the stories that got positive votes and selected as the best will be cross-pollinated across other channels. • Users can customize their own news feeds.		
Digg Roles		
All Digg users are volunteers. The Digg community is made up of users who play different, often overlapping roles.		
No	**Roles**	**Descriptions**
1	Casual reviewer	Looks for interesting stuff.
2	Reader	Makes up the majority of Digg user who reads and reaps the benefits of provided materials.
3	Submitter	Posts news and stories that s/he finds in different blogs, websites and random postings from around the Web.
4	Dedicated reviewer	Spends several hours each to check the stories, promote good ones, and report those are not good.
Relationship Between Roles in Digg		
• Digg provides a place for lively conversation, discussion, inquiry, and debate. Digg community can discuss the topics that they're passionate about. • Users can add friends and develop their relationships. • A user can block another user if doesn't like his submissions or comments. The blocked user may get banned from Digg if he is blocked by enough number of Digg users. • Users can create or develop a "Digg game" by submitting stories and digging them.		

(*continued*)

Table 2. (*continued*)

Digg Content Management
• All the content-related decisions are made by site's users. • When a user submits a story, its validity will be checked by the system. • When a submitted article is up for promotion to a category homepage or the front page, the system (karma) checks it to make sure the Diggs are valid. • Digg does not have editorial control on submissions, promotions, or burying. • Digg manages all things with a proprietary algorithm (de-promotion algorithm). When a story is Digged by certain number of users (at least 40 persons), the Digg system automatically will move it to the front page of the website. • The most popular stories of Digg are placed in the "Top News" section of the website. • The top news can be anything (e.g., fun content or serious news). • Digg is classified into different groups based on topics (e.g., business, technology, videos, and entertainment news). • Digg has tabs that let users filter or sort contents into news stories, videos, images and podcasts.

In summary, elements for a typical organizational structure for mass collaborative learning projects, as derived from the 14 studied examples, are displayed in Fig. 1.

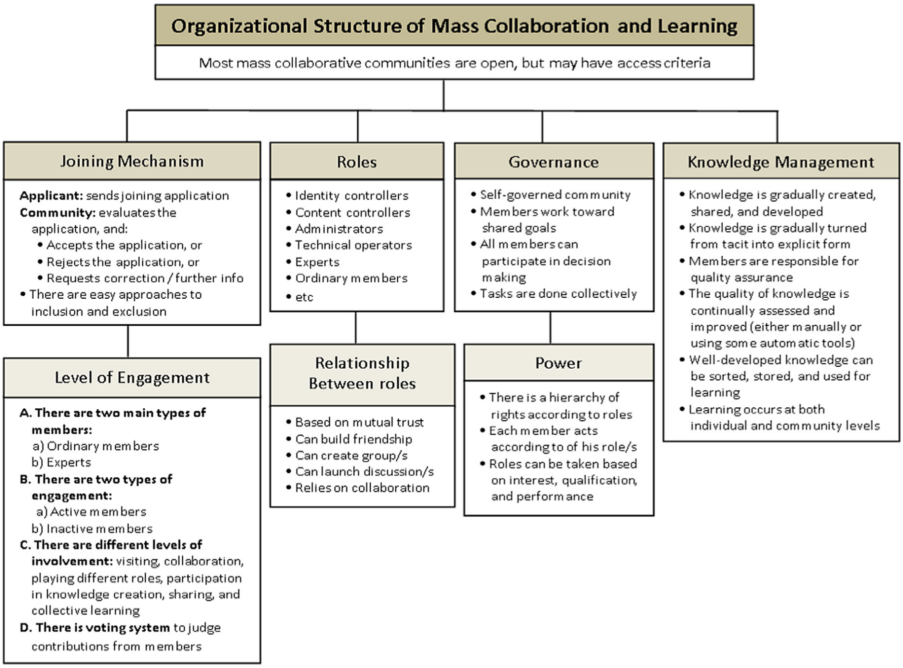


Fig. 1. Elements for an organizational structure for mass collaborative learning

Having critically analyzed the organizational structure of 14 studied examples of mass collaboration and also reviewed related papers, it is concluded that each organizational structure stands upon some building blocks and fundamental elements. In this sense, it is therefore suggested that the organizational structure of mass collaborative learning (for creating, developing, and servicing) should take into account at least four core elements and three supplementary elements. Core elements including, (A) the required mechanism for members to join the community, (B) the roles that can be taken and played by members, (C) the methods of governing the community, and (D) the way that possessed knowledge or information can be managed properly and efficiently. Supplementary element consists of, (a) the ways and levels that members can engage in different activities, (b) the ways that different roles can built and involved in interrelationships, and (c) the power, rights, and responsibilities that members can take. In this structure the role of supplementary elements is augmenting, clarifying, and facilitating core elements. This organizational structure depends on the specific situations and conditions of application. It is expected that the proposed structure in this work can contribute to the development of this field of study, and enrich the understanding of the complex organizational structures for mass collaboration and learning.

4 Conclusions and Future Work

Mass collaborative learning provides greater opportunities for distributed contributors to engage in virtual global learning and take advantage of powerful social communities of experts and peers to develop innovative solutions to major challenges. Despite of successful outcomes that mass collaborative learning already gained, we still need to clarify our understanding about the required organizational structures for this emerging phenomenon.

Evidences demonstrate that collaboration and innovation are not mutually exclusive; on the contrary, they feed and build upon each other. That is, collaboration brings and drives innovation, and innovation happens through collaboration [15–18]. Considering that fact, in this study, in order to identify appropriate organizational structures for mass collaborative learning projects, the organizational structure of 14 real examples of mass collaboration are reviewed.

This work is still ongoing, but it is expected that the preliminary findings of this review and the proposed organizational structure can provide communities and learners with helpful guidelines and directions for achieving effective mass collaborative learning.

Acknowledgements. This work was funded in part by the Center of Technology and Systems of Uninova and the Portuguese FCT-PEST program UID/EEA/00066/2013.

References

1. Ramaley, J.: Collaboration in an era of change: new forms of community problem-solving. Metrop. Univ. **27**(1), 3–24 (2016)
2. Zamiri, M., Camarinha-Matos, L.M.: Learning through mass collaboration - issues and challenges. In: Camarinha-Matos, L.M., Adu-Kankam, K.O., Julashokri, M. (eds.) DoCEIS 2018. IAICT, vol. 521, pp. 3–17. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-78574-5_1
3. Cress, U., Moskaliuk, J., Jeong, H. (eds.): Mass Collaboration and Education. CCLS, vol. 16. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-13536-6
4. Chiangmai, C.N.: Creating efficient collaboration for knowledge creation in area-based rural development. Kasetsart J. Soc. Sci. **38**(2), 175–180 (2017)
5. Galbraith, J.R.: Organization design. In: Lorsch, J.W. (ed.) Handbook of Organizational Behavior, pp. 343–357. Prentice Hall, Englewood Cliffs (1987)
6. Yaragarla, R.: Scenarios of collaborative approach in mass collaboration (2016). http://www.workmonkeylabs.com/scenarios-of-collaborative-approach-in-mass-collaboration/
7. Lindkvist, L.: Knowledge communities and knowledge collectivities: a typology of knowledge work in groups. J. Manage. Stud. **42**(6), 1189–1210 (2005)
8. Brogi, S.: Online brand communities: a literature review. Procedia Soc. Behav. Sci. **109**, 385–389 (2014)
9. Moser, Ch., Groenewegen, P., Huysman, M.: The influence of social norms on community learning - a theoretical framework to understand learning processes in online innovation communities (2010). https://pdfs.semanticscholar.org/919f/b258eae9c2cf2edc1c196e9540f58a7c7d32.pdf
10. Kaushik, A.: Evaluation of Internet resources: a review of selected literature. Braz. J. Inf. Sci. **6**(2), 61–83 (2012)
11. Gilda, Sh.: Evaluating machine learning algorithms for fake news detection. In: 15th Student Conference on Research and Development (2017). https://doi.org/10.1109/scored.2017.8305411
12. Arnold, S.D.: Assessing student learning online: overcoming reliability issues. In: IADIS International Conference on Cognition and Exploratory Learning in Digital Age (CELDA 2012) (2012)
13. https://www.maketecheasier.com/useful-tools-spot-fake-news/
14. Conroy, N.J., Rubin, V.L., Chen, Y.: Automatic deception detection: methods for finding fake news. In: Proceedings of the 78th ASIS&T Annual Meeting: Information Science with Impact: Research in and for the Community, Article No. 8 (2015)
15. Gressgård, L.J.: Virtual team collaboration and innovation in organizations. Team Perform. Manage. **17**(1/2), 102–119 (2011)
16. Levine, Sh.S., Prietula, M.J.: Open collaboration for innovation: principles and performance. Organ. Sci. **25**(5), 1414–1433 (2014)
17. Camarinha-Matos, L.M., Afsarmanesh, H., Galeano, N., Molina, A.: Collaborative networked organizations - concepts and practice in manufacturing enterprises. J. Comput. Ind. Eng. **57**, 46–60 (2009)
18. Rosas, J., Urze, P., Tenera, A., Camarinha-Matos, L.M.: Open innovation participants profiling: an archetypes approach. In: Camarinha-Matos, L.M., Afsarmanesh, H., Rezgui, Y. (eds.) PRO-VE 2018. IAICT, vol. 534, pp. 177–189. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-99127-6_16

Emerging Community Energy Ecosystems: Analysis of Organizational and Governance Structures of Selected Representative Cases

Kankam O. Adu-Kankam[1,2](✉) and Luis M. Camarinha-Matos[1]

[1] Faculty of Sciences and Technology and UNINOVA - CTS, Nova University of Lisbon, Campus de Caparica, 2829-516 Monte Caparica, Portugal
kankamadu@gmail.com, cam@uninova.pt

[2] School of Engineering, University of Energy and Natural Resources (UENR), P. O. Box 214, Sunyani, Ghana

Abstract. The quest to decarbonize and decentralize the current power grid has enabled high penetration of Distributed Energy Resources at the edge of the distribution network. The diversity, complexity and growing numbers of these energy resources currently pose a challenge to managing them. This has resulted in the emergence of various energy ecosystems which use diverse community-based organizational strategies and initiatives as forms of management techniques. There are also corresponding business models and governance structures that are innovative and technologically disruptive to the operations of the current grid. An analysis of five representative cases of these Renewable Energy Communities is performed using real-life projects as case studies. The focus areas considered in this study included organizational and governance structures, roles, and the relationship between key stakeholders/owners, how these ecosystems interact with the power grid, and the role of collaborations. The outcome of the study revealed that each category of ecosystem has similar organizational and governance structures although they may differently be constituted. In terms of ownership, energy cooperative, municipalities, and communities were found to own a greater share of these energy resources. Furthermore, most of these ecosystems were found to interact with the grid by supplying excess energy from the community to the power grid whilst others operate in isolation from the grid. Apart from one case, all the others showed elements of collaborations as an integral component of their mode of operations.

Keywords: Renewable energy communities · Energy cooperatives · Collaborative networks · Collaborative energy ecosystems

1 Introduction

According to [1], the future of energy rests on the foundations of cleaner, distributed, and intelligent networks of power. This notion is being advanced by the rapid innovations that are emerging at the edge of the power grid, which is driving a paradigm

Published by Springer Nature Switzerland AG 2019
L. M. Camarinha-Matos et al. (Eds.): DoCEIS 2019, IFIP AICT 553, pp. 24–40, 2019.
https://doi.org/10.1007/978-3-030-17771-3_3

shift, that is transforming the power network towards the 5D (digitalization, decentralization, decarbonization, democratization, and deregulation) [2]. These evolving technologies are characterised by high penetration of Distributed Energy Resources (DERs), artificial intelligence, cloud computing, blockchain technology and a host of other cutting-edge applications and services [3] that are facilitating these transformations in alignment with current trends on Industry 4.0 [4]. According to recent developments, DER asset owners, property owners (land and building) and other entities have become economic participants and key players in the way energy is generated, distributed, and consumed. Additionally, new stakeholders, such as individual prosumers, local community members, and local governments are also emerging with community-based energy concepts that are transforming and enhancing the energy value-chain in diverse ways. These developments are driving the rapid proliferation of new energy models such as Renewable Energy Communities (RECs), Peer to peer energy sharing, Integrated community energy systems, Energy hubs, IEEE smart villages, etc., which are kinds of emerging "energy ecosystems". These ecosystems fit into a broader framework of upcoming concepts like collaborative economy or sharing economy which involve a more decentralized approach to the exchange of goods and services. To compliment the rapid evolution of these concepts, scientific disciplines such as Collaborative Networks are emerging [5–7] which provide a comprehensive body of knowledge, including reference models [8] that enhances the understanding of collaborations within these emergent domains. Additionally, these disruptive concepts are redefining market relationships between traditional sellers and buyers and therefore expanding models of transaction and consumption, which invariably are impacting business models across diverse ecosystems, which includes energy [9].

This work focuses on the analysis of five emerging RECs using representative real-life cases. A sixth case, which is "Energy hubs", is mentioned briefly in this work but without any detailed analysis because the concept is currently at a theoretical stage. Three research questions are used for this study. These are: (1) RQ1. *How are these RECs organised?* (which can include as sub-questions: RQ1A. *What are the roles and responsibilities?* RQ1B. *How are these RECs governed?* RQ1C. *What relationship exists between roles?*); (2) RQ2. *How do these RECs interact with the power grid?* and (3) RQ3. What role does collaborations play in these RECs?

The outline of this paper is as follows: Sect. 1 introduces the research concepts and related research questions. Section 2 describes the relationship of this work to the conference theme, followed by Sect. 3 which focuses on the case study. The main contribution of this work is described in Sect. 4.

1.1 Related Works

This section of related works has a narrow scope because it is intended to address the related works in the context of the on-going research as described in Sect. 4. It is therefore focused on prosumers within RECs, also referred to as Prosumer Community Groups (PCGs). In [10], the authors performed a literature review to determine the evolution and future directions of key issues, challenges, and opportunities in PCGs. In this review, the authors proposed eight future trends and current research gaps in the area. Another related work presented in [11] explored the concept of PCGs and proposed a novel framework for the definition and formation of these groups by clustering prosumers based on their energy behaviour. Another dimension of this concept was explored by Rathnayaka et el in [12]. Here the authors propose a goal-oriented prosumer community group notion, that adopts the VPP concepts leading to a sustainable energy sharing process. From a different perspective, [13] organised prosumers into virtual clusters so that they could participate in the energy market as a single entity enabling a reduced total energy cost due to forecasting inaccuracies. In [14], a framework to integrate prosumer communities into the existing day-ahead and intraday markets is introduced. The authors proposed the Smart elecTricity Exchange Platform as an interface between the wholesale energy markets and prosumer communities. Furthermore, a comparative study is conducted in [15] to assess the technical and economic feasibility of household energy storage and centralized community energy storage.

2 Relationship to Innovation in Industrial and Service Systems

The next industrial revolution (Industry 4.0) and related services, characterized by digitalization of goods and services, resulting in enhanced value chains, and disruptive business models, require corresponding energy systems which are digitalized, distributed, decarbonized, democratized, and smart [4], [16]. According to the United Nations Industrial Development Organization, to achieve the international climate change mitigation targets, the industrial sector, which accounts for approximately one-third of global final energy use and 40% of global energy-related CO2 emissions (OECD/IEA 2009) [16], must reconsider the kinds and sources of energy they consume. Therefore, the transition towards more sustainable energy ecosystems and the digital transformation of industries could mutually benefit from each other by offering solutions, that help to optimise the integration of renewable/sustainable forms of energy on an industrial scale.

3 Case Study

3.1 Integrated Community Energy Systems/Solutions (ICES)

According to Koirala et al. (2016), there is no single definition nor typology to describe ICESs. However in general terms, they are tailored (in technology and size) to meet the needs of the respective community [17]. By taking a more holistic approach to energy demand and supply, ICES provide a good alternative to energy management for the benefit of an entire community. For instance, waste heat from industry could be used to heat buildings in community and activities that encourage energy conservation is also practiced. ICES approach to energy management help to achieve better overall energy savings in a community. An ICES community can be established in a rural area, a neighbourhood or a large city.

Table 1. Case 1: Feldheim, Germany [18, 19]

Key stakeholders	Roles			Governance structure	Grid interaction
	Key roles	Energy resource ownership	Characteristic & and relationship between roles		
Feldheim New Energy Forum Foundation	1. Project financiers 2. Management body 3. Education and information	Community owned assets: 1. Wind turbines 2. PV modules 3. Biogas 4. Biomass 5. CHP 6. Lithium-ion battery system	Characteristics of roles: –Public private partnership. Relationship between roles: –Distributed systems with centralized management and hierarchical governance	Governance type: Non-profit self-governance Structure: 1. Executive Board 2. General assembly with voting rights 3. One member one vote	Send 90% of energy produced in the community to the grid. The village uses only 10%
Municipality of Treuenbrietzen	1. Project financiers 2. Project initiation and implementation				
Residents of Feldheim and Treuenbrietzen	1. Project financiers 2. Centre of excellence 3. Energy consumers				
Feldheim Energiequelle GmbH	1. Project initiation and implementation 2. Project managers				
Farmers' cooperative	1. Project financiers 2. Land owners				

Summary of ICES Case Study

RQ1. Organization

- Roles and responsibilities: The key roles found include project financiers who provide financial input into the projects. Other financial partners include non-profit organizations, governmental agencies and community-owned cooperatives. There is also strong involvement of utility companies and grid operators who provide grid management and research services. Other actors in the ecosystem include academic/research institutions who provide research and development support for the ecosystem and finally environmentally sustainable groups who advocate for sustainable communities and energy use (providing energy advice to members). The community is key consumer of the energy produced.

Table 2. Case 2: Wildpoldsried, Germany [20–23]

Key stakeholders	Roles			Governance structure	Grid interaction
	Key roles	Energy resource ownership	Characteristic & relationship between roles		
The Wildpoldsried Municipal Council	Project financiers	1. Mixed ownership: i. Biogas (farmers and individuals) ii Biomass (individuals and Innovationspark Allgäu GmbH) 2. Community ownership: i. Village heating ii. Photovoltaics iii. Solar thermal 3. Individual ownership: i. Hydropower ii. Geothermal unit 4. Other ownership: i. Wind turbines owned by clubs	Characteristic of roles: Public private partnership Relationship between roles: Distributed systems with centralized management and hierarchical governance	Governance type: Non-profit self-governance The Environment and Energy Committee. (Composed of the Chairman and 4 honorary council members)	Community produces 469% more energy than needed. And This is sent back to the power grid
Germany's Ministry for Economic Affairs and Energy (BMWi)	Project financiers				
Siemens AG	Grid management				
RWTH Aachen University	Research partners				
ID.KOM Network Solutions	Smart grid markets solutions provider				
1. Kempton University of Applied Science 2. The Institute for High-Voltage Technology, RWTH Aachen University	Research partners				
AUW GmbH	Research and grid technology partner				
Wildpoldsried Village	Consumers				

- Governance: Types seen in these ecosystems are predominantly non-profit, non-governmental and self-governance. Board of directors are usually appointed to be managers of these energy ecosystems. The number of directors varies from case to case, however, there is a General Assembly that is constituted of the investors/community members and they all have equal voting rights which is: one man one vote, irrespective of the value/stake of their investment.
- Characteristics and relationships: Public private partnership. Distributed systems with centralized management and hierarchical governance is the relationship between key roles.

RQ2. Interaction with the power grid: In the cases considered, the community generates excess energy far beyond the demands of the community. This is sold to the power grid to generate revenue for community benefit. Some percentage of the revenue is re-invested in community projects whilst 5–10% is paid annually to investors as dividends. This possibility therefore makes these communities viable sources of long-term energy business investment.

RQ3. Role of collaboration: Collaborations between key stakeholder's results in the production of energy to meet the need of the community and the excess injected into the power grids. Stakeholders also collaborate to invest in social intervention projects such as conversion of street lights into energy saving LEDs, for the community using proceeds from energy sale.

Limitation of ICES

- Communities vary in size, structure, resources, laws, and opportunities, which leads to a wide array of organizational possibilities and no standardized approach to planning and deploying an ICES system.
- Limited ownership and participation by individuals. Most of the energy resources in these energy ecosystems are usually owned by municipalities, or governments, utility companies, factories and energy cooperatives. This is because of the scope, scale and technological complexity for planning and deployment of these ecosystems makes it unattractive for individuals.
- Development of these projects are usually confronted by the NIMBY (Not in my back yard) syndrome, which often opposes the installation of solar PVs and wind turbine in these communities.

3.2 Shared Renewable Energy Facility (SREF)

The Interstate Renewable Energy Council [24] defines a SREF as a facility that is located in or near the service territory of an Electricity Provider where the electricity generated by the facility is credited to participants of the facility. Typically, participating customers lease or purchase a share of the output of the facility. This facility may be owned by a utility company, a third party or participating customers [25]. As the facility generates electricity, participants receive a utility bill credit or direct payment, based on the electricity produced by their share of the facility. Current literature supports two types of SREF. These are: (1) Community shared solar, and (2) Community shared wind.

Since the concepts and principles behind these two technologies are similar, this study will focus on only the first category.

Community Shared Solar. A study conducted in 2008 by the National Renewable Energy Laboratory found that only 22 to 27% of residential rooftop areas is suitable for hosting an on-site photovoltaic (PV) system [26]. Therefore, to expand access to solar power, a new business model that enables those with shaded roofs, and those who choose not to install a residential system on their home for financial or other reasons, to rent or purchase a share of a solar facility. In a community shared solar project, customers subscribe to the program by purchasing a share of its energy output (solar). This is achieved either through an upfront payment or through ongoing monthly payments. This allows subscribers to benefit from a central commercial-scale solar plant through virtual net energy metering [27]. Literature such as [26] describes three models of Community shared solar projects: (1) Utility-Sponsored Model, (2) Special Purpose Entity Model, and (3) Non-profit Model. Studied cases:

– Case 1: Westmill Solar Co-operative (Special Purpose Entity Model) [28, 29]
– Case 2: Bristol Energy Cooperative (Special Purpose Entity Model) [30, 31]

Summary of Community Shared Solar

RQ1. Organization:

- Roles and responsibilities: These ecosystems are multi-stakeholder with diverse roles and responsibilities. Since these installations usually cover large parcels of land, land owners and local community/councils' have high stakes and play key ownership roles. Other emerging roles include partners for social marketing (e.g. Oggadoon) and advocacy groups like the Westmill Sustainable Energy Trust and the Climate Coalition play key roles to promote the sustainability of the energy community and climate change. Housing cooperation like Ambition Lawrence Weston are direct beneficiaries of the energy facility. In both cases considered, the facility is owned by the local cooperatives who are major consumers of the energy produced.
- Governance: The governance type here is also non-profit, non-governmental and self-governance. Just like the previous case in ICES, Board of directors are usually appointed to be managers of these ecosystems. The number of directors also varies from one case to another. However, there is a General Assembly that is constituted of the investors/community members and they all have equal voting rights which is one man one vote, irrespective of the value/stake of your investment.
- Characteristics and relationships: Public private partnership. Distributed systems with centralized management and hierarchical governance.

RQ2. Interaction with the power grid: These ecosystems also inject excess power into the grid and receive revenue through Feed-in-tariffs.

RQ3. Role of collaboration: The community collaborates to share a common resource which is a community owned solar PV installation. They also collaborate to consume locally generated energy thereby reducing total reliance on the power grid.

Limitations of Shared Renewable Energy Facilities

- As an emerging concept, it is likely to face developmental, implementation and policy barriers. Again, the technical, financial and socio-economic potential as well as challenges may take a while to understand.
- High upfront investment cost and long-term return on investments makes it attractive only to customers who are willing to pay high premium for renewable energy use, which may exclude moderate to low income customers.
- Regulations framework for these ecosystems are currently weak or non-existent.
- Interest here is predominantly based on the financial gains for the investors and not to promote a resilient energy community.

3.3 Community Microgrids (CM)

According to the Clean Collation, a Community Microgrid (CM) is a coordinated local grid area served by one or more distribution substations and supported by high penetrations of local renewables [32]. Although linked to the main electric grid, during power outage a CM can isolate from the broader grid and provide renewables-driven backup power to critical facilities [33]. Considering the constant threats posed by natural disasters like hurricanes, wildfires, floods tsunamis and cyber-attacks which have the potential to completely knock down the entire electrical infrastructure makes this energy concept very attractive. A key feature of a CM is the ability to serve thousands of customers with local renewable energy while achieving economies-of-scale and providing renewables-driven power backup to critical facilities and services during grid outages [34]. In [32], the authors further explained that CM represents a new approach for designing and operating the electric grid. It is purposely designed to rely heavily on DERs to enable a more sustainable, secure, and cost-effective energy system while generally providing renewables-driven power backup for prioritized loads over indefinite durations. Potential and ideal sites for CM implementation may include critical infrastructures that are useful for evacuations, provision of shelter, recovery staging, and life-sustaining services. Studied cases:

- Case 1: Long Island Community Microgrid Project (LICMP), East Hampton, NY [34–37].
- Case 2: Montecito Community Microgrid Initiative [38–40]

Summary of CM

RQ1. Organization:

- Roles and responsibilities: Key roles are funding partners who provide funding for these projects, research institutions who provide R&D activities and non-profit sustainable energy organization who promote environmentally sustainability in communities. In these ecosystems, energy resources are owned by the community whilst utility companies purchase the energy from the communities. In some instances, like the case of LICMP, the utility company is responsible for providing grid management service. In both cases considered, the Clean Coalition, a non-profit organization is key partner for project initiations. Long Island and Montecito Communities is the primary consumer of the energy produced in both CMs.

- Governance: Current literature does not provide adequate information, however there is ongoing debate as to whether these CM will be regulated as utility assets or products of the competitive market – or a hybrid of both? [41].
- Characteristics and relationships: Predominant characteristics are partnership between funding organizations, utilities, the communities, non-governmental organization and entities that host these installations. The key relationship between roles are distributed systems, with centralized management.

RQ2. Interaction with the power grid: These projects help to relief the power grid by facilitating energy self-sufficiency at the community level. They are usually designed such that they can provide emergency and backup services to critical infrastructure within the community during power outages. In the absence of outage, the micro-gird provides the needed power for the community.

RQ3. Role of collaboration: During power outages, the community, operators of the CM and emergency service providers collaborate to provide emergency power services to critical infrastructure within the community. In the absence of power outage, the community collaborate to share energy resources.

Limitations of CM

- Policy makers and regulators at various jurisdictions should consider policy changes that will enable utilities to own DER assets at least for the purpose of research and development. This will facilitate the advancement of these concepts as these utilities currently possess the capacities needed to run and test the practicability of these concepts.
- Community resistance to the mass installation of these facilities (e.g. Commercial-scale solar PVs on rooftops and parking lots) as they are perceived as aesthetic defects to the beauty of landscapes and an environmental nuisance to community settings.
- Currently in few jurisdictions, CM assets are owned by utilities (Utility 2.0, Utility of the future model) therefore community participation is limited and does not promote community sense [42].

3.4 Peer-to-Peer Energy Sharing (P2P)

According to [43], P2P energy sharing enables prosumers to directly trade energy with each other to achieve a win-win outcome. In [44] the authors described four different forms of P2P energy sharing networks. These are: (1) Platforms which are currently operating as proof-of-concept on a local scale and therefore target producers, prosumers and consumers; (2) P2P trading marketplace using platform-as-a-service to utility companies and retailers; (3) Cases that target prosumers and consumers on a platform with global scope; and (4) Cases that integrate two models such as local P2P platform to address producers, consumers and prosumers and also providing platform-as-a-service to utility companies and retailers. Both cases studied fall under the fourth category (P2P planform and platform-as-a-service):

– Case 1: Enosi ecosystem (P2P Platform and platform-as-a-service) [45, 46]
– Case 2: Pylon network (P2P Platform and platform-as-a-service) [41]

Summary of Peer to Peer Energy Sharing

RQ1. Organization:

- Roles and responsibilities: (i) IT partners - Provide web development and payment services; (ii) Energy partners-Management of generation and consumption; (iii) Energy Metering Oracles partners - Metering/data service Providers; (iv) Neo retailer - They provide tailor-made member management functionality; (v) Enosi platform - Platform used to trade electricity within Enosi ecosystem; (vi) GreenComputing, Smappee-Non-profit organizations that promote sustainable energy consumption and sustainable communities.
- Governance: For the Enosi ecosystem, the Enosi Decentralized Autonomous Organizations (EDAO)-has been proposed to provide ecosystem governance. In the case of Pylon Network, governance is supported through blockchain smart-contract technology.
- Characteristics and relationships: Dominant characteristic is partnerships. Key relationships are distributed systems with centralized platform for management (Offer platform-as-a-service).

RQ2. Interaction with the power grid: For the Enosi ecosystem there is 40–60% reduction in household electricity costs. The Pylon Network provides opportunity to monetize the surplus energy by selling to the grid and also enables a reduction of payback time of DER installation up to 25%. In the same network, Grid connected prosumers and large scale DER facilities are also rewarded for the energy injected into the grid.

RQ3. Role of collaborations: The trading platforms serve as coordination point to coordinate transactions and exchange of value in energy and related services. They offer platform-as-a-service and blockchain smart contracts to facilitate transactions. They do not focus on collaborations and the development of resilient communities.

Limitations of Peer to Peer Sharing

– The case study revealed that these energy trading platforms have weak governance structures. Other cases outside the study like [47–50] shows similar trend.
– These platforms adopt blockchain as a core technology enabler to provide trust-less transactions. However, literature such as [51] and [52] confirm prevailing flaws in blockchain technology which is likely to affect user confidence in terms of reliability, efficiency and safe-use.

– Currently, many jurisdictions, have challenges regulating these P2P platforms. The contention is whether to subject these prosumers to taxes and other regulations. Should they be considered as consumers or business entities? Judgment from a pending legal case at the court of Justice of the European Union that was forwarded from a Spanish court, Case C-434/15, [53] would finally answer some of these questions.
– The focus on these trading platforms is to maximize financial gains through exchange of value in energy and its related services. Interest and benefits are usually skewed towards individual stakeholders in the ecosystem and do not promote communal interest, which invariably affect the platforms resilience.

3.5 IEEE Smart Village (ISV)

ISVs are IEEE Foundation Signature Programs which provide renewable energy solutions and supporting funding to members of local communities, thereby enabling them build small-scale electricity companies, and partnerships that facilitate the establishment of sustainable businesses and training programs to communities [54]. Profits are often reinvested to empower the community through learning and development programs. Currently, ISVs serves over 50,000 people in 34 villages, and is prominent in countries such as Cameroon, Haiti, Kenya, South Sudan, and many more [55]. The SunBlazer II is the ISV flagship mobile solar power unit with a 1.5 kW base station. This is a highly adaptable unit that integrates rechargeable battery packs and home lighting kits with an auxiliary outlet for other DC powered devices [56]. Each kit lights up two rooms and operates auxiliary 12 V DC loads. Each station can charge 80 battery packs every three to four days to serve an estimated 500 people [55]. The generator is configurable to other load options such as schools, churches, community centres, clinics, coolers, small businesses, and internet connectivity. The unit is designed to work with a franchise-type business model. Studied cases:

– Case 1: Haiti [57, 58]
– Case 2: India [59, 60]

Summary of IEEE Smart Village

RQ1. Organization:

- The key roles here include funding partners, local energy entrepreneurs, and non-profit sustainable energy advocacy groups.
- Governance: Governance structures are not clearly defined, however, since these facilities are owned by local energy entrepreneurs, it could be a self-governance framework or arrangement.
- Characteristics and relationships: Key characteristics are franchising. Relationships found are distributed and stand-alone-systems.

RQ2. Relationship with the grid: These units are not connected to the grid. They are stand-alone units.

RQ3. Role of collaboration: Collaborations among stakeholders promote sustainable community businesses, capacity building through training and learning. The outcome of these collaborations results in improved livelihood and community development.

Limitations of IEEE Smart Village

- Limited generation capacity (serves few households at a time).
- Islanded units that cannot be connected to the power grid.
- Limited System scalability.

3.6 Energy Hubs

An energy hub is a multi-carrier energy system that integrates multiple energy technologies such as energy conversion systems, storage and network technologies. Energy hubs take many different forms and ranges from a single building to a larger geographical area. Energy hubs converts different types of energy in a hub to facilitates flexibility in energy provision. As such, energy hubs are particularly useful for enabling the integration of intermittent renewable energy sources such as solar and wind [61]. Currently this concept is being discussed within the research and academic circles. There are no real cases available yet.

4 Research Contribution and Relationship with Collaborative Virtual Power Plants

The main contribution of this work is that it provides grounds to implement our proposed concept of "***collaborative Virtual Power Plant***" [62, 63] which is an ongoing research intended to develop and implement key performance indicators that can be used to assess the global performance of RECs and their constituent prosumers when adopting a collaborative approach. In this regard, we have proposed and developed a new organizational structure for RECs by adopting knowledge and concepts from two disciplines. The two disciplines considered are: (a) Collaborative Networks (CN), and (b) Virtual Power Plants (VPPs). By merging concepts from these two areas, we introduce a new organizational structure that merges the functional characteristics and properties of both VPPs and CNs. In this context, our proposed CVPP ecosystem will be driven by prosumer collaborations that results in the aggregation of DER capacities of prosumers for the purpose of trading energy at the market. The proposed performance indicators are as follows:

1. **Net Economic-Value Indicator**

   This indicator is intended to quantify how much value collaboration can contribute to the economic benefits of the CVPP ecosystem. The indicator will be used to determine the profitability or otherwise of the ecosystem. Furthermore, it could also be used to compare the economic performance of different energy communities and may serve as an important tool for managers of RECs at various levels of aggregation. This indicator would also be used to measure the net economic value of the global CVPP ecosystem and also per prosumer, over a defined period of time.

2. **DER-Capacity Utilization Indicator**

   This indicator is intended to measure how much collaboration can enhance efficient utilization of generation capacity of DERs within our CVPP ecosystem. In the context of the ongoing research, generation capacity is constituted of: 1. DERs like household PV and battery storage system/electric vehicles. 2. Centralized community owned energy storage, and 3. Community owned Wind Turbine. The proposed indicator will be useful in determining how effective generation capacity is being utilized to achieve the objectives of the CVPP ecosystem. This indicator would be used to determine the capacity utilization of DERs at the global CVPP level and also at the prosumer level.

3. **Contribution Indicator**

   This indicator is intended to measure how much each prosumer within the ecosystem contributes towards collaborations. It is expected of each prosumer to contribute towards the sale of energy to the grid. However, it is envisaged that prosumer resources are not always available due to consumption preferences and behaviour. It is therefore useful to introduce an indicator that can measure how much contribution each prosumer made over a defined period of time.

Expected Outcome

By introducing these indicators, we envisage the following:

- Determine the set of conditions that can best influence/enhance the net economic value of the CVPP ecosystem and RECs in general;
- Determine the set of conditions that can help the enhancement of DER capacity utilization in the ecosystem;
- Determine the level of prosumer contributions in the ecosystem to enable appropriate design and deployment of incentives packages that can enhance prosumer collaborations.

Research Approach
The intended approach/methods are through simulations by using System Dynamics, Multi-agent System Technology and Discrete Events.

5 Conclusion

From the conducted study, a number of conclusions can be derived:

- RECs are promising energy sources that can help to reduce the demands on the power grid by providing off-grid power supply services to communities. Although these communities become self-sufficient in terms of energy supply, they also end up being environmentally sustainable communities in the sense that their activities promote sustainable lifestyle within the resident community.

– RECs provide an effective and less expensive alternative to the financing of electrical infrastructure which usually requires a huge initial cost of investment, therefore, making it attractive to only large utility companies.
– The concept of RECs helps to break the monopoly of big utility companies and facilitate energy democracy which is currently being advanced by grassroots activists in the United States and parts of Europe. The idea of energy democracy besides many other things also promote individuals within communities to become investors and economic beneficiaries of energy infrastructure and contribute to the governance and management of these infrastructures.
– RECs can help mitigate the impact of natural disasters like hurricane which can knock down significant power infrastructure, leaving thousands of homes stranded and without electricity for days. RECs like Community Microgrids, Community shared Solar, etc., can help to power critical infrastructure in such circumstances since their power sources are predominantly renewable and they are also designed to be self-sustaining.
– RECs are currently empowering communities to use a bottom-up approach to advance environmental sustainability issues and promote sustainable consumption behaviors at the local community level which is more effective than traditional methods.

Acknowledgements. The authors acknowledge the contributions of the Portuguese FCT-Strategic program UID/EEA/00066/2013 for providing partial financial support for this work. Furthermore, we extend our appreciations to Ghana Educational Trust Fund (GETFund), the University of Energy and Natural Resources and UNINOVA CTS for supporting this work with their research facilities and resources.

References

1. Lawrence, M., et al.: Energy Cloud 4.0. Capturing Business Value through Disruptive Energy Platforms, pp. 1–46 (2018)
2. Sumic, Z., Foust, N., Cohen, E.L., Harrison, K.: Top 10 Trends in 2018 Driving the Utility Industry Toward a Decarbonized, Distributed, Digital and Democratized Future, pp. 1–25, April 2018
3. Camarinha-Matos, L.M., Oliveira, A.I., Ferrada, F., Thamburaj, V.: Collaborative services provision for solar power plants. Ind. Manag. Data Syst. **117**(5), 946–966 (2017)
4. Camarinha-Matos, Luis M., Fornasiero, R., Afsarmanesh, H.: Collaborative networks as a core enabler of industry 4.0. In: Camarinha-Matos, Luis M., Afsarmanesh, H., Fornasiero, R. (eds.) PRO-VE 2017. IAICT, vol. 506, pp. 3–17. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-65151-4_1
5. Camarinha-Matos, L.M., Afsarmanesh, H.: Collaborative networks: a new scientific discipline. J. Intell. Manuf. **16**, 439–452 (2005)
6. Camarinha-Matos, L.M.: Collaborative smart grids-a survey on trends. Renew. Sustain. Energy Rev. **65**, 283–294 (2016)

7. Camarinha-Matos, Luis M., Afsarmanesh, H., Boucher, X.: The role of collaborative networks in sustainability. In: Camarinha-Matos, Luis M., Boucher, X., Afsarmanesh, H. (eds.) PRO-VE 2010. IAICT, vol. 336, pp. 1–16. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-15961-9_1
8. Camarinha-Matos, L.M., Afsarmanesh, H. (eds.): Collaborative Networks: Reference Modeling. Springer, Boston (2008). https://doi.org/10.1007/978-0-387-79426-6
9. Owyang, J., Tran, C., Silva, C.: The Collaborative Economy. Altimeter, p. 28 (2013)
10. Espe, E., Potdar, V., Chang, E.: Prosumer communities and relationships in smart grids: a literature review, evolution and future directions. Energies **11**(2528), 1–24 (2018)
11. Rathnayaka, A.J., Potdar, V.M., Dillon, T.S., Kuruppu, S.: Formation of virtual community groups to manage prosumers in smart grids. Int. J. Grid Util. Comput. **6**(1), 47–56 (2015)
12. Rathnayaka, A.J.D., Potdar, V.M., Dillon, T., Hussain, O., Kuruppu, S.: Goal-oriented prosumer community groups for the smart grid. IEEE Technol. Soc. Mag. **33**(1), 41–48 (2014)
13. Vergados, D.J., Mamounakis, I., Makris, P., Varvarigos, E.: Prosumer clustering into virtual microgrids for cost reduction in renewable energy trading markets. Sustain. Energy Grids Netw. **7**, 90–103 (2016)
14. Zepter, J.M., Lüth, A., Crespo Del Granado, P., Egging, R.: Prosumer integration in wholesale electricity markets: synergies of peer-to-peer trade and residential storage. Energy Build. **184**(2019), 163–176 (2018)
15. Jarnut, M., Wermiński, S., Waśkowicz, B.: Comparative analysis of selected energy storage technologies for prosumer-owned microgrids. Renew. Sustain. Energy Rev. **74**(March), 925–937 (2017)
16. Nagasawa, T., et al.: Accelerating clean energy through Industry 4.0: manufacturing the next Revolution. p. 56 (2017)
17. Acosta, C., Ortega, M., Bunsen, T., Koirala, B.P., Ghorbani, A.: Facilitating energy transition through energy commons: an application of socio-ecological systems framework for integrated community energy systems. Sustainability **10**(366), 1–15 (2018)
18. Neue Energien Forum Feldheim (2018). https://nef-feldheim.info/?lang=en. Accessed 06 Nov 2018
19. Renewable energy: The Feldheim model | VoxEurop.eu: European news, cartoons and press reviews. https://voxeurop.eu/en/content/article/4402611-feldheim-model. Accessed 06 Nov 2018
20. Wildpoldsried - Erneuerbare Energie. https://www.wildpoldsried.de/index.shtml?Energie. Accessed 04 Nov 2018
21. IREN2 research project. https://www.energy.siemens.com/apps/features/iren2/index.html#!/en/. Accessed 27 Nov 2018
22. Siemens, A.G.: How a distributed energy supply works-economically and reliably. pp. 1–2 (2016)
23. Wildpoldsried, Germany : Go100re.net (2014). http://www.go100re.net/properties/wildpoldsried-2/. Accessed 27 Nov 2018
24. Interstate Renewable Energy Council and The Vote Solar Initiative: Model rules for shared renewable energy programs. In: Interstate Renewable Energy Council, pp. 1–28 (2013)
25. Cleveland, M.: State policies for shared renewable energy (2017). http://www.ncsl.org/research/energy/state-policies-for-shared-renewable-energy.aspx. Accessed 04 Dec 2018
26. Coughlin, J., et al.: Guide to Community Solar: Utility, Private, and Nonprofit Project Development, pp. 1–76 (2012)
27. Augustine, P.: The time is right for utilities to develop community shared solar programs. Electr. J. **28**(10), 107–108 (2015)
28. Westmill Solar Co-operative: Rules of Westmill Solar Co-operative Ltd., pp. 1–10 (2011)

29. Westmill Solar Co-operative. http://westmillsolar.coop/. Accessed 07 Nov 2018
30. Bristol Community Energy Limited: Rules of Bristol Community Energy Limited. Great Britain, pp. 1–21 (2018)
31. Bristol Energy Cooperative - Renewable Energy. http://www.bristolenergy.coop/. Accessed 08 Nov 2018
32. Clean Coalation: Clean Coalation: Community Microgrids (2018). http://www.clean-coalition.org/our-work/community-microgrids/. Accessed 16 Nov 2018
33. PV magazine USA: Montecito Community Microgrid Initiative will provide energy resilience to the region (2018). https://pv-magazine-usa.com/press-releases/clean-coalition-facebook-twitter-linkedin-youtube-email-search-our-site-skip-to-content-about-us-our-work-resources-connect-news-events-montecito-community-microgrid-initiative-will-provid/. Accessed 16 Nov 2018
34. Clean Coalition: Feasibility Assessment for the Long Island Community Microgrid Project, pp. 1–116 (2016)
35. Clean Energy Coalition: Community Microgrid Initiative: Innovation for a Clean Energy Future, pp. 1–8 (2015)
36. Clean Coalition: Long Island Community Microgrid Project, pp. 1–2 (2015)
37. Clean Coalition: Long Island microgrid to set stage for huge savings, pp. 1–2 (2015)
38. Steepleton, S.: Montecito sets course for emergency energy independence, pp. 1–2 (2018)
39. Clean Coalation: Montecito Community Microgrid Initiative (2018). http://www.clean-coalition.org/our-work/community-microgrids/montecito-community-microgrid-initiative/#objectives. Accessed 16 Nov 2018
40. World Business Academy: SBR3 - Santa Barbara Microgrid - Safe Energy Project. https://safeenergyproject.org/santa-barbara-microgrid/. Accessed 07 Dec 2018
41. Utility Dive: The utility view of microgrids, pp. 1–12 (2014)
42. Trabish, H.K.: How utilities can leverage their grids to integrate solar faster and cheaper (2015). https://www.utilitydive.com/news/how-utilities-can-leverage-their-grids-to-integrate-solar-faster-and-cheape/393540/. Accessed 07 Dec 2018
43. Zhou, Y., Wu, J., Long, C.: Evaluation of peer-to-peer energy sharing mechanisms based on a multiagent simulation framework. Appl. Energy **222**(February), 993–1022 (2018)
44. Orlov, A., Bjørndal, M.H.: Blockchain in the Electricity Market: Identification and Analysis of Business Models. Norwegian School of Economics & HEC Paris, pp. 1–107 (2017)
45. Enosi: A distributed energy protocol: connecting and automating energy markets with distributed ledger technology, pp. 1–37 (2018)
46. Enosi. Providing consumers with choice, transparency and efficiency, pp. 1–39, April 2018
47. SolarCoin: SolarCoin: a blockchain-based solar energy incentive, pp. 1–13 (2018). https://solarcoin.org/
48. Consensys: GridPlus: welcome to the future of energy, Grid+Whitepaper v2.0. pp. 1–35 (2017)
49. The sun exchange: SUNEX Network Token Whitepaper: Monetizing Sunshine, pp. 1–48 (2018)
50. Vlux: Verv VLUX Whitepaper Version 2.0: The evolution of Energy, pp. 1–55 (2018)
51. FXCM: What Are The Blockchain's Limitations? – FXCM. FXCM Market Insights (2018). https://www.fxcm.com/insights/what-are-blockchains-limitations/. Accessed 27 Dec 2018
52. Berryhill, J., Bourgery, T., Hanson, A.: Blockchains Unchained: Blockchain Technology and its Use in the Public Sector. OECD Working Papers on Public Governance, no. 28, pp. 1–53 (2018)
53. European Court of Justice: Court of justice proceedings. Off. J. Eur. Union **C 72**(2), C 116/2–C 116/3 (2018)

54. Mackenzie, D.: IEEE Smart Village: Sustainable Development- A Global Mission. IEEE Power and Energy Society (2018). http://sites.ieee.org/pes-enews/2018/05/25/ieee-smart-village-sustainable-development-a-global-mission/. Accessed 21 Dec 2018
55. IEEE Smart Village: The Smart Village Initiative, pp. 1–2 (2017)
56. Larsen, R.S., et al.: Learning beyond the Light Bulb' among Least Developed Countries based on a sustainable PV solar utility model. In: Proceedings of the 4th IEEE Global Humanitarian Technology Conference, GHTC 2014, pp. 106–114 (2014)
57. IEEE Smart Village: Haiti - Integrating sustainable electricity, education, and entrepreneurial solutions to empower off-grid communities. IEEE Smart Village (2017). http://ieee-smart-village.org/communities/north-america/haiti/. Accessed 28 Dec 2018
58. IEEE Smart Village: Empowering Off-Grid Communities, pp. 1–2 (2017)
59. IEEE Smart Village: India - Integrating sustainable electricity, education, and entrepreneurial solutions to empower off-grid communities. IEEE Smart Village (2017). http://ieee-smart-village.org/communities/asia/india/. Accessed 28 Dec 2018
60. IEEE Smart Village: About IEEE Smart Village - IEEE Smart Village. IEEE Smart Village (2017). http://ieee-smart-village.org/about/. Accessed 28 Dec 2018
61. Energy Hub: Future Energy Efficient Buildings & Districts (2018). http://www.sccer-feebd.ch/research/energy-hub-definition-advantages-and-challenges/. Accessed 17 Nov 2018
62. Adu-Kankam, K.O., Camarinha-Matos, L.M.: Towards collaborative virtual power plants: trends and convergence. Sustain. Energy Grids Netw. **16**, 217–230 (2018)
63. Adu-Kankam, Kankam O., Camarinha-Matos, Luis M.: Towards collaborative virtual power plants. In: Camarinha-Matos, Luis M., Adu-Kankam, Kankam O., Julashokri, M. (eds.) DoCEIS 2018. IAICT, vol. 521, pp. 28–39. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-78574-5_3

Collaboration and Resilient Systems

Novel Approaches to Handle Disruptions in Business Ecosystems

Javaneh Ramezani[1,2(✉)] and Luis M. Camarinha-Matos[1,2(✉)]

[1] Faculty of Sciences and Technology, NOVA University of Lisbon, Campus da Caparica, 2829-516 Monte Caparica, Portugal
m.ramezani@campus.fct.unl.pt, cam@uninova.pt
[2] UNINOVA-CTS, Center of Technology and Systems, Caparica, Portugal

Abstract. Today's business world is continuously challenged by unexpected disruptive events, which are increasing in their frequency and effects. As a consequence, it is plausible to foresee future scenarios in which turbulence and instability are no longer considered as episodic crises, but rather somewhat the "norm" or the default status. This trend naturally raises the question of how organizations can strive and even gain in such disruptive environments, and which characteristics are required for combating disruptions. Resilience and antifragility are two emerging approaches to handle disruptions. Through a literature review, this paper identifies several strategies that contribute to business ecosystem's resilience or antifragility. Furthermore, it is also shown that contributions from a number of disciplinary areas, including Collaborative Networks, Systems Thinking, Thermodynamics, Management science, and ICT, can provide complementary views and support. A set of promising examples of applications of the discussed approaches are presented and briefly analyzed. Finally, a number of open questions and directions for further research are presented.

Keywords: Resilience · Antifragility · Disruptions · Business ecosystems · Cyber physical systems

1 Introduction

Due to the rise of ever tougher challenges in contemporary business environments characterized by high complexity and uncertainty, there is a clear need for finding new solutions to address unexpected issues and disruptions. Despite the increasing capabilities offered by technology in recent years, there are still significant engineering and philosophical problems that require to be handled to embrace the unknown future [1].

Traditional risk management approaches are not effective when a company is confronted with unexpected disruptions. Those approaches are based on the assumption of some stability of the business environments, and therefore disruptive events such as explosions, or natural disasters are hard to handle deviations. Traditional methods require risk identification and quantification, which rely on past experimental data. Therefore, those methods are not useful to handle low probability, high impact disruptions. Furthermore, traditional strategies are cost-based and narrow-focused and

Published by Springer Nature Switzerland AG 2019
L. M. Camarinha-Matos et al. (Eds.): DoCEIS 2019, IFIP AICT 553, pp. 43–57, 2019.
https://doi.org/10.1007/978-3-030-17771-3_4

therefore usually cause more vulnerability when dealing with unforeseeable events [2, 3]. Therefore, there is a need for the emergence of new paradigms to understand, measure, and respond to these stressful conditions. Resilience and antifragility are two promising concepts to deal with unexpected disruptions [1]. A resilient system survives shocks and stays the same or evolves to another stable state; an antifragile system thrives and improves with shocks [4].

The important question which emerges here is how some organizations can overcome stressful situations better than others. In fact, there are various strategies that companies can use to achieve resilience and/or antifragility in face of disruptions. However, existing studies in the business ecosystems area do not provide a comprehensive collection of strategies to cope with disruptions. Most of the previous studies have reviewed only a few resilience-oriented strategies, and there is no systematic analysis of antifragility-oriented strategies in the literature [5, 6].

This work addresses such lack of information by performing a literature analysis and proposing a categorization of disruption-coping strategies and capabilities in business ecosystems, considering: (1) their relation to resilience and antifragility, and (2) their relation to the phases of disaster management. This study addresses the above issues as follows: First, it provides a taxonomy of disruption-coping strategies. Second, it identifies the link between these strategies and resilience and antifragility. Third, it defines the relationship between the strategies and the three phases of disaster management. Fourth, it categorizes the essential capabilities, which can help organizations to achieve resilience and antifragility.

This research also seeks insights from various disciplines to identify different relevant tools, rules, and other features for a better understanding of the complex challenges in a stressful business environment. Another contribution is the identification of illustrative examples of promising approaches to resilience and antifragility. This survey is guided by the following research questions: What are emergent approaches to handle disruptions? Which knowledge areas are contributing to this issue? Which are promising/illustrative examples of approaches to handle disruptions?

The remaining of the paper is organized as follows: first, a short review of contribution to innovation in industrial and service systems is presented in Sect. 2; and then a set of related concepts to address disruptions in business ecosystems are explained in Section three. The main findings of the study are discussed in Section four, and finally, in Section five conclusions are provided.

2 Relation to Innovation in Industrial and Service Systems

Industry 4.0 is characterized by digitalization, automation and adaptation, communication, optimization and customization, value-added services and businesses. Furthermore, collaboration plays a central role in all dimensions of this industrial revolution [7]. In particular, the cyber-physical system (CPS) and Internet of Things promise new support to business ecosystems namely to allow real-time decisions and increased levels of system's intelligence. Resilience and antifragility are intrinsically connected to CPS, especially in a time where systems are becoming more complex. In fact, the innovations associated with the 4th industrial revolution need to be embedded disruptions-coping strategies, namely: (1) to enhance the ability of the business ecosystems to prepare for and adapt to disruptive market environments and withstand and recover rapidly from the impacts of unforeseen disasters, and (2) to stay competitive in volatile business environments. Therefore, the characteristics of resilience and antifragility including responsiveness, learning, adaptability, etc. need to be included in the design of advanced cyber-physical systems.

3 Base Concepts

In the business ecosystems literature, disruptions have been recognized as one of the critical issues that that can have severe impacts on business and in some cases even firms' closures. In this context, disruption refers to a "predictable or unpredictable event which affects the normal operation and stability of a business" [8, 9]. Generally, there are several ways in which an entity or a system responds to disruptions [4, 10]. In fact, firms face different consequences as a result of unexpected disruptive events: some fail, some survive, and some even grow [4, 11].

A number of relevant properties need to be distinguished here:

- **Fragility**. A system or entity that is vulnerable and can be easily broken when subjected to stress is known as fragile [12]. Fragility implies more to lose than to gain in case of perturbations or disruptions [4].
- **Robustness.** A system or entity that has is not easily affected by hazards and can stay unchanged even when subjected to shocks, is robust [4].
- **Resilience.** Some systems or entities can absorb shocks in such a way that although affected by hazards and temporarily change, they recover from those shocks; they are resilient [13].
- **Antifragility.** Some systems or entities not only have the capability of absorbing shocks and survive but even flourish and get better; they are called antifragile. Antifragility implies more gain than loss because shocks have a positive impact on these entities [4, 12].

While an earlier notion of resilience was focused on the absorptive coping capacity, the concept has been evolving to represent an adaptive and even transformative capacity – transformative resilience, which partially overlaps with the notion of antifragility [14]. In other words, a system does not necessarily return to its original state but may evolve to a slightly different but stable state. Our research work is mostly focused on transformative resilience and antifragility.

4 Results and Discussions

In this section we summarize emerging approaches to handle disruptions, main contributing knowledge areas, and also present some promising examples based on the performed survey.

4.1 Emerging Approaches to Handle Disruptions

Two main aspects are included here: strategies to address disturbances in the business ecosystems and capabilities that help a business to overcome disruptions.

i. **Strategies.** To help a business become less vulnerable to disturbances, there is a need to define adequate disruption-coping strategies [15]. Resilience and antifragility are two emerging approaches to deal with disruptions, enhancing traditional risk management strategies. At least at a conceptual level, they are able to cope with unpredictable uncertainties through their properties such as flexibility, agility, adaptability, etc. [2, 6]. Although there are several surveys about resilience and antifragility at the business ecosystems level the question "What are emergent approaches to handle disruptions?" is not yet answered. Therefore, in this section we propose a categorization of the resilience/antifragility-based strategies to deal with disruptions in line with three general phases of disaster management: Readiness, Response, and Recovery (Tables 1, 2 and 3).

Table 1. Examples of strategies to prevent disruptions and prepare for disaster response.

Strategies	Description	Approach		References
		Resil	Antif	
Forecasting	Refers to techniques based on advanced information systems to predict the market, demand, etc.	✓		[2, 5, 10, 18]
Mapping	Allows for acquiring knowledge about the current state and vulnerabilities of the ecosystem	✓		[19]
Barbell	Investing most of the assets conservatively while taking risks with the rest		✓	[4, 6, 12, 20]
Creating optionality	Giving the buyer/investor freedom to experiment and benefit from opportunities. Ex: lease		✓	[1, 4, 6, 21]
Buffering	Attempts to gain stability by establishing safeguards that protect a firm from disturbances	✓		[2, 5, 6, 8, 17, 22]
Infrastructure investment	Involves building infrastructure, and systemization. Ex: ICT adoption	✓		[5, 23]
Security Compliance	Increasing security using different policies, procedures, and technology against attacks	✓	✓	[2, 3, 5]
Hormesis	Exposing a system to low doses of a harmful "substance or agent" inducing higher resistance of the system when submitted in future to higher doses of the same stressor		✓	[4, 6, 24]

Table 2. Examples of strategies to cope with disruptions in immediate aftermath the event.

Strategies	Description	Approach		References
		Resil	Antif	
Acceptance of disruption risk	Do nothing when the mitigation costs are higher than disruption risks	✓		[15]
Postponement	Delaying operations or activities until the last possible moment to recognize and meet demand	✓		[17]
Revision	Substitution and revision of plan of sourcing, operations, and facility in response to disruptions	✓		[22, 25, 26]
Demand managing	Demand switching through different incentives, and pricing strategy	✓		[3, 15, 22]
Graceful degradation	Allowing limited interaction and avoid entire system's downtime	✓		[27]
Government lobbying	Attempts to influence government decisions by bringing attention to the long-term consequences of the catastrophe	✓		[28]

Table 3. Examples of strategies to return the affected system to a pre-disaster condition.

Strategies	Description	Approach		References
		Resil	Antif	
Integration	Integration and synchronization of individual functional capabilities such as system, resource, process, etc.	✓		[5, 17, 19, 23]
Cycle-time reduction	Service levels can be improved by reducing delivery lead times due to shorter product/process design and development time	✓		[17]
Insurance	Refers to financial risk sharing as it transfers the risk of compensable loss to the insurer	✓		[15, 22]
Customer service	Improve responsiveness to customers to ensure future customer loyalty	✓		[5, 17]
Knowledge management	Understanding business ecosystem disruptions with the capacity to learn from past disruptions to develop better preparedness for future events	✓		[17, 26]
Feedback mechanism	Prevent the same "attack"/shock from negatively affecting the system more than once. The negative feedback mechanism attempts to reverse the changes and restore the system to the normal state		✓	[20]

- **Readiness.** Disruption "readiness" involves investment in advance of a disaster to prevent, detect and eliminate the source of possible disruptions or mitigate system's vulnerability. For instance, mitigation strategies such as increasing security, and choosing facility location are proactive and focus on avoiding disruptions [16–18].
- **Response.** Disruption "response" is related to the immediate actions to take after a disaster occurs in order to bring the system into a temporarily acceptable operation state [10, 18]. This phase involves strategies that reactively cope with disruptions such as demand management, and multi-sourcing [15].
- **Recovery.** Disruption "recovery" is essentially concerned with a set of activities to return the system to a pre-disruption state or, preferably, improved levels of operation. This phase involves efforts to lead the system to a long-term stable state [10, 16–18].

ii. **Capabilities.** There are several critical capabilities that influence the resilience/antifragility ability of a business system to overcome disruptions [21, 29]. Table 4 provides a summary of relevant capabilities or attributes.

Table 4. Examples of attributes that enable a business to overcome disruptions.

Capabilities (Attributes)	Description	Approach		References
		Resil	Antif	
Flexibility	Ensures that the changes caused by a risk event can be absorbed through useful reactions	✓		[2, 23, 25, 30]
Redundancy	Having multiple assets/sources able to perform the same function	✓		[6, 17, 19, 22, 25]
Convexity	Things with positive asymmetries that expose to exponentially more benefit as uncertainty increases		✓	[1, 6, 12]
Agility	Ability of a business to rapidly respond to changes in environment	✓		[17, 25, 26, 30, 31]
Simplicity	Ensures decrease of complexity by removing fragile and harmful elements		✓	[4, 21]
Visibility	Traceability of products and the environment of a business ecosystem	✓		[2, 17, 23, 26, 30, 31]
Creativity	Flexibility of thinking, perceptiveness of problems, and redefine and elaborate ideas		✓	[12, 14, 24]
Financial strength	Capacity to retain volatility in cash flow	✓		[2, 3, 5, 22]

The above lists of examples are not exhaustive but nevertheless provide a global overview of existing proposals.

4.2 Contributing Knowledge Areas

Resilience and Antifragility in disruptive business ecosystems are related to or can benefit from different concepts introduced in various disciplines such as collaborative networks, systems thinking, thermodynamics, social science, biology, management science, and ICT. Table 5 further details how each discipline addresses different aspects of resilience and antifragility.

4.3 Promising Examples

Table 6 includes some representative examples of systems adopting some of the described strategies.

5 Future Direction

The current literature on resilience and antifragility mostly remains at a conceptual and some even at speculative level. The underlying mechanisms are not yet adequately understood, and thus the design of architectures and governance mechanisms for resilient/antifragile systems is not trivial. Therefore, it is necessary to move from these theoretical approaches to practical systems that enhance their strength through experience and error in complex and disruptive scenarios. Next stage of our planned work is to suggest a self-adaptive system that can help the creation of collaborative business ecosystems that can learn from turbulence and improve when facing disruptions (i.e., become transformative resilient or antifragile). However, one crucial issue is how to evaluate such approaches? Since the behavior of complex/chaotic systems is unpredictable/non-linear, we cannot apply conventional processes, and also waiting for disruptions to happen to learn from them is too risky. On the other hand, using the real business ecosystems to "play" with them and test new ideas is not feasible. Therefore, the planned approach is to develop a multi-agent-based and system dynamics-based modeling and simulation framework to allow testing new solutions and their effectiveness on the context of disaster scenarios. The simulation of the disaster response activity is achieved by modeling each element involved in the ecosystem as an agent. Every agent learns about its environment and collaborates with other agents. Agents execute autonomously and make their own decisions about future actions. Therefore, the ongoing work includes: (1) Analyzing the combination of different strategies/capabilities and how they influence the business ecosystems' resilience/antifragility. The challenge is how to design and select an appropriate combination of strategies to make a system resilient/antifragile?; (2) Collecting empirical evidence to identify the potential of cyber-physical systems usage in the business ecosystem resilience/antifragility; (3) Propose guidelines to build resilient/antifragile systems for volatile and complex environment: (4) Introducing a mechanism to measure the level of resilience/antifragility in the ecosystem.

Table 5. Summary of contributions from different knowledge areas

Area	Description	Refs
Collaborative networks	**Scope:** The area of collaborative networks focuses on the structure and behavior of networks of autonomous entities that collaborate to achieve common goals. **Main contributions:** Collaboration typically facilitates building resilience and antifragility in the business ecosystem in the following ways: - Through negotiation, and collaborative problem-solving. - Creating a culture of trust and knowledge sharing. - Identification of new opportunism by increasing communication. **Recent related approaches:** - Collective awareness: Which refers to supporting environmental awareness, to influence demand changes. - Knowledge co-production: Combining a plurality of knowledge sources to generate new knowledge to address a disruption. - Collaborative adaptive management: A kind of learning-based collaborative approach that links actors' knowledge and experience to respond to disruptions.	[29, 31–33]
Systems thinking	**Scope:** This area offers a holistic approach to problem-solving that attempts to view systems from a broad perspective rather than focusing only on specific events. **Main contributions:** Systems thinking tools allow a deeper understanding of the relationships, and interactions among the components of complex systems. **Recent related approaches:** - System dynamics: A useful approach in understanding the non-linear behavior of complex systems over time using stocks, flows, and internal feedback loops. - Systems engineering: An interdisciplinary collaborative approach that uses systems thinking principles and tools including modeling and simulation, to manage complexity. - Chaos theory: A useful conceptual framework to describe the unpredictability of business ecosystems. - Complexity theory: Understanding how organizations adapt to and cope with uncertainty (e.g. self-organizing).	[10, 21, 24, 26, 33]

(*continued*)

Table 5. (*continued*)

Area	Description	Refs
Thermodynamics	**Scope:** A branch of physics concerned with heat, temperature and their relation to energy. According the second law of thermodynamics entropy always moves from order to chaos (growth of entropy). According to the new meaning of the second law, open and non-linear systems like complex adaptive systems can overcome and recover from disruptions as they act far from equilibrium and can move from one stable state to another where entropy may decrease. **Main contributions:** - Describing the evolution/transformation behavior of complex systems. - Providing heuristics and adaptability-based modeling. **Recent related approaches:** - Phase transition: A useful concept to describe the transition of a complex and chaotic system from one phase to another. - Multi-equilibria: A way of modeling resilience as it keeps in play both stability and innovation through adaptation. - Path-dependence: An evolutionary approach related to the "hysteresis" concept that implies disruptions having persistence effects on the subsequent trajectory of chaotic systems.	[4, 24, 34, 35]
Management science	**Scope:** This discipline covers the application of systematic methods to solve problems and decision making in organizations. **Main contributions:** - Decision making under uncertainty. - Achieving sustainable competitive advantage. - Creating learning organizations. **Recent related approaches:** - Learning-by-doing: An approach towards adaptive governance to simultaneously manage and learn disruptions. - Y-management: Adopting a decentralized, participative management style which assumes that employees are self-motivated (tinkerers). - Convex heuristics: Heuristic based decision-making rules from risk management and antifragility metaphor.	[1, 4, 17, 36–38]

(*continued*)

Table 5. (*continued*)

Area	Description	Refs
Information and Communication Technologies (ICT)	**Scope:** ICT can be viewed as the application of computer science and engineering techniques in interconnected disrupted environments. **Main contributions:** - Improving information management and collective awareness of the disrupted environment. - Embedding resilience/antifragility-based mechanisms in all phases of disaster management (readiness, response, recovery). - Learning from past experiences using tools like machine learning. **Recent related approaches:** Two dimensions: 1- To facilitate adding resilience/antifragility to the business ecosystems: - Microservices: Providing autonomous structure with scalability, self-managing, and flexibility to test, and replace services. - Multi agent-based modeling: Providing an environment to test intelligent autonomous solutions and their effectiveness on the context of disaster response. - Chaos engineering: Experimenting on distributed systems to understand the behavior of systems in the face of disruptions. - DevOps: A software development methodology which focuses on combing software development (Dev) with information technology operations (Ops) to shorten the systems' development life cycle and move toward antifragile organizations. 2- To make the ICT systems more resilient/antifragile: - Automatic bug detection and repair: Refers to software that fixes own bugs using automatic runtime bug fixing capabilities. - Auto-scaling feature: A system with this feature can measure and then dynamically scale up or down to respond to demand stressors and changes. - Continuous deployment: A process with four sub-dimensions (deploy to production, verify the solution, monitor for problems, and respond and recover) in which features, and bug fixes are continuously released in production environment. - Failure-as-a-Service (FaaS): Allows cloud services to routinely perform large-scale, online failure in real deployments.	[1, 39–43]

Table 6. Summary of promising examples

Name	Description	Refs
Netflix	**What is it?** An on-demand media service provider that is also pioneer in the area of antifragile internet-based systems. **How is antifragility/resilience addressed?** Netflix applies the "simian army," a suite of tools referred to as "monkey services," that routinely generate real system failures with the aim of using lessons learned to prevent more massive disruptions and build up resistance against future stressors. **What are related strategies/capabilities?** - Fault injection: Failing fast using software tools to decrease the probability of unexpected response. - Modularity: Exploiting microservices architecture which is running in the Amazon Web Services. Each of the microservices focuses on their work individually with the benefit of flexibility and diversity in use. - Weak links: Using circuit breakers to ensure the services are weakly connected. A circuit breaker quickly detects a problem and breaks the weak link to stop failure propagation to other services. - Redundancy: Ensuring the availability and durability of data through redundant network and data storage. **Characteristics:** Fault tolerance and isolation, Graceful degradation, Learning, Scalability, Bottom–up tinkering, Fail fast.	[20, 39, 43]
ASCENS	**What is it?** Project to develop an integrated set of methods and tools to build autonomic service-component ensembles capable of local and distributed fault reasoning. **How is antifragility/resilience addressed?** Relying on a comprehensive approach to engineer autonomic service components which offers both pragmatic and formal theories and methods to support modeling, reasoning, monitoring and dynamic adaptation. These service-component ensembles are multi-agent systems in the form of intelligent swarms that can adapt at runtime, adjusting to the state of the environment and acquiring knowledge about themselves, other service components, and their environment. **What are related strategies/capabilities?** - Self-managing: Autonomic self-adaptive systems with self-managing "objectives provide autonomy features in the form of a system's ability to automatically discover, diagnose, and cope with various problems." - Feedback mechanism: Three different feedback loops that enables continuous improvement. - Self-healing: Nodes are self-aware of changes in load and of the network structure, which calls for self-healing properties. - Redundancy: Using redundant data storage to prevent data loss in cases where nodes drop out of the system. **Characteristics:** Swarming, Fragmentation, Self-aware, Self-adaptive, Self-organize.	[42]

(*continued*)

Table 6. (*continued*)

Name	Description	Refs
Tinkering school	**What is it?** "Tinkering" is at the core of student's educational philosophy and the basis of everything children do. Children with freedom to play learn life lessons through "tinkering" and gain experience of things such as uncertainty, and failure that prepare them for real world conditions. **How is antifragility/resilience addressed?** Tinkering school uses real tools, real materials, and practical problems in different ways to intentionally force students into disruption situations to encourage them to quick response and prototyping. Through different play and team working technics they learn about creativity, responsibility, persistence, adaptability, which are critical characteristics of resilient/antifragile people. **What are related strategies/capabilities?** - Learning by doing: Encourage children to create things bigger than usual in order to make mistakes and learn from them. - Trial and error: Putting children in situations that love mistakes (to become antifragile) by making numerous errors which are small in harm, even reversible and quickly overcome them. - Collaboration: Through trust and creativity leads to learning how to collaboratively design, build, solve problems to develop capable, adaptable citizens of the world. - Transformability: Related to changes of how the kids see the world and themselves through learning the ability to overcome stressors, thinking of non-obvious solutions (creativity), and confidence to change when things aren't working (adaptability). **Characteristic:** Adaptability, Tinkering innovation, and creativity, Self-reliance, Autonomy, Self-sufficiency, Growth mindset.	[4, 38, 42]
Kubernetes	**What is it?** An open-source platform to automate, deploy, and manage containerized applications and services. **How is antifragility/resilience addressed?** Kubernetes is a resilient container orchestration technology that can spread service instances on multiple nodes using an anti-affinity feature to reduce correlated failures. **What are related strategies/capabilities?** - Self-aware: Detecting the breakdown, routing around it, and taking corrective actions automatically. - Real-time monitoring: Performing regular health checks to detect failures in the services. - Fallback and graceful degradation: Restarting the container and bringing the application back to a healthy state when a failure is detected. **Characteristic:** Auto deployment, Auto-scaling, Observability, Self-healing, Self-management, Decentralization, Isolation.	[41]

6 Conclusions

Resilience and antifragility are critical properties needed in future business ecosystems to allow them to survive and improve in turbulent and disruptive environments. Although there is already a good amount of literature on this subject, we are still far from having effective solutions. Through a multidisciplinary literature review, this paper analyzed different perspectives, concepts and approaches to overcome disruptions in the business ecosystem. Specifically, "strategies," "capabilities," "knowledge areas," and "promising examples," related to the emerging concepts of resilience and antifragility under the light of industry 4.0 were collected and current limitations identified.

Further ongoing research work is focusing the development of a simulation environment where different architectures and combinations of strategies can be assessed.

Acknowledgments. This work was funded in part by the Center of Technology and Systems (CTS) and the Portuguese Foundation for Science and Technology (FCT) through the Strategic Program UID/EEA/00066/2013.

References

1. Russo, D., Ciancarini, P.: Towards antifragile software architectures. Procedia Comput. Sci. **109**(2016), 929–934 (2017)
2. Pettit, T.J., Croxton, K.L., Fiksel, J.: Ensuring supply chain resilience: Development and implementation of an assessment tool. J. Bus. Logist. **34**(1), 46–76 (2013)
3. Fiksel, J., Polyviou, M., Croxton, K.L., Pettit, T.J.: From risk to resilience: learning to deal with disruption. MIT Sloan Manag. Rev. **56**(2), 79–86 (2015)
4. Taleb, N.N.: Antifragile: Things that Gain from Disorder. Random House Publishing Group, New York (2012)
5. Chowdhury, M.M.H., Quaddus, M.A.: A multiple objective optimization based QFD approach for efficient resilient strategies to mitigate supply chain vulnerabilities: the case of garment industry of Bangladesh. Omega (U. K.) **57**, 5–21 (2015)
6. Derbyshire, J., Wright, G.: Preparing for the future: development of an 'antifragile' methodology that complements scenario planning by omitting. Technol. Forecast. Soc. Chang. **82**, 215–225 (2014)
7. Camarinha-Matos, L.M., Fornasiero, R., Afsarmanesh, H.: Collaborative networks as a core enabler of industry 4.0. In: Camarinha-Matos, L.M., Afsarmanesh, H., Fornasiero, R. (eds.) PRO-VE 2017. IAICT, vol. 506, pp. 3–17. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-65151-4_1
8. Hou, J., Zeng, A.Z., Zhao, L.: Coordination with a backup supplier through buy-back contract under supply disruption. Transp. Res. Part E Logist. Transp. Rev. **46**(6), 881–895 (2010)
9. Habermann, M., Blackhurst, J., Metcalf, A.Y.: Keep your friends close? Supply chain design and disruption risk. Decis. Sci. **46**(3), 491–526 (2015)
10. Chroust, G., Kepler, J., Finlayson, D.: Anticipation and systems thinking: a key to resilient systems. In: Proceedings of the 60th Annual Meeting, pp. 1–12 (2017)

11. Blos, M.F., Wee, H.M., Yang, W.: Supply chain risk management: introduction to risk in supply chain. In: Lu, J., Jain, L.C., Zhang, G. (eds.) Handbook on Decision Making, pp. 219–236. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-25755-1_12
12. Hespanhol, L.: More than smart, beyond resilient: networking communities for antifragile cities. In: Proceeding of the 8th International Conference Communities Technology, pp. 105–114 (2017)
13. Bhamra, R., Dani, S., Burnard, K.: Resilience: the concept, a literature review and future directions. Int. J. Prod. Res. **49**(18), 5375–5393 (2011)
14. Gotham, K.F., Campanella, R.: Toward a research agenda on transformative resilience: challenges and opportunities for post-trauma urban ecosystems. Crit. Plan. **17**(Summer), 9–23 (2010)
15. Tomlin, B.: On the value of mitigation and contingency strategies for managing supply chain disruption risks. Manage. Sci. **52**(5), 639–657 (2006)
16. Craighead, R.B., Blackhurst, C.W., Rungtusanatham, J., Handfiel, M.J.: The severity of supply chain disruptions : design characteristics and mitigation capabilities. Decis. Sci. **38**(1), 131–156 (2007)
17. Ponomarov, S.Y., Holcomb, M.C.: Understanding the concept of supply chain resilience. Int. J. Logist. Manag. **20**(1), 124–143 (2009)
18. David, P., McEntire, A.: Disaster Response and Recovery: Strategies and Tactics for Resilience. Wiley, Hoboken (2014)
19. Carvalho, H., Machado, V.C., Tavares, J.G.: A mapping framework for assessing Supply Chain resilience. Int. J. Logist. Syst. Manag. **12**(3), 354 (2012)
20. Devendorf, E., Zeliff, K., Jabbour, K.: Characterization of antifragility in cyber systems using a susceptibility metric. In: Computers and Information in Engineering Conference, pp. 1–11 (2017)
21. Esc, G., Ea, C.: Anti-Fragile Information Systems. In: ICIS 2015 Proceedings, pp. 1–19 (2015)
22. Tomlin, B., Wang, Y., Park, K., Vakharia, A.J., Yenipazarli, A., Tomlin, B.: Operational strategies for managing supply chain disruption risk. Manag. Sci. **December**(4), 639–657 (2011)
23. Atan, L., Snyder, Z.: Supply Chain Disruptions and Corporate Performance. Springer, New York (2013)
24. Dahlberg, R.: Resilience and complexity. J. Curr. Cult. Res. **7**(April), 541–557 (2015)
25. Carvalho, H., Barroso, A.P., MacHado, V.H., Azevedo, S., Cruz-Machado, V.: Supply chain redesign for resilience using simulation. Comput. Ind. Eng. **62**(1), 329–341 (2012)
26. Blackhurst, J., Dunn, K.S., Craighead, C.W.: an empirically derived framework of global supply resiliency. J. Bus. Logist. **32**(4), 374–391 (2011)
27. Vecchiola, C., et al.: Engineering resilient information systems for emergency management. IBM J. Res. Dev. **57**(5), 2:1–2: 12 (2013)
28. Diabat, A., Govindan, K., Panicker, V.: Supply chain risk management and its mitigation in a food industry. Int. J. Prod. Res. **50**(11), 3039–3050 (2012)
29. Camarinha-Matos, Luis M.: Collaborative networks: a mechanism for enterprise agility and resilience. In: Mertins, K., Bénaben, F., Poler, R., Bourrières, J.-P. (eds.) Enterprise Interoperability VI. PIC, vol. 7, pp. 3–11. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-04948-9_1
30. Helen, M.C.: Building the Resilient Supply Chain. Int. J. Logist. Manag. **15**(2), 1–14 (2004)
31. Soni, U., Jain, V., Kumar, S.: Measuring supply chain resilience using a deterministic modeling approach. Comput. Ind. Eng. **74**(1), 11–25 (2014)

32. Graça, P., Camarinha-Matos, L.M.: Performance indicators for collaborative business ecosystems—Literature review and trends. Technol. Forecast. Soc. Change **116**, 237–255 (2017)
33. Berkes, F.: Environmental governance for the anthropocene? Social-ecological systems, resilience, and collaborative learning. Sustainability **9**(7), 1232 (2017)
34. Gómez-Uranga, M., Etxebarria, G.: Thermodynamic properties in the evolution of firms and innovation systems. Ssrn, pp. 1–14 (2015)
35. Chen, S.H., Li, S.P.: Econophysics: bridges over a turbulent current. Int. Rev. Financ. Anal. **23**, 1–10 (2012)
36. Tseitlin, A.: The antifragile organization embracing failure to improve resilience and maximize availability. ACM Queue **11**(6), 1–7 (2013)
37. Jaaron, A., Backhouse, C.J.: Building antifragility in service organisations: going beyond resilience. Int. J. Serv. Oper. Manag. **19**(4), 491 (2014)
38. Moscovich, E.: Embracing the hydra. CA Technol. Exch. DISRUPTIVE Technol. **4**(3), 74–86 (2010)
39. Hole, K.J.: Anti-fragile cloud solutions. In: Anti-fragile ICT Systems, pp. 47–56. Springer (2016)
40. Hole, K.J.: Principles ensuring anti-fragility. In: Hole, K.J. (ed.) Anti-fragile ICT Systems, pp. 35–43. Springer, Cham (2016)
41. Ibryam, B.: From Fragile to Antifragile (2016). https://developers.redhat.com/blog/2016/07/20/from-fragile-to-antifragile-software/
42. Zwieback, D.: Antifragile Systems and Teams. O'Reilly Media, Sebastapol (2014)
43. Monperrus, M.: Principles of Antifragile Software. In: Proceeding of the International Conference on Software Engineering Programming, pp. 1–4 (2017)

Proposing a Risk Management Model in Construction of Combined-Cycle Power Plant Projects

Shakib Zohrehvandi[1(✉)] and Alexandra Tenera[2]

[1] Department of Industrial Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran
shakib.zohrehvandi@srbiau.ac.ir

[2] Department of Industrial and Mechanical Engineering, UNIDEMI, Faculty of Sciences and Technology (FCT), Universidade NOVA de Lisboa, Caparica, Portugal
abt@fct.unl.pt

Abstract. Companies are increasingly focusing on identifying risks and managing them even before they affect project executing. Risk management can play a key part in identifying problems and taking preventive measures in this regard. This research focused on risk management in the implementation of a combined cycle power plant project. This paper aims at exploring the role of risks and risk management practices in construction of this type of projects by considering how they can impact the project's time, quality, and cost. First, the project risk types were identified. The extent to which the project's objectives were achieved was then considered on the basis of three criteria: time, cost, and quality. Next, the risks were categorized and prioritized through a risk breakdown structure. The obtained results revealed the most influential risks in the project. Finally, the proposed risk management model was used in the project which led to strengthening the performance impact of risk management practices in the risky project environment.

Keywords: Project management · Risk management · Combined cycle power plant (CCPP) projects

1 Introduction

Electrical energy is the most convenient form of energy for most human uses. For this reason, demand for electricity is set to grow even further over the next few years. This research was focused on risk management in the implementation of Samangan combined cycle power plant (SCCPP) project. Samangan is the name of this project that located in Sirjan, Iran. The country's increasing demand for electricity has led to massive investment on building power plants. However, uncertainties, different layers of decision-making, and the increasing complication of technologies in this industry has created a multitude of risks and challenges in achieving the project's objectives [1]. The aim of project risk management is to decrease the likelihood of failure of the project. According to research, risk management in project portfolios must be carried

Published by Springer Nature Switzerland AG 2019
L. M. Camarinha-Matos et al. (Eds.): DoCEIS 2019, IFIP AICT 553, pp. 58–69, 2019.
https://doi.org/10.1007/978-3-030-17771-3_5

out by adopting a broader standpoint than that of the individual risk in the project [2]. Managing the project risks is one of the main responsibilities of the project manager. However, if risk management has not been properly done from the start of the project, the responsibility becomes complicated and even ineffectual. The goal of this research is to identify and prioritize the risks and explore the role of risks and risk management practices in construction of SCCPP project in terms of their impact on the project's time, quality, and cost. Based on the proposed approach applied in this research, the risks involved in the SCCPP project will be identified and analyzed. So, the research question is, does the risk management model used in SCCPP project lead to strengthening the performance impact of risk management practices in the project risky environments? Through the remainder of the paper, the relationship to industrial and service systems, literature review, methodology, case study, and the finding sections will be presented respectively. The paper continues with results, main discussion, and lastly, research conclusions.

2 Relationship to Industrial and Service Systems

The power generation industry and its service sectors are going through profound transformation towards digitalization and integration of new levels of smartness. Risks have significant impacts on power generation projects' execution, systems, products, value chains, and business models, including exceeding the estimated cost, failing to complete the project in the due date, and decline in quality and productivity. Considering the large amount of investment required in these projects, failing to reach the time, cost, and quality objectives in power plant projects can cause huge financial losses. This shows the importance of risk identification and management in this industry which is requiring increased level of real time data availability and exchange to continuously monitor and control risk changes and support smart risk plan implementations facing mostly human skills and knowledge deficiencies, improper contracts and import constraints and sanctions.

It is quite characteristic that electric power industry, as a key part of economy in any country, has a deep and versatile effect on the performance of industrial outfits. The power sector appears to be a crucial factor in enhancing universal corporate competitiveness and a driver of economic development [3]. So, in order to ensure the stepwise integrated growth of the power sector, it is essential to create an efficient sectorial risk management system to enable the power generating firms to instantly identify and minimize the influence of appearing threats [4]. This case study is carried out on a CCPP project by Mapna Group in Samangan project. Mapna is an umbrella name for a group of industrial companies tasked with construction and installation of power generation equipment, including wind, steam, and gas turbines, electrical generators, heat recovery steam generators, electrical and control systems, conventional boilers, and railway locomotives.

3 Literature Review

Risk management is indeed a fundamental component of project management. Many research projects have been conducted with different methods in various domains focusing on risk management. The results of these studies led to a categorization of risks within this domain. Nie et al. launched an interval-stochastic method for risk management to consider the notion of risk in stochastic programing and to control the resource cost variability. The model was formulated to identify the optimal power generation mix for Beijing's energy system. The obtained results revealed that the decision makers' risk attitude along with the inherent uncertainty of system components can significantly affect the power-generation schemes, the city's energy supply, the probabilistic penalty, and the system's total cost [5]. A new financing tool called hybrid bond was introduced by Lee and Zhong to launch projects related to renewable energy. The instrument can support the initial capital costs, and manage the risks involved in renewable energy investment [6]. Serpella et al. considered two kinds of risk management techniques: preventive risk management techniques applied prior to the start of a project to manage anticipated risks that are likely to occur during the execution of the project; and remedial risk management techniques used during the project's execution phase when a risk has occurred already [7]. Gatzert and Kosub identified recurrent risks and presented risk management solutions to be applied in renewable energy projects with the aim of identifying critical gaps in risk transfer, and thereby distinguishing between offshore and onshore wind parks with focus on the European market [8].

Hwang et al. investigated risk management in small scale projects in Singapore with respect to barriers, status, and the impact which risk management can have on project performance. They provided a detailed explanation of risk management in those projects and marked the main benefits of risk management for small projects participants [9]. Desai and Kashiyani reviewed construction projects works exposing the knowledge about Insurance as a risk management instrument in construction industry [10]. More recently Cheraghi et al. established a mathematical programing model for choosing risk response strategies in construction projects. The presented model is centered on the project classical triangle of time, quality, and cost in order to find the best risk response strategy in construction projects [11]. Liu and Zeng used the system dynamics method to study the risks involved in renewable energy investment. The results showed that the investment in the early developmental stage was mainly affected by policy risk, that gradually declines the same as technology risk, while market risk gradually becomes the main uncertainty that affects the investment in the mature developmental stage [12]. Qazi et al. considered identifying the critical risks and selecting appropriate risk mitigation strategies as decision problems at the start of a project, taking account of the decision maker's utility function regarding the significance of project goals and the holistic interaction between project risk and complexity [13].

Carvalho and Rabechini exposed how risk management contributes to project success, taking into account the contingent effect of project complexity [14]. Violante et al. highlighted the difficulties small construction companies have dealing with risk management and tried to describe how they manage risks and which aspects of their practice in risk management can be enhanced [15]. Titarenko et al. described the statistical and probabilistic methods applied in managing the investment risks in construction projects and emphasized the necessity of adopting robust procedures when dealing with conditions of significant uncertainty [16]. Reddy attempted to gain a general understanding of risks and their consequences in construction projects and the process through which they are managed. The impact of risk on project assessment is discussed as well as the risk management instruments and methods in construction industry [17]. Zhao et al. identified the critical obstacles to enterprise risk management implementation in Chinese construction companies and explored the interrelationships amongst these obstacles [18]. Kang et al. studied the current status of risk management practiced in the Malaysian construction industry in an attempt to assess the process, as well as the various techniques and tools currently used to manage the projects [19].

4 Methodology

As exposed, there are various risk management models for construction projects. Despite the considerable differences among them, they are all applied to serve the same purpose reduce or prevent negative results on the project objectives. In this study the risk management proposed model was carried out using the project operational requirements into account. The manner of directing risk management activities in the project was defined in risk management planning. Next, different types of risks involved in building power plants in the company's projects were identified based on experts' opinions, modeling, and examining previous projects. In this phase, which is referred to as risk identification, a list of events that might probably occur was developed. The project's success was then assessed on the basis of the three criteria of time, execution quality, and cost. Mahjoub et al. examined the role of strategic talent management in project success, considering the organizational commitment, job satisfaction and motivation as mediators [20]. The prioritization of the risks for further analysis and evaluation was carried out combining each probability of occurrence and their impact. The Probability and Impact Matrix, which is a prevalent risk analysis tool, was also used in this study. The impact of each risk will be separately examined using the probability and impact matrix; however, a simultaneous risk affects of the three criteria of time, quality, and cost of the project was also undertaken. In this research, the quantitative method for data collection were applied using interviews and questionnaires conducted by different project managers and experts (45 experts). First, the experts working on the sites of the projects, and the experts working in the project's central office were identified. The experts all had more than 12 years of experience in risk management and executing power plant construction projects with bachelor or master degree. The closed interviews have been conducted with the average duration of 30 min. The whole interview process took about 5 weeks.

The first and second levels of the risk breakdown structure of the SCCPP project is shown in Fig. 1 which represent an organized and hierarchical illustration of the main potential risks of the project, ordered in the form of sets and subsets. Generally, there are different classifications of the risks with regards to the project's objectives. So, the risk breakdown structure helps to identify the causes of occurrence and the sources of different types of project risks, if categorized according to their potential source of occurrence.

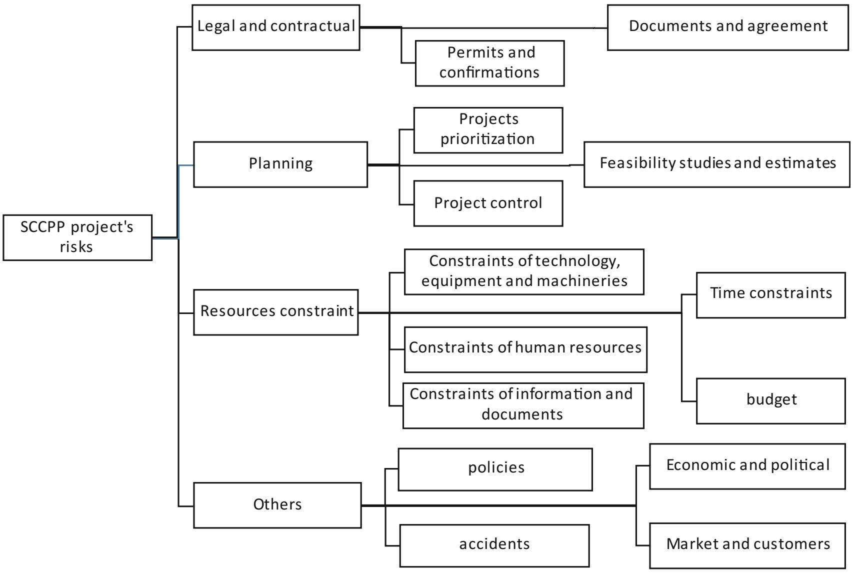


Fig. 1. The SCCPP project's risk breakdown structure

5 The SCCPP Case Study

The manner of directing the risk management activities in the SCCPP project was determined through risk management planning. Considering the fact that there should be no power outage throughout the power grid, projects must be conducted immediately within high quality level requirements, and based on the estimated cost. Therefore, the main goals of the projects under investigation pertained to quality, cost, and time. Hosseini et al. suggested a multi-objective mathematical model to discover the optimal balance amongst time, quality, and cost of construction projects which involved using reinforced concrete in underground structures according to existing risks [21]. It should be noted that the importance of these goals vary according to the type and nature of the projects and their particular circumstances.

The main phases of the SCCPP projects are as follows: preliminary studies, selection and occupation of land, selection of a consultant and signing a contract, selecting a contractor, project execution, pre-commissioning, commissioning, and final delivery. The nature of the problems in CCPP projects is generally related to investment, high technology, and high volume of projects. Zohrehvandi et al. proposed a procedure for going through the closing process group in mega projects. This research was done on a CCPP project as a case study [22]. Some experts believe that in risk management process, the most important phase is risk identification. Stosic et al. addressed the improvement of innovation project risk identification by applying Risk Breakdown Structure methods, and present a follow-up analysis of the subject [23]. The major output of this phase is identification and categorization of the risks and their sources. In this paper, the parties involved in the projects including the employers, contractors, supervisors, and consultants were interviewed with the aim of identifying the risks. The documents produced in the previously conducted projects were also examined in order to determine the reasons why some projects' objectives were not met. They were also studied for benchmarking purposes.

The examination of the SCCPP project led to the identification of a few risks, and each risk was assigned to a subset of the already prepared risk breakdown structure based on its source. In fact, Fig. 2 shows a more detailed level of risks compared to Fig. 1. In order to manage and properly confront the identified risks, they must then, categorized in order of importance. In this way the more serious risks, and their impacts in a greater focus on projects' risky aspects and parts, will be exposed. The methods of risk analysis and assessment are generally classified into two main groups: qualitative and quantitative methods. In qualitative risk analysis, typically the risks are categorized according to their probability of occurrence and their possible impact on the project, measured through a descriptive or relative approach. Therefore, qualitative approaches are a suitable tool for ranking and determining the relative priority of the risks for further actions. Therefore, in this research, the quantitative method for data collection were applied detailed in next section.

6 SCCPP Study Main Findings

Kırılmaz and Erol investigated the supply chain risk management process and suggested a procedure in the risk mitigation phase. The model was tested and analyzed by Probability and Impact Matrix [24], which is a prevalent risk analysis tool, also used in the case. This approach considers the probability of each risk and its possible impacts on project goals with respect to time, quality, and cost. In order to determine risk significance and its impact on project objectives, the risk assessment or Probability and Impact Matrix is applied. This matrix helps to determine the combination of probability and impacts in a way that the risks can be categorized in low, medium, and high-priority risks.

In this research, questionnaires were designed to gather the required data and information from SCCPP project experts and stakeholder and a quantitative method for data collection were applied. The risks impact on the criteria were examined through a five-point Likert scale (ranging from too little to too much). As an example, Table 1 shows the ranking of the risks impact on project duration.

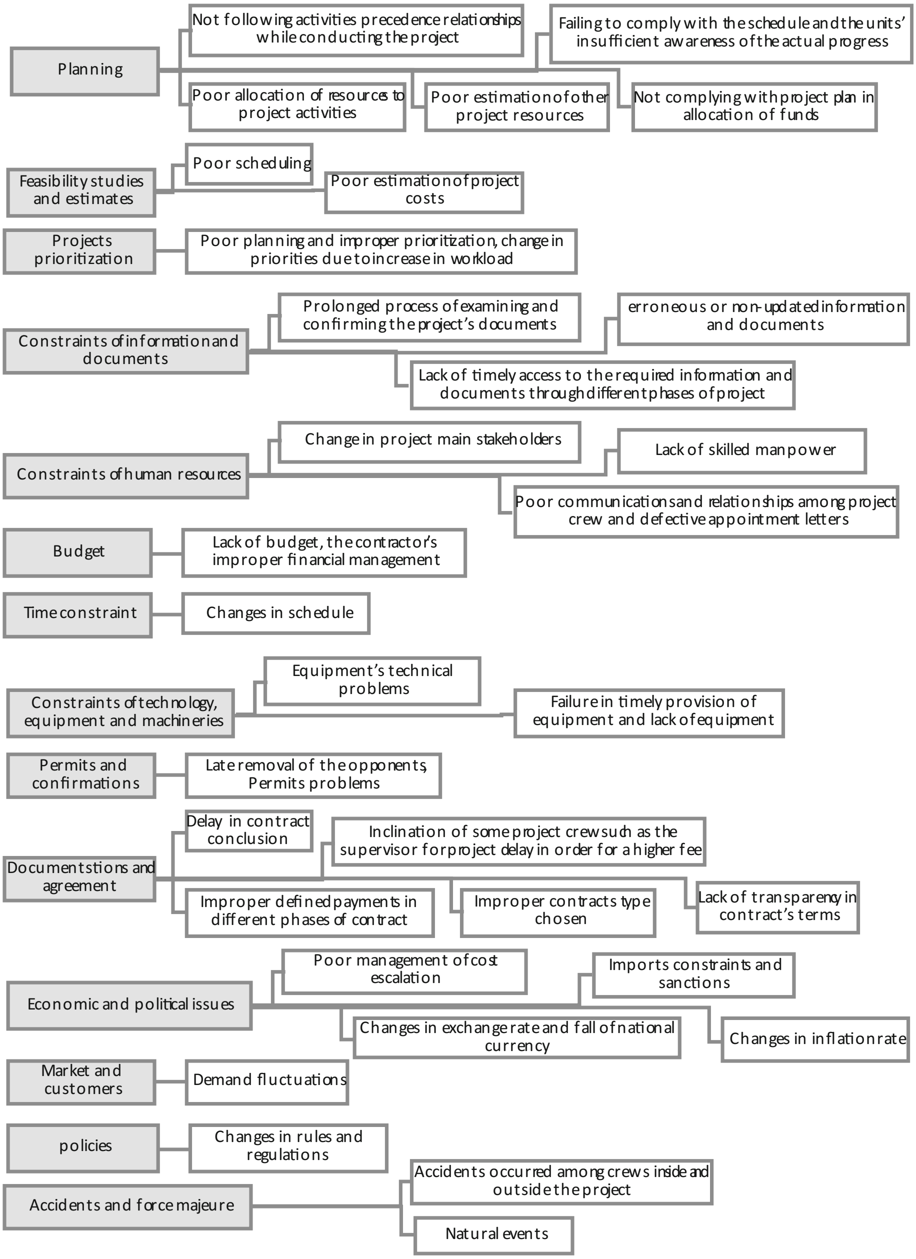


Fig. 2. SCCPP identified project risks

Table 1. SCCPP project's risk impact on project duration

Risk impact	Numerical values	Duration increase
Very low	1	Less than 10%
Low	2	10 to 20%
Moderate	3	20 to 35%
High	4	35 to 50%
Very high	5	More than 50%

It must be mentioned that the risks are mapped in different areas of the matrix according to the average value obtained from each criterion and the explanation of each area. The probability-impact matrix with regard to project duration for the SCCPP is shown in Fig. 3 Similar diagrams were also developed for the other two selected criteria.

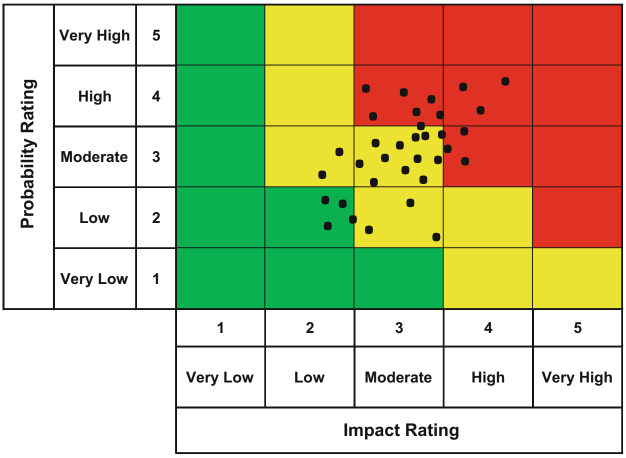


Fig. 3. SCCPP project probability-impact matrix (Color figure online)

As shown in Fig. 3, the red area represents the highest risks: the green area represents lowers, and the yellow area represents moderate risks. A risk with a high and very high probability of occurring and which will have a high impact on the project objectives will likely need further quantitative analysis and response plan. As also if a particular risk has a moderate probability with a higher estimated impact. A risk with a low and very low probability of occurring and which will have a low impact on the project objectives so no response plan will be developed. According to the probability of risks occurrence and their impact on SCCPP project duration, about 36 percent of the identified risks are in the area of high-priority, 48 percent in the area of medium-priority, and 12 percent are in the area of low-priority. This distribution shows that most of potential project risks might threat the project with completion delays. Pfeifer et al. created an optimization problem to maximize project delay due to certain stochastic task disruptions, and proposed a genetic algorithm to identify the critical tasks which may cause the maximum risk of delay in the project [25].

The probability and impact matrix examines the impact of each risk separately; however, a risk simultaneously affects the three criteria of project time, quality, and cost. Based on the specific conditions of a given project, one criterion of assessing a project's success might be considered relatively important. In these cases, the assessment and analysis of the risks impact on the project success criteria must be conducted simultaneously. Weighting is one of the common ways of performing multiple criteria analyses and determining the relative significance of variables in decision making. Normalized weighting was applied in this research to determine the relative significance of risk assessment criteria. Table 2 and Fig. 4 shows the relative weighting of the three criteria, determined according to interviews and questionnaires carried out by different project managers and experts as mentioned in methodology section.

Table 2. Relative significance of risk assessment criteria in SCCPP project

Impact of risk on project Quality	Impact of risk on project Cost	Impact of risk on project Duration
30%	33%	37%

To determine the risks' priority and final ranking, the average points of each criterion was multiplied by its weight, and the products were added together. The risks were divided into three groups of medium priority, high priority and low priority based on their final rankings. Also the obtained results show that the problems posed by the stakeholders match the rankings to a great extent. In SCCPP project, Lack of skilled manpower, Improper contracts type chosen and Imports constraints and sanctions are the high-priority risks according to Fig. 5. Therefore, a smarter planning is deemed necessary to resolve these issues.

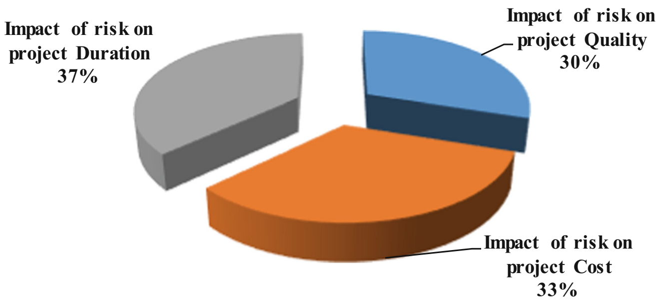


Fig. 4. The relative weighting of the three criteria

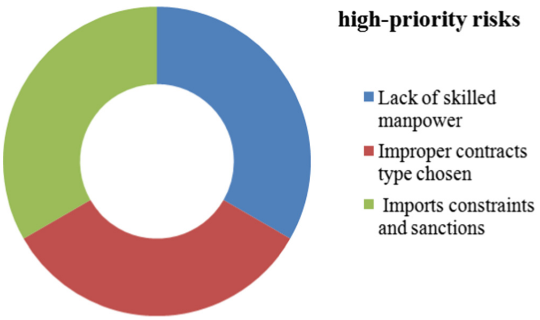


Fig. 5. The high-priority risks in the project

7 Main Conclusions

Risk management is a process through which the threats to an organization's earnings and capital are identified, assessed and controlled. The main purpose in risk management is to understand, analyze, and address risk in order to make sure that organizations and firms achieve their objectives. The goal of this study was to identify and prioritize the risks and explore the role of risks and risk management practices in construction of SCCPP project in terms of their effect on the project's time, quality, and cost. As a part of the project risk management process, different types of risks were identified after the planning phase of risk management. The results of these examinations led to making the projects' risk breakdown structure in groups of planning, resource constraint, legal and contractual and others. Each of these main groups were classified into subgroups. The identified risks were put in their respective groups. The results indicated that in SCCPP project, significant risks such as Lack of skilled manpower, Improper contracts type chosen and Imports constraints and sanctions were among the high-priority risks which. Considering the results of this study project managers and stakeholders should take appropriate measures to reduce the undesirable risks' impacts by this risks developing proper mitigation risk strategy plan development in order to achieve project higher resilience on SCCPP projects. Eventually, according to the research question, SCCPP project complemented and strengthened the performance impact of classical risk management practices in risky environments. For future research, the scope of the study can be expanded to examine the risk management of similar projects in different countries. It is also suggested that the handling of risk management in other types of projects be considered seriously by academics and project practitioners as an area in need of further research.

Acknowledgements. The authors are grateful to all of those with whom they had the pleasure to work during this and other related projects. This work has been partially supported by UNIDEMI - Research and Development Unit for Mechanical and Industrial Engineering by the Portuguese FCT-PEST program UID/EMS/00667/2019.

References

1. Ashtiyani, Z., (translator): A Guide to the Project Management Body of Knowledge, 4th edn. Adineh Pub. Co., Iran (2008)
2. Teller, J., Kock, A., Gemünden, H.G.: Risk management in project portfolios is more than managing project risks: a contingency perspective on risk management. Proj. Manag. J. **45**(4), 67–80 (2014)
3. Domnikov, A., Khomenko, P., Chebotareva, G.: A risk-oriented approach to capital management at a power generation company in Russia. WIT Trans. Ecol. Environ. **1**, 13–24 (2014)
4. Domnikov, A., Chebotareva, G., Khodorovsky, M.: Development of risk management for power generating companies in developing countries. WIT Trans. Ecol. Environ. **193**, 859–870 (2015)
5. Nie, S., Li, Y.P., Liu, J., Huang, C.Z.: Risk management of energy system for identifying optimal power mix with financial-cost minimization and environmental-impact mitigation under uncertainty. Energy Econ. **61**, 313–329 (2017)
6. Lee, C.W., Zhong, J.: Financing and risk management of renewable energy projects with a hybrid bond. Renew. Energy **75**, 779–787 (2015)
7. Serpella, A.F., Ferrada, X., Howard, R., Rubio, L.: Risk management in construction projects: a knowledge-based approach. Procedia Soc. Behav. Sci. **119**, 653–662 (2014)
8. Gatzert, N., Kosub, T.: Risks and risk management of renewable energy projects: the case of onshore and offshore wind parks. Renew. Sustain. Energy Rev. **60**, 982–998 (2016)
9. Hwang, B.G., Zhao, X., Toh, L.P.: Risk management in small construction projects in Singapore: status, barriers and impact. Int. J. Proj. Manag. **32**(1), 116–124 (2014)
10. Desai, A., Kashiyani, B.: Role of insurance as a risk management tool in construction projects. Int. J. Adv. Res. Eng. Sci. Manag. **2**(3/3), 247–256 (2015)
11. Cheraghi, E., Khalilzadeh, M., Shojaei, S., Zohrehvandi, S.: A mathematical model to select the risk response strategies of the construction projects: case study of Saba tower. Procedia Comput. Sci. **121**, 609–616 (2017)
12. Liu, X., Zeng, M.: Renewable energy investment risk evaluation model based on system dynamics. Renew. Sustain. Energy Rev. **73**, 782–788 (2017)
13. Qazi, A., Quigley, J., Dickson, A., Kirytopoulos, K.: Project Complexity and Risk Management (ProCRiM): towards modelling project complexity driven risk paths in construction projects. Int. J. Proj. Manag. **34**(7), 1183–1198 (2016)
14. Carvalho, M.M.D., Rabechini Junior, R.: Impact of risk management on project performance: the importance of soft skills. Int. J. Prod. Res. **53**(2), 321–340 (2015)
15. Violante, A., Dominguez, C., Paiva, A.: Risk management in construction projects: are small companies prepared? MOJ Civ. Eng **4**(1), 00090 (2018). https://doi.org/10.15406/mojce.2018.04.00090. Risk Management in Construction Projects: Are Small Companies Prepared, 2(7)
16. Titarenko, B.P., Lubkin, S.M., Eroshkin, S.Y., Kameneva, N.A.: Robust and traditional methods of risk management in investment and construction projects. In: 2017 Tenth International Conference Management of Large-Scale System Development (MLSD), pp. 1–4. IEEE, October 2017
17. Reddy, A.S.: Risk management in construction industry-a case study. Int. J. Innov. Res. Sci. Eng. Technol. **4**(10), 10058–10067 (2015)
18. Zhao, X., Hwang, B.G., Pheng Low, S., Wu, P.: Reducing hindrances to enterprise risk management implementation in construction firms. J. Constr. Eng. Manag. **141**(3), 04014083 (2014)

19. Kang, B.G., Fazlie, M.A., Goh, B.H., Song, M.K., Zhang, C.: Current practice of risk management in the malaysia construction industry-the process and tools/techniques. Int. J. Struct. Civ. Eng. Res. **4**, 371–377 (2015)
20. Mahjoub, M., Atashsokhan, S., Khalilzadeh, M., Aghajanloo, A., Zohrehvandi, S.: Linking "Project Success" and "Strategic Talent Management": satisfaction/motivation and organizational commitment as mediators. Procedia Comput. Sci. **138**, 764–774 (2018)
21. Hosseini, S.A., Akbarpour, A., Ahmadi, H., Aminnejad, B.: Balance of cost, time, and quality related to construction projects regarding the reinforced concrete of underground structures using a meta-heuristic algorithm. Arch. Civ. Eng. **63**(4), 103–121 (2017)
22. Zohrehvandi, S., Khalilzadeh, M., Hajizadeh, M., Cheraghi, E.: Planning project closure phase in combined cycle power plant projects. Procedia Comput. Sci. **121**, 274–281 (2017)
23. Stosic, B., Mihic, M., Milutinovic, R., Isljamovic, S.: Risk identification in product innovation projects: new perspectives and lessons learned. Technol. Anal. Strat. Manag. **29**(2), 133–148 (2017)
24. Kırılmaz, O., Erol, S.: A proactive approach to supply chain risk management: shifting orders among suppliers to mitigate the supply side risks. J. Purch. Supply Manag. **23**(1), 54–65 (2017)
25. Pfeifer, J., Barker, K., Ramirez-Marquez, J.E., Morshedlou, N.: Quantifying the risk of project delays with a genetic algorithm. Int. J. Prod. Econ. **170**, 34–44 (2015)

voteChain: Community Based Scalable Internet Voting Framework

Ricardo L. Almeida[1(✉)], Laura Ricci[2], and Luis M. Camarinha-Matos[1]

[1] Faculty of Sciences and Technology, UNINOVA-CTS, Nova University of Lisbon, Campus de Caparica, 2829-516 Monte da Caparica, Portugal
rdl.almeida@campus.fct.unl.pt, cam@uninova.pt

[2] Department of Computer Science, University of Pisa, Pisa, Italy
laura.ricci@unipi.it

Abstract. Most democratic countries still use the traditional systems of paper ballots and voting boxes. As technology develops, new electronic voting systems have been proposed to modernize and facilitate the voting process. Most e-voting systems are based on centralizing models, i.e. client-server structures, which have been proved to be unreliable and prone to be affected by the same problems of any centralized computer system: Denial of Service attacks, server hacking, etc. The advent of cryptocurrencies in recent years has shed light on their underlying technology – blockchain – as a powerful decentralizing technological paradigm that keeps finding new areas of application outside this implementation, such as electronic voting. In this paper, we present a proposal for a voting framework based on blockchain technology and analyze its potential to improve current voting systems as well as the implementation drawbacks.

Keywords: Electronic voting · Cryptography · Blockchain · Smart contracts · Ethereum · Communities · Collaborative networks

1 Introduction

Most countries in the world still rely on analogue methods to perform elections. Save for about 50 countries (as of 2018) [1], most of the world still uses paper ballots filled out with pen and cast into a ballot box which is nothing more than a closed box with a slit on top - hardly a secure and tamper proof system. The integrity of such systems depends exclusively of the people that run the voting process since the system itself has no built-in security features. In most cases, the privacy and anonymity of the voter are only ensured by the cloth curtain on the voting booth.

As an example, the current voting system in Portugal is significantly outdated. In 2017, during elections for local councils, the rules and security measures were the same ones that were developed a few months after the 1974 revolution [2]. There has not been any improvements or updates for over 40 years. Elections and referendums always take place on a Sunday. This model was well adapted to the Portuguese society in the 70's: many had the weekend off from work and people rarely had to move away

Published by Springer Nature Switzerland AG 2019
L. M. Camarinha-Matos et al. (Eds.): DoCEIS 2019, IFIP AICT 553, pp. 70–80, 2019.
https://doi.org/10.1007/978-3-030-17771-3_6

from their registered voting parish councils. As such, it was safe to assume that most of the voting population could be expected to be minutes away from a polling station on the election day. This is a common voting system found in many other democratic countries.

It only took less than half a century for this scenario to change substantially. With increased mobility, greater international exposure and flexible working schedules, people are still close enough to a ballot in an election Sunday, just not the one that they can use to vote. In Portugal a citizen needs to register with a parish council before an election to be able to cast his or her vote there. This registration however locks that parish, as a physical location, as the only place in which the voter can cast his or her choice. A voter can always change his voting parish but that needs to be done pre-emptively and the process is not immediate. Rigid rules in such flexible society create unexpected difficulties in the democratic act. There are other factors to consider in this matter, but voter turnout has been steadily decreasing from 92.5% in 1975 to 55.9% in 2015 [3] in this country.

These problems can easily affect other nations that retain similar voting systems. As voters became increasingly mobile due to professional and/or social changes, so does the probability of missing the next election [4]. It is very easy to miss out an election nowadays. All it takes is an unscheduled work trip, a sick co-worker that needs his weekend shift covered or a transport malfunction. Scheduling elections on Sundays may have helped in the past but in current 7-working days, shift-oriented society, it may be causing more harm than good. For example, public transportation may be harder to use in a Sunday to travel to the voting place, and if it is not, that means that other people need to work on that day to assure it, risking their own voting availability in the process.

Voting today requires time and money from the voters themselves, in most cases. This investment is enough to prevent some users, especially those in less than stable professional and economic situations (relocated students and workers, shift workers, emergency caretakers, etc.) from exerting their constitutional right and implicitly skewing the election's results by removing parts of the population from the election process.

As an alternative, it is worth exploring the use of information and communication technologies (ICT) to support more flexible voting systems. Blockchain is an emergent technology with high disruption potential by its distributed architecture that introduces novel ways to secure and transmit data over unsecure channels. Considering this, we considered the following question: what can be a suitable way of using this blockchain technology in the development of an electronic voting framework such that it is able to increase voter turnout in elections?

In this context, this work aims at exploring the use of blockchain as an effective approach for flexible and secure voting – the voteChain framework. This research will be developed over the hypothesis that, if the native cryptographic and data hashing capabilities of blockchain technology are used to ensure vote security, transparency and anonymity in and electronic voting framework, while also addressing voter equality and mobility by providing a tool that can be used to cast votes remotely over unsecure communication channels, then this framework can be used in elections to increase voter turnout.

The remainder of this paper includes the relation of this subject to the conference theme in Sect. 2 and a literature review on the main concepts approached follows in Sect. 3. The technical details and description of the voteChain concept is presented in Sect. 4. Section 5 presents our research plan main points and the paper finishes with the concluding remarks in Sect. 6.

2 Relationship to Innovation in Service Systems

Communities are often facing decisions that affect a significant number of their members. From choosing a different electricity provider in an apartment building, deciding on local council's excess budget expenses or to elect public officials, the democratic voting process has been the preferred choice to find a solution that suits the majority of the community's members. This has created space for third party solutions that provide voting as a service. From free platforms such as Doodle [5] to more commercial ones, such as Polyas [6] or eBallot [7], these solutions provide centralized server-client solutions but in which they run proprietary platforms in which voter transparency and anonymity are ensured through opaque proprietary protocols.

Our solution offers an innovative approach to the concept of voting as a service by providing an open framework based on a decentralized and transparent approach that can be scaled according to the community's needs.

3 Literature Review

3.1 Electronic Voting Systems

As the Internet matured into a reliable social interactive platform, the traditional voting system was put up for an upgrade. Several countries took upon that effort, although in an individual effort, with each country developing its own system with its own problems and advantages. More than twenty years and several election cycles after and we are still waiting for an electronic voting system that can provide the same level of functionality as the traditional ones, plus the speed, mobility and reliability that characterize current online applications.

Most systems saw only an update in the method in which the vote was cast by using voting machines instead of the traditional paper ballot. Countries such as Brazil, Germany, India, The Netherlands and USA only upgraded the vote counting process - But Switzerland, Estonia, Norway and Canada went as far as implementing truly online voting methods with various degrees of success [8].

Electronic Voting Machines allow faster vote counts and recounts while keeping the voting process under a controlled environment that can be overseen by election officials if needed. Yet they don't address voter mobility since voters are still required to physically travel to the voting station in which they were registered.

True Internet Voting or Remote Voting systems allow the voter to cast his or her choice over unsecure channels and in uncontrolled environments such as through a mobile application or a web portal. The voter must previously register to vote and needs to authenticate himself before casting the vote. Current approaches use centralized server-client models to build their voting platforms. Some notable examples of this approach were implemented by the Estonian [9] and Swiss [10] governments. This approach gives support to mobility and ease of access to voters, specially physically impaired ones, since mobile applications, personal computers systems and access to Internet are increasingly available, even in underdeveloped countries. But this advantage came with a cost in system complexity and security since it needs to ensure secure transmission of data at every step of the process (authentication and vote submission) while becoming vulnerable to the attacks that centralized platforms are prone for, namely Denial of Service, server hacking, Man-in-the-Middle attacks, etc. True online voting is still not available to most voters and, when it is, it normally requires several pre-conditions to be met. The most recent attempts had to deal with significant software issues or were over complicated [1, 11, 12]. Electronic voting systems based on a typical centralized server-client model are notoriously weak regarding voter's privacy, anonymity, ballot irrevocability and process transparency [13, 14].

3.2 Blockchain Based Electronic Voting

Blockchain is a data structure in which a sequence of data blocks is connected using cryptography. Each block contains a timestamp, transaction data and a cryptographic hash of the complete previous block. This method assures data integrity for the whole chain since it is practically impossible to add a falsified block since this block would have a different hash and therefore wouldn't match the cryptographic hash already present in the next block of the chain. The robustness of this system derives from its distributed implementation. The data belonging to the blockchain, in the form of transactional data inserted into blocks cryptographically chained together, is distributed through all the machines or nodes that compose the blockchain network [15]. This distributed database is monitored by the nodes that support it and changes made to it, namely the addition of new blocks, needs to be agreed among the majority of nodes in the network [16].

There are two possible scenarios in which, theoretically, it is possible to introduce a false block in the chain: (1) one can replace all blocks from the false one to the beginning of the chain, thus ensuring that the cryptographic hashes are all correct. But this approach is unrealistic due to the high rate of blocks being added to the blockchain. Currently, the Bitcoin blockchain adds a new block every 10 min approximately while newer blockchains, such as Ethereum for example, add a new block to its chain every 15 s on average. Whenever a new block, fake or otherwise, is to be added to the chain, a computer intensive cryptographic puzzle needs to be solved first as a prerequisite to produce the cryptographic hash. To be able to add false blocks to a rate fast enough to validate this false chain, the falsifier would need such amount of computing power and energy that would simply rend the process economically unfeasible. (2) A fake block that produces the same cryptographic hash as the block that intends to replace can be produced. Hash collisions, i.e., when two or more pieces of distinct data produce the same hash string when run through the hash function, are theoretically possible although highly improbable. Currently, the only way to achieve that is through "brute force", i.e., trying all possible combinations until a match is found. As an example, for a hash function with a 256-bit output, in the worst case it would have to be computed $2^{256} + 1$ times. In a computer that calculates 10,000 hashes per second it would take 10^{27} years to do just that. That's more than an octillion years [17]. And if, by luck alone, it happened that a match was found in the first few computations, chances are that the data that produces that hash has no use whatsoever since the hashing process is oblivious to the structure of it (it would be a random string of bits).

The probability of any of the above happening is extremely small but it is not zero and that is why, for the sake of argument, blockchain cannot be considered completely tamper proof. Nevertheless, since its inception, blockchain has drawn attention to its potential and it is currently a hotspot of research [18].

Regarding online voting, the inherent properties of blockchain brought a fresh perspective to an old problem. By decentralizing the whole system and relying on asymmetric cryptography to secure information transfers, blockchain can guarantee the security and transparency that has been creating problems to its centralized counterparts. There are already several solutions for online voting using blockchain [13, 14, 18–20]. The reason behind this surge in blockchain applied to e-voting was thoroughly summarized in [1], which enumerates the following issues in any online voting system: (i) Identification of voters, (ii) Voting details storage, (iii) Ballot counting, (iv) Voter's anonymity, and (v) Encryption key management. Implementations of this approach so far are incomplete [13], possess major limitations [14] or are still early in their conceptual phase [19].

3.3 Blockchain Types and Choice of Platform

For the current stage of the voteChain project we choose the Ethereum blockchain platform. The reason behind this choice is its capability of running Smart Contracts, which are applications whose code resides in the Ethereum blockchain itself. Smart Contracts are run through an Ethereum Virtual Machine (EVM), a distributed computing platform provided to any member of the network to run their Smart Contracts. We intend to use these Smart Contracts to implement the functionalities of our voting platform, such as voter validation, casting, count, etc.

Another key concept to this project proposal is the Distributed Application (DApp). The DApp is a natural extension of the Smart Contract and the long term plan for the Ethereum protocol is to create a new paradigm in regard to these DApps, namely to publish them in a "App Store" like environment [21], without the need to download anything locally and with virtually complete availability due to a decentralized storage of Smart Contract code and its respective front-end code, for a more user-friendly usage.

4 The voteChain Proposal

4.1 Implementation

Most solutions analyzed either avoid using the blockchain all together or go for an extreme approach in which the blockchain is used in every step, removing any centralizing elements from the equation. This project plans to position itself in the middle of the spectrum. We understand that any extreme approach in this sense is bound to remove important elements essential for a viable solution. The electronic voting framework proposed is structured into the following logic steps:

Voter authentication. One advantage brought forward by the plethora of cryptocurrencies developed in the last decade was to enable an online behavior that people were already able to do with hard currency: anonymous currency exchanges between parties. Transaction anonymity is one of the strong points behind cryptocurrencies, but it is undesirable in a voting process.

All democracies require their voters to be properly identified and that implies the existence of a centralizing authority that can validate a user for voting through verification of official records. This centralizing step is essential and unavoidable and as so, it is going to be included explicitly in the framework. Once a user can authenticate himself before the Central Authority, he receives a voting token from it. The usage and transfer of valueless tokens in the voting process is one of the novel approaches taken by this project. This token is a key element on recording information about the process in the blockchain (Fig. 1).

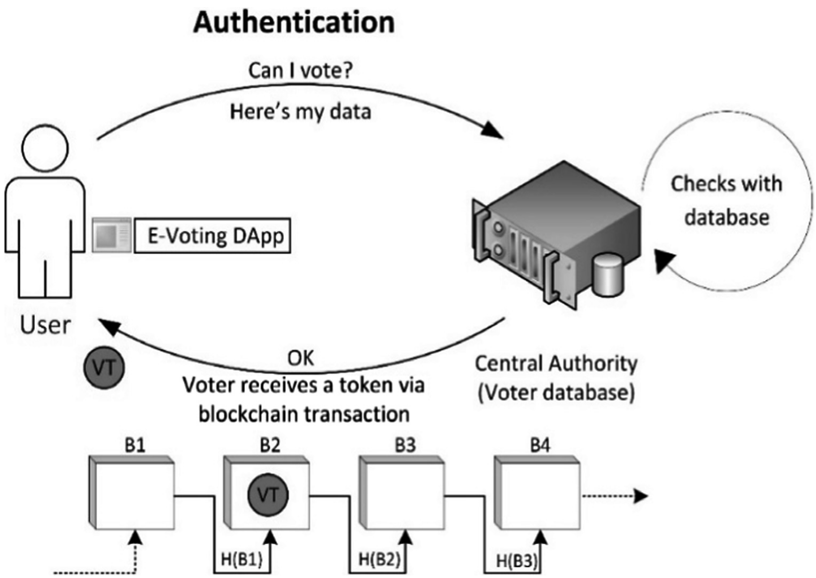


Fig. 1. Voter authentication through a trusted central authority

The transaction form allows extra data to be added to the transaction's own. It is this custom added data capability behind blockchain transactions that makes it flexible for other applications like the voteChain. The blockchain uses asymmetrical encryption to protect the contents of the transaction from unauthorized access. All blockchain transactions, and the data contained in them, are encrypted in blockchains' public interface.

Vote casting. This step introduces the Regional Voting Ballot (RVB) element. Instead of polling all votes into a single storage instance, thus creating a single point of failure in the process, the virtual ballots in which the voting tokens can be redeemed are distributed entities but with limited access. Their main objective is to hold the votes themselves, look for double vote attempts and store votes in an encrypted format.

Continuing with the token exchange paradigm, the voter sends it voting token to the RVB as a proof of vote. The voter exchanges his public encryption keys with the RVB during the validation stage, and then uses the RVB key to encrypt his identification and vote data, thus restricting the access to this information to the RVB alone and securing it for transfer over an unsecure connection. The RVB receives the two encrypted data bundles and decrypts only the voter identification. The vote data remains encrypted to maintain vote secrecy. The encrypted vote is then stored internally in the RVB as a value in a hash table whose key is determined by hashing the voter identification data. This way voter anonymity is preserved to an extent since the only connection between the vote, which is still encrypted, and the voter is a hashed version of his ID data.

Double voting becomes easier to detect when using this method. If a voter somehow submits another vote under the same ID data, the RVB can detect that the provided voter ID hash already as a value under it in the internal hash table. RVB operations are based on Smart Contracts, adding transparency to the process since the code for these is publicly accessible in the blockchain and any user can verify the integrity of these contracts regarding the voting process, namely, to make sure that there are no security flaws that can reveal the identity of the voter or his choice. If all these operations are successful, the RVB returns a receipt token to the voter. This token serves two purposes: it gives the voter an assurance that his or her vote was successfully cast, and it is going to be counted and publishes another transaction on the blockchain in which we use the custom data fields to state that "voter X with information Y has voted successfully". This data is previously encrypted with the voter's public encryption key, which means that only the voter can decrypt and see it using his private key (Fig. 2).

Vote counting. At the counting stage votes are finally decrypted from the RVBs internal hash tables in which they were stored. They do not contain or connect to any information that could point to the original voters - and thus achieving voter anonymity without sacrificing proof of vote - and are currently stored under multiple RVBs. This adds a layer of security by decentralizing the process further (Fig. 3).

The final tally is determined with the Central Authority querying each RVB for its partial count and then add it all together. Continuing with the same logic, the partial and final counts are then published unencrypted in the blockchain for public consultation.

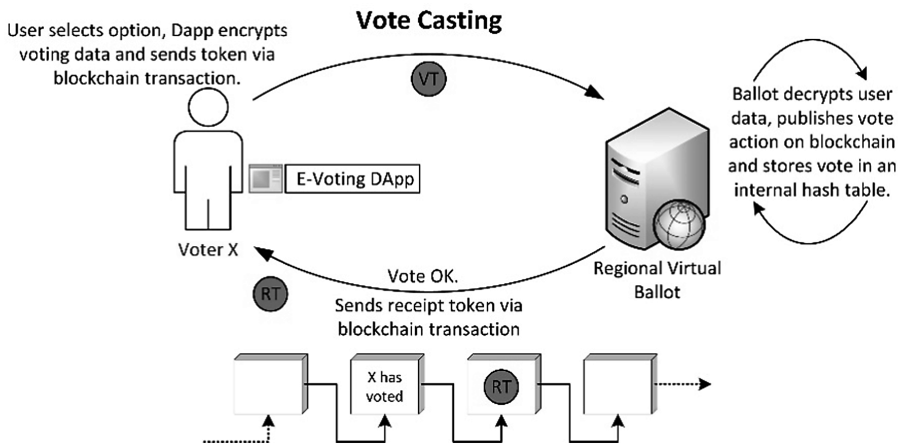


Fig. 2. The voter casts his vote to his assigned RVB and gets a receipt token if the operation is successful.

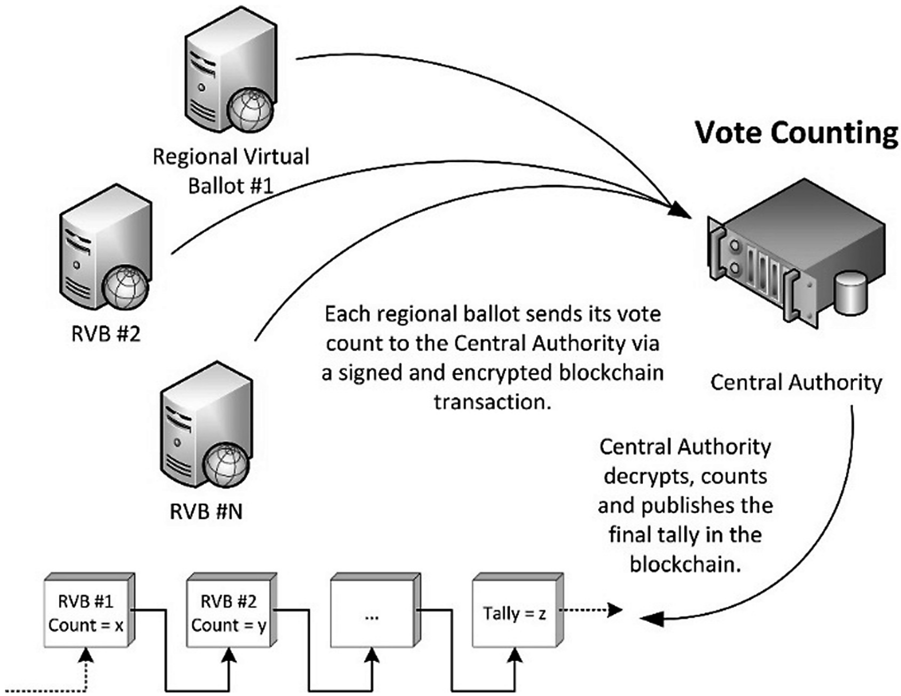


Fig. 3. Final tally determined by counting all the votes cast into the RVBs.

5 Future Research

5.1 Multiple Vote Casting

Using electronic methods to collect votes simplifies and speeds up the voting process greatly. This allows for novel approaches to some know problems that occur with traditional voting methods, namely vote buying and cohesion.

These tactics are only effective due to the inherent rigidity of traditional methods, which allow for only one vote per person. A strategy to counter these problems with electronic voting systems consists in allowing people to cast multiple votes in which only the last one submitted gets counted. It also allows voter to correct erroneous voting. It may be rare occurrence, but it is currently impossible to correct in a traditional voting system. The Estonian government employed this feature in their 2005 e-voting study [22].

Implementing this feature in the voteChain framework is technically possible but that needs to be studied, specifically regarding its impact with maintaining voter anonymity.

5.2 Possible Attacks to the System

Voter authentication, a key step in our framework, is done through interaction with a centralizing actor, therefore we need to protect it against the typical cyber-attacks that affect this type of elements, e.g. Denial of Service, Server hacking, SQL code injections, Man-in-the-middle attacks, etc., as for those that are specific to democratic processes, e.g. double voting and/or Sybil attacks, vote buying, etc. that potentially affect the system as a whole.

5.3 Smart Contracts Need Ether to Run

A voting system should be free to use. Even though nowadays people may spend some money indirectly when they decide to vote, it is not realistic to set up a fee upfront to provide a constitutional right.

The Ethereum protocol rules imply that some money, in the form of its cryptocurrency, Ether or Wei (1 Ether = 1×10^{18} Wei), needs to be spent to execute a smart contract on the EVM. This was implemented to prevent malicious smart contracts to be run infinitely on the EVM. Having to spend some value of the cryptocurrency to run code on the EVM, it is expected that honest developers keep their Smart Contracts optimized and simultaneously prevent Denial of Service attacks from dishonest ones by essentially trying to bankrupt the attackers in the process. But that also implies that some money, regardless of how little, needs to be spent whenever a vote is cast.

On the other hand, traditional voting systems are inherently expensive, which means that some money is always needed to be used to finance the process regardless. Even if voting using this platform does need some financing to run the associated Smart Contracts, the values involved are substantially lower than even the smallest of traditional elections. Nevertheless, this is an issue requiring further analysis.

5.4 Communities Context

Although previous sections described the voting process at a country level, similar processes are needed at the level of smaller communities such as collaborative networks/business ecosystems [23, 24]. Traditional approaches in this context usually rely on trusting a network manager that centralizes the voting process. Even when electronic institutions such as e-notary [25] have been introduced, the issue of trusting the network manager remains.

Being these communities relatively small, they also provide a good context to experiment and validate the proposed approach and mechanisms.

6 Conclusion

Existing electronic voting projects either avoid using the blockchain at all or go for the opposite approach, using the blockchain on every single step of the process. The blockchain was created to be used as a ledger in financial transactions and when this concept is forced into a different cast, the results are often poor.

The proposed voteChain framework uses a mixed approach in order to find the optimal usage of the blockchain for this context. The voting system cannot be adapted to a financial application in verbatim because the systems are different. The proposal presented in this document introduces an ongoing multi-year research project based on a transformative but recent technology that has the potential to significantly change the existing paradigm regarding Remote Internet Electronic Voting.

Acknowledgments. This work has been funded in part by the Center of Technology and Systems (CTS – Uninova) and the Portuguese FCT-PEST program UID/EEA/00066/2013.

References

1. Kim, H.R., Min, K., Hong, S.: A study on ways to apply the blockchain-based online voting system. Int. J. Control Autom. **10**(12), 121–130 (2017)
2. de Eleições, C.N.: Current Portuguese Election System - History. http://www.cne.pt/content/historia
3. PORDATA, INE, and DGS/MS: Taxa de abstenção em eleições para a Asssembleia da Républica, Portugal 1975–2015. PorData (2015). https://www.pordatahttps//www.pordata.pt/DB/Municipios/Ambiente+de+Consulta/Gráfico
4. Krimmer, R., Volkamer, M.: Challenges Posed by Distant voting in General - Posting voting and in Particular, e-Voting, no. July, Vienna & Passau, p. 11 (2007)
5. Doodle. https://doodle.com/
6. POLYAS - Secure Online Voting. https://www.polyas.com/
7. I. Votenet Solutions: eBallot. https://www.eballot.com/
8. T. E. K. Network: Countries with e-voting projects. Ace Project. http://aceproject.org/ace-en/focus/e-voting/countries
9. Madise, Ü., Martens, T.: E-voting in Estonia 2005. The first practice of country-wide binding Internet voting in the world. In: Electronic Voting 2006, pp. 15–26 (2006)

10. Braun, N., Brändli, D.: Swiss E-Voting pilot projects: evaluation, situation analysis and how to proceed. In: Electronic Voting 2006, pp. 27–36 (2006)
11. Phillips, A.: Utah Republicans are holding a first-ever online presidential primary. And it's not going so well. Washington Post (2016)
12. Kuo, L.: Electronic voting is failing the developing world while th US and Europe abandon it. Quartz (2013)
13. Bartolucci, S., Bernat, P., Joseph, D.: SHARVOT: secret SHARe-based VOTing on the blockchain. In: WETSEB 2018 IEEE/ACM 1st International Workshop on Emerging Trends in Software Engineering for Blockchain, pp. 1–5, March 2018
14. Faour, N.: Transparent Voting Platform Based on Permissioned Blockchain. Higher School of Economics (National Research University), Russian Federation (2017)
15. Swan, M.: Blockchain Blueprint for a New Economy. O'Reilly, Sebastopol (2015)
16. Nakamoto, S.: Bitcoin: A Peer-to-Peer Electronic Cash System (2009). https://bitcoin.org/
17. Narayanan, A., Bonneau, J., Felten, E., Miller, A., Goldfeder, S.: Bitcoin and Cryptocurrency Technologies. Princeton University Press, Princeton (2016)
18. Wang, B., Sun, J., He, Y., Pang, D., Lu, N.: Large-scale election based on blockchain. Proc. Comput. Sci. **129**, 234–237 (2018)
19. Hanifatunnisa, R., Rahardjo, B.: Blockchain based E-Voting recording system design. In: 2017 11th International Conference on Telecommunication Systems Services and Applications, October 2017
20. Shaheen, S.H., Yousaf, M., Jalil, M.: Temper proof data distribution for universal verifiability and accuracy in electoral process using blockchain. In: 2017 13th International Conference on Emerging Technologies, December 2017
21. Dannen, C.: Introducing Ethereum and Solidity. Apress, Brooklyn (2016)
22. Volkamer, M., Grimm, R.: Multiple cast in online voting: analysing chances. In: Electronic Voting 2006, pp. 97–106 (2006)
23. Camarinha-Matos, L.M., Afsarmanesh, H., Galeano, N., Molina, A.: Collaborative networked organizations - concepts and practice in manufacturing enterprises. Comput. Ind. Eng. **57**(1), 46–60 (2009)
24. Graça, P., Camarinha-Matos, L.M.: Performance indicators for collaborative business ecosystems — Literature review and trends. Technol. Forecast. Soc. Change **116**, 237–255 (2017)
25. Oliveiral, A.I., Camarinha-Matos, L.M., Pouly, M.: Agreement negotiation support in VO creation. In: Camarinha-Matos, L.M., Picard, W. (eds.) PRO-VE 2008. ITIFIP, vol. 283, pp. 107–118. Springer, Boston (2008). https://doi.org/10.1007/978-0-387-84837-2_11

Decision Systems

UAV Downwash-Based Terrain Classification Using Wiener-Khinchin and EMD Filters

João P. Matos-Carvalho(✉), André Mora, Raúl T. Rato, Ricardo Mendonça, and José M. Fonseca

Center of Technology and Systems - CTS/UNINOVA, FCT, NOVA University of Lisbon, Caparica, Portugal
{jp.carvalho, atm, rtr, rmm, jmf}@uninova.pt

Abstract. Knowing how to identify terrain types is especially important in the autonomous navigation, mapping, decision making and detect landings areas. A recent area is in cooperation and improvement of autonomous behavior between robots. For example, an unmanned aerial vehicle (UAV) is used to identify a possible landing area or used in cooperation with other robots to navigate in unknown terrains. This paper presents a computer vision algorithm capable of identifying the terrain type where the UAV is flying, using its rotors' downwash effect. The algorithm is a fusion between the frequency Wiener-Khinchin adapted and spatial Empirical Mode Decomposition (EMD) domains. In order to increase certainty in terrain identification, machine learning is also used. The system is validated using videos acquired onboard of a UAV with an RGB camera.

Keywords: Image processing · Wiener-Khinchin · EMD · IMF · Machine learning · Terrain classification · UAV

1 Introduction

Nowadays, Unmanned Aerial Vehicles (UAVs) have a huge impact in the area of research and industry. In the research area since it is a recent topic and therefore little studied. In the area of industry, being a lightweight and low-power robot, it can be used in various situations, such as precision agriculture, emergency landings and rescue missions.

In this case, the task of the UAV is to cooperate with the Unmanned Surface Vehicle (USV) so that it reaches its intended destination. This is the purpose of this paper.

With an RGB camera mounted on the UAV it is possible to receive the images from the outside from the effect of the downwash caused by the UAV. This effect is only felt at low altitudes and has unique effects on the terrain (e.g. the circular wave effect on the water terrain).

Two different techniques were merged to classify the terrain type: Wiener-Khinchin Theorem and Empirical Mode Decomposition (EMD). These techniques, although well developed for 1-D signals, are seldom applied in image processing domain. Thus, many of the following questions are still unknown: How to apply both algorithms for

Published by Springer Nature Switzerland AG 2019
L. M. Camarinha-Matos et al. (Eds.): DoCEIS 2019, IFIP AICT 553, pp. 83–90, 2019.
https://doi.org/10.1007/978-3-030-17771-3_7

image processing? Is it necessary to evaluate the whole image or may it be necessary to find methods to reduce computational complexity? Knowing that the EMD technique decomposes the signal into several intrinsic mode functions (IMFs), which of the IMFs best classifies the type of terrain? How can the Wiener-Khinchin theorem and the EMD algorithm complement each other? These challenges are addressed throughout this work and how they have been resolved to have the best results possible.

2 Contribution to Industrial and Service Systems

The behavior of different terrains when exposed to wind turbulence (downwash effect) is discernable and, by processing that behavior using Wiener-Khinchin (W-K) and EMD filters, we can generate singular results to identify a terrain type. Exploiting this knowledge, a novel terrain classifier is proposed and presented as a scalable service that can be used simultaneously by several applications and systems. Therefore, the herein presented work is focused on the Industrial and Service Systems domain, specifically, in Robotics and Integrated Manufacturing and Signal Processing topic.

3 Related Work

UAV applications is a topic that lately has arisen a lot and is becoming more important in several areas and is predicted that it will soon have a great impact on society [1]. One of the areas that is causing a great impact is in the area computer vision with the aim of making an UAV in an autonomous robot. For this, and concretely in this paper, the UAV must be able to capture, process and analyze the images of the environment in order to classify the type of terrain where it is flying over. Many researchers have proposed several methods in the classification of terrain, such as texture algorithms [2–5], color information [6] and other types of sensors that help in the classification decision, such as LIDAR type sensors [7–9]. Currently, on terrain classification using the downwash effect concept there are two recent works [10, 11]. The first one, used the concept of optical flow to detect the dynamic features of an image (Lucas Kanade algorithm [10]), from an RGB camera that was placed in an UAV. The objective of this work was to decide whether the terrain under study was water or not water. However, it takes at best 4 s to classify whether the terrain is water-type or not. Thus, it forces the UAV to hover in a steady position to classify the terrain. Another work that exploits the downwash effect to classify the type of terrain, takes advantage, not only of the dynamic part but also the static part of the image. In relation to [10], this work in addition to classifying water-type terrains can also differentiate vegetation and sand terrains in real time (less than 0.03 s). To classify the terrain type, the algorithm uses the Gabor concept [11] to remove the static texture of the terrain, complementing with the concept of optical flow (Farneback algorithm [11]) to remove the dynamic part of the terrain under study.

Color information was also used in [6] to classify four types of terrain. Converting each pixel of the image by the square root of the three channels of the pixel itself, the result will emphasize the color that most represents the type of terrain (e.g., blue for water). Other methods that rely on laser scanners also proved to differentiate between water and non-water terrain with good accuracy [7–9]. However, shallow water terrain will increase the decision error due to laser reflection and thus the sensor classifies as non-water-type terrain.

4 Proposed Model

Two algorithms were used in the proposed classifier: Wiener-Khinchin (W-K) and Empirical Mode Decomposition (EMD). Since these two techniques are independent of each other (one works in the spatial domain and the other in the frequency domain), it is possible to operate in parallel mode to increase the processing speed. From the proposed classifier model, as depicted in Fig. 1, it is possible to identify six fundamental stages:

1. *Rectified Image:* For the proposed classifier to be generic enough to work for all RGB cameras (depending on the resolution), it is first necessary to calibrate the image and mitigate its distortion;
2. *Dimensionality reduction:* To increase the classification speed only a small piece (i.e., region delimited by the red square) of information was extracted in the image (in this case the information was taken in a square format);
3. *Empirical Mode Decomposition (EMD) filter*: In the spatial domain, the EMD algorithm is applied in the image;
4. *Computation of the first Three IMFs*: Having the EMD obtained previously, the first three Intrinsic Mode Functions (IMFs) are taken to analyze the signal coming from the terrain under study;
5. *Wiener-Khinchin (W-K) filter*: Vertical and horizontal projections are applied to the thresholded image, extracting unique features that help differentiate the different types of terrains;

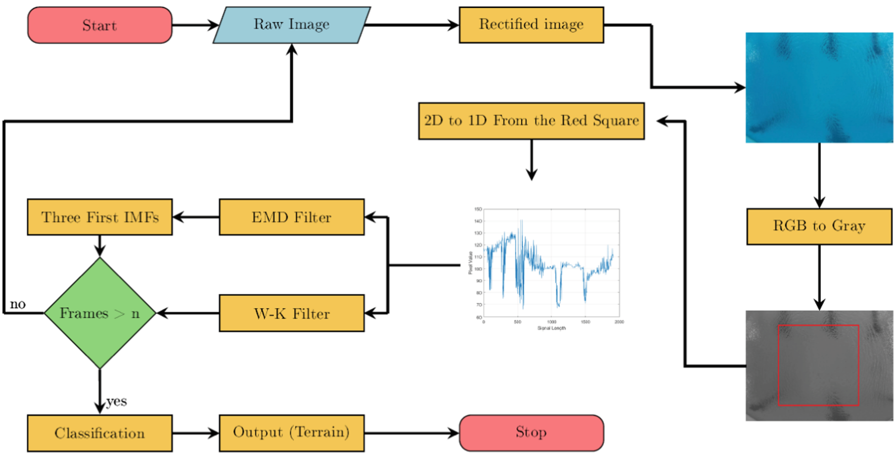


Fig. 1. Proposed terrain classifier model.

6. *Terrain Classification*: Certain decisions can be improved by using machine learning [12–14]. It will contribute to a more automated and reproducible process and possibly to increase the decision accuracy.

Five different terrain types representing the most common natural landscapes were selected to be the output of the proposed classifier: lake, pool, vegetation, dirt and sand terrains. When exposed to wind turbulence (downwash effect) distinct behaviors emerge according to the type of terrain. For instance, in water-like terrains, a circular wave movement outwards a center point is observed. In the vegetation, the movement occurs linearly from the inside to the outside. In sand and dirt, the movement is almost static.

5 Experimental Results

5.1 Empirical Mode Decomposition

According to Huang [15, 16], the principle of EMD is to decompose the signal into several frequencies, called Intrinsic Mode Functions (IMFs) and each IMF represents a zero-mean AM-FM function [17]. EMD is performed as follows:

1. Initialize vector **h**(t) and **k**(t) for a given image ***I***;
2. Get local maxima and minima from **h**(t) to build an upper $\mathbf{s}^{+}$(t) and a lower **s**(t) envelopes;
3. Calculate the mean envelope:

$$mean_{envelope} = 0.5 \times [s^{+}(t) + s^{-}(t)] \quad (1)$$

4. The mean is subtracted from signal **c**(t) (initially **c**(t) is equal to vector **h**(t)):

$$c(t) = c(t) - mean_{envelope} \quad (2)$$

5. c(t) is verified against a defined threshold. If it is above it, then h(t) = c(t) and repeat step 2–4. If it is below, then an IMF is found and h(t) = k(t) − c(t); k(t) = c(t); and repeat the steps 2–4 to find the next IMF.

Only the first three IMFs and the baseline are taken. Next, two heuristics are computed to identify the terrain: the number of local maximums given by an IMF and the number of times the IMF has crossed zero along the vector. These two heuristics were calculated for a set of 332 frames. From Fig. 2 it is possible to identify 5 clusters representing a terrain type. Despite the positive results, there are a few outliers. To improve accuracy, EMD was complemented with the Wiener-Khinchin filter.

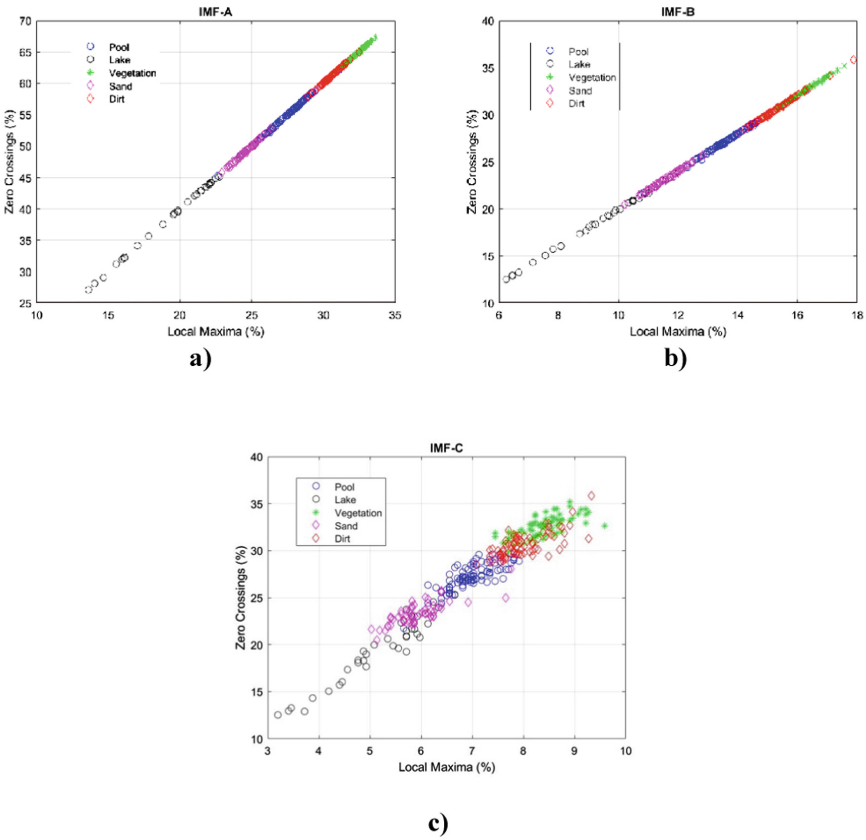


Fig. 2. Intrinsic Mode Function – The zero crossings in relation to the total number of local maxima. (a) IMF-A; (b) IMF-B; (c) IMF-C.

5.2 Wiener-Khinchin Filter

The Wiener-Khinchin filter (W-K) is performed in the frequency domain while the algorithm in Sect. 5.1 is presented in the spatial domain. To increase computational speed, the algorithm is not used in the total image but only in sub-region of it. To compare different types of terrain, it is necessary to compute the average value from that sub-region and normalize it apply general thresholds. As the Wiener-Khinchin theorem uses the concept of energy in the frequency domain, it becomes intuitive to normalize according to energy. These same values were then passed to the frequency domain. Next, frequency values are signal crossed with themselves. Finally, an inverse transform is performed. In last, it is possible to use the same heuristic characteristics that were used in Sect. 5.1. This algorithm was also applied to the same set of 332 frames (90 frames are pool water, 35 frames are lake, 74 frames are vegetation, 55 frames are sand and 78 frames are dirt). As depicted in Fig. 3, with the Wiener filter it is also possible to clearly identify the five terrain types. However, lakes present greater uncertainty. Due to this situation, to increase the probability of a correct classification, machine learning was also added to the proposed model.

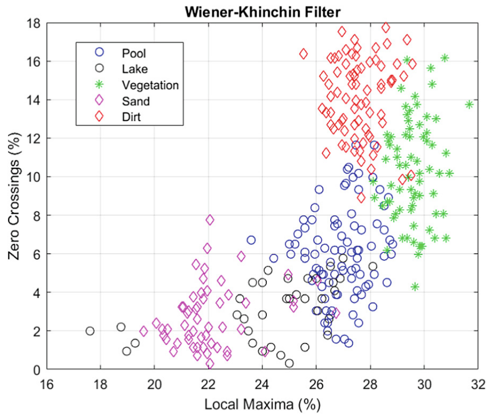


Fig. 3. Wiener-Khinchin Filter – Number of times the signal crosses the zero in relation to the total number of local maxima.

5.3 Classification

To automate the classification process, feed-forward neural network (NN) was used. Several tests were done to assess the best combination between the Wiener-Khinchin filter and Empirical Mode Decomposition. The first step was to create a neural network composed of two layers: the first layer is hidden and contains 10 neurons; the second layer contains 5 neurons that represents the terrain types.

A sigmoidal function was used as activation function and the final classification of the terrain under study is validated by the neuron that gives a greater result (the training dataset was composed by 334 samples, from which 70% were for training, 15% for testing and 15% for validation). Table 1 was produced showing the classification accuracy in all possible combinations. The best was the Wiener-Khinchin filter with the first IMF, yielding a 93% terrain classification accuracy.

Table 1. Combination of the two algorithms developed for the proposed classifier.

Combination	W-K	IMF-A	IMF-B	IMF-C	Result
1	x				90.0%
2		x			88.0%
3			x		74.0%
4				x	74.0%
5	x	x			93.0%
6	x		x		77.0%
7	x			x	82.0%
8		x			79.0%
9		x		x	85.0%
10			x	x	84.0%
11	x	x			82.0%
12	x	x		x	81.3%
13	x		x	x	80.0%
14		x	x	x	83.3%
15	x	x	x	x	84.0%

6 Conclusions

The goal of this work was to design a terrain classifier using two algorithms that exploit the UAV's downwash effect. The two algorithms are the Wiener-Khinchin filter and Empirical Mode Decomposition. It was proved throughout this work that both algorithms worked well together by presenting good results in the terrain classification. Regarding the EMD algorithm, it was emphasized that, despite the high accuracy in the NN decision, the probability of correctly classifying a terrain worsens from IMF to IMF and therefore it is always better to use the first IMF. However, it is always possible to improve certainty in the decision and therefore the EMD algorithm was complemented with the theory of Wiener-Khinchin. With this addition, the certainty in the classification of terrains managed to extend to 93%.

Acknowledgments. This work was partially funded by FCT Strategic Program UID/EEA/00066/203 of the Center of Technologies and System (CTS) of UNINOVA - Institute for the Development of new Technologies.

References

1. Bestaoui Sebbane, Y.: Intelligent Autonomy of UAVs: Advanced Missions and Future Use. CRC Press, Boca Raton (2018)
2. Linderhed, A.: Image Empirical Mode Decomposition: A New Tool For Image Processing. Adv. Adapt. Data Anal. **01**(02), 265–294 (2009)
3. Feng, Q., Liu, J., Gong, J.: UAV remote sensing for urban vegetation mapping using random forest and texture analysis. Remote Sens. **7**(1), 1074–1094 (2015)

4. Khan, Y.N., Komma, P., Bohlmann, K., Zell, A.: Grid-based visual terrain classification for outdoor robots using local features. In: IEEE SSCI 2011: CIVTS 2011 (2011)
5. Pietikäinen, M., Hadid, A., Zhao, G., Ahonen, T.: Computer Vision Using Local Binary Patterns. CIVI, vol. 40. Springer, London (2011). https://doi.org/10.1007/978-0-85729-748-8
6. Ebadi, F., Norouzi, M.: Road Terrain detection and Classification algorithm based on the Color Feature extraction. In: Artificial Intelligence and Robotics, pp. 139–146. IEEE (2017)
7. Yan, W.Y., Shaker, A., El-Ashmawy, N.: Urban land cover classification using airborne LiDAR data: a review. Remote Sens. Environ. **158**, 295–310 (2015)
8. Wallace, L., Lucieer, A., Malenovsky, Z., Turner, D., Vopěnka, P.: Assessment of forest structure using two UAV techniques: a comparison of airborne laser scanning and structure from motion (SfM) point clouds (2016)
9. GruszczynSki, W., Matwij, W., Ćwiąkała, P.: Comparison of low-altitude UAV photogrammetry with terrestrial laser scanning as data-source methods for terrain covered in low vegetation. ISPRS Photogramm. Remote Sens. **126**, 168–179 (2017)
10. Pombeiro, R., et al.: Water detection from downwash-induced optical flow for a multirotor UAV. In: OCEANS 2015, pp. 1–6. IEEE (2015)
11. Matos-Carvalho, J.P., Fonseca, J.M., Mora, A.D.: UAV downwash dynamic texture features for terrain classification on autonomous navigation. In: Proceedings of the 2018 Federated Conference on Computer Science and Information Systems. Annals of Computer Science and Information Systems, vol. 15, pp. 1079–1083. IEEE (2018)
12. Mora, A., et al.: Land cover classification from multispectral data using computational intelligence tools: a comparative study. Information **8**, 147 (2017)
13. Heung, B., Ho, H.C., Zhang, J., Knudby, A., Bulmer, C.E., Schmidt, M.G.: An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping. Geoderma **265**, 62–77 (2016)
14. Giusti, A., et al.: A machine learning approach to visual perception of forest trails for mobile robots. IEEE Robot. Autom. Lett. **1**(2), 661–667 (2016)
15. Huang, N.E., et al.: The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proc. Roy. Soc. Lond. A **454**, 903–995 (1998)
16. Oonincx, P.J.: Empirical mode decomposition: a new tool for S-wave detection. In: CWI Reports of Probability, Networks and Algorithms (PNA) (2002). PNA-R0203, ISSN 1386–3711
17. Rato, R.T., Ortigueira, M.D., Batista, A.G.: On the HHT, its problems, and some solutions. Mech. Syst. Sig. Process. **22**(6), 1374–1394 (2008)

A Markov Process-Based Approach for Reliability Evaluation of the Propulsion System in Multi-rotor Drones

Koorosh Aslansefat(✉), Francisco Marques, Ricardo Mendonça, and José Barata

CTS-UNINOVA, Universidade Nova de Lisboa (UNL), Caparica, Portugal
{koorosh.aslansefat, fam, rmm, jab}@uninova.pt

Abstract. Autonomous multirotor drones as a popular type of Unmanned Aerial Vehicles (UAVs) have a tremendous potential to facilitate activities such as logistics, emergency response, recording video, capturing special events, and traffic management. Despite the potential benefits the possibility of harming people during operation should be considered. This paper focuses on modeling the multirotor drones' propulsion system with Markov chains. Using the proposed model, both reliability and Mean Time To Failure (MTTF) of the propulsion system are evaluated. This study proposes a fault detection and recovery system based on a Markov Model for mission control of multirotor drones. Concretely, the proposed system aims to reduce potential injuries by increasing safety.

Keywords: Markov models · Mission planning · Decision making · Reliability evaluation · Flight control · UAV

1 Introduction

Reliability can be defined as the probability of multirotor functioning correctly during a given timespan. Safety, on the other hand, is the probability that a multirotor either functions correctly or stops its operation in a safe manner without causing injuries. To design a system that is safe and reliable, multirotor faults and failures need to be studied. Furthermore, performance modeling of the multirotor in the presence of faults and failures needs also to be done. By modeling multirotors for performance and functionality evaluation, bottlenecks and weak points of the system can be determined and strengthened. Having the Markov model of the multirotor is not only beneficial for reliability evaluation. But also, for proposing a mission control and diagnostics system to reduce possible injuries caused by faults and failures in multirotors' systems and subsystems. As failures in the propulsion system can potentially cause a complete loss of control of the UAV it was chosen as the focus of this paper. In the following paragraphs, a brief literature survey on the research of faults and failures in UAVs is provided. In (Murtha 2009) an FTA based procedure is proposed for reliability improvement in a cost-effective way. Dempster-Shafer Theory is also used to reduce the uncertainty in failure data. However, while the aforementioned references have

Published by Springer Nature Switzerland AG 2019
L. M. Camarinha-Matos et al. (Eds.): DoCEIS 2019, IFIP AICT 553, pp. 91–98, 2019.
https://doi.org/10.1007/978-3-030-17771-3_8

used FTA that is limited to modeling static characteristics of the systems we propose to cover the dynamic behavior of the multirotors by leveraging on a Markov model. Furthermore, while in (Olson and Atkins 2013), only the qualitative failure analysis for the Michigan UAV has been addressed we will focus on the quantitative failure analysis. In this paper, the remaining useful life (RUL) of the UAV is estimated through fault tree analysis and has been used for health-based task allocation (Shi et al. 2016), based on the controllability degree, the reliability of multirotor with different configurations have been studied. In (Belcastro et al. 2017) experimental flight test techniques with different possible sets of hazard-based test scenarios are proposed in order to evaluate the safety of sUAS operations. In spite of intensive research conducted so far, there is no published work that we are aware of on Markov modeling for fault detection and recovery for multirotor robots. This paper introduces the Markov model of a hexacopter and evaluates its reliability and MTTF. The organization of this paper is as follows. In Sect. 2, the relationship to industrial and service systems will be addressed briefly. Section 3 introduces Markov modeling of the propulsion system in multirotors and Sect. 4 illustrates numerical results obtained from introduced Markov models. In Sect. 5, a brief discussion on the idea of Markov-based fault detection and recovery is briefly explained. The paper ends with some concluding remarks and possible future avenues.

2 Relationship to Industrial and Service Systems

Nowadays, multirotors and small unmanned aircraft systems (sUAS) are being developed for consumer and commercial applications in urban areas that encompass logistics, emergency response, filming, traffic monitoring, agriculture monitoring, search and rescue, railway surveillance, and corrosion inspection in bridges. Many applications entail operating in populated areas where a failure of the multirotor can endanger humans. Therefore, reliability and safe landing of multirotors are important and challenging because of their dynamic behavior and complex models (Sadeghzadeh et al. 2011).

3 Markov Modelling of the UAV's Propulsion Systems

In this section, the proposed Markov model and its solution will be provided for reliability and MTTF calculation. In addition, the simplification of Markov model for multirotors will be introduced.

3.1 Simple Markov Model and Its Continuous-Time Solution

Figure 1 shows the Markov reliability model of a simple system with one operational and one failure states. States and transitions between states are first recognized, and the rates of transitions are allocated. With exponential failure rate distribution function, the reliability of a system after period of time is then expressed by (1), and its unreliability (failure probability) is given by (2).

$$e^{-n\lambda t} = 1 - n\lambda\Delta t + n^2\lambda^2\Delta t^2 - \ldots \tag{1}$$

$$1 - e^{-n\lambda t} = 1 - \left(1 - n\lambda\Delta t + n^2\lambda^2\Delta t^2 - \ldots\right) \to n\lambda\Delta t \tag{2}$$

Thus, the probability of the system at the first state after elapsing Δt time is given by: $1 - \lambda\Delta t$. In this figure, Lambda stands for the rate of failure.

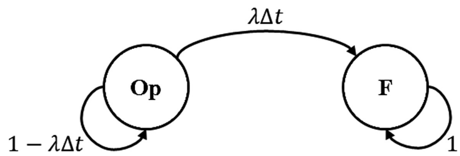


Fig. 1. Reliability Markov model of a simple system.

3.1.1 Reliability Evaluation

Based on the Markov model illustrated in Fig. 1 and according to the Markov theorem, (3) can be written as follows (Dubrova 2013).

$$P(t + \Delta t) = M.P(t) \tag{3}$$

where P is the "probability states vector" and M is the discrete state transition matrix.

$$P(t) = [P_{OP}(t), P_F(t)]^T \tag{4}$$

$$M = \begin{bmatrix} 1 - \lambda\Delta t & \lambda\Delta t \\ 0 & 1 \end{bmatrix} \tag{5}$$

Equation (3) can be recursively solved if as an initial probability vector is known. The result for different times is given by (6).

$$P(n\Delta t) = P^n.P(0) \tag{6}$$

Equation (7) is the continuous form of (3).

$$\dot{P}(t) = A.P(0) \tag{7}$$

where A is the continuous Markov transition matrix in the form of (8).

$$A = \frac{\partial M^T}{\partial \Delta t} = \begin{bmatrix} -\lambda & 0 \\ \lambda & 0 \end{bmatrix} \tag{8}$$

Solving (7) gives the probability of system states at any time t.

$$P_{OP}(t) = e^{-\lambda t} \tag{9}$$

$$P_F(t) = 1 - e^{-\lambda t} \tag{10}$$

The reliability of a system can be obtained from the probability of states in Markov model those which represent the operational condition of the system. Hence, the reliability of a simple system is now calculated as equations and (11). In this figure, each Lambda stands for the failure rate of a motor. For example, Lambda a stands for the failure rate of the motor a (Fig. 2).

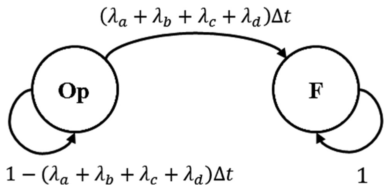


Fig. 2. Reliability Markov model of a Quadcopter

Having the Markov model for a quadcopter, the reliability expression of the system can be obtained easily by (12).

$$R_{Quadcopter}(t) = P_{OP}(t) = e^{-(\lambda_a t + \lambda_b t + \lambda_c t + \lambda_d t)} \tag{11}$$

3.1.2 MTTF Calculation

Same Markov model for Mean Time To Failure (MTTF) calculation can be used as in Fig. 1. This Markov model is an arbitrary absorbing Markov chain which is described by a transition matrix P. If there are r absorbing states and t transient states, the transition matrix will have the following canonical form. The MTTF Markov model has only one absorbing state and so in this case (I = 1).

$$P = \begin{bmatrix} Q & R \\ 0 & I \end{bmatrix} \tag{12}$$

Consider $N = (I - Q)^{-1}$ as the fundamental matrix of P and Let t_i be the expected number of steps before the chain is absorbed (goes to state A), given that the chain starts in state s_i and let TTF be the column vector whose i^{th} entry is TTF_i. The column vector of TTF can be written as follows:

$$TTF = \begin{bmatrix} t(NA_1) & t(NA_2) & \ldots & t(NA_n) \end{bmatrix}^T = NC \tag{13}$$

where C is a column vector all of whose entries are one. Once the system starts from state Op (Operational), the expected number of steps to be in the failure state (F) can be achieved by (14).

$$MTTF = TTF(Op) \tag{14}$$

3.2 Simple Markov Model and Its Continuous-Time Solution

Figure 3 illustrates a simple hexacopter configuration with rotors labeled from a to f. In this configuration, motors are PNPNPN. P stands for positive clockwise direction and N stands for negative anti-clockwise direction (Shi et al. 2016).

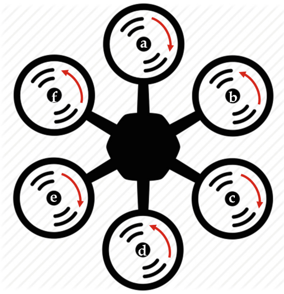


Fig. 3. A simple hexacopter with rotors labeled from a to f (PNPNPN configuration).

Based on the hexacopter illustrated in Fig. 3, the Markov model of Fig. 4 can be obtained. In this model, 19 states are considered. The first state in this model is “a b c d e f” where all rotors are fully operational. If during the mission one of the rotors fails (for example the one labeled “a”), then with the failure rate of $\lambda_a \Delta t$ the system goes to the “b c d e f” state. In the state: b c d e f”, if the rotor “f” or “b” fails, then the systems goes to the “Failure” state. With all these possible failures in the hexacopter, the following model will be constructed. (For more information about the operational and failure modes based on the controllability theory please refer to (Shi et al. 2016)).

3.3 Model Simplification

Assuming that “the failure rates of all rotors are the same” the model presented in Fig. 4 can be simplified into the one depicted in Fig. 5 the first state in this model labeled with “6” and “M” means that in this state all six rotors are operational. In case of having only four rotors operational, two states can be considered. The star state demonstrates the system with operational rotors in a quadcopter configuration while the second one covers the other possible configurations. With the assumption of having same failure rates of all rotors, the 19 states Markov model is simplified to an understandable and less complicated 6 states Markov model.

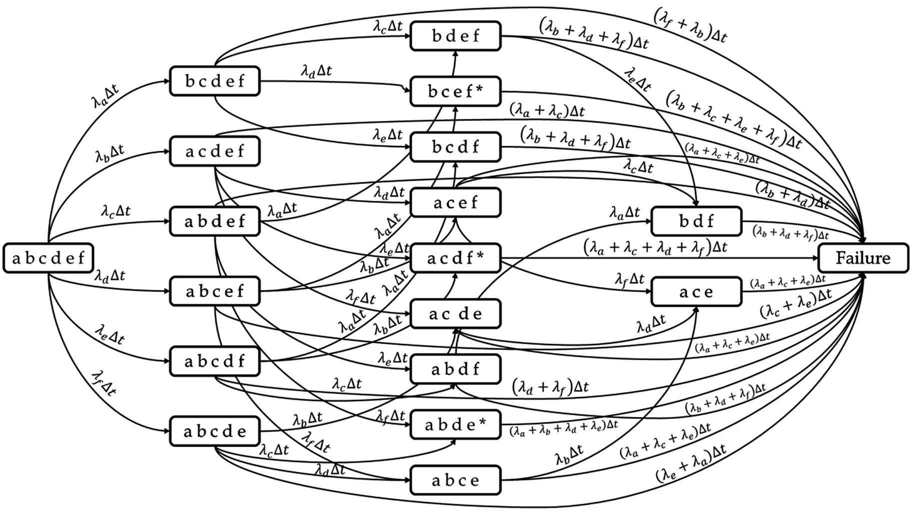


Fig. 4. Reliability Markov model of a hexacopter with PNPNPN configuration.

4 Numerical Results

For understanding the system's behavior in terms of reliability and MTTF, numerical results are provided in this section. Figure 6 shows the evaluation of the reliability of a quadcopter and hexacopter with two kinds of rotors configuration versus time. In this figure, the failure rate of each rotor is assumed as 0.04 failures per hour. The reliability of hexacopter with PNPNPN configuration is higher than the reliability of hexacopter with PPNNPN configuration and both are higher than the reliability of quadcopter. In (Shi et al. 2016), the reliability of hexacopter with both configurations and without yaw control has been considered with the same expression. However, the result achieved from Markov model shows that they have different values of reliability vs. time in PPNNPN configuration.

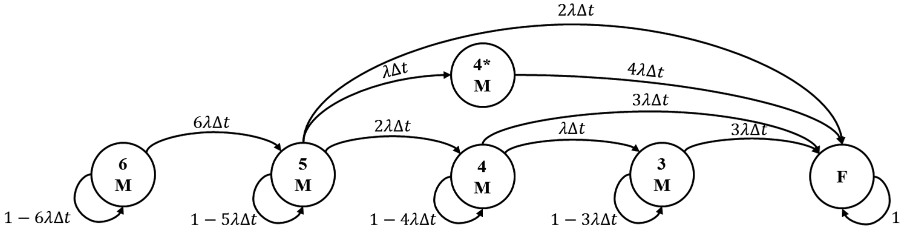


Fig. 5. Simplified reliability Markov model of Hexacopter with PNPNPN configuration.

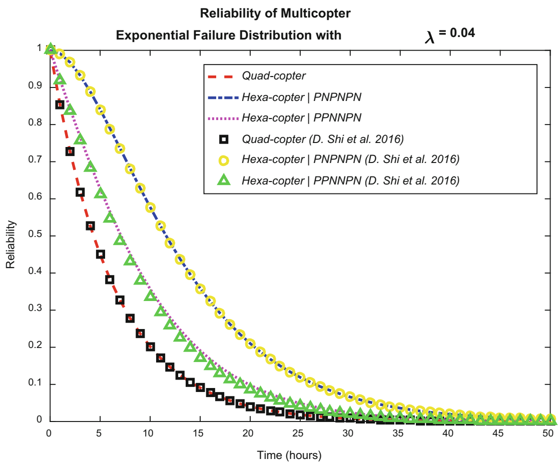


Fig. 6. Reliability evaluation of Multirotors (Quadcopter and Hex-copter) vs. time with failure rate of 0.04 failure/hour

5 Discussion

Markov modeling of multirotors is the first step in designing the fault detection and recovery system. A fault diagnosis module will be needed to specify the current state of the multirotor in its Markov model. Hence, it is assumed that there is a fault detection and diagnosis module that can detect the failure of each rotor. Based on the proposed Markov model and detection of a failed rotor, the mean time to failure and the probability of failure from a related state can be estimated. If both probabilities of failure and MTTF are less than desired value, the robot will continue the mission. However, if the probability of failure or MTTF is more than the desired value the robot will abort the mission or perform a safe landing.

6 Conclusions

The higher maneuverability of multi-rotors compared with other types of UAVs makes them very popular. However, safety when operating multi-rotors in populated areas remain one of the most important challenges. In this sense, a Markov model was herein proposed as a fault detection and recovery module for the propulsion system of multi-rotors and was validated through Monte Carlo simulations. Reliability and MTTF of the robot were also evaluated via the developed models. In particular, it was shown that the reliability of a hexacopter configuration is higher than a quadcopter. Moreover, between the two hexacopter configurations studied, the PNPNPN configuration has higher reliability and MTTF when compared to the PPNNPN.

References

Belcastro, C.M., Klyde, D.H., Logan, M.J., Newman, R.L., Foster, J.V.: Experimental flight testing for assessing the safety of unmanned aircraft system safety-critical operations. In: 17th AIAA Aviation Technology, Integration, and Operations Conference, Denver, Colorado (2017)

Dubrova, E.: Fault-Tolerant Design. Springer, Berlin (2013). https://doi.org/10.1007/978-1-4614-2113-9

Murtha, J.F.: Evidence theory and fault-tree analysis to cost-effectively improve reliability in small UAV design (2009)

Olson, I., Atkins, E. M.: Qualitative failure analysis for a small quadrotor unmanned aircraft system. In: AIAA Guidance, Navigation, and Control (GNC) Conference, Boston, MA (2013)

Sadeghzadeh, I., Mehta, A., Zhang, Y.: Fault/damage tolerant control of a quadrotor helicopter UAV using model reference adaptive control and gain-scheduled PID. In: AIAA Guidance, Navigation, and Control Conference, Portland, Oregon (2011)

Shi, D., Yang, B., Quan, Q.: Reliability analysis of multicopter configurations based on controllability theory. In: IEEE 35th Chinese Control Conference, Chengdu, China (2016)

A New Approach for Crop Rotation Problem in Farming 4.0

Bruno S. Miranda[1](✉), Akebo Yamakami[1], and Priscila C. B. Rampazzo[2]

[1] School of Electrical and Computer Engineering (FEEC-UNICAMP), Av. Albert Einstein, 400, Campinas 13083-852, Brazil
miranda.bsm@live.com, akebo@dt.fee.unicamp.br

[2] School of Applied Sciences (FCA-UNICAMP), R. Pedro Zaccaria, 1300, Limeira 13484-350, Brazil
priscila.rampazzo@fca.unicamp.br

Abstract. Technology and innovations have long improved farming over the world and, as Industry 4.0 quickly spread, farmers have embraced high-level automation and data exchange, driving a transformation called Farming 4.0. Consequently, precise and even real-time field information have become easily accessible. Though, analyzing all this information requires great skills and tools, like mathematical knowledge and powerful computational algorithms to reach farmers expectations. This research explores the Crop Rotation Problem (CRP) and its relevance for the integration of Precision Agriculture (PA) and farm management. This paper presents a new mathematical approach for the CRP based on the nutrient balance and crop requirements, increasing the sustainable appealing of the problem. A real-encoded genetic algorithm (GA) was developed for optimization of the CRP. The results indicate good performance in mid and long-term crop scheduling.

Keywords: Crop Rotation Problem · Farming 4.0 · Precision Agriculture · Genetic Algorithm · Farm management

1 Introduction

Technological advances have been widely shaped agriculture. Innovations in sensor devices and embedded systems have improved tillage and crop yields. Efficiency in resources management and autonomous data acquisition turned out to be essential among farmers. As Industry 4.0 thoughts have pushed ahead manufacturing units, agriculture also moves toward a transformation called Farming 4.0, or also referred to as Agriculture 4.0. In general, important developments in automation have been noticed, related to Precision Agriculture expansion [1, 2].

Precision agriculture (PA) follows from the integration of crop management and information technologies. The field measurement of crop requirements and the proper supply of these demands are promoted by PA and intend to improve crop production and resource consumption, approaching sustainable ideas and reducing environment impacts [3, 4].

Published by Springer Nature Switzerland AG 2019
L. M. Camarinha-Matos et al. (Eds.): DoCEIS 2019, IFIP AICT 553, pp. 99–111, 2019.
https://doi.org/10.1007/978-3-030-17771-3_9

Industry 4.0 relies on digital technologies, such as Big Data, Internet of Things (IoT) and Artificial Intelligence (AI), to enhance production systems and reach the highest level of cooperation and mobility. In agriculture, connectivity has been expanding allied with IoT and transforming the whole infrastructure. This trend has been noticed in farm equipment: connected tractors, spray devices with route optimization devices and other advanced machines, which are able to gather data and even share information with their manufactures for maintenance purposes [5].

Besides farm technologies stand as advanced as industrial applications, production planning remains a challenging task in agriculture. Aside from monoculture, scheduling crops in a large set of fields rely on market outcomes, weed, and pest control or weather forecasts. Searching for strict mathematical approaches remains quite unusual, relying much on the farmer expertise and short-term goals.

According to USDA, Latin American agricultural GDP growth ought to recover strength in a short-term period marked by recession and slow growth, which relates mainly to Brazil and Argentina economic instability [6]. Brazil is one of the greatest food producers in the world. Investments in agrarian researches and development are quite expressive, they reached 1.82% of Brazilian GDP in 2013 [7]. Therefore, agrarian researches present as much potential as industrial ones for Latin America economies.

This research paper acknowledges the importance of CRP in the Farming 4.0 developments. The proposed mathematical model for CRP bases on fertilization management techniques, such as nutrient budgets.

The main contribution of this paper relies on the relevance of the proposed model supporting farm management decisions. It achieves profitable solutions and evaluates the resource balance, providing valuable insights in the crop sequence.

To accomplish a broad range of CRP scenarios, optimization techniques ought to provide solutions for small and sizeable instances of the problem. The proposed evolutionary algorithm presents good results in the large instances where the deterministic method takes large amounts of time. The stochastic approach in this paper combined with the deterministic method ensures that the proposed model can assist agrobusiness management providing solutions for many scenarios.

The Subsect. 1.1 presents a review about the CRP approaches in the literature. Section 2 establishes the relationship to industrial and service systems. Section 3 presents the proposed model and its details. The optimization techniques are discussed in Sect. 4 and the attained results in Sect. 5. The conclusion and further steps of this research presents in Sect. 6.

1.1 Discussion on Previous Works

The CRP has been extensively researched. Distinguished approaches and mathematical models have enriched widely the available literature. The complexity related to this NP-hard problem is the main reason for continuous innovation. Heuristic and metaheuristic approaches are quite common among applied techniques.

A review of the stochastic and deterministic methods applied to the CRP presented in [8] and pointed out that the deterministic approach has been successful in some specific models, though lacks in efficient over sizeable instances of the problem.

Although there is no agreement among researches about the best method, evolutionary and hybrid algorithms are promising, due to the high flexible structure of genetic algorithms.

Concerned about a large amount of consumed resources and the environmental impact of monoculture farming, the research presented in [9] discussed an optimization model for CRP, based on organic farming concepts. A new column generation algorithm combined with a greedy heuristic was developed and generated solution for this proposed model of the CRP.

The research presented in [10] analyzed hybrid metaheuristic algorithms in search of quality solutions for the CRP. Attaining feasibility presented to be a hard task on the proposed model, to overcome this hurdle, the initial population was generated by a heuristic procedure. The hybrid algorithms with local search and with Simulating Annealing presented good results in this related work.

Multi-objective approaches for the CRP are quite contemporary. A bi-objective approach presents in [11], it acknowledges that profitability and diversity of crop rotation are conflicted goals and provides a bi-objective model that explores both objectives.

2 Relationship to Industrial and Service Systems

Based on a vast area of expertise, Operational Research deals with problems regardless of the context in which they arise. Problems in which the objective is to determine, according to one criterion, the best choice within a set of alternatives. The development of the areas of engineering, computing, and economics has been characterized by the increasing use of optimization models as paradigms for representation and resolution of decision-making problems.

Scheduling Problems represent a class of significant decision-making problems in optimization. Numerous companies and service systems face problems regarding task sequencing, which can be caused by improper allocation of resources and poorly defined processes. The areas affected by these problems are diverse: industry, manufacturing, agriculture, process management, transportation, among others. We can optimize processes by performing task sequencing planning, which results in improved production flow control, meeting deadlines, and scheduling tasks to better utilize available resources.

Solutions of this type of problem are not obtained in closed forms. Instead, they are determined by algorithms: a sequence of procedures applied repeatedly to the problem until the best solution is obtained. According to the adopted formulation (fundamentally dependent on the parameters), we can solve the problem exactly in polynomial time, or even deal with NP-class problems. In these cases, it is necessary to abandon the search for an optimal solution to seek a quality solution, through heuristic procedures.

The present research aims to provide a valuable method to analyze information only acquired by the current state of interconnection between physical and cyber world, such as actual crop nutrient requirements. And, although the proposed methodology applies to agrarian management, the concepts of Industry 4.0 are embraced in this research. The perception of this work exceeds agricultural environment by the integration of planning, profit and production requirements.

3 Crop Rotation Problem

The proposed mathematical model for the CRP is described ahead and bases on mathematical approaches presented in [9, 10]. The nutrient balance concept presented in the model relates to researches in [12, 13] and bases on the surface nutrient budget idea.

- N: size of the crop set;
- N_f: number of crop families;
- F_p: the set of crops from the family p;
- M: number of periods;
- L: number of plots;
- $area_k$: available tillage area of each plot k (acre);
- l_i: profitability of crop ‘i’ ($);
- t_i: crop ‘i’ production cycle, from planting date to the harvesting;
- p_i: crop ‘i’ average production per acre;
- I_i: crop planting interval $\{ I_i^1, \ldots, I_i^n \}$;
- D_i: demand for crop ‘i’ (units / M periods);
- S_k: adjacent plots of the plot k;
- $F_{N_{\alpha k}}, F_{P_{\alpha k}}, F_{K_{\alpha k}}$: Quantity of nitrogen, phosphorus and potassium fertilizer applied in the plot k over the interval α;
- R_{N_i}, R_{P_i}, R_{K_i}: requirement of nitrogen, phosphorus and potassium per area unit for crop ‘i’;
- F_{min}, F_{max}: Fertilization limits;
- c_N, c_P, c_K: fertilization costs ($/ unit);
- β: sequence crop restriction factor;
- θ: interval of fertilization balance.

$$max \sum_{i=1}^{N} \sum_{j=1}^{M} \sum_{k=1}^{L} area_k \cdot l_i \cdot p_i \cdot x_{ijk} - \sum_{\alpha \in \Omega} \sum_{k=1}^{L} F_{N_{\alpha k}} \cdot c_N + F_{P_{\alpha k}} \cdot c_P + F_{K_{\alpha k}} \cdot c_K \tag{1}$$

$$s.a \sum_{i \in F_p} \sum_{r=0}^{t_i - 1} \sum_{v \in S_k} x_{i(j-r)v} \leq L \cdot \left(1 - \sum_{i \in F_p} \sum_{r=0}^{t_i - 1} x_{i(j-r)v} \right), p = 1, \ldots, N_f, j = 1, \ldots, M, k = 1, \ldots, L \tag{2}$$

$$\sum_{i \in F_p} \sum_{r=0}^{t_i + \beta} \sum_{v \in S_k} x_{i(j-r)k} \leq 1, p = 1, \ldots, N_f, j = 1, \ldots, M, k = 1, \ldots, L \tag{3}$$

$$\sum_{i=1}^{N}\sum_{r=0}^{t_i-1} x_{i(j-r)k} \leq 1, j = 1,\ldots,M, k = 1,\ldots,L \tag{4}$$

$$F_{N_{\alpha k}} - \sum_{i=1}^{N}\sum_{j=1+(\alpha-1)\cdot\theta}^{\alpha\cdot\theta} x_{ijk}\cdot area_k \cdot R_{N_i} \geq 0, k = 1,\ldots,L, \alpha \in \Omega \tag{5}$$

$$F_{P_{\alpha k}} - \sum_{i=1}^{N}\sum_{j=1+(\alpha-1)\cdot\theta}^{\alpha\cdot\theta} x_{ijk}\cdot area_k \cdot R_{P_i} \geq 0, k = 1,\ldots,L, \alpha \in \Omega \tag{6}$$

$$F_{K_{\alpha k}} - \sum_{i=1}^{N}\sum_{j=1+(\alpha-1)\cdot\theta}^{\alpha\cdot\theta} x_{ijk}\cdot area_k \cdot R_{K_i} \geq 0, k = 1,\ldots,L, \alpha \in \Omega \tag{7}$$

$$\sum_{j=1}^{M}\sum_{k=1}^{L} area_k \cdot p_i \cdot x_{ijk} \geq D_i, i = 1,\ldots,N \tag{8}$$

$$\sum_{k=1}^{L}\sum_{j\notin I_i} x_{ijk} = 0, i = 1,\ldots,N \tag{9}$$

$$x_{ijk} \in \{0,1\}, i = 1,\ldots,N, j = 1,\ldots,M, k = 1,\ldots,L \tag{10}$$

$$\mathrm{F}_{\mathrm{N}_{\alpha k}} = \{\mathrm{F}_{\mathrm{N}_{\alpha k}} \in R^+ | F_{max} \geq \mathrm{F}_{\mathrm{N}_{\alpha k}} \geq F_{min}\}, k = 1,\ldots,L, \alpha \in \Omega \tag{11}$$

$$\mathrm{F}_{P_{\alpha k}} = \{\mathrm{F}_{P_{\alpha k}} \in R^+ | F_{max} \geq \mathrm{F}_{\mathrm{P}_{\alpha k}} \geq F_{min}\}, \qquad k = 1,\ldots,L, \alpha \in \Omega \tag{12}$$

$$\mathrm{F}_{\mathrm{K}_{\alpha k}} = \{\mathrm{F}_{\mathrm{K}_{\alpha k}} \in R^+ | F_{max} \geq \mathrm{F}_{\mathrm{K}_{\alpha k}} \geq F_{min}\}, \qquad k = 1,\ldots,L, \alpha \in \Omega \tag{13}$$

$$\Omega = \{\alpha \in N^* | \alpha\cdot\theta \leq M, \theta \in N^*\} \tag{14}$$

The objective function presented in Eq. 1 evaluates crop schedule profits and fertilization costs. Besides crop budgets already cover fertilization costs, the objective function also incurs these costs to minimize the number of consumed resources.

The constraint set in Eq. 2 prevents adjacent areas to hold the same crop family scheduled in the same period and the constraint set in Eq. 3 ensures that crops from the same family shall not be scheduled in sequence on each crop field.

The constraint set in Eq. 4 prevents scheduling more than one crop in the same period, in other words, it is a spatial restriction. Fertilization balances are established in Eqs. 5, 6 and 7, following the concept of surface nutrient budget. The production requirements for each crop are evaluated by the constraint set in Eq. 8. The constraint set in Eq. 9 ensures that crop scheduling happens just in the proper planting period, denying allocation outside this window. The decision variables x_{ijk} are Boolean type and each one represents the schedule of the crop "i" in the period "j" on the field "k" and the fertilization variables ($\mathrm{F}_{\mathrm{N}_{\alpha k}}$, $\mathrm{F}_{\mathrm{P}_{\alpha k}}$ and $\mathrm{F}_{K_{\alpha k}}$) are real values.

In order to enlighten the mathematical model and its constraints, a sample solution of the CRP is presented in Table 1.

Table 1. A crop schedule solution attained by the proposed model for the CRP. The left side presents the crop schedule in a 24 periods interval (each period corresponds to a 15-day interval). The right side presents the related families to the sequence.

<table>
<tr><th rowspan="2">Periods</th><th colspan="3">CROPS</th><th colspan="3">FAMILIES</th></tr>
<tr><th>Plot 1</th><th>Plot 2</th><th>Plot 3</th><th>Plot 1</th><th>Plot 2</th><th>Plot 3</th></tr>
<tr><td>1</td><td rowspan="4">Cabbage for Fresh Market</td><td rowspan="6">Strawberries for Fresh Market, Spring</td><td rowspan="4">Leaf Lettuce for Fresh Market, Winter</td><td rowspan="4">Mustard</td><td rowspan="6">Rose</td><td rowspan="4">Lettuce</td></tr>
<tr><td>2</td></tr>
<tr><td>3</td></tr>
<tr><td>4</td></tr>
<tr><td>5</td><td rowspan="10">Carrots for Fresh Market, Summer</td><td>***</td><td rowspan="10">Carrot</td><td>***</td></tr>
<tr><td>6</td><td>***</td><td>***</td></tr>
<tr><td>7</td><td rowspan="4">Leaf Lettuce for Fresh Market, Summer</td><td rowspan="4">Strawberries for Fresh Market, Summer</td><td rowspan="4">Lettuce</td><td rowspan="4">Rose</td></tr>
<tr><td>8</td></tr>
<tr><td>9</td></tr>
<tr><td>10</td></tr>
<tr><td>11</td><td rowspan="4">Tomato for Fresh Market, Summer</td><td>***</td><td rowspan="4">Nightshade</td><td>***</td></tr>
<tr><td>12</td><td rowspan="8">Watermelons for Fresh Market, Summer</td><td rowspan="8">Cucurbit</td></tr>
<tr><td>13</td></tr>
<tr><td>14</td></tr>
<tr><td>15</td><td rowspan="4">Spinach for Fresh Market, Fall</td><td>***</td><td rowspan="4">Beet</td><td>***</td></tr>
<tr><td>16</td><td rowspan="8">Carrots for Fresh Market, Winter</td><td rowspan="8">Carrot</td></tr>
<tr><td>17</td></tr>
<tr><td>18</td></tr>
<tr><td>19</td><td rowspan="4">Cabbage for Fresh Market</td><td rowspan="4">Mustard</td></tr>
<tr><td>20</td><td rowspan="4">Leaf Lettuce for Fresh Market, Winter</td><td rowspan="4">Lettuce</td></tr>
<tr><td>21</td></tr>
<tr><td>22</td></tr>
<tr><td>23</td><td>***</td><td>***</td></tr>
<tr><td>24</td><td>Spring Onions for Fresh Market</td><td>Strawberries for Fresh Market, Spring</td><td>Spring Onions for Fresh Market</td><td>Lily</td><td>Rose</td><td>Lily</td></tr>
</table>

The constraint set in Eq. 2 prevents the same family in adjacent crops. This can be verified in the sample solution in Table 1, according to the adjacent plots presented in Table 2.

Table 3 shows that the crops from the same family can be schedule again in the same plot after a period defined by β, which satisfy the constraint set in Eq. 3.

Nutrient constraints in Eq. 5, 6 and 7 are quite intuitive and rely on supplying the required inputs from the scheduled crops on the interval θ. in each plot "k". Reaching production requirements follows from constraint set in Eq. 8 and scheduling in the proper planting period from Eq. 9. The fertilization inputs and the attained production from sample solution are not presented for assumption of easy understandability.

Table 2. The adjacency among plots for this solution presents on the left side. Each row represents adjacent plots from the row index. Additionally, the right side shows each plot area.

Adjacent plot matrix				Field area (acre)
	Plot 1	Plot 2	Plot 3	
Plot 1				1
Plot 2				1.5
Plot 3				0.8

Table 3. Partial view of the previous sample solution. It shows how the problem fits the constraint set.

Periods	CROPS	FAMILIES
	Plot 1	Plot 1
1	Cabbage for Fresh Market	Mustard
2		
3		
4		
5	β	β
6		
...		
17		
18		
19	Cabbage for Fresh Market	Mustard
20		
21		
22		
23	***	***
24	Spring Onions for Fresh Market	Lily

The presented CRP model requires data about the crop features and the market. Usual planting dates were gathered from [14], average prices and production statistics based on [15]. Crop budget for the established profitability were mainly taken from [16] and, in some cases, updated with production and average prices from [15]. The crop nutrient requirements were acknowledged from [17].

4 Methodology

Evolutionary algorithms are computational procedures for solving problems, resulting from the iterative application of heuristic techniques, which makes these algorithms capable of promoting the search for a solution in a huge space of possibilities, in a very flexible way [18]. Evolutionary Algorithms work with a population formed by a set of individuals. The coding of the individuals (representation of possible solutions to the problem) is the most important step of the algorithm. The adaptation function allows assigning each element of the search space (individual) a value that is used as a measure of performance. In optimization problems, this function incorporates all aspects of the objective function. During the generations, this population is evaluated.

The most suitable ones tend to be selected and can undergo modifications in their characteristics through crossover and mutation operators, generating descendants; finally, the more adapted go to the next generation, resulting in individuals getting more fit, while others tend to disappear.

Some characteristics of these methods deserve to be highlighted: flexibility, generality, ability to escape from great locations, ability to deal with complex problems for which it is not possible or difficult to obtain a detailed description, as well as being less susceptible to form or continuity [19–21].

For the development of the approach proposed in this work, we chose the Genetic Algorithm (GA) [22], because it is a classic method, one of the first Evolutionary Algorithms proposed in the literature and easily adapted to any problem, besides being extensively applied in the different areas of engineering. The functionalities of the algorithm proposed in this work were implemented with the concern of respecting the physical and operational characteristics of the problem. The coding technique and the main operators to approach this work are described ahead.

4.1 Encoding and Decoding

The binary decision variables in the CRP might appear rather suitable for binary encoded genetic algorithms, but evaluations showed that performance decreases on mid-term and long-term planning. Also, the binary GA requires hybrid approaches to improve performance [9, 10]. The proposed GA based on the real encoding technique presented ahead.

The left side of Table 4 presents the data structure which holds all the feasible combinations of crop, period and plot. The right side presents a chromosome that holds two memory positions, one has a real number and the second one is an integer index related to the previous data structure.

Table 4. Real-encoded information in the GA. The left side presents a support data structure with all the possible crop allocations and the right side shows the chromosome filled with floating numbers.

Index	Crop	Period	Plot
1	1	17	1
2	1	18	1
3	1	19	1
4	1	20	1
5	1	17	2
6	1	18	2
7	1	19	2
8	1	20	2

Chromossome	Index
0.3527983	1
0.8572196	2
0.0049846	3
0.4896225	4
0.4130837	5
0.3991828	6
0.1883861	7
0.3622583	8

Decoding process initializes with sorting the chromosome according to the real numbers. The highest priority returns an index that represents the first crop to be allocated, then the process continuous until there are no more possible allocations (Table 5).

Table 5. Decoding process. The chromosome is sorted according to the floating values. The second column of the chromosome holds the index related to the support data structure. Decoding follows down this priority list until there is no more availability in the schedule.

Index	Crop	Period	Plot
2	1	18	1
4	1	20	1
5	1	17	2
6	1	18	2
8	1	20	2
1	1	17	1
7	1	19	2
3	1	19	1

Chromossome	Index
0.8572196	2
0.4896225	4
0.4130837	5
0.3991828	6
0.3622583	8
0.3527983	1
0.1883861	7
0.0049846	3

Period	Plot 1
17	
18	1
19	1
20	1
21	1
22	
23	

The fertilization variables have direct correspondence to the chromosome values, apart from the schedule priority list and without requiring any decode technique.

4.2 Genetic Operators

The presented GA uses the tournament selection as a technique to sort out individuals from the population. Tournaments happens with "k" random selected individuals; the winner is the highest fitness among them and shall be placed in the mating pool, which will generate a new offspring.

The tested crossover and mutation operators are presented in [23, 24]. A brief description of Laplace Crossover technique and Power Mutation follows ahead.

- Laplace Crossover:

  (1) Random numbers: $u_i, r_i \in [0, 1]$

$$\beta_i = \begin{cases} a - b \cdot \log(u_i), r_i \leq 0.5 \\ a + b \cdot \log(u_i), r_i > 0.5 \end{cases} \rightarrow \begin{cases} y_i^1 = x_i^1 + \beta_i \left| x_i^1 - x_i^2 \right| \\ y_i^2 = x_i^2 + \beta_i \left| x_i^1 - x_i^2 \right| \end{cases}$$

- Power Mutation:

  1. Random number: $s_1, r \in [0, 1]$
  2. $s = s_1^p,\ t = \frac{\bar{x} - x^l}{x^u - \bar{x}} \rightarrow x = \begin{cases} \bar{x} - s \cdot \left(\bar{x} - x^l\right), t < r \\ \bar{x} + s \cdot (x^u - \bar{x}), t \geq r \end{cases}$

4.3 Overview of the Proposed Genetic Algorithm

A random procedure generates the initial population. At first, the current population is the same as the initial population. Then, the genetic operators (selection, crossover and mutation) produce the so-called new population. The mixed population combines the current population and the new population, sorting individuals according to their fitness. Individuals with high fitness among the mixed population are selected to fill the current population and compose the next generation. The code was developed in C language to reach high efficiency. Figure 1 summarizes the characteristics of the GA.

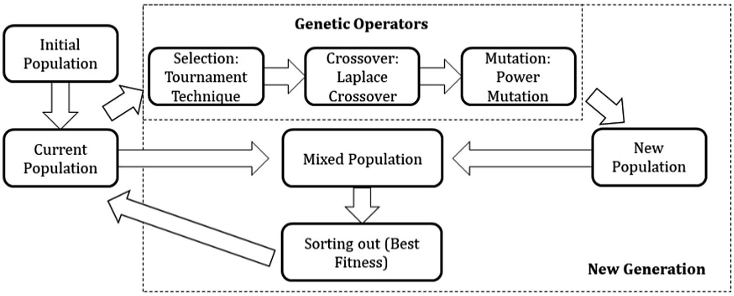


Fig. 1. A flowchart presents the proposed genetic algorithm structure.

The coding and decoding generate solutions which already fit constraint set in Eq. 2, 3, 4 and 9, decreasing infeasibility workarounds. To increase performance, there is a penalty method applied to the nutrient balance constraints.

5 Discussion on Results

The proposed model has been evaluated by a deterministic approach and by the presented GA. Glop linear solver, which is a powerful linear optimization solver developed by Google, handled the deterministic evaluations. According to Table 2, achieving exact solutions requires exhaustive computational time even in mid-term planning.

Performance of the GA was tested with 48 periods and with 72 periods in comparison with the Glop solver results. In each case, the GA ran 20 times to attain the presented results, this stochastic procedure requires several executions to validate its proper performance. The initialization process generates 1000 random individuals. Each further generation has 400 new individuals. The current population and the new population generate the mixed population, which is sorted out based on fitness and the best individuals from this population shall be inserted in the current population.

Crossover and mutation probability are 1.00 and 0.15, respectively. The total number of generations is 300. These parameters were selected after tuning and provided the best results for the tested instances.

This research presents the proposed model as potential tool for farm management. In order to ensure the reliability of this mathematical approach, there should be optimization strategies able to manage sizeable instances of the CRP. The proposed GA and the deterministic approach are complementary techniques in this work. The first one turned out to be quite suitable for the large data analyzes, whereas the exact searches may seem extremely high-cost alternatives.

The results in Tables 6 and 7 analyzes the proposed GA in terms of repeatability and stability, which are very important to ensure that the solution attained in the large instances are reasonable. Although the computational time required by the GA exceeds Glop's one, the performance of the Glop solver is undoubtedly better in the small-size problems, besides GA provides sub-optimal solutions, and so, the proposed evolutionary approach is designed to work together with deterministic methods.

Table 6. Results from executions in a 60-crop database over 48 periods (2-year schedule) and 7 plots.

	Fitness ($)	Gap %	Execution Time (s)
Optimum (Glop Solver)	463960.0	-	4589
Maximum Fitness(GA)	434067.1	6.44%	685
Minimum Fitness(GA)	386763.4	16.64%	610
Average Fitness (GA)	400880.4	13.60%	639

Table 7. Results from executions in a 60-crop database over 72 periods (3-year schedule) and 7 plots.

	Fitness ($)	Gap %	Execution Time (s)
Optimum (Glop Solver)	695087.0	-	34637
Maximum Fitness (GA)	631495.7	9.15%	1009
Minimum Fitness (GA)	563338.5	18.95%	1006
Average Fitness (GA)	587355.9	15.50%	1007

6 Conclusion

The proposed mathematical approach for the CRP enlightens agriculturally sustainable appeal without turning back on profits as the main goal of the problem. The developed algorithm achieved prospect results in a reasonable computational time and provided quality solutions for the large instances of the CRP.

This research acknowledges the potential for further developments related to CRP in sustainability and multi-objective approaches. Field crop information will increase fast and become more reliable due to observation and monitoring techniques which are driven by Precision Agriculture. Consequently, crop nutrient balance holds an important role in further developments.

The real-encoding technique presents to be very powerful in the long-term analyzes of the CRP, avoiding exceeding computational time related to the infeasibility workarounds. Establishing the relationship between crop scheduling and nutrient balance in the same model is the main contribution of this research and requires several agricultural cases to acknowledge the extent of this approach on the farm management.

Based on the presented results, the proposed mathematical model is reliable and provides quality solutions through the deterministic approach or the proposed GA, each according to the data complexity. The proposed approach for this scheduling problem and the characteristics of the GA presented, such as the real-encoding technique, could suit many applications in the Industry 4.0.

References

1. Braun, A.-T., Colangelo, E., Steckel, T.: Farming in the era of industrie 4.0. In: 51st CIRP Conference on Manufacturing System, pp. 979–984 (2018)
2. Pereira, A., Romero, F.: A review of the meanings and the implications of the Industry 4.0 concept. In: Manufacturing Engineering Society International Conference 2017, pp. 1206–1214 (2017)
3. Pandey, G., Weber, R.J., Kumar, R.: Agricultural cyber-physical system: in-situ soil moisture and salinity estimation by dielectric mixing. IEEE Access **6**, 43179–43191 (2018)
4. Far, S.T., Rezaei-Moghaddam, K.: Impacts of the precision agricultural technologies in Iran: an analysis experts' perception & their determinants. Inf. Process. Agric. **5**, 173–184 (2018)
5. Bonneau, V., Copigneaux, B., Probst, L., Pedersen, B.: Industry 4.0 in agriculture: focus on IoT aspects. In: European Commission, Digital Transformation Monitor. https://ec.europa.eu/growth/tools-databases/dem/monitor/content/industry-40-agriculture-focus-iot-aspects
6. United States Department of Agriculture (USDA): USDA Agricultural Projections to 2027. https://www.usda.gov/oce/commodity/projections/USDA_Agricultural_Projections_to_2027.pdf
7. Empresa Brasileira de Pesquisa Agropecuária (Embrapa). https://www.embrapa.br/busca-de-noticias/-/noticia/13128392/brasil-lidera-investimentos-em-pesquisa-agricola-na-america-latina
8. Memmah, M., Lescourret, F., Yao, X., Lavigne, C.: Metaheuristics for agricultural land use optimization. A review. Agron. Sustain. Dev. **365**, 975–998 (2015)
9. Santos, L.M.R., Michelon, P., Arenales, M.N., Santos, R.H.S.: Crop rotation scheduling with adjacency constraints. Ann. Oper. Res. **190**, 165–180 (2008)
10. Aliano, A., Florentino, H., Pato, M.: Metaheuristics for a crop rotation problem. Int. J. Metaheuristics **3**, 199–222 (2014)
11. Aliano, A., Florentino, H, Pato, M.: Metodologias de escalarizações para o problema de rotação de culturas biobjetivo. In: Proceeding Series of the Brazilian Society of Applied and Computational Mathematics, vol. 6 (2018)
12. Watson, C., et al.: A review of farm-scale nutrient budgets for organic farms as a tool for management of soil fertility. Soil Use Manag. **18**, 264–273 (2002)
13. Berry, P., et al.: N, P and K budgets for crop rotations on nine organic farms in the UK. Soil Use Manag. **19**, 112–118 (2003)
14. United States Department of Agriculture (USDA): Vegetables Usual Planting and Harvesting Dates. USDA (2007). https://naldc.nal.usda.gov/download/CAT30992961/PDF
15. United States Department of Agriculture (USDA): Vegetables 2017 Summary. USDA (2017). https://www.nass.usda.gov/Publications/Todays_Reports/reports/vegean17.pdf
16. Oregon State University: Oregon Agricultural Enterprise Budgets. http://arec.oregonstate.edu/oaeb/
17. Mohler, C.L., Johnson, S.E.: Crop Rotation on Organic Farms: a Planning Manual. Natural Resource, Agriculture, and Engineering Service, Ithaca (2009)
18. Knowles, J., Corne, D., Deb, K.: Multiobjective Problem Solving from Nature: From Concepts to Applications. Springer-Verlag, Berlin (2008). https://doi.org/10.1007/978-3-540-72964-8

19. Bäck, T., Fogel, D.B., Michalewicz, Z.: Evolutionary Computation 1: Basic Algorithms and Operators. Institute of Physics Publishing, Bristol (2000)
20. Bäck, T., Fogel, D.B., Michalewicz, Z.: Evolutionary Computation 1:2 Advanced Algorithms and Operators. Institute of Physics Publishing, Bristol (2000)
21. Coello, C.A.C.: Evolutionary Multiobjective Optimization: Current and Future Challenges. Advances in Soft Computing. Springer, London (2003)
22. Michalewicz, Z.: Genetic Algorithms + Data Structures = Evolution Programs. Springer-Verlag, Berlin (1996)
23. Deep, K., Singh, K.P., Kansal, M., Mohan, C.: A new crossover operator for real coded algorithms. Appl. Math. Comput. **188**, 895–911 (2007)
24. Deep, K., Singh, K.P., Kansal, M., Mohan, C.: A real coded genetic algorithm for solving integer and mixed integer optimization problems. Appl. Math. Comput. **212**, 505–518 (2009)

Optimization Systems

Application of Monte Carlo Methods in Probability-Based Dynamic Line Rating Models

Levente Rácz(✉), Dávid Szabó, Gábor Göcsei, and Bálint Németh

Budapest University of Technology and Economics, 18 Egry József str., Budapest 1111, Hungary
racz.levente@vet.bme.hu

Abstract. Due to the growing demand for electrical energy, the use of alternative transfer capacity-enhancing methods such as Dynamic Line Rating (DLR) become more and more significant. However, there are some challenges regarding the prediction of the DLR value, which are still unresolved. In the last few years several DLR pilot projects have been constituted resulting a big database of the measured environmental and load parameters. One aim of this article is to introduce how different Monte Carlo methods could be applied in probability-based DLR models to predict the DLR value and the operational safety risk factor. Based on simulations, it is possible to implement a smart DLR system in the future, which will be able to set the model parameters from time to time using big data. In order to demonstrate the advantages, relevance and limitations of the Monte Carlo simulations, a case study is presented for a genuine transmission line.

Keywords: Dynamic Line Rating · DLR · Monte Carlo methods · MCMC · Transmission line · Probability-based models

1 Introduction

Today one of the most current issues is to meet growing energy needs while reducing expenditure and maintaining operational safety in the transmission system. To meet these requirements, Dynamic Line Rating (hereinafter referred to as DLR) calculation method can be used to increase the transmission capacity of suitable transmission lines cost-effectively. The essence of DLR is that it monitors the changes of the environmental parameters in a real-time manner, and constantly adjusts the ampacity to the current state in contrast to the conventional transfer capacity calculations. In case of using DLR method it is important to have appropriate quantity of real-time data available for the calculations and provide a precise DLR prediction for the system operator in real time. The main advantages of the DLR prediction are that it encourages the integration of the renewable energy sources into the grid and makes the system more flexible, so that disastrous black-outs can be avoided. In the first sections of the article the possibility of Monte Carlo simulation is detailed. In the second part of the paper a simulation is carried out to show how the model works under real circumstances and what are the application limits. In this simulation, half year of recorded data are used to determine and tune the

Published by Springer Nature Switzerland AG 2019
L. M. Camarinha-Matos et al. (Eds.): DoCEIS 2019, IFIP AICT 553, pp. 115–124, 2019.
https://doi.org/10.1007/978-3-030-17771-3_10

main parameters of a probability-based DLR model using Markov Chain Monte Carlo (MCMC) simulation. After it, a general Monte Carlo sampling is applied to determine the DLR value at a given operational risk factor value [1, 2].

2 Relationship to Industrial and Service Systems

Several pilot projects have been started in recent years in which sensors and weather stations are installed on conductors and power line towers. In most cases sensors take place on the conductor, due to these devices needs direct contact to measure the temperature of the wire while weather stations are installed on the steel structure of the high voltage towers. These devices continuously record environmental and load data in real-time, thus forming a big data system that can be used to perform further computations and explore statistical relationships between these parameters. DLR calculation based on real-time measurement is radically different from the currently used transfer capacity-determining methods. Until now, the ampacity calculation of a transmission line was based on the so-called static way, which means that the transfer capacity is calculated from the worst-case scenario of the environmental factors. Accordingly, in this way the real capacity of each transmission lines cannot be exploited [3] (Fig. 1).

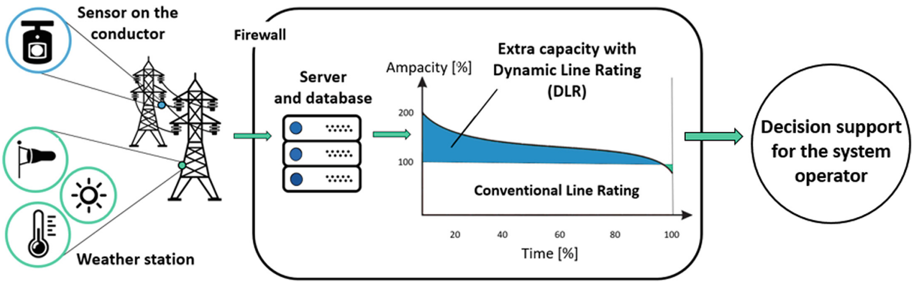


Fig. 1. Concept of Dynamic Line Rating system

Measured data from the sensors and weather stations are transmitted wirelessly to a central server which forwards them to a database. In this way it is possible to use these data for further calculations such as setting the parameters of the different environmental factors (wind speed, wind direction etc.) in the prediction models by using MCMC methods. Due to the use of the data of the weather stations a more accurate temporal and spatial resolution could be achieved in the different predictions.

3 Probability-Based DLR Models

Although in the literature mostly deterministic models are presented such as CIGRE or IEEE ones, several probability-based models have been developed in recent years. One of their major advantages is that these models work with distributions describing environmental and load conditions more accurately. Another big advantage of them is that a risk factor could be associated for such probabilities, which contains additional information for the system operator. However, these models are appropriate for the determination of conductor temperature and DLR, the real challenge is to predict these values [3, 4].

3.1 Prediction Models

The conductor temperature is significantly influenced by five factors. Four of them are environmental variables (wind speed, wind direction, ambient temperature and solar radiation) and the last one is the load variable (current of the conductor). In order to calculate the conductor temperature and the DLR, it is necessary to give prediction for these environmental variables [3, 4] (Table 1).

Table 1. The main properties of forecasting models [7]

Predicted variable	Prediction model	Distribution	Tune of the parameters
Wind speed	Probability-based	Weibull	Expectation-Maximization
Wind direction	Probability-based	von-Mises	Markov-Chain Monte Carlo
Solar radiation	Deterministic	-	-
Ambient temperature	Deterministic	-	-

In case of ambient temperature and solar radiation there is no need for probability-based prediction models, due to the high accuracy of the meteorological predictions. However, in case of wind speed and wind direction the temporal and spatial resolution of the meteorological predictions are not proper, therefore different time series models can be used to get more precise results [3].

3.2 Bayesian Approach

These wind prediction models contain various parameters that need to be tuned from time to time based on historical big data. At this point, applying Bayesian approach can lead to result for these model parameters. The essence of Bayesian approach is to determine the posterior distribution of the parameters based on the measured data and the a-prior knowledge. The Bayes formula is described by Eq. (1) [5–7].

$$P(\theta|Y) = \frac{P(Y|\theta)P(\theta)}{\int P(Y|\theta)P(\theta)d\theta} = \frac{P(Y|\theta)P(\theta)}{P(Y)} \tag{1}$$

Where Y is an observed variable, θ is a parameter of the model, $P(\theta)$ is the a-prior distribution of the parameter, $P(\theta|Y)$ is the posterior distribution of the parameter and $P(Y,\theta) = P(Y|\theta)P(\theta)$ is the joint distribution. However, in most of the cases the $P(Y)$ is unable to be calculated or even estimated, but MCMC method or EM (Expectation-Maximization) method could lead to result [5–7].

4 Monte Carlo Methods

The Monte Carlo method is a collective term, which means a variety of procedures, techniques and sampling forms. The common feature of these methods is that they are based on random number generation, thus facilitating the resolution of analytically difficult problems. In case of DLR, different Monte Carlo methods can be applied for the prediction of the parameters and for the calculation of the operational risk factor [6].

4.1 Markov Chain Monte Carlo (MCMC)

The essence of MCMC methods is that the posterior distribution of the parameters of the first-order time series models can be determined. As its name indicates, this method is based on the characteristics of Markov chains [6].

Markov chain is a series of probability variables over a state space where the probability of being in each state depends on only the previous state. If a Markov chain is aperiodic and irreducible it is called ergodic, so that it has a stationary distribution named π. Directed graphs where every change in the states has a probability forming a matrix (named P) are good representations for these kind of chains. However, the operation of MCMC contrasts with Markov chains, since in this case the P matrix is not known, but the π distribution to which the process converges is clearly declared [5] (Fig. 2).

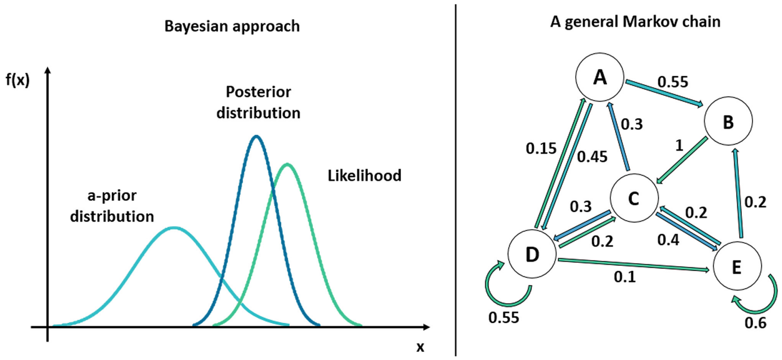


Fig. 2. Bayesian approach of model parameters and a graph representation of a general Markov chain based on [11]

There are various algorithms for the construction of Markov chains. The most common is Metropolis-Hastings and its special case named Gibbs sampling. The latter is also used in the algorithm in the wind direction prediction model [5–7].

4.2 Monte Carlo Sampling (MCS)

Using MCMC and EM methods, a full prediction of wind direction and wind speed for the time $t + 1$ can be specified. From the meteorological data there is a deterministic prediction for solar radiation and ambient temperature. If these parameters are known for the time $t+1$, then the temperature of the wire could be calculated by using the CIGRE model [4, 7].

$$c_p m \frac{d\theta_C}{dt} = P_j + P_S - P_r - P_C \tag{2}$$

Where c_p is the heat capacity of the conductor [J/kg · K], m is the specific mess of the conductor [kg/m], $d\theta_c$ is the change in the temperature of the conductor [K], dt is the period of the change in conductor temperature [s], P_j is Joule-heating [W/m], P_S is solar heating [W/m], P_c is convective cooling [W/m] and P_r is radiative cooling [W/m] (Fig. 3).

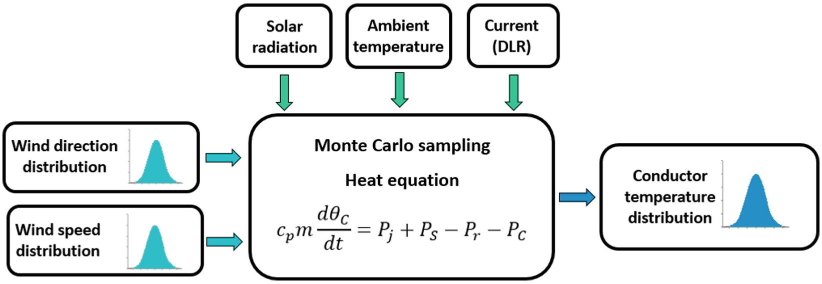


Fig. 3. Schematic diagram of the conductor temperature model [11]

Since the wind direction and wind speed are represented with a distribution it is possible to sample from them using Monte Carlo Sampling method. The essence of MCS is that the wind speed and direction values that are more likely to occur are sampled more frequently. Thus, the result of the calculation is a distribution for the conductor temperature on which an operational risk factor can be defined [7, 8].

For each wire there is a maximum temperature value for which the particular conductor has been designed. By dividing the number of instances in which this maximum temperature was exceeded with the number of the simulations, an operational risk factor can be determined with the use of Eq. (3) [7, 8].

$$P(\theta_c > \theta_{max}) = \frac{N_f}{N} \tag{3}$$

Where N_f is the number of times when the maximum temperature has been exceeded and N is the number of all simulations (Fig. 4).

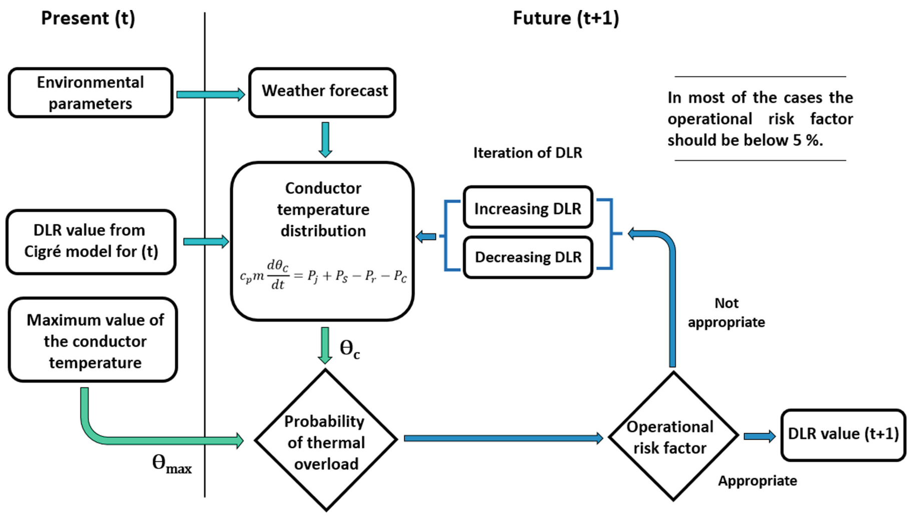


Fig. 4. DLR forecasting based on a probability-based model based on [8]

However, in reality, a reverse situation arises because a system operator strictly defines the safety factor and the ampacity of the line should be adjusted to it. In this way, the DLR value for the time $t + 1$ can be determined.

5 Case Study

While the mentioned models are known for a while, there are quite a few real-time simulations that make broader validation more difficult. As a result of DLR pilot projects, it is possible to find out the limits of the practical application of a probability-based DLR model by presenting a case study. The major aim of the case study is to predict DLR value with the operational safety factor 95% for 15 min ahead. In order to calculate this value there is a need for a 15-minute weather forecast. The wind direction is predicted by MCMC method, the wind speed is estimated with Weibull-distribution, and for the ambient temperature and solar radiation there are exact forecast from meteorology [9].

The case study included the parameters and measurement data of a 110 kV transmission line in Eastern Central Europe. On the line an ACSR wire 240/40 mm^2 is installed, and the main parameters are in Table 2. The static line rating of the line is 530 A.

Table 2. The main parameters of the overhead-line

Parameter	Value	Parameter	Value
Type of the conductor	ACSR 240/40 mm^2	Emission factor	0.55
Maximum temperature	40 °C	Resistance at 25 °C	0.120 Ω/km
Outer diameter	21.9 mm	Resistance at 75 °C	0.144 Ω/km
Outer strand diameter	3.45 mm	Average mass per unit lengths	0.987 kg/m

The line is equipped with a sensor that measures the temperature of the conductor in every 15 min with direct temperature measurement method. On the steel structure of the tower there is a weather station that includes an anemometer, a thermometer and a solar radiation measurement point. The weather station also provides real-time data in every 15 min. The operating temperature range of the sensor is between −40.0 °C and +85.0 °C and its power supply is a current transformer. The conductor temperature measurement resolution is 0.5 °C, while the temperature measurement deviation is ±2.0 °C in range from −20 °C to +100 °C. The communication with the datacenter is based on GSM/GPRS protocol [10].

5.1 Wind Direction Prediction with MCMC

The wind direction model is a first-order autoregressive Bayesian time series model, so that MCMC method could be proper to make a prediction for the next 15 min. For the simulation, all 15 min of the last 6 days have been examined [7, 8].

In the wind direction model there are 3 parameters $\alpha_1, \alpha_0, \kappa$. The α_1, α_0 has a normal a-prior distribution, while κ has a gamma a-prior distribution. For the MCMC simulation it is necessary the define the log-posterior distributions of each parameter in order to simulate the posterior distributions. According to the log-posterior distributions MCMC algorithm constructs the Markov chain with Gibbs-sampling [6, 7].

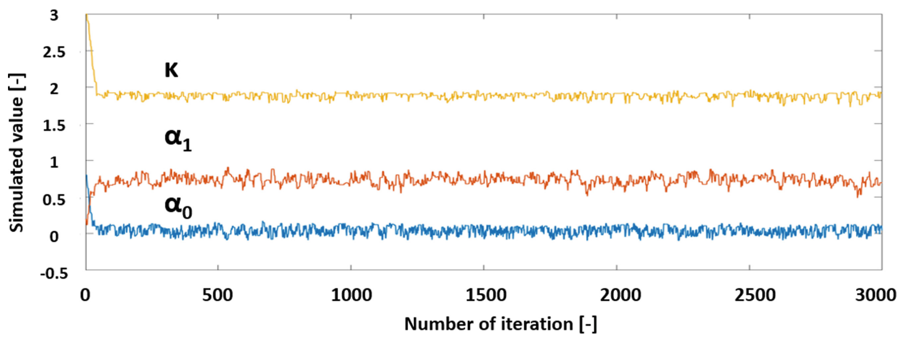


Fig. 5. Trace plots of the parameter as a result of MCMC simulation [11]

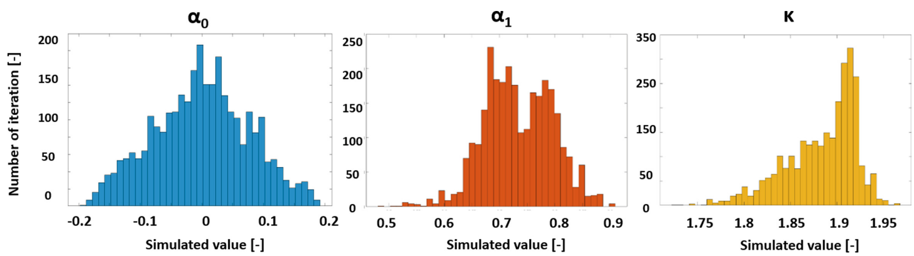


Fig. 6. Posterior distribution of the parameters at iteration number of 3000 [11]

According to Fig. 5, it can be seen from the trace plots that the MCMC simulation is successful, and the convergence of the parameters is independent from the starting value. From the trace plots the posterior distributions are illustrated in Fig. 6.

Using these parameters, it is possible to make a prediction for the wind direction. This predicted wind direction fits a von Mises distribution with a mean value 229°. However, the expected mean value of the wind direction was 13°. Although, it is also important to mention that this 13° was a struggler value, the five former and later wind direction was between 210–240°. The reason is why struggler values could occur is that in this area there is no dominant wind direction resulting extremely stochastic changes in the wind. Based on this statement, the model needs further clarification if the wind speed is under 5 m/s.

5.2 Predicting DLR Value with an Operational Safety Factor of 95%

Due to the inaccuracy in wind direction prediction, artificial distributions were generated in Matlab for also wind speed and wind direction based on the weather station measurements. The aim of this section is to present the tune of the DLR value using Monte Carlo Sampling from these distributions at a given operational safety level.

For this simulation, Eq. (2) was applied where the time step was 15 min. The first simulation was carried out to validate the model using the real-time measurement of the sensor. The SCADA system measured 99 A for that period, and the temperature of the conductor was 12.1 °C. The maximum conductor temperature of the conductor is 40 °C.

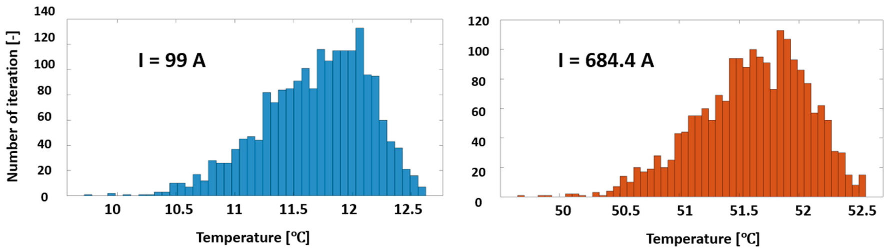


Fig. 7. The distribution of conductor temperature at 99 A and 684.4 A [11]

The first section of Fig. 7 shows that the simulated temperature approximates the measured value well, so that the model can be applied for further simulations. The second section of Fig. 7 shows the simulated temperature for time $t+1$ with DLR value calculated for time t. According to this, the reduction of current is necessary due to the high operational risk factor value.

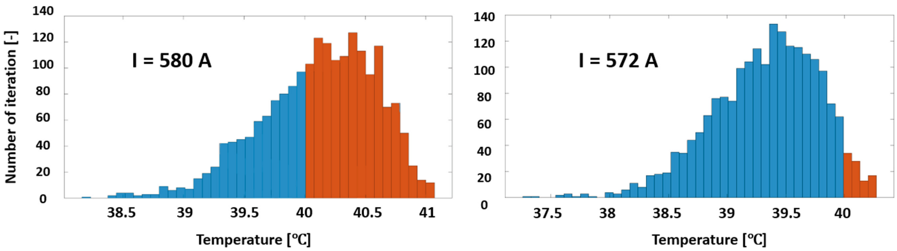


Fig. 8. The distribution of conductor temperature at 580 A and 572 A [11]

In Fig. 8, it can be seen how operational risk factor decreases with the change of the current value. According to the simulations, the 95% of operational safety level is achieved at 572 A, so that the real DLR value for the time $t + 1$ (next 15 min) is 42 A more than the static line rating. This is a nearly 10% significant increase in the transfer capacity.

6 Conclusion

In the last few years several DLR pilot project were started resulting great amount of measured data about the environmental and load parameters. The main aim of this article was to present how probability-based DLR models and different Monte Carlo simulation could be linked under real circumstances. It can be stated that Monte Carlo methods are proper for the prediction of environmental parameters and also for the calculation of the operational risk factor. If the operational risk factor is determined by the system operator the prediction of the DLR value for the next 15 min is also possible.

In the first part of the case study it was demonstrated how the parameters in the wind direction model can be tuned by MCMC method. While these simulations are promising, it is worth to mention that there are some applicability limits. If the wind speed is under 5 m/s or the territory does not have a dominant wind direction the level of the prediction accuracy decreases significantly. In the second part of the case study the major aim was to make a prediction for the DLR value in the next 15 min, while the operational safety factor maintains at 95%. As a result of the simulation, the transfer capacity become higher with 42 A than the 530 A static rating, which is almost 10% increase in the ampacity.

Generally, Monte Carlo methods could be linked to DLR calculation, but the models and simulations need to be fine-tuned in the future. Overall, it can be said that the results of the simulations are a good basis for further research.

Acknowledgement. This work has been developed in the High Voltage Laboratory of Budapest University of Technology and Economics within the boundaries of FLEXITRANSTORE project, which is an international project. FLEXITRANSTORE (An Integrated Platform for Increased FLEXIbility in smart TRANSmission grids with STORage Entities and large penetration of Renewable Energy Sources) aims to contribute to the evolution towards a pan-European transmission network with high flexibility and high interconnection levels.

EMBERI ERŐFORRÁSOK MINISZTÉRIUMA

Supported by the ÚNKP-18-2-I New National Excellence Program of the Ministry of Human Capacities.

References

1. McCall, J.C., Servatius, B.: Enhanced economic and operational advantages of next generation dynamic line rating systems, Paris (2016)
2. Electric Power Research Institute: Evaluation of instrumentation and dynamic thermal ratings for overhead lines (2013)
3. Rácz, L., Szabó, D., Németh, B., Göcsei, G.: Grid management technology for the integration of renewable energy sources into the transmission system. In: 2018 7th International Conference on Renewable Energy Research and Applications, Paris (2018)
4. Cigré Working Group B2.43: Guide for thermal rating calculations of overhead lines (2014)
5. Antal, P., et al.: Intelligent Data Analysis (in Hungarian), Typotex (2014)
6. van Ravenzwaaij, D., Cassey, P., Brown, S.D.: A simple introduction to Markov Chain Monte–Carlo sampling. Psychon. Bull. Rev. **25**(1), 143–154 (2018)
7. Zhang, J., Pu, J., McCalley, J.D., Stern, H., Gallus Jr., W.A.: A Bayesian approach for short-term transmission line thermal overload risk assessment. IEEE Trans. Power Deliv. **17**(3), 770–778 (2002)
8. Dong-Min, K., Jong-Man, C., Hyo-Sang, L., Hyun-Soo, J., Jin-O, K.: Prediction of dynamic line rating based on assessment risk by time series weather model. In: 9th International Conference on Probabilistic Methods Applied to Power Systems, KTH, Stockholm (2006)
9. Ringelband, T., Schäfer, P., Moser, A.: Probabilistic ampacity forecasting for overhead lines using weather forecast ensemble. Electr. Eng. **95**(2), 99–107 (2012)
10. Gubeljak, N., Banić, B., Lovrenčić, V., Kovač, M., Nikolovski, S.: Preventing transmission line damage caused by ice with smart on-line conductor monitoring. In: 2016 International Conference on Smart Systems and Technologies (SST) (2016)
11. Rácz, L., Rácz, D.A.: Application of Monte Carlo methods in Dynamic Line Rating models, Local Student's Scientific Conference – BME 2018, Budapest, Hungary (2018)

Theory of Constraints Thinking Processes on Operational Lean Programs Management Improvement: An Energy Producer Company Case

Margarida Gaspar[1], Luis Cristovão[2], and Alexandra Tenera[1,3(✉)]

[1] Faculty of Sciences and Technology (FCT), Universidade NOVA de Lisboa, Caparica, Portugal
ms.gaspar@campus.fct.unl.pt, abt@fct.unl.pt
[2] ToC Expert, Goldratt Consulting, Lisboa, Portugal
cristovl@hotmail.com
[3] Department of Industrial and Mechanical Engineering, UNIDEMI, Caparica, Portugal

Abstract. Several theories and associated models are arising in the field of systems' continuous improvement, focused on structured solutions to face the internal or external factors that affect them. Within these theories stands out Theory of Constraints (TOC), proposed to manage the most relevant constraints that exist within an organization. In this case study, the TOC Thinking Processes approach will be applied to the Management of Operational Lean Programs, on a pilot investigation at the organization were the Case Study took place. This application aims to analyze their Lean Program's Management in order to find the constraint that inhibits the system to reach its best level of performance, in order to support the development of robust improvement solutions that can solve the identified restrictions and sustain the proposed changes over time reaching a process of ongoing improvement.

Keywords: Theory of Constraints · Thinking Processes · Lean programs · Continuous improvement

1 Introduction

Nowadays, companies, whether on production or services fields, focus mostly on their levels of competitiveness on the global markets where they operate. So, they need to adopt strategies to guarantee their success amongst clients and stakeholders. The first step on choosing this strategy is to focus on the perception of the business processes flow [1]. On other hand, in order to step up the operational results, reduce costs and improve operations, it's usual to try implementing a culture of continuous improvement at the company as part of the business plan. One of these continuous improvement approaches is the Theory of Constraints that is considered by some as a management philosophy [2, 3] and by others as a methodology [4–6].

Published by Springer Nature Switzerland AG 2019
L. M. Camarinha-Matos et al. (Eds.): DoCEIS 2019, IFIP AICT 553, pp. 125–142, 2019.
https://doi.org/10.1007/978-3-030-17771-3_11

Theory of Constraints (TOC) was introduced to the scientific field on the 80's. Since then, it has been disseminated as a methodology or a philosophy to solve problems and improve systems' performance on many fields, by applying many structured TOC solutions. The less implemented solution of TOC is the Thinking Processes (TP), which can be verified by the published literature over the years. This solution was constructed to be applied mostly at complex systems, giving managers tools to think structurally about problems and how to solve them, providing sustainable solutions for organizations.

Thus, this research aims to give a contribution to the scientific field, providing a real Case Study developed at a top company, where the Thinking Processes tools were applied, showing their potential on an integrated application. This research effort intends to result on a valuable contribution to TOC-TP, providing increased results, since for the past 5 years, only few researches were published at the scientific level, as it will be exposed further. Notice also that on this Case Study, the application of TOC-TP tools didn't follow its traditional approach, having been adapted to the company's reality, as it is stated further on chapter four. This adaptation proves that Thinking Processes tools can be orchestrated to answer any company's needs, providing adapted solutions with the same level of quality as any other continuous improvement approach. Also, considering all the revisited literature, this is one of the few Case Studies were all the tools were applied, showing the full potential of TOC-TP, not being limited to the application of one or two tools as a complement of a major analysis.

At an organizational level, the Case Study contributes to the company's Operational Lean Programs Management Improvement, introducing a new methodology, new improvement tools and a systematic approach to examine problems, turning this into a pilot investigation at the organization. What is unique about this investigation, regarding to the organization's reality, is the attempt to improve management practices inside a continuous improvement lean program itself. By doing so, it is possible to question why the management of a lean program is not as lean as it should be. As stated by Taveira [7] on a previous evaluation of the system, the company where the Case Study took place, saw lean as a mean to reduce the expenses and wasn't considering other major benefits on their daily routine and on its implementation and, nowadays, the problem seems to persist. With this research we aim to give the company a systematic reflection approach to analyze an organizational system and come up with robust solutions to improve it, applying all known Goldratt's TOC-TP tools, which rarely are put into practice as a whole.

Therefore, the main research question brought to discussion is: how a lean organization can benefit from TOC, and how it can provide reasonable and logic solutions for the effective and efficient improvement of the Operational Lean Programs Management? In order to answer the question, this paper is divided into five sections. The present section provides general insights about the Case Study, the subject's main highlights and the major research contributions, the following section explains the relationship of the paper to industrial and service systems, followed by a literature review on the TOC-TP topic on section number three. The fourth section details and explains how the Case was conducted and developed. The last chapter provides the discussion and conclusions about the obtained results, ending with some suggestions of further work on the field.

2 Subject Relationship to Industrial and Service Systems

Planning is the core phase of a Plan-Do-Check-Act (PDCA) cycle, requiring a high level of effort. However, nowadays it is crucial to answer efficiently to the fast changes of global markets, and often organizations act before taking the suitable time to plan. By doing so, many companies fail on their plans and must rework their strategies, wasting resources and losing service quality. Thus, on industrial and service systems, when applying continuous improvement methodologies, such as Lean, Six Sigma or Theory of Constraints (TOC), it is fundamental to follow the inherent cycles of improvement, to assure the follow up of the planning as well as the fulfillment of the other phases of the cycles.

From the Theory of Constraints perspective, one of the most common applied tools among industrial and operations environments is the Drum-Buffer-Rope (DBR), which has been applied on many fields and types of industries. However, the same doesn't happen yet with the reflexive logical tool of TOC, the Thinking Processes, although it is recognized as the TOC tool with more improvement potential [8]. Its application to real industrial and service paradigms is not very wide and its potential as a continuous improvement methodology hasn't been fully explored.

The TOC Thinking Processes reflexive tools, when applied to an organization, imply systematic thinking following cause-effect logic and reflection mechanisms. By doing so, companies can track their organizational problems, think about solutions, construct future realities and come up with possible obstacles that they can discuss and solve before the implementation of the solution. Thus, the planning is all structured beforehand any operational action takes place, so organizations will be aware of which is the most effective plan they must stick to. This approach to planning will require a higher organizational interconnection of the players, to be supported on new levels of creative and intelligent business decision systems, to be able to produce less expensive, faster and valuable decisions for highly smart organizational decision making.

This Case Study tries to contribute to the expansion of Theory of Constraints Thinking Processes (TOC-TP) as a strong player on its methodological field, giving a full and updated analysis of how TOC-TP is applied nowadays and how it can be managed on a real and practical application on a service system. By applying TOC-TP to an energy producer company, that has many years of experience on lean programs, it is intended to show how an organization can benefit from TOC, and how it can provide reasonable and logic solutions for problems that tend to persist over the years, elevating its performance on lean applications.

3 Theory of Constraints Thinking Processes (TOC-TP)

3.1 Theory of Constraints and Thinking Processes Review

In 1984, Goldratt published "The Goal" [9], a novel that introduces TOC to the management field, based on the story of Alex Rogo, his production plant and how he saves it from disaster with the help of his mentor Jonah. The basis of TOC goes by the premise that every system has at least one active constraint, which means, a systemic problem. On TOC way, constraints are seen as positive things, representing an opportunity to improve the given system. As the constraint sets the performance of the system, an improvement of it will improve the system's results.

In 1994, Goldratt also published "It's Not Luck", another novel which introduces the Thinking Processes and associated tools, trying to provide a systematic way to address the identification and resolution of problems concerning management policies [10]. According to Cox and Robinson [11], Goldratt proposed Thinking Processes as a way to apply the scientific method to the resolution of business problems, seeking to identify and analyze the inherent simplicity of a system through cause-effect logic. So, the use of the Thinking Processes tools should contribute to achieve system's improvements, to search for a graphical perception of its logical relations and to adopt a simple method to identify, analyze, understand and communicate problems, as well as develop logical solutions.

The identification of the constraint is very important, particularly when it is a nonphysical/intangible one because physical constraints are more tangible and much easier to identify and address. Also notice that many times just by changing relevant policies or performance measures, the existing main constraints can disappear, which results on exponential benefits towards the system's goal [12]. So, for this to happen, and to ensure a significant state of improvement on a system, we should answer three Goldratt's basic improvement questions, i.e. the TOC fundamental questions [13, 14]: What to change? What to change to? How to change?, which later were expanded by Cox et al. [15] to the current five: Why to change? What to change? What to change to? How to change? How to measure and sustain the change? So, while answering these questions, TOC's continuous improvement actions should be developed according to the identified main constraint of the system [12], thought a prescribed process of continuous improvement (POOGI), consisting in Five Focusing Steps (5FS): (1) identify the system constraint, (2) decide how to exploit the constraint, (3) subordinate the rest of the system to the identified constraint, (4) elevate the constraint and (5) return to step one and don't allow inertia to become the constraint [13]. As the three basic improvement questions evolved, on the TOC's 5FS were added two pre-requisites processes to implement improvement actions, as it will be considered at the present research. So, according to Pass and Ronen [5], before the identification of the constraints we should also: (1) Stablish what is the organization's goal and (2) Define global performance metrics.

3.2 Overview of Thinking Processes' (TP) Recent Literature

The TOC Thinking Processes unlikely other TOC tools, as DBR, isn't so deeply developed in terms of practical and real situations. As referred at section one, to check the scientific current research level of the TOC-TP approach, a bibliography study was conducted, using research engines such as B-ON and Scopus. This research was developed considering the periods of time between 2008–2013 and 2014–2019 and defining a set of terms to conduct the investigation, using quotation marks to restrict joint terms such as "Thinking Processes". Also, to identify the scientific works directly related to the main subject (TOC) and the Goldratt term was also added to each research term. On Table 1, the results are identified on the research terms by "terms" and "Goldratt" search results. Also, in the scope of this search all sources and all types of information available were included, regarding to keywords and titles of conferences, papers, academic and scientific journals, reports and books.

Table 1. Theory of Constraints publications of the last seven and five years

Research engine	Research term	Number of publications			
		2008–2013		2014–2019	
		By keywords	By title	By keywords	By title
B-ON	Theory of Constraints/*Goldratt	2735/*141	160/*47	2453/*47	170/*72
	Theory of Constraints Case Study/*Goldratt	2/*0	6/*4	1/*0	8/*2
	Theory of Constraints Thinking Processes/*Goldratt	1/*0	2/*2	0/*0	0/*0
	Thinking Processes Case Study/*Goldratt	0/*0	2/*1	0/*0	3/*0
	Thinking Processes/*Goldratt	3/*0	32/*3	3/*0	28/*1
	Thinking Processes Tools/*Goldratt	1/*0	12/*3	0/*0	8/*1
Scopus	Theory of Constraints/*Goldratt	310/*205	201/*86	253/*184	95/*74
	Theory of Constraints Case Study/*Goldratt	2/*2	5/*5	4/*3	10/*8
	Theory of Constraints Thinking Processes/*Goldratt	17/*15	5/*5	10/*9	4/*4
	Thinking Processes Case Study/*Goldratt	2/*1	3/*2	4/*3	3/*1
	Thinking Processes/*Goldratt	174/*18	53/*10	124/*14	75/*10
	Thinking Processes Tools/*Goldratt	36/*11	13/*7	19/*3	11/*3

Analyzing Table 1, regarding to B-ON search, it's possible to verify that, generically, Theory of Constraints has been more explored on recent years, unlikely Thinking Processes, which numbers decreased over the past few years. Regarding Case Studies, they have been more case publications between 2014 and 2019 than between 2008–2013, but when adding the term "Goldratt" the number of available publications decreased to less than half on most cases. The same seems to happen with Thinking

Processes Case Studies. Overall, although with few results were found, the research conducted by B-ON shows some consistency on the number of publications related to Theory of Constraints and Thinking Processes over the past 10 years, even though there's no results on some cases, according to the researched terms.

Regarding to the investigation conducted in the research engine Scopus, it is possible to find that numbers aren't much different from the scenario provided by B-ON, although Scopus provides a bigger amount of publications on every aspect (possibly because of research options provided by both engines). Generically, the number of publications also increased over the recent years as on the case of Case Studies publications. Although TOC publications numbers are still low, comparing to other continuous improvement methodologies, as Lean or Six Sigma. Concerning the identified Case Studies on Thinking Processes we could highlight the "Revolutionizing blood bank inventory management using the TOC thinking process: An Indian case study" (2017) by Lowalekar and Ravi [16], "The thinking process of the theory of constraints applied to public healthcare" (2019) by Bauer, Vargas, Sellitto, Souza and Vaccaro [17], "A process improvement approach based on the value stream mapping and the theory of constraints thinking process" (2014) by Librelato, Lacerda, Rodrigues, Veit [18].

Although Scopus provides a higher number of publications on most cases and having in consideration the cited publications, the reduced research on TOC-TP remains, showing the need to increase research and applications on this subject and on TOC area in general.

3.3 Overview of Thinking Processes Tools

Goldratt developed a set of Thinking Processes tools to put TOC-TP on practice. The following tools were developed in a format of logic trees and a cloud: Current Reality Tree (CRT), Evaporating Cloud (EC), Future Reality Tree (FRT), Negative Branch Reservation (NBR), Prerequisite Tree (PRT) and Transition Tree (TT). Later, Dettmer [19] also added the Goal Tree (GT) and Goldratt as well added, the Strategic and Tactic Tree (S&T), which is today the last contribute for the TP tools and it is also considered the last tool of the sequence that should be implemented. Figure 1 represents the existent connection between the five questions, the inherent application of the five steps and the TP tools and how they all work together to establish a full practical investigation on the Theory of Constrains Thinking Processes.

Concerning to the roles of TP tools, the Goal Tree answers the first question "Why to change?", identifying the organizational model proposed to achieve the goal. Before a company can focus, it is necessary to have in mind what is the goal. Once the answer gets clear, the Necessary Conditions (NC) and Critical Success Factors (CSF) to achieve the goal are defined to identify which actions will be used to manage the progress. For the second question, CRT answers the question "What to change?", which allows to diagnose what, in the system, needs to be changed [4], which is a logical structure designed to illustrate the actual state of the system, isolating what needs to change, identifying the major problems (the undesirable effects, UDE) and draw the path to the root causes and the core problem.

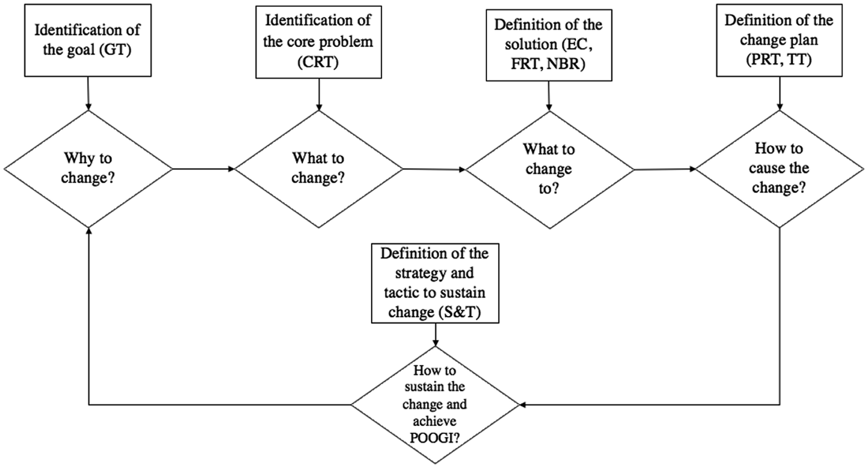


Fig. 1. Framework of the TOC-TP tools application

Once the CRT is fully designed, the constraint emerges and the best way to manage it is through commitment. So, in order to solve the core problem, TOC defines that behind every problem there is a conflict that blocks any robust solution. In order to evaporate this conflict and get plausible solutions (the injections) Goldratt suggests the EC as the perfect TP tool to serve this purpose, answering the question "What to change to?" [4].

The fourth tool of the TP set is the Future Reality Tree (FRT) which allows to construct a solution that, when implemented, replaces the Undesirable Effects (UDE) that are the basis for the CRT, by Desirable Effect (DE) in the FRT. The FRT is designed to ensure that all UDE will be eliminated by the solutions founded at the Evaporating Cloud (EC), providing a valid alternative for the future. However, when trying to get buy-in of the system's owners it's crucial to make sure that there is no potential resistance related with the proposed solutions. This type of thought is considered when referring to the Negative Branch Resolution (NBR), which determines if there are negative branches on the new solutions. By trimming these branches with new injections, should be possible to come out with a more robust solution for the system we are trying to improve.

The next TP tool is the Prerequisites Tree (PRT) that aims to answer the question "How to cause the change?" by identifying the obstacles that prevent the chosen injection(s) from being implemented [4]. The transition Tree (TT) is also used to give specific actions to the injections implementations i.e. combining each provided effect successively with the subsequent specific actions to produce new improved effects, contributing to answer the same question as a PRT. According to Dettmer [19, 20], the PRT can be combined with the TT. Together they should be capable to provide a more powerful and robust tool to structure the injections that are going to be implemented on the future reality of the system, having into account the possible obstacles on the way and eliminating them.

The Strategic and Tactic tree (S&T) is used to identify and communicate the identified strategies and tactics to ensure Management's attention at all levels to be synchronized and focused on the highest priority changes. This last tool has been specially designed to help Top Management prevent common mistakes that result in failures of change initiatives within organizations. It is possible to verify that for each strategy (S) there must be a tactic (T), so an S&T tree consists of a number of pairs S and T, presented at various levels, each level represents one layer of responsibility within the organization [14].

4 Applying TOC-TP to Operational Lean Programs Management

4.1 Background to the Case Study

According to Saunders [21], the strategy to develop this investigation was defined as a Case Study, suiting the definition given by Robson [22]. This author defines a Case Study as a strategy for doing research which involves an empirical investigation of a particular contemporary phenomenon within its real-life context using multiple sources of evidence.

Thus, the company where the Case Study took place is part of a Group which the main business is the production, distribution and commercialization of energy. This company in specific is responsible to manage the production of thermic and hydric energy, on Production Centers across Portugal. Being decentralized, with more than 40 production points, it is difficult to manage all the operations. To avoid resources' waste (money, time, quality) and to be as more efficient as possible, in 2006, the company launched the Lean Program. First, this Program was a pilot on a chosen Center and then it was replicated to many others. Nowadays the Local Programs are at the sustention phase, being often monitored and evaluated, answering to the KPI that are suggested by the managers.

To evaluate the system "as-is" was necessary to interview the company staff based on a design thinking approach and on brainstorming. This evaluation, which details are referred at Table 2, was carried out by the company on a lean event. The evaluation was based on the steps of a lean initiative life cycle to understand both positive and negative opinions of the lean practitioners at each stage, also considering all the geographies where exists a Lean Program. Based on this empirical data, the first analysis to the system was made, based on the given negative feedback in order to analyze, at a first stage, the main problems of the system.

The data collected comes from various sources, which were combined to analyze the system, in order to get a fully understanding of it. They may include, for example, interviews, observation, documentary analysis and questionnaires [21].

Using tools such as Pareto Diagram and Ishikawa Diagram, it was possible to scrutinize a large amount of data. These tools allowed to group the main issues identified by the staff, in categories such as, Difficulty and Resistance to apply Lean, Unavailability of staff, Communication problems, which are the main concerns of the company. Later, these topics where addressed on the application of TOC-TP.

Table 2. Details of the Lean event

Lean day	
Date	November 2017
Duration	One day
People	Approx. 10 teams of 10 company employees (different roles)
Roles	Multiskilled employees (managers, technicians, engineers, operational collaborators)
Purpose	Evaluate the status of Operational Lean Programs, based on staff opinions
Goal	Improve the management of Operational Lean Programs and the satisfaction of the staff with the Lean Program
Main approach	Design Thinking and Brainstorming
Type of data collected	Empirical, collected by semi structured interviews

The following subsections show some diagrams that represent the trees and clouds developed for his problem as well as its explanation.

4.2 Why to Change? The Application of the GT

After the empirical analysis done to the data provided by the company, the main goal of the investigation was set to start the development of the TOC-TP logical tools. This is also considered the first step to develop the first Thinking Processes tool, the Goal Tree. Thus, with the agreement of the company, the main goal of the Case Study was defined as to identify the main actions to follow "to improve the effectiveness and efficiency of the Operational Lean Programs Management (OLPM)". In fact, the Goal Tree shows the main actions to have into account when trying to reach the proposed goal. The many inputs given by the company were crucial to identify which were the Critical Success Factors (CSF) and the Necessary Conditions (NC) that needed to exist to achieve the Goal. These factors and necessary conditions are defined to represent how the company must perform if the goal was achieved (although it hadn't been, due to the undesirable effects that create the gap between the reality and the desired state). The CFS are the fundamental aspects that are considered crucial to achieve the proposed Goal and the NC are the conditions that must exist to enable the Critical Success Factors. The organization's staff had an active role helping to construct the Goal Tree, as their opinions, feelings and inputs were collected from a brainstorming session. By actively construct the Goal Tree with the company, this tree can provide an accurate answer to the question to "Why to change?", giving the perception that the problem really exists and there's a need and a reason to create a positive change on the system. As the desired reality was constructed, the same session was used to inquire the participants about what is preventing the organization from achieving the Goal, constructing a list of 18 UDE. This list confirmed the main issues revealed on the first analysis to the system and those inputs were the basis of the following step: What to Change? That is explained in detail on the following section.

4.3 What to Change? The Application of the CRT

The reverse line of thought was taken into account to think about the undesirable effects (UDE) that prevent the system from achieving the desired state. Between the eighteen UDE identified on the system, the CRT (Fig. 2) shows the situation as perceived at the time, answering to the question “what to change?”. This diagram was constructed, from top-down and its read bottom-up, linking all the intermediate effects to the UDE, using inputs like the data analysis and team brainstorming, to develop logical cause-effect connections. At CRT, these cause-effect connections use “IF… THEN” statements, for example IF “Lean Programs aren’t included in the QMS” OR “Improvement is seen as a time consumer” THEN “Improvement isn’t considered as a management system”.

Analyzing the CRT is possible to state that the intermediate effects that connect UDE are numbered above the boxes and UDE are identified by their numbers (e.g. UDE 1.2). Therefore, the entities “A”, “B”, “C” and “L” represent long connections between entities. For example, entity “A” means that UDE 5.2 is connected to the UDE 1.1 by the ellipse (which represent an “AND”). However, “L” means that there is a negative reinforcing loop at the tree, meaning that when connecting those effects, repeated negative consequences occur at a faster pace.

From Fig. 2 it is also possible to see that the main problems on the system are related to “Training”, “Means to report lean initiatives” and “Alignment of metrics” factors, which are the identified Critical Root Causes (CRC). These CRC can be identified by spotting the UDE that have no entry connections (UDE 4.1) and by counting which UDE connects with most of other entities at the tree, for example, UDE 5.2 and UDE 5.3. These CRC are the problems that mainly contribute to the existence of core problem identified at the CRT. Therefore, these CRC must be scrutinized to be evaporated by EC. Not surprisingly, the major problem (core problem) that worried all participants including managers was, still, the “Lack of Commitment with Lean and Lean’s culture”, which is consistent with the problem found by Taveira [7]. This entity is represented at the bottom box of the CRT (box 55), has no entry connections and leads to most of the UDE (directly or indirectly) from bottom to top until reaching the Goal (the upper entity).

By developing the CRT, it is possible to conclude that a major and transversal problem has been affecting the company and the management of Lean Programs over the years and there wasn’t no solution at that time. How can an organization be Lean and have a healthy Lean Program if it isn’t felt on their culture and on a daily basis? How can a long term Lean Program have such structural problems as training, reporting and metrics? These paradigms were taken into consideration from the application of the CRT until the end of the research to suggest ways to solve it. The identified core problem becomes the focus of the cloud EC, revealing the dilemmas.

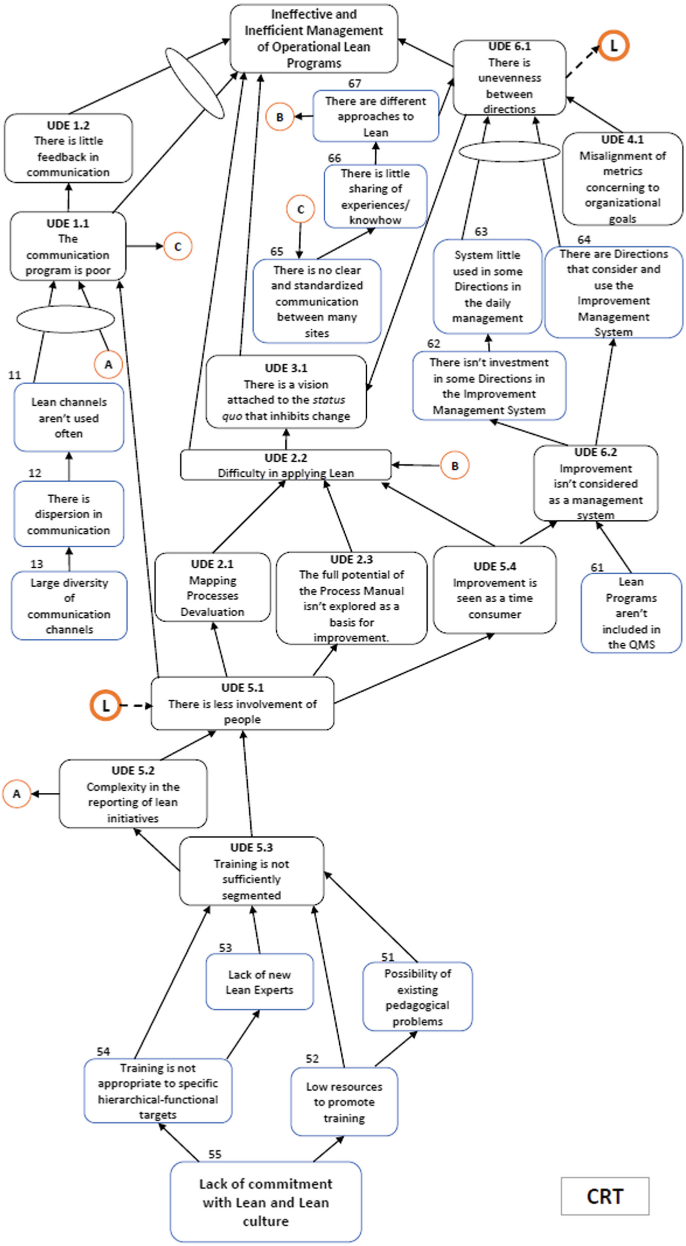


Fig. 2. Current Reality Tree

4.4 What to Change to? The Application of EC, NBR and FRT

After the analysis of the CRT we are now able to study more deeply each of the main problems identified on the system applying the EC, NBR and FRT. The combination of these tools answers the third question "What to Change to?", giving a possible reality of the system after the change. Also, on this phase it is possible to come up with suitable new injections to mitigate the systemic problems.

To apply the Evaporating Cloud and to clearly point the dilemmas or conflicts regarding to the TOC-TP language, four clouds were developed, one of them to the core problem and the others to each of the identified CRC. Figure 3 represents the relationship between all the developed clouds, which focus is identical and represents the proposed Goal exposed at the first diagram. So, it is possible to state that entity A is the same for every cloud, representing the "Effective and efficient management of Operational Lean Programs".

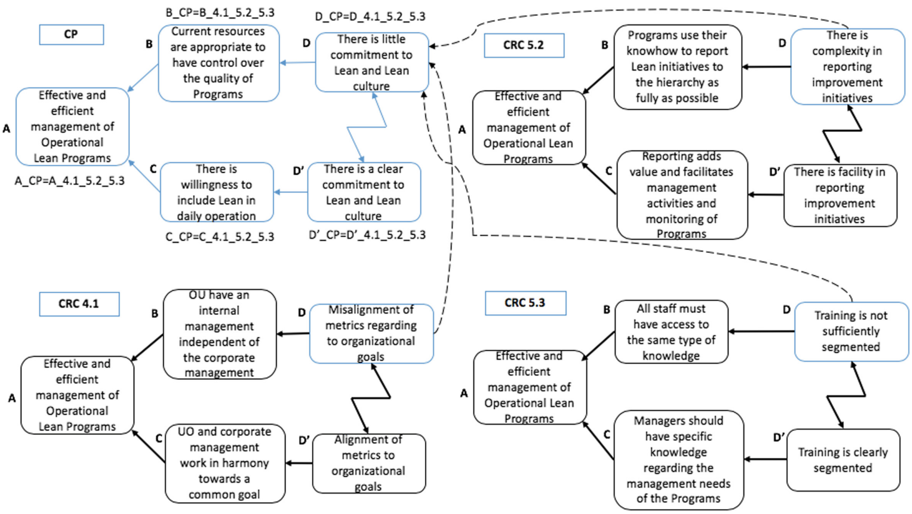


Fig. 3. Evaporating Clouds (EC)

As constructing each cloud, we can identify the weakest necessity cause linkage in the EC by creating assumptions that would be able to break it. These assumptions are created by necessity logical relations as "In order to have A, we need B". When having some assumptions for a given link it is necessary to find one statement weak enough to be broken by an injection that will dissolute the conflict. So, after the selection of the link to break, possible injections are created to evaporate the conflict and the cloud.

The "CP" cloud at upper left corner of Fig. 3 represents the EC for the core problem. Its dilemma can be interpreted as: to have an "effective and efficient management of Operational Lean Programs" we need "willingness to include Lean in daily operation".

In order to have a "willingness to include Lean in daily operation" we need to obtain "a clear commitment with Lean and Lean Culture". On the other hand, for an "effective and efficient management of Operational Lean Programs" we need "the current resources to be appropriate to have control over the quality of Programs". And for this should be "little commitment to Lean and Lean culture" (since the current resources aren't considered enough). This is in direct conflict with the lower branch, which is represented by an elbow double arrow. Hence, we can't have much and low commitment with Lean at the same time.

At this Case Study, a total of 14 injections were brought to discussion with the representative organization manager to validate the findings. This process was repeated for every cloud and all the injections were analyzed with the support of an Efficiency/Difficulty matrix elaborated for the purpose, which was crucial to study the impact of each injection on the system. By this analysis, two injections were declined, four injections were approved by the organization to mitigate de training problem, four injections for the metrics issue and tree injections were proposed for the reporting. Although these injections will have an indirect effect on the core problem, contributing to its mitigation, one more injection was approved to directly mitigate the core problem.

Even though we are analyzing the EC, by chance, this application converged to what is called the Three Clouds Method, that is an alternative approach to the conventional application of the CRT [23]. At this investigation we chose to apply the conventional method of the CRT as we consider that is the most robust solution for the purpose. The advantage of having the architecture of the three-cloud method applied to the EC, as is represented at Fig. 3, is that we can study the system as a whole by linking the entities on the clouds to each other's by their common points, as is shown by the arrows that converge from entity "D" to the entity "D" at the "CP" cloud.

The analysis of the EC concludes, as expected from the previous feedback, that the system needs a relevant improvement on what concerns to Lean commitment and cultural change. This change should be wider and deeper than just to mitigate the identified problems to align visions, practices, and to work in harmoniously on a Lean organization. However, we expect the proposed injections to help substantially on this matter.

After the final selection of the eligible injections to be implemented, the organization's manager selected the more volatile injection being "Linking programs and goals that translate into financial rewards/other benefits". This injection was scrutinized using the NBR as this tool that can induce negative side effects to the system. Next, NBR was trimmed by the division of the volatile injection into two "Use only non-financial and career progression benefits" and "Use other company benefits", that placed strategically at the NBR, could invalidate the negative effects, preventing the system from new UDE and the change, possibly, from failing.

The last step to finish answering the basic question "What to change to?" is the application of the Future Reality Tree. This tree comes up through the positive effects generated on consequence of the injections proposed at the EC, showing what would change if the system improved its effectiveness and efficiency on the management of Operational Lean Programs. Usually, these effects are the opposite of the negative effects identified at the CRT. So, the FRT shows, by drawing the logical paths, how the injections implemented on the system can create positive effects that will converge on the goal's achievement.

4.5 How to Change? The Application of the PRT+TT

After the definition of the future reality it is necessary to induce the change by the application of the PRT and the TT, answering the question “How to change?”. These two trees will be applied as a single tool, as justified by Dettmer [19], since the extension and complexity of the Case Study doesn’t justify its independent utilization. Ideally, this tree would be constructed for each injection created previously, to analyze the details of its implementation on the system. However, on this paper we give only one example to illustrate how an injection can be explored by TOC-TP. Later, it is expected the company to construct the rest of the PRT+TT diagrams, to plan the implementation of the improvement actions.

Thus, a list was created to specify the intermediate objectives (IO) on the implementation of the injection. These intermediate objectives are the steps to follow to achieve the execution of the injection. Reading the diagram bottom-up, the IO are represented by boxes with curved edges, on a necessary condition thinking: “In order to have A we need B”. For example, on Fig. 4, “In order to analyze comparatively Lean tools with company’s Lean tools we need to detail Lean tools according to the usage criteria”. Then, another list was created to identify the obstacles regarding to the implementation of the chosen injection. These obstacles are assigned to the links that may have associated negative effects and can be interpreted as obstacles that prevent the system from the correct execution of the action. After, the PRT was created by linking the obstacles with the corresponding IO, as shown at Fig. 4.

But, as stated before, this isn’t the final stage of the tree, the next step is to trim those obstacles with specific actions (SA) to finish the implementation plan. These actions are necessary in order to prevent the improvement to fail and to avoid inertia. For example, to prevent the Obstacle 1 (O1) “Unawareness of the detail of the performance of functions” from happening the Specific Action 1 (SA1) “Contact the company and clarify any doubts that may exist” must be executed. Regarding the structure of the PRT+TT (see Fig. 4), a plan of implementation for this specific injection was defined for the organization to follow. It is expected the organization to implement the rest of the injections taking into account the example given on the research and in order to promote a sustainable change, closing the first cycle of change by TOC-TP.

4.6 How to Measure and Sustain the Change?

To answer this question, at first, we would have to build a S&T tree. But, considering the lack of enough knowledge by the research team about companies grounding on this subject to develop this tree, the research team decided not to pursuit with its construction. The S&T tree was idealized to be implemented by managers, on project management subjects, as a tool to help them on the definition of the organizational strategy, which action is outside of the main scope of the present study. Nevertheless, S&T should be implemented on projects and should be orchestrated by top management as a macro tool to manage the main lines of a bigger strategy.

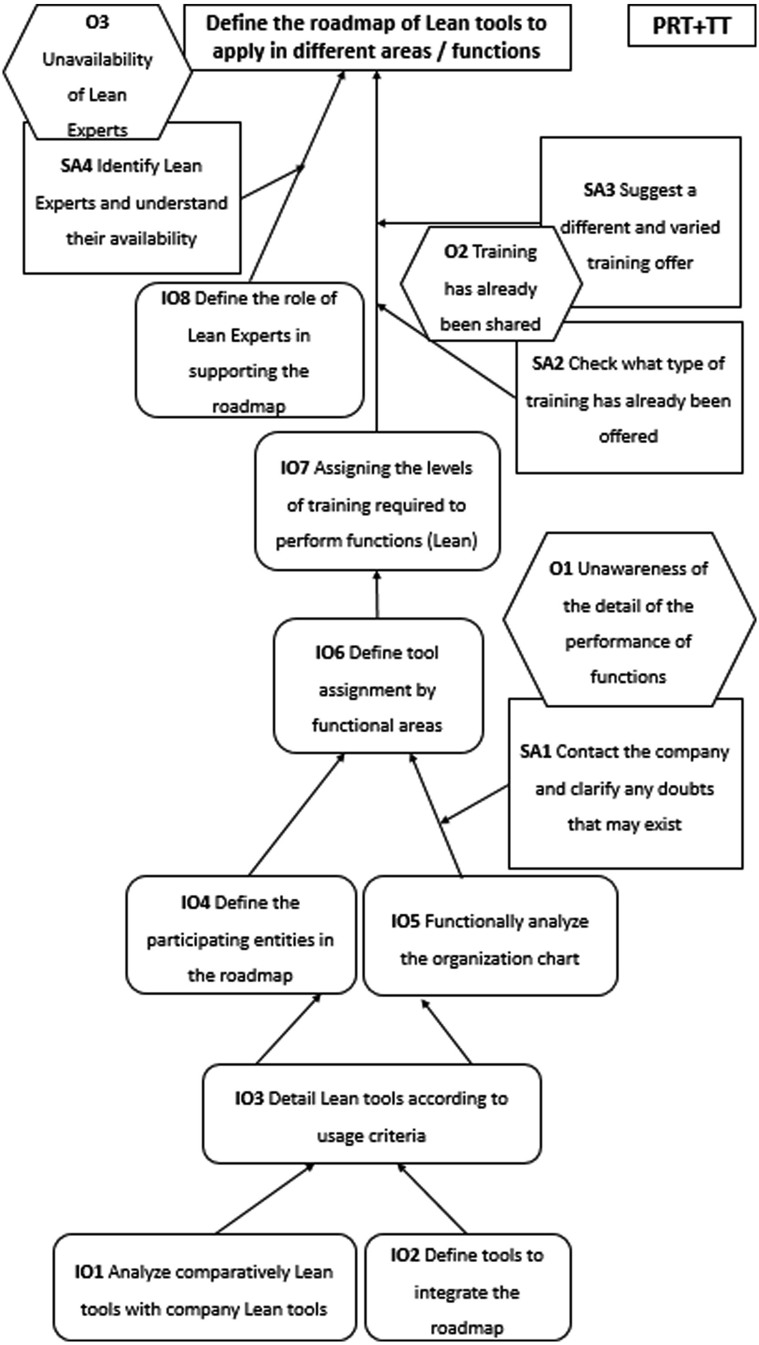

PRT+TT
Define the roadmap of Lean tools to apply in different areas / functions
O3 Unavailability of Lean Experts
SA4 Identify Lean Experts and understand their availability
IO8 Define the role of Lean Experts in supporting the roadmap
SA3 Suggest a different and varied training offer
O2 Training has already been shared
SA2 Check what type of training has already been offered
IO7 Assigning the levels of training required to perform functions (Lean)
O1 Unawareness of the detail of the performance of functions
IO6 Define tool assignment by functional areas
SA1 Contact the company and clarify any doubts that may exist
IO4 Define the participating entities in the roadmap
IO5 Functionally analyze the organization chart
IO3 Detail Lean tools according to usage criteria
IO1 Analyze comparatively Lean tools with company Lean tools
IO2 Define tools to integrate the roadmap

Fig. 4. PRT+TT

5 Main Conclusions and Further Work

This research allowed to consider solutions for an organizational problem that persists over the past years on a lean based company. The problem, lack of enough commitment with Lean and Lean's culture, is a long-term issue at the organization that is felt at the *modus operandi* of staff and at the organizational culture. In order to solve this problem, twelve injections were approved by the organization to put in practice. As an example, one of them was structured and prepared for the organization to implement.

The research allowed to implement GT to find the goal, the CRT to find the core problem, the EC to come up with injections, the NBR plus the FRT to structure and prepare the future reality, and the PRT+TT to give a figure of what one of the implementations could be. Given the full implementation of TOC-TP, the organization has now a systematic ground to promote the resolution of major problem that affects the Management of their Lean Programs. However, this Case Study doesn't validate the Theory and its suppositions, it just represents a contribution to the empirical application. Although the systematic method to apply Thinking Processes is concrete and has strict rules, its application is empirical and mostly qualitative, which produces qualitative results, that can have many interpretations. The lack of investigation on the field, doesn't provide yet a solid ground of exploration, giving few examples of application to better analyze systemic problems.

It is considered that the research question was mainly answered on its purpose as the research allowed to put into practice a full framework of basic TOC-TP tools, giving a sense of full accomplishment of the process of ongoing improvement, by using cause-effect logic and reflecting about what could be improved on a given system. However, the present investigation stills lacks on giving proof and analysis of the implementation of the injections on the company, measuring its results and, perhaps, starting another cycle of improvement to find if the problem sustains. Another limitation of this research approach is that the case study results can't be extrapolated to other realities since every company has a different reality. Due to the type of data collected and to the exposed problem on the Case Study, tools such as S&T tree, TOC's Throughput Accounting and the verification of KPI weren't considered on this case study report, which could be developed in order to assure that, company could continually monitor the change and the solutions implementation impacts. Also, it is recommended to continually measure the satisfaction indexes of the staff regarding to lean, especially after the implementation of the improvement actions.

On this Case, TOC was used to provide a framework for the change needed on the system, regarding to the identified core problem and to provide a clear logic-based argument for the changes proposed, which are essential to ensure the sustainability of the Lean Programs. Also, this Case Study contributed with an innovative problem-solving method for the company and to increase TOC-TP exploration on the scientific field, contributing to encourage pairs to further explorations of TOC-TP. This research also adds to verify the potential and robustness of this method on smart organizational improvement, by giving unique and maybe out of the box solutions to organizations which require new levels of creative and intelligent business decision tools in order to be able to produce less expensive, faster and valuable decision i.e. highly smart organizational decision making.

Acknowledgements. This work has been partially supported by UNIDEMI - Research and Development Unit for Mechanical and Industrial Engineering by the Portuguese FCT-PEST program UID/EMS/00667/2019. The authors are grateful to the supporter research company's collaborators which the researcher had the pleasure to work during the development of this research work.

References

1. Şimşit, Z.T., Günay, N.S., Vayvay, Ö.: Theory of constraints: a literature review. Proc. Soc. Behav. Sci. **150**, 930–936 (2014)
2. Gupta, M., Boyd, L.: Theory of constraints: a theory for operations management. Int. J. Oper. Prod. Manag. **28**(10), 991–1012 (2008)
3. Šukalová, V., Ceniga, P.: Application of the theory of constraints instrument in the enterprise distribution system. Proc. Econ. Financ. **23**, 134–139 (2015)
4. Mabin, V.: Goldratt's "Theory of Constraints" thinking processes: a systems methodology linking soft with hard. In: History, pp. 1–12 (1999)
5. Pass, S., Ronen, B.: Management by the market constraint. Int. J. Prod. Res. **41**(4), 713–724 (2003)
6. Taylor III, L., Rekha, A.: Applying Theory of Constraints principles and Goldratt's Thinking Process to the problems associated with inventory control. Franklin Bus. Law J. **2016**(4), 83–104 (2016)
7. Taveira, A.: Avaliação da Sustentação da Metodologia Lean numa Organização: Caso de Estudo na EDP Produção, Universidade Nova de Lisboa (2015)
8. Goldratt, E.M., Cox, J.: The Goal: A Process of Ongoing Improvement, 1st edn. North River Press, Great Barrington (1984)
9. Chaudhari, C., Mukhopadhyay, S.: Application of Theory of Constraints in an integrated poultry industry. Int. J. Prod. Res. **41**(4), 799–817 (2003)
10. Cox III, J.F., Schleier Jr., J.G. (eds.): Theory of Constraints Handbook. McGraw-Hill, New York (2010)
11. Cox III, J., Boyd, L., Sullivan, T., Reid, R., Cartier, B.: TOCICO Dictionary, pp. 1–135 (2012)
12. Tenera, A.: Contribuição para a melhoria da gestão da incerteza na duração dos projectos através da teoria das restrições, Universidade Nova de Lisboa (2006)
13. Goldratt, E.M., Cox, J.: The Goal: A Process of Ongoing Improvement, 3rd edn. The North River Press Publishing Corporation, Great Barrington (2004)
14. Schragenheim, E., Dettmer, H.W.: Manufacturing at Warp Speed: Optimizing Supply Chain Financial Performance, 1st edn. CRC Press, Boca Raton (2000)
15. Cox III, J., Robinson, E.: Applying Goldratt's thinking processes to prevent mistakes. Hum. Syst. Manag. **36**(4), 315–340 (2017)
16. Lowalekar, H., Ravi, R.: Revolutionizing blood bank inventory management using the TOC thinking process: an Indian case study. Int. J. Prod. Econ. **186**, 89–122 (2017)
17. Bauer, J., Vargas, A., Sellitto, M., Souza, M., Vaccaro, G.: The thinking process of the theory of constraints applied to public healthcare. Bus. Process Manag. J. (2019). https://doi.org/10.1108/BPMJ-06-2016-0118
18. Librelato, T., Lacerda, D., Rodrigues, L., Veit, D.: A process improvement approach based on the value stream mapping and the theory of constraints thinking process. Bus. Process Manag. J. **20**(6), 922–949 (2014)

19. Dettmer, H.W.: The Logical Thinking Process a Systems Approach to Complex Problem Solving. ASQ Quality Press, Milwaukee (2007)
20. Dettmer, H.W.: Goldratt's Theory of Constraints: A Systems Approach to Continuous Improvement. ASQC Quality Press, Milwaukee (1997)
21. Saunders, M., Lewis, P., Thornhill, A.: Research Methods for Business Students, 5th edn. Pearson Education, London (2009)
22. Robson, C.: Real World Research: A Resource for Social Scientists and Practitioner-Researchers, 2nd edn. Blackwell, Oxford (2002)
23. Burton-Houle, T.: Field Guide to the Theory of Constraints Thinking Process. Goldratt Institute, New Haven (2000)

Dynamic Search Tree Growth Algorithm for Global Optimization

Ivana Strumberger[1](✉), Eva Tuba[1], Miodrag Zivkovic[1], Nebojsa Bacanin[1], Marko Beko[2,3], and Milan Tuba[1]

[1] Faculty of Informatics and Computing, Singidunum University, Danijelova 32, 11000 Belgrade, Serbia
{istrumberger, etuba, mzivkovic, nbacanin}@singidunum.ac.rs, tuba@ieee.org
[2] COPELABS/ULHT, Lisbon, Portugal
[3] UNINOVA, Caparica, Portugal
mbeko@uninova.pt

Abstract. This paper presents dynamic version of the tree growth algorithm. Tree growth algorithm is a novel optimization approach that belongs to the group of swarm intelligence metaheuristics. Only few papers addressed this method so far. This algorithm simulates the competition between the trees for resources such as food and light. The dynamic version of the tree growth algorithm introduces dynamical adjustment of exploitation and exploration search parameters. The efficiency and robustness of the proposed method were tested on a well-known set of standard global unconstrained benchmarks. Besides numerical results obtained by dynamic tree growth algorithm, in the experimental part of this paper, we have also shown comparative analysis with the original tree growth algorithm, as well as comparison with other methods, which were tested on the same benchmark set. Since many problems from the domains of industrial and service systems can be modeled as global optimization tasks, dynamic tree growth algorithm shows great potential in this area and can be further adapted for tackling many real-world unconstrained and constrained optimization challenges.

Keywords: Tree growth algorithm · Swarm intelligence · Global · Unconstrained · Metaheuristics · Dynamically adjusted parameters

1 Introduction

1.1 Motivation

In the todays era of digitalization, Internet of Things (IoT) and Cloud Computing, the industry and service sectors are transforming by introducing many innovations, where the most emphasized features of these innovations are so-called "exponential technologies". This transformation is usually being expressed with the terms Industry 4.0, Smart Manufacturing and Economy 4.0.

In order to implement and to establish innovations in industry and service sectors, many problems need to be addressed. Most of such problems can be mathematically

Published by Springer Nature Switzerland AG 2019
L. M. Camarinha-Matos et al. (Eds.): DoCEIS 2019, IFIP AICT 553, pp. 143–153, 2019.
https://doi.org/10.1007/978-3-030-17771-3_12

modelled and can be categorized as NP hard problems. For example, in the domains of Industry 4.0, cloud computing and IoT, many scheduling problems exist, that are characterized with NP hardness [1]. Also, the implementation of the wireless sensor networks (WSN) is very important in conducting innovations of industry and service sectors. Many problems, such is the problem of localization, from this domain belong to the NP hard group [2].

Another area which belongs to the group of NP hard problems is data clustering, along with data mining. Modern industry often relies on large amount of data. Large datasets are stored on the clusters, and focus of the algorithms is partitioning any dataset into an optimal number of groups through one run of optimization [3]. Modern industry also heavily depends on telecommunication networks. In order to keep the cost of the network reasonable, it is not possible to directly connect all communication nodes. Routing process is responsible for selection of the path between source and destination, by optimizing the objectives together with constraints. Any routing problem has the main objective to maximize the network performance [4].

For solving NP hard problems, it is not enough to use classic deterministic algorithms, since the results cannot be generated in a reasonable amount of time. For such purposes, it is better to employ stochastic methods like metaheuristics. Metaheuristic algorithms can be roughly divided into those who are inspired by the nature, and those who are not nature inspired. Nature-inspired metaheuristics can be further divided into evolutionary algorithms (EA) and swarm intelligence (SI).

Swarm intelligence is population-based, stochastic and iterative search methods that try to improve the population of candidate solutions in a predetermined number of iterations. These algorithms simulate the groups of natural organisms such as flock of birds and fish, groups of fireflies and bats, etc. One of the first swarm intelligence approaches was particle swarm optimization (PSO) that was developed in 1990s [5]. After the PSO, many other swarm intelligence algorithms emerged that were adapted for solving benchmark, as well as real-world NP hard optimization tasks. Some of the examples include artificial bee colony (ABC) [6], firefly algorithm (FA) [7], fireworks algorithm (FWA) [8], monarch butterfly optimization (MBO) [9], and many others [10].

According to the literature review, many swarm intelligence algorithms were successfully applied to the domains that are crucial for innovations of industry and service sectors. The PSO metaheuristics obtained satisfying results for solving job-shop scheduling problem in the era of industry 4.0 [1], as well as in some problems from the domain of IoT [11] and cloud computing [12], MBO algorithm can be applied to localization of unknown sensor nodes in some wireless sensor network topologies [13], etc.

Besides mentioned, swarm intelligence algorithms were also successfully applied to the domains of neural networks training [14], and image processing [15]. In recent years, the adaptations of swarm intelligence methods for image classification with convolutional neural networks (CNN) emerged [16]. These implementations may be crucial for digitalization.

Since the role of metaheuristics will be important in solving NP hard problems that will have to be tackled in the process of transformation of industry and service sectors, and due to the fact that many of such problems can be formulated as bound-constrained and constrained problems global optimization problems, our motivation behind this work is to improve recently proposed tree growth algorithm (TGA) [17] for global optimization assignments.

In this paper, we propose dynamic version of the TGA swarm intelligence metaheuristics for solving global optimization problems. The approach was tested on a set of standard global benchmarks, and with only few modifications, enhanced TGA can be successfully applied to NP hard problems from the domain of innovations of industry and service sectors.

1.2 Research Question

According to the literature survey [17, 18], we concluded that the TGA can be successfully applied to solving range types of NP hard problems. The research question, which is the stepping stone of this paper, can be formulated as:

How to design and implement an efficient and robust TGA algorithm that will be capable of solving bound constrained and constrained global optimization NP hard problems and that will outperform other similar approaches tested for the same problem formulations?

In order to address the research question, we have formulated the following hypothesis:

TGA metaheuristics, as novel and promising method, can be further improved for solving global optimization NP hard problems. Once improved, this method can be easily adapted for solving different kinds of problems that innovations in industry and service systems face.

1.3 Related Work

TGA algorithm is novel approach that was developed in 2017 by Armin Cheraghalipour and Mostafa Hajiaghaei-Keshteli [17]. The algorithm was tested on standard global optimization benchmarks and on some engineering (constrained) benchmarks [17]. According to the literature review, only two papers that deal with this specific method exist in the literature.

On the other hand, since many NP hard problems that are crucial for innovations in industry and service systems were successfully tackled by employing swarm intelligence methods [3, 11, 12], and according to our analysis of the TGA metaheuristics, we concluded that the TGA can be further improved, and in the future research can be adapted for solving various kinds of problems from the domain innovations of industry and service systems.

1.4 Contributions

In this paper, we show an implementation of the improved tree growth algorithm (TGA) adopted for solving global optimization problems. Due to the fact that many problems from the domains of industry and service sectors innovation can be formulated as global NP hard problems, the improved TGA in the unmodified, or slightly adjusted implementations, can be easily applied to these problems as well, and this represents the clear contribution of this paper for this research domain. Also, we should emphasize that the TGA was not explored well, since only two papers that deal with this approach exist in the literature.

2 Relationship to Technological Innovation for Industrial and Service Systems

Currently, the 4th industrial revolution is taking place. The idea of this new revolution is represented by the terms Industry 4.0, Smart Manufacturing and Economy 4.0. During this revolution, many innovations of industrial and service systems will take place. Cloud computing, Internet of things, WSNs, and the integration of "exponential technologies" play an important role in these innovations. Also, during these innovations, increasing digitalization and interconnection of systems, products, value chains, and business models is needed.

However, in order for innovations to be successfully implemented, many problems should be solved. As already, stated in the Sect. 1, most of such problems can be mathematically modelled and such they belong to the category of NP hard optimization. Since the swarm intelligence methods have proved to be efficient in solving NP hard problems, these methods can be applied for tackling NP hard problems when innovating industrial and service systems.

For example, cloud computing may be the production platform for technological innovations for industrial and service systems. According to the literature review, load balancing [19], as well as scheduling problems [20] in cloud computing environments were successfully tackled by using swarm intelligence metaheuristics. Swarm intelligence has also been successfully applied in numerous optimizations in telecommunications and routing systems [4].

3 Basic and Enhanced Tree Growth Algorithm

TGA is novel swarm intelligence metaheuristics developed in 2017 by Armin Cheraghalipour and Mostafa Hajiaghaei-Keshteli [17]. TGA models the competition between the trees in the nature for acquiring light and food sources.

Inspired by the nature, one iteration of the TGA is divided into four phases. In the first phase, N_1 better solutions (individuals) in the population perform the local search process. In the second phase, N_2 solutions are moved to the distance between the close best solutions under different α angles. Third phase is conducted by discarding N_3 worst solutions from the population, which are replaced by randomly generated solution from the feasible domain of the search space. Finally, in the fourth phase, N_4 new solutions are generated, when each newly created solution is modified by the mask operator respect to the best solution in the set N_1 that consists of best solutions in the population.

One execution (run) of the TGA can be summarized in the eight following steps:

Step 1: Generate random initial population of N candidate solutions within the values of lower and upper parameters' bound and calculate fitness of each solution. In the case of minimization problems, the fitness is reverse proportional to the value of objective function.
Step 2: Sort population according to the value of fitness and find the current best solution T_{GB}^{j}, where j-th denotes the current iteration.
Step 3: For N_1 better solutions in the population perform the local search for all solutions' parameters according to the Eq. (1) [17].

$$\boldsymbol{T_i^{j+1}} = \frac{\boldsymbol{T_i^j}}{\boldsymbol{\theta}} + \boldsymbol{rT_i^j} \tag{1}$$

where T_i^j and T_i^{j+1} denote the i-th solution in the population in the iterations j (old solution i) and $j + 1$ (new solution i), respectively, θ is the reproduction rate parameter, and r is uniformly distributed pseudo-random number between 0 and 1. When the new solution is generated, the selection between old and new solution is performed using greedy mechanism (the solution with the higher fitness value is retained).
Step 4: N_2 solutions from the population are moved towards the two closest better solutions from the subpopulation N_1 under different α angles. To find the closest best solutions, the following equation is applied for all solutions' parameters [17]:

$$d_i = \sqrt{\sum_{i=1}^{N_1} \left(T_{N_2}^j - T_i^j\right)^2}, \text{where } d_i = \begin{cases} d_i & \text{if } T_{N_2}^j \neq T_i^j \\ \infty & \text{if } T_{N_2}^j \neq T_i^j \end{cases} \tag{2}$$

When the distance is calculated, then two solutions x_1 and x_2 with the minimal distance (d_i) from each solution $T_{N_2}^j$ are selected for producing linear combination by using the equation [17]:

$$y = \lambda x_1 + (1 - \lambda) x_2 \tag{3}$$

where λ represents the TGA control parameter and its value is between 0 and 1. Finally, all solutions $T_{N_2}^j$ from the subpopulation N_2 are moved between two adjacent solutions with α different angles by employing Eq. (4) [17].

$$T_{N_2}^j = T_{N_2}^j + \alpha_i y \tag{4}$$

Step 5: The N_3 worst solutions are discarded from the population and they are replaced with the randomly generated solutions from the feasible region of the search space.
Step 6: New population N is created, where $N = N_1 + N_2 + N_3$

Step 7: In this step, new set of N_4 randomly distributed solutions are created and each solution is modified by using the mask operator respect to the best solution from the subpopulation N_1. Solutions from the N_4 subpopulation are then added to the population N.

Step 8: Population $N + N_4$ is sorted according to the fitness and the best N solutions are chosen as the initial population for the next iteration of the algorithm. In this step, tournament or roulette wheel selection process is employed.

In the presented steps, N, N_1, N_2, N_4, θ and λ are the control parameters of the TGA.

3.1 Dynamic Tree Growth Algorithm

Two basic processes of any swarm intelligence algorithm, that guide the search, are the exploitation (intensification) and exploration (diversification). To obtain the satisfactory results, the balance between these two processes should be adjusted well.

If the balance between exploration and exploitation in not properly balanced in favor to exploitation, the algorithm may converge to suboptimal solutions in early iterations, and may be trapped in the local optimum, which is called the premature convergence. At the other hand, if this balance is in favor of exploration, it can happen that the algorithm may find the right part of the search space, but may not be able to fine tune around the current best solutions. As a consequence, the algorithm cannot find the optimum solution.

The exploration and exploitation in the original TGA are controlled by two control parameters θ and λ, and the number of solutions N_3, which are discarded from the population in each iteration. The θ parameter directs the local search process, while the λ parameter directs solutions towards the better solutions in the population. Both parameters conduct the process of exploitation. At the other hand, by adjusting N_3, the process of exploration is conducted. All three parameters are static during the whole course of algorithm's execution.

By conducting empirical tests, we concluded that the TGA performance can be improved by dynamically adjusting the values of θ and N_3. In early iterations of algorithm's execution, the value of the θ parameter should be lower, and the newly generated solution should be further from the current solution (Eq. (1)). Basic assumption is that the algorithm in early iterations has not found the right part of the search space. However, in later iterations, the value of the θ parameter should be higher, with the assumption that the algorithm has found the proper part of the search space. The value of the parameter λ should be static, since the nature of Eq. (3).

Also, the number of solutions that are discarded from the population (parameter N_3) should be higher in early iterations, since in this phase more exploration power is needed. The change of N_3 parameter is directly influenced by the values of N_1 and N_2.

By incorporating dynamic behavior of θ and N_3, we devised dynamic search TGA (dynsTGA) metaheuristics. Pseudo-code of dynsTGA is shown in Algorithm 1.

Algorithm 1. Pseudo-code of the dynsTGA metaheuristics

Initialization. Generate pseudo-random population, set the iteration counter $t=1$, maximum iteration number *MaxIter*, and initial values for θ and N_3 control parameters
while $t < MaxIter$ **do**
  Evaluate population and sort all solutions according to their fitness
  for all solutions in N_1
    Perform local search by using Eq. (1)
    Apply greedy selection between old and new solution
  end for
  for all solutions in N_2
    Move solutions towards the closest best solutions in N_1 by using Eqs. (2) – (4)
    Apply greedy selection between old and new solution
  end for
  Discard N_3 worst solutions from the population and replace them with pseudo-random solutions
  Generate N_4 randomly distributed solutions and modify each solution in respect to the best solutions in N_1 by using the mask operator
  Evaluate population and sort all solutions according to their fitness
  Choose N solutions for the new iteration
  Adjust the values of θ and N_3 along with N_2 and N_1 control parameters
end while
return the best solution in the population

4 Empirical Results and Discussion

For the purpose of the research presented in this paper we devised our own software framework by using .NET 4.5 technology and C# in Visual Studio 2017 Integrated Development Environment. In the framework we incorporated both, original TGA and dynsTGA, along with the benchmark functions.

To evaluate the performance of the dynsTGA, we used similar parameter values as in [18] and [17]. The size of initial population (N) was set to 100, while N1 and N2 were set to 20. In the beginning of the algorithm's execution, the value of N3 was set to 60 (N3 = N − (N1 + N2)). The value of N4 was set to 30 and it was fixed during the whole course of the algorithm's execution. The maximum iteration number (MaxIter) was set to 250, and the algorithm was executed in 30 independent runs.

In the first 210 iterations, the values of N_1, N_2 and N_3 were fixed. Then, in each of the following iterations (from iteration 211 to iteration 250), the value of N_3 was decremented by one, and the values of N_1 and N_2 were incremented by 1 in even and odd iterations, respectively. The value of the λ parameter was set to 0.5, while the initial value of θ was set to 0.2. In each iteration, the value of θ was adjusted according to the following expression: $\theta^{j+1} = \theta^j \cdot 1.002$, until the threshold value of 1.5 is reached.

We tested dynsTGA on a standard set of global optimization benchmarks and compared results with the original TGA, as well as with other state-of-the-art approaches that were tested on the same benchmark test. Details of benchmark functions are given in Table 1.

Table 1. Benchmark function details

ID	Name of the problem	Dim.	Type	Parameter range
F1	Ackley's Problem (ACK)	10	Multimodal	(−30, 30)
F2	Aluffi–Pentini's Problem (AP)	2	Multimodal	(−10, 10)
F3	Becker and Lago Problem (BL)	2	Multimodal	(−10, 10)
F4	Easom Problem (EP)	2	Unimodal	(−10, 10)
F5	Rastrigin Problem (RG)	10	Multimodal	(−5.12, 5.12)
F6	Rosenbrock Problem (RB)	10	Multimodal	(−30, 30)
F7	Goldstein and Price Problem (GP)	2	Multimodal	(−2, 2)
F8	Gulf Research Problem (GRP)	2	Unimodal	(0, 100)

Algorithm was tested in 30 independent runs. We performed comparative analysis with outstanding stochastic algorithms from the literature that were tested on the same benchmark set, and compared obtained values for best, mean and standard deviation performance indicators. Comparative analysis is presented in Table 2, where the best results for each performance indicator and each benchmark test are bolded.

Table 2. Simulation results for performance indicators and comparative analysis.

ID	Indicator	IBPA	LADA	TS	WSA	TGA	dynsTGA
F1	Best	0.00815	0.00088	0.14185	0.888E−15	0.00	0.00
	Mean	0.02260	0.00473	0.38528	0.888E−15	0.00	0.00
	StdDev	0.01021	0.00157	0.07488	1.0029E−31	0.00	0.00
F2	Best	−0.35238	−0.35238	−0.35238	−0.35238	−0.35239	−0.35239
	Mean	−0.35238	−0.35238	−0.35238	−0.35236	−0.35239	−0.35239
	StdDev	1.067E−6	5.576E−7	2.183E−5	8.761E−6	6.09E−07	**3.53E−09**
F3	Best	3.217E−9	1.259E−9	3.955E−7	5.589E−8	1.07E−08	**0.00**
	Mean	2.826E−7	2.486E−7	7.637E−6	1.267E−7	3.70E−07	**2.04E−09**
	StdDev	2.838E−7	2.704E−7	6.302E−6	3.877E−8	5.84E−07	**5.11E−10**
F4	Best	−0.99999	−0.99999	−0.99999	−0.99999	−0.99999	−0.99999
	Mean	−0.83334	−0.99999	−0.46667	−0.99957	−0.99999	−0.99999
	StdDev	0.379010	**2.885E−6**	0.507330	2.025E−4	1.59E−06	8.45E−05
F5	Best	0.08790	0.00606	4.58753	0.00	0.00	0.00
	Mean	0.29275	0.01584	6.35541	0.00	0.00	0.00
	StdDev	0.12481	0.00554	0.89405	0.00	0.00	0.00
F6	Best	1.6578	13.1161	24.7395	8.9167	0.5653	**0.5361**
	Mean	12.1420	26.4740	66.1024	8.9449	**0.8231**	0.8502
	StdDev	14.9202	14.9521	19.1763	0.0160	0.3419	**0.3137**
F7	Best	3.00000	3.00000	3.00000	3.00000	3.00207	3.00000
	Mean	5.70001	10.00710	3.00053	**3.00032**	3.11253	3.00068
	StdDev	8.23847	16.46670	**5.751E−4**	1.622E−4	1.34E−01	7.305E−4
F8	Best	5.399E−6	8.124E−5	3.120E−5	32.83	1.52E−05	**8.45E−06**
	Mean	**0.00157**	5.362E−4	2.047E−4	32.83	7.05E−02	3.50E−02
	StdDev	**0.00162**	3.456E−4	1.382E−4	1.445E−15	2.60E−01	1.72E−01

First thing that we want to emphasize is that from the Table 2, it is clear that the dynsTGA outperforms original TGA metaheuristics. For example, in F3 (BL Problem) benchmark, dynsTGA showed better performance for best, mean, as well as for standard deviation indicator. Similar, in the case of F8 (GRP) test, dynsTGA showed significantly better robustness than the basic TGA. The improvements over unmodified TGA are also significant in the F6 test, where dynsTGA obtained better values.

As can be seen from Table 2, comparative analysis with IBPA [17], LADA [17], TS [17] and WSA [17] was performed. In average, dynsTGA approach performs better than all other metaheuristics that are included in comparative analysis. The superiority of dynsTGA can be best pictured in the F3 (BL Problem) benchmark, where dynsTGA completely outperformed all other approaches for best, mean and standard deviation indicators.

5 Conclusion

In this paper, a modified and improved version of relatively new tree growth algorithm (TGA) metaheuristics was presented. Only few papers from the literature deal with the TGA approach. Original TGA was enhanced by introducing dynamical adjustment of exploitation and exploration search parameters. The performance of enhanced TGA metaheuristics was measured on standard set of global optimization benchmarks. In the experiments, similar parameter setup as in [17] was applied.

To prove the robustness and solutions' quality of the dynamic TGA, comparative analysis with the basic TGA, as well as with other state-of-the-art algorithms for global optimization was conducted. From the presented side-by-side comparison, it is obvious that the dynamic TGA significantly improves original TGA and performs better than other algorithms that were included in analysis.

Since many problems from the domains of industrial and service systems can be modeled as global optimization tasks, dynamic TGA metaheuristics shows great potential in this area and can be adapted for tackling many real-world unconstrained and constrained optimization challenges.

Acknowledgements. This research is supported by the Ministry of Education, Science and Technological Development of Republic of Serbia, Grant No. III-44006. The work of M. Beko was supported in part by Fundação para a Ciência e a Tecnologia under Projects UID/MULTI/04111/0213 and UID/MULTI/04111/0216, UID/EEA/00066/2013 and foRESTER PCIF/SSI/0102/2017, and Grant IF/00325/2015.

References

1. Leusin, M.E., Frazzon, E.M., Maldonado, M.U., Kück, M., Freitag, M.: Solving the job-shop scheduling problem in the Industry 4.0 era. Technologies **6**(4) (2018). https://doi.org/10.3390/technologies6040107
2. Strumberger, I., Beko, M., Tuba, M., Minovic, M., Bacanin, N.: Elephant herding optimization algorithm for wireless sensor network localization problem. In: Camarinha-Matos, L.M., Adu-Kankam, K.O., Julashokri, M. (eds.) DoCEIS 2018. IAICT, vol. 521, pp. 175–184. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-78574-5_17
3. Abraham, A., Das, S., Roy, S.: Swarm intelligence algorithms for data clustering. In: Maimon, O., Rokach, L. (eds.) Soft Computing for Knowledge Discovery and Data Mining, pp. 279–313. Springer, Boston (2008). https://doi.org/10.1007/978-0-387-69935-6_12
4. Ducatelle, F., Gianni, A.D., Luca, M.G.: Principles and applications of swarm intelligence for adaptive routing in telecommunications networks. Swarm Intell. **4**(3), 173–198 (2010)
5. Kennedy, J., Eberhart, R.: Particle swarm optimization. In: Proceedings of ICNN 1995 - International Conference on Neural Networks, Perth, WA, Australia, pp. 1942–1948 (1995). https://doi.org/10.1109/icnn.1995.488968
6. Bacanin, N., Tuba, M.: Artificial Bee Colony (ABC) algorithm for constrained optimization improved with genetic operators. Stud. Inform. Control **21**(2), 137–146 (2012)
7. Yang, X.-S., He, X.: Firefly algorithm: recent advances and applications. Int. J. Swarm Intelligence **1**(1), 36–50 (2013). https://doi.org/10.1504/IJSI.2013.05580
8. Strumberger, I., Tuba, E., Bacanin, N., Beko, M., Tuba, M.: Bare bones fireworks algorithm for the RFID network planning problem. In: 2018 IEEE Congress on Evolutionary Computation (CEC), Rio de Janeiro, pp. 1–8 (2018). https://doi.org/10.1109/cec.2018.8477990
9. Wang, G.-G., Deb, S., Cui, Z.: Monarch butterfly optimization. In: Neural Computing and Applications, pp. 1–20 (2015)
10. Tuba, M., Bacanin, N.: Hybridized bat algorithm for multi-objective radio frequency identification (RFID) network planning. In: 2015 IEEE Congress on Evolutionary Computation (CEC), Sendai, pp. 499–506 (2015). https://doi.org/10.1109/cec.2015.725693
11. Nouiri, M., Jemai, A., Ammari, A.C., Bekrar, A., Trentesaux D., Niar, S.: Using IoT in breakdown tolerance: PSO solving FJSP. In: 2016 11th International Design & Test Symposium (IDT), Hammamet, pp. 19–24 (2016). https://doi.org/10.1109/idt.2016.7843008
12. Masdari, M., Salehi, F., Jalali, M., et al.: A survey of PSO-based scheduling algorithms in cloud computing. J. Netw. Syst. Manage. **25**(1), 122–158 (2017). https://doi.org/10.1007/s10922-016-9385-9
13. Strumberger, I., Tuba, E., Bacanin, N., Beko, M., Tuba, M.: Monarch butterfly optimization algorithm for localization in wireless sensor networks. In: 2018 28th International Conference Radioelektronika (RADIOELEKTRONIKA), Prague, pp. 1–6 (2018). https://doi.org/10.1109/radioelek.2018.8376387
14. Tuba, M., Alihodzic, A., Bacanin, N.: Cuckoo search and bat algorithm applied to training feed-forward neural networks. In: Yang, X.-S. (ed.) Recent Advances in Swarm Intelligence and Evolutionary Computation. SCI, vol. 585, pp. 139–162. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-13826-8_8
15. Tuba, E., Alihodzic, A., Tuba, M.: Multilevel image thresholding using elephant herding optimization algorithm. In: 2017 14th International Conference on Engineering of Modern Electric Systems (EMES), Oradea, pp. 240–243 (2017). https://doi.org/10.1109/EMES.2017.7980424

16. França da Silva, G.C., Valente, T.L.A., Silva, A.C., Cardoso de Paiva, A., Gattass, A.: Convolutional neural network-based PSO for lung nodule false positive reduction on CT images. Comput. Methods Programs Biomed. **162**, 109–118 (2018). https://doi.org/10.1016/j.cmpb.2018.05.006
17. Cheraghalipour, A., Hajiaghaei-Keshteli, M.: Tree Growth Algorithm (TGA): an effective metaheuristic algorithm inspired by trees' behavior. In: 13th International Conference on Industrial Engineering, vol. 13 (2017)
18. Cheraghalipour, A., Hajiaghaei-Keshteli, M., Paydar, M.M.: Tree Growth Algorithm (TGA): a novel approach for solving optimization problems. Eng. Appl. Artif. Intell. **72**, 393–414 (2018). https://doi.org/10.1016/j.engappai.2018.04.021
19. Li, D., Li, K., Liang, J., Ouyang, A.: A hybrid particle swarm optimization algorithm for load balancing of MDS on heterogeneous computing systems. Neurocomputing (2018, in press). https://doi.org/10.1016/j.neucom.2018.11.034
20. Kalra, M., Singh, S.: A review of metaheuristic scheduling techniques in cloud computing. Egypt. Inform. Journal **16**(3), 275–295 (2015). https://doi.org/10.1016/j.eij.2015.07.001

Assistive Systems

Intelligent HCI Device for Assistive Technology

Rui Azevedo Antunes[2,3](✉), Luís Brito Palma[1,3], Hermínio Duarte-Ramos[1], and Paulo Gil[1,3]

[1] Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, 2829-516 Caparica, Portugal
{lbp,psg}@fct.unl.pt, hduarteramos@gmail.com
[2] Escola Superior de Tecnologia de Setúbal, Instituto Politécnico de Setúbal, 2914-761 Setúbal, Portugal
rui.antunes@estsetubal.ips.pt
[3] Centre of Technology and Systems (CTS-Uninova), Caparica, Portugal

Abstract. This paper presents a new intelligent Human-Computer Interaction (HCI) device for Assistive Technology. The developed device can be used as a mouse or as a gamepad, moving a part of the human body, typically the head, for hands-free computer access tasks. The state-of-the-art hardware uses an Advanced RISC Machine 32-bit microcontroller and a microelectromechanical 9-degree motion sensor, including a System-in-Package (SiP) accelerometer, gyroscope and magnetometer. The hardware/software device with a human-in-the-loop controller can be identified as a Cyber-Physical intelligent system to be incorporated in the "Industry 4.0" trend. Results reveal that the embedded controller of the HID device allows the improvement of the user's performance, decreasing the effort and the execution time of the hands-free computer tasks.

Keywords: Human-in-the-Loop (HuIL) control · Assistive Technology (AT) · Microelectromechanical Systems (MEMS) · Microcontrollers

1 Introduction

The main goal of this work is to test an intelligent Cyber-Physical System (CPS) device, developed in [1], with several microelectromechanical sensors, to allow hands-free computer access for persons with disabilities. The simplified architecture of the designed Human Interface Device (HID) is presented in Fig. 1.

The human operator performs head movements, with the HID module, to use the computer, replacing the traditional mouse, keyboard and gamepad/joystick. The HID module dimensions are only: 42 × 21 × 9 mm.

State-of-the-art hands-free human-computer assistive devices nowadays in the market are the Quha Zono gyroscopic mouse [2] and the GlassOuse assistive device [3]. There are also other non-gyroscopic commercial solutions, not so accurate, since are based on vision, eye-gaze and speech recognition technologies, such as the Tobii Dynavox PCEye Plus [4] and the AbleNet TrackerPro [5] assistive devices. Recent advances in brain-computer interaction and human-computer accuracy, using micro-electrodes implanted into the brain [6], have the disadvantages and potential risks of

Published by Springer Nature Switzerland AG 2019
L. M. Camarinha-Matos et al. (Eds.): DoCEIS 2019, IFIP AICT 553, pp. 157–168, 2019.
https://doi.org/10.1007/978-3-030-17771-3_13

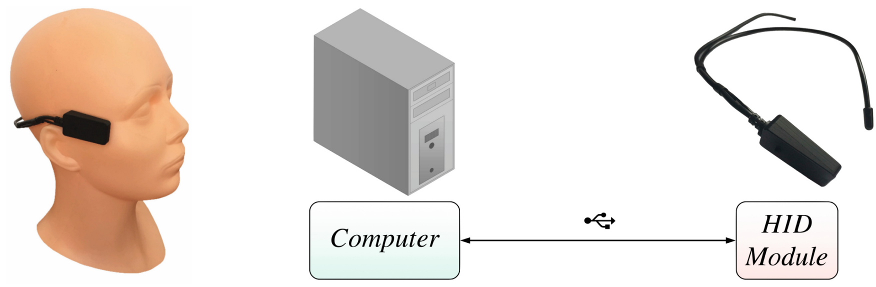


Fig. 1. HID module placed in the head (*left*). Simplified architecture of the Human-Computer Interface device in USB connection mode (*right*).

being intrusive experimental techniques. The assistive devices mentioned are mainly based on open-loop electronic systems.

The proposed approach in this work is focused on a closed-loop assistance scheme with error feedback, to assist the operator and reduce human effort. It is based in the Human Adaptive Mechatronics (HAM) concept [7, 11], and consists in adding an embedded electronic assistance-controller, tuned from a simplified mathematical model that represents the main characteristics and handicaps of each individual user with disabilities. Hence, the main research question of the PhD thesis [1], in which the proposed prototype is described in detail, is: "*Can performance of a system integrating the human being be improved, if an assistance-controller is used, that takes into account the behaviour and constraints of the human operator and the machine?*", and the research hypothesis, related to this part of the PhD work, is: "*If following (Point-To-Point) tasks are performed, then linear modelling is a feasible method for the design of assistance control systems*".

The new HID module, with an "intention" feedback PID controller, tuned by a nature-inspired metaheuristic algorithm, is able to sense the head/body movements. The developed hardware/software allows a human user to perform, hands-free, the computer's cursor movements, the gamepad/joystick movements, the virtual keyboard and the mouse/gamepad/joystick buttons activation, the single click and double click mouse configurations, and also the "drag and drop" and "scroll" functionalities.

The electronic device incorporates the 32-bit ARM Cortex-M4 NXP-Freescale MK20DX256VLH7 microcontroller and the 9-degree MEMS motion sensor TDK InvenSense MPU-9250 [8, 9]. The field of application for the developed device includes the assistance for several motor and sensorial impairments, resulting from tetraplegia, quadriplegia, multiple sclerosis, cerebral palsy, amyotrophic lateral sclerosis, cranioencephalic and vertebral-medullary injuries, muscular dystrophy, the carpal tunnel or the Rett syndrome, where a human user has reduced or no control of his hands, and is commonly unable to use a traditional Human-Computer Interface.

2 Relationship to Industrial and Service Systems

The fourth Industrial Revolution results nowadays from the integration of innovative production systems and information technologies. The concept of Additive Manufacturing (AM) [10] is considered the new paradigm of the "Industry 4.0" and the "Economy 4.0" trends.

The developed hardware/software can be considered as an intelligent Cyber-Physical System (CPS) with a dedicated human-in-the-loop assistance architecture, based on the human operator "intention", providing several applications for Assistive Technology and allowing the increase of the user's autonomy and productivity in computer access for service, industrial systems, active learning and entertainment, complying the Web Content Accessibility Guidelines (WCAG). The aim of this work is to make a valid contribution, adopting CPS, to a more inclusive society.

3 Assistance-Controller Project

According to previous works [11], one of the simplest human controller model is a Proportional and Derivative (PD) controller, defined as

$$H(s) = \frac{K_{\mathrm{dh}}s + K_{\mathrm{ph}}}{T_{\mathrm{N}}s + 1} e^{-\tau s}. \tag{1}$$

where K_{dh} and K_{ph} are the derivative and the proportional gains of the human brain controller, T_{N} is a neuromuscular time constant and τ is a reaction time-delay. These parameters change according to human status and the operation. The experiments conducted by Suzuki et al. [11] revealed that the overcoming progress from learning could be considered a slow and time-variant process. Hence, the time-constant τ can be discarded for identification purposes, and the simplified human model becomes

$$H(s) \approx \frac{K_{\mathrm{dh}}s + K_{\mathrm{ph}}}{T_{\mathrm{N}}s + 1}. \tag{2}$$

The human controller $H(s)$ transfer function (2), can be rewritten in the form

$$H(s) \approx K_1 \frac{s + a_{\mathrm{h}}}{s + b_{\mathrm{h}}} \qquad a_{\mathrm{h}}, b_{\mathrm{h}}, K_1 \in \mathbb{R}^+. \tag{3}$$

$s = -a_{\mathrm{h}}$ and $s = -b_{\mathrm{h}}$ are the zero and the pole of $H(s)$, and K_1 a system constant. The proportional gain K_{ph} and the derivative gain K_{dh} of the human PD controller are defined by

$$K_{\mathrm{ph}} = \frac{a_{\mathrm{h}}K_1}{b_{\mathrm{h}}}. \tag{4}$$

$$K_{\mathrm{dh}} = \frac{K_1}{b_{\mathrm{h}}}. \tag{5}$$

The human neuromuscular time-constant T_N is obtained by

$$T_N = \frac{1}{b_h}. \quad (6)$$

With the $H(s)$ (3) static gain K_0 computed by

$$K_0 = \frac{a_h K_1}{b_h} = K_{ph}. \quad (7)$$

A simplified block diagram of the human-in-the-loop (HuIL) system, formed by the computer access hardware that features the cursor movements (m_0/s), for each axis, and the $H(s)$ human user model is presented in Fig. 2, with the Operating System (OS) Enhanced Pointer Precision (EPP) option disabled.

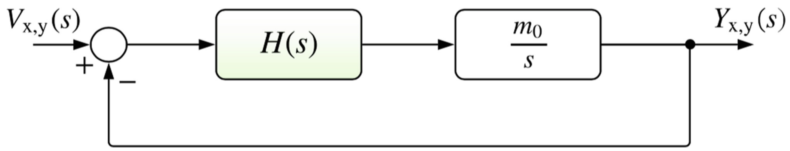


Fig. 2. Block diagram of the Human-Computer Interface (HCI), in "mouse" mode.

The transfer function $G_{x,y}(s)$ of the resulting closed-loop system, that aggregates the human operator and the HID device, is

$$G_{x,y}(s) = \frac{Y_{x,y}(s)}{V_{x,y}(s)} = \frac{\frac{m_0}{s} H(s)}{1 + \frac{m_0}{s} H(s)}. \quad (8)$$

Resulting, from (3) and (8)

$$G_{x,y}(s) = \frac{m_0 K_1 (s + a_h)}{s^2 + (m_0 K_1 + b_h)s + m_0 K_1 a_h}. \quad (9)$$

Which leads, for each axis, to a second-order stable linear system, formed by two negative real poles and a zero, admitting the condition $2m_0 K_1 (2a_h - b_h) \leq b_h^2 + (m_0 K_1)^2$.

The resulting $G_{x,y}(s)$ system can be identified, for each axis, from an Auto-Regressive Linear Model with Exogenous Input (ARX)[1] discrete time structure, with $na = 2$, $nb = 2$ and $nd = 1$, neglecting the delay, which corresponds to

[1] The MATLAB *ident* command from the System Identification toolbox, provides a Graphical User Interface (GUI), allowing the identification of this type of ARX structure by Least Squares (LS), returning the T_{P_1}, T_{P_2} and T_z parameters of the equivalent closed-loop resulting model $G_{x,y}(s)$ (10), for each screen axis, that aggregates the operator dynamics and the HID device.

$$G_{x,y}(s) = \frac{1 + T_z s}{(1 + T_{p_1} s)(1 + T_{p_2} s)}. \tag{10}$$

With

$$a_h = \frac{1}{T_z}. \tag{11}$$

$$b_h = \frac{T_{p_1} + T_{p_2} - T_z}{T_{p_1} T_{p_2}}. \tag{12}$$

$$K_1 = \frac{T_z}{m_0 T_{p_1} T_{p_2}}. \tag{13}$$

Resulting, from (3), (11), (12) and (13)

$$H_{x,y}(s) = \frac{T_z}{m_0 T_{p_1} T_{P_2}} \left(\frac{s + \frac{1}{T_z}}{\frac{T_{p_1} + T_{p_2} - T_z}{T_{p_1} T_{p_2}} + s} \right). \tag{14}$$

The parameter $m_0 = 0.1458$ of (8) is experimentally obtained, being proportional to the computer's cursor speed, moving in one screen direction (X) with the electronic HID device configured for mouse operating mode. The m_0 constant corresponds to the inverse of the time duration for the cursor to reach the half screen distance using one axis (from normalized positions 1 to 0, or normalized positions −1 to 0). This value represents the effortless horizontal head rotation angle, affecting the proportional gain of the computer's cursor dynamics and also the resulting closed-loop system transfer function L_{HCR}, obtained in (17). The cursor speed was settled to correspond to an (effortless) horizontal head rotation (yaw) of $\pm 10° = \pm\pi/18$ rad, without EPP. It was experimentally verified that when m_0 is too high, it could lead to instability. Nevertheless, this effect could still be compensated, by adjusting directly from the software, the computer's mouse speed. On the other hand, if m_0 is too low, the Point-To-Point (PTP) time response increases, resulting in a worse performance.

The block diagram of the closed-loop computer access assistance system (Fig. 3), includes a Proportional-Integral-Derivative (PID) controller, tuned by the Firefly Algorithm (FA), described in Sect. 3.1. The control architecture is based on a cascade control scheme, that uses the operator's cursor "intention". The PID controller C_{PID}

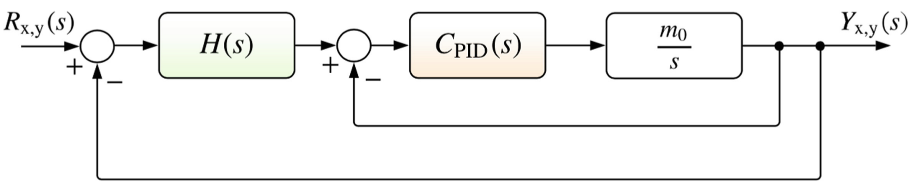


Fig. 3. Cascade control architecture, for each axis, of the hands-free computer access system.

with anti-windup is tuned by the nature-inspired optimization FA algorithm, which takes into account the estimated human model $H(s)$, an expert operator specification and a set of constraints.

The transfer function $L_{\mathrm{HCR}}(s)$ of the resulting closed-loop system, for each screen axis (X and Y), is obtained by

$$L_{\mathrm{HCR}_{x,y}}(s) = \frac{Y_{x,y}(s)}{R_{x,y}(s)} = \frac{m_0 H(s) C_{\mathrm{PID}}(s)}{s + (H(s)+1) m_0 C_{\mathrm{PID}}(s)}. \tag{15}$$

With

$$C_{\mathrm{PID}}(s) = K_{\mathrm{p}}\left(1 + \frac{1}{T_{\mathrm{i}} s} + T_{\mathrm{d}} s\right). \tag{16}$$

From (14), (15) and (16), the resulting transfer function $L_{\mathrm{HCR}}(s)$, for each axis, is obtained by

$$L_{\mathrm{HCR}_{x,y}}(s) = \frac{\frac{K_{\mathrm{p}} T_{\mathrm{z}}}{T_{\mathrm{p}_1} T_{\mathrm{p}_2}}\left(\frac{s + \frac{1}{T_{\mathrm{z}}}}{\frac{T_{\mathrm{p}_1} + T_{\mathrm{p}_2} - T_{\mathrm{z}}}{T_{\mathrm{p}_1} T_{\mathrm{p}_2}} + s}\right)\left(1 + \frac{1}{T_{\mathrm{i}} s} + T_{\mathrm{d}} s\right)}{s + K_{\mathrm{p}}\left(\frac{T_{\mathrm{z}}}{T_{\mathrm{p}_1} T_{\mathrm{p}_2}}\left(\frac{s + \frac{1}{T_{\mathrm{z}}}}{\frac{T_{\mathrm{p}_1} + T_{\mathrm{p}_2} - T_{\mathrm{z}}}{T_{\mathrm{p}_1} T_{\mathrm{p}_2}} + s}\right) + m_0\right)\left(1 + \frac{1}{T_{\mathrm{i}} s} + T_{\mathrm{d}} s\right)}. \tag{17}$$

3.1 Controller Tuning with the Firefly Algorithm

The goal for the PID controller tuning is to find the values of K_{p}, K_{i} and K_{d} that minimize the specification for $L_{\mathrm{HCR}}(s)$, given by

$$min\{0.3|oversh - 10| + 0.5|ris\_time - 3| + 0.2|set\_time - ris\_time - 2|\}$$

which characterize the step response of $L_{\mathrm{HCR}}(s)$ with the (expert) operator A, without EPP. The variables *oversh*, *rise_time* and *set_time* stand respectively for percentage overshoot, rise time and settling time.

The gain constraints of the PID assistance-controller in the three-dimensional search space ($d = 3$) take into account that the proportional and the integral gains K_p and K_i are not greater than 0.5 and the derivative gain K_d is less than 0.02, so that the step response of the human-computer control system does not become too fast (which could generate instability), and to allow to minimize the rise time and the settling time.

As the attractiveness β of a given firefly in the search space is proportional to its luminous intensity, recognized in the environment by the adjacent fireflies, β will be obtained by

$$\beta = \beta_0 e^{-\gamma r^2}. \tag{18}$$

With γ representing the light absorption coefficient.

At distance $r = 0$, the attractiveness of the firefly is given by β_0. The Cartesian distance r_{ij} in the d-space between two fireflies i and j, located at $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$, is defined by

$$r_{ij} = \left\|\boldsymbol{x}_i - \boldsymbol{x}_j\right\|. \tag{19}$$

That is

$$r_{ij} = \sqrt{\sum\nolimits_{k=1}^{d} \left(x_{i,k} - x_{j,k}\right)^2}. \tag{20}$$

When a given firefly i, at the search space location $\boldsymbol{x}_i$, is attracted to another brighter firefly j, its new $\boldsymbol{x}_i'$ position in the d-Euclidean space is computed, at iteration $iter$, by

$$\boldsymbol{x}_i' = \boldsymbol{x}_i + \beta_0 e^{-\gamma r_{ij}^2}\left(\boldsymbol{x}_j - \boldsymbol{x}_i\right) + \alpha_0 \upsilon^{iter} \mathbf{z}_i. \tag{21}$$

$\propto_0 \in [0, 1]$ is the randomization parameter, $\mathbf{z}_i$ is a d-dimensional array of random numbers with uniform distribution, and $\upsilon \in [0, 1]$ is the mutation coefficient employed.

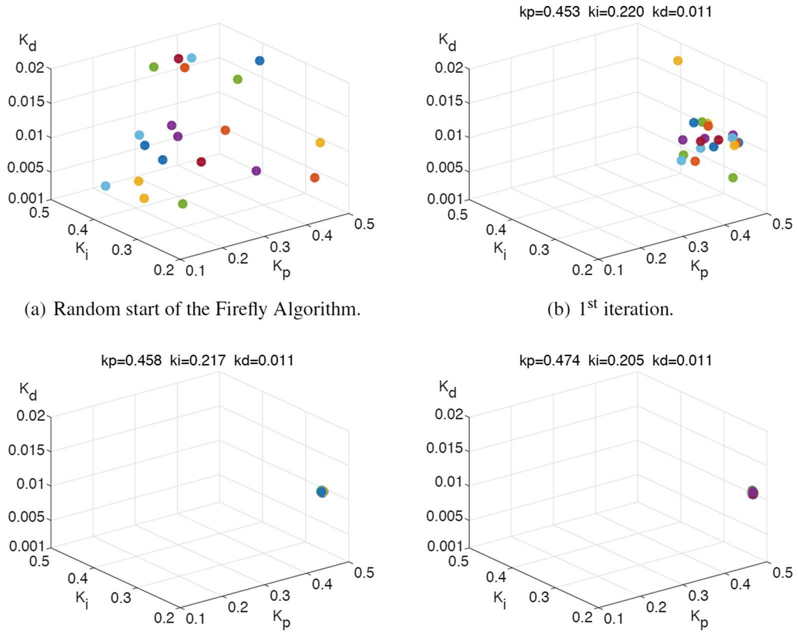
The Proportional-Integral-Derivative (PID) controller tuning method is described in the Algorithm 1 [12, 13], taking into account the light absorption $\beta_0 = 1.0$, the randomization parameter $\alpha = 0.2$ and the mutation coefficient $\upsilon = 0.98$.

Algorithm 1: Firefly Algorithm ([12]).

Define the objective function $f(\boldsymbol{x})$, with $\boldsymbol{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_d \end{bmatrix}^T$.
Create the initial population of n fireflies $\boldsymbol{x}_{i=1,2,\ldots,n}$.
The light intensity I_i at $\boldsymbol{x}_i$ is obtained by $f(\boldsymbol{x}_i)$.
Define the light absorption coefficient γ.
Define the attractiveness β_0 for zero distance ($r = 0$).
Define the randomization parameter α.
Define the maximum number of iterations Max_iter.
while $iter < max\_iter$ **do**
  for $i = 1 : n$ **do**
    for $j = 1 : n$ **do**
      if $I_i < I_j$ **then**
        Move the firefly i to j.
      end if
      Vary attractiveness with the distance r via $e^{-\gamma r}$.
      Evaluate the new solutions and update the light intensity.
    end for
  end for
  Rank the n fireflies and find the global best $\boldsymbol{g}^*$.
end while
Process and visualize results.

An example of the Firefly Algorithm (FA) evolution, with 20 fireflies and 10 iterations, is presented in Fig. 4, taking into account the human operator A $H(s)$ model (14), with $T_{p_1} = 0.777$, $T_{p_2} = 0.874$ and $T_z = -0.245$.

The proposed nature-inspired optimization methodology ensures improved performance in the hands-free task movements of the computer's cursor, as shown in the experimental results.



(a) Random start of the Firefly Algorithm.

(b) 1st iteration.

(c) 3rd iteration.

(d) 10th iteration.

Fig. 4. Convergence of the Firefly Algorithm (FA).

4 Discussion of Results and Critical View

A total of 500 Point-To-Point (PTP) random sequences were performed using the head-mounted HID device, connected to the computer via USB cable, without and with the Proportional-Integral-Derivative (PID) controllers. Each PTP task, with the cursor and the EPP option enabled, consists in following a random PTP target in the computer screen, at distance, in pixels (px), $A = \{300\,\text{px}, 400\,\text{px}, 500\,\text{px}, 600\,\text{px}, 700\,\text{px}\}$, and with the target diameter $W = 100$ px, whose position on the screen changes every 10 s.

The PID controller, with anti-windup, is tuned from the Firefly Algorithm. The choice for the number of fireflies, iterations and attempts took into account a good convergence-speed compromise. The FA algorithm results revealed convergence with 20 fireflies, 10 iterations and 5 attempts, resulting in an average time duration of 5 min and 30 s, using a Skylake dual-core processor Intel® Core i5-6300U, running at 2.40 GHz, and the simplified mathematical model for each human participant (A to E), consisting of an equivalent Auto-Regressive Model with Exogenous (ARX) structure, with $na = 2$, $nb = 2$ and $nd = 1$, neglecting the delay (14), with an operating frequency of 50 Hz ($T_s = 20$ ms), the filter coefficient $N_{\text{PID}} = 10$ and the anti-windup time constant $T_t = 1$ s.

The discrete implementation of the real-time, anti-windup, PID controller structure into the embedded microcontroller of the Human Interface Device (HID) module was programmed in C programming language. Examples of acquired signals during Point-To-Point (PTP) random tasks, performed by the human participant B, using the hands-free HID developed device, without and with the PID assistance controller, are depicted in Fig. 5.

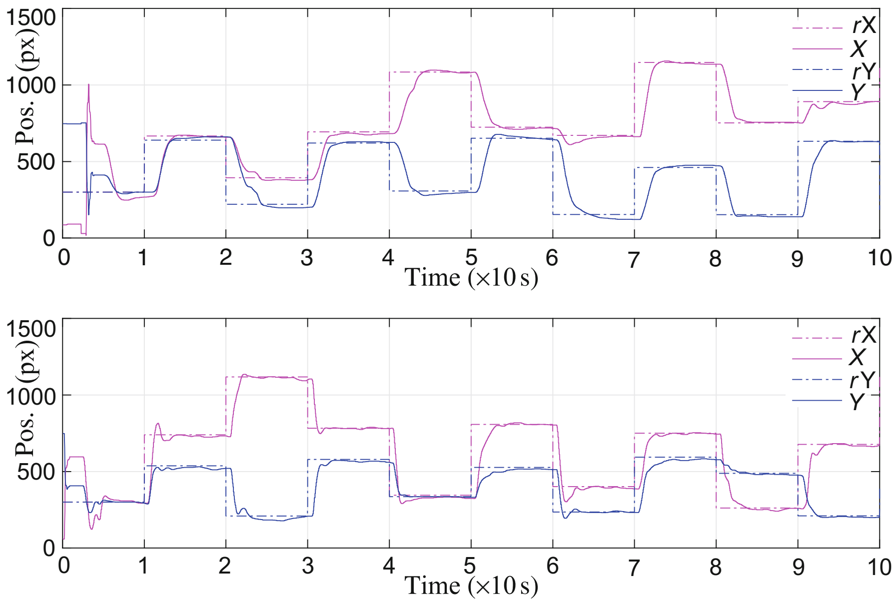


Fig. 5. PTP target signals (r_X, r_Y) and correspondent responses (X, Y), for X and Y axis of human participant B: without controller (*top*); with the controller (*bottom*).

4.1 Results and Critical View

The obtained experimental results concerning the PTP tasks time duration T_{PTP} for human participants A to E are presented in Table 1, without and with the assistance-controller, using the Mahony et al. sensorial fusion methodology [14, 15]. Best results are depicted in italic.

It can be confirmed that the assistance-controllers improve the human-computer performance in the PTP tasks with the developed hands-free HID device, resulting in a decrease of 44% in the average PTP time duration.

Table 1. PTP results, without and with the controllers, for a sequence of 10 PTP random tasks, with target distance A.

T_{PTP} (×10 s)					
Without controller	A = 300 px	A = 400 px	A = 500 px	A = 600 px	A = 700 px
Participant A	1.97	2.28	2.98	3.46	3.54
Participant B	2.61	3.17	3.35	3.33	3.37
Participant C	7.41	7.09	9.36	8.12	9.51
Participant D	3.45	2.87	3.45	3.39	4.10
Participant E	3.40	4.42	5.54	5.44	6.17
With controllers	A = 300 px	A = 400 px	A = 500 px	A = 600 px	A = 700 px
Participant A	*1.23*	*1.50*	*1.91*	*1.82*	*2.43*
Participant B	*1.74*	*2.26*	*2.33*	*2.52*	*2.73*
Participant C	*3.75*	*6.31*	*4.59*	*6.36*	*6.74*
Participant D	*2.33*	*2.58*	*2.72*	*2.89*	*2.63*
Participant E	*3.27*	*3.32*	*2.99*	*4.22*	*3.82*

5 Conclusions and Further Work

The hardware/software device with the user assistance-controller improves the hands-free human-computer performance, allowing the replacement of computer access interfaces that require manual actions/movements. The relevance of the human hands can be proven by the Motor *Homunculus* model (Fig. 6), initially proposed in 1937 by Penfield & Boldrey [16], consisting of a representation of how the human body would be, if the size of its parts grew in proportion to the area of the brain related with the correspondent motor functions in the central sulcus of the frontal cortex.

Fig. 6. The Motor *Homunculus* 3D model.

The controller of the developed hardware/software device can be configured, in a customized way, for the specific needs of each human user, allowing the adjustment of the PID parameters and also their storage in the EEPROM of the HID microcontroller, using a C# configuration application, also built. The ability to customize the computer access device, tailored to the behaviour and the relevant capabilities of each human user, modelled by a simplified mathematical transfer function, represents a novel trend in Assistive Technology (AT).

The hardware/software developed in [1] for Assistive Technology, which can be shown working, may in future be used for other relevant areas, such as biomedical engineering, rehabilitation and sports medicine, physiotherapy and interface systems engineering.

References

1. Antunes, R.A.: Controlo em Sistemas de Interação Humano-Máquina. PhD thesis, Universidade Nova de Lisboa (2018)
2. Quha Zono Gyroscopic Mouse: Quha Assistive Technology Product Catalog (2019). www.quha.com/wp-content/uploads/2018/10/Quha_Product_catalog_2019_for_web.pdf
3. GlassOuse Assistive Device (2018). https://www.glassouse.com/
4. Tobii Dynavox 2017, PCEye Plus, eye tracker for individuals facing mobility challenges. https://www.tobii.com/group/news-media/press-releases/2017/5/tobii-dynavox-launches-pceye-plus-the-next-generation-eye-tracker-for-individuals-facing-mobility-challenges/
5. TrackerPro, AbleNet (2019). https://www.ablenetinc.com/trackerpro
6. Nuyujukian, P., et al.: Cortical control of a tablet computer by people with paralysis. PLoS ONE **13**(11), e0204566 (2018). https://doi.org/10.1371/journal.pone.0204566
7. Laurikkala, M., Suzuki, S., Vilkko, M.: Predicting operator's cognitive and motion skills from joystick inputs. In: IECON 2016 - 42nd Annual Conference of the IEEE Industrial Electronics Society, Florence, Italy, pp. 5935–5940 (2016)
8. InvenSense MPU-9250 Product Specification Revision. https://www.invensense.com/wp-content/uploads/2015/02/PS-MPU-9250A-01-v1.1.pdf
9. MPU-9250 Nine-Axis (Gyro + Accelerometer + Compass) MEMS Motion Tracking Devices. www.invensense.com/products/motion-tracking/9-axis/mpu-9250/
10. Meboldt, M., Klahn, C. (eds.): Industrializing Additive Manufacturing - Proceedings of Additive Manufacturing in Products and Applications - AMPA 2017. Springer (2018). https://doi.org/10.1007/978-3-319-66866-6
11. Suzuki, S., Harashima, F., Kurihara, K.: Assist control and its tuning method for haptic system. In: 9th IEEE International Workshop on Advanced Motion Control (AMC 2006), Istanbul, Turkey, pp. 374–379 (2006)
12. Yang, X.S.: Nature-Inspired Metaheuristic Algorithms. Luniver Press (2010)
13. Yang, X.-S. (ed.): Cuckoo search and firefly algorithm. SCI, vol. 516. Springer (2014). https://doi.org/10.1007/978-3-319-02141-6
14. Mahony, R., Hamel, T., Pflimlin, J.: Nonlinear complementary filters on the special orthogonal group. IEEE Trans. Autom. Control **53**(5), 1203–1218 (2008)

15. Mahony, R., Hamel, T., Pflimlin, J.M.: Complementary filter design on the special orthogonal group SO(3). In: Proceedings of the 44th IEEE Conference on Decision and Control, Seville, Spain, pp. 1477–1484 (2005)
16. Penfield, W., Boldrey, E.: Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation. Brain **60**(4), 389–443 (1937)

Real-Time Human Body Pose Estimation for In-Car Depth Images

Helena R. Torres[1(✉)], Bruno Oliveira[1], Jaime Fonseca[1], Sandro Queirós[1], João Borges[1], Nélson Rodrigues[1], Victor Coelho[3], Johannes Pallauf[2], José Brito[4], and José Mendes[1]

[1] Algoritmi Center, University of Minho, Guimarães, Portugal
helena.ribeiro.torres@gmail.com
[2] Bosch, Abstatt, Germany
[3] Bosch, Braga, Portugal
[4] 2Ai, Polytechnical Institute of Cávado and Ave, Barcelos, Portugal

Abstract. Over the next years, the number of autonomous vehicles is expected to increase. This new paradigm will change the role of the driver inside the car, and so, for safety purposes, the continuous monitoring of the driver/passengers becomes essential. This monitoring can be achieved by detecting the human body pose inside the car to understand the driver/passenger's activity. In this paper, a method to accurately detect the human body pose on depth images acquired inside a car with a time-of-flight camera is proposed. The method consists in a deep learning strategy where the architecture of the convolutional neural network used is composed by three branches: the first branch is used to estimate the confidence maps for each joint position, the second one to associate different body parts, and the third branch to detect the presence of each joint in the image. The proposed framework was trained and tested in 8820 and 1650 depth images, respectively. The method showed to be accurate, achieving an average distance error between the detected joints and the ground truth of 7.6 pixels and an average accuracy, precision, and recall of 95.6%, 96.0%, and 97.8% respectively. Overall, these results demonstrate the robustness of the method and its potential for in-car body pose monitoring purposes.

Keywords: Autonomous driving · Deep-learning · Depth images · Pose estimation

1 Introduction

In recent years, the concept of autonomous driving vehicles is emerging owing to an increment on the development of advanced driver-assistance systems [1]. In fact, it is expected that fully automated driving will be the next goal for the automobile industry, which will extinguish the role of the driver. Thus, without the necessity of driving the car, the passengers can spend their time doing other types of activities [2]. In this sense,

H.R. Torres and B. Oliveira—Both authors contributed equally.

Published by Springer Nature Switzerland AG 2019
L. M. Camarinha-Matos et al. (Eds.): DoCEIS 2019, IFIP AICT 553, pp. 169–182, 2019.
https://doi.org/10.1007/978-3-030-17771-3_14

the need for monitoring all occupants in the car becomes crucial to analyze the passengers' behavior and, therefore, to ensure their safety.

There are different types of visual sensors that can be used for monitoring purposes, including monitoring persons through the detection of their pose. The most common sensor consists of RGB cameras that allow to retrieve visual information of the interior of the car and its occupants. However, these sensors only produce 2D images which hampers the interpretation of the human activity that is enrolled in the 3D world and is susceptible to brightness variations [3]. A depth sensor, which is robust to light variations, can be used to solve this problem, as it also provides information about the distance between an object and the camera, giving a 3D information about the scene. However, they not give textural information, which is useful to detect different parts of the body. Recently, several works have focused on the integration of RGB and Depth sensors (RGB-D), which allow merging the advantages of both sensors [4], which could potentially increase the robustness of human pose detection. However, these sensors are more expensive in comparison with RGB and depth sensors.

To detect the human body pose in images acquired with camera sensors and, thus, the activity of the car's occupants, a robust method for human body pose estimation is needed. Specifically for the in-car environment, few strategies were previously proposed. Indeed, the images acquired inside a car have several occlusions, making somejoints undetectable and so making the task of estimating these joints hard. This problem hampers the direct usage of traditional human pose estimation algorithms and makes this task more challenging than open space pose estimation scenes. The focus of the present work was to develop a method for human body pose estimation in depth images acquired inside a car. For it, a state-of-the-art method developed for open space human pose estimation in RGB images, Part Affinity Fields (PAF), was extended for an accurate and robust detection of human body pose inside a car using depth images [5]. The main requirements of the method were its accuracy in detecting the pose along with its robustness to deal with large variability between different people (*i.e.* regarding body shape or size). Moreover, to deal with fast human movements, it was also important to guarantee the real-time capability of the pose detection. Overall, the current work introduces three main contributions: (1) extension of the PAF method for the detection of the presence/absence of a specific joint in the image; (2) a data augmentation strategy for depth datasets; and (3) a dataset for human body pose detection inside a car with depth images.

The rest of the paper is organized as follows. In Sect. 2, the relationship of the present work with industrial and services systems is presented. In Sect. 3, the state-of-the-art for human body pose estimation is detailed. The proposed framework for the human body pose estimation is described in Sect. 4. In Sect. 5, some implementation details related to the proposed method are outlined, being the results presented in Sect. 6. In Sect. 7, the method's performance is discussed, with the main conclusions of this paper presented in Sect. 8.

2 Relationship to Industrial and Service Systems

With the progress in technology, it is expected that autonomous cars will become part of our life, and so, this new paradigm has high influence in industrial and service systems. In fact, partial and fully autonomous driving can bring social-economic advantages [6]. One of the advantages will be an improvement in road safety, since most of current accidents is caused by errors committed by the driver. Moreover, it is also expected that the traffic flow and the mobility could be improved in the autonomous driving scenario. Another advantage of these systems is related to the increase of the driver's comfort, which by not being focused on driving will have time to do other types of tasks [7]. However, this technological evolution not only concerns personal aspects but also affects public services. In fact, it is expected that shared autonomous vehicles (SAV) will be the next revolution in public transportation. Similarly to personal autonomous vehicles, the SAV also brings several advantages, such as the improvement of passengers' mobility in an easy and economical way [8].

Despite all the above mentioned advantages in autonomous driving, removing the responsibility of the driver/passengers in the driving activity may not be straightforward as it may seem. Indeed, the safety of both car's occupants and vehicle must always be ensured. Regarding personal vehicles, the driver should have the capability to take control of the car when the autonomous driving option is not safe. Concerning the SAVs, the monitoring of the car environment is crucial to maintain the integrity of the vehicle and to ensure the safety of all passengers. This is intrinsically related to the concept of resilient systems, where a system has the capability to recover from perturbations that may affect its normal functioning, allowing the reduction of the vulnerability of the system. A higher capability of maintaining a good performance of the autonomous vehicles can be achieved by monitoring the occupants of the car, being the proposed method of high interest for industrial and services systems.

3 Related Work

To recognize human body pose in images acquired by camera sensors, and, therefore, the activity of the car's occupants, a robust method for human body pose estimation is required. There are two main classes of algorithms for human body pose detection: generative approaches and discriminative approaches. Generative approaches are designed to fit and deform a model to match the image and detect the pose [9–12]. Discriminative approaches are designed to learn a mapping from image features to a body pose, using only the information of the image [13–17]. Regarding generative ones, the main advantage is the robustness, once these methods fit one or more previous models to the image, allowing to introduce shape prior information to the method. However, this class of methods uses error minimization functions, being susceptible to be trapped in local minima and requiring high computational cost. In opposition, discriminative methods are fast during inference, allowing to more easily achieve real-time detection. Moreover, these methods are capable to deal with large body shape variations. However, they are inherently limited by the amount and quality of the training data. Nevertheless, discriminative approaches, namely deep learning strategies, have shown to be more robust and accurate for human body pose detection than other type of algorithms [18–20].

Owing to the success of discriminative approaches for human body pose estimation, several methods have been developed using these type of strategies, specifically deep learning methods. In [21], a graphical model for human pose estimation was used where convolutional neural networks (CNN) were used to learn the different body parts and their spatial relationships (represented by a mixture model over several possible relations). In [22], the graphic model for the human body pose was obtained using the Markov Random Fields approach. Despite its accuracy in human body pose estimation, the previous methods rely on graphic models which can fail to model complex human poses. To overcome this issue, in [23], a dual source CNN was used to detect both full body and local body parts to achieve the final body pose estimation. In [24], another method for human pose estimation using CNNs was proposed. In this method, a detection followed by a regression cascade strategy was used, where in a first stage heatmaps are inferred to detect the human body parts and then a regression of these heatmaps is performed to learn body relationships. In this work, heatmaps are confidence maps that represent the belief that a particular body part occurs at each pixel of the image. In [5], this heatmaps concept was also used, where a network with two branches was designed to learn both part locations (heatmaps learning) and their associations (Part Affinity Fields - PAF). In this sense, instead of using regression of the heatmaps to learn body relations, the association of the different body parts are simultaneous learned with the heatmap inference.

Although several methods have been proposed for human body pose detection, these methods were applied mostly for pose estimation in open spaces, with the in-car scenario having received little attention in the research world [9, 25, 26].

4 Methods

The main goal of the proposed method was to accurately detect the human body pose of passengers inside a car in depth images. Figure 1 presents an overview of the proposed method, which is based on the deep learning-based body pose estimation method presented in [5]. The proposed method uses as input a depth image of the driver acquired from a time-of-flight camera (Fig. 1A) and as output the location of each joint of the different human body parts (Fig. 1D). To obtain the joint positions, a CNN is used to simultaneous predict a set of heatmaps (one for each body part joint, Fig. 1B) and a set of part affinity field vectors that represent the association between the different parts (Fig. 1C). Since the in-car environment produces occlusions of some body parts and the Field-of-View (FoV) of commercial depth sensors may not be enough to visualize all joints once the driver stands near the camera, some joints may not be detectable in the images. Thus, the proposed network also predicts if a joint is presented in the image or not (henceforward called as label detection branch), which may boost the method's robustness.

In Fig. 2, the architecture used for the convolutional neural network is presented. As shown in the figure, the first part of the convolutional network consists in the first ten layers of the VGG-19 [27], which are used to perform a first analysis of the image, generating a set of feature maps $\boldsymbol{F}$. Next, the network is split in three branches for the simultaneous learning of the heat maps, the part affinity fields, and the label detection.

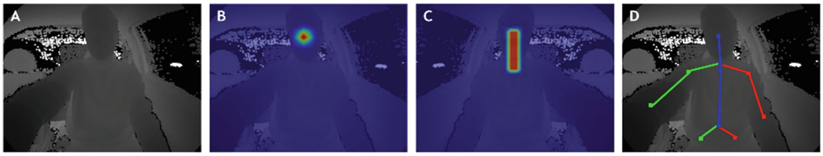


Fig. 1. Overview of the proposed method. (A) Input depth image; (B) Heat map (output of first branch) for the head joint; (C) Part affinity fields (output of second branch) for the association between head and neck joints; (D) final human body pose estimation

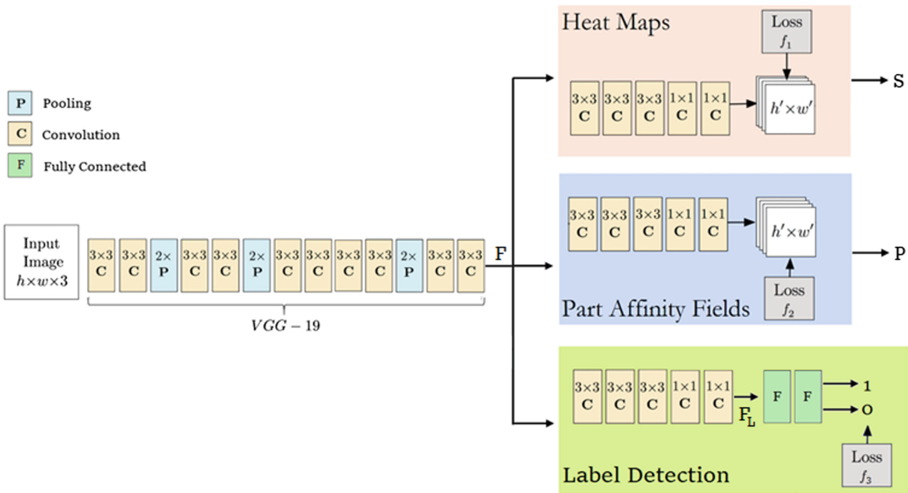


Fig. 2. Architecture of the convolutional network used in the proposed method. The coral branch concerns the learning of the heat maps for body parts' detection and the blue branch concerns the part affinity fields for body part associations. Finally, the green branch is related to the label detection for joint categorization. Adapted from [5]. (Color figure online)

4.1 Heat Maps for Body Parts' Detection

One of the branches of the convolutional network is used to predict confidence maps of each human body part (coral branch showed in Fig. 2). As previously stated, a confidence map represents the belief that a body joint occurs in a given image pixel, being assigned to each pixel a probability of being a body joint. In this sense, a confidence map can be seen as a 2D gaussian-like function, where the maximum of the gaussian map represents the ideal joint position.

To train the method to predict heat maps, a loss function f_{heat} was applied in the end of this branch to calculate the difference between predictions and ground truth. In this case, the ground truth for the confidence maps was generated using a manual labeling of the joint positions and constructing a gaussian map around the joint locations. The loss function for this branch is given by:

$$f_1 = \sum_{j=1}^{J} ||S_j - S_j^*||, \tag{1}$$

where J represents the number of joints, and S_j and S_j^* are the prediction and ground truth maps for part j, respectively.

In the test phase, the joint position for each body part is given by the maximum of the respective confidence map, *i.e.* its peak, after a non-maximum suppression.

4.2 Part Affinity Fields for Body Part Association

To increase the accuracy of the body part detection, a second branch that measures the association between each pair of body parts was added to the convolutional network (blue branch in Fig. 2). This association is given by part affinity fields, which consists in a vector field between two body joints that encodes the direction between one body joint to another (see Fig. 1). To better understand how the part affinity fields are generated, the reader is kindly directed to [5].

The loss function associated to this branch is given by:

$$f_2 = \sum_{c=1}^{C} ||P_c - P_c^*||, \tag{2}$$

where C represents the number of connections between the different body parts,P_c is the prediction of a part affinity field and P_c^* is the ground truth for the association.

Besides refining the inference of the confidence maps during training, owing to the backpropagation scheme used during it, the part affinity fields are useful when there is more than one person in the image. Indeed, in this scenario, a confidence map by itself may not be enough for an accurate detection because several peaks for the same joint may be detected (one for each person in the image). In this sense, the association between body parts are crucial to understand which joints belong to the same person. However, in this work, and given the FoV of the camera used, we only focused in the detection of one person, and therefore, we only used the part affinity fields to refine the heat map prediction.

4.3 Label Detection for Joint Categorization

Owing to the limited size of an in-car environment and to the reduced FoV of the cameras used for monitoring in this environment, there is a higher probability of certain body joints being outside of the image, specially the extremity limbs (*e.g.* the driver can have its arm outside of the lateral window and the associated joint is therefore not present in the image). It is thus important to understand if the joint is present or not in the image to increase the accuracy of the human body pose estimation. To deal with this problem, we added a third branch to our network. This third branch allow us to categorize the joint with a different label according to its existence in the image, i.e. the joint has the label 0 if it is outside of the image and label 1 if it is inside.

This label detection was achieved by using a set of fully-connected layers to learn non-linear combinations of the features F_L extracted by the convolutional layers (green branch in Fig. 2). Afterwards, a softmax layer was used to assign a probability for the label detection, by taking the output of the fully-connected layers and transforming it into a vector with two prediction scores (one for each class i: presented in the image or not). Each prediction score is given by:

$$p_{i,j} = \frac{e^{a_{i,j}}}{\sum_{k=1}^{K} e^{a_{k,j}}}, i = 1, \ldots, K \tag{3}$$

where a_i represents the output of the last fully-connected layer for class i, K corresponds to the number of classes (in this case $K = 2$), and J represents the number of joints. Please note that the sum of the probabilities is equal to 1 and, therefore, the joint's presence label is given by selecting the class with higher score. The loss function for this branch is given by Eq. (4), where t_k is the ground truth for the probability of each class (0 or 1).

$$f_3 = \sum_{j=1}^{J} -\frac{1}{K} \sum_{k=1}^{K} \log(p_k, t_k). \tag{4}$$

At test time, the prediction of this third branch is used to verify which heat maps predicted in the first branch should be evaluated. If the label detection branch predicts that the joint is not present in the image, it is considered not detected and the respective heap map is not evaluated. Otherwise, it is assumed that the joint is presented in the image and its position is given by the maximum of the respective heat map, as stated in Sect. 4.1.

The final human body pose estimation is obtained by combining the output of the three branches. Thus, the overall objective function is given by summing the loss of each branch:

$$f = f_1 + f_2 + f_3. \tag{5}$$

5 Experiments

5.1 Dataset Creation

Due to the inexistence of public datasets of depth images in an in-car scenario, it was needed to create our own dataset. Owing to the deep learning nature of the method, a massive amount of data is needed as training data, and therefore, it was needed to create a large dataset. Moreover, besides the high number of training images, the training dataset must also be variable enough to include the large number of actions possible in this scenario. In fact, the accuracy of the method is very dependent of the quality of the dataset. In this sense, we constructed our dataset by acquiring depth images using a time-of-flight camera placed near the windshield in front of the driver. For the construction of the dataset, ten different cars were used to achieve the desired variability in terms of image background. For each one of the ten cars, five subjects acted as driver, performing different actions inside a car (*e.g*. driving, putting the seat belt, picking up the phone, and others) to give the robustness needed for the dataset. The combination of the different cars and different drivers allowed to construct a dataset with 12200 depth images. The dataset was then divided in training, validation, and testing set with 8820, 1730, and 1650 images each, respectively. In this work, the training dataset was used to train the method, the validation set was used to test the progress of the performance of the method during training, and the testing set was used for the final validation of the proposed method. Note that the cars in each set of images differ from each other to achieve an unbiased evaluation.

Concerning the ground truth for the different body parts, it was constructed by manual labelling of the joint positions. Fourteen joints were used, namely: head (H), neck (N), right/left shoulder (RS/LS), right/left elbow (RE/LE), right/left wrist (RW/LW), chest (C), pelvis (P), and right/left hip (RH/LH). Note that the performance of the method was only evaluated for these upper body joints once the lower body parts are naturally occluded when simulating a driving position. Moreover, the joint categorization (present or not) was also manually defined per image.

5.2 Data Augmentation

As above-mentioned, the ground-truth for the training dataset was obtained manually, which could represent a limitation in respect to the number of training images once manual labelling is a tedious and time-consuming task. In this sense, besides the real dataset constructed, and as common in deep learning strategies, a data augmentation layer was implemented, allowing to generate more training images that the ones initially labelled. Moreover, this data augmentation strategy allowed to increase the variability of the training dataset. Traditionally, data augmentation strategies rely on image flip, rotation, and scaling. Although it is an effective way of increasing image variability during training, such strategies do not modify the intensity information of the image. Although such feature is not so problematic for RGB images, it can for depth ones, as changing the intensity of these images can be a useful way to simulate different camera positions in the world (i.e. the distance between the camera and objects). In this work, the implemented data augmentation layer can simulate these changes in terms of camera's position, allowing to create images where the objects (i.e., the driver) are closer or farther from the camera than they were in fact.

The first step of our data augmentation approach is to convert the 2D depth image in a 3D point-cloud, using the extrinsic parameters of the camera. Upon obtaining the 3D camera coordinates, all points can be transformed by applying a given translation, moving the 3D point cloud in any direction and in any axis. The final step consists in using the intrinsic camera parameters to transform the translated point-cloud into a new 2D depth image. In Fig. 3, it is possible to visualize two examples of the result of our data augmentation strategy: one simulates the camera closer to the driver (Fig. 3B) and the other simulates the camera located farther from the driver (Fig. 3C).

Fig. 3. Data augmentation strategy. (A) Original depth image; (B) Augmented image that simulates the positioning of the camera closer to the driver; (C) Augmented image that simulates the camera being located farther from the driver

5.3 Implementation Details

In the implementation of any deep learning method, the definition and optimization of certain training parameters for each specific problem are fundamental and can have a significant importance in the final accuracy of the method. In this work, experiments were carried out to get the optimal settings for these parameters. Concerning the learning rate and the momentum, these parameters were experimentally set to 0.0004 and 0.9, respectively. For the model optimization, the Adam solver was used with a regularization term of 0.01. For the batch size, 10 images were used. Note that these parameters were chosen by evaluating the method in the validation dataset.

6 Results

One important task to correctly implement a deep learning strategy is to evaluate the progress of the training in the validation dataset. Besides being useful to conduct experiments related with the best parameters to be used in the deep learning strategy, the validation dataset is also needed to avoid problems like overfitting to the training data, which may cause failure of the method when applied in a different set of images. In this sense, evaluating the method's performance during training in the validation dataset allow us to detect when the training converges, avoiding overfitting problems. Figure 4 presents an example graph showing the progression of the loss during training in both training and validation datasets.

To evaluate the performance of the proposed method in the testing set, different evaluation metrics were used. The first metric consists in the distance (D) between the detected joint and the ground truth in pixels. Moreover, the method's accuracy (Ac, percentage of correctly detected joints), the precision (Pr, fraction of correctly detected joints among all the correctly or wrongly detected joints) and the recall (Re, fraction of correctly detected joints among all the correctly detected joints and the wrongly non-detected joints) were also evaluated, being defined by Eqs. (6) to (8).

$$Ac = \frac{T_P + T_N}{N} \times 100 \quad (6)$$

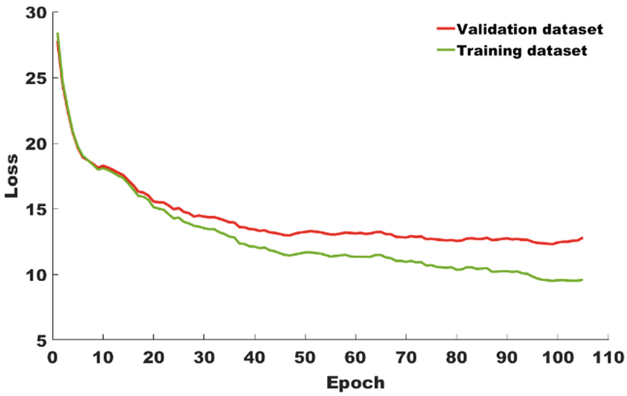


Fig. 4. Method's performance (loss in function of epochs) during training in the training (green line) and validation (red line) datasets. (Color figure online)

$$Pr = \frac{T_P}{T_P + F_P} \times 100 \tag{7}$$

$$Re = \frac{T_P}{T_P + F_N} \times 100 \tag{8}$$

where T_P,T_N, F_P, and F_N correspond to the true positives, true negatives, false positives, and false negatives, respectively, and N is the total number of testing images.

Table 1 summarizes the method's performance for each joint on the testing dataset. The presented values for the distance metric correspond to the median of the distance errors obtained. Note that the information if a joint is correctly detected or not used in the estimation of the accuracy, precision, and recall is given by the third branch of the proposed network. Moreover, the results showed in the table were obtained using the model obtained in the ideal epoch for early training stopping, which were estimated using Fig. 4. In Fig. 5, the boxplots for the distance metric are presented to analyze the detection performance distribution in terms of percentiles and outlier points. Figure 6 presents the evolution of the accuracy, precision, and recall for different distance thresholds (i.e. not using the information provided by the label categorization). In this case, a joint was considered correctly detected only if its distance error is bellow a given threshold. Finally, in Fig. 7, some example results of the proposed strategy for human body pose estimation are presented.

Table 1. Method's performance in the testing dataset, assessed in terms of distance error (*D*, pixels), accuracy (*Ac*, %), precision (*Pr*, %), and recall (*Re*, %).

	H	N	RS	RE	RW	RS	RE	RW	C	P	RH	LH	Mean
D	6,0	5,1	5,4	6,7	10,8	6,7	6,4	10,0	9,2	7,2	8,6	9,1	7,6
Ac	98,8	99,6	99,8	97,8	85,0	98,5	96,8	88,1	99,1	98,8	92,3	92,4	95,6
Pr	99,1	99,7	99,8	98,7	82,6	98,8	97,8	83,9	99,4	99,0	96,8	96,6	96,0
Re	99,7	99,9	100,0	99,1	96,0	99,8	98,9	90,4	99,8	99,8	95,1	95,5	97,8

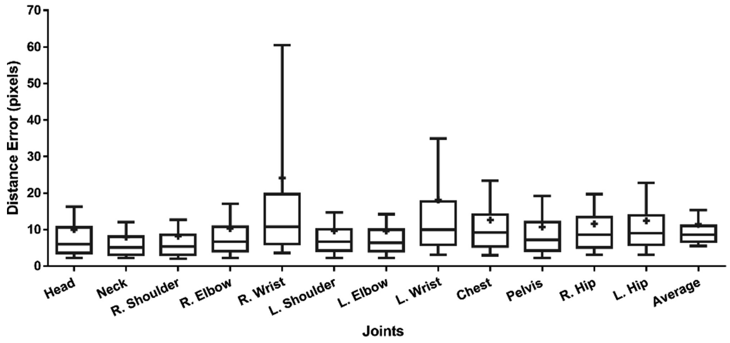


Fig. 5. Boxplots for the distance errors obtained by the proposed method per joint. The ends of the whiskers represent the 10th and 90th percentiles and the crosses represent the mean values.

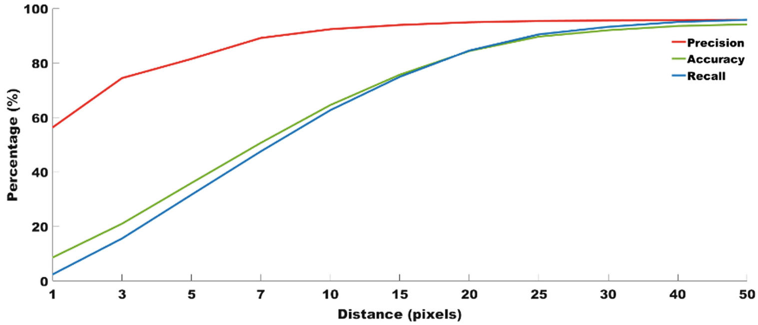


Fig. 6. Variation of accuracy (green line), precision (red line), and recall (blue line) for different distance thresholds in pixels. (Color figure online)

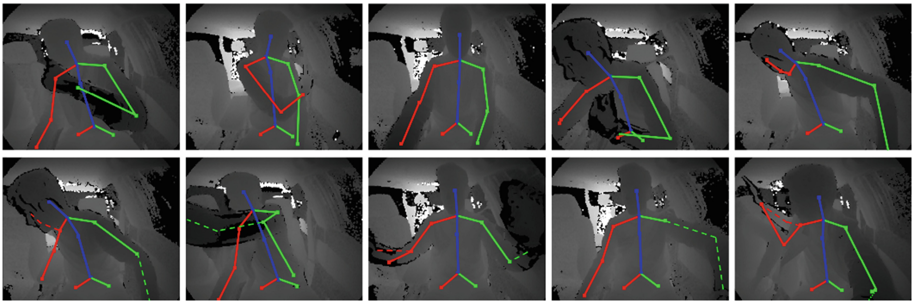


Fig. 7. Qualitative results of the proposed human body detection method. The first row presents examples of good results. In the second row, some examples where the pose estimation failed for a few joints are illustrated, with the ground truth pose shown in dashed lines

7 Discussion

This paper proposed a method for human body pose estimation in an in-car scenario. As stated, an important study to be performed during the implementation of a deep learning strategy concerns the evolution of the training. For that, its performance during training must be evaluated both in the training and validation datasets. Analyzing Fig. 4, it is possible to visualize that the ideal timing for stopping the training would be approximately around the 60th epoch. After this point of the training, the graph suggests that the model may start suffering from overfit to the training dataset, resulting in a very good performance for the images presented in this dataset but a lower or equal performance in the validation dataset.

After analyzing the ideal epoch for early stopping, the model obtained in this epoch was used for the final validation of the method in the testing set. Analyzing Table 1 it is possible to verify the good performance of the human body pose detection method, with a mean distance error for all joints of 7.6 pixels. Moreover, high values for the accuracy, precision, and recall were obtained, showing the method's robustness. The worst results were obtained for the wrist joints. This can be explained by the fact that these joints are frequently near the image's limits, which can lead to a lower accuracy of the deep learning strategy. Moreover, owing to the proximity of the camera to the driver, these joints are not always present in the camera' field-of-view, which hampers the training process for these joints. This less accurate detection for these extremity joints can also be visualized in Fig. 5, where it is possible to verify the larger mean and interquartile range for these joints. Nevertheless, Fig. 5 corroborates the overall good performance of the method, with narrow boxplots being obtained for the majority of the joints. In Fig. 6, it is shown that the accuracy, precision, and recall are improved with the increase of the threshold for the distance, as expected. Analyzing this graph, one can verify that for a threshold distance of 15 pixels, these evaluation metrics present acceptable values.

Concerning the computational time required by the method, the proposed human body pose detection methods takes approximately 80 ms per image, which gives a performance of 12 frames per second, which proves the nearly real-time capability of the method. Note that this runtime analysis was performed using a laptop with a NVIDIA GeForce GTX-1060 GPU.

One important aspect to take into account in the proposed method is its application in a real setup. In fact, to achieve a real-time estimation of the human body pose inside a car, a high computational power is needed, which can be a limitation for an in-car scenario. However, the constant growth in computation capability of the technology will allow the application of the proposed method in a real scenario. Another aspect to take into account is that the application of software in autonomous cars should follow some existing standards to ensure its applicability in a real scenario. Nevertheless, all these issues will be address in the future.

8 Conclusions and Future Work

This paper presented a framework to detect the human body pose in depth images acquired inside a car. This method consists in a deep learning strategy where a new convolutional network configuration with three branches was used for simultaneous learning of confidence maps for each joint position, body parts' associations, and joint categorization (regarding its existence in the image). The proposed framework was validated in 1650 depth images, achieving an average distance error of 7.6 pixels and an average accuracy, precision, and recall of 95.6%, 96.0%, and 97.8% respectively. Overall, the proposed human body pose estimation method proved to be successful and accurate to detect the driver's pose, while showing a nearly real-time performance.

In future work, we intend to expand the human body pose estimation to more than one person, allowing to monitor not only the driver but all the car' occupants. In addition, we also intend to modify the method to exploit the 3D information provided by the depth image to infer the 3D joint position.

Acknowledgements. This work is supported by: European Structural and Investment Funds in the FEDER component, through the Operational Competitiveness and Internationalization Programme (COMPETE 2020) [Project nº 002797; Funding Reference: POCI-01-0247-FEDER-002797].

References

1. Levinson, J., et al.: Towards fully autonomous driving: systems and algorithms. In: IEEE Intelligent Vehicles Symposium, pp. 163–168 (2011)
2. Banks, V.A., Stanton, N.A.: Analysis of driver roles: modelling the changing role of the driver in automated driving systems using EAST analysis of driver roles: modelling the changing role of the driver in automated driving systems using EAST. Theor. Issues Ergon. Sci. 1–17 (2017)
3. Regazzoni, D., De Vecchi, G., Rizzi, C.: RGB cams vs RGB-D sensors: Low cost motion capture technologies performances and limitations. J. Manuf. Syst. **33**(4), 719–728 (2014)
4. Shao, L., Han, J., Xu, D., Shotton, J.: Computer vision for RGB-D sensors: Kinect and its applications. IEEE Trans. Cybern. **43**(5), 1314–1317 (2013)
5. Cao, Z., Simon, T., Wei, S.-E., Sheikh, Y.: Realtime multi-person 2D pose estimation using part affinity fields. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7291–7299 (2017)
6. Casner, B.Y.S.M., Hutchins, E.L., Norman, D.O.N., Promise, A.C.: The challenges of partially automated driving. In: Communications of the ACM, pp. 70–77 (2016)
7. Fagnant, D.J., Kockelman, K.: Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations. Transp. Res. Part A **77**, 167–181 (2015)
8. Krueger, R., Rashidi, T.H., Rose, J.M.: Preferences for shared autonomous vehicles. Transp. Res. Part C Emerg. Technol. **69**, 343–355 (2016)
9. Demirdjian, D., Varri, C.: Driver pose estimation with 3D Time-of-Flight sensor. In: 2009 IEEE Workshop on Computational Intelligence in Vehicles and Vehicular Systems, pp. 16–22 (2009)
10. Ye, M., Yang, R.: Real-time simultaneous pose and shape estimation for articulated objects using a single depth camera. In: CVPR 2014, pp. 2345–2352 (2014)

11. Ye, M., Wang, X., Yang, R., Ren, L., Pollefeys, M.: Accurate 3D pose estimation from a single depth image. In: 2011 International Conference on Computer Vision, pp. 731–738 (2011)
12. Sigalas, M., Pateraki, M., Trahanias, P.: Full-body pose tracking? The top view reprojection approach. IEEE Trans. Pattern Anal. Mach. Intell. **38**(8), 1569–1582 (2016)
13. Shotton, J., et al.: Real-time human pose recognition in parts from single depth images. Commun. ACM **56**(1), 116–124 (2013)
14. Shotton, J., et al.: Efficient human pose estimation from single depth images. IEEE Trans. Pattern Anal. Mach. Intell. **35**(12), 2821–2840 (2013)
15. Tsai, M.-H., Chen, K.-H., Lin, I.-C.: Real-time upper body pose estimation from depth images. In: 2015 IEEE International Conference on Image Processing (ICIP), pp. 2234–2238 (2015)
16. Buys, K., Cagniart, C., Baksheev, A., De Laet, T., De Schutter, J., Pantofaru, C.: An adaptable system for RGB-D based human body detection and pose estimation. J. Vis. Commun. Image Represent. **25**(1), 39–52 (2014)
17. Haque, A., Peng, B., Luo, Z., Alahi, A., Yeung, S., Fei-Fei, L.: Towards Viewpoint Invariant 3D Human Pose Estimation, pp. 160–177. Springer, Cham (2016)
18. Belagiannis, V., Zisserman, A., Group, V.G.: Recurrent human pose estimation. In: 12th IEEE International Conference on Automatic Face & Gesture Recognition (2017)
19. Chu, X., Yang, W., Ouyang, W., Ma, C., Yuille, A.L., Wang, X.: Multi-context attention for human pose estimation, pp. 1831–1840. arXiv preprint arXiv:1702.07432
20. He, K, Gkioxari, G, Dollár, P, Girshick, R.: Mask R-CNN. In: Computer Vision (ICCV), pp. 2980–2988 (2017)
21. Chen, X., Yuille, A.: Articulated pose estimation by a graphical model with image dependent pairwise relations. In: Conference on Neural Information Processing Systems, pp. 1–9 (2014)
22. Tompson, J., Jain, A., Lecun, Y., Bregler, C.: Joint training of a convolutional network and a graphical model for human pose estimation. In: Advances in Neural Information Processing Systems, pp. 1–9 (2014)
23. Fan, X., Zheng, K., Lin, Y., Wang, S.: Combining Local Appearance and Holistic View: Dual-Source Deep Neural Networks for Human Pose Estimation
24. Bulat, A., Tzimiropoulos, G.: Human pose estimation via convolutional part heatmap regression. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9911, pp. 717–732. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46478-7_44
25. Borghi, G.: POSEidon: Face-from-depth for driver pose estimation. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5494–5503 (2017)
26. Murthy, P., Kovalenko, O., Elhayek, A., Gava, C., Stricker, D.: 3D Human Pose Tracking inside Car using Single RGB Spherical Camera (2017)
27. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: Computer Vision and Pattern Recognition, pp. 1–14 (2014)

Treatment of Ventricular Assist Device Test Bench Data for Prediction of Failures and Improved Intrinsic Reliability

Jeferson C. Dias(✉), Jônatas C. Dias(✉), Edinei Legaspe(✉), Rodrigo Lima Stoeterau(✉), Fabrício Junqueira(✉), Newton Maruyama(✉), Lucas Antônio Moscato, Paulo E. Miyagi(✉), and Diolino J. Santos Filho(✉)

Polytechnic School of the University of São Paulo, São Paulo, SP, Brazil
{jefersondiasl, rodrigo.stoeterau, fabri, maruyama, pemiyagi, diolinos}@usp.br, jxdias@ymail.com, edineilegaspe@hotmail.com

Abstract. This article regards over analytics of reliability of ventricular assist devices (VAD) used as therapy for advanced heart failure conditions in the face of malfunction related adverse events. This question directs research and the search for a solution proposal, even if prospective, but that promotes the longevity of these devices, increasing the intrinsic reliability. An "In Vitro" test bench is used to obtain variations of dynamic behavior over time; by means of a set of variables and the deviations (failures) compared between the standard and tested devices; since these devices are systems that vary in time. An intelligent systematics obtained through the automation of the test bench, using sensors and actuators to control the independent variables, and the data collection and analysis using the technologies present in the industry 4.0 completes the increase of the reliability of the VAD.

Keywords: Cardiovascular · Ventricular Assist Device (VAD) · In Vitro · Big Data Analytics · Industry 4.0 · Test bench

1 Introduction

Among cardiovascular diseases (CVDs), heart failure (HF) has a remarkable prevalence and a great impact on morbidity and mortality worldwide [1]. HF is considered to be a syndrome that renders the heart unable to supply oxygen at an adequate rate to the tissues, or at the expense of increasing the filling pressure (preload). Generally, HF results from structural or functional dysfunction of the heart, which compromises the ability to fill with blood or to eject it [2].

For refractory HF cases and when no further treatment is available, cardiac transplantation (CT) is considered the most effective therapeutic modality to significantly prolong the patient's life [1, 3, 4]. However, there is a fundamental problem in the practice of organ transplantation, according to the "Thirty-fourth Adult Heart Transplantation Report—2017" [5], the number of cardiac surgeries done for transplants in

Published by Springer Nature Switzerland AG 2019
L. M. Camarinha-Matos et al. (Eds.): DoCEIS 2019, IFIP AICT 553, pp. 183–190, 2019.
https://doi.org/10.1007/978-3-030-17771-3_15

Europe, for example, does not grow significantly. For the rest of the world, just as in the US inclusive, there is a growing trend, but it is not enough to balance supply and demand. Drug therapy alone for refractory HF continues to have disappointing results. Because of this, VADs have emerged as a therapy for refractory HF and are used, for example, as a long-term use strategy, to provide conditions so that the patient can receive the heart transplant at the expected time, a practice called by "bridge to the transplant", or if the patient is not able to perform the surgery because of a high probability of rejecting the transplant, VAD is recommended as a permanent practice known as "destination therapy" [1].

Although VADs have become an advantageous opportunity in refractory HF therapy, this alternative has opportunities for improvement in the face of adverse events caused by the use of these devices. For example, some of the main adverse events are: bleeding, infection, stroke and malfunction, the latter event being directly related to the failure of the device functionality, according to the *Interagency Registry for Mechanically Assisted Circulatory Support*[1] in its eighth annual INTERMACS report [6]. This leads us to reflect on the need to record reliability data and improve the reliability of these devices.

2 Relation with Industrial and Service Systems

The technological evolution observed in recent years; mainly related to: (a) the digitalization that transforms the way in which we collect data to transform into useful information for decision making, (b) the creation of increasingly smaller sensors from microelectronics with innovations in microchips and (c) the technologies of communication networks wireless; has provided the development of products and processes in various market segments, from consumer goods to industrial automation systems [7].

This prospective article proposes the use of this technological evolution in the improvement of mechatronic devices, as is the case of VAD. The technological evolution based on multidisciplinary concepts that created the mechatronic devices, when applied in a project, and in this case for VADs, with resources that allow the communication through Internet (IoT), robotics, software and manufacturing accompanied by an integrated digital system, it was possible to provide the creation of smarter devices [8].

This paper highlights how this technological innovation can be used to improvement of VADs of continuous flow and with the mechanism of operation of the pump of the centrifugal type. The importance of technological application to improve these devices is justified by two reasons: (a) To predict and avoid adverse events attributed to malfunction and (b) The increasing use of VADs as destination therapy.

[1] INTERMACS - Interagency Registry for Mechanically Assisted Circulatory Support contemplates a database with a total of 22,866 patients who received a mechanical circulatory support device approved by the FDA. With registration from 06/23/2006 to 12/31/2016 of 185 participating hospitals.

3 Objective

To propose and test a method of improving the intrinsic reliability of ventricular assist devices (VADs) using: (i) an automated test bench and (ii) a knowledge base, allowing data-based decision making using the technologies present in the industry 4.0 such as: Cloud computing, Internet of Things (IoT) and technologies linked to Artificial Intelligence (AI) present in tools of Big Data Analytics and Data Mining.

The experimental performance, according to the proposed method, will result in: (a) a knowledge base, coming from the data of failure of the behavior of a VAD throughout the tests, acquired knowledge, will allow intelligent control to predict future consequences and (b) a knowledge analysis obtained will result in a set of improvements, which will be applied to the design of new devices, making these devices more durable.

4 Materials and Methods

Intrinsic reliability is not simple to achieve in time-varying devices, such as a VAD. For this reason, a five-stage cyclic process was developed, as proposed in Fig. 1. "4. VAD Test" is described in detail in the following sections. A VAD project, step 1, is drawn from demand requirements, in which it is evaluated in step "2. Risk Analysis" in such a way as to enable it to provide sufficient reliability for rapid prototyping production, provided in step "3. Production (3D-PETG)", of a bench test device in step "4. VAD Test on Bench".

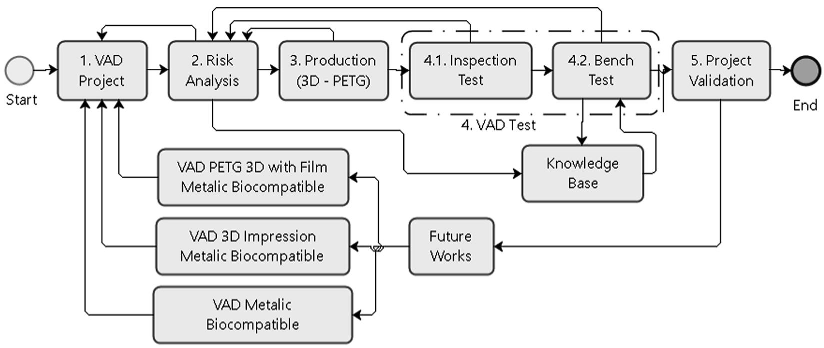


Fig. 1. VADs validation process. **Source**: Dias, J. C. 2019 (Author)

In these steps 1 through 4, the risk analysis evaluates how reliable the project, production, or testing is to then allow the project to be validated in step "5. Project Validation". This last stage guarantees the intrinsic reliability of the device, as well as its respective design, allowing a future verification to verify the reduction of adverse events due to the malfunction of these devices. A "Knowledge Base" in the cloud collects information for adjusting control parameters of the actuators in the test bench, so that the "Bench Test Procedure" described in the section below can be controlled [9].

4.1 VADs Test Bench

The test bench used in this work consists of the reservoirs 1 and 2 which store the liquid used is transported from tank T1 to tank T2, performed from a pump (VAD), which imposes energy on the system, according to Fig. 2.

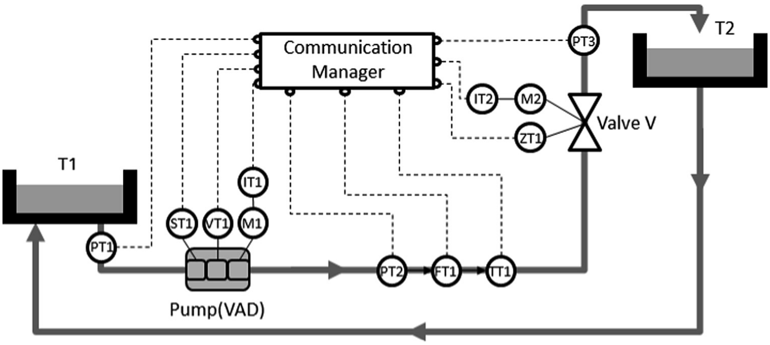


Fig. 2. Scheme of the "Test Bench" existing in the Department of Mechatronics of USP. **Source**: Dias, J. C. 2019 (Author)

Sensors perform the collection of fluid pressure at the output of reservoirs 1, rotation, vibration and electrical current of the motor of the VAD, fluid pressure, flow and temperature at the output of the VAD, position and the electric current of the V-valve motor 2 and the pressure of the liquid at the inlet of the reservoir 2. The test bench has its own motor to ensure control of the rotation of the blood pump and the motor 2 that contours the valve V, so it is possible to control the flow of the liquid into the system.

4.2 Collection and Processing of Data

To collect information from the "context" [10] of the assisted device, emitted by the sensors in the "Test Bench", Fig. 2, and impose the new references to the actuators, a communication manager module is required for distribution and targeting of the collected data to the cloud knowledge base. These data from the sensors are processed for real-time error checking (failures), at which time the adjustment decisions are made. This knowledge base is handled by a standards analyzer for knowledge discovery using "Big Data Analytics". In this article "context" is any information that can be used to characterize the situation of an entity (object of interest), it is also known as sentient computing [10].

4.3 Knowledge Base Data Analysis

Big Data Analytics approaches to pattern discovery are based on: (i) DESCRIPTIVE ANALYSIS: It is based on real-time understanding of events to make immediate decisions. Descriptive analysis works with historical data, cross-information to generate a clear and precise understanding of the relevant topics for the present moment, it is not necessary to relate past or future patterns [11]. (ii) DIAGNOSTIC ANALYSIS: The objective is to understand the cause and effect relationship over time. The diagnostic analysis works based on the collection of data related to a certain subject and crosses information to understand which factors influenced the current outcome [11]. (iii) PREDICTIVE ANALYSIS: It uses data mining as: statistical and historical data to know future trends [11]. (iv) PRESCRIPTION ANALYSIS: Prescriptive analysis commonly confused with predictive analysis, because it works with the same logic. However, for different purposes, while the predictive analysis identifies future trends, the prescriptive delineates the possible consequences of each action [11].

4.4 Inspection Test Procedure

The inspection test, procedure 4.1 of Fig. 1, required for continuity in the following tests, is performed by means of roughness recognition through a top image. The method uses top images of calibrated roughness patterns. From these images computational information is obtained that allows the characterization of the surface and recognition of the roughness, according to technique of surface profilometry with and without contact.

4.5 Test Bench Procedure

The procedure “4.2 Test Bench”, Fig. 1, occurs in three phases:

4.5.1 Standard Curve Survey

The VAD is fixed in the test bench to perform data collection of independent variables (velocity and rotor flow) that are manipulated with progressive values adjustment, impacting the behavior of the dependent variables (pressure, vibration, temperature and energy consumption). The data collection cycle (dependent variable values) begins with the fixed rotation speed for all flow variations. Thus, for each variation of the flow, the respective values of the dependent variables are collected.

After having performed all the variations of values referring to the flow variable; with the respective collection (reading) of the dependent variables (already mentioned); a new configuration is performed on the rotor speed variable and a new cycle is started. A pre-programming is established in the system that controls the execution of tests on the devices; for example: configuration speed ranges and flow variations. At the end of the test, the device is able to perform the next tests. Gathering information during this step is useful for improving the default curve for this type of device.

Once the various standard curve tests have been performed considering the same type of device, the knowledge base for the behavior of dependent and independent variables will be used to observe deviations in the following test phases.

4.5.2 Short Term Testing

The short duration test consists of continuous repeatability of the standard curve test until the first failure is obtained, at which time the failure occurrence is recorded in the database. Thereafter, the pump is disassembled, so that each pump part goes through the surface inspection protocols. The defective part is replaced, just as the identified fault is recorded in the database. After identifying the causes of failure and risk analysis, the pump returns to the bench test until at least one failure is recorded for each pump part (VAD).

4.5.3 Long-Term Testing

Once the standard curve of the same type of device is defined, within its variables any observed deviations will be recorded and plotted on the graph. For the operating limits of the independent variables, only the speed limits are defined. Thus, tissue perfusion values simulated by speed limits will be appropriate for the human. Throughout the test deviations of the behavior of the variables can occur in relation to those registered in the standard curve.

Before the simulated tissue perfusion (independent variables) in the test bench, during the long-term tests become impracticable, through self-adjustments of the dependent variables, or a critical failure occurs, an AI manager module interrupts the device. The purpose of the long-term test is to obtain as much fault data as possible within the operating range of dependent and independent variables for a simulated human perfusion.

5 Results and Discussion

At the moment of the implantation of the method proposed in this work, according to Fig. 1, a VAD project was already underway, and a VAD made of aluminum with subtractive technology had already undergone a bench test; with the occurrence of failure. This is because the magnetic rotor had scraped on the upper casing, damaging the VAD. For this reason, the validation process of VADs was put into practice in the process “2. Risk Analysis” in which he submitted the damaged VAD parts to the “4.1 Inspection Test process”.

The analytical procedure presents in the process “2. Risk Analysis”, identified a fault in the bearing of the lower housing. A callosity at the center of the bearing was identified as the fault that caused the magnetic rotor to misalign, damaging the upper housing of the tested device.

An initial report was provided with improvement items, as Table 1, indicating the main points for increasing intrinsic reliability.

The improvement items also caused a change in the design of the VAD, process “1. VAD design” resulting in the construction of the base of the mobile cradle for adjusting the lower bearing clearance, with tightening torque connection, as Fig. 3.

Table 1. Improvement items proposed during phase "2. Risk analysis"

PROBLEM	EFFECTS	CAUSE	SOLUTION
Bearing failure	Engine rotation error Change of pump flow Motor torque friction Rotor Locking	1. Bearing inclination	**R1.1** Ensure the production according to the project. **R1.2** Ensure compliance with the assembly procedure. **R1.3** Overall geometric and dimensional surface analysis with test protocols.
		2. Clearance between bearing and cradle	**R2.1** Make the cradle base mobile for adjusting the lower bearing clearance with wrench connection for clearance adjustment. **R2.2** Ensure compliance with the assembly procedure.
		3. Deformity in bearing cradle (manufacturing flha)	**R3.1** Ensure manufacturing according to design. **R3.2** Perform surface, geometric and dimensional analysis with test protocols.

Source: Dias, J. C. 2019 (Author)

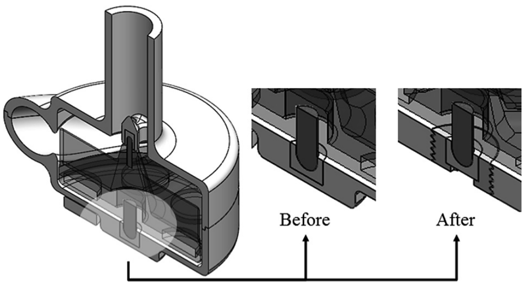


Fig. 3. Result of project improvements. **Source**: Dias, J. C. 2019 (Author)

6 Conclusions

This prospective project, still in progress of its finalization, has already been able to demonstrate its positive effects, according to results section. The next steps include the implementation of intelligent control of the test bench with the use of the knowledge base for the prescription and prediction of events for the decisions of adjustment of the variables of the bench to the device in test. Thus, during every test step, the collected data is stored for an evaluation of the failures. The data collected, when analyzed using Big Data Anlytics, allow the prediction of failures for analyzed sample. Thus, when performing the evaluation in the new devices tested for fault predictivity, based on the information acquired, it is possible to make a decision before the failures appear. This assists in the longevity of the device being tested, in a similar human perfusion condition.

With regard to the gap presented in the INTERMACS report on adverse events related to device malfunction, this work seeks to improve reliability by reducing failures that may cause malfunction of these devices. Thus, the tangible contributions are first divided into a test bench with an intelligent control system, improving from the collected data, and assisting in a geographical expansion in the treatment of failures of other distributed devices. And the second contribution is a verification and validation of devices, from the knowledge of stored reliability data.

Acknowledgements. The authors thanks FAPESP, CNPQ e CAPES for supporting this research.

References

1. Magalhães, C.C., Serrano Jr., C.V., Consolim-Colombo, F.M., Nobre, F., Fonseca, F.A.H.: Tratado de Cardiologia SOCESP, 3rd edn. Manole, Barueri (2015)
2. Martins, M.A., Carrilho, F.J., Alves, V.F., Castilho, E.: Clínica Médica, Volume 2: Doenças Cardiovasculares, Doenças Respiratórias, Emergências e Terapia Intensiva, 2nd edn. Manole, Barueri (2015)
3. Moreira, M.V., Montenegro, S.T., Paola, A.A.V.: Livro-texto da Sociedade Brasileira de Cardiologia, 2nd edn. Manole, Barueri (2015)
4. Azeka, E., Jatene, M.B., Jatene, I.B., Horowitz, E.S.K., Branco,. K.C.: I Diretriz de Insuficiência Cardíaca (IC) e Transplante Cardíaco, no Feto, na Criança e em Adultos Com Cardiopatia Congênita, da Sociedade Brasileira de Cardiologia. Arq. Bras. Cardiol. **103**, 144 (2014)
5. Chambers, D.C., et al.: The registry of the international society for heart and lung transplantation: thirty-fourth adult lung and heart-lung transplantation report—2017; focus theme: allograft ischemic time. J. Hear. Lung Transplant. **36**(10), 1047–1059 (2017)
6. Kirklin, J.K., et al.: Eighth annual INTERMACS report: special focus on framing the impact of adverse events. J. Hear. Lung Transplant. **36**(10), 1080–1086 (2017)
7. Faceli, K., Lorena, A.C., Gama, J., de Carvalho, A.C.P.L.F.: Inteligência Artificial - Uma Abordagem de Aprendizado de Máquina. LTC, Rio de Janeiro (2011)
8. Atoche, A.C., Marrufo, O.P.: Nuevas Tendencias en Sistemas Mecatrónicos. Ingeniería **15** (2) (2011)
9. Dias, J.C., Dias, J.C., Barbosa, M., Miyagi, P.E., Filho, D.J.S.: In vitro test bench with intelligent behavior to ventricular assist devices. In: 15th International Conference on Informatics in Control, Automation and Robotics (ICINCO 2018), vol. 1, pp. 127–134 (2018)
10. de Andrade, L.A.: Jogos Digitais, Cidade e (Trans)mídia: A Próxima Fase, Curitiba (2015)
11. Marquesone, R.: Big Data: Técnicas e tecnologias para extração de valor dos dados. Casa do Código, São Paulo (2016)

Smart Environments

Basis for an Approach to Design Collaborative Cyber-Physical Systems

Artem A. Nazarenko(✉) and Luis M. Camarinha-Matos

Faculty of Sciences and Technology and UNINOVA-CTS,
Nova University of Lisbon, 2829-516 Monte Caparica, Portugal
a.nazarenko@campus.fct.unl.pt, cam@uninova.pt

Abstract. Nowadays Cyber-Physical Systems gain more and more attention in regard to the Industry 4.0 or Digital Transformation in general. These systems imply the tight integration of physical and software components and are becoming more complex, forming highly inter-connected systems-of-systems. Furthermore, as components and subsystems are becoming more intelligent, there is a need for a paradigm shift towards considering them as ecosystems of collaborative entities with growing levels of autonomy. There is, however, the lack of proper methodologies and support frameworks for the design of such systems. In this context a contribution to an approach for the development of Collaborative Cyber-Physical Systems is proposed. It introduces some core definitions, organizational and architectural aspects. The proposed approach is in line with the design science research methodology and is illustrated with some examples.

Keywords: Collaborative Cyber-Physical Systems · Industry 4.0 · Design methodology

1 Introduction

Cyber-Physical Systems (CPS) are becoming complex systems, based on interconnected and tightly coupled hardware and software components with growing level of intelligence. Modern CPS might contain a large number of components, in fact, being a system-of-systems. These new systems are characterized by high levels of intelligence and autonomy of their components and requiring collaboration among those components as well as interaction with outer systems, which led to emergence of the Collaborative Cyber-Physical Systems (CCPS) concept [1]. CCPS focus on improving the networking among components and their joint efforts for accomplishing complex tasks, bridging not only the cyber and physical parts, but also the social one. The physical part of a CCPS consists of sensing and actuating capabilities deployed in a certain environment, while the cyber part ensures the needed abstraction level to integrate the physical devices together and provide high-level functionalities, while also allowing users to be an active part of CCPS.

Modern CCPS face not only technical challenges, but also architectural and organizational issues. Thus the borders of modern systems are extending such that often a CPS can be considered as a complex system or a collaborative ecosystem

Published by Springer Nature Switzerland AG 2019
L. M. Camarinha-Matos et al. (Eds.): DoCEIS 2019, IFIP AICT 553, pp. 193–205, 2019.
https://doi.org/10.1007/978-3-030-17771-3_16

incorporating various elements with different aims, characteristics and even nature. While the smart objects become smarter and more able to accomplish complex tasks, conventional centralized approaches to organize them might limit their autonomy and thus their capabilities. This raises several issues in terms of organization and management of the smart autonomous entities that are combined to form a system. It also requires paying attention to the aspects of collaboration between them and users.

As some smart objects are mobile, their movements will affect the conditions of the whole system or environment, making it difficult to define strict borders for the smart environment to which they are assigned to, leading to a notion of changing or "fluid" borders. Moreover, for the cases of complex structures containing several autonomous environments/systems an approach to manage the mobility of objects, not only in terms of physical positioning, but also in terms of logical belonging is of particular importance. This also raises the issue of organizing collaboration at different scales and hierarchical layers. Another aspect is that systems need to adapt to changing environments through periodic analysis of those changes and context, developing an appropriate adaptation strategy. One small change may cause a "domino" effect on the other systems within the same ecosystem or affecting the successful accomplishment of a certain task.

Those topics are not well-developed in the literature, as most of the research works on CPS and Internet of Things are focused on low-level aspects, such as protocols, communication technologies, energy optimization, etc., and also on specific applications. Due to the variety of possible organizational approaches and involved technologies, the design process of complex CPS might be very challenging. As such there is a need for a methodological and framework support to help designing new CPS/CCPS. In fact, nowadays each time developers build a new system they start from scratch, possibly facing the same difficulties as in other systems. As an example, [2] identifies such difficulties in the telemedicine CPS solutions, stating the need for an environment that supports multirole distributed teams with "interface units" being used to connect monitoring devices to a "surveillance centre" which serves as hub. The proposed approach in this example goes back again to the classical master-slave paradigm, limiting the system's capabilities. In another work devoted to CPS implementation in the energy sector [3], the same approach with one aggregation point collecting all the data from devices deployed all over a building is presented. In these examples, selected just for illustration purposes, developers went through similar difficulties, namely determining how the devices have to be managed, how to support the human-system interaction, coping with different hierarchical layers, etc. In most cases mobility is not even considered. In this context, this work is a part of initiative to provide CPS designers with supportive techniques, strategies, and tools for the development, integration and evaluation of CPS. The aimed methodology and support framework needs to be general in order to be applicable to different application areas.

Based on above mentioned aspects, the following research question is defined:

What could be a suitable set of models and organizational structures to support the increasing complexity of evolving CCPSs?

In order to answer this question, the following hypothesis is set for this work:

The increasing complexity of evolving CCPS can be effectively supported if an appropriate collaborative ecosystem is provided that combines a suitable organizational structure with intelligent adaptation and integration methods. Moreover, a hierarchical organization, supporting horizontal and vertical collaboration can increase autonomy and provide logical sections with higher level of independence, context awareness, enhanced responsibilities and mobility support.

The following sections of the paper present preliminary results in this direction and some discussion on ongoing and future research.

2 Relation to Innovation in Industrial and Service Systems

Cyber-Physical Systems are an integral part of the Industrial and Service systems being utilized in a wide variety of application areas. One of the main innovation trends here is observed in the area of Smart Manufacturing, which strongly relies on the adoption of CPS. From the market point of view, the implementation of CPSs in manufacturing allows more flexible adaptation to demand changes, to product lifetime reduction, as well as support for the collaboration processes among all parties involved in manufacturing and related supply chains [4]. Furthermore, using CPS in manufacturing offers new capabilities for companies to provide services throughout the life cycle of products [5], e.g. post sale support will take new and more effective forms with the introduction of CPSs, which will also induce new business models for these companies. In particular, the notion of smart supply chain covers the entire range of issues related to both provision of the necessary materials and delivery of products to the customers or distributors, which requires effective CPS.

Many related research works appear more focused on industrial use cases, covering for instance the challenges of smart predictive systems [6], human-based [7], fog industrial computing systems [8] and other relevant topics responding to principles set by the Industry 4.0 paradigm. The concept of Industry 4.0 can be described by a number of dimensions, including: vertical integration, horizontal integration, through-engineering, acceleration of manufacturing, digitalization of products and services, and customer involvement in co-creation process [9]. CPS and collaboration aspects are present in all these dimensions, thus a clear case for CCPS.

Another application area for the ideas proposed in this work is the smart home, which is a good example of Service System. A smart home can be seen as an ecosystem containing various services provided for its inhabitants and visitors, while also facilitating the interaction with other service systems in its vicinity, eventually connected to the concept of smart city, smart energy grid, etc.

3 Design Approach

As CPS become larger and more complex, various authors have pointed out the need for proper design methodologies and support tools [10–13]. As such, various proposals have been put forward for classical CPS design, with origin in different areas, e.g. embedded systems, software engineering, control, and thus reflecting the concerns of those areas. Service-based design, platform-based design, contract-based approach, waterfall and spiral models, hardware-software co-design, V-model, model-based development, standards-based design, are just a few examples.

Of special relevance is the identification of design challenges [11], [10] namely coping with the following characteristics: hybridity, heterogeneity, interoperability, distributed nature, dynamicity, complexity, adaptiveness/evolving nature, human-centricity, etc. According to [11], a design methodology should support the cross-domain developments and be component-based, learning-based, time-aware, trust-aware and human-centric. As we move towards cognitive and collaborative CPS, other challenges become relevant e.g. (i) learning capability, already identified in [11], but becoming more important in the context of big data, and (ii) collaborative facets, for which it is relevant to seek contributions in the area of Collaborative Networks/Business Ecosystems [1].

On the other hand, cognitive and collaborative CPS represent a new "unexplored territory" and thus engineering such systems becomes more and more like an engineering research activity. As such, our proposed approach for Collaborative Cyber-Physical system design is based on the Design Science Research Methodology for Information Systems Projects [14]. The underlying conceptual framework considers three main cycles Relevance, Rigor and Design, and three pillars (Fig. 1):

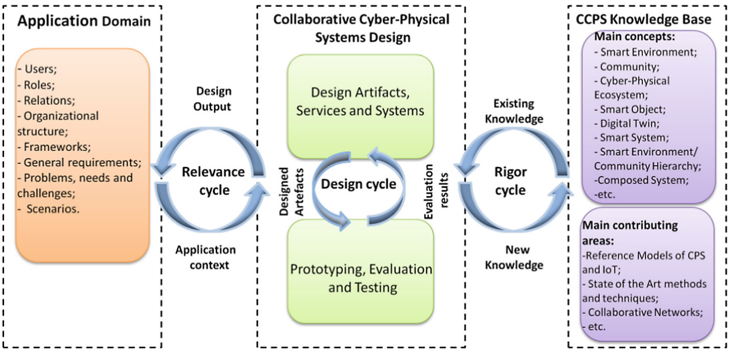


Fig. 1. CCPS design methodology inspired on design science

– Application Domain (e.g. Smart Manufacturing, Smart Home, Elderly Care, etc.), which contains needs, limitations and challenges that need to be considered throughout the design cycle. It also identifies the main actors and the scenario where these actors are living in and the way they are communicating and organised within the CCPS.
– Research pillar, which concentrates on the implementation of the given scenario by designing the cyber and physical artefacts with further evaluation and testing.
– CCPS Knowledge Base, which provides access to base concepts, theories, and best practices for artefacts building.

To interconnect the pillars of the CCPS Design Methodology, and following the design science approach three cycles are adopted:

– Relevance cycle, which is responsible for bridging the Application Domain and the CCPS Research in terms of providing the input for the design phase, but also in terms of identifying formal criteria for assessing the output of the design Cycle inside the application area.
– Design Cycle, during which the artefact's design undergoes constant checking for concordance with stated requirements and needs, in order to reach a satisfying solution to the stated problem.
– Rigor Cycle, that assures the necessary link between the research process and CCPS Knowledge Base containing both the application-based expertise as well as existing generic solutions or artefacts. Moreover, this cycle accumulates the knowledge generated during the artefacts' design phase to be considered in further iterations.

Of special relevance in this context is the identification of the key concepts for the CCPS knowledge base that should reflect the collaborative perspective, which will be addressed in the following sections.

4 Base Concepts

In this section some of the main concepts that will populate the CCPS Knowledge Base are introduced. To be able to design a complex CCPS there is a need to consider various abstractions levels [15]. Moreover some inspiration for describing and designing complex CPS can also be found in the Nature [16], which leads to adoption of the concepts of Ecosystem, Environment and Users Community.

Thus, a definition of a Smart Environment is proposed as follows:

> *"**Smart Environment** is a surrounding, with changing characteristics enriched with sensing, actuating and computing capabilities focusing on improving its inhabitants' experiences through increased contextual awareness and extended interaction capabilities."*

A Smart Environment comprises physical as well as cyber artefacts which are deployed in the environment and provide a variety of services. Physical artefacts are represented, in fact, by Smart Objects:

> *"**Smart object** is a physical entity with a cyber identity, some computing power, storage, and communication capabilities. It might be equipped with sensing and/or actuation capabilities, able to monitor the surrounding environment and/or change it."*

Cyber artefacts might be of a several types, as for instance, the applications which are discovering and enabling the capabilities of the physical artefacts and ensuring bidirectional communication. Another important cyber artefact is a digital twin:

> *"**Digital twin** is a virtual representation, in the cyberspace, of a physical asset, process or system, possessing real-time information about the represented entity and a model of its behavior.*
> *Each smart object can have a digital twin in the cyber part of a CCPS."*

As the CCPS often presume that the end-users of a provided service can be humans and not only other sub-systems, these systems might be assigned to a certain group or "Users Community":

> *"**Users Community** is a group of humans tied together by some social bond that creates in its members a "sense of belonging". The cohesion provided by this social bond usually leads to some sharing of intent or common interest, resources, preferences, risks, etc., materializing some notion of collective identity."*

Both "Smart Environment" and "Users Community" are the constituents of a Cyber-Physical Ecosystem:

> *"**Cyber-Physical Ecosystem** is a complex structure composed of a group of interconnected entities forming communities of users and sets of digital twins, the environment where they live in and the complex types of relationships between them considering mutual influence on each other both on cyber as well as physical layers."*

It also considers collaboration between humans and systems, including identification of users' needs and response with corresponding service, conflicts resolution, policy implementation, and adaptation to changing circumstances. Moreover the CPS Ecosystem as well as the Smart Environment can be considered under a hierarchical composition point of view:

> *"A **Hierarchy** assumes decomposition of some entity on smaller structural blocks/units which are characterised by some common aspect (e.g. location, devices types) or function/aim (e.g. energy management, air control). Those in turn might contain further structural units organised following the same principle while remaining affiliated to a parent entity and thus inheriting its attributes, rules and constraints."*

As the complexity and autonomy of CPS components increase, the CPS can be considered as System of Systems, consisting of a variety of building blocks or modules, which can be combined together in order to respond to users' needs bringing different functionalities in run time to create an "integrated service". This view is aligned with the notion of "*CPS components being freely composable*" as stated in [17]. Through combining resources and functionalities of smart, distributed entities we might achieve the concentration of necessary competencies to solve a certain problem which is not possible by a single component. This process needs to be dynamic, namely possessing the ability to quickly respond to the need with an appropriate "solution" and, if necessary, to add new components or replace those which do not deliver the required functionality. This idea is similar to the earlier notion of "coalition" of manufacturing components introduced in [18] and later revisited in other works such as [19]. While that initial idea faced several obstacles due to the lack of modularity of the physical manufacturing components, recent progress towards smart objects, represented by digital twins, makes it a promising approach. The notion of coalition or consortium gets its inspiration in the area of Collaborative Networks (CN) [20], namely from the concept of Dynamic Virtual Organisation (VO). Dynamic VOs are able to quickly respond to a changing market situation and are formed for a limited time. This idea presents a good potential in terms of application within CPS, as the components can be reused lately in order to respond to another need, thus this will allow to spare deployment of additional components with similar functionality. Thereby combining principles from CN, Coalition of Manufacturing Agents and the CPS background we define a notion of Coalition of Smart Components as:

> *"**Coalition of Smart Components** is a temporary association of a set of cyber entities – digital twins, representing physical components, which can be dynamically configured and adjusted to a changing surroundings/demand in order to provide an integrated solution."*

While in CNs partners for a VO are typically chosen from a pool of members of a Virtual Organizations Breeding Environment (VBE), here the components to create the coalition are selected from the pool of digital twins associated to a smart Environment.

Some of the definitions given above might be refined during the next stages of this ongoing research and some new might be added. However, these ones allow understanding the scope of the research and establishing the common basis. Three key concepts, Cyber-Physical Ecosystem, Smart Environment and Users Community, are intended to give the necessary abstraction level during the design and deployment phases. Cyber-Physical Ecosystems may also be small or large scale. An example of a small scale ecosystem can be a Smart Home, while a large scale example can be a Smart City. It is important to mention that not all smart objects necessarily require a digital twin, as simple sensors providing some measurements can be deployed as a part of more complex objects, while complex robots containing various actuators and sensors controlling and reacting on a variety of events might benefit from introducing the digital twin representing functionality on higher abstraction levels. Nevertheless, higher abstraction levels allow focusing on the collaborative and organizational aspects paying less attention to the low level technical details, which have been thoroughly addressed by other works.

5 High-Level Modeling of Basic Concepts

This section provides a model for some relevant concepts, such as Cyber-Physical Ecosystem, Users Community and Smart Environment, which were defined in the previous section. The UML diagram (Fig. 2) shows the main relations among these concepts. The Cyber Physical Ecosystem is a "bridging point" between the Smart Environment, responsible for managing the cyber and physical components, and the Users Community, focusing on users' organization. It is important to mention, that the Cyber-Physical Ecosystems, as well as the Smart Environments can consist of several sub-ecosystems or sub-environments building a hierarchy.

Each Users Community has a number of members; some of them, however, can be at the same time members of the other Users Communities. For instance, in a Smart House a person providing some repair services might be a member of the visitors' community and of the service providers' community. Being a member of a Users Community allows getting access to certain capabilities with privileges which depend on the assigned role. Thus, in fact, the role defines the access rights for each particular member. One open issue related to Users Community is the problem of "ownership", since although members may be owners of components, also a particular component might be the property of a Users Community and not of a specific member. This is not yet considered in the UML diagram.

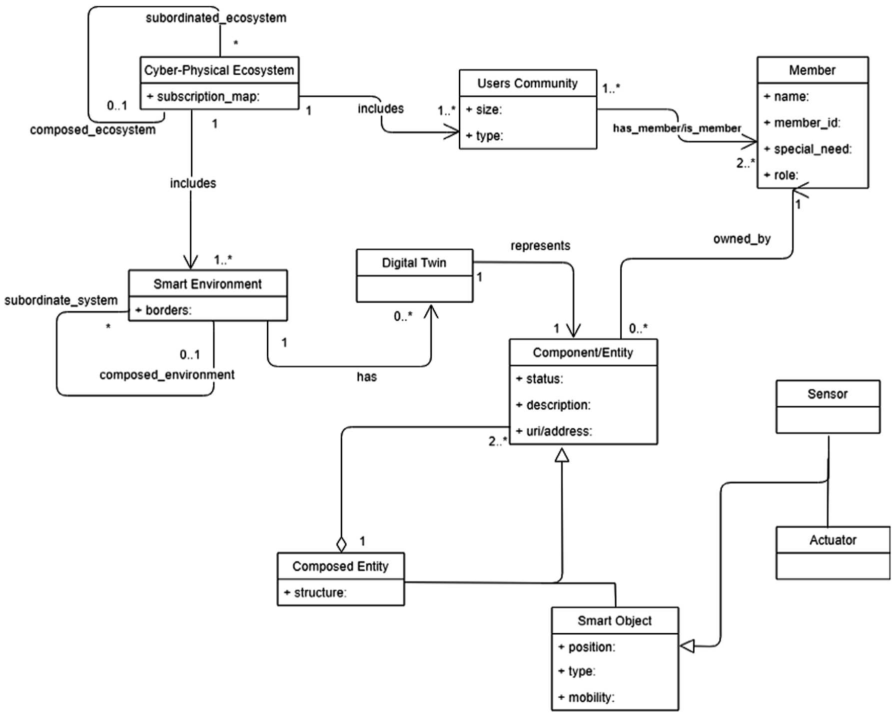


Fig. 2. A partial UML model of relevant concepts

The Smart Environment includes software entities being digital replicas of the physical components, the Digital Twins. Physical components are of two types, namely Composed entities and Smart Objects such as sensors and actuators. Furthermore, components can be of two types: mobile or static. Mobile components can move among different Smart Environments, as well as among Cyber-Physical Ecosystems.

The mobility of components sets another challenge regarding "borders", which means, when the mobile component leaves the host environment it does not possess a link to this environment anymore, but to the new one. One possible representation of this aspect is presented in Fig. 3:

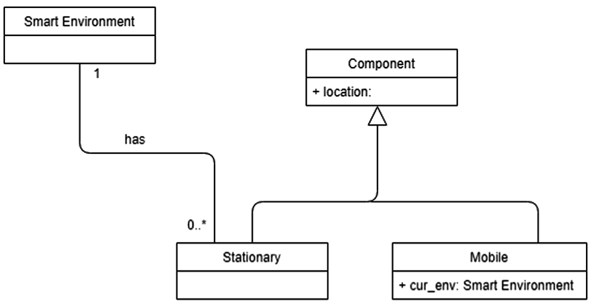


Fig. 3. Mobile and stationary components

Complementarily, and for further clarification of the stated concepts, the i* modeling language [21] is also used. It supports two modeling concepts, namely:

– Strategic Dependency Model, reflecting the agents or actors and the relationships among them;
– Strategic Rationale Model, which allows capturing the context behind the agents or actors through some basic elements, as tasks, goals, resources and soft goals.

In Fig. 3 the key concepts are presented in a Strategic Dependency model as agents showing the main relations among them. There are three main types of relations in this type of diagram: Goal, Resource and Soft Goal. This model allows giving an answer to some basic issues: “Which goals need to be achieved by an agent?”, “Which goals an agent asks to be achieved by the other agents?”, “Which resources are required?” (Fig. 4).

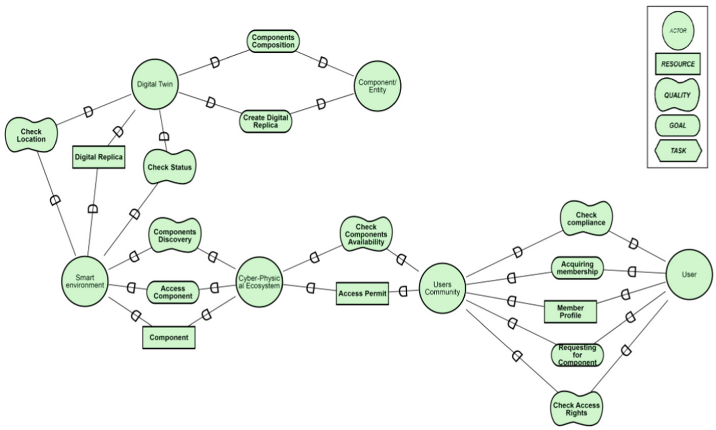


Fig. 4. Partial Strategic Dependency Model of key concepts

For illustration of the Strategic Rationale Model (SRM), one particular focusing point, namely the way of acquiring the membership to a Users Community, was chosen. SRM allows decomposition of tasks into sub-tasks. In Fig. 5 two actors are represented, one Member asking for a membership and the Users Community processing the request.

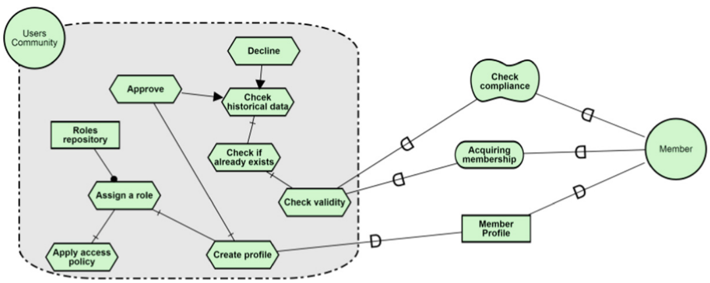


Fig. 5. SRM Model of membership acquiring

6 Collaborative Facets in CCPS Design

As mentioned above, the convergence of ideas developed within the area of Collaborative Networks and the foundations of the CPS is central point for this work. Namely the aspects related to VBEs and VOs are particularly relevant. Virtual Organisations Breeding Environment can be described as a pool of organisations possessing the "potential and will to cooperate with each other and that are prepared to rapidly join a VO in response to a business opportunity" [20]. Analogously, a Cyber-Physical Ecosystem contains a pool of ready to cooperate entities/components. The counterpart to a VO is a Coalition of Smart Components (CSC). In order to define the stages of the life cycle of the CSC, the corresponding model developed for VO creation can be adopted. In the simplest case, the VO creation process involves: establishing the VBE, partners search and suggestion, negotiation, launching and dissolution. Table 1 attempts to map the steps of VO creation process to the creation of CSC:

Table 1. Mapping the steps for VO Creation to CSC Creation

VO creation	CSC creation
VBE – is "a strategic alliance of organisations which adhere to long term cooperation and adopting common operating principles having the goal of increasing their chances towards collaboration in potential VO"	In the proposed approach, the Cyber Physical Ecosystem is analogous to a VBE representing a pool of Components (through their corresponding digital representations or Digital Twins), which provides a basis for building coalitions
Partners' search and suggestion – finding partners based on some criteria, such as fitness of competences, trust level, etc.	Formalizing the challenges and identifying necessary competencies to respond to the problem/collaborative opportunity. Build a taxonomy/structure of needed capabilities and how they can be combined. Discovery of components with defined capabilities (e.g. offered services) and selection based on some criteria, such as availability (free/occupied), location, (if in the required location or if it can move to desired location), etc.
Negotiation – an iterative process aiming at reaching agreements among partners and adjust needs with offers	Negotiations can be made within a Cyber-Physical Ecosystem considering the needs of components owners, scheduling plans and the desired output
VO launching involves ICT infrastructure configuration, orchestration of collaboration spaces, resources allocation, informing of involved members and manifestation of new VO in the VBE space	On this stage the responsibilities for each component are set, the structure of the Coalition and its representation in the ecosystems are properly defined
VO Dissolution – dissolution occurs when the VO has fulfilled its goal or the fulfilment seems not possible anymore due to various circumstances	When the task is finished for which the Coalition was established it is either dissolved or evolved to respond to another collaborative opportunity

As evaluation scenario, our ongoing work focuses on the area of home automation or smart home. Home automation includes a variety of possible applications. For instance, let us consider one of the core tasks of smart home, which is the provision of a comfortable living environment. For ensuring the comfortable living environment various components, with sensing and actuating capabilities, are involved and for this purpose a Coalition of Smart Components needs to be established. This CSC will include components such as temperature, humidity, air quality, and smoke sensors, and corresponding actuators as air conditioner, thermostat, lights, etc. Within the CSC these components are cooperating providing an integrated impact to the Users Community, i.e. jointly contributing to provide the service desired by the user. Moreover, as different users can have different preferences considering temperature, humidity, etc., other components possessing the ability to recognise the presence of a particular user may join the CSC.

For the stated scenario we have already started initial developments, for which Prolog language and a frame engine developed on top of it are used as the development environment.

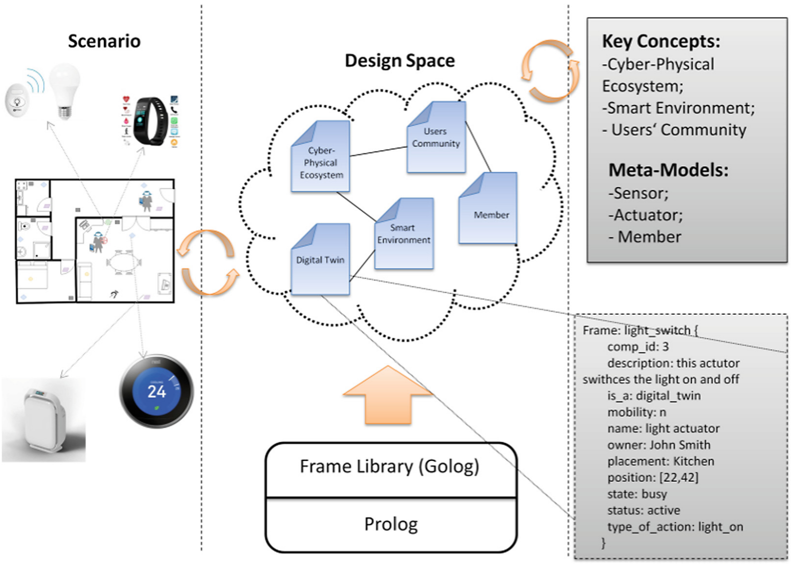


Fig. 6. General view for the planned developments

Following the previously introduced CCPS methodology, the approach for further developments is illustrated in Fig. 6. The framework considers 3 segments, the first one focuses on the target scenario (in this example a smart home), in the centre is the design space supported by a Frame Engine and Prolog which allows usage of the object-oriented and logic-based approach, and the left side represents the knowledge base, which provides the meta-model of relevant concepts. Frames are used as the base modelling formalism.

7 Conclusion and Further Work

The convergence of theories and ideas developed in the area of Collaborative Networks with the base CPS background led to emergence of the concept of Collaborative CPS which is intended to improve the potential of conventional CPS to adapt to changing users' needs and benefit from increasing intelligence level of CPS components. This collaborative perspective facilitates the development of Service Systems, as they require dynamic configuration of resources in order to produce an integrated impact. In this regard, one challenging aspect is to establish a CCPS design methodology for which an approach is highlighted in this work. Basic concepts and definitions for CCPS, combining ideas from CN and CPS, were also proposed in the scope of this work. Moreover, the idea of establishing Coalitions of Smart Components, based on background for VO creation, was proposed and a mapping of the VO creation to the CSC creation was established.

Some of open challenges are the aspects related to components mobility and corresponding issues of ownership/belonging and borders for the Cyber-Physical Ecosystem. Ongoing work focuses on the application of the proposed concepts and approach to a real use-case in the area of home automation.

Acknowledgments. This work was funded in part by the Center of Technology and Systems (CTS) and the Portuguese Foundation for Science and Technology (FCT) through the Strategic Program UID/EEA/00066/2013.

References

1. Nazarenko, A.A., Camarinha-Matos, L.M.: Towards collaborative cyber-physical systems. In: International Young Engineers Forum on Electrical and Computer Engineering (YEF-ECE), Costa da Caparica, Portugal, pp. 12–17. IEEE Xplore (2017). https://doi.org/10.1109/yef-ece.2017.7935633
2. Jezewski, J., Pawlak, A., Horoba, K., Wrobel, J., Czabanski, R., Jezewski, M.: Selected design issues of the medical cyber-physical system for telemonitoring pregnancy at home. Microprocess. Microsyst. Part A **46**, 35–43 (2016). https://doi.org/10.1016/j.micpro.2016.07.005
3. Munoz, D.-J., Montenegro, J.A., Pinto, M., Fuentes, L.: Energy-aware environments for the development of green applications for cyber-physical systems. Future Gener. Comput. Syst. **91**, 536–554 (2019). https://doi.org/10.1016/j.future.2018.09.006
4. Atos: Smart Factory. Connecting data, machines, people and processes – delivering the next generation of manufacturing, White Paper (2014)
5. Herterich, M.M., Uebernickel, F., Brenner, W.: The impact of cyber-physical systems on industrial services in manufacturing. Proc. CIRP **30**, 323–328 (2015). https://doi.org/10.1016/j.procir.2015.02.110
6. Chen, X., Wang, L., Wang, C., Jin, R.: Predictive offloading in mobile-fog-cloud enabled cyber-manufacturing systems. In: 2018 IEEE Industrial Cyber-Physical Systems (ICPS), Saint-Petersburg, Russia, pp. 167–172 (2018). https://doi.org/10.1109/ICPHYS.2018.8387654

7. Pacaux-Lemoine, M.-P., Berdal, Q., Enjalbert, S., Trentesaux, D.: Towards human-based industrial cyber-physical systems. In: 2018 IEEE Industrial Cyber-Physical Systems (ICPS), Saint-Petersburg, Russia, pp. 615–620 (2018). https://doi.org/10.1109/ICPHYS.2018.8390776
8. O'Donovan, P., Gallagher, C., Bruton, K., O'Sullivan, D.T.J.: A fog computing industrial cyber-physical system for embedded low-latency machine learning Industry 4.0 applications. Manuf. Lett. Part B **15**, 139–142 (2018). https://doi.org/10.1016/j.mfglet.2018.01.005
9. Camarinha-Matos, Luis M., Fornasiero, R., Afsarmanesh, H.: Collaborative networks as a core enabler of Industry 4.0. In: Camarinha-Matos, L.M., Afsarmanesh, H., Fornasiero, R. (eds.) PRO-VE 2017. IAICT, vol. 506, pp. 3–17. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-65151-4_1
10. Sangiovanni-Vincentelli, A., Damm, W., Passerone, R.: Taming Dr. Frankenstein: contract-based design for cyber-physical systems. Eur. J. Control **18**(3), 217–238 (2012). https://doi.org/10.3166/ejc.18
11. Seshia, S.A., Hu, S., Li, W., Zhu, Q.: Design automation of cyber-physical systems: challenges, advances, and opportunities. IEEE Trans. Comput. Aided Des. Integr. Circ. Syst. **36**(9), 1421–1434 (2017). [7778207]. https://doi.org/10.1109/TCAD.2016.2633961
12. Stokic, D., Scholze, S., Decker, C., Stöbener, K.: Engineering methods and tools for collaborative development of industrial cyber-physical based products and services. In: 2014 12th IEEE International Conference on Industrial Informatics (INDIN), pp. 594–599. IEEE Xplore. https://doi.org/10.1109/indin.2014.6945580
13. Jung La, H., Dong Kim, S.: A service-based approach to designing cyber physical systems. In: 9th IEEE/ACIS International Conference on Computer and Information Science, pp. 895–900. IEEE Xplore (2010). https://doi.org/10.1109/icis.2010.73
14. Hevner, A.R., March, S.T., Park, J., Ram, S.: Design science in information systems research. MIS Q. **28**(1), 75–105 (2004)
15. Gerostathopoulos, I., Skoda, D., Plasil, F., Bures, T., Knauss, A.: Tuning self-adaptation in cyber-physical systems through architectural homeostasis. J. Syst. Softw. **148**, 37–55 (2019). https://doi.org/10.1016/j.jss.2018.10.051
16. Camarinha-Matos, L.M., Afsarmanesh, H.: Roots of collaboration: nature-inspired solutions for collaborative networks. IEEE Access **6**(1), 30829–30843 (2018). https://doi.org/10.1109/ACCESS.2018.2845119
17. Cyber Physical Systems Public Working Group: Framework for Cyber-Physical Systems, Release 1.0 (2016). https://s3.amazonaws.com/nist-sgcps/cpspwg/files/pwgglobal/CPS_PWG_Framework_for_Cyber_Physical_Systems_Release_1_0Final.pdf
18. Barata, J., Camarinha-Matos, L.M.: Coalitions of manufacturing components for shop floor agility - The CoBaSA architecture. Int. J. Netw. Virtual Organ. **2**, 50–77 (2003). https://doi.org/10.1504/IJNVO.2003.003518
19. Wang, L., Törngren, M., Onori, M.: Current status and advancement of cyber-physical systems in manufacturing. J. Manuf. Syst. Part 2 **37**, 517–527 (2015). https://doi.org/10.1016/j.jmsy.2015.04.008
20. Camarinha-Matos, L.M., Afsarmanesh, H.: Collaborative networks: a new scientific discipline. J. Intell. Manuf. **16**, 439–452 (2005). https://doi.org/10.1007/s10845-005-1656-3
21. Yu, E.S.: Social modeling and *i**. In: Borgida, A.T., Chaudhri, V.K., Giorgini, P., Yu, E.S. (eds.) Conceptual Modeling: Foundations and Applications. LNCS, vol. 5600, pp. 99–121. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-02463-4_7

A Group Evacuation Method for Smart Buildings

Qasim Khalid[1(✉)], Alberto Fernández[1], Marin Lujak[2], and Arnaud Doniec[2]

[1] CETINIA, Universidad Rey Juan Carlos, Madrid, Spain
{qasim.khalid,alberto.fernandez}@urjc.es
[2] IMT Lille Douai, Univ. Lille Unité de Recherche Informatique Automatique, 59000 Lille, France
{marin.lujak,arnaud.doniec}@imt-lille-douai.fr

Abstract. Mass evacuation of people in buildings during emergencies is still a burning question regarding safety and reliability. Due to lack of information, crowd prefers to evacuate in the form of random groups. This random grouping increases the possibility of panic among evacuees due to human behavioral factors like herding and stampeding. As a result, congestion may occur resulting in unnecessary casualties. For this purpose, we propose a multi-agent evacuation architecture, that not only collects all the information from evacuees and events occurring in the building on a real-time basis but also provides the evacuation routes to evacuees. In this paper, we discuss a group formation module of our proposed architecture and use an example as a test case to check the functionality of our proposed group formation approach.

Keywords: Multi-agent system · Evacuation · Semantic technology · Situation Awareness · Smart buildings

1 Introduction

We would like to draw your attention to the fact that it is not possible to modify a paper in any way, once it has been published. This applies to both the printed book and the online version of the publication. Every detail, including the order of the names of the authors, should be checked before the paper is sent to the Volume Editors.

Emergency evacuation is performed in case of fire, terrorist threat, earthquake or any other hazardous situation in buildings. The ultimate goal of evacuation for all the occupants is to move out of a specific area or a building safely and as quick as possible. In general, a static route provided in an evacuation plan is followed, which is developed during the erection stage of a building. However, static evacuation plans lack coordination between evacuees and administration as it does not consider contingencies that may occur during evacuation. Due to the absence of any coordination during the whole evacuation process, several problems may occur, e.g., congestion along evacuation routes or at exit points or locations that are frequently used, and as a result, casualties may occur. Among important factors associated with an emergency evacuation are evacuees' behavior, their mobility characteristics, and their coordination in following

Published by Springer Nature Switzerland AG 2019
L. M. Camarinha-Matos et al. (Eds.): DoCEIS 2019, IFIP AICT 553, pp. 206–213, 2019.
https://doi.org/10.1007/978-3-030-17771-3_17

the evacuation route related to a type of emergency and relevant evacuation strategy. Vermuyten et al. have conducted a literature survey in which they discussed various evacuation strategies and crowd behavior, crowd flow, optimized paths, bottlenecks, etc. (see, e.g., [1]). These strategies have somehow improved the evacuation process, but most of the studies are static. Due to this reason, real-time status of buildings and evacuees is not known in case of an emergency and during the evacuation process. Therefore, there is still a need for an intelligent system that could detect hazardous situations and handle evacuation problem on a real-time basis.

To this end, we propose a real-time multi-agent architecture for mass evacuation in smart buildings. The novel concept for this study is that the information of events and evacuees is processed on a microscopic level and the evacuation is done in the form of groups on a macroscopic level. These groups are created on the basis of momentary location, possible family ties with other evacuees, and mobility characteristics of each evacuee. Later, groups evacuate the building under the supervision of their assigned group leaders. These group leaders use hand-held devices that provide all the information needed for the evacuation process. At this point, our system is designed for the office spaces and the places where people are slightly aware of topology of the buildings. Due to the scope of this paper, we briefly discuss our proposed architecture and its functioning. Also, we specifically focus on the group formation module of the proposed architecture.

The rest of the paper is organized as follows. Section 2 states the contribution of this research to service systems. Related works in the area of crowd evacuation and smart buildings with their limitations are discussed in Sect. 3. The structure of the proposed architecture is explained in Sect. 4. The group formation methodology is discussed in Sect. 5 and we give a functioning example in Sect. 6. In the end, we conclude the paper with further research directions in Sect. 7.

2 Relationship to Industrial and Service Systems

In a service system, both technology and manpower work together to fulfill the requirements of its user(s). As service systems maximize the output as well as ensure the safety and reliability of the system, therefore its applications are found everywhere such as in the sectors of engineering, banking, supermarkets, and even on internet.

The main objective of service systems is to serve its customers and provide them what they want in minimum time. Customer satisfaction is also an important factor in this type of systems. A service system must be safe and reliable enough so that its users can use it without any risk or danger. When it comes to the cyber-physical world where machines perform most of the work, while humans are involved mostly as users, then the question of user satisfaction and safety is of special importance.

In this study, we propose an intelligent architecture for evacuation of buildings. The purpose of this system is to minimize the involvement of a specialized personnel for evacuation management during any emergency situation, while providing decision support for evacuation of a multitude of people according to their evacuation requirements and preferences, such as mobility characteristics and mutual family ties. Since the lives of people are dependent on it, the reliability of such a system is of utmost priority. In this respect, the proposed decision support architecture for emergency evacuation may be seen as a service system whose aim is to minimize the evacuation related risks considering the mobility requirements of its users.

3 Related Work

Multiple approaches have been proposed for the emergency evacuation of buildings (see, e.g., [2–5]). Vermuyten et al. performed a literature survey in which they have classified the research topics according to the phases and requirements of the whole evacuation system, e.g., pedestrian dynamics, evacuation design, optimization of evacuation routes, crowd management, model movement, and crowd flow (see, e.g., [1]).

There are generally two kinds of crowd evacuation models: macroscopic and microscopic one. Huges explained the significance and the applicability of both approaches for handling the crowd in [6]. The macroscopic model is used at a crowd level to implement the rules over a whole crowd whereas the microscopic model is used to model every individual participating in the evacuation process. Here, some authors have proposed a hybrid approach to use both strategies as they complement each other (see, e.g., [7]).

On the other hand, Liu et al. explained that the evacuation process is not only individual action, but it is a group activity where the efficiency of the whole process is dependent on everyone involved. Hence, coordination and cooperation of evacuees is needed to make that evacuation process efficient (see, e.g., [8, 9]). In this manner, researchers have proposed some interesting strategies to observe the behavior of groups during an evacuation. For example, [10] developed a genetic algorithm with a combination of a cellular automata model to evacuate groups of people. In this model, each group is assigned to a specific exit which evacuates within a specific time because the next group has to evacuate afterwards. The best evacuation plan is found after finding the route having minimum evacuation time.

On the other hand, when the crowd evacuates in the form of groups, synchronization and coordination issues arise. In case of any wrong decision by the group leader, even worse disaster could happen as the time is a most critical thing during that process (see, e.g., [11]). However, it is also found that the evacuation time increases because of the presence of groups in the crowd (see, e.g., [12]).

A hybrid evacuation strategy that could assist the evacuees in the formation of groups as well as in choosing an optimal route for them is still an open challenge that we treat in this paper.

4 Evacuation Architecture

In this section, we explain the architecture of our proposed system for evacuation. The system consists of three agents namely, Situation Awareness Agent, Route Optimization Agent, and Group Management Agent. *Situation Awareness Agent* handles and processes the information of events coming from sensors on a real-time basis.

Various sensors are installed that detect fire, smoke, terrorist attack or threat, quarrel or any other risk that could lead to an emergency evacuation (see, e.g., [13–15]). This agent also communicates with other agents to assign an alternate safe route to evacuees in case of any blockage or congestion in the existing provided path. *Route Optimization Agent* chooses the optimal exit path for each group. For this purpose, the building is represented as a graph and divided into different sections called nodes, and the paths that connect these sections called edges or links. Later, the safest and fastest route is calculated with the help of the live information received from the sensors. *Group Management Agent* creates groups which are supposed to evacuate together under the supervision of (human) group leaders. Groups are created using evacuees' live locations, family ties and physical characteristics whereas group leaders are trained occupants of a building chosen beforehand on a voluntary basis. This agent provides the formation of groups with their relevant collection points and the allocation of groups' leaders. The block diagram of the architecture is shown in Fig. 1.

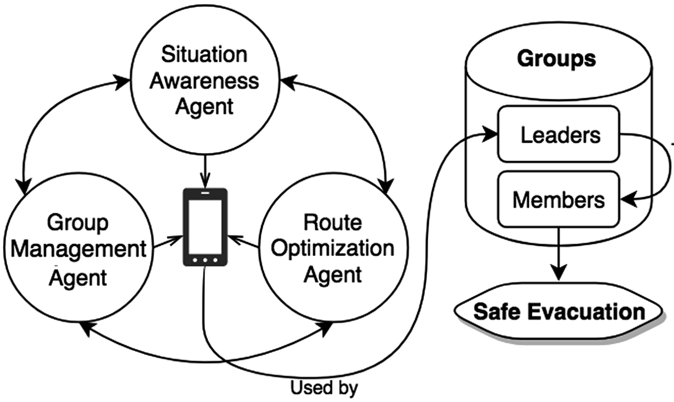


Fig. 1. Evacuation architecture.

Personal information of evacuees related to evacuation is crucial for well-functioning of the proposed system and should be obtained beforehand. As this information is static, it is stored in the database with the help of an ontology. Upon entering the details into the system, a special RFID or a class of Bluetooth low energy (BLE) device (e.g., iBeacon) tag is provided to each person which they have to wear all the time during the evacuation so that all the movements and locations of people are monitored.

As it is discussed in [16], the evacuation process is a multi-phase procedure. Therefore, we divide the whole evacuation process into two phases: notifying phase and evacuation phase. Firstly, during the notification phase, we use an audio or a text message or an email to inform every evacuee about his/her assigned group collection point. Also, this information (based on the momentary location) is shown on public screen displays. During the evacuating phase, real-time location of every occupant is processed individually, and groups are updated on the basis of this information. Later, the groups gather at their relevant collection points, and then group leaders start the evacuation process under their supervision. Note that the group leaders use smart devices with an app where all information of group members are shown along with their calculated evacuation route plans.

5 Group Formation

In this section, we focus on the group formation module of group management agent of the proposed architecture.

In general, groups are formed by combining the static and dynamic information of evacuees. Static information includes information of mobility characteristics and family ties with other evacuees. Dynamic information consists of live locations of the people which are continuously detected by sensors. Every person in the building has a RFID or iBeacon tag which is provided before entering into the building or during the registration process. These tags have a unique ID by which the data of every person can be accessed.

If any hazard or risk is detected and evacuation is necessary, evacuees are provided by a location (known as collection points) where they have to gather and wait for next instructions from their group leaders. These collection points are assigned individually in line with the mobility characteristics of each evacuee. In this regard, a procedure for group creation is given in the following. This procedure also provides collection points to evacuees individually.

The procedure requires the following information: (i) an updated momentary location and mobility characteristics and family ties for each evacuee, (ii) location and capacity of each collection point, (iii) available evacuation routes with their free-flow capacities from each one of the collection points, (iv) maximum evacuation time given by the evacuation authority.

The procedure runs in the following steps that assign each evacuee to its collection point considering the capacity of the collection points and their evacuation routes.

Step 1: Calculate the shortest evacuation time of each person from his/her momentary location to each collection point;

Step 2: Assign each evacuee whose evacuation time is larger than the maximum evacuation time to the closest collection point with the shortest evacuation time;

Step 3: Assign each mobility impaired person to his/her evacuation leader and a collection point with sufficient capacity and the shortest evacuation time;

Step 4: Assign evacuees with family ties (members of the same family) to the same collection point. In case they are not located in the same area, choose a collection point that is the closest to the same distance from all of them;

Step 5: Assign the rest of the evacuees to their nearest collection points based on the available capacity.

6 Case Study

In this section, we use an example as a test case to check the functionality of our proposed algorithm for group formation of people and their allocation to their feasible collection points. For this purpose, the output of the algorithm is shown in Fig. 2.

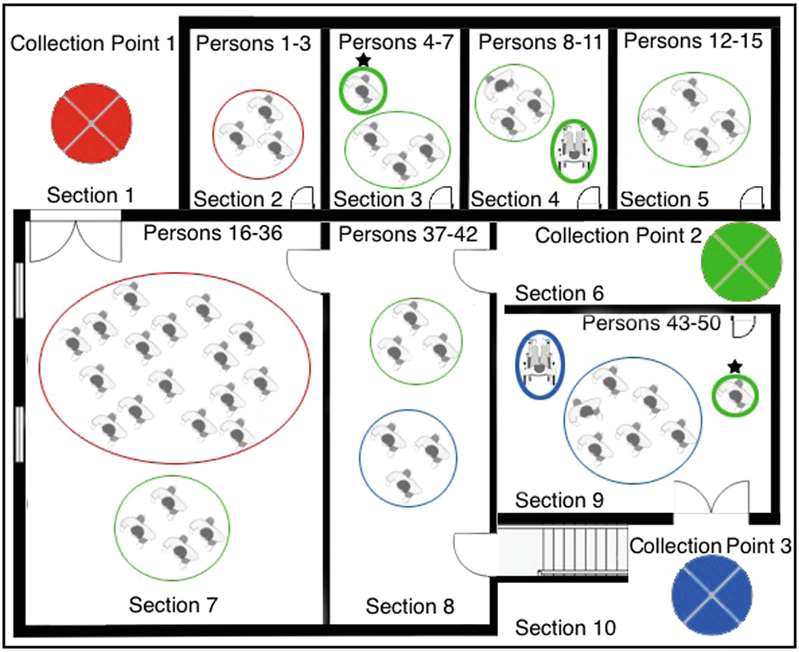


Fig. 2. Final grouping with respect to their feasible collection points. (Color figure online)

Given is a layout of a floor in which people are located in possibly different sections. These sections can be rooms and/or corridors, and all sections are connected to every collection point. For simplicity, we assume that the geo-positioning of evacuees is flawless.

In our example in Fig. 2, there are three collection points and 50 persons of which two are mobility impaired (P8, P43) and other two have family ties (P5, P45). We assume that all people are located in the areas that are within the maximum evacuation time from safe exits. Therefore, the priority will be given to mobility impaired people and people with family ties. As a result, P8 and P43, who are mobility impaired are assigned to their nearest collection points, CP2 and CP3 respectively. Also, persons with family ties, P5 and P45, are assigned to the same group/collection point, CP2, because collection point CP2 is the nearest one to both of them. The rest of the persons are allocated to their nearest collection point based on the available capacity. Once the capacity of a collection point is reached, remaining persons are allocated to the nearest collection point with available capacity.

In Fig. 2, collection points are colored in green, red, and blue. People are grouped in groups based on their assigned collection points (red, green, and blue) so that they have to gather at their allocated collection point. People who are mobility impaired have a bigger ring around them to represent their priority. Similarly, people who have family ties also have a bigger ring around them as well as a star over them to distinguish them from able-bodied and mobility impaired people.

The output of the proposed procedure will be in the form of a couple person's ID, collection point of that person (assigned group).

Following are the sets of persons who are allocated to a specific collection point on the basis of knowledge acquired by them. Red, green, and blue sets are formed according to the capacity of collection point 1, collection point 2 and collection point 3 respectively.

Red set = *{P1, P2, P3, P16, P17, P18, P19, P20, P21, P22, P23, P24, P25, P26, P27, P28, P29, P30, P31, P32}*

Green set = *{P5, P45, P8, P4, P6, P7, P9, P10, P11, P12, P13, P14, P15, P33, P34, P35, P36, P40, P41, P42}*

Blue set = *{P45, P37, P38, P39, P44, P46, P47, P48, P49, P50}*

7 Conclusion

In this study, we proposed an evacuation architecture that coordinates the evacuation process from the detection of the hazard or emergency to the evacuation of people in buildings. Our system is based on three different agents that perform their evacuation-related tasks. In our proposed strategy, evacuation is done in the form of small groups, each one led by a group leader. Each group leader gets live information from an app installed on a smart device specifically designed as an interface to our proposed architecture. In this paper, we discussed the module of group management agent, a module of our proposed architecture together with group formation and allocation of collection points. Also, due to the hierarchy of this solution proposed for buildings, people may have some training for specific scenarios.

As a future work, we plan to develop other agents of our proposed architecture. Also, we will integrate the group formation module, developed in this study, with other parts of the architecture. We also plan to compare the performance of the proposed system with other state-of-the-art evacuation coordination systems.

Acknowledgments. Work partially supported by the Autonomous Region of Madrid (grant "MOSI-AGIL-CM" (S2013/ICE-3019) co-funded by EU Structural Funds FSE and FEDER), project "SURF" (TIN2015-65515-C4-4-R (MINECO/FEDER)) funded by the Spanish Ministry of Economy and Competitiveness, and through the Excellence Research Group GES2ME (Ref. 30VCPIGI05) co-funded by URJC-Santander Bank.

References

1. Vermuyten, H., Beliën, J., De Boeck, L., Reniers, G., Wauters, T.: A review of optimisation models for pedestrian evacuation and design problems. Saf. Sci. **87**, 167–178 (2016)
2. Lujak, M., Billhardt, H., Dunkel, J., Fernández, A., Hermoso, R., Ossowski, S.: A distributed architecture for real-time evacuation guidance in large smart buildings. Comput. Sci. Inf. Syst. **14**(1), 257–282 (2017)
3. Lujak, M., Ossowski, S.: Intelligent people flow coordination in smart spaces. In: Rovatsos, M., Vouros, G., Julian, V. (eds.) EUMAS/AT - 2015. LNCS (LNAI), vol. 9571, pp. 34–49. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-33509-4_3
4. Wagner, N., Agrawal, V.: An agent-based simulation system for concert venue crowd evacuation modeling in the presence of a fire disaster. Expert Syst. Appl. **41**(6), 2807–2815 (2014)
5. Zheng, X., Zhong, T., Liu, M.: Modeling crowd evacuation of a building based on seven methodological approaches. Build. Environ. **44**(3), 437–445 (2009)
6. Hughes, R.L.: A continuum theory for the flow of pedestrians. Transp. Res. Part B Methodol. **36**(6), 507–535 (2002)
7. Xiong, M., Tang, S., Zhao, D.: A hybrid model for simulating crowd evacuation. New Gener. Comput. **31**(3), 211–235 (2013)
8. Liu, H.: Context-aware agents in cooperative design environment. Int. J. Comput. Appl. Technol. **39**(4), 187–198 (2010)
9. Liu, H., Sun, Y., Li, Y.: Modeling and path generation approaches for crowd simulation based on computational intelligence. Chin. J. Electron. **21**(4), 636–641 (2012)
10. Abdelghany, A., Abdelghany, K., Mahmassani, H., Alhalabi, W.: Modeling framework for optimal evacuation of large-scale crowded pedestrian facilities. Eur. J. Oper. Res. **237**(3), 1105–1118 (2014)
11. Oxendine, C., Sonwalkar, M., Waters, N.: A multi-objective, multi-criteria approach to improve situational awareness in emergency evacuation routing using mobile phone data. Trans. GIS **16**(3), 375–396 (2012)
12. Li, Y., Liu, H., Liu, G., Li, L., Moore, P., Hu, B.: A grouping method based on grid density and relationship for crowd evacuation simulation. Phys. A **473**, 319–336 (2017)
13. Sentinel, Accuware. https://www.sentinelcv.com/. Accessed 15 Feb 2019
14. Temperature sensors, Texas Instruments incorporated. http://www.ti.com/sensors/temperature-sensors/overview.html. Accessed 15 Feb 2019
15. Intelligent smoke detectors, Kidde-Fenwal Systems. https://kidde-fenwal.com/Public/SystemDetails/Kidde-Fire-Systems/SmartOne-Intelligent-Smoke-Detectors. Accessed 15 Feb 2019
16. Talebi, K., Smith, J.M.: Stochastic network evacuation models. Comput. Oper. Res. **12**(6), 559–577 (1985)

Optimized Electrification of Subsea Oil & Gas Infrastructures Based in Genetic Algorithm

Tiago A. Antunes[1(✉)], Rui Castro[2(✉)], P. J. Santos[3,4(✉)], A. J. Pires[5,6(✉)], and Matthias Foehr[7(✉)]

[1] Siemens SA, Energy Management Transmission Solutions, Lisbon, Portugal
tiago.antunes@siemens.com
[2] INESC-ID/IST, University of Lisbon, Lisbon, Portugal
rcastro@tecnico.ulisboa.pt
[3] INESC Coimbra, Coimbra, Portugal
[4] Department of Electrical Engineering, EST Setúbal, Setúbal, Portugal
paulo.santos@estsetubal.ips.pt
[5] CTS/UNINOVA, Caparica, Portugal
[6] Department of Electrical Engineering, EST Setúbal/IPS, Setúbal, Portugal
armando.pires@estsetubal.ips.pt
[7] Siemens AG, Energy Management Transmission Solutions, Erlangen, Germany
matthias.foehr@siemens.com

Abstract. Offshore field development relies on multiple optimization techniques targeting a feasible and cost-effective production solution yet are focused on the field itself. While so, advancements in offshore engineering bring increasingly complex subsea infrastructures to depths in the excess of 3,500 m. Many offshore production topsides which currently rely on costly and harmful onboard thermal-based power generation are turning to high voltage power-from-shore electrification solutions to cope with the challenges being brought by subsea infrastructures. An optimal electrification of these subsea templates is a challenge on its own as the seafloor morphology and well distribution is far from consistent. This paper presents a combined k-means and genetic-algorithm optimization to assess how the combined deployment of high voltage umbilical, wellheads and subsea substations can be optimized for the lowest cost possible. Results show a significant improvement in optimization of the total umbilical length as well as the substation positioning on the seabed.

Keywords: Subsea · Substations · Oil & Gas · Optimization · Genetic algorithm

1 Introduction

The International Energy Agency (IEA) stresses that projects in the main areas such as the Gulf of Mexico and the North Sea, were once characterized by a break-even ratio of 60–80 USD/barrel. Equivalent projects are made viable and as robust as break-even oil prices of 25–40 USD/barrel [1]. This can only be achieved via proper simplification, modularization and downsizing; Electrification is at the core [2, 3].

Published by Springer Nature Switzerland AG 2019
L. M. Camarinha-Matos et al. (Eds.): DoCEIS 2019, IFIP AICT 553, pp. 214–223, 2019.
https://doi.org/10.1007/978-3-030-17771-3_18

Offshore power production is based in thermal combustion systems whether this is accomplished via diesel or gas and that can reach several dozen megawatts [4, 5]. Modularization and optimization mean that owners can increase the complexity of the templates, reduce the size of the topside and increase the overall efficiency and resilience. Furthermore, unmanned operations, expected to dramatically reduced Operational Expenditure (OpEx), will require a reliable power supply with a large maintenance window, possible by means of power-from-shore (PFS) solutions [6].

In subsea facilities commonly ranging up to 100 MW [7] the matter is addressed on a point-to-point or direct in-place analysis of the most suitable electrical infrastructure and not on the optimization of the actual electrical grid on the seafloor. The authors have reviewed field development cases [8, 9], yet those references are focused on the subsea production setup and not on the actual electrification.

The authors were able to review several case studies where satisfactory results were obtained by means of GA [10–12] – which sustains the suitability of the approach. The same algorithm was also used in other examples, such as in the design of S/S's earthing grids [13], switchgear internal design [14] or in life-cycle assessment of electrical S/S [15]. References show that simulated annealing can also be applied for this sort of optimization, often times combined with GA [16, 17].

The rest of the paper is organized as follows. On the second chapter, a connection is established between this research and the resilient systems topic. The third chapter provides an overview of the main industry challenges, which shall serve, on the forth chapter, to support the model design. The results and conclusions are presented on the fifth and sixth chapter, respectively.

2 Relationship to Industrial and Service Systems

Although PFS on itself is not an innovation, and field development is a consolidated environment, no references to support the use of Genetic Algorithms (GA) or others for optimized subsea electrification were found. However, the application of GA on the optimization of power systems, more specifically, in the substation (S/S) expansion planning (SEP) is widespread and quite common. Companies like SIEMENS want to drive large-scale deployment of PFS solutions as an alternative to onboard thermal allowing utilities to take full advantage of the massive renewable capacity available.

Having so, and whilst the authors assess new ways to power Oil & Gas (O&G) infrastructures from shore, this paper presents their novelty approach towards bringing a much-needed solution for the cost optimization of the subsea electrical setup. It is designed by means of a Python-based mixed Kernel-Means (KM) and GA, considering simultaneously the layout of the field and the electrification concept to get the most cost-effective solution.

3 Industry Challenges

- An analysis of over one hundred offshore sites showed electrification costs ranging between 0.3–0.5 MUSD/kboed[1] despite the variations on the resource or water depth. In the case of subsea infrastructures, the ratio was 0.5;
- Subsea production facilities are characterized by largely dispersed and small-sized equipment. An analysis of all the possibilities can become time-consuming and the GA does not require the evaluation of all possibilities [18];
- Industry leaders are working towards subsea electrical power standardization, which supports the claim that robust subsea electrical systems are both new and indeed needed at a fast pace [19];

On one hand, unsupervised methods such as KM clustering bring a straightforward method to analyze the positioning of the wells without initial candidate locations [17]. As it does not require initial mapping it serves as a solid method to suggest individual candidates on the GA application [11]. On the other hand, GA are known for providing high quality and performance results, especially for a high number of possibilities [20]. In addition, they show enough flexibility to accommodate several restrictive measures – such as the number of power outlets allowed on a S/S and the maximum cable distance.

4 Proposed Method

4.1 Initial Assumptions

An extensive analysis of the offshore portfolio was made. Wellheads are distributed along a squared area, representing field section. The number of wells is based on current projects. Focus was given to the evaluation of the best possible electrification for any given field development considering the following portfolio restrictions:

- Electrical distribution is made typically using voltages ranging between 3.3 and 36 kV [7, 19] and oil-filled, pressure-balanced and coated electrical distribution units (EDU or SDU) are used to power the subsea systems. These units rely on templates pre-assembled on surface [21];
- Electrical distribution manifolds (EDM) - a subsea substation, can also be used to provide power to multiple locations at the same field whilst the number of connectors is kept to a minimum to ensure overall system resilience [21];
- One power transformer is planned on each S/S and the technology used, high voltage alternated current (HVAC) or direct current (HVDC), is not compared as it depends on the distance to shore; And this optimization serves for both;
- The number of power distribution outlets of the EDU is estimated between 3 and 6 at a given voltage level whereas subsea wellheads are assumed to be rated the same [3];
- The number of S/S is estimated based on the quantity of wellheads, as the cost of the PDU is largely higher than the HV umbilical an estimation based solely in the number of connections to do was used [21];

[1] One thousand barrels of oil equivalent per day.

- The initial positioning of the S/S is done via the K-means clustering method whereas the GA populates multiple candidate substation positioning. The size of the solution array is constant which means that all subsea templates are powered by one and only one of the subsea S/S;
- The solutions (individuals) share traits (S/S connections) between each other and the GA application is able, not only to foster the better combination of those traits, but also, to short-list the better candidate solutions [22].

A joint approach on both hydraulic and electrical power subsea distribution was not made although it may lead to further improvements. The geography of the seabed and fields are outside the focus of this paper. The objective is an universal canvas.

4.2 Process Flow

As said, KM is used as an unsupervised learning method for clustering the field data. This algorithm provides each individual with a candidate position and the distance to each wellhead to which respects to. Please refer to Fig. 1. for a process flow example. Subsea wells are identified in blue whereas the S/S are identified in green. The links are mapped accordingly. This is the baseline for the GA, representing one individual from the sense of the S/S location.

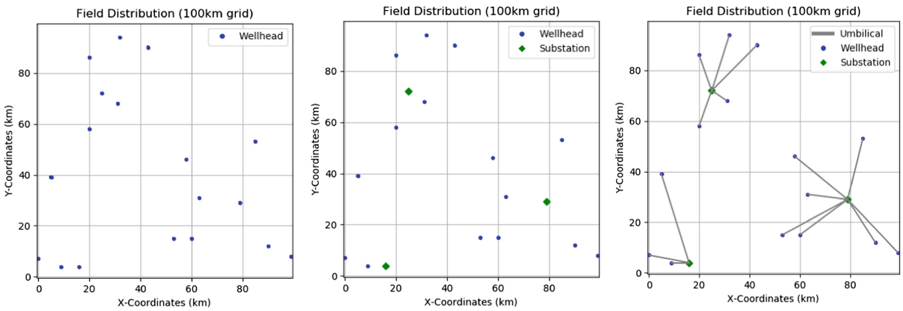


Fig. 1. Deployment of wellheads, substations and umbilical (algorithm process)

For any well and subsea S/S set, the objective is to minimize the total cost of the electrification solution (cumulative cost of the subsea cabling for each well-substation link). The cost objective function (1) to be minimized is presented next, where L represents each cable distance, C and C_{KM} the total and unitary cost, respectively:

$$C = \Sigma L_{CABLE-N} * C_{KM} \tag{1}$$

The connections are represented via a single-digit code which states the reference of the respective S/S. That allows for a direct and natural encoding. The same node (position) cannot be completed by neither none nor more than one S/S - an example is

1	2	2	2	3	3	1	2	1	3	3	2	2	1	3

Fig. 2. Example of an individual as used on the GA

shown on Fig. 2. Script-wise, each individual is made of an integer array with a length equal to the number of load centres (wells).

The distance is calculated as the shortest straight line between the two-points (S/S and well). Such calculation outputs a similarly-structured array. The decision variable is the total cost for an individual (composed of the cumulative cost of each wellhead/substation tuple). Typically, for this analysis, the restrictions are the spatial representation of the wells, S/S and connection links. All individuals are deemed acceptable and the tuples are selected within the number of wells and S/S proposed.

4.3 Characterization of the GA Population

The population size was set so that a balance between both the number of S/S and templates to power is reached. Sensibility tests were made to assess the appropriate setup. The best compromise was found sizing a population equal to the number of wellheads (links) with a medium crossover rate and a very low mutation rate. The iteration rate was kept high during most cases, although fitness is achieved quite early. The population size is the most critical variable although the mutation rate must be kept to a minimum (up to 10%) given the large population size [7] (Table 1).

Table 1. Constrains of the genetic algorithm

Rate designation	Variable	Value	Units
Population (individuals)	*ind_rate*	1	[-]
Iterations of the GA	*ite_rate*	200	[-]
Top fitness rate	-	1	[-]
Crossover rate	*cross-rate*	40	%
Mutation rate	*mut_rate*	10	%

4.4 Ranking, Crossover and Mutation

The total cost per each individual is calculated by means of the sum of the length of all links combined and the unitary cost. Figure 3 presents an example of the elite selection by addressing the total cost (per individual) array of the population.

10	22	35	12	14	18	19	20	12	23	25	17	15	20	22

Fig. 3. Example of a cost array as used on the GA

A portion of the remaining population is selected for cross-over. An even number of individuals are selected, and a proper parent-children pair can be selected. The cross-over is done using the sexual reproduction principal with two individuals, done via cross-over pointing, where the two new sub-set strings are swapped and will form the children individuals. An example is presented in Fig. 4.

Parent A	1	2	2	2	3	3	1	2	1	3	3	2	2	1	3
Parent B	1	3	1	1	2	2	3	3	2	2	2	1	3	3	2

Child A	1	2	2	2	3	3	3	3	1	3	3	2	2	1	3
Child B	1	3	1	1	2	2	1	2	2	2	2	1	3	3	2

Fig. 4. Example of a cross-over operation as used on the GA

The parents are randomly selected, and each tuple provides one children. For that reason, the number of cross-over iterations matches the number of parents (population size remains unchanged). There is no preference on the parents and the two new individuals from each round proceed for the mutation pool.

The mutation is done to the pool of all individuals with an exception for the top fitness one, thus including the untouched individuals, as well as, the cross-over children. The low rate mutation is applied on a single random point of an individual. One link (S/S reference) is changed in that percentage of individuals. In order not only to keep accurate results and avoid a recalculation of the costs, but more importantly, to reanimate possible important solutions, the "mutated" individual receives a gene (link and distance tuple) from another individual of the pool, also randomly selected. The process is shown in Fig. 5.

Original	1	2	2	2	3	3	1	2	1	3	3	2	2	1	3
Mutated	1	2	2	2	3	3	1	3	1	3	3	2	2	1	3

Donor	1	3	1	1	2	2	3	3	2	2	2	1	3	3	2

Fig. 5. Example of a mutation operation as used on the GA

The distance is already available, taken also from the equivalent position on the distances' array. This avoids the need to recalculate all distances of the current population and optimizes computation time. The standard deviation is used to assess the performance of the GA. This process is iterative and is repeated for a given number of times.

5 Results

5.1 Initial Conditions

The test conditions are presented in Table 2.

The mapping is reviewed against the subsea grids and O&G fields known. The cost for the umbilical is indicative and the number of S/S is selected based on maximum

Table 2. Key data of the case study

Constrain description	Value	Units
Map resolution	100	km
Number of substations	3	[-]
Number of wellheads	15	[-]
Number of iterations	1,000	[-]
Unitary umbilical cost	5,000	USD/km

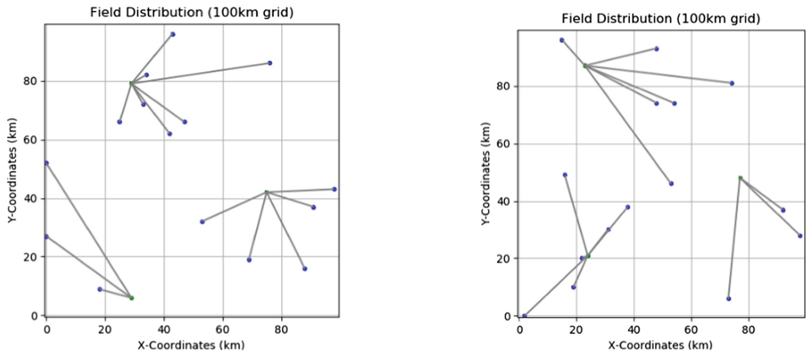


Fig. 6. Field distribution with substations deployed (Case studies)

number of connections of each unit (PDU), in this case 5 wellheads each. The optimized deployment is shown on Fig. 6. The grey line represents the HV umbilical laid between the wellheads and the respective S/S. The authors acknowledge that the solutions proposed are not perfect and additional measures should be taken to avoid dispersion of the S/S along the field during GA execution.

5.2 Cost Minimization

The results of the total distance and cost for each iteration are presented in Fig. 7.

In the same manner, the standard deviation is presented in Fig. 8.

There is a quick convergence towards the best cost position. However, as already commented, the higher rate of iterations was defined so that a sustained convergence of the overall population was possible. Results are summarized on Table 3.

In some scenarios, the same cost was found for two simultaneous positions. Yet, these are a minor occurrence on the overall spectrum of scenarios addressed. The routine is, however, able to identify the individuals which match this best fitness performance.

In fact, using a KM before the GA implementation dramatically reduces the computation time for the optimization as the centroids are calculated (via the Euclidean distance) close to their final positions in the majority of cases. Increasing the threshold was used to accommodate the lack of convergence, in multiple cases, of the standard deviation. Efforts were made to adjust the cross-over and mutation ratios to improve such convergence.

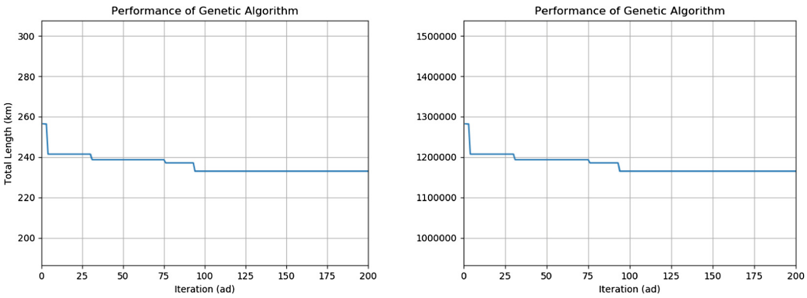


Fig. 7. {Cost, distance} per iteration

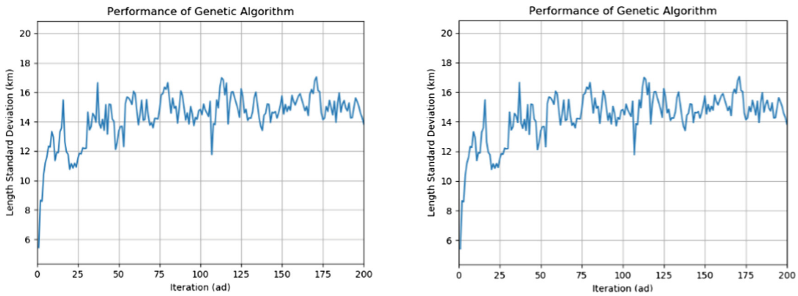


Fig. 8. {Standard distance deviation, cost deviation} per iteration

Table 3. Final results

Variable identification	Value	Units
Total of iterations	1,000	-
Initial length	258	km
Final length	235	km
Initial cost	1,280	MUSD
Final cost	1,170	MUSD

An increase on the mutation rate is known to allow for an optimization of the computation of the solution which supports the claim that associative optimization problems can be served with better results by using mutation rates higher than cross-over rates. The increased cross-over ration introduces a very high disturbance on the standard deviation, which should be avoided. On the other hand, the increase mutation factor causes the abrupt convergence of the least cost having a similar deviation conversion as of the initial case scenario.

6 Conclusions

It was concluded that the paper presents an appropriate solution for an optimized deployment of a subsea infrastructure by means of a KM and GA-based routine. A satisfactory optimization of the total cost was achieved, and an appropriate narrowing of the individual's fitness was also accomplished. The authors also acknowledge that there is enough room to innovate as multiple deployment restrictions have to be accounted for in such models.

The authors trust that a successful tool will certainly include additional optimization tiers (still using the GA or other suitable algorithm) so that topics like the substation positioning vs the HV umbilical, avoidance of unsuitable paths for the links or maximum tie-in length topics are addressed.

Acknowledgements. Rui Castro was supported by national funds through Fundação para a Ciência e a Tecnologia (FCT) with reference UID/CEC/50021/2019.

This work was supported by national funds through FCT – Fundação para a Ciência e a Tecnologia, CTS – Centre of Technology and Systems, with reference UID/EEA/00066/2019.

References

1. IEA: Offshore Energy Outlook (2017)
2. Rajashekara, K., Krishnamoorthy, H.: Power electronics for subsea systems: challenges and opportunities. In: Proceedings of the International Conference on Power Electronics and Drive Systems (2018)
3. Rajashekara, K., Krishnamoorthy, H.S., Naik, B.S.: Electrification of subsea systems: requirements and challenges in power distribution and conversion. CPSS Trans. Power Electron. Appl. **2**, 259–266 (2017). https://doi.org/10.24295/CPSSTPEA.2017.00024
4. Shi, G., Peng, S., Cai, X., et al.: Grid integration of offshore wind farms and offshore oil/gas platforms. In: Conference Proceedings - 2012 IEEE 7th International Power Electronics and Motion Control Conference - ECCE Asia, IPEMC 2012, pp 1301–1305 (2012)
5. Hamdan, H.A., Kinsella, B.: Using a VSC based HVDC application to energize offshore platforms from onshore - a life-cycle economic appraisal. Energy Proc. **105**, 3101–3111 (2017)
6. Yamamoto, M., Almeida, C.F.M., Angelico, B.A., et al.: Integrated subsea production system: an overview on energy distribution and remote control. In: 2014 IEEE Petroleum and Chemical Industry Conference - Brasil, PCIC Brasil (2014)
7. Hazel, T., Baerd, H.H., Legeay, J.J., Bremnes, J.J.: Taking power distribution under the sea: design, manufacture, and assembly of a Subsea electrical distribution system. IEEE Ind. Appl. Mag. (2013). https://doi.org/10.1109/MIAS.2012.2215648
8. Wang, Y., Gu, J., Duan, M., et al.: A new partition model for the optimization of subsea cluster manifolds based on the new definition of layout cost. Proc. Inst. Mech. Eng. Part M J. Eng. Marit. Environ. (2016). https://doi.org/10.1177/1475090215590466
9. Wang, Y., Duan, M., Feng, J., et al.: Modeling for the optimization of layout scenarios of cluster manifolds with pipeline end manifolds. Appl. Ocean Res. (2014). https://doi.org/10.1016/j.apor.2014.02.006

10. Taroco, C.G., Carrano, E.G., Neto, O.M., et al.: A faster genetic algorithm for substation location and network design of power distribution systems. In: 2012 IEEE Congress on Evolutionary Computation, pp. 10–15 (2012)
11. Vahedi, S., Banejad, M., Assili, M.: Pseudo-dynamic substation expansion planning using hybrid heuristic and genetic algorithm. In: 2015 4th International Conference on Electric Power and Energy Conversion Systems (EPECS) (2015)
12. Gen, M., Syarif, A.: Hybrid genetic algorithm for multi-time period production/distribution planning. Comput. Ind. Eng. **48**, 799–809 (2005)
13. Ghania, S.M.: Optimum design of grounding system inside high voltage substations for transient conditions. In: ICHVE 2016 – 2016 IEEE International Conference on High Voltage Engineering and Application (2016)
14. Hinow, M.: Genetic algorithm based methodology for optimisation of innovative switchgear design. In: Conference Record of IEEE International Symposium on Electrical Insulation (2008)
15. Hinow, M., Mevissen, M.: Substation maintenance strategy adaptation for life-cycle cost reduction using genetic algorithm. IEEE Trans. Power Deliv. (2011). https://doi.org/10.1109/TPWRD.2010.2065247
16. Adler, D.: Genetic algorithms and simulated annealing: a marriage proposal. In: IEEE International Conference on Neural Networks - Conference Proceedings (1993)
17. Hongbo, S., Shuxia, L.I., Degang, G., Peng, L.: Genetic simulated annealing algorithm-based assembly sequence planning. In: International Technology and Innovation Conference, ITIC 2006 (2006)
18. Rexhepi, A., Maxhuni, A., Dika, A.: Analysis of the impact of parameters values on the Genetic Algorithm for TSP. IJCSI Int. Conf. Comput. Sci. Issues **10**, 158 (2013)
19. Asa, S., Noel, C., Alford, M., Lanier, S.: Subsea electrical power standardization. In: PCIC Europe 2013, pp. 1–7. IEEE (2013)
20. Rothlauf, F.: Representations for Genetic and Evolutionary Algorithms, 2nd edn. Springer, New York (2006)
21. Bai, Y., Bai, Q.: Subsea Engineering Handbook. Elsevier, Burlington (2010)
22. Rothlauf, F.: Representations for evolutionary algorithms (2012)

Smart Manufacturing

Production Scheduling Requirements to Smart Manufacturing

Duarte Alemão(✉), André Dionísio Rocha, and José Barata

UNINOVA, CTS, Faculdade de Ciências e Tecnologia,
NOVA University of Lisbon, Caparica, Portugal
{d.alemao,andre.rocha,jab}@uninova.pt

Abstract. The production scheduling has attracted a lot of researchers for many years, however most of the approaches are not targeted to deal with real manufacturing environments, and those that are, are very particular for the case study. It is crucial to consider important features related with the factories, such as products and machines characteristics and unexpected disturbances, but also information such as when the parts arrive to the factory and when should be delivered. So, the purpose of this paper is to identify some important characteristics that have been considered independently in a lot of studies and that should be considered together to develop a generic scheduling framework to be used in a real manufacturing environment.

Keywords: Production scheduling · Manufacturing systems · Smart manufacturing · Scheduling requirements

1 Introduction

The manufacturing systems are going through a significant transformation since market has evolved to a more diversified and customized products demand during the last decades. To deal with this process, factories should adapt themselves and adopt a more dynamic and agile paradigm, able to quickly reconfigure and adapt to the market requirements. The raising of the 4th Industrial Revolution (4IR), encompassing characteristics and paradigms such as Cyber-physical Production Systems (CPPS), Internet of Things (IoT), Cloud-based manufacturing or Radio Frequency Identification (RFID), is allowing to develop more dynamic and agile approaches which improve the efficiency of the manufacturing systems.

In order to achieve the needed productivity on the shop-floor, it is crucial to develop an efficient schedule of what, where and when is going to be processed within the factory. Nevertheless, the scheduling process can be a hard task to perform in a real manufacturing system, regarding all the complex constraints and limitations that can arise from a dynamic system.

To deploy a generic scheduling solution in a factory it is necessary to consider and combine some features that are sometimes ignored in the literature, mostly due to the complex implementation of them. Although there are functional implementations in real manufacturing systems, those are focused on the specific problem. On the other hand, most studies are focused on testing benchmarks and not developing strategies to

Published by Springer Nature Switzerland AG 2019
L. M. Camarinha-Matos et al. (Eds.): DoCEIS 2019, IFIP AICT 553, pp. 227–237, 2019.
https://doi.org/10.1007/978-3-030-17771-3_19

be deployed in real environments as demonstrated in Çaliş and Bulkan [1]. Nowadays, with the 4th Industrial Revolution it is necessary to understand which functionalities the factories need in this context. This paper intends to highlight some technical characteristics that are usual in real environments however are sometimes ignored in literature. Bring them together can contribute to develop a robust generic solution, able to be implemented in a real manufacturing system.

The paper is organized as followed: in Sect. 1 is given an overview of the paper objective; Sect. 2 does the link between the paper content and the Industrial and Service Systems; Sect. 3 provides a brief overview of smart manufacturing; Sect. 4 consists in the analysis of scheduling solution; in Sect. 5 is done a summary of the analysis; in Sect. 6 a conclusion of the work, as well as the further work, is presented.

2 Relation to Industrial and Service Systems

Industrial manufacturing systems are evolving, and the trending paradigm of mass production has been changed to mass customization. The smart factories have emerged, constituted by smart entities along the shop-floor, as well as a lot of sensors, and able to improve processes through self-optimization. By combining production, data and communication technologies it allows the integration of the entire manufacturing supply chain. To deal with the rigorous market demand it is necessary to have an excellent coordination and organization in manufacturing systems.

This paper aims to point out some features that should be considered when developing scheduling solutions and which are often ignored or considered as having no value in a lot of studies. Sometimes these characteristics are not considered due to the raise in complexity that involves considering such a number of details, and this also creates a problem in terms of the performance of the algorithms.

Therefore, instead of worrying with developing better algorithms in terms of performance, this work focusses on searching for requirements that will make possible to develop a generic approach that can be used in real manufacturing systems.

3 Smart Manufacturing

The factories have been obliged to adapt their systems to respond efficiently to the current trend of highly customized products. Thus, they evolved from Dedicated Manufacturing Lines (DML), dedicated to mass production, to Flexible Manufacturing Systems (FMS) which were adopted to provide responsiveness and agility to the systems, once it was able to ease the changes in products and resources [2]. However, the huge demand from the customers led to an insufficient production rate which did not provide the necessary response to the market [2, 3].

The necessity for more responsiveness systems able to launch new products and to react quickly and cost-effectively to market changes, as well as order changes and unexpected failures led to a new manufacturing approach known as Reconfigurable Manufacturing Systems (RMS) [4, 5]. As stated in [2] RMS "are not constrained by capacity or by functionality, and are capable of changing over time in response to changing market circumstances". The crucial scopes of RMS are [5]: modularity, integrability, flexibility, scalability, convertibility and diagnosability.

Following the trend of the 4IR, with the development of IoT, Information and Communications Technology (ICT) and CPPS, factories are becoming more digital and smarter, by connecting and networking the manufacturing units with each other [3]. The use of more sensors combining with the collaboration between entities makes it possible to extract more data from the system, which should be converted in useful information. Thus, the traditional machines are able to become aware of the environment and provide self-optimization. This enables the companies to deal with the challenge of producing more customized and high-quality products demanded by the market [3, 6–8].

4 Scheduling

The production scheduling plays an important role in smart manufacturing systems. To keep it reliable it is crucial to have an online representation of the production and maintenance processes, which can dynamically adapt to the environment, since unexpected events may occur, machines may not be available all the time, and setup times could be sequence-dependent [11].

The development of CPPS, as well as the IoT, brought a strong connection between hardware and software to the factories, through the development of sensor networks equipped with computational intelligence. It allows the real-time access to the system data, providing the information to build and rebuild a schedule faster than ever before, improving the agility and robustness of manufacturing systems. The use of cloud computing and web services can also make the systems more dynamic, reconfigurable, scalable and modular, making possible to develop more robust and adaptable scheduling solutions. The cloud manufacturing, which relies on the is supported by the cloud computing, IoT, virtualization and service-oriented technologies allows to smartly manage the resources and products on the shop-floor and makes it easy to deal with scalability [8, 12].

Using new technologies, such as the use of sensors in robots, integrated with an efficient analysis of the collected data can provide the basis to produce an intelligent schedule of both preventive and corrective operations [3]. Besides the production and maintenance scheduling are often executed independently in real manufacturing systems, sometimes the systems can benefit if they are performed together [18].

Currently, and responding to the change in manufacturing paradigm, enterprises have been focus in factors such as time, quality, environment, society, knowledge, and many others, instead of just focusing in reduce the cost production and increase the yield, which can create and add more value to the company [9, 10].

4.1 Scheduling Requirements

In manufacturing systems, it is possible to find a lot of requirements and constraints to run the system efficiently. Regarding the scheduling process, there are numerous requirements that can be listed to achieve a robust and efficient solution to deal with the factory constraints. Besides basic and common assumptions such as only one operation of each job may be processed at a time, no machine may process more than one operation at a time, an operation can only be performed once and the operation precedence within a product should be respected [13], here other important requirements that should be considered when developing a scheduling solution are presented.

There is a large range of aspects considered when developing a scheduling system, such as economic, managerial, societal and environmental aspects. Most of them focus on classical aspects such as flexibility, transportation time, processing time variations, setup times or maintenance activities [10, 14]. More recently, the research community have been focusing on solving problems such as energy cost, carbon emission or other pollutant gas emissions [14].

Chou, Cao and Cheng [15] developed a dynamic scheduling considering machine breakdown, and able to recover from its faults and prevent them from happening. It was proposed an agent-based solution to solve a distributed job-shop scheduling problem. The system is able to self-configure and to manage itself with barely human interaction, and deal with possible system disturbances. Furthermore, the solution tries not only to minimize the makespan but also the energy consumption. Shahrabi, Adibi and Mahootchi [16] proposed a dynamic scheduling approach considering both machine breakdowns and random products arrival to the system. Kaplanoğlu [11] proposed a dynamic a scheduling approach for a single machine considering dynamic arrival of jobs without the need to re-solving all the problem, due to the flexibility of decentralized control systems. Moreover, it considers maintenance activities, both periodic and non-periodic, where the machine needs a maintenance operation when a deterioration threshold is reached, and a setup time is required before each operation starts and it is based on the previous processed product in that machine. In Zhang et al. [17] was applied the concept of Internet of Manufacturing Things and Real-time Production Scheduling with real-time maintenance operations considering the energy consumption of maintenance, the total maintenance cost and the loading rate of workers to develop a real-time data-driven production scheduling which gets real-time status and information from the shop-floor components, as well as information about unexpected events. Thus, it is possible to dynamically monitor the systems to provide optimal updated schedule solutions. The approach tries to minimize the assembly costs, the energy consumption, and the tardiness delivery rate, as well as to improve the resources management. Liu, Dong and Chen [18] developed a scheduling solution considering both production and maintenance planning in a single machine. The deterioration state of the machine is considered at each time and a predictive maintenance task can be allocated. The solution combines analytical information, such as machine deterioration, and the available resources to get an optimal scheduling considering predictive maintenance activities. This will improve the machine lifetime.

Shen, Dauzère-Pérès and Neufeld [19] considered a flexible scheduling to solve a job-shop problem with flexible machines, where each machine is able to execute more than one operation. The solution also considers sequence-dependent setup times. The setup times can be dependent on the machine or on the sequence of operations processed. The conducted experiments showed that the integration of structural properties in scheduling problems is extremely important. In Lu et al. [20] was developed an approach to solve a flexible job-shop scheduling where an operation of a product can be processed by more than one machine. Chang and Liu [21] proposed an approach to solve a distributed (considering different factories) and flexible job-shop scheduling problem, where products can be assigned to different factories and different machines in each factory. Each machine is able to execute more than one operation, and each job can be performed in a set of possible machines. The solution considers the delivery times of the products, after the process is finished. In Han et al. [22] was proposed a hybrid particle swarm optimization algorithm to solve the deadlock-free scheduling problem in a flexible manufacturing system considering both routing flexibility and machine flexibility.

Alotaibi, Lohse and Vu [23] proposed a multi-agent system solution which eases the reconfigurability and adaption of the schedule in response to unpredictable changes. The main goal of the agents is to reach a global objective from local agent solutions. It aims not only to minimize the tardiness but also to minimize the total energy consumption. A particle swarm optimization together with chaos theory was used in Petrović et al. [24] to solve the process planning and scheduling problem considering the transportation time between different machines. The solution contemplates the production time which considers the processing time of the operation, the transportation time, the total tool change time and the total setup change time. Karimi et al. [13] developed a solution which includes the transportation times between the machines in a flexible job-shop scheduling problem, where the transportation is dependent of the distance between machines and the type of job to be transported. However, it is considered that the number of transporters is infinite, which is not true in real manufacturing systems, and so there is no transportation delays. Jia et al. [25] used an objective function to solve the job-sop scheduling problem in a distributed manufacturing system which considers the transportation cost of the parts between factories.

Yazdani et al. [26] proposed a solution for solving the job-shop scheduling problem with multiple machines where is introduced an objective function with the sum of maximum earliness and tardiness criteria. In Bürgy and Bülbül [27] was developed an approach to solve the job-shop scheduling problem where the storage, earliness and tardiness of products are penalized with different costs, with the objective of finish them just in time. Both release and due dates are considered, so the parts are not always available. Kuhpfahl and Bierwirth [28] developed a solution to solve the total weighted tardiness objective in a flow-shop problem.

Chung and Kim [29] try to solve the scheduling problem in order to the makespan on a single machine with step-deterioration and rate-modifying activities which affect the processing times of the operations. If a value is higher than a given deterioration threshold, then a step deterioration is added to the processing time. A cloud-based solution was developed in Helo, Phuong and Hao [30] to solve a flow-shop scheduling problem. This allows to perform near real-time analysis and collect and update the processing times of each machine very quickly. In Zarook and Abedi [31] was developed an approach to solve a parallel-machine scheduling problem where the time affects the performance of the machines until they are subjected to maintenance. The time to operate a task increases due to the wear caused by the already processed products, that is why it is necessary to consider flexible processing times. The objective is to minimize the earliness and tardiness, and the maintenance cost which is also considered. Just-in-time is pursued by operations managers for achieving both customer and manufacturer satisfaction simultaneously.

Zhang, Gao and Li present [32] present a scheduling solution to solve a job-shop problem, considering random job arrivals and machine breakdowns. Thus, an operation can be interrupted due to machine breakdown. The remaining processing time of the operation is then equal to the total time minus the time already processed. The objective is to minimize the makespan and the stability after performing the reschedule. Nikolakis [33] adopted a strategy to reschedule a task (which is divided in operations) in case of occurrence of unexpected events. The task can be rescheduled entirely or from the beginning of a specific operation. The main goal is to minimize the mean flow time, but "there is no restriction on the kind and number of criteria that can be selected".

Xiong et al. [34] developed a scheduling approach to solve a dynamic job-shop problem, where new jobs can arrive during the process, using dispatching rules considering the precedence constraint between products, and to minimize the tardiness. Furthermore, release and due dates of the parts are considered, which reflects and important aspect in manufacturing systems.

Chan, Choy and Bibhushan [35] developed a solution to solve the job-shop scheduling problem in a flexible shop with the objective of minimizing the makespan. It the study was considered that the parts have a release and due date, and setup times were taken into account.

Azami, Demirli and Bhuiyan [36] developed a solution to solve a specific real-world hybrid flow-shop problem, where machines are arranged into different stages in series. Some constraints were considered such as a limited number of workstations in each cell, the parts cannot stay too long in the buffer area, the buffer has a limited capacity and different parts need to be grouped before being sent for the cure process. The parts also have a due date to be finished. Gao et al. [37] studied a real situation of a remanufacturing enterprise to perform the rescheduling process in a job-shop. Thus, they considered new job insertions during the process, as well as flexible machines able to perform different operations. The objective was to minimize the instability regarding the first scheduling, as well as to minimize the makespan or the total flowtime or the machine workload or the total machine workload.

5 Analysis and Discussions

Based on the analyzed papers, it is possible to come out with some common requirements which are considered when developing scheduling solutions. The percentage of each characteristic present in the articles is shown in Fig. 1. The percentage of the number of characteristics studied together in each paper is shown in Fig. 2.

The real manufacturing environments need to deal with **dynamic events** that sometimes are not expected and may occur anytime. Those unexpected disturbances that can happen during the manufacturing process may be changes in the shop-floor, the arrival of new products to the factory or unexpected machine breakdowns, and a dynamic schedule should be able to respond to it in real-time [15, 16]. Furthermore, to respond to the market demand for more customized products, manufacturing systems

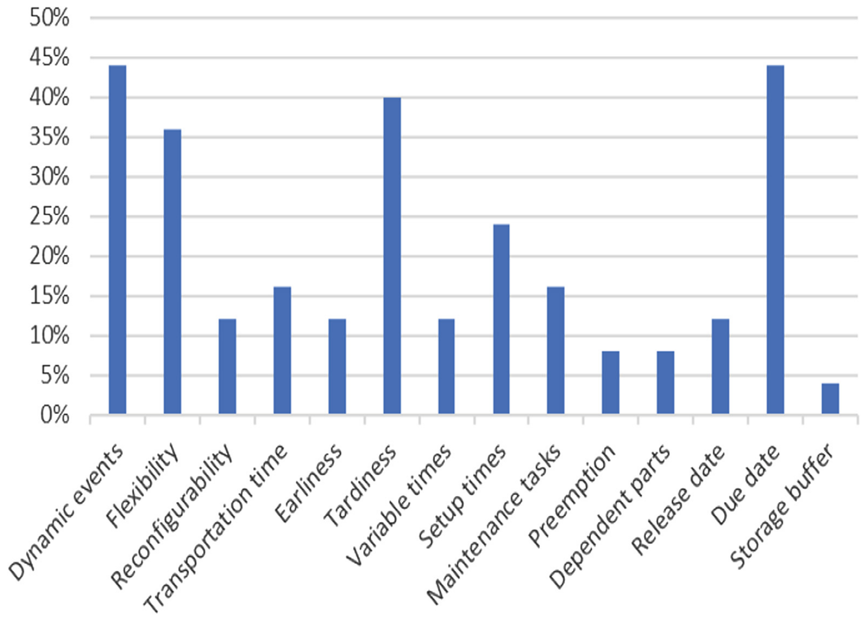


Fig. 1. Percentage of characteristics present in analyzed papers

are becoming more **flexible**, able to produce different products by sharing tools. In a flexible system an operation can be executed in more than one machine (routing flexibility) or each machine can be able to perform more than one operation by sharing resources (machine flexibility). More factories are adopting flexible machines, able to perform more than a unique task, thus a scheduling considering flexible machines is important in the context of the 4IR [13, 20, 22, 38]. Likewise, dynamic events and the uncertainty during the processes makes the way for **reconfigurable** systems, able to respond autonomously and deal with those events [23]. Most of the times machines and tools are subject to deterioration. So, **maintenance** activities play a crucial role in manufacturing systems and should be considered in scheduling solutions in order to achieve a better performance of the system, either a pre (to avoid) or post-failure (to recover) solution [11, 17, 18]. A less common requirement, but which can be important in some situations is **preemption**. Although it could be hard to handle, due to the implementation complexity, sometimes could be important to restart a process from the same point when an operation has been interrupted, so it could improve the performance of some production system. Another not commonly discussed characteristic is the **precedence** constraints among the operations of different products. For example, in the assembly process of two or more parts [34]. However, it is hard to find this constraint being considered in manufacturing scheduling studies.

Some other features present in real environments, often ignored in literature, are: the transportation time of the parts within the factory, a crucial feature to plan an efficient production that can vary a lot during the process [13]; most of the times the products need to be ready for delivery at some time and it could be crucial to not overcome those dates (**tardiness**). On the other hand, it could be important to not finish the products too soon (**earliness**) as well, since it can lead to some wear in the parts or involve storage costs [26, 28]; **processing times may vary** in real manufacturing systems due to the most diverse situations, and they mainly increase over time. It can be caused by resources deterioration, a fault in the setup, or the surrounding conditions [29, 31]; the **setup times**, which play a crucial role in real manufacturing systems. In some cases setup times are required amongst operations and may depend on the precedent process [11, 19]; **release** and **due dates** can be crucial in manufacturing systems and are often present in real situations. An operation should be scheduled to a date after the part is available. On the other hand, there are due dates to be met and when a product should be ready. Also, finishing a product before the due date means that it needs to be stored in the factory and besides taking storage space it may be subject to some wear. The **buffer** storage, often ignored when developing algorithms, is also a crucial feature in manufacturing environments, since factories do not have infinite storage to hold the products.

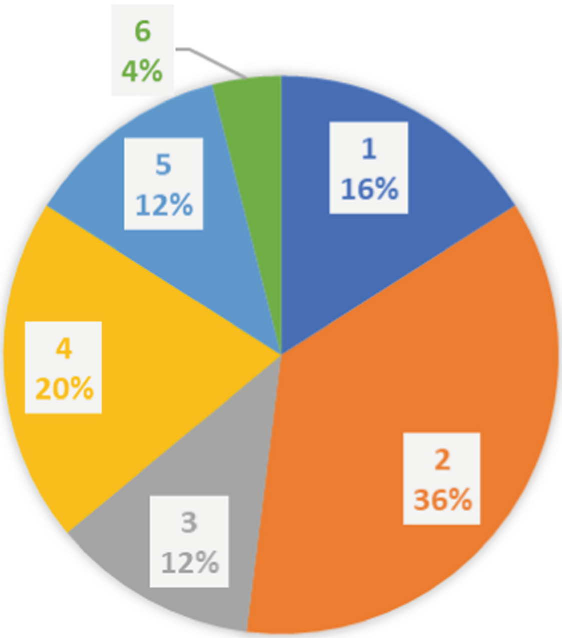


Fig. 2. Number of characteristics taken into account in the same study based on the analyzed papers.

6 Conclusion and Further Work

This paper conducts a comprehensive review of classical requirements of scheduling that need to be considered simultaneously in order to develop a generic scheduling framework. By analyzing the literature, it is possible to identify common requirements that are important to consider to scheduling manufacturing approaches. However, most studies tend to neglect some of these characteristics, trying to achieve better simulation results, instead of validating solutions on industrial environments. Even though the complexity of the solution can increase a lot when considering all of these conditions, it is important to develop approaches able of being transferred to real manufacturing environments.

This paper presents a collection of some crucial requirements important to consider when developing a production scheduling solution for manufacturing systems. It is critical nowadays, in an era where the factories are becoming more digital and have more data to be processed than ever before, to have more organized and well-structured manufacturing solutions, starting by the scheduling of the production.

All the requirements presented before are important and should be considered in order to develop a generic scheduling solution to be used in a real manufacturing system, even if some of them are not so important in a specific case. However, a complete and robust generic approach should not ignore and not be restricted to these characteristics only. The 4IR makes it possible to develop flexible, reconfigurable and efficient production systems, and acquire information provided by the sensors, making possible to react to environment disturbances in real-time.

It is important to put some effort in connect and relate what already exists instead of only try to create and improve algorithms to test on benchmark and focus in just a few aspects of the scheduling.

In further work it is necessary to study the importance of characteristics such as societal and environmental characteristics, as well as to understand how the 4IR will help the scheduling process, by autonomously and dynamically adapt the scheduling to the current environment in a smoother way. The development of a framework able to deal with all these different kind of aspects in a generic way, making it possible to use the same scheduling approach for different manufacturing scenarios will be considered.

References

1. Çaliş, B., Bulkan, S.: A research survey: review of AI solution strategies of job shop scheduling problem. J. Intell. Manuf. **26**(5), 961–973 (2015)
2. Koren, Y., Shpitalni, M.: Design of reconfigurable manufacturing systems. J. Manuf. Syst. **29**(4), 130–141 (2010)
3. ElMaraghy, H., ElMaraghy, W.: Smart adaptable assembly systems. Procedia CIRP **44**, 4–13 (2016)
4. Koren, Y., et al.: Reconfigurable manufacturing systems. CIRP Ann. **48**(2), 527–540 (1999)
5. Ribeiro, L., Barata, J.: Re-thinking diagnosis for future automation systems: an analysis of current diagnostic practices and their applicability in emerging IT based production paradigms. Comput. Ind. **62**(7), 639–659 (2011)

6. Gao, Q., Shi, R., Wang, G.: Construction of intelligent manufacturing workshop based on lean management. Procedia CIRP **56**, 599–603 (2016)
7. Lee, J., Kao, H.A., Yang, S.: Service innovation and smart analytics for Industry 4.0 and big data environment. Procedia CIRP **16**, 3–8 (2014)
8. Zhong, R.Y., Xu, X., Klotz, E., Newman, S.T.: Intelligent manufacturing in the context of industry 4.0: a review. Engineering **3**(5), 616–630 (2017)
9. Tao, F., Cheng, Y., Zhang, L., Nee, A.Y.C.: Advanced manufacturing systems: socialization characteristics and trends. J. Intell. Manuf. **28**(5), 1079–1094 (2017)
10. Koren, Y., Gu, X., Badurdeen, F., Jawahir, I.S.: sustainable living factories for next generation manufacturing. Procedia Manuf. **21**, 26–36 (2018)
11. Kaplanoğlu, V.: Multi-agent based approach for single machine scheduling with sequence-dependent setup times and machine maintenance. Appl. Soft Comput. J. **23**, 165–179 (2014)
12. Yao, X., Zhou, J., Lin, Y., Li, Y., Yu, H., Liu, Y.: Smart manufacturing based on cyber-physical systems and beyond. J. Intell. Manuf. 1–13 (2017). https://doi.org/10.1007/s10845-017-1384-5
13. Karimi, S., Ardalan, Z., Naderi, B., Mohammadi, M.: Scheduling flexible job-shops with transportation times: mathematical models and a hybrid imperialist competitive algorithm. Appl. Math. Model. **41**, 667–682 (2016)
14. Hongying, F., Qian, L., Dan, S.: A survey of recent research on optimization models and algorithms for operations management from the process view. Sci. Program. **2017**, 1–19 (2017)
15. Chou, Y.C., Cao, H., Cheng, H.H.: A bio-inspired mobile agent-based integrated system for flexible autonomic job shop scheduling. J. Manuf. Syst. **32**(4), 752–763 (2013)
16. Shahrabi, J., Adibi, M.A., Mahootchi, M.: A reinforcement learning approach to parameter estimation in dynamic job shop scheduling. Comput. Ind. Eng. **110**, 75–82 (2017)
17. Zhang, Y., et al.: The 'Internet of Things' enabled real-time scheduling for remanufacturing of automobile engines. J. Clean. Prod. **185**, 562–575 (2018)
18. Liu, Q., Dong, M., Chen, F.F.: Single-machine-based joint optimization of predictive maintenance planning and production scheduling. Robot. Comput. Integr. Manuf. **51** (January), 238–247 (2018)
19. Shen, L., Dauzère-Pérès, S., Neufeld, J.S.: Solving the flexible job shop scheduling problem with sequence-dependent setup times. Eur. J. Oper. Res. **265**(2), 503–516 (2018)
20. Lu, P.-H., Wu, M.-C., Tan, H., Peng, Y.-H., Chen, C.-F.: A genetic algorithm embedded with a concise chromosome representation for distributed and flexible job-shop scheduling problems. J. Intell. Manuf. **29**(1), 19–34 (2018)
21. Chang, H.-C., Liu, T.-K.: Optimisation of distributed manufacturing flexible job shop scheduling by using hybrid genetic algorithms. J. Intell. Manuf. **28**(8), 1973–1986 (2017)
22. Han, L., Xing, K., Chen, X., Xiong, F.: A Petri net-based particle swarm optimization approach for scheduling deadlock-prone flexible manufacturing systems. J. Intell. Manuf. **29** (5), 1083–1096 (2015)
23. Alotaibi, A., Lohse, N., Vu, T.M.: Dynamic agent-based bi-objective robustness for tardiness and energy in a dynamic flexible job shop. Procedia CIRP **57**, 728–733 (2016)
24. Petrović, M., Vuković, N., Mitić, M., Miljković, Z.: Integration of process planning and scheduling using chaotic particle swarm optimization algorithm. Expert Syst. Appl. **64**, 569–588 (2016)
25. Jia, H.Z., Fuh, J.Y.H., Nee, A.Y.C., Zhang, Y.F.: Integration of genetic algorithm and Gantt chart for job shop scheduling in distributed manufacturing systems. Comput. Ind. Eng. **53** (2), 313–320 (2007)
26. Yazdani, M., Aleti, A., Khalili, S.M., Jolai, F.: Optimizing the sum of maximum earliness and tardiness of the job shop scheduling problem. Comput. Ind. Eng. **107**, 12–24 (2017)

27. Bürgy, R., Bülbül, K.: The job shop scheduling problem with convex costs. Eur. J. Oper. Res. **268**(1), 82–100 (2018)
28. Kuhpfahl, J., Bierwirth, C.: A study on local search neighborhoods for the job shop scheduling problem with total weighted tardiness objective. Comput. Oper. Res. **66**, 44–57 (2016)
29. Do Chung, B., Kim, B.S.: A hybrid genetic algorithm with two-stage dispatching heuristic for a machine scheduling problem with step-deteriorating jobs and rate-modifying activities. Comput. Ind. Eng. **98**, 113–124 (2016)
30. Helo, P., Phuong, D., Hao, Y.: Cloud manufacturing – scheduling as a service for sheet metal manufacturing. Comput. Oper. Res. **0**, 1–12 (2018)
31. Zarook, Y., Abedi, M.: JIT-scheduling in unrelated parallel-machine environment with aging effect and multi-maintenance activities. Int. J. Serv. Oper. Manag. **18**(1), 99 (2014)
32. Zhang, L., Gao, L., Li, X.: A hybrid genetic algorithm and tabu search for a multi-objective dynamic job shop scheduling problem. Int. J. Prod. Res. **51**(12), 3516–3531 (2013)
33. Nikolakis, N., Kousi, N., Michalos, G., Makris, S.: Dynamic scheduling of shared human-robot manufacturing operations. Procedia CIRP **72**, 9–14 (2018)
34. Xiong, H., Fan, H., Jiang, G., Li, G.: A simulation-based study of dispatching rules in a dynamic job shop scheduling problem with batch release and extended technical precedence constraints. Eur. J. Oper. Res. **257**(1), 13–24 (2017)
35. Chan, F.T.S., Choy, K.L., Bibhushan, : A genetic algorithm-based scheduler for multiproduct parallel machine sheet metal job shop. Expert Syst. Appl. **38**(7), 8703–8715 (2011)
36. Azami, A., Demirli, K., Bhuiyan, N.: Scheduling in aerospace composite manufacturing systems: a two-stage hybrid flow shop problem. Int. J. Adv. Manuf. Technol. **95**(9–12), 3259–3274 (2018)
37. Gao, K., Yang, F., Zhou, M., Pan, Q.: Flexible job-shop rescheduling for new job insertion by using discrete Jaya algorithm. IEEE Trans. Cybern. **49**(5), 1944–1955 (2019)
38. Lei, H., Xing, K., Han, L., Gao, Z.: Hybrid heuristic search approach for deadlock-free scheduling of flexible manufacturing systems using Petri nets. Appl. Soft Comput. **18**(2), 240–245 (2017)

Open Modular Components in the Industry Using vf-OS Components

Joao Giao(✉), Joao Sarraipa, and Ricardo Jardim-Gonçalves

CTS, UNINOVA, Dep.º de Eng.ª Electrotécnica,
Faculdade de Ciências e Tecnologia, FCT,
Universidade Nova de Lisboa, Caparica, Portugal
{jgs, jfss, rg}@uninova.pt

Abstract. The increasing complexity in industrial and information technology in the last decades forced the manufacturing process to adapt to these new evolution trends. Factories are using the advancement of the Internet of Things (IoT) to have multiple platforms with sensorial information into their factory processes, to improve and reduce the costs of their products and increase their profit. To address these complexes and useful computations, the vf-OS (Virtual Open Operating System) aims to facilitate the development of applications using their individual components as well as FIWARE enablers.

Keywords: Industry · IoT · Sensor · vf-OS

1 Introduction

An increasing number of heterogeneous smart objects communicate and interconnect through the internet. However, to use the full potential of these smart objects, it is necessary to transcend the interoperability issues between them. In the past years, there has been efforts to overcome this issue at system and network level. At system level, there can be a lack of system's interoperability, which makes necessary rewriting the software code for every new smart component that is inserted into the old system. At the network level, the smart components cannot communicate with each other because they have different interfaces and API protocols to interconnect with the surrounding environment [1].

Nowadays, it has been emerged a transversal discipline called CPS (Cyber Physical Systems), which integrates various intelligent components that has helped software architects and designers to deal with complex and difficult systems. CPS are combinations of computations and physical processes that enable the control of physical processes with computing and communication infrastructures [2, 3]. CPS allows sensors and actuators to be controlled by a control unit, usually composed by more than one microcontroller, where physical processes affect computations and in the other way around as well. The design of these systems requires from experts a complete knowledge about the computers dynamics (e.g. software or networks) and the physical processes to deal with big and complex data flows. In order to ease the development and deployment of new services that use such kind of data processes, smaller operating systems have been developed to help in the building of applications that use IoT

Published by Springer Nature Switzerland AG 2019
L. M. Camarinha-Matos et al. (Eds.): DoCEIS 2019, IFIP AICT 553, pp. 238–246, 2019.
https://doi.org/10.1007/978-3-030-17771-3_20

devices, like the vf-OS (Virtual Factory Open Operating System). The vf-OS system has been developed by the European Commission co-funded project called vf-OS [4]. vf-OS provides the necessary software tools to connect a factory and their internal devices by managing a factory related computer hardware and software resources, providing mutual services for factory computational programs.

Therefore, the main research question that emerges is:

Can modular systems contribute to the development and deployment of CPS in manufacturing systems?

One additional motivation is to use open-source software to speed up the software development and minimize the costs. As such, this paper proposes an architecture that explores the potential usage of open source software components applied in the industry, using some of the vf-OS system components already available. Accordingly to the Open Source Initiative [5], an open source software is a software that can be used without costs, changed and shared with anyone. Additionally, IoT components together with open standard protocols, can both assure interoperability and hence present a unified and a stable software infrastructure.

2 Related Literature

In smart factories, condition monitoring systems (CMS) represent the business logic that monitor the state variables and create diagnostics of current and future factory' errors. However, most factories have several tasks depending on the associated production machine or process, which results in varying the requirements for cyber-physical computing architectures. Notwithstanding, in [6] the authors proposed a general processing steps for CMS: (1) State detection - measuring and store machine parameters that represent the current status of the factory. For this stage, sensors are used for monitoring and quality assurance; (2) State comparison – comparison of the current data with predetermined nominal data. After data collection from, e.g., sensors from loads, temperatures, machine parameters and environment conditions, these values are compared with the manufacturer machine supervisor. This state relates with CPS, by having multiple types of sensors for perceiving the factory' environment and efficient mechanisms to process information and send events when is necessary; (3) Diagnosis – Specific error diagnostics. Based on the previous states, it is possible to detect potential errors or anomalies.

2.1 CPS Architectures

A CPS model contains models of software and physical processes, intelligent computation platforms and networks. It requires a multi-disciplinary development, not only on the distinct computational and physical components, but also on their integration and interaction. Accordingly to [7], CPS research has been focused on identifying needs, challenges and industrial sectors opportunities, encouraging multidisciplinary collaborative research between academia and industry. The main goal between this synergy is the development of new engineering techniques and methods to build high-confidence systems, intelligent, secure and integrated at all scales.

In a near future, a factory will have millions of sensors and actuators controlled by thousands of CPUs. To have this sensorial adaptation, the factories' system needs to flexibly adapt to an ever-changing external environment. System's architecture dictates how the passive or active elements can be organized and behave, accordingly to the external environment and the internal conditions defined by controllers [8].

2.2 Modularity

Modularity allows specialized functionalities that can be achieved through a plug-and-play fashion, opening opportunities for innovation and specialized software that allows developing heterogeneous and complex systems. To have this seamless integration of different components, the interfaces of the components must be standardized to facilitate the interoperability between them [7]. Adaptability is one advantage of these modular components. Since factory's machines need to be enhanced over time to add and improve their functionalities, their hardware/software must also adjust to these changes, which can be easily made through changing/adding new components on the process workflow.

Yelamarthi et al. in [9] present that most of the IoT architectures are not prepared to be used in various domains. To overcome this customization problem, the authors proposed a modular IoT architecture that can be configured for different domains, such as smart home, city, industry for environment control, optimizing energy usage and automating applications, agriculture for soil and temperature monitoring, healthcare for monitoring patient physiological parameters or remote motion tracking.

2.3 Warehouse Management Systems

Accordingly to Lee et al. [10], the major challenge in modern warehouse management is the inventory accuracy, space utilisation, process management and optimize when raw materials are ordered. To overcome these challenges, it is necessary an agile supply chain, which can be done with CPS network, that will connect users, objects and physical processes in a warehouse operation. This improvement requires the integration of technological and administrative innovations, which includes the proper selection of CPS technologies, ambient intelligence, timely information flow and agility.

3 vf-OS

IoT platforms are designed on the aspect of collecting data and using it in edge computing platforms for cloud computing. To be used in real-life scenario, the existing software need to be able to simulate and process parts of the production processes, which can be done through vf-OS. vf-OS system will provide a set of manufacturing applications and tools for manufacturing users and software developers, respectively. Thus, they provide a set of services for factory computational programs, which can be used for managing factory related computer hardware and software resources as well as to provide services for factory computational programs. As an open operating system, it has components that allow industrial users to download and upload the applications (vApps) developed with vf-OS. The vApps, as illustrated by Fig. 1, can use the vf-OS components, as a software layer extended over multiple factories, to acquire sensorial information and use the actuators over them.

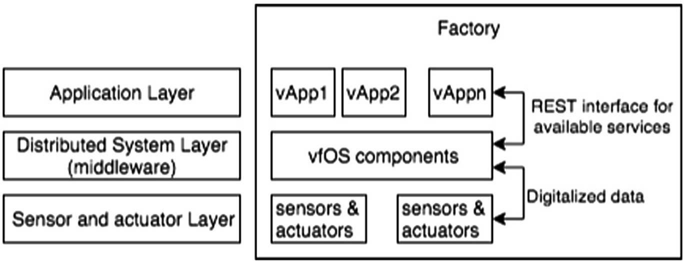


Fig. 1. High-level layers interaction and information flow in a proposed modular architecture for an industrial scenario.

First, the physical entities need to be sensed by a transducer and the electrical property will be converted into a numeral value, done by the sensorial and actuator layer. Then, through the distributed system layer, this raw data will be acquired from different remote locations in the factory' environment to be filtered and send to a common storage to be processed. This same layer will also be responsible to perform the necessary algorithms and operations that are necessary to give new information and functionalities to the factories, which is done afterwards by the application layer.

3.1 vf-OS Modular Components

vf-OS project uses a Service Oriented Architecture (SOA) approach in which all components perform individual functionalities and can connect with each other to create more complex services. This interoperability can be achieved taking into consideration that all components implement and publish a REST interface allowing a common data interface. This has the advantage of being easy to test and modify the service' workflow, by adding/removing components as needed. As presented in [11], the vf-OS building blocks were already used in an industrial implementation scenario to orchestrate a vApp through the vf-OS component "Process Designer", a process model in form of a BPMN (Business Process Model and Notation). Some components able to be used in such BPMN component are presented in the following:

- Enablers Framework (EF) - Acts as a bridge between service provider and service consumer, providing support for enablers' integration, installation and management of their instances. The main added value of this component is to ensure an unique interface and common access for all registered enablers, acting as a wrapper engine for the different enablers and the vApps. For the European commission, these enablers, developed from FIWARE, can be used in multiple domains, e.g., connect to IoT devices, process data and media in real-time, perform Big Data analysis [12], ensuring portability, interoperability and openness of services across Europe, by contributing to the Digitisation of Europe Industry and European Cloud Initiative [13]. EF module represents an advantage for the existing enablers due to the fact that it facilitates the enablers' usage through unique and easy REST API;

- Enablers - Open-source software that exposes heterogeneous services. They can be classified as FIWARE Generic Enablers, Manufacturing enablers and vf-OS enablers. The FIWARE enablers were developed under the FIWARE foundation on the context of the Future Internet Private Public Partnership (FI-PPP) that was part of the FP7 (Seventh Framework Programme) of the European Commission [14]. Their goal was to provide services in different domains such as IoT, data context management, Cloud Hosting, security platforms, etc. [15]. Manufacturing enablers are enablers that were developed with the purpose of providing functionalities for manufacturing industries, such as the developed enablers under the FITMAN (Future Internet Technologies for Manufacturing Industries) project [16]. vf-OS enablers were developed to complement/improve the specific applications developed to be used by the vf-OS users;
- Messaging-Publish/Subscribe (PubSub) - Data infrastructure middleware responsible for the data communication between different components. This component has an open source message broker, RabbitMQ [17], that creates a distributed communication layer for messaging handling with Advanced Message Queuing Protocol (AMQP) [18] standard protocol. It also allows to configure the security policies in order to define which modules have access to specific message queues;
- Datastorage - Scalable data storage system that is capable of handling real-time sensor data and events through a common API (REST). It encapsulates a relational database (for structured information), time series database (for sensorial data) and document oriented database (for unstructured information);
- Drivers - This component provides a collection of libraries that allows to collect data from, and send commands to, physical devices such as industrial automation devices (RFID readers, PLCs, smart sensors, etc.). This software module can connect to sensors and actuators through physical interfaces, communication protocols or intermediary systems such as field bus controllers, providing a flexible and modular approach to interact with devices [15].

3.2 Proposed Architecture and Prototypical Implementation of an Industrial Scenario

In this paper, a generic architecture to integrate different modules in order to be used by a manufacturing factory, is presented. The developed work presented in this paper is in the context of the vf-OS project, where the PhD student, author of this paper, developed the module "Enablers Framework" and the libraries to use the "Messaging and PubSub" services. These modules are central components that enables integration and interoperability required in system modules. Further information about these components is described in next section.

The goal is to allow two different factories to communicate with each other in order to coordinate their production and verify if the factories have stored enough amount of spare parts for future production, as can be seen by Fig. 2.

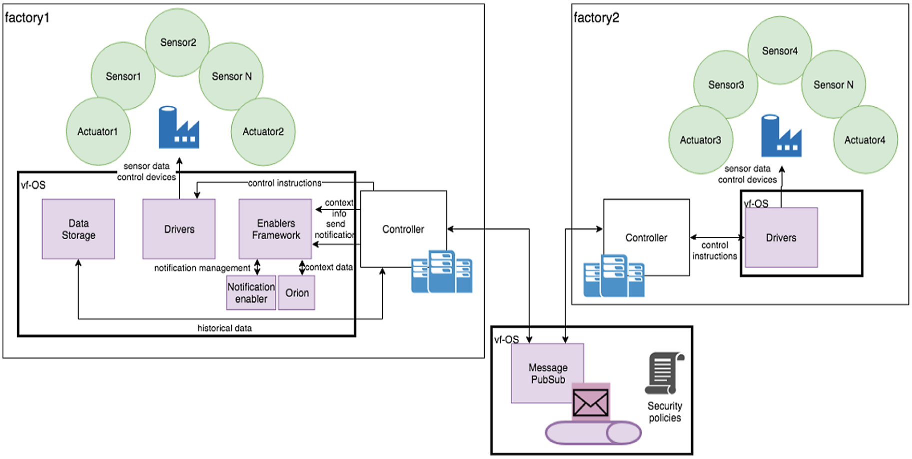


Fig. 2. General overview of the proposed architecture between two different factories and a communication channel for inter-factory resources' planning.

The required modules for the proposed architecture are the following: (1) data storage, a software component responsible to save the historical data of the factory's sensors. A predictive analytics can be used through historical and current data to predict future outcomes [19]; (2) drivers, to interact with physical industrial devices to integrate sensors' data and send commands to them; (3) Enablers Framework, Orion & Notification Enablers. Orion is an enabler that is used for a data/context scenario to update context information, queries and subscriptions. This context-aware component allows to model and gain access to context information connected to sensor networks, users and/or existing industrial applications allowing, for e.g., to receive notifications when some pre-defined condition is accomplished. The Notification enabler is a vf-OS enabler that provides the components of the system an easy to use notification system, to alert when the stock is off limits for each spare-part of the factory. To facilitate the usage of enablers, the EF is used by interpreting the enablers' REST and NGSI [20] protocols into an easy to use and understand REST interface; (4) controller, this module will control the production of the factory, accordingly to the available resources and the output of the other factory; (5) messaging-pub/sub, that will responsible for the communication between the factory' modules and inter-factory communication. By defining a security policy, it is possible to secure the communication channels, granting the access to the authorised components.

4 Industrial Scenario

The presented industrial scenario is related with manufacturing, automation and logistics, which uses an application named "vfStockPolicies" (Pilot 1, application 3) [21]. The main goal of this vApp is to manage spare-parts that each industrial company needs, monitor their stocks and alert when the stock is lower than a minimum defined for each spare-part. Contrarily to other IoT applications described in [9], the goal of this application is not to have a low-power system, but rather having a reliable and wearable devices attached to the existing factory devices. Since this application requires users from different factories, it is crucial that a RBAC (Role based access control) is used. Currently, the messaging-pub/sub component already has integrated an external service which allows the configuration of the authentication (process of ascertaining the user identification though a password) and authorisation (specification of rules that determine who is allowed to use the messaging channels). After the authentication and authorisation permissions, the factory developer must configure the data access of the sensors/PLCs to perform the device integration with the industrial vApp. Taking into consideration that the Driver's module already contains a complete working copy of the IO component binaries as well as many common functionalities for input and output of the sensors and actuators, respectively, the factory developer only needs to develop the software algorithms [22]. This variant has the advantage of the software being transferable and it doesn't require rewrite of the whole code. Another configuration to be made is to define the context rules on Orion enabler and Notification Enabler, which will be used to evaluate a specific case and later validate if there are materials to create a product inside both factories. The communication channels also need to be configured inside the message and PubSub component for the message channel between factories and between the factory's internal components.

The controller will use the ORION context broker to reason the previously formalized rules for the context and validate if both factories have the required spare parts to create the final product. If the stock of the spare parts is lower than a certain threshold, the Notification Enabler will be used to alert the maintenance manager of the company about this situation in order to acquire new materials.

The future works concentrate on the total experimentation of the proposed architecture in the real industrial environment under the last phase of the vf-OS project, represented on.

Figure 3. This application will be executed in different industrial systems in Spain in order to manage the spare parts of two different factories that use materials from each factory to create a new final product.

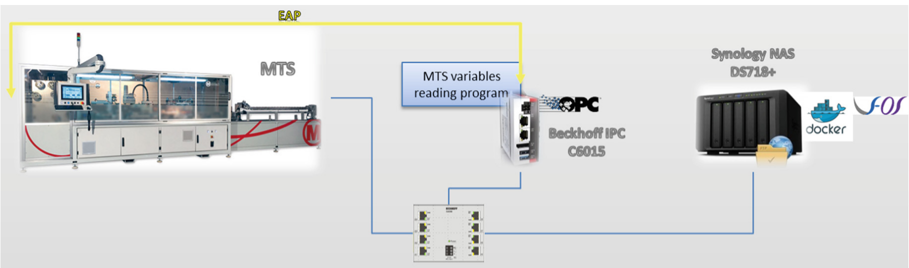


Fig. 3. vfStockPolicies - Pilot 1 - Application 3 from vf-OS project use case.

Another crucial aspect to take into consideration about the implementation of this architecture in a real factory is the security. Factories have sensitive data that must be protected for unauthorized use, so it is mandatory to correctly define a user management to regulate user permissions, especially if there is a communication between different factories.

5 Conclusions

This paper defines an architecture aimed at the software design using open-source modular components to develop an application to be used in the industrial field. It allows researchers to speed and quickly develop, without monetary costs, their IoT applications, without having to understand how each component works and communicate with each other. This modular component approach provides a flexible way of building industrial applications which decouples the complexity of constructing complex CPS into simpler subsystems. Thus, the proposed architecture works as a building CPS catalyst tool. It facilitates the development of intelligent systems that actuate accordingly to its applied environment. In this case, it was used the proposed architecture with an application to remotely determine which spare parts are available on a different factory and schedule the production accordingly to such lack of materials. Furthermore, the application can be used to warn maintenance managers with sufficient time to prevent stock breakout and the consequent repair delay in forthcoming maintenance operations.

Due to the use of such modular components the presented work contributes directly at the system level, where multiple components with different structures interoperate with each other to perform a workflow of services and is able to be customized for a diverse range of applications. On the other hand, at the network level, the solution proposed just uses a communication standard protocol to facilitate the communication between different systems.

Acknowledgments. This work was funded in part by the Center of Technology and Systems (CTS) and the Portuguese Foundation for Science and Technology (FCT) through the Strategic Program UID/EEA/00066/2013.

References

1. Rosenkranz, P., Wählisch, M., Baccelli, E., Ortmann, L.: A distributed test system architecture for open-source IoT software. In: Proceedings of the 2015 Workshop on IoT challenges in Mobile and Industrial Systems - IoT-Sys 2015, pp. 43–48 (2015)
2. Derler, P., Lee, E.A., Vincentelli, A.S.: Modeling cyber–physical systems. Proc. IEEE **100** (1), 13–28 (2012)
3. Jazdi, N.: Cyber physical systems in the context of Industry 4.0. In: 2014 IEEE International Conference on Automation, Quality and Testing, Robotics, pp. 1–4 (2014)
4. vf-OS Wiki. Vf-OS, vf-OS consortium. https://www.vf-os.eu. Accessed 10 Jan 2019
5. Open Source Definition. Open Source Initiative. https://opensource.org/. Accessed 10 Jan 2019

6. Fleischmann, H., Kohl, J., Franke, J.: A modular architecture for the design of condition monitoring processes. Procedia CIRP **57**, 410–415 (2016)
7. Baheti, R., Gill, H.: Cyber-physical Systems, from: The Impact of Control Technology (2011). www.ieeecss.org
8. Hehenberger, P., Vogel-heuser, B., Bradley, D., Eynard, B., Tomiyama, T., Achiche, S.: Design, modelling, simulation and integration of cyber physical systems: methods and applications. Comput. Ind. **82**, 273–289 (2016)
9. Yelamarthi, K., Aman, M.S., Abdelgawad, A.: An application-driven modular IoT architecture. Wirel. Commun. Mob. Comput. **2017** (2017)
10. Lee, C.K.M., Lv, Y., Ng, K.K.H., Ho, W., Choy, K.L.: Design and application of Internet of things-based warehouse management system for smart logistics. Int. J. Prod. Res. **56**(8), 2753–2768 (2018)
11. Corista, P., Giao, J., Sarraipa, J., Almeida, R., Perales, O.G., Nejib, M.: Enablers framework: an approach to develop applications using FIWARE. In: Enterprise Interoperability (2018)
12. FIWARE Foundation: from research to those digital services you will love. European Comission (2016) https://ec.europa.eu/digital-single-market/en/news/fiware-foundation-research-those-digital-services-you-will-love. Accessed 01 Feb 2019
13. The Future Internet platform FIWARE. European Comission (2018). https://ec.europa.eu/digital-single-market/en/future-internet-public-private-partnership. Accessed 01 Feb 2019
14. Corista, P., Ferreira, D., Giao, J., Sarraipa, J., Goncalves, R.J.: An IoT agriculture system using FIWARE. In: 2018 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC), pp. 1–6 (2018)
15. D2.1: Global Architecture Definition - version 1.2.2. Vf-OS, vf-OS consortium (2017)
16. Fitman project. Fiware for Industry. http://www.fiware4industry.com/?page_id=979. Accessed 10 Jan 2019
17. Pivotal: RabbitMQ. https://www.rabbitmq.com/. Accessed 10 Jan 2019
18. AMQP - Advanced Message Queuing Protocol. https://www.amqp.org/. Accessed 10 Jan 2019
19. Gandomi, A., Haider, M.: Beyond the hype: big data concepts, methods, and analytics. Int. J. Inf. Manage. **35**(2), 137–144 (2015)
20. NGSI-9/NGSI-10 information model. Fiware. https://forge.fiware.org/plugins/mediawiki/wiki/fiware/index.php/NGSI-9/NGSI-10_information_model. Accessed 10 Jan 2019
21. D1.3 Providers Scenarios Characterisation - version 1.0.1. Vf-OS, vf-OS consortium (2017)
22. D4.1.2a: WP4 Virtual Factory I/O Umbrella Deliverable - Vs: 1.0.8. Vf-OS, vf-OS consortium (2017)

Big Data on Machine to Machine Integration's Requirement Analysis Within Industry 4.0

Felipe A. Coda(✉), Rafael M. Salles(✉), Henrique A. Vitoi(✉), Marcosiris A. O. Pessoa(✉), Lucas A. Moscato, Diolino J. Santos Filho(✉), Fabrício Junqueira(✉), and Paulo E. Miyagi(✉)

Escola Politécnica da Universidade de São Paulo, Cidade Universitária, Av. Prof. Mello Moraes, 2231, São Paulo, SP, Brazil
{coda.felipe, rafaelmachadosalles, hvitoi, marcosiris, lamoscat, diolinos, fabri, pemiyagi}@usp.br

Abstract. One of the foundations for Industry 4.0 is the integration of various industrial elements (i.e. sensors, machines, and services) so that these devices can decide in a relatively autonomous way the level of integration which will be adopted. Thus, it is important to understand how the communication Machine to Machine is effectively realized and how these data can be explored and used to enhance the manufacturing process. The exchange of information between machines in the industrial process represents a potential to acquire and analyze a mass of data characterized as "big data", which can be perceived as an opportunity to discuss the paradigms of the industrial systems. Therefore, the purpose of this research is to identify the requirements for the Machine to Machine communication and the use of this data/information for more complexes analyzes using big data and analytics techniques. The KAOS methodology was utilized to model these requirements.

Keywords: Industry 4.0 · Requirements engineering · KAOS modeling · Big data · Machine to Machine · Interoperability

1 Introduction

Industry 4.0 (I4.0), or fourth industrial revolution, is an expression that was displayed in 2011 at the Hannover fair and started in April 2013 with the project "Plattform Industrie 4.0" [1]. Basically, Industry 4.0 uses technologies such as RFID (Radio Frequency IDentificator), to identify and track all of its elements or assets, such as machines, tools, and products. These identifiers [2]. Ploner [3] describes how the title "Industry 4.0" indicates an industrial revolution marked by the connection of three main technologies: Internet of Things (IoT), Cyber-Physical Systems (CPS), hings (IoT) and Information and Communication Technology (ICT).

Computational elements are used to control or monitor cyber-physical systems. In these systems, the physical and software elements are interconnected and also act closely integrated with the Internet and its users. Organization assets can be defined as a CPS with their physical and virtual functionalities, and the communication between

Published by Springer Nature Switzerland AG 2019
L. M. Camarinha-Matos et al. (Eds.): DoCEIS 2019, IFIP AICT 553, pp. 247–254, 2019.
https://doi.org/10.1007/978-3-030-17771-3_21

the functionalities of these assets is carried out by IoT and ICT. From these functionalities, Machine to Machine (M2M) communication also can be achieved through IoT, which must satisfy all the needs of I4.0, relating the efficiency and monitoring criteria of these machines. Through Big Data analytical techniques, it is possible to understand how the machines work and how their performance can be optimized.

The Machine to Machine communication is accomplished by integrating any machine to any machine that shares a common data exchange service. For this, sensors, actuators, and identifiers are coupled on these machines, thus representing communication through the IoT [4].

For the development of this communication and the data analysis, it arises the need to develop criteria for its implementation. The first stage of this process begins with the analysis of requirements, responsible for specifying objectives and possible obstacles. In order to obtain a cohesive requirement analysis, three points must be considered [4]:

- The viewpoint of various stakeholders and development engineers should be analyzed in order to have no subsequent problems with wrong or incomplete requirements [5];
- Establishment of methods for requirements management and validation;
- Creation of organized processes that lead to a specification with a natural language.

In this context, the focus of work is the modeling of requirements in both communications between machines and big data systems.

A hypothesis raised by the authors is that modeling the requirements described by initiatives associated with the I4.0 like the RAMI 4.0 (Reference Architectural Model for Industry 4.0) [6] using the KAOS (Keep All Objectives Satisfied) [7] methodology can assist in developing systems that effectively reflect the needs of the industry.

2 Relationship to Industrial and Service Systems

As described [4], requirements analysis remains the first step of any type of project. It has the responsibility of defining goals and predicting hindrances and therefore assisting on creating a resilient system.

It is necessary that the requirements for a system are accurately identified, enabling its development to be effectively oriented to fulfill the necessities of the involved stakeholders (the individuals committed with the benefits of the company's business) associated with the I4.0 context.

In recent years, efforts to establish models and standards for the structure of the Industry 4.0 were made. Some of them decompose the properties of a productive system in layers, from the physical assets, at the bottom, to the business model and regulatory conditions, at the top. According to [6], the most advanced of these structures is the German initiative "RAMI 4.0".

Within this context, previous works [4, 8] characterize the objectives, requirements, and challenges from the viewpoints of stakeholders associated with the Industry 4.0 context. Using the layers described in RAMI 4.0 as a base, [4] presents the modeling of the interaction between machines and [8] alludes to the data analysis functionalities and information capacities needed.

This work relates to industrial and service systems, providing a take the first step towards the development of systems capable of dealing with the heterogeneity of the entities in the Industry 4.0. Another valuable contribution of this research is to propose the discussion of the current efforts surrounding Industry 4.0 from the angle of the requirements engineering, orienting the productive system development towards the solving of the existing problems.

3 Literature Review

The foundation concepts utilized in the present work are displayed in this section. A comprehensive review of the concepts of Machine to Machine, Big Data, and the KAOS methodology can be found in previous works [4, 8].

3.1 RAMI 4.0

A reference architecture is a document that represents the structures and integrations recommended to form a solution. It incorporates industry-accepted practices and therefore answers the most common questions that arise during the development of solutions and technologies [2, 9].

The use of reference architectures is important for orienting the technological development towards the needs and requirements of the stakeholders of a production process, ensuring that the solutions developed are focused on solving existing problems.

The German Electrical and Electronic Manufacturers Association (ZVEI) established the RAMI 4.0 (Reference Architectural Model for Industry 4.0) as a standard design for I4.0 [6]. The RAMI 4.0 was conceived from current standards of the productive sector aiming to group different specifications on a single three-dimensional model (Fig. 1) to match the integration, both vertical, horizontal, and of distinct stages within the engineering process [2].

With these three dimensions, all significant features of I4.0 can be outlined, allowing the elements like the machines to be classified according to the model. The concepts of high flexibility proposed by I4.0 can be represented and implemented using RAMI 4.0, enabling a step-by-step departure from the current industry to the Industry 4.0 [10].

Each dimension can be summarized as:

- The dimension "**Layers**" is used to represent asset characteristics, making it possible for them to be fully mapped virtually.
- The dimension "**Life Cycle and Value Stream**" is used to visualize and normalize the relationships of Industry 4.0 components along their life cycle.

The dimension "**Hierarchical Levels**" describes the integration of business control systems. It was based on the ISA-95 standard, which defines the interface between the control functions and the other functions of the enterprise [2].

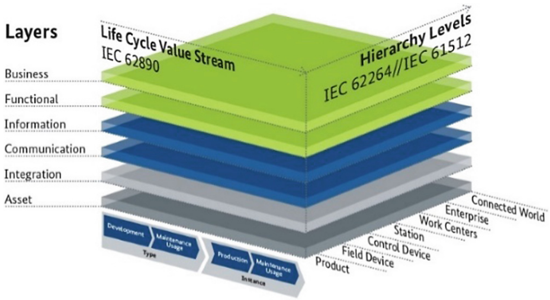


Fig. 1. RAMI 4.0 [6].

According to [2], the dimension *Layers* is divided into six levels to illustrate the breakdown of machines and physical entities, mapping those elements to their respective virtual representations, this virtual represented entity can also be referred as Industry 4.0 Component.

Using these levels, an overview to data mapping, functional descriptions, communication protocols, hardware and assets, and business processes can be represented. Each level and its interrelationships can be described as [6]:

- **Business:** Contains the rules that the system must follow as legal and regulatory conditions, mapping the business model of the system and orchestrating the services of the functional level;
- **Functional:** Contains a formal description of all the functionalities of a system and it is the stage responsible for their horizontal integration;
- **Information:** Contains the necessary data source for the control of assets, ensuring that the data of an asset is treated and made available. In this layer Big Data analysis is performed according to the specific application;
- **Communication:** Guarantees that all the data directed to the above layer have a uniform format, allowing the access of information;
- **Integration:** It is the virtual mapping of the real system. Here, the information on the physical assets is provided by the sensors as RFID (Radio-Frequency IDentification) and the Human Machine Interface (HMI);
- **Asset:** Represents physical components such as machines, actuators, parts, documents, and people.

This "Layer" dimension on RAMI 4.0 represents the main change on the way to view and understand a physical asset. Here the asset is seen as a CPS and the IoT is the tool that enables the connection among the physical asset and its digital counterpart on the Integration Layer.

3.2 KAOS Methodology

The KAOS methodology is a "goal-oriented requirements engineering approach" elaborated at the Universities of Oregon and Louvain. The benefit of KAOS over other methods is its ability to align goals to its requirements, improving the odds of the modeling adding value to the business [8, 11].

The KAOS modeling uses a graph with:

- **Goals:** Central Work Objective (white parallelograms);
- **SoftGoal:** Secondary objectives to achieve the central objective (gray parallelograms);
- **Requirements:** Establish the requirements for achieving the Goals (bold white parallelograms);
- **Operation:** Act that shows the procedures for obtaining the result (ellipses);
- **Actors:** Responsible for system action (hexagons);
- **Event:** Effect of the operation and demonstrated requirements (pentagons).

Figure 2 represents an example of how the requirements modeling are divided in this paper:

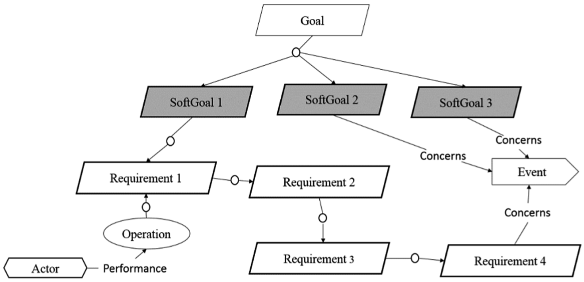


Fig. 2. Example of modeling.

4 Requirements Modeling

In this section, the objectives, operations, entities, and actors are demonstrated and discussed.

4.1 Goals

With the purpose of understanding the problem, it is first necessary to design the objectives. In this paper, the central goal is to implement Machine to Machine integration by monitoring the performance of products and machines with Big Data analysis. Thus, some points must be considered:

- The proper use of IoT for communication;
- Interoperability between the machines;
- Integrate M2M within CPS steps, using RAMI 4.0;
- In the Big Data analysis, it must be verified whether the decision-making speed for the manufacturing process is satisfactory; if the amount of data transmitted is enough; if there is a possibility of working with various types of data; if there is reliable information; and if the analysis can improve production [8].

4.2 Interaction Between Tiers of "Layers" Dimension

Machine to Machine integration and Big Data analysis happens on RAMI 4.0 "Layers" dimension. The operation begins when physical assets are initially represented as virtual assets. This step in RAMI 4.0 is called integration.

After the virtualization of the assets, the standardization of the data occurs in the Communication level, in order to guarantee that they are standardized before passing them to other levels.

The level of Information is next. At this level, machines can be evaluated to improve production performance. Big Data analytical methods can be employed to direct the production to the machine with the best performance for the required procedure.

The last level considered is the Functional one. It is at this level that commands are directed to the selected machine. At the end of the process, the machine receives the information and executes the required command. In this level, in fact, that occurs the interaction between the machines (M2M), supported by Big Data techniques. The requirements model is exemplified in Fig. 3.

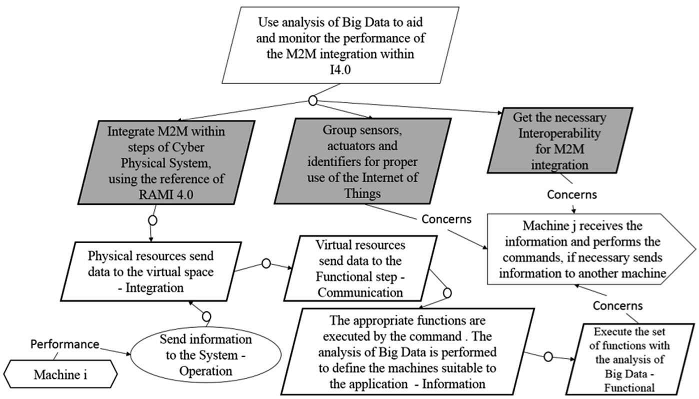


Fig. 3. Goals on the Machine to Machine integration and Big Data analysis.

4.3 Discussion About the Model

It is considered in the model, the steps of asset, integration, communication, information and functional. Machine to Machine communication starts at the asset, goes through the integration, collects information, and performs the function. Big Data occurs in the Information step, where the data are studied for the highest performance of the process.

It is noted in [12] that the communication between elements can occur not only within a single company but also throughout a "partner network". In the context of risk assessment, the model presented in this work can be improved by considering the concepts presented in [13] for risks involving "virtual enterprises" (i.e. a temporary entity created through a collaborative network of enterprises for the fulfillment of a goal).

Another important point to note is that there is no level of "Business" in the model. The legal and regulatory conditions and the orchestration of the level of functional services are represented in the level of "Business". The model represents each company with its operating rules guided by its level of "Business". In addition, each company has specifics functions, communication structures, databases, physical and virtual resources. Each company has its "partner network" who can communicate with each other, if necessary, and share a database or even Machine to Machine communication if there is interoperability.

4.4 Model Checking

For validation, it is necessary to know if all the requirements are correctly defined in the model. This problem can be solved with interactive questions directed to the stakeholders that aim to show if all the requirements have been fulfilled according to the viewpoint of each stakeholder. Some of these questions are verified in the papers [7] and [14]. The key issues addressed in [7] and [14] that are relevant to the present paper can be summarized in:

- Does the model need to further detail the system requirements?
- Would you like to check out the model's critical properties?
- Is it necessary to determine the scope of the system?
- Does the model need to consider non-functional system requirements that are difficult to quantify?

5 Final Remarks

The requirements analysis is an essential step required to understand the global and initial aspects of a project. The adoption of the methodology to model these requirements depends on the intricacy of the system to be included.

With the KAOS methodology, the requirements of the big data analysis on the M2M integration are accurately recognized, supporting its development to be adequately oriented to answer the needs of the interested stakeholders associated within the Industry 4.0 context. Validation of the requirements can be further performed with the statement of the aforementioned questions to be answered by the stakeholders and end users involved in a specific business model within the Industry 4.0 context. In this case, further details will be added to the model by incorporating characteristics sought by a specific application or system.

An interactive process must be carried out if any point not previously considered is identified on the model when answering the questions. A new solution is then proposed with a further corrected interpretation of requirements. The cycle must be carried out until the model is completely defined without issues.

Acknowledgement. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal no Nível Superior - Brazil (CAPES), Fundação de Amparo à Pesquisa do Estado de São Paulo - Brazil (FAPESP), and Conselho Nacional de Desenvolvimento Científico e Tecnológico - Brazil (CNPq).

References

1. Kagermann, H., Helbig, J., Hellinger, A., Wahlster, W.: Recommendations for implementing the strategic initiative INDUSTRIE 4.0: securing the future of German manufacturing industry; final report of the Industrie 4.0 Working Group (2013)
2. Pisching, M.A., Pessoa, M., Junqueira, F., Santos Filho, D., Miyagi, P.E.: An architecture based on RAMI 4.0 to discover equipment to process operations required by products. Comput. Ind. Eng. **125**, 574–591 (2018)
3. Ploner, L.: Industry Insiders Report: Industry 4.0. Renesas Edge (2014). http://br.renesas.com/edge_ol/global/12/index.jsp
4. Salles, R.M., Coda, F.A., Silva, J.R., Santos Filho, D.J., Miyagi, P.E., Junqueira, F.: Requirements analysis for machine to machine integration within Industry 4.0. In: IEEE International Conference on Industry Applications (INDUSCON) (2018)
5. Liu, X.F., Yen, J.: An analytic framework for specifying and analyzing imprecise requirements. In: Proceedings of the 18th International Conference on Software Engineering. IEEE Computer Society [S.1.], pp. 60–69 (1996)
6. Adolphs, P., et al.: Reference architecture model Industrie 4.0 (RAMI 4.0). ZVEI and VDI, Status report (2015)
7. Horkoff, J., Yu, E.: Analyzing goal models: different approaches and how to choose among them. In: Proceedings of the 2011 ACM Symposium on Applied Computing, pp. 675–682. ACM (2011)
8. Coda, F.A., Salles, R.M., Junqueira, F., Santos Filho, D., Silva, J.R., Miyagi, P.E.: Big data systems requirements for Industry 4.0. In: IEEE International Conference on Industry Applications (INDUSCON) (2018)
9. DIN; DKE; VDE: German Standardization Roadmap Industrie 4.0 - DIN/DKE Roadmap (2016). https://www.din.de/blob/65354/57218767bd6da1927b181b9f2a0d5b39/roadmap-i4-0-e-data.pdf
10. Hankel, M., Rexroth, B.: The reference architectural model Industrie 4.0 (RAMI 4.0). ZVEI, April 2015
11. Lapouchnian, A.: Goal-Oriented Requirements Engineering: An Overview of the Current Research, Department of Computer Science, University of Toronto, pp. 1–30 (2005)
12. Polyantchikov, I., Shevtsenko, E., Karaulova, T., Kangilaski, T., Camarinha-Matos, L.M.: Virtual enterprise formation in the context of a sustainable partner network. Ind. Manag. Data Syst. **117**(7), 1446–1468 (2017)
13. Mahmood, K., Shevtsenko, E., Karaulova, T., Otto, T.: Risk assessment approach for a virtual enterprise of small and medium-sized enterprises. Proc. Est. Acad. Sci. **67**, 17–27 (2018)
14. Lopes, N., Esteves, M.G.P., de Souza, J.M., Prado, P.: A checklist for peer knowledge validation in project-based organizations. In: 2015 IEEE 19th International Conference on Computer Supported Cooperative Work in Design (CSCWD), pp. 54–59. IEEE (2015)

Water Monitoring Systems

Artificial Neural Networks Application to Support Plant Operation in the Wastewater Industry

Ivan Pisa[1,2](✉), Ramon Vilanova[1], Ignacio Santín[1], Jose Lopez Vicario[2], and Antoni Morell[2]

[1] Advanced Systems for Automation and Control (ASAC) Group, Escola d'Enginyeria, Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain
{Ivan.Pisa,Ramon.Vilanova,Ignacio.Santin}@uab.cat
[2] Wireless Information Networking (WIN) Group, Escola d'Enginyeria, Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain
{Jose.Vicario,Antoni.Morell}@uab.cat

Abstract. This communication presents the main aim, contextual and development framework of the PhD that is being conducted by the first author. In this PhD, main aim is the application of data driven methods to industrial processes in order to improve and support industrial operations. In this case, Wastewater Treatment Plants (WWTPs) are adopted as the industry where data driven methods will be applied. WWTPs are industries devoted to managing and process residual water coming from urban and industrial areas. Those type of industries apply highly-complex and nonlinear processes to reduce the contamination of water. Therefore, among the different data driven methods, in this PhD we will focus on the application of Artificial Neural Networks (ANNs) in order to improve and support the operations performed in this type of industries. ANNs are considered due to their ability in the modeling of highly-complex and nonlinear processes such as the WWTPs processes.

Keywords: Wastewater industries · Data driven methods · Artificial Neural Networks

1 Introduction

Nowadays, the incursion of Industry 4.0 has arisen as a paradigm becoming true, industrial systems are becoming able to communicate and cooperate with other systems and also with humans by means of Wireless Sensor Networks (WSN) in order to take decentralized decisions to manage the industrial plants [1]. Consequently, large amount of measurements and data from industrial processes will be stored and processed [2]. Thus, the aim of this research is to support and improve the performance of certain industrial processes by means of data driven methods and especially Artificial Neural Networks (ANNs). In addition, these methods will be applied over data coming from Wastewater Treatment Plants (WWTPs) which are industrial plants devoted to reducing the pollutant levels of residual waters.

Published by Springer Nature Switzerland AG 2019
L. M. Camarinha-Matos et al. (Eds.): DoCEIS 2019, IFIP AICT 553, pp. 257–265, 2019.
https://doi.org/10.1007/978-3-030-17771-3_22

As introduced, investigated prediction schemes will be based on ANNs, whose most popular tasks when dealing with industrial data correspond to predictive tasks such as predictive maintenance or anomaly detection [3]. Thus, predictions can support industrial processes operations by providing them with information about future industrial events or measurements. In this work, predictions will support the WWTP plant operations by feeding control strategies providing information to them to anticipate actuations and improve the overall WWTP performance.

2 Relationship to Industrial and Service Systems

As it has been stated, Industry 4.0 is becoming a reality. Industries have been provided with systems able to communicate and cooperate with others. Thus, the connectivity and communication among industrial systems will be performed by means of WSNs which will generate thousands of bytes of information per day. Consequently, data can be processed in order to extract and generate value from it.

In terms of WWTPs industries, the ones in which this work is based, WSN are adopted in order to measure incoming water ratios, pollutant concentrations, oxygen demand and other parameters which characterize the WWTPs' behavior [4]. Therefore, large amount of concentration measurements is stored during a day and consequently, the application of data driven methods to these measurements is widely encouraged in order to extract certain characteristics of the plant. For instance, in [5] the WWTP's influent ratio profile is processed to determine the expected weather and therefore prepare the WWTP to deal with all the incoming flow rate.

In that sense, ANNs are widely adopted as service systems which feed industrial systems or simply obtain certain measurements which gives information about the industry where they are deployed. In addition, their adoption is motivated by the fact that they only require input and output information of the industrial processes to generate a mathematical model able to yield information of the industrial process. This information will be used to complement other systems and consequently, generate value from the stored data. Thus, ANNs can be seen as soft-sensors adopting data-driven methods which work as service systems. Their purpose is to model high-complex and nonlinear systems in order to yield either prediction or classification of process outputs which will fulfill other industrial processes or drive monitoring and decision modules. In that manner, the overall complexity of the industrial processes can be simplified just requiring input and output data of these processes.

3 State-of-the Art

The increasing interest in the Internet of Things (IoT) and Industry 4.0 [6] has motivated the adoption of ANNs for forecasting purposes in industrial systems. ANNs have been considered in different works where predictions are adopted to complement different industrial processes and systems. For instance, the adoption of neural networks as service systems dealing with sewer systems' data is proposed in [7, 8].

In WWTPs context, ANNs have been considered to perform different tasks. For instance, ANNs have been considered as soft-sensors predicting hard-to-measure WWTPs' values. Predictions of measurements such as Chemical Oxygen on Demand (COD) and aeration tank's dissolved oxygen concentration (SO) are performed in [9] showing a Percentage Error (MAPE) in the prediction around 4.48%.

Other approaches are based on the adoption of ANN in the reduction of pollution effluent concentrations. This process is performed by means of highly-complex and nonlinear biochemical and biological processes which are defined by the Activated Sludge Model No.1 (ASM1) [10]. Those are devoted to reducing the pollutants derived from nitrogen and phosphorus in order to reduce the pollution of the environment where treated water flows. Besides, local administrations have established limits to pollutants such as ammonium ($S_{NH,e}$) and total nitrogen ($S_{Ntot,e}$) present at the WWTP effluent. Violations of them implies economical punishments to the WWTP [11].

In order to avoid violation, predictions can be adopted by different control strategies to let them actuate before violations are really produced. For instance, in [12] a hierarchical control strategy based on Model Predictive Control (MPC) and Fuzzy Logic control are proposed to reduce the concentrations of ammonium ($S_{NH,e}$) and total nitrogen ($S_{Ntot,e}$) concentrations. Both strategies are fed with effluent concentration predictions performed by Multilayer Perceptron (MLP) ANN. They predict the maximum effluent concentration that will be observed during a day. Moreover, MLPs do not consider the time correlation which characterize industrial signals. In [13], the adoption of ANNs to forecast WWTPs effluent concentrations and feed control strategies is presented. However, the considered ANN does not take into account the time correlation between influent and effluent measurements. Instead, it adopts the previously observed effluent concentration as an input.

4 Research Problem

Aforementioned, ANNs predictions feeding some control strategies have been applied in previous works. However, there exist pollutant violations as shown in [12], where the hierarchical control strategy is not able to totally remove the violations of total nitrogen. They are produced during a total of 8.72 days during a year. Some of these violations are produced due to the fact that the control strategy does not have enough time to process the pollutant.

In order to solve this, data driven methods and specially ANNs considering the time correlation between influent and effluent data are proposed in the realization of this PhD. In such a context, those ANNs are adopted at WWTPs industries to complement the control strategies and, therefore improve the performance of the overall WWTP operability exploiting the time-correlation of data. In the case of this work, we will focus on the nitrogen removal process. Thus, $S_{NH,e}$ and $S_{Ntot,e}$ concentrations in the effluent are the process outputs to be forecasted by the considered ANNs. A violation is prone to occur whenever the prediction exceeds the limits shown in Table 1.

Table 1. Effluent limits. Exceeding these limits is punished with high economical fines.

Concentration	Value
Ammonium ($S_{NH,e}$)	4 mg/L
Total Nitrogen ($S_{Ntot,e}$)	18 mg/L

5 Materials and Methods

In accordance to what has been seen in the previous section, the purpose of this work is to adopt ANNs to perform predictions of the WWTP effluent pollutant levels to complement the control strategies. Thus, the required materials and methods are described in this section: (i) Benchmark Simulation No.2 will be considered in the generation of input and output data, whereas (ii) ANNs and especially Long-Short Term Memory cells will be adopted to perform predictions.

5.1 Benchmark Simulation Model No.2

BSM No.2 corresponds to a simulation scenario which models the default structure of a WWTP [14]. This default structure can be split into two differentiated parts: (i) the water line and (ii) the sludge treatment. The water line corresponds to five biological reactor tanks (Activated Sludge Reactors) where pollutant concentrations reduction process is performed (See Fig. 1).

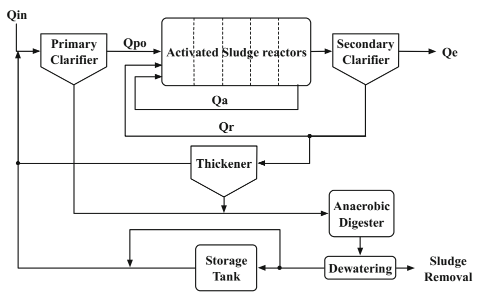


Fig. 1. BSM No.2 model where the different WWTP elements are observable.

First two tanks are anoxic (working without oxygen) whereas the remaining three correspond to aerated ones. An internal recycle flow transporting sludge from the last aerobic tank to the first anoxic one is considered. In addition, two clarifiers where the sludge sedimentation process is performed are considered. Concerning the Sludge treatment part, it corresponds to those elements devoted to treat the WWTPs' sludge. This part encompasses elements such as the Thickener, the Storage tank, the Dewatering and the Anaerobic Digester elements. Different flows are considered in BSM No.2 scenario: Qin corresponds to the influent, Qpo is the primary clarifier overflow, Qa is the internal recycle flow rate, Qr is the sludge internal recycle flow rate and finally Qe is the effluent flow rate.

5.2 Long-Short Term Memory Cells

ANNs consist in a set of layers where each layer considers a certain number of hidden neurons. These neurons are characterized by an activation function which relates the input of the neuron with the output. Each layer is connected to its neighbors by means of weights and biases which are obtained in the ANNs' training process.

Moreover, among the different ANNs, we will consider the adoption of Recurrent Neural Networks (RNN) and specially the Long-Short Term Memories (LSTM) due to their performance dealing with time series and highly-time correlated signals [15]. Thus, if influent and effluent profiles are considered as highly-time correlated parameters, the application of LSTM is widely suitable. LSTMs, as other gated RNNs, are designed to alleviate the well-known RNN's problem: the vanishing and exploding gradients [16]. They replace the RNN's hidden neurons by memory cells with four gates: the forget, the input the state update and the output gate. Forget gate is in charge of resetting the LSTM's memory cell which will be filled with the information coming from the input gate (2) and the state update (3).

$$\mathbf{f_t} = \sigma\left(\mathbf{W}_f^T\mathbf{x_t} + \mathbf{U}_f^T\mathbf{h_{t-1}} + \mathbf{b_f}\right). \tag{1}$$

$$\mathbf{i_t} = \sigma\left(\mathbf{W}_i^T\mathbf{x_t} + \mathbf{U}_i^T\mathbf{h_{t-1}} + \mathbf{b_i}\right). \tag{2}$$

$$\mathbf{c'_t} = \tanh\left(\mathbf{W}_c^T\mathbf{x_t} + \mathbf{U}_c^T\mathbf{h_{t-1}} + \mathbf{b_c}\right). \tag{3}$$

Output gate determines if the output of the memory cell will influence on another LSTM cell or not [15, Chapter 10].

$$\mathbf{o_t} = \sigma\left(\mathbf{W}_o^T\mathbf{x_t} + \mathbf{U}_o^T\mathbf{h_{t-1}} + \mathbf{b_o}\right). \tag{4}$$

In all cases **W, U** correspond to the weights of each gate, **b** to their biases and **x** and **h** to the input and cell state values. LSTM structure is observable in Fig. 2.

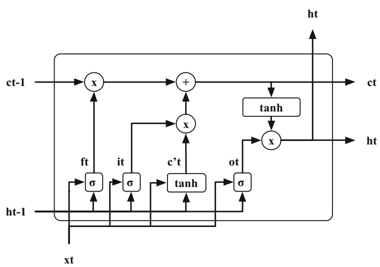


Fig. 2. LSTM structure. f_t, i_t and o_t correspond to the forget, input and output gate. σ and tanh corresponds to the activation function adopted at each gate. Each gate is formed by different ANNs with the same number of hidden neurons. c_t corresponds to the hidden state of the memory cell. h_t is the cell's output.

Its training process is performed by means of the backpropagation through time algorithm, which is based on the iterative update of the network weights towards the opposite of the gradient of the loss function w.r.t the weights. Training process is performed considering the well-known split of data (70% for training the ANN and 30% for testing purposes).

6 Expected Outcomes

6.1 Prediction Approach

This PhD is based on the application of ANNs to improve and support WWTPs. Therefore, the reduction of violations and the overall operational cost is sought. A first approach adopting WWTP's effluent predictions corresponds to an ANN-based system, which is able to predict the WWTP's effluent concentrations. It will generate an alarm whenever a violation is predicted. Further details on the ANN-based system can be found in [17]. The steps followed by the system are:

1. Gather available measurements of the WWTP plant (in real time)
2. Normalize the data to reach ANN converge towards an optimal solution
3. The system will order measurements considering the time correlation
4. Effluent concentrations will be predicted by means of LSTM-based ANNs
5. Predictions will be contrasted with the effluent limits to determine is violations will be performed or not.
6. The effluent limits can be modified by WWTP's operator to set the accuracy of alarm detection process by means of ROC Curves.

6.2 Prediction Results

Predictions of the proposed ANN-based system [17] are performed by means of two LSTM-based prediction structures: (i) Ammonium prediction structure which considers two stacked LSTM cells with 50 hidden neurons per gate; (ii) and the Total nitrogen prediction structure which considers two stacked LSTM cells with 10 hidden neurons per gate. Table 2 shows the performance of both prediction structures where Mean Absolute Percentage Error (MAPE) and the false positive (P_{fa}) and miss-detection (P_{miss}) probabilities are computed [17]. Predictions are observable in Fig. 3.

Table 2. Prediction structures' performance.

Prediction structure	MAPE	P_{fa}	P_{miss}
Ammonium	7.93%	0.03	29.85
Total Nitrogen	4.69%	0.44	31.75

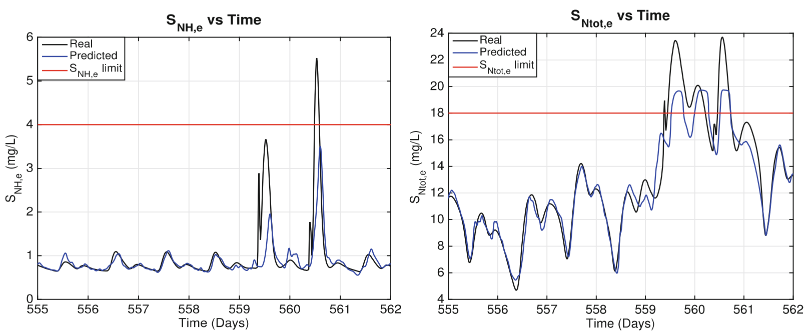


Fig. 3. Performed predictions. First figure corresponds to Ammonium predictions whereas second figure corresponds to the Total Nitrogen predictions.

Although performance shows P_{fa} and P_{miss} around 29.85% and 31.75%, they can be enhanced lowering the limits at which effluent predictions are considered a violation. Receiver Operating Characteristic (ROC) curves, where P_{fa} and P_{miss} are faced, showing an Area under the Curve (AuC) around 0.99 (1 means perfect performance).

Results show that LSTM cells predict accurate effluent concentrations in real time considering the time correlation between values. Thus, predictions performed in previously commented work [12] are improved with the deployment of the proposed system by considering the time correlation between input and output data.

6.3 Future Work

The future research will be oriented along the following main streams:

- On one side, predictions improvement by adopting K-Fold as the training algorithm in order to deal with the data unbalanced problem (the largest part of values is below the effluent limits). Moreover, predictions will be adopted to feed control strategies and consequently reduce the number of violations.
- Direct use of predicted output data for computing of predictive control actions. In this case, the use of data driven predictive control algorithms is sought
- The third step will be to explore the possibility to directly generate control action on the basis of ANN. In this case, ANN could be trained to do the work of the predictive controller or, instead, explore the use of the actor-critic scenario where all the players are implemented as ANN.

7 Conclusions

The adoption of ANNs and especially LSTM cells as service systems applied at industrial environments has been studied in this work. Here, two LSTM-based structures to predict effluent concentrations at WWTPs have been considered. They show a good performance yielding low MAPE values around 4.69% for $S_{Ntot,e}$ and 7.93% for $S_{NH,e}$ concentrations. Moreover, P_{fa} and P_{miss} probabilities and ROC curves have been also computed showing an AuC around 0.99 (1 is perfect performance).

Acknowledgments. This work has received the support from the Spanish Ministry of Economy and Competitiveness program under MINECO/FEDER grant DPI2016-77271-R and also from La Secretaria d'Universitats i Recerca del Departament d'Empresa i Coneixement de la Generalitat de Catalunya i del Fons Social Europeu under FI grant. Authors belong to the recognized research groups SGR 1202 and SGR 1670 by the Catalan Government.

References

1. Sarvari, P.A., Ustundag, A., Cevikcan, E., Kaya, I., Cebi, S.: Technology roadmap for Industry 4.0. In: Industry 4.0: Managing The Digital Transformation. Springer Series in Advanced Manufacturing, pp. 95–103. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-57870-5_5
2. Rashid, B., Rehmani, M.H.: Applications of wireless sensor networks for urban areas: a survey. J. Netw. Com. Appl. **60**, 192–219 (2016)
3. Martí, L., Sanchez-Pi, N., Molina, J.M., Garcia, A.C.B.: Anomaly detection base on sensor data in petroleum industry applications. Sensors **15**, 2774–2797 (2015)
4. Fernandez de Canete, J., Del Saz-Orozco, P., Baratti, R., Mulas, M., Ruano, A., Garcia-Cerezo, A.: Soft-sensing estimation of plant effluent concentrations in a biological wastewater treatment plant using an optimal neural network. Exp. Syst. Appl. **60**, 8–19 (2016)
5. Qiao, J.F., Hou, Y., Zhang, L., Han, H.G.: Adaptive fuzzy neural network control of wastewater treatment process with multiobjective operation. Neurocomputing **275**, 383–393 (2018)
6. Wollschlaeger, M., Sauter, T., Jasperneite, J.: The future of industrial communication: automation networks in the era of the Internet of Things and Industry 4.0. IEEE Ind. Electron. Mag. **11**, 17–27 (2017)
7. Zhang, D., Martinez, N., Lindholm, G., Ratnaweera, H.: Manage sewer in-line storage control using hydraulic model and recurrent neural network. Wat. Resour. Mang. **32**, 2079–2098 (2018)
8. Zhang, D., Hølland, E.S., Lindholm, G., Ratnaweera, H.: Hydraulic modeling and deep learning based flow forecasting for optimizing inter catchment wastewater transfer. J. Hydrol. **567**, 792–802 (2017)
9. Güçlü, D., Dursun, Ş.: Artificial neural network modelling of a large-scale wastewater treatment plant operation. Biopro. Biosyst. Eng. **33**, 1051–1058 (2010)
10. Henze, M.: Activated sludge model No.1. IAWPRC Sci. Tech. Reports 1 (1987)
11. Vilanova, R., Santín, I., Pedret, C.: Control y Operación de Estaciones Depuradoras de Aguas Residuales: Modelado y Simulación. RIAI – Rev. Iberoam. Autom. e Inform. Ind. **14**, 217–233 (2017)
12. Santín, I., Pedret, C., Vilanova, R., Meneses, M.: Advanced decision control system for effluent violations removal in wastewater treatment plants. Cont. Eng. Prac. **49**, 60–75 (2015)
13. Foscoliano, C., Del Vigo, S., Mulas, M., Tronci, S.: Predictive control of an activated sludge process for long term operation. Chem. Eng. J. **304**, 1031–1044 (2016)
14. Jeppsson, U., et al.: Benchmark simulation model no 2: general protocol and exploratory case studies. Water Sci. Technol. **56**, 67–78 (2007)

15. Goodfellow, I., Bengio, Y., Courville, A.: Deep Learning. MIT Press, Cambridge (2016)
16. Bengio, Y., Simard, P., Frasconi, P.: Learning long-term dependencies with gradient descent is difficult. IEEE Trans. Neural Networks **5**, 157–166 (1994)
17. Pisa, I., Santín, I., Vicario, J.L., Morell, A., Vilanova, R.: A recurrent neural network for wastewater treatment plant effluents' prediction. In: XXXIX Jornadas de Automática, pp. 621–628 (2018)

Towards a Practical and Cost-Effective Water Monitoring System

João Marques[1], Brígida Lopes[1], Carlos Ferreira[1,2], Henrique Pinho[1,2], Manuel Barros[1,2], Pedro Granchinho[1,2(✉)], and Pedro Neves[1]

[1] Department of Engineering, E.S.T.T., Polytechnic Institute of Tomar, Quinta do Contador - Estrada da Serra, 2300-313 Tomar, Portugal
{jomarques, bablopes, cferreira, hpinho, fmbarros, granchinho, pedroneves}@ipt.pt

[2] Smart Cities Research Center (Ci2), Polytechnic Institute of Tomar, Quinta do Contador - Estrada da Serra, 2300-313 Tomar, Portugal

Abstract. In recent years, there has been increasing awareness of the preservation, protection and sustainable use of natural resources. Water resources, being one of the most important, face major threats due to contamination by pollutants of various types and origins. Maintaining the quality of water resources requires more robust, reliable and more frequent monitoring than traditional data collection techniques based on manual sampling methods. This article, which is the result of ongoing research, proposes a practical and cost-effective solution for a surface water monitoring system, using a robotics platform and cloud services. The proposed solution allows for scalability and will accommodate a wide range of end-user specifications. To allow for continuous operation in longer activities, the design of a versatile real-time water quality monitoring system should also take into consideration the question of its energy requirements and self-sufficiency.

Keywords: Water quality monitoring · Wireless sensor networks · Autonomous vehicles · Environment protection · Resources management · IoT · Environmental robotics · Cloud computing

1 Introduction

Society needs fast and reliable environmental monitoring platforms. Conventional processes involve the manual collection at various points from a sampling network, followed by laboratory testing and analysis. These methods can be seen as ineffective, as they are cumbersome, time-consuming, and lack real-time results to promote a proactive response to water contamination sources [1].

This work provides an outline for developing a system that solves the need for monitoring the surface water quality in natural and artificial freshwater systems of different types, whilst mitigating the normal entry-level barriers for the academia, local authorities, businesses and general population. In future works, a requirements specification plan will be introduced, which will be necessary for developing a working prototype.

Published by Springer Nature Switzerland AG 2019
L. M. Camarinha-Matos et al. (Eds.): DoCEIS 2019, IFIP AICT 553, pp. 266–272, 2019.
https://doi.org/10.1007/978-3-030-17771-3_23

Two key areas of environmental monitoring are particularly important in the context of this work, especially those concerning surface water:

Sensor Networks. Environmental sensor networks (ESN) have recently emerged as a promising technology for monitoring aquatic environments [2]. ESNs consist of a small group of small networked devices that allow real-time monitoring of relevant variables at various fixed and predetermined points [3]. ESNs have the advantage of allowing data collection in multiple locations simultaneously [4] and with greater fidelity. A major limitation of ESNs is that they usually need to have fixed sampling locations [3] and therefore do not have the ability to self-reconfigure in response to unexpected events or to increase coverage in the area of interest.

Environmental Robotics. Standalone robots for environmental monitoring are a promising solution to overcome the limitations of ESNs, adding mobility and thus increasing their potential. Robot groups can collect data from several places simultaneously, allowing to dynamically increase the spatial-temporal area of data collection that would be impossible to achieve with a single robot or a network of static sensors [3]. Some examples of this technology include the ROAZ II system [5] developed at the Polytechnic Institute of Porto (ISEP) an and the THA1592 project implemented in the Polytechnic Institute of Tomar [6]. Unmanned Aerial Vehicles (USVs) are often unable to achieve continuous operation because of their limited movement and energy autonomy.

2 Relationship to Industrial and Service Systems

To meet the growing global need for fresh water sources, cloud computing can help manage data and information related to water supply and quality. Unfortunately, industry is not leveraging this new technology as fast and as widely as possible. The democratization of solutions, are not always in the interest of turnkey solutions providers (e.g. Consultants).

The research team also believes that in order to achieve a truly resilient system, a more agile mindset is required to face future and unexpected challenges. Water quality management is also more complex than, and potentially as important to the future of mankind as air quality management.

In the coming years, as sources of fresh water become scarcer worldwide, and more entities compete ever more vigorously for this resource, different businesses will be challenged in different ways. What is certain is that all businesses will be affected either directly or indirectly by the coming changes. Scarcity usually encourages better management of resources or forces extinction due to inadequate or inappropriate responses. Cloud Computing-based environmental solutions represent the best way for companies to take ownership of their data, satisfy looming reporting requirements, and tackle their water quality management challenges.

3 Rapid Prototyping and Deployment for Environmental Robotics

In this section, the components for a robotic water monitoring system archetype are described. In order to achieve fast development speeds, a simple but robust robotics platform was selected. Sensor integration and communication also needs to be addressed. For these reasons, cloud services were chosen to achieve fast deployment speeds and reliability.

Figure 1 presents, schematically, the three main functional blocks of the proposed system: the robotics and IoT hardware and the cloud-based data storage and retrieval. For simplicity, the two systems will be functioning separately in a complementary fashion. The robotics module will allow the vehicle to execute user-defined missions, while the IoT sub-system gathers and processes the data from the attached sensors.

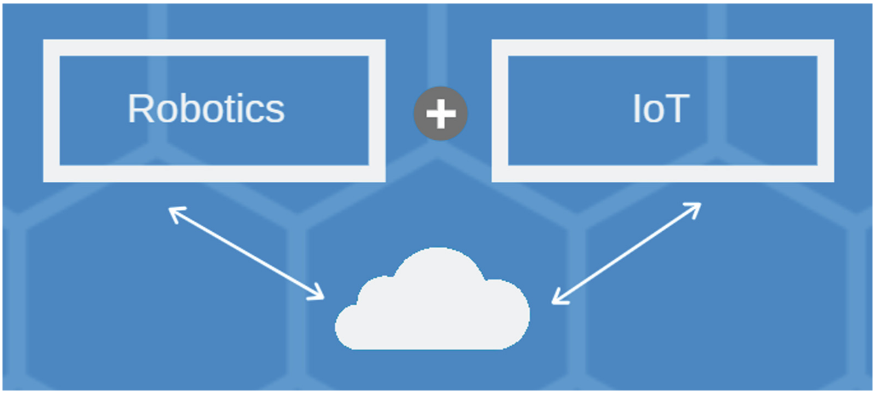


Fig. 1. Concept - Integration of robotics, IoT and cloud services technologies.

3.1 Robotic Platform Overview

The system consists of an autonomous, mobile and robotic monitoring system unit (called LIBÉLULA) with onboard sensors for the acquisition of water quality parameters (pH, temperature, conductivity, etc.), a long-range wireless radio communication module and a microcontroller board with a GPS and compass unit. The control of the robotic platform will be based on a modular software architecture provided by an advanced, full-featured and reliable open source autopilot software, called Ardupilot. Ardupilot is open-source and has been developed over more than five years by a team of diverse professional engineers and computer scientists, being affirmed by the community as a well-tested and proven autopilot software. It provides services for navigation, localization, simulation, and integration with the well know robotic framework, ROS; which, in turn, allows remote monitoring of relevant water quality data and their dissemination in real time to a central server and cloud platforms [7].

3.2 IoT Platform Overview

Cloud computing, on-demand computing or Internet based computing is a low-cost technology that offers fast and reliable processing, data storage, and several developer and analytical tools that allow both easy integration of physical devices and advanced data analysis. In this work, cloud computing services are to be used as the infrastructure backbone for the system as they provide fast deployment, continuous service, security and scalability. Although any provider of cloud services can be used, Fig. 2 shows a diagram consisting of an interconnected suite of applications made available from IBM, specifically for IoT.

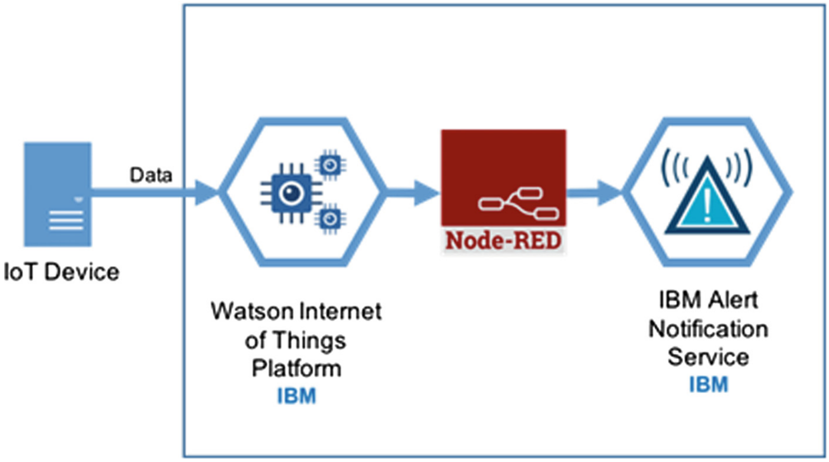


Fig. 2. Data flow for IoT system in the cloud (source: IBM)

4 Water Quality Analysis and Forecasting

In order to adequately characterize the aquatic environment, the monitoring system needs to acquire a specific set of environmental data. Furthermore, to identify potential problems in water quality, some computational methods must be introduced to detect subtle deviations in the data. The most used of the general physical-chemical parameters are: water temperature, pH, electrical conductivity, dissolved oxygen and total suspended solids. Following the theme of this work, whose main aim is to help the authorities in the "decision-making" process, different methods have been identified in the literature. Some of them are more in the field of traditional statistical techniques, such as principal component analysis (PCA), regression analysis, and cluster analysis [8] but in the age of never-ending streams of data there's also a growing tendency to use artificial intelligence algorithms and machine learning. A recent study has been made employing fuzzy logic, machine learning and adaptive filtering to successfully predict measurements a day ahead, as well as techniques to incorporate the window of past values in order to be able to make a more precise prediction [9, 10].

5 Energy Management Requirements

Energy in mobile robots scenarios is an important issue. Fast charging methods and power converters should be used to power supply the system. Due to exposition to environmental conditions inductive contactless power transfer is a possible solution. It is a low maintenance solution and is safer by providing arc-free energy transfer.

Since the operation will normally be in remote locations, in terms of power, renewable energy sources should be considered in order to achieve autonomy for the system. Typically, probable power sources are water stream, converted to electricity through a hydrogenerator or, more generally, photovoltaic panels that generate power from sunlight. The harvested energy should be stored in a battery for permanent availability. Locating the charger station and precise docking are problems to be addressed.

6 Building a Case-Study

To validate the proposed platforms described in Chapter 4, field tests will be realized in the Castelo do Bode reservoir, one of Portugal's main water reserves. Figure 3 shows a sample mission, created using software for ground station control [11], which shall be executed by a robotic prototype developed by the research team from IPT. This is a very simple test and is intended as a proof of concept.

Fig. 3. Mission Planner software (source: Ardupilot)

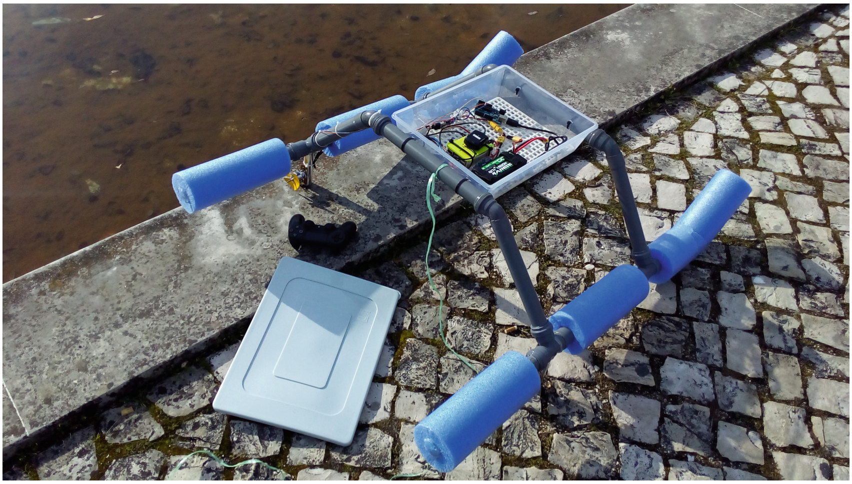


Fig. 4. Working prototype for a low-cost unmanned surface vehicle, developed by the research team from IPT.

In Fig. 4, a first prototype is shown for a low-cost, light-weight and durable catamaran type. In the future, work will be directed towards establishing a higher benchmark by developing a more robust working prototype, with the use of better materials as well as CAD and professional machining processes. During field testing, water parameters data is to be collected, stored and visualized using cloud services. Adding obstacle avoidance or the use of semantic segmentation for detection of navigable space, are desirable features that provide an important layer of intelligence and added mission autonomy. The proposed system has many potential applications and use-cases like surface water monitoring, wastewater management, research and development departments, small companies and individuals, local authorities and administrations, academic research, among others.

7 Conclusion

The present work defines the technical requirements of a capable system of surface water quality monitoring for natural and artificial freshwater systems. A novel alternative is proposed, that builds on commercial and community (open-source) ventures like Ardupilot (auto-pilot software) and cloud services, made available by some of the key players in the information technology sector (e.g. Microsoft, Amazon, Google, IBM, Oracle, etc.). The acquired data by such a system will allow the identification of pollution sources or the discharge and dispersion of contaminants along watercourses. State-of-the-art robotics and artificial intelligence technology are very powerful but remain difficult to implement by the less technically adept. Robotic monitoring systems are still a niche industry, with very high price tags/rates making them inaccessible to several interested parties who would undoubtedly benefit from said technologies This project aims to be a versatile, robust and reliable real-time water quality monitoring system, thus contributing to the management and protection of water resources, increasing security in the water supply of populations for a more efficient and sustainable management of resources.

Acknowledgements. This work was supported by Centro 2020, Portugal 2020 and European Union (EU) under the grants, CENTRO-01-0145-FEDER-024052E – Libélula: Mobile robotic surface water quality monitoring system.

References

1. Pule, M., Yahya, A., Chuma, J.: Wireless sensor networks: a survey on monitoring water quality. J. Appl. Res. Technol. **15**, 562–570 (2017)
2. Ballesteros-Gomez, A., Rubio, S.: Recent advances in environmental analysis. Anal. Chem. **83**(12), 4579–4613 (2011)
3. Dunbabin, M., Marques, L.: Robots for environmental monitoring: significant advancements and applications. IEEE Robot. Autom. Mag. **19**, 24–39 (2012)
4. Xu, G., Shen, W., Wang, X.: Applications of wireless sensor networks in marine environment monitoring: a survey. Sensors **14**, 16932–16954 (2014)
5. Ferreira, H., Almeida, C., Martins, A., Almeida, J., Dias, N., Dias, A.: Autonomous bathymetry for risk assessment with ROAZ robotic surface vehicle. In: Proceedings of the OCEANS 2009 IEEE conference, pp. 1–6. IEEE Press, New York (2009)
6. Raimundo, D., Leite, T., Silva, A.: Projeto de um Veículo Autónomo de Superfície para Monitorização de Águas Fluviais. EECS Project, IPT, Tomar (2016)
7. ArduPilot Open Source Autopilot. http://ardupilot.org
8. Bassiliades, N., Antoniades, I., Hatzikos, E., Vlahavas, I., Koutitas, G.: An intelligent system for monitoring and predicting water quality. In: Proceedings of the European Conference Towards Environment, pp. 534–542, Prague (2009)
9. Bhat, S., Meraj, G., Yaseen, S., Pandit, A.K.: Statistical assessment of water quality parameters for pollution source identification in Sukhnag stream: an inflow stream of lake Wular (Ramsar Site), Kashmir Himalaya. J. Ecosyst. (2014). https://doi.org/10.1155/2014/898054
10. Thiyagarajan, K., Pappu, S., Vudatha, P., Niharika, A.V.: Intelligent IoT based water quality monitoring system. Int. J. Appl. Eng. Res. **12**, 5447 (2017)
11. Mission Planner Documentation. http://ardupilot.org/planner/index.html

Smart Cities: Non Destructive Approach for Water Leakage Detection

Lucas Nunes Monteiro[1,2], Felipe Crispim da Rocha Salvagnini[1,2(✉)], Edinei Peres Legaspe[2], Sidney José Montebeller[1], Andréa Lucia Braga Vieira Rodrigues[1], Fernando Garcia Deluno[1], and Diolino J. Santos Filho[2]

[1] Centro Universitário da Faculdade de Engenharia de Sorocaba, Sorocaba, Brazil
felipe.salvagnini@facens.br
[2] Escola Politécnica da Universidade de São Paulo, São Paulo, Brazil
diolinos@usp.br

Abstract. Natural resources management is essential, especially of water distribution within cities. In Brazil, water losses in distribution systems go up to around 38%. In the context of “Smart Cities”, technologies that use “The Internet of Things” can be applied to reduce such losses. The present article shows that leakages produce distinctive noise ranging from 100 Hz to 1000 Hz. Through digital signal processing techniques, such as the Discrete Fourier Transform and Goertzel Transform, the spectral signals are decomposed, revealing their components of frequency such as the intensity. An architecture that performs the communication between slave nodes through a TCP/IP network is then proposed. The slave nodes are responsible for data collection for leakage identification. The collected data is then sent to the data master where there is greater computing power. The data master will perform the processing according to the paradigm of edge computing, thus obtaining frequency responses and the identification of the leakage itself. It will also make data available through OPC-UA, a standard “Internet of Things” communication protocol widely used in the industrial context.

Keywords: Leakage detection · OPC-UA · Water management · Smart Cities · Water loss control

1 Introduction

Water scarcity is a major problem concerning development requirements, leading to major clashes and potential unrest around the world. The situation on developing countries is more critical as they still need to review their policies related to the rational and sustainable use of water [1]. In 2015, the World Resources Institute (WRI) conducted an analysis based on global climate change data, establishing a water stress rank for 167 countries, projecting it to 2020, 2030 and 2040. According to their study, 33 countries (14 of which are located in the Middle East) will present some worrying water stress by 2040 [2].

Published by Springer Nature Switzerland AG 2019
L. M. Camarinha-Matos et al. (Eds.): DoCEIS 2019, IFIP AICT 553, pp. 273–280, 2019.
https://doi.org/10.1007/978-3-030-17771-3_24

Currently, Brazil is a privileged country, where water is an abundant resource. Brazil holds more than half of the available water in South America, summing up 13% of the world's supplies. However, Brazil faces problems due to faulty distribution, degradation caused by pollution, mismanagement and poor distribution networks maintenance, both preventive and corrective [3, 4]. Figure 1 shows the amount of water consumption per capita and percentage of water loss in the distribution network from 2008 to 2013 [5, 6].

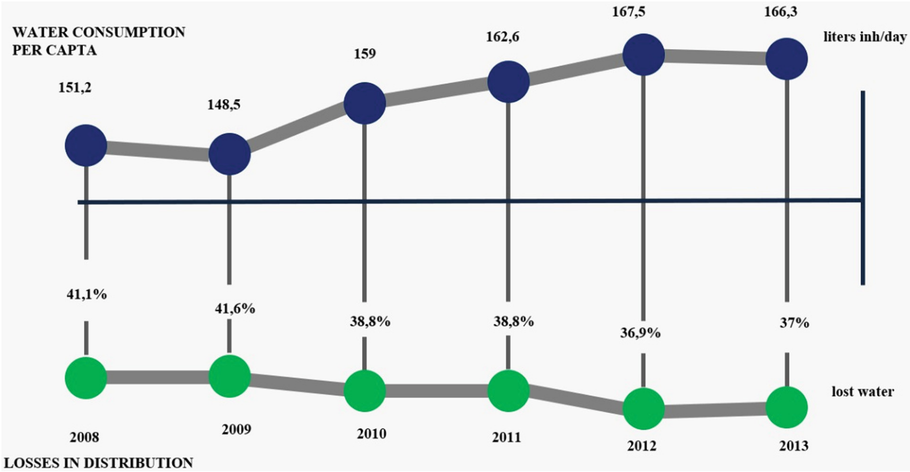


Fig. 1. Amount of *water consumption per capita* and *losses in the distribution* network

The chart on Fig. 1 represents the emergency scenery, pointing out a water resources loss of 37% in distribution by 2013, when every individual of the population consumed 166,3 L of water per day, approximately 151% of the amount considered necessary to each person, according to the World Health Organization (110 L/day) [5].

Motivated by the scenario above, this article presents the development of an autonomous non-intrusive leakage measurement system connected to a data network using OPC-UA protocol and Edge Computing concepts based on the platform Industry 4.0. Therefore, the article presents the initial state of a doctoral project, whose main objective is the implementation of a non-invasive water-leakage detection system, service oriented (SOA). At the present time, the acquisition and spectral analysis was validated, as it will be demonstrated on this work. The next steps comprehend the development of the proposed architecture, through the study of the protocol OPC-UA, delivering a solution based on SOA. Furthermore, the implementation of protocols to attend the necessity of low consumption on the edge will be studied, i.e. protocols regarding Low Power Wide Area Networks (LPWAN).

This paper is organized as the following: Sect. 1 presents an introductory look at geophones for leak detection, Internet of Things (IoT) and OPC-UA, which are the concepts that support this solution proposal model. Section 2 presents the contribution of this work in the context of systems directed to services for the industry. Section 3 details theoretical concepts for leak detection in a non-invasive way, concerning the identification of the frequencies components in an acquired signal based on Fourier Discrete Transform and the Goertzel's Algorithm. In the Sect. 4, the detailed solution is presented: the concept of edge computing and the architecture of a system to monitor the water distribution network of a given system. In Sect. 5, results and analysis are made available. Finally, in the Sect. 6 are presented the final conclusions and future works.

1.1 Electronic Geophone

Electronic geophones are devices responsible for detecting vibrations generated by water leaks, transforming them into sounds called leakage noises, which are monitored by a human operator. The leakage noise is generated by vibrations from damaged pipes in the distribution network. If there is a leak in a pressurized network, water is expelled from the pipe in direct contact with the soil causing vibration transmitted to the surface [7]. The geophone, which resembles a stethoscope, picks up the vibrations on the surface caused by such leaks. Buried leaks are usually in the range of 200 to 600 Hz and leaks in the distribution network are in the range of 600 to 2,000 Hz [8].

There are two inspection methods using geophones. One method is to use them in direct contact with valves, hydrants, faucets, etc., applying the electronic listening stick to check whether there is leakage noise or not. Another method is to listen to the surface noise using ground sensors [8].

The electronic geophone requires an operator who will classify the detected noise as a leak or not. Therefore, human error can cause uncertainty in some cases. For example, where there are cars or pedestrians in the area. The noise produced by those elements may confuse the operator in his evaluation. Thus, leakage detection depends mainly on the operator's auditory sensitivity [8].

1.2 "The Internet of Things" and OPC-UA

"The Internet of Things" (IoT), a term created by Kevin Ashton, refers to the fusion of the Internet and the physical world, when the frontier between the digital and the physical worlds becomes ever more tenuous [9]. This is due to the constant growth in the number of intelligent machines and objects which communicate and process information autonomously [9]. The use of "IoT" makes the development of collaborative autonomous systems possible, encompassing thematic areas such as artificial intelligence, edge computing and communication protocols in distributed systems, which are the core themes of this work. When it comes to "communication protocols", one of the most used worldwide is the OPC. The standardization of the OPC protocol was defined in 1996 and was originally based on the Microsoft OLE model. OPC is an acronym for OLE for Process Control, which is basically a communication protocol between Windows applications [10]. In 2017, a new OPC specification was developed based on the open XML standard, which is a type of communication with content similar to HTML pages, where data is structured hierarchically [11]. This new open standard allowed the execution of OPC in systems not only based on the Windows Operating System [11].

This new OPC specification based on XML is called OPC-UA. Today, OPC-UA, defined in the standard IEC 62541 as Openness, Productivity Collaboration [12, 13].

2 Relationship to Industrial and Service Systems

Due to the ever growing demand for water, mainly as a vital element for human survival and for the operation of several production systems, efficient control of water losses caused by non-visible leaks becomes essential.

Some leaks may be easy to detect when they are on the surface. However, when the leak is not visible, that is, when it does not appear on the surface, it is only detectable by changes in the consumption measurement showing extrapolation of the expected consumption average. This way, due to the lack of effective and efficient control solutions, leaks may be recurring for months until they are detected. Thus, in order to achieve a sustainable use of water, it is necessary to develop a control system for leakage inspection that optimizes maintenance process and makes it predictive.

The autonomous system proposed in this work can identify malfunctions caused by problems related to non-visible water leaks at the exact moment they happen, allowing detection in a non-intrusive way. The application of the autonomous system in businesses, industries and residential condominiums optimizes action and repair time when such leaks take place. It also optimizes control of losses in the distribution network. The necessary information will be provided through protocols based on "IoT" and OPC-UA which enable the technological implementation of the system.

3 Fundamental Concepts

The development of the sensors is based on the use of a piezoelectric transducer and an electronic signal conditioning board which will provide the frequencies of the leakage vibrations in the form of acoustic signals. By analyzing the frequencies of such leakage noise, it is possible to identify leakage frequencies through signal processing techniques.

3.1 Identification of Frequencies

Sound signals detected by microphones or vibrations captured by piezoelectric sensors can be converted to frequency by means of Discrete Fourier Transform, described by Eq. 1 [14, 15].

$$X_k = \sum_{n=0}^{N-1} x_n e^{-\frac{j2\pi kn}{N}} \tag{1}$$

In the context of optimization, the Discrete Goertzel Transform takes place; a more simplistic approach of the DTFT, according to Eq. 2.

$$X[k] = \sum_{n=0}^{N-1} x[n] W_N^{nk}, W_{N=} e^{-j\frac{2\pi}{N}nk} \quad (2)$$

The Goertzel Transform, or Goertzel's algorithm, enables the investigation of a spectral range, opposite to the whole spectrum. That is, in the case being studied, it would be possible to analyze the signal in detail, in search of signals ranging from 100 to 1000 Hz only, which, according to the presented results, are characteristic of leaks. In addition, there is no component phase calculation, thus reducing computational complexity [16].

4 Proposed Architecture

The proposed architecture developed in this work consists in the utilization of devices able to detect the vibrations originating from leakage in the pipes. To attend the features of a monitoring system, we propose the utilization of such devices positioned over a water distribution system. Every device is responsible for the local monitoring of the vibrations utilizing piezoelectric sensors. It is worth emphasizing that the mentioned devices are sensor elements connected to low-consumption microcontrollers, acting like nodes on a network. Figure 2 presents in focus the piezoelectric sensor along with the microcontroller.

In this architecture, the data processing and the generation of useful information is carried out in an application server, based on a microprocessor type Cortex A. To attend this specification, every monitoring device is composed of connectivity modules, which enable the information be sent to the server.

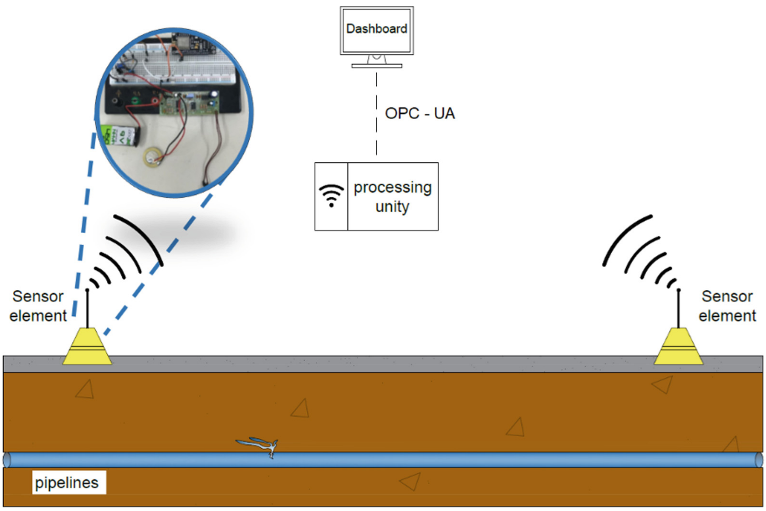


Fig. 2. Proposed architecture and prototype used for the signal acquisition

On the server, the vibration data is processed and analyzed on the domain frequency. The data analysis on this domain allows the identification of the characteristics of a pipeline leakage. Finally, it becomes possible to utilize the information on a monitoring system, generating alarms and notifications according to what is mentioned on the next section. In this case, the protocol OPC-UA was chosen due to the standard proved by the IEC 62541, a well-stablished standard protocol in the Industry 4.0.

5 Results and Analysis

The noise produced in the field may present different frequency components. The materials used in this experiment were HDPE (High Density Polyethylene) and PVC (Vinyl Polychloride) and sizes 1/2″ and 3/4″.

5.1 Frequency Results

To evaluate the proposed system, two test scenarios were created. The first one contains a PVC pipeline and the second one a HDPE pipeline. Then, a signal analysis (Goertzel's algorithm explained in Sect. 3.1) was applied with the acquired signals with and without leakage.

Figure 3 illustrates the frequency spectrum of leakage noise in a water distribution network with PVC pipes.

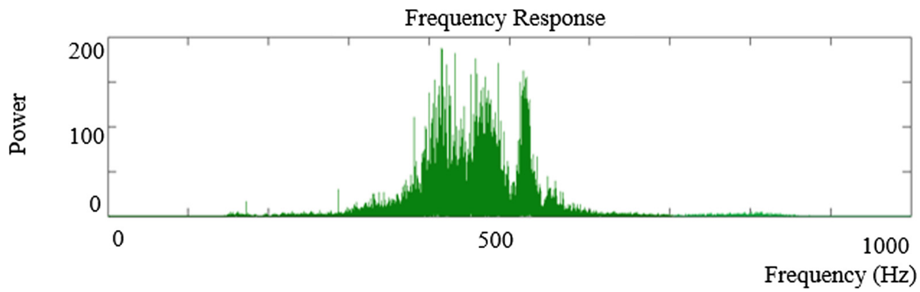


Fig. 3. Frequency spectrum of water leakage noise, amplitude of noise record and *frequency of response power.*

It was observed that the frequency range of the noise analyzed is within 100 Hz to 1,000 Hz. Different types of soil interfere in the signal strength, but not in the frequency. In Fig. 4 the material used was HDPE.

Figure 5 shows the case of the system with HDPE pipes without leaks.

The results presented above show that the frequency components obtained from the Goertzel's algorithm represent a useful feature for the leak detection. Regardless the materials tested (PVC and HDPE), for situations with anomalies in the pipelines, it is possible to observe that the components have a module of higher intensity between 100 and 1000 Hz.

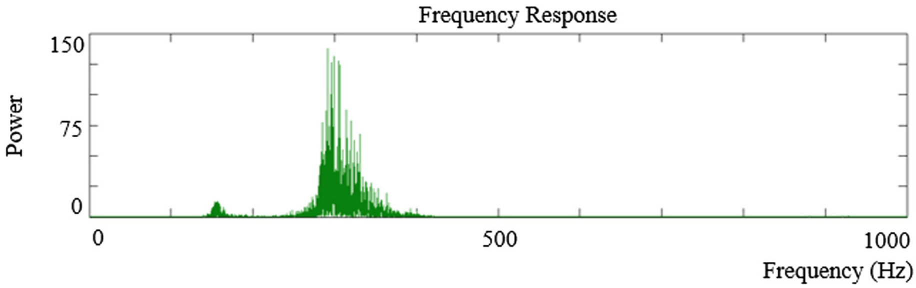


Fig. 4. Frequency spectrum of leakage noise in a water distribution network with HDPE pipe.

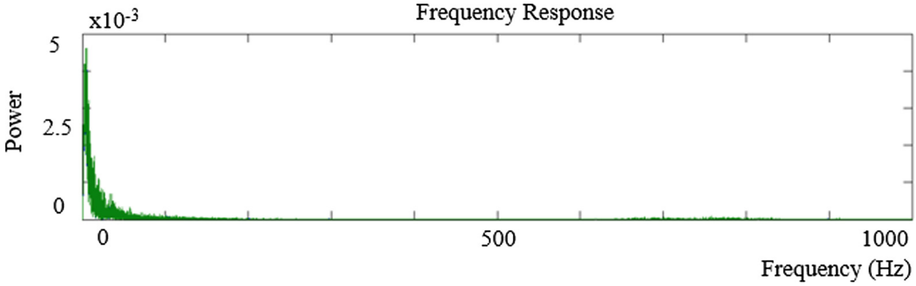


Fig. 5. Frequency spectrum of noise in a distribution network with HDPE pipes that does not present water leaks.

6 Conclusion

This article presents an architecture model for a distributed network of sensors, capable of detecting and identifying noise from water leaks contributing to the proposal of a monitoring system for water loss reduction, based on the concepts of signal analysis and processing and the use of technologies based on "IoT" and Industry 4.0. The stages of acquisition and processing of the signal, whose results were presented, lay on the foundation of the proposed architecture, with the future implementation of the protocol OPC-UA between the central unit and the dashboard, and the studies of LPWAN between the sensor devices and the central unit.

The distribution of devices and the use of edge computing allow detection of problems throughout the water distribution network in a systematic way. Therefore, it contributes directly to control the use of water in industrial systems or residential condominiums, optimizing maintenance and repairs. The concepts of applied "IoT" and "OPC-UA Communication" enable the implementation of the autonomous system within the scope of Industry 4.0.

This study allowed the validation of a concept that the authors can use in the future and evolve to the development of an autonomous system using able to identify the exact location of water leaks.

Acknowledgment. We would like to thank the computer engineering department of FACENS, CAPES, CNPq, FAPESP e PET for the development of this study.

References

1. Cirilo, J. A.: Crise hídrica: desafios e superação. Revista USP, São Paulo, n. 106, pp. 45–58 (2015)
2. Maddocks, A., Young, R.S., Reig, P.: Ranking the World's Most Water-Stressed Countries in 2040. World Resources Institute (2018). http://www.wri.org/blog/2015/08/ranking-world%E2%80%99s-most-water-stressed-countries-2040
3. ABES: Controle e Redução de Perdas nos Sistemas Públicos de Abastecimento de Água. São Paulo, SP, 100 p. (2015)
4. Barifouse, R.: Maior crise hídrica de São Paulo expõe lentidão do governo e sistema frágil. BBC Brasil. https://www.bbc.com/portuguese/noticias/2014/03/140321_seca_saopaulo_rb (2014)
5. Eduardo, A.: Água no Brasil. Folha de S. Paulo. São Paulo, SP. http://www1.folha.uol.com.br/infograficos/2015/01/118521-agua-no-brasil.shtml (2015)
6. Oliveira, G., Marcato, F.S., Scazufca, P., Pires, R.C.: Perdas de Água 2018 (SNIS 2016): Desafios para Disponibilidade Hídrica e Avanço da Eficiência do Saneamento Básico. Instituto Trata Brasil (2018). http://www.tratabrasil.org.br/images/estudos/itb/perdas-2018/estudo-completo.pdf
7. Sardinha, J., Serranito, F., Donnelly, A., Marmelo, V., Saraiva, P., Dias, N., Guimarães, R., Morais, D., Rocha, V.: Active WaterLossControl: EPAL TechnicalEditions. EPAL – Grupo Águas de Portugal. 2ª Edição. Portugal (2017)
8. Xuesong, S., Min L., Chenwei H.: Application of fuzzy inference in the confidence analysis on the sound wave data of water leakage. In: 8th International Conference on Electronic Measurement and Instruments. Xi'an, China (2017)
9. Miorandi, D., Sicari, S., Pellegrini, F., Chlamtac, I.: Internet of Things: Vision, Applications and Research Challenges. Elsevier, Varese (2012)
10. Imtiaz, J., Jaspemeite, J.: Scalability of OPC-UA Down to the Chip Level Enables "Internet Of Things". IEE, Lemgo, Alemanha (2017)
11. Schwarz, M.H., Borcsok, J.: A survey on OPC and OPC-UA. In: XXIV International Conference on Information, Communication and Automation Technologies (ICAT), IEE. Kassel, Alemanha (2013)
12. Bonomi, F., Milito, R., Zhu, J., Addepalli, S.: Fog computing and its role in the internet of things. In: Proceedings of the First Edition of the MCC Workshop on Mobile Cloud Computing, pp. 13–16. Helsinki, Finland (2012)
13. Gezer, V., Ruskowski, M., Um, J.: An extensible edge computing architecture: definition, requirements and enablers. In: The Eleventh International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies. Barcelona, Espanha (2017)
14. Smith, S.W.: The Scientist and Engineer's Guide to Digital Signal Processing. San Diego, California (1998)
15. Lathi, B.P.: Sinais e Sistemas Lineares. 2ª edição. 856 p., Bookman. Rio Grande do Sul, Brasil (2017)
16. Joe, F.C., Kilabi, M.T.: A sliding goertzel algorithm. Signal Process. **52**(3), 283–297 (1996)

Communication Systems

Energy Efficient Massive MIMO Point-to-Point Communications with Physical Layer Security: BPSK vs QPSK Decomposition

David Borges[1(✉)], Paulo Montezuma[1,2], Rui Dinis[1,2], and Pedro Viegas[2]

[1] DEE, FCT Universidade Nova de Lisboa, Caparica, Portugal
d.borges@campus.fct.unl.pt
[2] IT, Instituto de Telecomunicações, Av. Rovisco Pais, Lisbon, Portugal

Abstract. Massive multiple-input multiple-output systems (mMIMO) are the most prevalent candidates for the next generation of wireless communication. Yet even with mMIMO systems the joint optimization of spectral and energy efficiencies can be only attained by combining high order signal constellations and efficient power amplification. In order to push this limitation, the transmitter can spread the information into several amplification branches, which are the result of the decomposition of multilevel constellation symbols into quasi constant envelope signals. Nevertheless, the high number of antennas involved in this type of communication leads to an increase of the channel matrix's size and therefore the complexity of the equalization process can create drawbacks for the power consumption and latency. In this paper we will study the combination of a multi-layer transmitter with a low complexity receivers based on an iterative block decision feedback equalizer (IB-DFE). These receivers avoid the matrix inversion operation in the equalizer the feed-forward by replacing it with an equal gain combiner (EGC) or a maximum ratio combiner (MRC) module. Results show that can be used without penalties on performance provided that the number of antennas involved is high.

Keywords: Massive MIMO systems · Efficient power amplification · Constellation shaping · Low complexity detection

1 Introduction

Massive multiple-input multiple-output (mMIMO) systems allow many improvements in the communication performance bounds [1]. Additionally, to maximize spectral efficiency, high order multilevel constellations can be employed. However, high order multilevel constellations have high peak-to-average power ratio (PAPR) which restricts the efficiency of the amplification [2]. In order to improve the power amplification

This work was supported in part by FCT - Portuguese Foundation for Science and Technology through the PhD scholarship SFRH/BD/131093/2017, IT UID/EEA/50008/2013 (plurianual founding and project GLANCES), EnAcoMIMOCo EXPL/EEI-TEL/2408/2013 and UID/EEA/50008/2013 - MM5G.

Published by Springer Nature Switzerland AG 2019
L. M. Camarinha-Matos et al. (Eds.): DoCEIS 2019, IFIP AICT 553, pp. 283–295, 2019.
https://doi.org/10.1007/978-3-030-17771-3_25

efficiency, the symbols with high envelope fluctuations can be decomposed into low order modulation symbols, which benefit of low envelope fluctuations [3]. As example, we can decompose one 16 Quadrature Amplitude Modulation (QAM) symbol into four Binary Phase Shift Keying (BPSK) symbols or two Quadrature Phase Shift Keying (QPSK) symbols. By doing this, each component can be amplified separately without distortion by NonLinear (NL) power amplifiers, since the PAPR of a BPSK or a QPSK is much lower than the PAPR of the original constellation. Using not oversized amplifiers and working in their optimal operation zone, energy efficiency of each amplifier can be highly improved. In this architecture, each power amplifier is directly connected to a single antenna. This means that each antenna will transmit a single component from the decomposed higher order symbol, i.e. a BPSK or QPSK symbol. At the channel level, the original symbol can be obtained through the combination of all the components. In other words, at the receiver side, each antenna will receive the signal from all transmitting antennas which will add up to form the original high order constellation symbol. Moreover, a multi-layer system could be created using this approach. The first layer is composed by the multi-branch amplification structure, the second layer is composed by the antenna array connected to each amplifier's output and the third layer associated to spatial multiplexing is composed by several layers 1 and 2 in parallel.

However, the high number of antennas employed in the multi layered transmitter and receiver communication process will imply an increasing size of the channel matrix, which means a higher complexity in the equalization step due to matrix inversion operation. In order to overcome this problem, a low complexity receiver, combining an Iterative Block Decision Feedback Equalization (IB-DFE) and a Maximum Ratio Combiner (MRC) or an Equal Gain Combiner (EGC) is implemented for systems based on these multi layered transmitters and analyzed in this paper. These simplified receivers avoid the usual matrix inversion inherent to mMIMO conventional detection scheme, but as we shall see the conditions for such approach must be fulfilled to avoid any impact on system performance.

The rest of this paper is organized as follows: Sect. 2 characterizes the system's architecture; Sect. 3 contains the description of the system; Simulation results and performance analysis are presented in Sect. 4; Sect. 5 resumes the paper.

2 System Architecture and Characterization

It was already shown that high order constellations can be decomposed in polar components with the constellation symbols expressed as a function of the corresponding bits [4]. Let $\mathcal{S} = \{s_0, s_1, \ldots, s_{(\mathrm{N}-1)}\}$, a constellation with M points (i.e., $\#\mathcal{S} = M$), where $s_n \in \mathbb{C}$. To each constellation point s_n we associate a set of $\mu = \log_2(M)$ bits in polar form $\mathcal{B} = \{b_n^0, b_n^1, \ldots, b_n^{(\mu-1)}\}$, with $b_n^{(i)} = \pm 1 = 2\beta_n^{(i)} - 1$ and $\beta_n^{(i)} = 0$ or 1. The set of μ bits can be decomposed in $M = 2^\mu$ different subsets $\mathcal{B}_m$, $m = 0, 1, \ldots, M-1$.

Having M constellation points in $\mathcal{S}$ and M different subsets of $\mathcal{B}$, $\mathcal{B}_0, \mathcal{B}_1, \ldots, \mathcal{B}_{M-1}$, we can write

$$\mathrm{s_n} = \sum_{\mathrm{m}=0}^{\mathrm{M}-1} \mathrm{g_m} \prod_{\mathrm{b_n^{(i)}} \in \mathcal{B}_\mathrm{m}} \mathrm{b_n^{(i)}}, \mathrm{n} = 0, 1, \ldots, \mathrm{M} - 1, \tag{1}$$

which corresponds to a system of M equations (one for each s_n and M unknown variables g_m). Moreover, by doing $m = \left(\gamma_{(\mu-1,m)}, \gamma_{(\mu-2,m)}, \cdots, \gamma_{(1,m)}, \gamma_{(0,m)}\right)$ we can associate m to its corresponding binary representation with μ bits, which turns possible to define $\mathcal{B}_m$ as the set of bits where the bit $b_n^{(i)}$ is included if and only $\gamma_{(i,m)}$ is 1. Based on that we may write

$$s_n = \sum_{m=0}^{M-1} gm \prod_{i=0}^{\mu-1} \left(b_n^{(i)}\right)^{\gamma(i,m)}. \tag{2}$$

The transmitter layered structure of Fig. 1 takes advantage of this decomposition in polar components, and uses a mMIMO scheme with $N_v \times N_m \times N_b$ antenna elements, arranged in N_v sets of $N_m \times N_b$ antennas. Conventional beamforming schemes could be implemented by a layer 2 with N_b antenna elements connected to each one of N_m amplification branches. For spatial multiplexing a layer 3 with a set of $N_v \times N_m$ antennas is needed, where the N_m antennas of layer 1 are associated to the signal components of the constellation symbol and N_v sets of N_m antennas are used to transmit simultaneously N_v different constellation symbols (an example of a transmitter based on a layer 1 and layer 3 combination can be seen in Fig. 2).

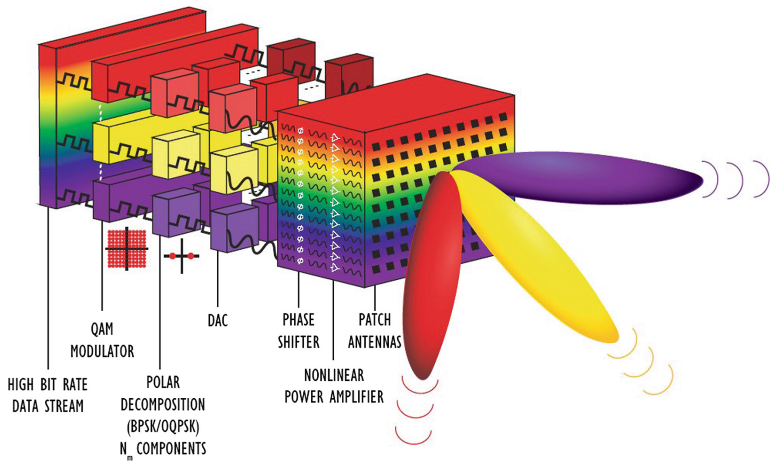


Fig. 1. Transmitter layered structure.

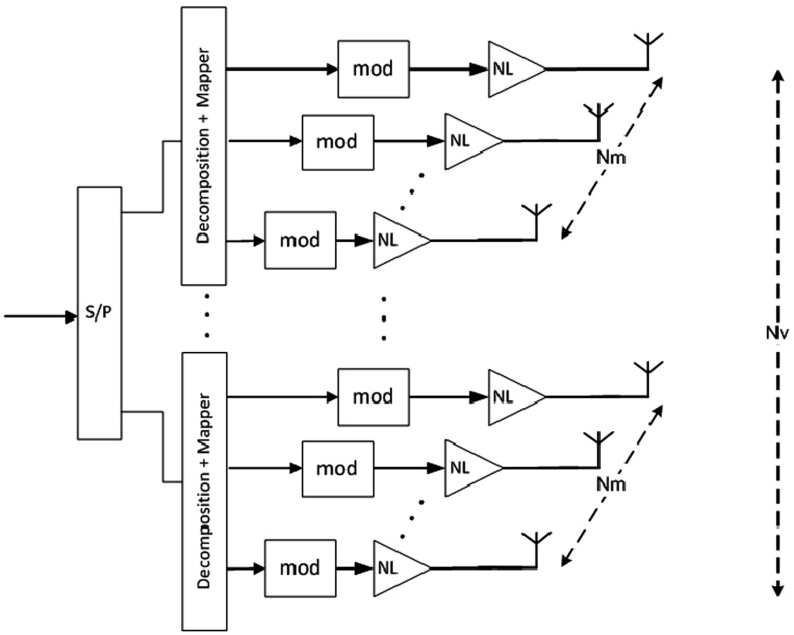


Fig. 2. Layered transmitter structure with layer 1 and layer 3.

It follows that power efficiency comes improved due the lower PAPR of the component signals and the possible use of NL amplifiers in layer 1 [3, 5]. Another important aspect lies on uncorrelated signals associated to each RF chain, which means an omnidirectional radiation pattern for each set of N_m antennas. But it is important to mention that each set of N_m antennas implements directivity at information level, since the constellation points of the transmitted signal maintain their positions at the desired direction Θ but are scrambled for other directions.

In Fig. 3 it can be seen that the transmitted components will arrive to the receiver with different delays, related with the spacing d between antennas and the desired transmission angle Θ. However, these delays are compensated by introducing phase shifts between antennas related to the spacing between the amplification branches in the transmitter. This means that in direction Θ the receiver sees a symbol constellation without any distortion due to phase rotations of signal components. In other directions, phase shifts between antennas do not compensate the phase shifts due the different propagation delays and the received symbol suffers distortions since it results from the combination of signal components with the wrong phases. Obviously antenna spacing can be adjusted to assure the desired configuration of information directivity for a specific direction. Therefore, any user can demodulate the received signal if is in the desired direction while other users in other directions receive a scrambled constellation. Hence, to decode the information, each j user needs to be in the right direction Θ_j.

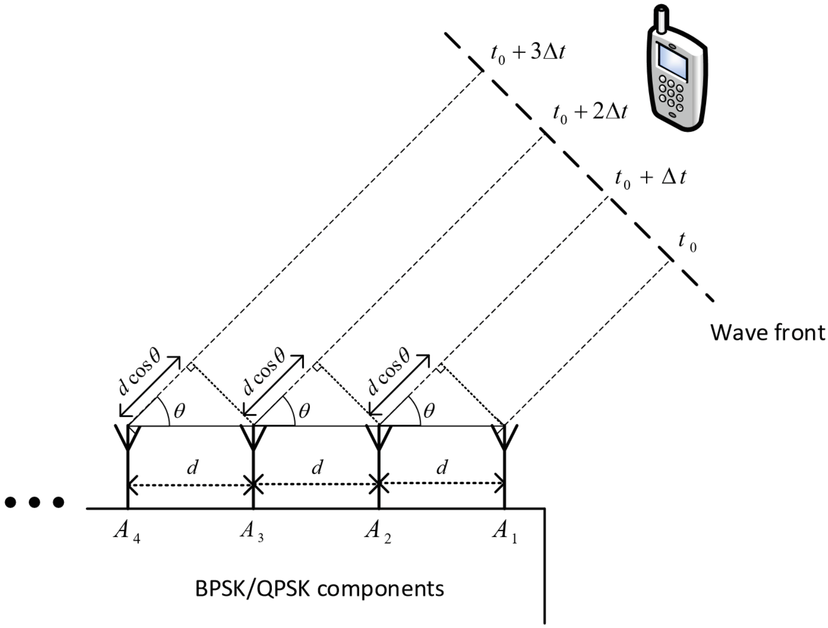


Fig. 3. Constellation rotation due to delay.

3 System Description

Consider now the MIMO scenario shown in Fig. 4 characterized by a point-to-point communication link between a transmitter with $T = N_v \times N_m$ antennas and a receiver with $R \geq T$ receive antennas. The transmitter's configuration has the layered structure shown in Fig. 2, but were Layer 3 is composed by a set of N_v transmitters each one composed by N_m amplification branches in parallel (for a 16-quadrature amplitude modulation (16-QAM) decomposition into QPSK or BPSK components we have $N_m = 2$ and $N_m = 4$, respectively). This means a third layer with $T = N_v \times N_m$ antennas that transmits simultaneously N_v symbol constellations (more exactly the components of N_v symbols from a 16-QAM constellation). The transmitter configuration is known and the receiver is assumed to be in the right direction Θ (in which the constellation is optimized).

Coupling effects among antennas are avoided by a horizontal spacing of $\lambda/2$ between antennas of layer 1 and a vertical spacing of λ between each set of N_m antennas. It is assumed that all antennas at transmitter are assigned to an authorized receiver with R antennas.

It is assumed that we have frequency slective channels linking transmit and receive antennas and a SC-FDE block transmission technique. To compensate the channel's frequency selectivity it can be adopted an IB-DFE receiver with the structure shown in Fig. 3 [6, 7].

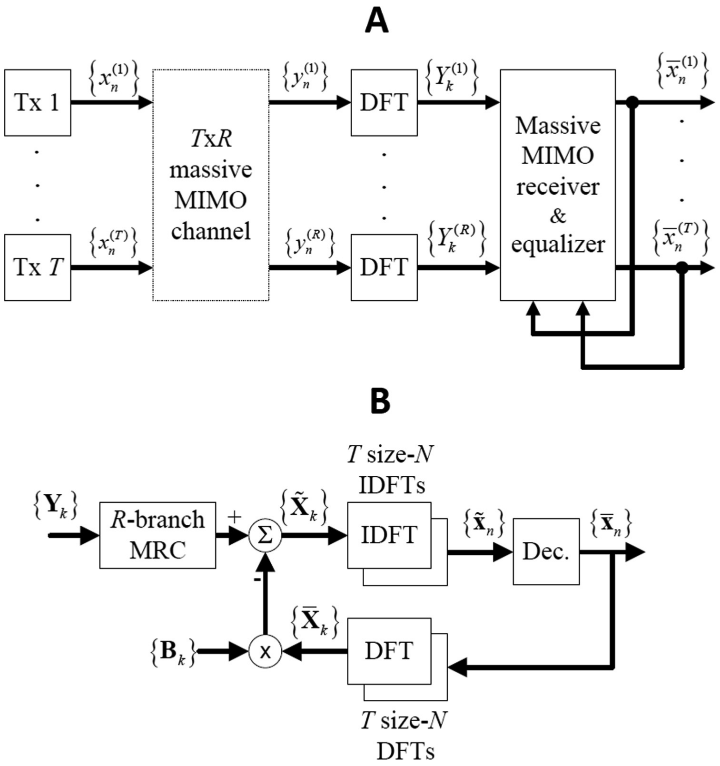


Fig. 4. Massive MIMO system architecture for SC-FDE schemes **(A)** and detail of the low complexity massive MIMO receiver and equalization **(B)**.

At transmitter the t^{th} antenna sends the block of N data symbols $\{x_n^{(t)}; n = 0, 1, \ldots, N - 1\}$ being $\{y_n^{(r)}; k = 0, 1, \ldots, N - 1\}$ the received block at the r th receiver's antenna. As usual, a cyclic prefix with duration higher than the overall channel delay is appended to each transmitted block and removed at the receiver. Under these conditions, the corresponding frequency-domain received block $\left\{Y_k^{(r)};\ k = 0, 1, \ldots, N - 1\right\}$ is given by

$$\mathbf{Y}_k = \left[Y_k^{(1)} \ldots\ Y_k^{(R)}\right]^T = \mathbf{H}_k \mathbf{X}_k + \mathbf{N}_k, \tag{3}$$

where $\mathbf{H}_k$ denotes the $R \times T$ channel matrix for the k th frequency, with (r, t)th element $H_k^{(r,t)}$, $\mathbf{X}_k = \left[X_k^{(1)} \ldots X_k^{(T)}\right]^T$ and $\mathbf{N}_k$ represents the channel noise.

In an iterative Minimum Mean Squared Error (MMSE) receiver the data symbols in each iteration are obtained through the inverse discrete Fourier transform (IDFT) of the block $\left\{\tilde{X}_k^{(t)}; k = 0, 1, \ldots, N - 1\right\}$, and are given by

$$\tilde{\mathbf{X}}_k = \left[\tilde{X}_k^1 \ldots \tilde{X}_k^{(R)}\right]^T = \mathbf{F}_k \mathbf{Y}_k - \mathbf{B}_k \overline{\mathbf{X}}_k, \quad (4)$$

(see details in [8], e.g.). Interference cancelation is done using $\overline{\mathbf{X}}_k = \left[\overline{X}_0 \ldots \overline{X}_{N-1}\right]$, with $\overline{X}_k$ denoting the frequency-domain average values conditioned to the FDE output for the previous iteration, which can be computed as described in [8, 9]. Coefficients F_k, B_k and the correlation coefficient ρ can be computed as described in [4, 7, 8]. In first iteration no information is available about the transmitted symbols and $\overline{\mathbf{X}}_k = \mathbf{0}$, which means that this receiver is equivalent to the linear frequency-domain MMSE receiver. Subsequent iterations use the average values conditioned to the receiver's output from previous iteration to remove the residual intersymbol interference (ISI).

Due to matrix inversions the IB-DFE receiver has a computational complexity of $O(N_R^3)$. To avoid this problem other two low-complexity iterative frequency-domain receivers, denoted as maximum ratio detection (MRD) and equal gain detection (EGD) are also considered [10]. Similarly to [10] the ratios R/T between receiving and transmitting antennas considered here are at least equal or higher than 4.

As mentioned in [10], the MRD receiver (see Fig. 3 B) is characterized by

$$\tilde{\mathbf{X}}_k = \mathbf{\Psi} \mathbf{H}_k^H \mathbf{Y}_k - \mathbf{B}_k \bar{\mathbf{X}}_k, \quad (5)$$

and

$$\mathbf{B}_k = \mathbf{\Psi} \mathbf{H}_k^H \mathbf{H}_k - \mathbf{I}, \quad (6)$$

where $\mathbf{\Psi}$ denotes a diagonal matrix in which the (t,t)th element is given by $\left(\sum_{k=0}^{N-1} \sum_{r=1}^{R} |H_k^{(r,t)}|^2\right)^{-1}$, takes advantage of the fact that

$$\mathbf{H}_k^H \mathbf{H}_k \approx R\mathbf{I}, \quad (7)$$

which is accurate when $R \gg 1$ and the correlation between different channels have small correlation (i.e., different transmit and receive antennas). On the other hand, the EGD characterized by

$$\mathbf{B}_k = \mathbf{\Psi} \mathbf{A}_k^H \mathbf{H}_k - \mathbf{I}, \quad (8)$$

where $\mathbf{\Psi}$ denotes a diagonal matrix whose (t,t)th element is given by $\left(\sum_{k=0}^{N-1} \sum_{r=1}^{R} |H_k^{(r,t)}|\right)^{-1}$, takes advantage of the fact that for mMIMO systems with $R \gg 1$ and small correlation between the channels associated to different transmit and receive antennas, the elements outside the main diagonal of

$$\mathbf{A}_k^H \mathbf{H}_k \quad (9)$$

are much lower than the ones at its diagonal, where $(i,i')th$ element of the matrix $\mathbf{A}$ is $[\mathbf{A}]_{i,l} = \exp(j\arg([\mathbf{A}]_{i,l}))$.

4 Results

Bit error rate (BER) results are obtained through Monte Carlo experiments (at least 150 error events are considered for each value of BER). A 16-QAM constellation is adopted in Layer 1, where the decomposition applied can be based on QPSK or BPSK components. This means that in first case we have $N_m = 2$ branches with QPSK components and in second case we have $N_m = 4$ BPSK components. In both cases the transmitted symbols s_n are selected with equal probability from the original constellation. It is assumed linear power amplification at the transmitter and perfect time and frequency synchronization conditions at the receiver. The results are expressed as function of $\frac{E_b}{N_0}$, where $N_0/2$ denotes the noise variance and E_b represents the energy of the transmitted bits. The number of users is always settled at $T = 5$. For EGD, MRD and IB-DFE receivers the number of iterations is 4 (previous study and showed than there is no advantage to consider higher number of iterations, since there are no significant improvements in performance) [10].

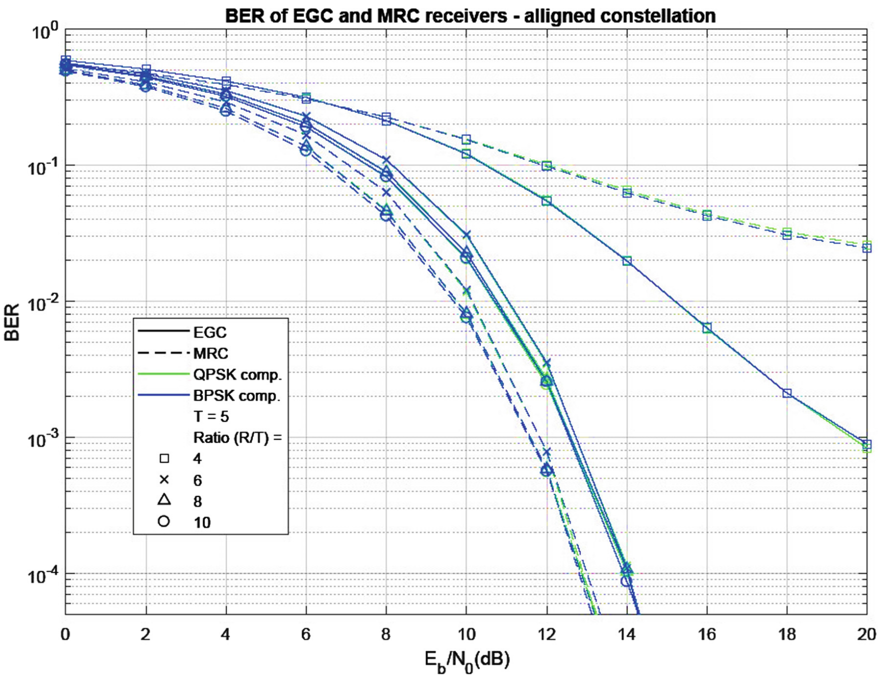


Fig. 5. BER performance for MRD and EGD with $\Delta\Theta = 0^o$ and variant R/T.

Results shown in Figs. 5 and 6 refer a layer 1 with both $N_m = 2$ and $N_m = 4$ and angular separation to the direction Θ of $\Delta\Theta = 0^o$ (i.e. the user is placed in the correct direction where the components add in phase). Values of R/T between 4 and 10 are also adopted. From the comparison of the results, it is clear that when the ratio R/T

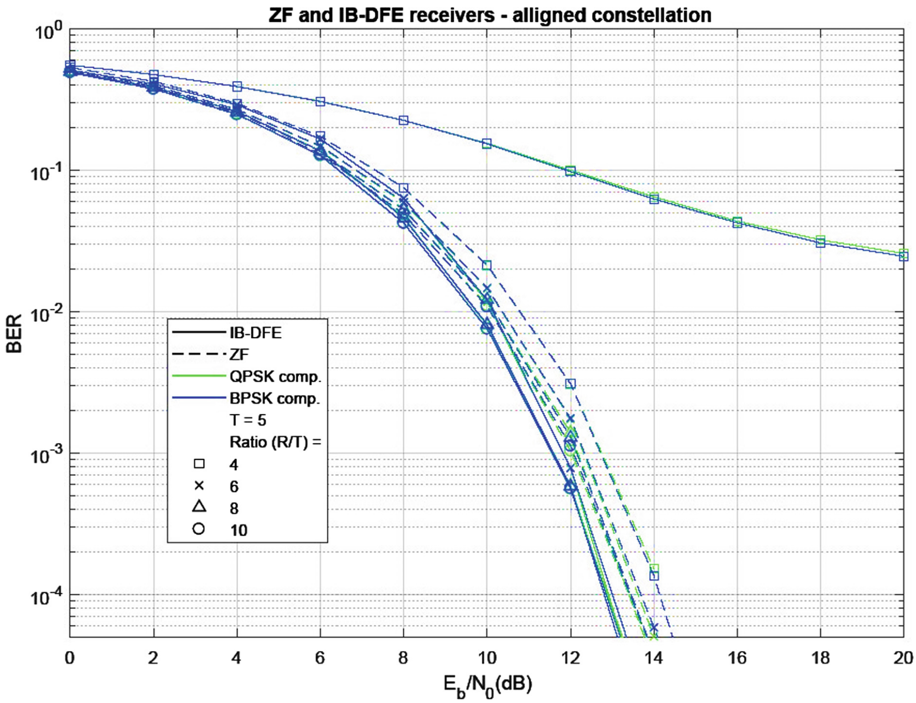


Fig. 6. BER performance for IB-DFE and ZF with $\Delta\Theta = 0^o$ and variant R/T.

grows, the performance improvement in ZF receiver can be considered as negligible. However for the other receivers only values of R/T higher than 6 turn performance improvements to be considered as negligible (this means that the validity conditions for the approach done to avoid the matrix inversion stands for R/T equal or higher than 6). As expected ZF and EGD perform worst that IB-DFE and MRD receivers, with the two later ones showing similar performances. Another important conclusion is the fact there is no performance disparity between both cases adopted for layer 1, since performances for the decomposition into BPSK and QPSK components are the same. It is important to mention, that the results for $\Delta\Theta = 0^o$ are the same obtained by a conventional spatial multiplexing scheme, which means that combination of layer 1 and layer 3 can be employed without compromising system's performance.

To analyze the impact of errors in the alignment of the users with the direction Θ, we consider a fixed ratio of $R/T = 6$ and two values of $\Delta\Theta$ (0º and 5^o). The results are presented in Figs. 7, 8 and 9 for different receivers compared with ZF. From the comparison of figures, it becomes clear that when the ratio R/T grows, the gap between the MRD and the EGD receiver vanishes (as we can see in Figs. 7 and 8 performance curves of receivers overlap which is not the case of Fig. 5). It is also clear that despite the performance improvement achieved by the higher value of R/T, the performance impact of constellation shaping due to the decomposition of the constellation into BPSK components is higher than QPSK case. Indeed, when at layer 1, it is adopted a decomposition in BPSK components the data link becomes compromised for the ZF receiver, due to a BER of 0.5. Even the other receivers only have reasonable values of BERs for the decomposition into QPSK components. It is also observed that the IB-DFE exhibits more robustness to the constellation shaping, even for the angular error of $\Delta\Theta = 5^o$. On the other hand, both MRD and EGD are severely affected by the angular error. Also, as expected the IB-DFE and the MRD receivers perform better than the EGD one. From the results of Figs. 7, 8 and 9 it is clear that both IB-DFE and MRD receivers have similar performances. Also, as expected, these two receivers perform better than the EGD one. On the other hand, better security is assured. This can be seen in the results regarding the ZF and EGD receivers where it is obvious the higher impact on system performance of any angular error for $N_m = 4$. For $\Delta\Theta = 5^o$ the BER near 0.5 means that users can't decode any useful information.

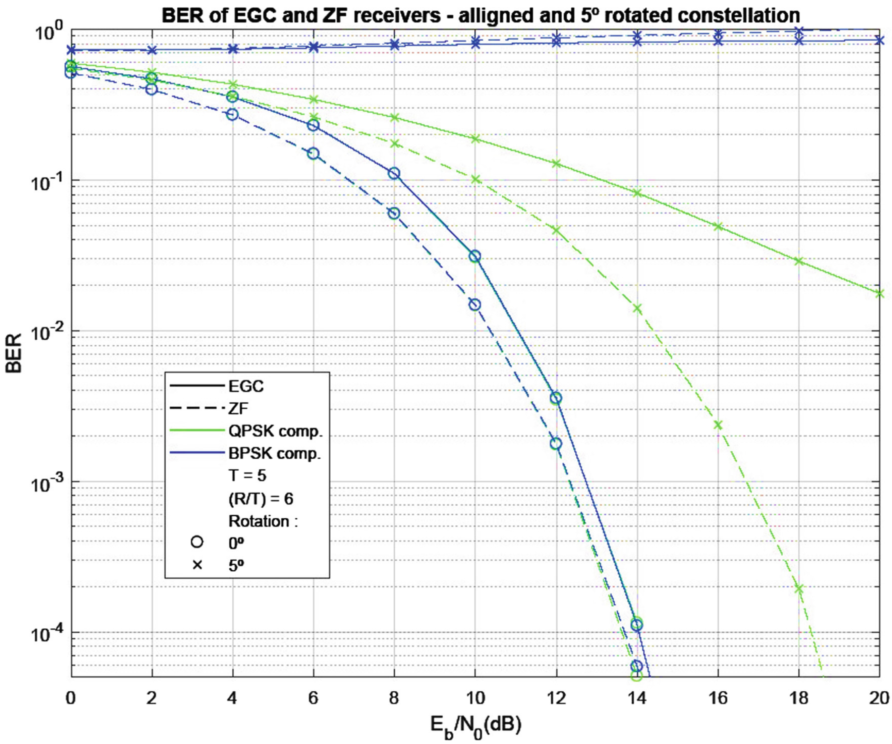


Fig. 7. BER performance comparing ZF and EGD with different $\Delta\Theta$.

Regarding the security level assured by this transmitter, it is evident that the decomposition into BPSK components applied in layer 1 will be more effective than the decomposition into QPSK components. Performance results also show that security due to constellation shaping is achieved independently the type of adopted receiver, since the performance of all receivers is severely affected when $\Delta\Theta \geq 5^o$ for $N_m = 4$.

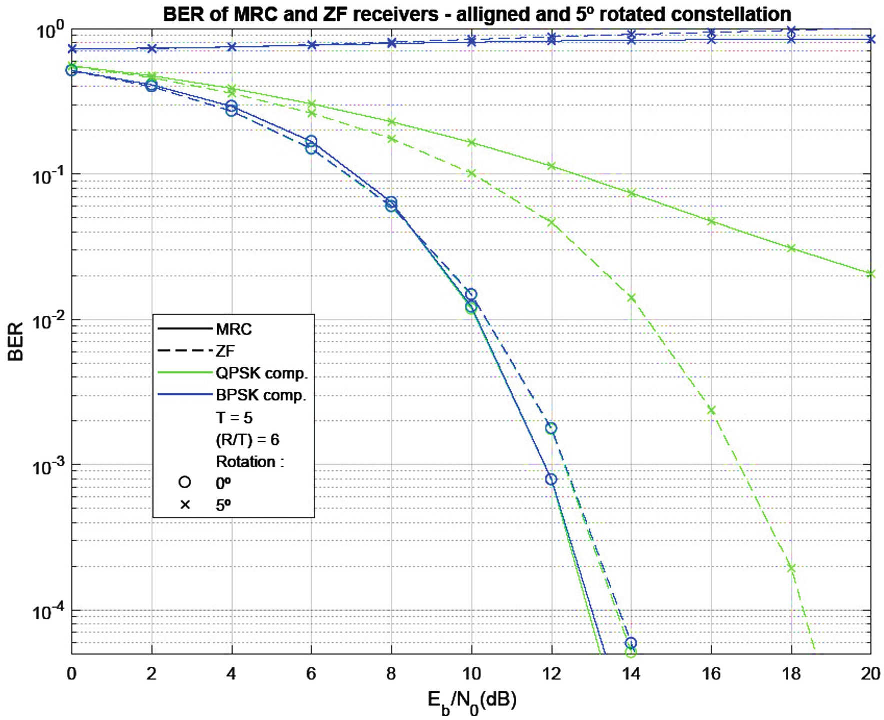


Fig. 8. BER performance for ZF and MRD with $\Delta\Theta = 5^o$ and variant R/T.

More interestingly, good physical security can be assured in point-to-point communication links with only 4 antennas (for $N_m = 2$ previous results showed good tolerance against $\Delta\Theta \leq 5^o$, which can compromise the security). Thus, angular shaping of the transmitted symbol performed in layer 1 can be used to separate user streams and achieve security between them (clearly the angular separation between users should be equal or higher than 5^o or the number of components could be increased). Security can be even increased if beamforming is adopted at layer 2, which will be analyzed in further works.

It should be mentioned that in modern communication systems such as long term evolution (LTE) or further 5G system, various kinds of channel coding and error correction techniques are employed in different layers (mainly in physical and logical layers) to correct residual errors associated to a specific degree of BER. Thus, the BER level of 10^{-4} the bit error rate attainable by the proposed transmission schemes is more than sufficient to achieve the desired quality of service, since when combined with an error correction we may see much lower error rate than the case without error correction (shown in previous figures). Block error rate (BLER) of frame error rate (FER) are commonly used to assess the system performance, since they result from the sum of contributions of the several error correction techniques employed in the layers. This also means that the previous BER results are within the performance ranges employed in physical layer of modern communication systems.

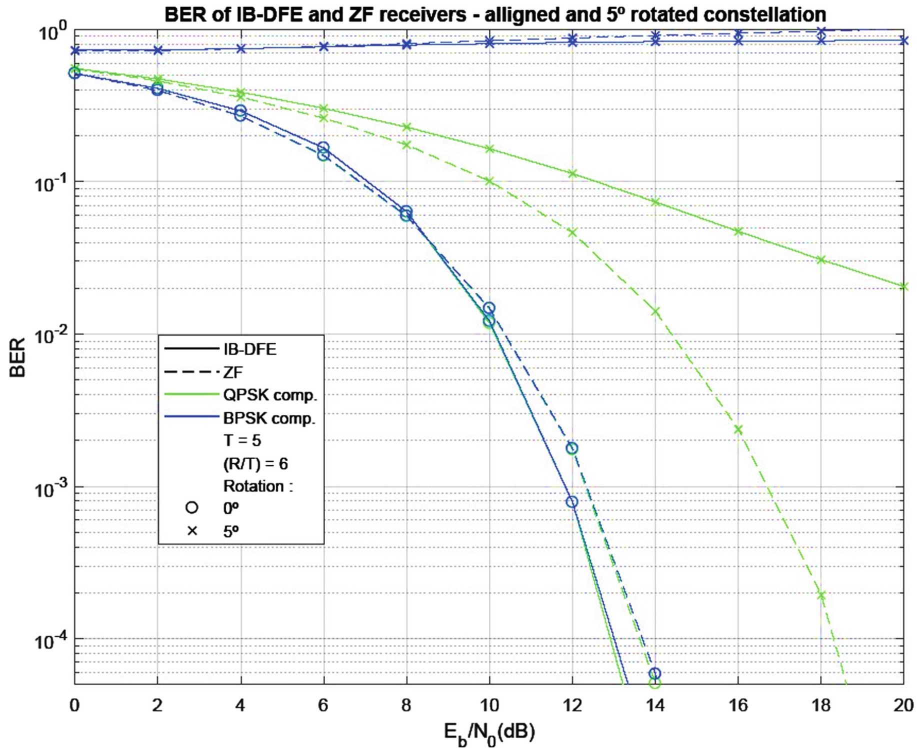


Fig. 9. BER performance for ZF and IB-DFE with $\Delta\Theta = 5^o$ and fixed R/T.

5 Conclusions

In this paper it was shown that a double layer structure with information directivity and spatial multiplexing could be employed at the transmitter to improve power amplification efficiency and to assure physical security. For an aligned receiver $(i.e.\Delta\Theta = 0^o)$, since BER results achieved by both decompositions options are similar to those of classical spatial MIMO multiplexing, we may say that both higher power amplification efficiency and physical layer security come without sacrificing system's performance. Despite of using more antennas than the traditional scheme, in further millimetric wave communication systems this factor will be not a restriction.

References

1. Larsson, E.G., Edfors, O., Tufvesson, F., Marzetta, T.L.: Massive MIMO for next generation wireless systems. IEEE Commun. Mag. **52**(2), 186–195 (2014)
2. Dinis, R., Montezuma, P., Gusmao, A.: Performance trade-offs with quasi-linearly amplified OFDM through a two-branch combining technique. In: Proceedings of IEEE VTC96, Atlanta, Georgia, US, May 1996

3. Astucia, V., Montezuma, P., Dinis, R., Beko, M.: On the use of Multiple grossly Nonlinear amplifiers for Higly Efficient Linear amplification of multilevel constellations. In: Proceedings of IEEE VTC2013-Fall, Las Vegas, NV, US, September 2013
4. Dinis, R., Montezuma, P., Souto, N., Silva, J.: Iterative frequency-domain equalization for general constellations. In: IEEE Sarnoff Symposium, Princeton, USA, April 2010
5. Montezuma, P., Gusmão, A.: Design of TC-OQAM schemes using a generalised nonlinear OQPSK-type format. IEE Elect. Lett. **35**(11), 860–861 (1999)
6. Benvenuto, N., Tomasin, S.: Block iterative DFE for single carrier modulation. IEE Electron. Lett. **39**(19), 1144–1145 (2002)
7. Dinis, R., Kalbasi, R., Falconer, D., Banihashemi, A.: Iterative layered space-time receivers for single-carrier transmission over severe time-dispersive channels. IEEE Commun. Lett. **8** (9), 579–581 (2004)
8. Silva, P., Dinis, R.: Frequency-Domain Multiuser Detection for CDMA Systems. River Publishers, Aalborg (2012)
9. Silva, F., Dinis, R., Montezuma, P.: Estimation of the feedback reliability for IB-DFE receivers. ISRN Commun. Netw. **2011**, Article No. 30 (2011)
10. Borges, D., Montezuma, P., Dinis, R.: Low complexity MRC and EGC based receivers for SC-FDE modulations with massive MIMO schemes. In: 2016 IEEE GlobalSip, Washington, DC, USA, December 2016

V-GRADIENT: A Density-Aware Geocast Routing Protocol for Vehicular Delay-Tolerant Networks

Henrique Nascimento and Paulo Rogério Pereira(✉)

INESC-ID, Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal
henrique.nascimento@tecnico.ulisboa.pt, prbp@inesc.pt

Abstract. Vehicular Delay-Tolerant Networks (VDTNs) are networks of vehicles that communicate wirelessly, where there are no permanent end-to-end connections. VDTNs have a highly variable topology, with frequent partitions, and possibly low node density. Thus, delay-tolerant routing adopts a store-carry-and-forward message transfer paradigm, where messages have a useful Time To Live (TTL) and are stored until a good contact opportunity arises. Multiple message replicas can be generated to improve delivery probability at the cost of increasing network congestion. In this paper, we propose the V-GRADIENT geocast routing protocol that monitors node density and buffer occupancy, to adapt dynamically the forwarding techniques used to disseminate messages within the geographic region of interest. Simulation results show that V-GRADIENT is capable of controlling network congestion and efficiently deliver messages resulting in better delivery ratios (13–99%) and lower latencies when compared with existing protocols.

Keywords: Wireless communications · Geocast routing · Vehicular delay-tolerant networks

1 Introduction

Delay Tolerant Networks (DTNs) are characterized by long or variable delays, intermittent connectivity, asymmetric data rates and high error rates. In this type of networks, routing strategies rely on the store-carry-and-forward paradigm, i.e., on the dissemination of messages to intermediate nodes that can retain them, transport them and deliver them either to the final destination or to other intermediate nodes. The complexity of the routing decisions is, for that reason, closely related with the selection of a proper message replication mechanism.

The geocast concept refers to a delivery scheme that filters the eligibility of the network's target nodes using their location as a criterion. With this in mind, the diversity of geocast environments encompassed by DTNs suggests the demarcation of the strategies according to generic properties such as the definition of regions of interest or the replication mechanisms applied.

Moreover, the coupling between the advances in wireless communications technologies and the car industry makes Vehicle-to-Vehicle connectivity an enabler for a

Published by Springer Nature Switzerland AG 2019
L. M. Camarinha-Matos et al. (Eds.): DoCEIS 2019, IFIP AICT 553, pp. 296–304, 2019.
https://doi.org/10.1007/978-3-030-17771-3_26

broad range of applications that can be grouped into two categories: safety-related and commercial-related. Therefore, in light of the advances mentioned, this communication paradigm has received a lot of research attention and has motivated the design of routing strategies in Vehicular Delay Tolerant Networks (VDTNs) that can properly handle the following challenges of such distinguishable environment.

In this sense, the design of strategies in VDTNs that can efficiently deliver messages to a group of recipients that would particularly benefit from the reception of such messages has become a widespread topic of discussion. To that extent, the implementation of algorithms that exploit information from the vehicles' navigation system to route messages within specific Regions of Interest (ROIs) motivates the development of the V-GRADIENT protocol.

2 Contribution to Industrial and Service Systems

The proposed V-GRADIENT geocast routing protocol is the basis for a message dissemination service within a ROI in vehicular highly dynamic or sparse networks. Applications include Vehicle-to-Vehicle (V2V) commercial of safety-related message dissemination. Examples are disseminating fuel prices, current weather or incident warnings. The protocol provides the following contributions: mechanisms to monitor node density and node's buffer occupancy; mechanisms to control message replication avoiding network congestion and buffer exhaustion; efficient message delivery within the ROI.

3 State of the Art

3.1 Unicast Routing

As described in [1], the Direct Delivery protocol is the simplest way to conceive a unicast delivery scheme. The node that carries the message forwards it if and only if it establishes a direct contact opportunity with the target node.

The Spray and Wait protocol [2], in its turn, is based on the dissemination of a limited set of replicas of a given message to distinct contacted nodes during a first phase called spray phase. Afterwards, during the wait phase, relay nodes assume the responsibility of delivering the message to the final recipient directly. However, with regard to the spray phase, several heuristics can be envisioned to distribute properly the messages.

Finally, the Epidemic [3] protocol is classified as a flooding-based scheme, since the idea behind it is based on the provision of unlimited message replication throughout the network by exchanging all possible messages whenever contact opportunities arise.

3.2 Geocast Routing

On the one hand, protocols such as the GSAF [4] and the Geoopp [5] focus on distributing messages by all nodes located inside a region set remotely, relatively to the position of the nodes that generate such messages. This implies, however, that these algorithms break down the routing process into the two following phases: (1) forwarding (and carrying) the messages to the destination region; (2) delivering the message to all nodes inside the ROI. While, regarding the latter phase, both protocols mentioned follow an epidemic-based behaviour, their strategy differs with respect to the first phase. For instance, when using the GSAF protocol a spraying procedure is applied in a first phase. Messages are assigned a predetermined number of tickets T that denote the number of times a given message can be forwarded to encountered nodes outside the destination region. Thus, each time a message is replicated, the T field is sequentially decreased on both messages until T equals zero.

On the other hand, geocast protocols can be applied to distribute content in ROIs whose center is defined by the position of the nodes responsible for creating the respective messages at the generation instant. In this sense, the authors of the Floating Content algorithm [6] include in the messages' headers not only the center coordinates and radius of the circularly shaped ROI but also an additional radius parameter to determine the extension of a buffer zone. The buffer zone is defined as an extension of the ROI where carrier nodes may keep the messages storage beyond the boundaries of the ROI.

4 Geocast Strategies Comparison

This work focuses on distributing messages, during a specified Time To Live (TTL), by all nodes that come across an invariant circularly shaped ROI defined around the node which creates the messages. Moreover, the development of the intended message dissemination algorithm is performed by grasping which features of the considered schemes yield better results under a broad variety of simulations.

In this study, in order to drive such conclusions, the following routing strategies are taken into account, which result from the extension of the definitions regarding the aforementioned unicast routing protocols to a geocast context: Multiple Copy GeoDirectDelivery (MC-GeoDirectDelivery), GeoSpray-and-Wait (GeoS&W), GeoEpidemic and GeoEpidemic Constrained.

Firstly, when using the MC-GeoDirectDelivery protocol, on the one hand, messages are delivered to and kept by nodes residing inside the respective ROI at the delivery instant. On the other hand, messages are retained when nodes leave the ROI, as they may return to the ROI in the future. Because messages are only dropped when their TTL expires or the nodes' buffers are filled up, all nodes that receive a message also become responsible to deliver it to the remaining recipients, enabling the presence of multiple copies of the same message in the network.

Secondly, the GeoS&W scheme couples the principles of the Spray and Wait and the GSAF protocols. In this sense, during the spray phase, messages are replicated, even when nodes are located outside the ROI, using the spraying heuristic adopted by the GSAF protocol. Regarding the wait phase, the GeoS&W behaves as the MC-GeoDirectDelivery protocol.

The Epidemic Protocol nature may assist message distribution in a geocast context as it already distributes an unbounded number of copies of the same messages by multiple receivers. However, to distinguish the protocol's operation in unicast scenarios from its operation in a geocast routing paradigm, this protocol will be referred as the GeoEpidemic protocol throughout the article. Likewise, the GeoEpidemic Constrained scheme follows the Epidemic protocol's principles as well. However, in this case, the routing scheme under consideration constrains the dissemination of messages within the boundaries of the ROI, since messages are dropped as soon as the nodes carrying them leave this region.

Additionally, in this line of reasoning, it is also crucial to make explicit the following performance assessment metrics utilized: delivery ratio and delivery latency. The delivery ratio is computed using the formulation developed in [4]. By monitoring the recipients that reside inside a given message's ROI during its TTL, it is possible to compute a ratio between the number of successful receivers and total number of nodes belonging to the list of recipients. Obviously, this ratio only translates itself into a per message result. For that matter, to obtain an overall delivery ratio estimation, the per message metric is extended to all created messages through an average operation. Latency is also monitored, in a first instance, taking into account a single message. However, in this case, the per-message metric is computed by summing the amount of time it takes for the message under consideration to reach each one of its successful receivers and, then, by dividing this outcome by the same number of recipients.

5 V-GRADIENT Design

The major points of improvement attained by benchmarking the different geocast strategies relate not only with a moderated replication process applied outside the messages' regions of interest but also with a dropping policy that can dynamically estimate if it is beneficial to keep messages being carried by nodes located outside the ROI. With this in mind, on the one hand, with regard to the replication process, the V-GRADIENT algorithm incorporates the principles of the Spray-and-Wait protocol [2] in a geocast context and, during the spray phase, the methodology adopted by the GSAF protocol [4]. On the other hand, the deployed strategy follows the idea behind the Floating Content Algorithm [6], i.e., defines a buffer zone whose range varies according to the network conditions.

For such purposes, to select dynamically the buffer zone range, the V-GRADIENT algorithm takes advantage of estimations of the network density level and of the buffer occupation level, updated periodically. Because in this study it is assumed that nodes do not have access to any knowledge oracles, vehicles rely only on information exchanged during contact opportunities to perceive the surrounding network conditions and, for that reason, both awareness metrics are kept as dictionaries in the nodes' buffers in a decentralized manner. Moreover, the mathematical formulation used to compute both metrics follow an Exponentially Weighted Moving Average (EWMA). For the density level metric ρ_t, the EWMA is performed every 30 s taking into account the number of distinct contacts n_c established, as shown in Eq. 1. α was set to 0.25.

$$\rho_t = \alpha \cdot n_c + (1 - \alpha) \cdot \rho_{t-1}. \tag{1}$$

The buffer occupation metric β_t is determined by averaging the fractions of the buffer occupancy β_{level}, varying from zero to one, retrieved from neighbouring nodes in 30 s intervals with an EWMA, as shown in Eq. 2, where γ was set to 0.7.

$$\beta_t = \gamma \frac{\sum^{n_c} \beta_{level}}{n_c} + (1 - \gamma)\beta_{t-1}. \tag{2}$$

In order to carry out a reliable implementation of the mechanism responsible for adjusting the extent of the buffer zone, a wide variety of simulations were executed using the ONE simulator [7] under different density scenarios, considering ROIs with different radius values, and in environments that led to serious buffer congestion issues, as shown in Table 1. Accordingly, in each simulation an additional dropping policy was included that periodically removed messages from storage if the nodes were located outside of the buffer zone range, defined as the product between the ROI radius, R_{ROI}, value and a threshold R_{+ratio}.

Table 1. Assigned parameters during the simulation procedure.

Map	Downtown part of the city of Helsinki [4500 m × 3400 m]
Simulation time	12 h
Movement model	Shortest path map-based movement
Type of nodes	Vehicles (cars)
Nodes' speed interval	10–50 km/h
Nodes' buffer size	5 MB
Nodes' wait time	5–30 s
Message size interval	500 kB–1 MB
Message generation interval	30–90 s
Interfaces' data rate	250 kbps
Interfaces' transmission range	50 m
Initial number of copy tickets	3
Message's TTL	150 min
Number of nodes	[50;100;200;400] nodes
ROI radius	[250;500;750;1000] meters
R_{+ratio}	Vector of 100 evenly spaced points between 1 and 10

Furthermore, to exhibit the impact of the network density conditions and the ROIs' radius values on the additional dropping policy formulation, Table 2 includes the optimal R_{+ratio} values, bearing in mind the delivery ratio results obtained. It can also be seen in the same table that, for a fixed number of nodes deployed in the network, the optimal R_{+ratio} expression can be modeled by an equation obtained through a power regression. The profile of such equations, since it also follows a trend depending on the number of nodes and consequently on the density level, also suggests a mathematical formulation of a variable threshold value associated with the extension of the buffer zone as shown in Eq. 3.

$$R_{+ratio}(R_{ROI}, \rho_t) = g(\rho_t) \cdot R_{ROI}^{h(\rho_t)}. \tag{3}$$

With respect to the formulation of the $g(\rho_t)$ and the $h(\rho_t)$ expressions, by selecting their corresponding target parameters presented in Table 2, it can be observed, as depicted in Fig. 1, that both describe a well modeled decay relatively to the average density level ρ_{avg} measured in simulations deploying a different number of nodes.

Table 2. Optimal R_{+ratio} values obtained for simulations considering different density conditions and distinct radius values defining the ROI.

	50 Nodes	100 Nodes	200 Nodes	400 Nodes
	$\rho_{avg} = 0.3952$	$\rho_{avg} = 0.8329$	$\rho_{avg} = 1.6470$	$\rho_{avg} = 3.2884$
R_{ROI} = 200 m	9.5	8.6	8.1	7.9
R_{ROI} = 400 m	3.4	3	2.8	2.5
R_{ROI} = 600 m	2	1.7	1.5	1.4
R_{ROI} = 800 m	1.5	1.2	1.1	1.1
R_{ROI} = 1000 m	1	1	1	1
$\hat{R}_{+ratio}$	$12872 \cdot R_{ROI}^{-1.366}$	$10970 \cdot R_{ROI}^{-1.360}$	$9460.3 \cdot R_{ROI}^{-1.348}$	$7649.4 \cdot R_{ROI}^{-1.320}$
R^2 score	0.9964	0.9932	0.9827	0.9706

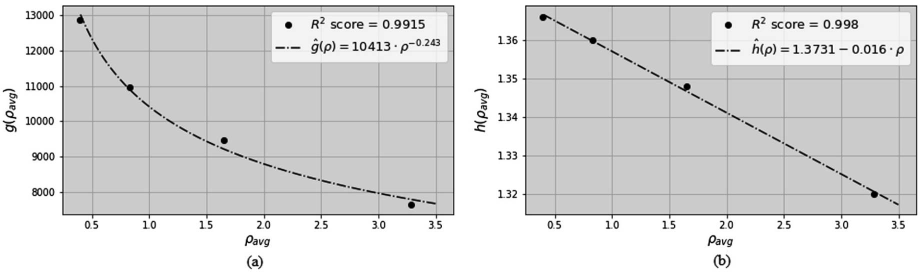


Fig. 1. Regression-based model for the $g(\rho_t)$ and the $h(\rho_t)$ expressions.

In this sense, in scenarios characterized by extreme buffer congestion issues, the V-GRADIENT algorithm allows a dynamic adjustment of the range of the buffer zone for all messages carried by a particular node according to the estimated threshold expression presented in Eq. 4, as a result from the best fit from Fig. 1. Additionally, the buffer occupation level metric plays an important role on detecting the occurrence of congestion, providing nimbleness to the dropping policy operation. If we let ourselves picture a scenario where only a few messages are generated per hour, it can be concluded that, because no congestion issues arise, the drop policy should be less restrictive. For this reason, a logistic function $f(\beta_t)$ responsible for regulating the stringency of the policy, with $\beta_t = 0.5$ defining the sigmoid's midpoint, is included in the final definition of R_{+ratio}.

$$R_{+ratio}(R_{ROI}, \rho_t, \beta_t) = \frac{1}{f(\beta_t)} 10413 \cdot \rho_t^{-0.243} \cdot R_{ROI}^{1.3731-0.016 \cdot \rho_t}. \quad (4)$$

6 Evaluation and Results

In this section, the results regarding the different strategies described in Sect. 3 are compared against the ones concerning the V-GRADIENT algorithm, considering a simulation scenario following most of the parameters' configuration presented in Table 1. However, in this case, different message generators are deployed in each simulation simultaneously. Additionally, the nodes' buffer size is doubled to 10 MB and four different seed values are used for the movement models' pseudo random number generator to reinforce the statistical significance of the results obtained. The corresponding 95% confidence intervals are presented in the following graphs.

Moreover, to assess the performance of the V-GRADIENT strategy under different density conditions, results are presented in parallel according to the number of nodes deployed in the network. Regarding the delivery ratio assessment, as depicted in Fig. 2, the V-GRADIENT algorithm yields better results than the remaining protocols. In fact, for the simulations executed with 150 nodes deployed, the V-GRADIENT algorithm shows a 13% increase on the delivery rate relatively to the GeoS&W protocol, 15% relative to the MC-GeoDirectDelivery, 39% relative to the GeoEpidemic Constrained and 99% relative to the GeoEpidemic. Furthermore, the remaining protocols' outcome also reinforce the choices related with the V-GRADIENT design. For instance, the pronounced upward trend shown by the GeoEpidemic Constrained protocol's results, as the density level increases, validates the exercise of a dropping policy that renders itself stricter in high density conditions.

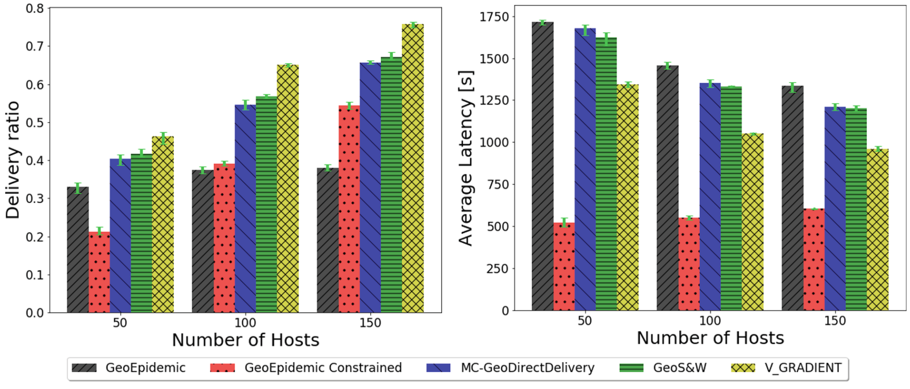


Fig. 2. Delivery ratio and delivery latency as a function of node density and protocol.

When analyzing latency results, one should also bear in mind the delivery ratio results to draw respectable conclusions, since only the delivered messages are taken into account when it comes to measuring latency. Accordingly, in spite of the GeoEpidemic Constrained achieving better results, it can be assumed that, because messages are deleted from the nodes' buffers as soon the carrier nodes leave the ROI, they only live a fraction of their total assigned TTL, which compromises the delivery ratio efficiency. In this sense, the V-GRADIENT provides a moderate and adequate solution, putting a limit on latency, by means of a dynamic dropping policy, while still achieving good performance regarding the delivery ratio: 20% better than GeoS&W.

7 Conclusions and Future Work

We proposed the new V-GRADIENT geocast routing protocol for VDTNs that monitors node density and buffer occupation to disseminate efficiently messages within a ROI centered in the source node. The protocol achieves a better delivery ratio (13–99% with 150 nodes) and lower latency than the existing protocols by controlling network congestion. As future work, we intend to extend the operation of the V-GRADIENT protocol to an environment encompassing distinct geocast groups associated with several types of messages' applications, taking into account the penetration rate of such groups in the network. We also intend to test other scenarios and evaluate the protocol overhead.

Acknowledgments. This work was supported by national funds through Fundação para a Ciência e a Tecnologia (FCT) with reference UID/CEC/50021/2019.

References

1. Spyropoulos, T., Psounis, K., Raghavendra, C.S.: Single-copy routing in intermittently connected mobile networks. In: First Annual IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks (SECON), pp. 235–244. IEEE, October 2004
2. Spyropoulos, T., Psounis, K., Raghavendra, C.S.: Spray and wait: an efficient routing scheme for intermittently connected mobile networks. In: ACM SIGCOMM Workshop on Delay-Tolerant Networking (WDTN), pp. 252–259. ACM, New York (2005)
3. Vahdat, A., Becker, D.: Epidemic routing for partially-connected Ad Hoc networks. Technical report, CS-200006. Department of Computer Science, Duke University, Durham, NC (2000)
4. Rajaei, A., Chalmers, D., Wakeman, I., Parisis, G.: GSAF: efficient and flexible geocasting for opportunistic networks. In: IEEE 17th International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM), pp. 1–9. IEEE, June 2016
5. Lu, S., Liu, Y.: Geoopp: geocasting for opportunistic networks. In: IEEE Wireless Communications and Networking Conference (WCNC), pp. 2582–2587. IEEE (2014)

6. Ott, J., Hyytiä, E., Lassila, P.E., Vaegs, T., Kangasharju, J.: Floating content: information sharing in urban areas. In: IEEE International Conference on Pervasive Computing and Communications (PerCom), pp. 136–146. IEEE (2011)
7. Keränen, A., Ott, J., Kärkkäinen, T.: The ONE simulator for DTN protocol evaluation. In: SIMUTools 2009: Proceedings of the 2nd International Conference on Simulation Tools and Techniques. ICST, New York (2009)

Simulation of an Early Warning Fire System

Paulo Lourenço[2,3(✉)], Alessandro Fantoni[1,3], and Manuela Vieira[1,3]

[1] ISEL - Instituto Superior de Engenharia de Lisboa,
Instituto Politécnico de Lisboa, Rua Conselheiro Emídio Navarro, 1,
1959-007 Lisbon, Portugal
[2] Departamento de Engenharia Eletrotécnica,
Faculdade de Ciências e Tecnologia, FCT, Universidade Nova de Lisboa,
Campus da Caparica, 2829-516 Caparica, Portugal
pj.lourenco@campus.fct.unl.pt
[3] CTS-UNINOVA, Departamento de Engenharia Eletrotécnica,
Faculdade de Ciências e Tecnologia, FCT, Universidade Nova de Lisboa,
Campus da Caparica, 2829-516 Caparica, Portugal

Abstract. In this paper, we will be using separate software tools (wireless network and Finite Differences Time Domain based simulators) to simulate the implementation of a wireless sensor network model based on low-rate/power transmission technology. The system operates in an unlicensed frequency range and the sensing nodes rely on surface plasmon resonance phenomenon for the detection of combustion by-products. More specifically, our simulations contemplate a system for early detection of fire in densely forested areas, which will then issue a warning in an automated way. As late detection of these events usually leads to severe flora, terrain, wild life and societal impact, an early warning system will provide better event assessment conditions, thus enabling efficient resources allocation, adequate response and would certainly be a promising improvement in minimizing such disruptive impairments.

Keywords: FDTD simulations · MMI coupler · Surface plasmon resonance · WSN · Fire detection

1 Introduction

Contemporaneous technologies on wireless mobile communication, embedded systems, cloud computing and sensors, have contributed for many examples of successful applications where intelligent systems have been employed. Implementations of such systems have occurred in many sectors of our economy, from public security systems to logistics platforms and many other, and it is predictable a worldwide proliferation of these devices/systems in many aspects of our daily lives, providing ubiquitous and seamless access to a smart world. This transition will require the development of an enormous quantity of such devices, all with energy and manufacturing requirements, hence all efforts must be made to assure cost-effective and highly efficient approaches at both their production and usage/operation cycles [1].

The paradigm associated to a smart world, the internet of everything (IoE), makes reference to a *network* of interconnected *things*. Here, the *network* is usually

Published by Springer Nature Switzerland AG 2019
L. M. Camarinha-Matos et al. (Eds.): DoCEIS 2019, IFIP AICT 553, pp. 305–317, 2019.
https://doi.org/10.1007/978-3-030-17771-3_27

contemplated as the TCP/IP network we all rely on at present time and the *things* being a myriad of sensors and actuators. All of these devices incorporate telecommunications, storage and processing capabilities, as depicted in Fig. 1, thus being able to provide human to machine and/or machine to machine interactions. They are also able to aggregate themselves as a network of smart nodes which is then bridged with traditional communications networks or the Internet through a gateway, enabling this way remote operation, control and management [2].

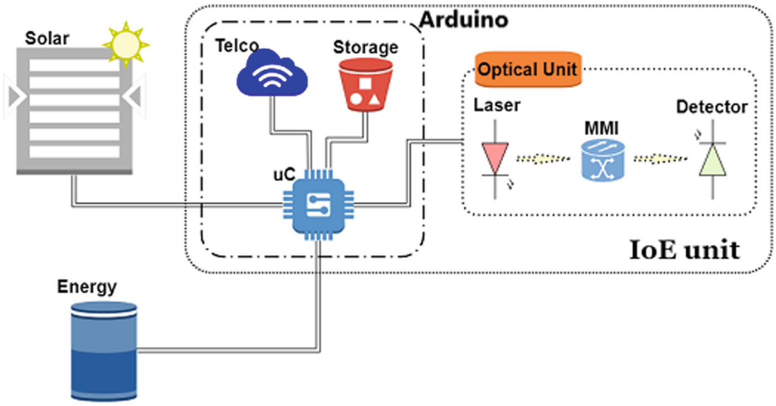


Fig. 1. Proposed block diagram of a single unit.

This document intends to initiate the steps for the development of a highly integrated, efficient and cost effective Information and Communication Technology (ICT) system for early detection of wild fire in forested areas. Towards that end and through available software tools [3, 4], the simulation of the three components of this work have been conducted. Namely:

- The SPR based sensor which is able to sense minute variations of its surrounding environment refractive index. As such, the resulting by-products of burning wood (e.g. soot – carbon atoms agglomerates) will be detected as an increase in the refractive index of the environment in the near proximity of the sensing structure;
- The 3 dB optical coupler that will provide the required optical paths for the detection and reference fields of our device. These EM fields will be analysed, processed and quantified, and the obtained results will be monitored and interpreted;
- The Wireless Sensor Network (WSN) that enables the deployment of these devices over an arbitrary area, allowing control, monitoring and interconnection between all the network nodes, hence providing an early warning system for wild fire in forests, in an automated way.

The remaining of this paper is organized as follows:

– Next section presents how this document relates to the context of Innovation in Industry and Service Systems, which is the main focus of DoCEIS2019 Doctoral Conference;

- Then follows the Surface Plasmon Resonance section, where numerical simulations of the sensing structure of each individual device are presented and the results obtained show the ability to detect refractive index variations of the surrounding environment. The refractive index increase will result from the burning wood smoke carrying, amongst other compounds, unignited and condensed micro-meter sized agglomerates of carbon atoms (soot), which move away from fire and will increase the refractive index of the environment near the sensing structures. This soot production results from unburned carbon atoms that moving away from the top of the flame and is still an active field of research [5].
- Follows Multimode Interference (MMI) section, where will be described a 1×2 MMI structure consisting on a 3 dB coupler that provides the required optical paths to feed the sensing device and to yield the reference arm signal for further analysis and processing.
- Afterwards comes the Wireless Sensor Network (WSN) section. In this section will be presented a software tool [6] that assisted on the deployment of the nodes on the map location, the network structure organization, the communication between nodes, the communication with the external network (i.e. TCP/IP) and the whole network behaviour simulation in a fire event.
- Finally, there is the Conclusions section where obtained results are gathered, analysed and conclusions are drawn and reported. Here, future areas of related research will also be discussed. Namely, the experimental implementation of simulated device and the optimization and development of software functions in the WSN simulator tool.

2 Contribution to Innovation in Industry and Service Systems

ICT systems have been playing an important and increasing role for several years, as far as reliable control of automated tasks is concerned and in many sectors of contemporaneous society. Even in traditional sectors of our economy, technological proliferation is a fact and the trend seems to indicate an increasing penetration in every aspect of our lives. Industry will be no exception to this tendency and is forecasted to undergo a major transformation already on its way. This is expected to be a profound industrial revolution that will take manufacturing processes automation to a whole new level of efficiency and mass production. Although the debate about how to designate this transformation is still in progress, the term "Industry 4.0" has been agreed upon amongst many western European companies, academics and technological practitioners.

According to [7], the most often cited terms when referring to Industry 4.0 are Cyber-Physical Systems (CPS), Internet of Things (IoT), Smart Factory (SF), and Internet of Services (IoS). CPS aims the integration of physical systems with computation, IoT is mainly about technological systems being able to exchange data while sensing their surroundings, IoS intends to wrap existing and future connected mobile devices as gateways to services provided by the manufacturer and SF aggregates CPS and IoT to seamlessly assist machines and humans to perform their manufacturing tasks.

In this article we would like to point out Innovation also as a key element in Industry 4.0 for all these interacting systems certainly provide a fertile platform in which originality will thrive. To this end, our simulations envisions an automated early warning fire detection system consisting on a WSN with integrated optical devices based on SPR for the detection of burning wood by-products, namely soot. To the best of our knowledge, this fire detection system represents a novel approach to wildfire events prevention in forests and their severe environmental, financial and societal impact.

The automated early warning fire detection system is a clear implementation of real world virtual integration where live data is gathered, exchanged, analysed and decisions are made based on this process, thus entailing the set of recommendations for Industry 4.0 referred in the literature. Moreover, the same working principle of this system might be applied in similar devices but to sense other physical quantities, as long as they can be translated into refractive index variations of the environment in the vicinity of the sensing structure.

3 Surface Plasmon Resonance

Optical sensing devices based on the excitation of plasmonic waves on metallic surfaces are usually referred as Surface Plasmon Resonance (SPR) sensors and their operation relies on the detection of refractive index variations in the electromagnetic (EM) mode supported by the free electrons of the metallic surface. These structures have been extensively reported as valid applications for the detection of gases, chemical and biological elements.

The scientist who first described this phenomenon used a structure that has been named after him, the Kretschmann configuration [8]. This method consists on a quartz prism where an incident EM field is refracted on a facet and reflected at the base where a thin metallic film has been deposited. Under the right conditions, the energy of the incident field is transferred to the free electrons on the metallic surface and triggers their oscillation. These oscillations generate surface plasmon waves that propagate along the metal/dielectric interface.

Surface plasmon waves are highly sensitive to the refractive index of the surrounding environment. This wave is a single transverse magnetic (TM) mode and its magnetic field intensity vector is parallel to the metal/dielectric interface plane and perpendicular to the propagation direction. Its propagation constant can be defined as:

$$\beta_{SP} = k_0 \sqrt{\frac{\varepsilon_m \varepsilon_d}{\varepsilon_m + \varepsilon_d}}. \tag{1}$$

where β_{SP} is the propagation constant of the surface plasmon wave, k_0 is the vacuum wavenumber and ε_m and ε_d are the dielectric permittivity for metal and insulator, respectively. As one can infer through the above equation, this propagation constant is highly sensitive to minor variations on the refractive index of the surrounding environment [9].

Our numerical simulations consisted on a Silicon Carbide (SiC) waveguide deposited over a layer of silica (SiO_2) and where a 50 nm high layer of Aluminium (Al) is deposited. The structure was dimensioned to assure single mode operation at 633 nm wavelength at initial waveguide dimensions, then its cross section is increased to create a bimodal waveguide as shown in Fig. 2. The working principle is, to some extent, similar to the device developed by Zinoviev et al. [10] for the interferometric biosensor.

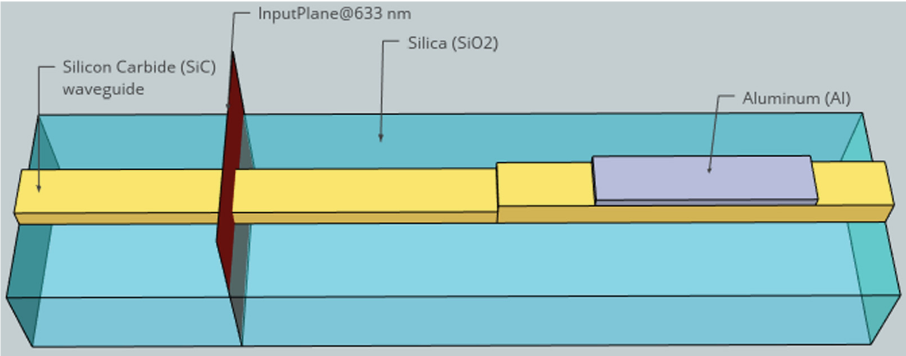


Fig. 2. Diagram of the simulated structure.

The reasons justifying the selection of SiC as the material for the waveguide and Al for metallic layer were its almost non-existent absorption at visible wavelengths for the former and its cost effectiveness for the latter. Moreover, the minute dimensions of the structure enable the integration of all components in one single device, thus fully complying with the lab-on-chip concept [11].

Numerical simulations of this structure based on the Finite Differences Time Domain algorithm were carried out. This 3D structure is an invariant refractive index structure on the third axis and the waveguide is much wider than thick. Thus, one may assume the modal fields having the same behaviour along its width, hence a 2D planar waveguide approach might be considered [12]. Nevertheless, given the high contrast involved and being a rectangular waveguide, the dual effective-index method [13] was the considered approach and the results obtained were successfully compared with the ones provided by analysis for this structure [14].

The equivalent 2D simulation workspace is depicted in Fig. 3, where in **(a)** is depicted a longitudinal cut of simulated structure and in **(b)** one can identify the refractive index of the core waveguide, the two propagation modes (i.e. TM_0 and TM_1) allowed in the wider cross section of this structure, the modes profile and their amplitudes.

Note that the mode amplitude of TM_0 is approximately half the amplitude of TM_1, therefore contribution of the fundamental mode for transmission is minimal. Moreover,

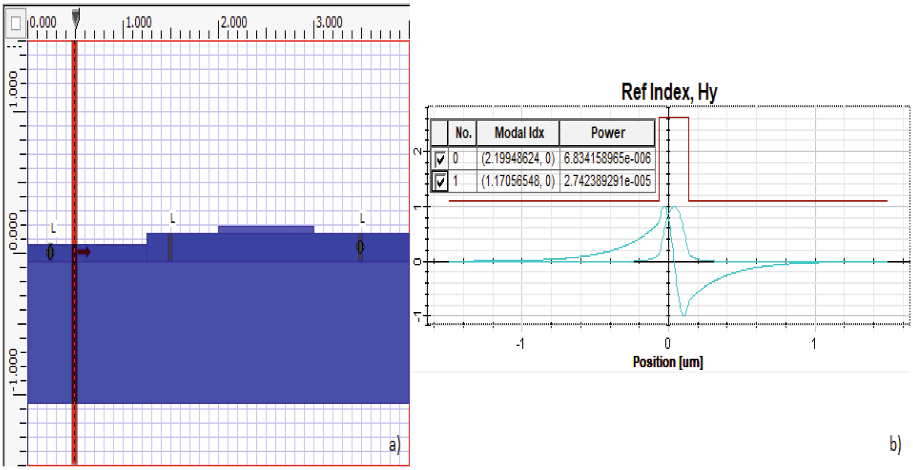


Fig. 3. **(a)** OptiFDTD simulation workspace; **(b)** Refractive index profile, modes propagated and respective power amplitudes.

simulations show that TM_0 is highly dispersed into the SiO_2 substrate and only reminiscent power reaches the detection area. With this in mind, following procedure and discussion considered only the propagation of TM_1 mode.

The Al layer has been simulated considering the Lorentz-Drude dispersive model from OptiFDTD [3] library, thus providing the closest possible conditions to experimental environment. Figure 4**(a)** shows a close up of the Poynting vector evolution along the propagation direction. It is noticeable the propagation of the SPR wave along the Al layer. After the sensing area, the field amplitude decays approximately 6 dB, as can be verified in **(b)** where are shown the obtained field profiles on the observation lines located before and after this zone.

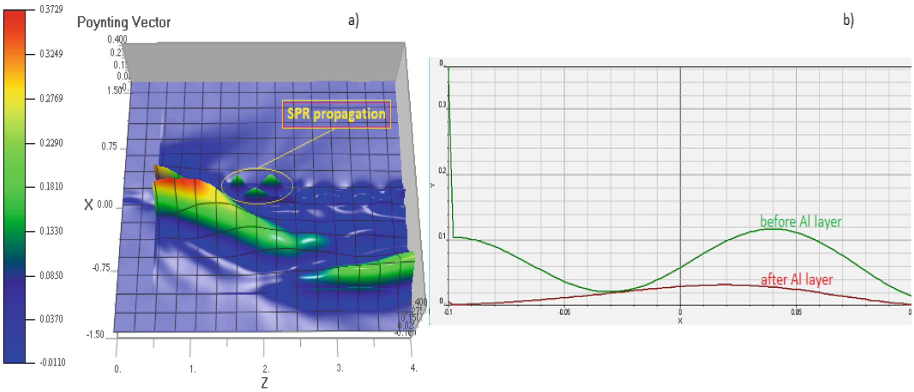


Fig. 4. **(a)** Poynting vector amplitude evolution; **(b)** Field profiles before and after detecting zone.

This sensing structure must be able to detect slight refractive index variations in its close vicinity. To evaluate its detection capabilities, a similar conditions simulation took place only this time with an increase of 0.1 in the refractive index of the surrounding environment. The results obtained are presented in Fig. 5, where are shown the transmission spectral response for both default and increased refractive indexes and has been verified a red shift of ~7.4 nm for a variation of 0.1 refractive index units.

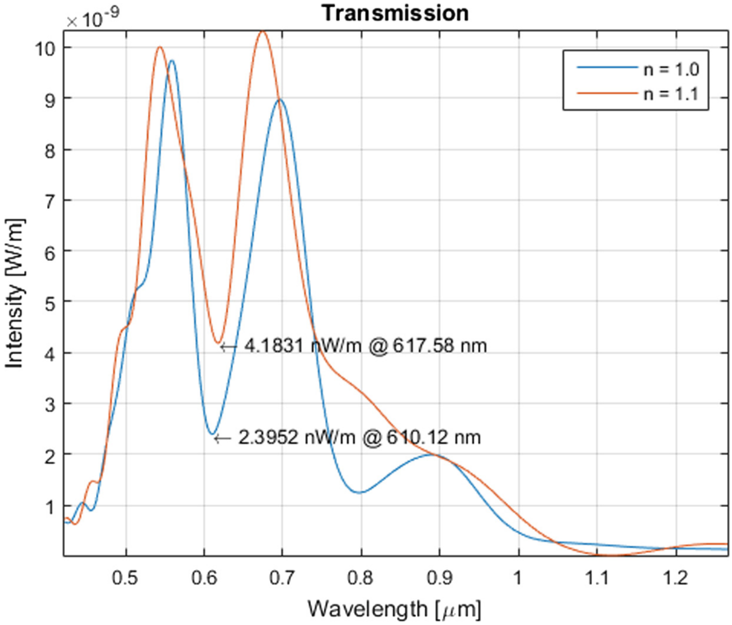


Fig. 5. Transmission spectral response for both refractive indexes.

Burning wood produces large quantities of smoke carrying droplets of soot that is formed by unburned and condensed micro-meter sized agglomerates of carbon atoms [5]. Carbon refractive index is 2.4105 at the operating wavelength of 633 nm [15] thus, production of these carbon agglomerates (soot) that result from the combustion of wood will contribute for the refractive index increase in the surrounding environment of the detection unit.

4 Multimode Interference

Our intent was to obtain two identical EM field profiles resulting from an initial fundamental mode input profile. One of the EM fields will be affected by a sensing area and the other one will represent a comparison reference for further processing.

The available options evaluated were single mode Y-junctions, two-mode interference (TMI) and MMI devices, but they all presented disadvantages when compared to the latter structure. Namely, close proximity of access waveguides on these devices usually leads to unwanted modal coupling and separation aperture filling, due to the limited resolution of lithographic process, hence affecting TMI section length arbitrarily and causing performance degradation [16]. According to [17–19], operation of MMI devices relies on the self-imaging principle which states that single or multiple images of a given input field profile are replicated periodically in space, as the electromagnetic field propagates through the waveguide. The propagation constant of a mode β_m $(m = 0, 1, 2, \ldots)$ propagating in a high contrast step index multimode device shows an approximate quadratic dependence to the mode number m:

$$\beta_m \cong k_0 n_{core} - \frac{(m+1)^2 \pi \lambda_0}{4 n_{core} W_{eff}^2}, \tag{2}$$

where k_0 is the vacuum wavenumber, n_{core} the refractive index of the structure, λ_0 the vacuum wavelength and W_{eff} the effective width of the MMI waveguide (considering the Goos-Hähnchen shift on the device boundaries). In high refractive index contrast devices, the penetration depth of the EM field beyond the inner walls of the device is practically non-existent, hence:

$$W_{eff} \cong W \tag{3}$$

The spatial location of single/multiple and direct/mirrored images, resulting from the propagation modes interference, is directly related to the beat length of the two lowest order modes, L_π:

$$L_\pi = \frac{\pi}{\beta_0 - \beta_1} \cong \frac{4 n_{core} W_{eff}^2}{3\lambda_0} \tag{4}$$

Single mirrored and direct images from the input field profile form at $3L_\pi$ and $2(3L_\pi)$, respectively, while two-fold images form at $\frac{1}{2}(3L_\pi)$ and $\frac{3}{2}(3L_\pi)$. The single images are approximately the same amplitude as the input EM field and each of the two-fold images is affected by an attenuation factor of 3 dB, thus offering the ideal conditions for a 3 dB coupler device, similar to the structure diagram depicted in Fig. 6.

The simulated device consisted on a silicon carbide (SiC) based structure, where the input and output waveguides were dimensioned to support single mode operation at 633 nm vacuum wavelength. First simulations of our structure did not show the intended results, mainly due to fine misplacement of the output waveguides, thus an optimization algorithm had to be executed on the simulation tool.

The results obtained are shown in Fig. 7, where one is able to find the optical field profile spatial distribution on the top half of the figure and, at the bottom half, the normalised amplitudes along the monitored paths leading to the output waveguides. Note that, at the latter, there are two evolving lines that reflect each of the already mentioned paths. These are hardly distinguishable due to the optimization algorithm that balanced the amplitudes of both output waveguides.

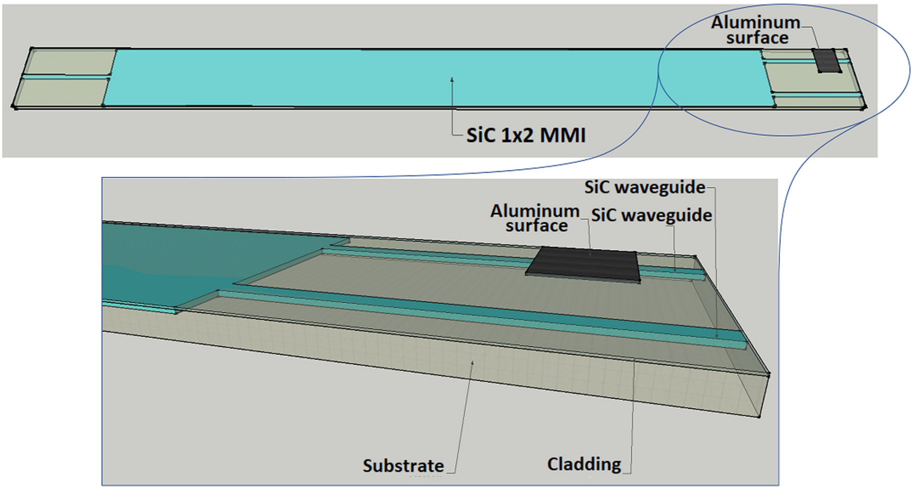


Fig. 6. MMI structure and sensing area diagram.

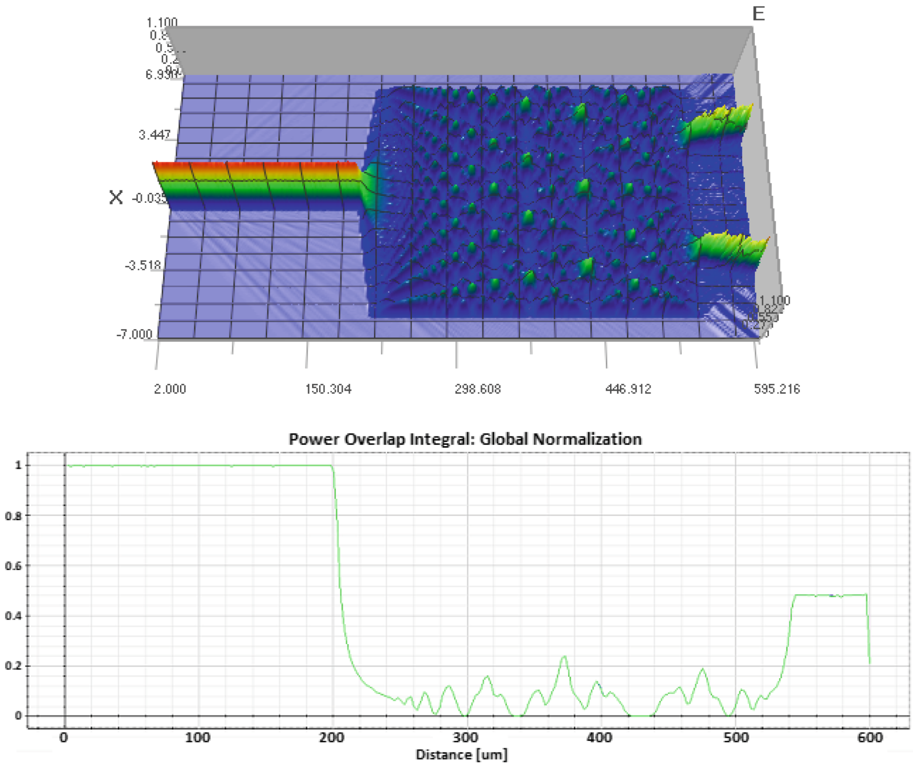


Fig. 7. Optical field profile distribution (top) and field amplitudes along the monitoring paths (bottom).

5 Wireless Sensor Network

Contemporaneous state-of-the-art technology in telecommunications and electronics enables the development of a myriad of sensor structures, each with capabilities of sensing specific events in its proximity, communicating with its neighbours, data storage and processing power. With this in mind, a large number of these devices deployed over an arbitrary area can collaborate to form an *ad-hoc* network of sensing nodes, which will be referred throughout this document as a Wireless Sensor Network (WSN).

The trend direction seems to point for these WSNs to be present in every aspect of our lives as the predominant sensing technology [2], with unprecedented impact in every sector of our economy and in society. Implementation and deployment of a WSN is not burden free and its cost is directly related to the number of nodes in the network. Hence, it is of vital importance to define an implementation strategy prior to deployment. To this end, network simulators are of great assistance for they enable the simulation and refinement of the network architecture, operation and behaviour of our design before going into the implementation phase.

To elaborate this document, several network simulators were evaluated, namely Packet Tracer [20], NS-2 [21] and CupCarbon [4]. After consulting the available documentation for each tool and careful consideration, the selected simulator was the latter one for its appropriateness over our intended task: - WSN simulation for early fire detection. CupCarbon [22] is a simulator that operates at the application layer of the nodes and will be next demonstrated adequate for network behavioural monitoring when in presence of a fire event. It integrates radio propagation and interference models which allows fast and accurate evaluation of link quality based on transmission conditions [23]. Sensor nodes can be placed on an actual geographical map through Open Street Map (OSM) framework.

Being a multi-agent and discrete event simulator it enables monitoring several node related events, such as tracking mobility, energy consumption, gas and fire detection, and more, while simulation is running [24]. Moreover, this software is able to simulate IEEE 802.15.4 (Zigbee) [22], WiFi and LoRA, and each node has data processing capabilities through SenScript commands. It is an open source Java programmed tool, source code is available and may be altered to include custom code, functions or algorithms, e.g. an improved Dijsktra algorithm [25] or data mining techniques for fire detection [26].

Thus, CupCarbon was the selected simulator to model an automated early detection of fire in forests based on a WSN. The selected radio protocol was Zigbee for this is a low-rate, low-traffic network and distance between nodes is not limited by radio but sensing range. To this end, 178 nodes with 50 m sensing range were deployed over a densely forested area, all with data storage and processing capabilities. Nodes distribution was maximized on tree areas over bushes or clear sight terrain. Figure 8 shows the deployed WSN where each orange dot is a sensing node, the orange triangle is the sink and the white arrows symbolize bi-directional links.

A gateway (sink) was placed within reach of some nodes to assure interaction with the external network (TCP/IP), thus allowing remote control, monitoring and

Fig. 8. Deployed WSN.

management of all network elements. Scripts were developed for both nodes and gateway, according to their operational functions:

- Nodes are always checking the radio data buffer for an incoming message to route and, at every given time set in the script, the sensing device is monitored for an alarm condition. They only transmit or forward a message once and wait for the reset message from the gateway;
- Gateway receives messages from nodes and forwards them to the external network. This is signalled by a yellow circle around the triangle. Once an alarm message is received from all neighbours, the sink sends the reset message to nodes. Sink also provides application layer connection for remote control, monitoring and management of all WSN.

Simulation was conducted and, at some point in time, a fire event has been modelled. Figure 9(a) depicts the start of the fire event detected by the sensing structure. Then, the alarm message is sent out the node to its neighbours and starts propagating to the rest of the network. As the message evolves throughout the network, after 61 received messages it reaches the gateway, as represented in Fig. 9(b), and is forwarded

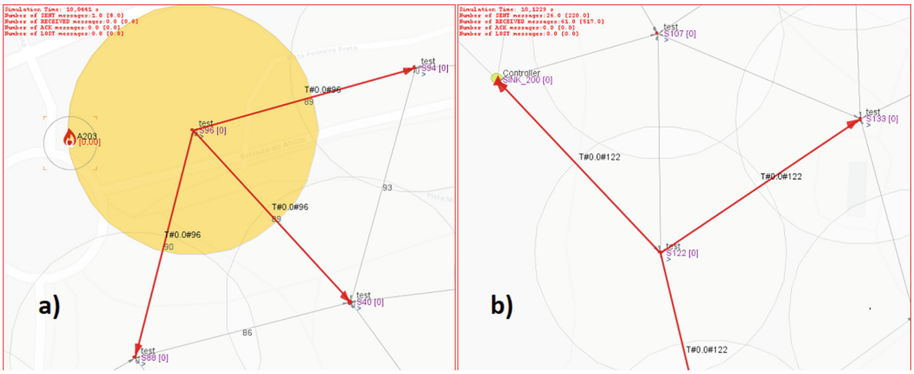


Fig. 9. **(a)** Fire event start and initial alarm message propagation; **(b)** Alarm message propagation reaches gateway. (Color figure online)

through an available technology (e.g. 3G/4G, cable, optical fibre, ADSL, etcetera) to the remote control centre for evaluation.

6 Conclusions

The simulated SPR based sensing device, has been designed on a SOI platform with cost-effective materials such as Al for the active metal and SiC for the waveguide and MMI elements. This structure was able to detect refractive index variations of the surrounding environment at visible wavelengths, thus its features can be incorporated on an experimental sensing device. This structure has been modelled with dispersive materials, including losses, to reflect as much as possible experimental conditions. Device integration of one or many of these structures is achievable given design simplicity and involved minute dimensions, thus complying with lab-on-chip concept [11]. Simulations of a network operating under these conditions have shown that an early wild fire detection system is feasible when implemented on a WSN.

Future work will consist on the experimental implementation (through Plasma Enhanced Chemical Vapour Deposition facilities at ISEL) of the sensing device to validate the results obtained, setting up a comparison benchmark between mentioned results and the ones obtained with noble metals and will also consider development of Java based code in the WSN simulation tool, in order to optimize energy monitoring capabilities in a more consistent way and to develop a fire spreading algorithm.

Acknowledgments. This research was supported by EU funds through the FEDER European Regional Development Fund and by Portuguese national funds by FCT – Fundação para a Ciência e a Tecnologia with projects PTDC/NAN-OPT/31311/2017, SFRH/BPD/102217/2014 and by IPL IDI&CA/2018/aSiPhoto. A special word of recognition to professors Miguel Fernandes and Yuri Vygranenko, whom have contributed with their invaluable expertise.

References

1. Saha, H.N., Mandal, A., Sinha, A.: Recent trends in the Internet of Things. In: 2017 IEEE 7th Annual Computing and Communication Workshop and Conference, CCWC 2017, pp. 15–18 (2017)
2. Kaur, S., Grewal, V.: Wireless sensor networks- recent trends and research issues. Indian J. Sci. Technol. **9**(48), 1–5 (2017)
3. Design software for photonics. https://optiwave.com. Accessed 29 Dec 2018
4. CupCarbon, WSN simulator. http://www.cupcarbon.com/. Accessed 29 Dec 2018
5. Emmons, H.W., Atreya, A.: The science of wood combustion. Proc. Indian Acad. Sci. Sect. C Eng. Sci. **5**, 259–268 (1982)
6. Virtualys, IEMN, IRCICA, Xlim, Lab-STICC: CupCarbon User Guide (2017)
7. Hermann, M., Pentek, T., Otto, B.: Design principles for Industrie 4.0 scenarios: a literature review. Technische Universität Dortmund, no. 01 (2018)
8. Kretschmann, E., Raether, H.: Radiative decay of non radiative surface plasmons excited by light, pp. 0–1, November 1968

9. Homola, J.: Surface plasmon resonance sensors for detection of chemical and biological species. Chem. Rev. **108**, 462–493 (2008)
10. Zinoviev, K.E., González-Guerrero, A.B., Domínguez, C., Lechuga, L.M.: Integrated bimodal waveguide interferometric biosensor for label-free analysis. J. Light. Technol. **29** (13), 1926–1930 (2011)
11. Haeberle, S., Mark, D., Von Stetten, F., Zengerle, R.: Microfluidic platforms for lab-on-a-chip applications. Microsyst. Nanotechnol. **9783642182**(9), 853–895 (2012)
12. Buus, J.: The effective index method and its application to semiconductor lasers. IEEE J. Quant. Electron. **18**(7), 1083–1089 (1982)
13. Chiang, K.S.: Dual effective-index method for the analysis of rectangular dielectric waveguides. Appl. Opt. **25**(13), 2169 (1986)
14. Computational Photonics. https://www.computational-photonics.eu/eims.html. Accessed 29 Dec 2018
15. Refractiveindex.info database (2017). http://refractiveindex.info/. Accessed 29 Dec 2018
16. Soldano, L.B., Veerman, F.B., Smit, M.K., Verbeek, B.H., Dubost, A.H., Pennings, E.C.M.: Planar monomode optical couplers based on multimode interference effects. J. Light. Technol. **10**(12), 1843–1850 (1992)
17. Soldano, L.B., Pennings, E.C.M.: Optical multi-mode interference devices based on self-imaging: principles and applications. J. Light. Technol. **13**(4), 615–627 (1995)
18. Honbun: Chapter 2 Multimode Interference Theory, pp. 15–64 (2006)
19. Sosa, A.: Design of silicon photonic multimode interference couplers. Opt. Express **19**(2), 88 (2012)
20. Cisco Packet Tracer. https://www.netacad.com/courses/packet-tracer-download/?p_auth=h6jOI2lZ&p_p_auth=iJXrHgDA&p_p_id=resendscreenname_WAR. Accessed 29 Dec 2018
21. NS-2. https://www.isi.edu/nsnam/ns/. Accessed 29 Dec 2018
22. Bounceur, A.: CupCarbon: a new platform for designing and simulating Smart-City and IoT Wireless Sensor Networks (SCI-WSN). In: Proceedings of the International Conference on Internet Things Cloud Computing, p. 1:1, March 2016
23. Bounceur, A., et al.: CupCarbon: a new platform for the design, simulation and 2D/3D visualization of radio propagation and interferences in IoT networks. In: CCNC 2018 - 2018 15th IEEE Annual Consumer Communications & Networking Conference, April, pp. 1–4, January 2018
24. Mehdi, K., Lounis, M., Bounceur, A., Kechadi, T.: CupCarbon: a multi-agent and discrete event wireless sensor network design and simulation tool. In: Proceedings of the Seventh International Conference on Simulation Tools and Techniques (2014)
25. Lopez-Pavon, C., Sendra, S., Valenzuela-Valdes, J.F.: Evaluation of CupCarbon network simulator for wireless sensor networks. Netw. Protoc. Algorithms **10**(2), 1 (2018)
26. Saoudi, M., Bounceur, A., Euler, R., Kechadi, T.: Data mining techniques applied to wireless sensor networks for early forest fire detection. In: Proceedings of the International Conference on Internet of Things and Cloud Computing - ICC 2016, no. Dm, pp. 1–7 (2016)

Energy Markets

Integration of Renewable Energy in Markets: Analysis of Key European and American Electricity Markets

Hugo Algarvio[1,2(✉)], Fernando Lopes[1], and João Santana[2,3]

[1] LNEG−National Laboratory of Energy and Geology, Estrada do Paço do Lumiar 22, 1649-038 Lisbon, Portugal
fernando.lopes@lneg.pt

[2] Instituto Superior Técnico, Avenida Rovisco Pais 1, 1049-001 Lisbon, Portugal
{hugo.algarvio,jsantana}@tecnico.ulisboa.pt

[3] INESC-ID, Rua Alves Redol, 9, 1000-029 Lisbon, Portugal

Abstract. Electricity markets are systems for affecting the purchase and sale of energy. Most existing markets are built on well-established principles of competition and transparency. However, their designs are based on centralized power plants with a small participation of end-use customers. During the past years, the share of electricity produced by renewable sources increased significantly. This paper analyses the structure and operation of two European markets and two American markets. The analysis highlights that the design, rules and characteristics of most markets are still not completely adapted to power systems with high levels of variable renewable energy. Accordingly, the paper proposes some recommendations to foster the integration of renewable generation.

Keywords: Energy markets · Market operation · Renewable generation · Market design

1 Introduction

In the world, power systems have evolved from traditional vertically monopoly structures to liberalized markets that encourage competition between suppliers and consumers of electricity. Deregulation led to a completely different industry with novel competitive wholesale and retail markets based on open-access rules (see, e.g., [1]).

Most existing energy markets are built on well-established principles of competition and transparency. In particular, the design of European markets is founded on the measures of the Third Energy Package, approved in 2009 [2].

Broadly speaking, the Third Energy Package was very important for market participants. Yet, its rules are based on fuel-based power plants—that is, the predominant generation technologies of the last decade. The electricity system of Europe changed

This work was supported by "Fundação para a Ciência e Tecnologia" with references UID/CEC/50021/2019 and PD/BD/105863/2014 (H. Algarvio).

The original version of this chapter was revised: The year mentioned in the funding information has been corrected. The correction to this chapter is available at https://doi.org/10.1007/978-3-030-17771-3_37

Published by Springer Nature Switzerland AG 2019
L. M. Camarinha-Matos et al. (Eds.): DoCEIS 2019, IFIP AICT 553, pp. 321–328, 2019.
https://doi.org/10.1007/978-3-030-17771-3_28

profoundly since 2009. The share of electricity produced by variable renewable sources is expected to grow up to 27% in 2030. Accordingly, the Clean Energy for all Europeans adopted at the end of 2016 introduces further modifications in the energy market design [3].

Against this background, this paper analyses two key European markets: Nord Pool [4] and MIBEL [5]. Nord Pool is the Nordic electricity exchange, which includes three different markets: Elspot, Elbas and Eltermin. Elspot and Elbas are markets for the physical trade of energy, while Eltermin is a market for financial trade of energy. The Iberian Electricity Market (MIBEL) is the result of the cooperation between the Portuguese and the Spanish Governments with the aim of promoting the integration of the electrical systems of both countries.

The paper also analyses two key American markets: NYISO [6] and PJM [7]. NYISO (the New York Independent System Operator) coordinates the markets from the heart of Manhattan to the remote stretches of the Adirondack wilderness. PJM (the system operator of Pennsylvania, New-Jersey and Maryland) coordinates the markets from Delaware and Illinois to West Virginia and the District of Columbia.

All of these markets are unique and quite complex. Also, they were initially designed without taking into account the increasing levels of variable renewable energy (see, e.g., [8, 9]). Accordingly, this paper analyzes several key features of these markets to determine, at least to a certain extent, if their rules are appropriate to the new market realities.

The remainder of the paper is structured as follows. Section 2 describes the structure and operation of real-world energy markets generally. Section 3 describes in detail the four aforementioned energy markets. Section 4 analyses the key features of the four markets. Finally, Sect. 5 presents some concluding remarks and describes avenues for future work.

2 Relationship to Industrial and Service Systems

As noted, this work analyses four real-world markets in terms of structure and operation. Accordingly, the relationship of the work with industrial and service systems is clear and natural. It is also worth mentioning that the participants in wholesale energy markets are typically industrial companies.

European Markets. In terms of structure, European markets usually include day-ahead markets, intra-day markets and bilateral markets (see, e.g., [10]). Day-ahead markets (DAMs) are markets based on auctions. Energy is traded at 12 noon of the day before the day of operation. The pricing method is based on the marginal pricing theory. Intra-day markets may be based on auctions or operate continuously. These markets allow the parties to refine their positions a few hours ahead of real-time.

Bilateral markets allow to hedge, at least in part, against the risk inherent to day-ahead and intra-day markets. Market participants may negotiate the price (and "adjust" the quantity) of various types of standardized contracts, notably forward contracts, future contracts, options and contracts for difference [10].

In addition to the aforementioned markets, the so-called "balancing markets" allow the participation of specific market participants at short notice. Such markets represent the key mechanism for fixing the imbalances relative to the physical settlement on both day-ahead and intra-day markets. Typically, the transmission system operator (TSO), a non-commercial organization, independent of commercial players, constitutes the demand side. Balance responsible parties (BRPs) represent the supply side and sign a pre-agreement on balance responsibility with the TSO. BRPs with available generation capacity submit up-regulation orders to the TSO. Likewise, BRPs able to reduce consumption submit down-regulation orders to the TSO. All regulating power orders are ranked with increasing price. In case the TSO is procuring up regulation, the up-regulation orders with lowest prices are activated until the procured quantity is reached. The price of the last up-regulated order sets the up-regulation price. The orders with prices below the up-regulation price have a profit, defined by the difference between the final price and the offered price. A similar procedure is used to set the down-regulation price.

Following the definition of up-regulation and down-regulation prices, the TSO performs an imbalance settlement process. Basically, the participants that cause system imbalances should contribute to the payment of the costs for re-establishing the balance of the system. In this way, players should ensure that they trade electrical energy according to specific plans. Deviations from plans add up to the net imbalances and typically involve the payment of penalties.

American Markets. Most American markets include a DAM and a short term market, generally referred to a real-time market (RTM). This design reflects a two-settlement system. The DAM defines prices and schedules for an entire day, one day in advance, based on the locational marginal pricing theory. The RTM is a 5-min market where participants can refine their positions a few minutes before real-time. The average prices of the DAM and RTM should converge, to avoid speculation and a preference for any of the two markets (see, e.g., [11]).

3 Key Electricity Markets

3.1 European Markets: Nord Pool and MIBEL

In Nord Pool, the day-ahead market (Elspot) accepts several types of orders, including single hourly orders, block orders, exclusive groups and flexi orders. Customers can use any type of orders or a combination of different types to meet their requirements. The market closes at 12:00 CET and area prices (as well as the system price) are calculated for each delivery hour. It allows aggregated bids. The intra-day market (Elbas) supplements the day-ahead market and provides some flexibility to participants by offering 15-min, 30-min, hourly and hour-blocks products. This market is based on continuous trading, meaning that the orders are matched automatically when concurring offers are registered in the trading platform, typically up to one hour before the delivery hour [12, 13].

Nord Pool includes a balancing market that consists of a regulating power market and a balancing power market. In the regulating power market, TSOs buy/sell power from/to BRPs in the delivery hour taking into account the bids performed for upward and downward regulation. Bids refer to the next day of operation (*d*) and may or may not cover the entire day. Typically, bids are submitted until 17:00 of day $d - 1$ and can be adjusted after the clearing of the intraday market, up to 45-min before the delivery hour. The price of regulating power (RP price) is determined according to the marginal pricing theory and calculated on an hourly basis [14].

On the balancing power market, TSOs buy/sell power from/to market participants in order to balance the system. The settlement basis for balancing power is constituted by BRP notifications submitted to TSOs. Notifications are submitted during day $d - 1$ and may be adjusted during the day of operation (*d*) as a result of intra-day and bilateral trades. The balance settlement is based on the difference between production (and/or consumption) compared to the submitted notifications. Two models for settlement of balancing power are used: the one-price model and the two-price model [14].

The Daily Iberian market (day-ahead market) sets prices for electrical energy every day at 12 noon, by considering EUPHEMIA, an algorithm based on the marginal pricing model adopted by the European Union. Generating companies, retailers and other market participants may trade energy regardless of whether they are in Spain or in Portugal. Bids are accepted according to the merit order, until the interconnection between Portugal and Spain is fully occupied. In such a case, there will be a different price of electricity for Portugal and Spain. Otherwise, the price will be the same for both countries. MIBEL does not allow aggregated bids. The intra-day market runs after the DAM and involves six trading sessions, based on auctions. This market permits agents to readjust their commitments for purchasing and selling energy up to 4 h ahead of real time. As of that moment, there are other markets managed by the TSOs in which a balance is struck at all times between production and consumption [15].

The system operators of Portugal and Spain identify separately situations when they consider necessary to modify the schedules to ensure the quality and security of the electricity supply, and also to guarantee a constant balance of the power system. To this end, they make use of ancillary services, which may be roughly divided into: (i) mandatory services, namely primary frequency and voltage control and (ii) voluntary services, notably secondary and tertiary frequency control, voltage control and service refurbishment. In Portugal, the submission of offers for tertiary regulation should be done until 8:00 p.m. of day $d - 1$. The offers may be adjusted up to 20 min after the publication of the final hourly program of the different intra-day sessions [16].

3.2 American Markets: NYISO and PJM

NYISO manages a day-ahead market and a real-time market, resulting in a two-settlement power system that defines prices and quantities of electricity based on the locational marginal pricing theory. The DAM closes at 5:00 a.m. of day $d - 1$ and the prices resulting from a security constrained unit commitment process are posted until 11:00 a.m. of the same day. The RTM closes 75-min prior to the operating hour and is based on real-time offers. The clearing of the market is settled 45 min before the delivery hour. The RTM allows agents to refine their positions in relation to the DAM settlement. The real-time dispatch involves blocks of 5 min (for each hour *h*), i.e., executes periodically on a 5-min basis and defines the corresponding prices. DAM schedules are balanced against "actual" usage and differences are settled taking into account real-time prices [17].

NYISO considers various ancillary services, notably regulation and frequency response, operating reserve and energy imbalance. Regulation service qualified suppliers submit bids into the day-ahead and real-time regulation services markets. For the day-ahead regulation services market, NYISO establishes the regulation capacity requirements according to the standards established by the North American Electric Reliability Council. Ancillary services are scheduled on an hour basis. Bids may be submitted up to 75-min prior to the delivery hour [18].

The energy market of PJM is divided in two markets, a day-ahead market and a real-time market. The DAM closes at 10:30 of the day before the day of operation and the results are announced at 13:30. The RTM is a 5-min market based on actual real-time operation. Participants may submit bids until 65-min prior to the delivery hour [19].

The Regulation Market of PJM allows participants to trade specific ancillary services. Daily regulation offers must be submitted prior to 14:15 (of the day before the day under consideration) and cannot be negative. These offers can be changed until 65-min prior to the delivery hour [19].

4 Analysis of Key Markets

Tables 1 and 2 summarize the most important features of the day-ahead and intra-day/real-time markets of Nord Pool, MIBEL, NYISO and PJM. The harmonization of European day-ahead markets is at an advanced stage. The main differences between most markets are related to price caps and the possibility of submitting aggregated bids. However, the harmonization of European intra-day markets is still at an early stage. The main differences between most markets are related to the gate closure, the market time unit, the trading type, the pricing mechanism, the price caps and the possibility of submitting aggregated bids.

Table 1. Key day-ahead markets: summary of features.

Feature	Elspot	MIBEL	NYISO	PJM
Gate closure	12–37 h ahead	12–37 h ahead	19–43 h ahead	13.5–37.5 h ahead
Time unit	1 h	1 h	1 h	1 h
Trading type	Auction	Auction	Auction	Auction
Pricing mechanism	Locational marginal pricing	System marginal pricing	Locational marginal pricing	Locational marginal pricing
Product type	Capacity	Capacity	Capacity	Capacity

Table 2. Key intra-day and real-time markets: summary of features.

Feature	Elbas	MIBEL	NYISO	PJM
Gate closure	5–60 min ahead	2–3 h ahead	75 min ahead	65 min ahead
Time unit	15–60 min	1 h	5 min	5 min
Trading type	Continuous	Auction	Auction	Auction
Pricing mechanism	Pay-as-bid (auction also for Germany)	System marginal pricing	Locational marginal pricing	Locational marginal pricing
Product type	Capacity	Capacity	Capacity	Capacity

Table 3 summarizes the main features of key balancing markets associated with Nord Pool and MIBEL. The harmonization of European balancing markets is also at an early stage. The differences between markets are significant.

Table 3. Key balancing markets: summary of features.

Feature	Nord Pool	REN	NYISO	PJM
Gate closure	7–31 h ahead	1–24 h ahead	75 min ahead	1 h ahead
Bid adjustment	45 min ahead	1–3 h ahead	1 h	1 h
Trading type	Auction	Auction	Auction	Auction
Product type	Capacity	Capacity	Capacity	Capacity

Now, variable renewable energy (VRE), such as solar and wind power, has increased significantly during the past years. VRE brings uncertainty to the net load, and also near-zero variable costs, but significant fixed costs. To accommodate the increasing levels of renewable generation, electricity markets should ideally account for the following:

1. aggregated bids (e.g., bid from various wind power producers);
2. a gate-closure close to real-time operation (e.g., a few minutes ahead);
3. the participation of wind power producers in balancing markets;
4. a shorte time unit (e.g., 15 min);
5. new market products adapted to the increasing levels of renewable generation.

The aforementioned European and American markets do not comply with all of these characteristics. For instance, European markets consider reasonable gate-closures (ranging from 3 h, in the intra-day European market, to 5 min, in Elbas), but large time units (ranging from 4 h, in the German balancing market, to 15 min, in Elbas). Also, American markets should consider gate-closures closer to real-time operation.

5 Conclusion

This paper has analyzed two important European electricity markets: Nord Pool and MIBEL. The paper has also described two key American markets: NYISO and PJM. Several conclusions can be drawn from the analysis. The harmonization of day-ahead markets, particularly for the case of European markets, is already at an advanced stage. However, the harmonization of European intra-day markets is still at an early stage. In relation to the European balancing markets, there are still significant differences between them, and the harmonization is also at an early stage.

We can state, at least to some extent, that the design, rules and characteristics of the analyzed markets are still not (completely) adapted to power systems with large levels of variable renewable energy. Accordingly, State Regulators should focus on the following five aspects:

1. The active participation of variable generation in markets.
2. The possibility of aggregated bids.
3. The consideration of gate-closures closer to real-time.
4. Shorter markets time units.
5. New products adapted to variable generation.

These aspects can be considered very important to increase the efficiency of power systems with increasing levels of variable renewable energy.

References

1. Lopes, F., Coelho, H. (eds.): Electricity Markets with Increasing Levels of Renewable Generation: Structure, Operation, Agent-based Simulation, and Emerging Designs. SSDC, vol. 144. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-74263-2. Accessed 15 Feb 2019
2. EC: Directive 2009/72/EC of the European Parliament and of the Council. Off. J. Eur. Union L 211/55–L 211/93 (2009)
3. EC: Proposal COM (2016) 861 for a Regulation of the European Parliament and of the Council on the Internal Market for Electricity. Brussels (2017)
4. Nord Pool. https://www.nordpoolspot.com/
5. Iberian Electricity Market. http://mibel.com/en/home_en/
6. New York Independent System Operator. http://www.nyiso.com/
7. Pennsylvania-New-Jersey-Maryland Independent System Operator. https://www.pjm.com/
8. Skytte, K., Grohnheit, P.E.: Market prices in a power market with more than 50% wind power. In: Lopes, F., Coelho, H. (eds.) Electricity Markets with Increasing Levels of Renewable Generation: Structure, Operation, Agent-based Simulation, and Emerging Designs. SSDC, vol. 144, pp. 81–94. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-74263-2_4. Accessed 15 Feb 2019
9. Ela, E., Milligan, M., Bloom, A., Botterud, A., Townsend, A., Levin, T.: Incentivizing flexibility in system operations. In: Lopes, F., Coelho, H. (eds.) Electricity Markets with Increasing Levels of Renewable Generation: Structure, Operation, Agent-based Simulation, and Emerging Designs. SSDC, vol. 144, pp. 95–127. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-74263-2_5. Accessed 15 Feb 2019

10. Lopes, F.: Electricity markets and intelligent agents Part I: market architecture and structure. In: Lopes, F., Coelho, H. (eds.) Electricity Markets with Increasing Levels of Renewable Generation: Structure, Operation, Agent-based Simulation, and Emerging Designs. SSDC, vol. 144, pp. 23–48. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-74263-2_2. Accessed 15 Feb 2019
11. Ela, E., et al.: Overview of wholesale electricity markets. In: Lopes, F., Coelho, H. (eds.) Electricity Markets with Increasing Levels of Renewable Generation: Structure, Operation, Agent-based Simulation, and Emerging Designs. SSDC, vol. 144, pp. 3–21. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-74263-2_1. Accessed 15 Feb 2019
12. Nord Pool: Day-ahead Market Regulations. Nord Pool AS, 15 August 2017
13. Nord Pool: Intraday Market Regulations. Nord Pool AS, 12 June 2018
14. Entsoe: Nordic Balancing Philosophy. European Network of Transmission System Operators, 16 June 2016
15. OMIE: Day-Ahead and Intraday Electricity Market Operating Rules. OMI-Polo Español S. A., 11 May 2018
16. ERSE: Manual de Procedimentos da Gestão Global do Sistema do Setor Elétrico. Entidade Reguladora dos Serviços Energéticos, Junho 2018
17. NYISO: NYISO Markets: New York's Marketplace for Wholesale Electricity, New York Independent System Operator (2018)
18. NYISO: Ancillary Services Manual. New York Independent System Operator, April 2018
19. PJM: Manual 11 - Energy & Ancillary Services Market Operations. Pennsylvania-New Jersey-Maryland Independent System Operator, October 2018

A New Approach to Provide Sustainable Solutions for Residential Sector

Ricardo Santos[1,2](✉), João C. O. Matias[2,3,4], and Antonio Abreu[5,6]

[1] University of Aveiro, Aveiro, Portugal
ricardosimoessantos84@ua.pt
[2] GOVCOPP - University of Aveiro, Aveiro, Portugal
jmatias@ua.pt
[3] Department of Economics, Management, Industrial Engineering and Tourism (DEGEIT), University of Aveiro, Aveiro, Portugal
[4] C-MasT – University of Beira Interior, Covilhã, Portugal
[5] ISEL - Instituto Superior de Engenharia de Lisboa, Instituto Politécnico de Lisboa, Lisbon, Portugal
ajfa@dem.isel.ipl.pt
[6] CTS Uninova, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Lisbon, Portugal

Abstract. An energy-efficient appliance normally presents a lower energy consumption compared to a less efficient one, with a higher initial investment, although this not always happens. Additionally, each appliance, presents very different features, leading to some difficulties on its choice by the consumer (decision-agent).

On the other hand, each consumer, has specific and distinguished needs from other consumers, namely of social, economic or environmental nature. Even by adopting these criteria, this is not an isolated guarantee of an optimal solution for the consumer. It is then necessary to complement this approach with multicriteria, combined with optimization techniques. Evolutionary Algorithms (EA), could be used as an optimization technique, to provide sustainable solutions to the consumer, from the market. In this paper, it's presented an approach that integrates both concepts, where at the end, it shall be presented a case study, to demonstrate the application of the proposed method.

Keywords: Energy efficiency · Electrical appliances · Life cycle cost analysis · Multi-Attribute Value Theory (MAVT) · Multi-objective optimization · Evolutionary algorithms

1 Introduction

The residential sector, plays an important role, in achieving Sustainable Development (SD), with buildings accounting for about 40% of the energy consumed [1, 2].

In the last years, there was made some energy efficiency improvements, regarding electrical household appliances. One of such measures, was the mandatory labeling [3, 4], which allows to inform the consumer about important issues, specifically

Published by Springer Nature Switzerland AG 2019
L. M. Camarinha-Matos et al. (Eds.): DoCEIS 2019, IFIP AICT 553, pp. 329–342, 2019.
https://doi.org/10.1007/978-3-030-17771-3_29

regarding each appliance (e.g. energy consumption, noise, capacity (fridge), etc.), promoting therefore a suitable use, adjusted to its needs [5].

However, and given the several options available on market (brands and models) as well as the appliances' features, it's difficult to analyze their benefit-cost ratio and therefore, what's the best solution to adopt, to satisfy the consumer needs [6, 7].

In this sense, Multiple-Attribute Value Theory (MAVT), could be used as a method, based on a set of criteria, to define the space decision and both objective functions.

Furthermore, the use of optimization techniques, combined with MAVT, can support the decision-agent (consumer), by achieving sustainable solutions, through the household appliances to be acquired.

Given the previous work from [14], Evolutionary Algorithms (EA), have been successfully applied to solve this kind of optimization problem with less time than other algorithms, given their stochastic nature and global search ability [14, 15].

Therefore, this work aims to contribute to the following PhD research question, by proposing an integrated approach, based on MAVT and Non-dominated Sorting Genetic Algorithm II (NSGAII), to provide sustainable solutions to the consumer and from the market, concerning Water, Energy and CO_2 savings, satisfying at the same time his different needs according to a set of criteria.

Research Question
Is it possible to develop a holistic model (Economic, Social and Environmental dimensions) to support decision making, based on evolutionary algorithms (EA), that allows the decision agent to obtain sustainable solutions?

Hypothesis
If by using Multi-objective Evolutionary Algorithms (EA), combined with Multicriteria analysis, it's possible to achieve several sustainable solutions to the consumer, by selecting household appliances from the market.

2 Relationship Between the PhD Work to Industrial and Service Systems

As it mentioned before, the industry and service sectors are going through profound transformation towards digitalization and integration of new levels of "smartness", originating therefore the 4^{th} industrial revolution.

This transformation, is led by terms such as Industry 4.0, Smart Manufacturing and Economy 4.0, giving therefore an interdisciplinary character, expressed by an increasing digitalization and interconnection of systems, products, services and business models. The link between the physical and the cyber worlds, as well as the integration of the "exponential technologies", are key features of this innovation trend.

The paper presented here, is part of a methodology, which is being developed on behalf of a PhD work, to be applied in the context of the 4^{th} Industrial Revolution.

Given its multidisciplinary character, by establishing the interconnection between several and different concepts (e.g. energy efficiency, investment decisions, environmental impact, evolutionary algorithms, product life cycle, among others) as well as being a link between physical systems and cyber worlds, this work can make some contributions specially at a sustainable level, maximizing therefore, the environmental, social and economic wellbeing for each decision agent (e.g. household consumers, companies, public institutions, etc.).

The example of application, presented in this work, can be extended to electrical appliances, regarding industry services, as well as other devices, applied in industry.

The approach presented here, can be suitable at the same time, to the world's different changes (e.g. prices, technology innovation, legislation, etc.), being therefore, a technological Innovation for Resilient Systems.

Therefore, the main goal is to support the decision-agent decisions, with a methodology, implemented by an app, that by making the interconnection between physical systems and cybernetic world, can provide sustainable solutions, regarding a set of criteria pre-established.

3 State of the Art/Related Literature

Methods like simulation (e.g. [18]) based on what if analysis, are usually employed to investigate a limited number of an alternative options.

There are some approaches, which are mainly economical, allowing therefore to obtain highest energy savings, for the same initial investment (e.g. [19, 20]). Other approaches, explores several issues like benefit-cost analysis, initial investment, CO_2 savings, among others, related to retrofitting measures (e.g. [21]), where some of them, are even combined with measures and technologies too (e.g. [20]).

However, this type of approaches is considered somehow limited, since it does not account other important factors (e.g. environmental, energy labelling, legal, social factors, among others) to find solutions, suitable to the occupant needs, as well as, they don't consider the different criteria regarding each household appliance, available on the market, and according to the number of household occupants.

Nowadays, some works have developed multicriteria decision making (MCDM) models to support professionals to solve problems, associated with the retrofitting of buildings, by taking into consideration factors, such as the degradation of building elements, energy efficiency, and internal environment comfort. (e.g. [16]), although others, are based on the ranking of alternative solutions (e.g. [17]).

Although MCDM, allows to choose the best alternative on each set of viable options, the criteria adopted are usually conflicting on nature, giving therefore a solution that it is impossible to be optimal against all criteria.

In the same context, there are also other MCDM models, as well as Multiple−attribute value theory (MAVT) models, found on literature, that combines optimization with multicriteria techniques in order to obtain feasible solutions, by exploring many alternative measures/solutions, pre-selected, according to a set of criteria, suitable to the consumer needs (e.g. [18–21]).

However, such approaches don't consider the different criteria regarding each household appliance, available on the market for each dwelling and its occupants.

Methods based on metaheuristics, have been also applied into energy problems, as an efficient tool to provide a set of feasible solutions, such as Particle Swarm Optimization (PSO) (e.g. [9]) and Genetic Algorithms (GA) (e.g. [13]), among others.

However, none of this methods have been integrated into a combined approach that allows to select efficient appliances to a decision-agent, according to a set of criteria.

4 Research Contribution and Innovation

The approach presented here, was developed to support a Decision-Agent (DA), who wants to acquire a set of electrical appliances (energy services) from the market.

On Fig. 1, it is presented an approach to provide an optimal set of appliances, regarding each energy service, needed by the DA (e.g. Consumer).

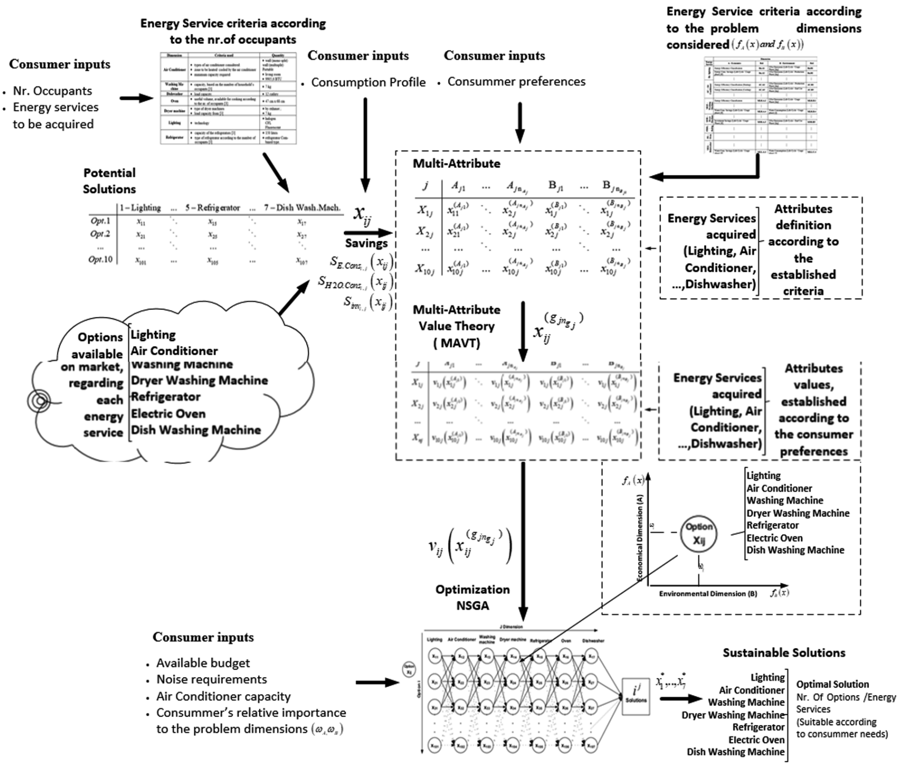


Fig. 1. Conceptual model

At first, a set of potential solutions (x_{ij}) are pre-selected from the market, according to specific criteria, pre-established, based on the number of occupants. The criteria are the same, although the value of the correspondent attributes, can change according to the building number of occupants. An example of such table, is given on ANNEX I (Table A1), regarding the case study presented here.

The pre-selection, allows to reduce the decision space, accounting only the solutions, suitable to the consumer needs, as well as to increase NSGAII efficiency, by achieving optimal solutions with less time.

Each one of this potential solution $\left(x_{ij}\right)$, is then formulated as an option i, regarding energy service j, that can be acquired by the DA (consumer) from the market.

Given a DA (e.g. consumer) consumption profile (see example on ANNEX I (Table A2)), Life Cycle Cost Assessment (LCCA) is then preformed to achieve, for each appliance, the corresponding savings, regarding energy consumption ($S_{E.Cons_{i,j}}\left(x_{ij}\right)$), water consumption ($S_{H2O.Cons_{i,j}}\left(x_{ij}\right)$) as well as the initial investment ($S_{inv_{i,j}}\left(x_{ij}\right)$). Both parameters, are savings, obtained from the comparison between the efficient and the correspondent "standard solution" (less efficient one).

Given the diversity of features, regarding each solution, as well as the DA's economic, social and environmental concerns, it was defined a set of attributes according to the consumer preferences and regarding each energy service, for the two problem dimensions; A-Economics, B-Environment. These attributes are presented on Table 1.

The consumption profile was performed, by making a set of assumptions based on the hours, which was then extrapolated to a weekly and year base. However, the decision-agent (consumer) can also define its usage profile according to its needs, or by using the profile, considered in the case study presented here, as a default.

As it referred before, MAVT is used to support the DA, by evaluating a set of alternative solutions, according to a set of criteria/attributes established (Table 1).

Based on criteria from Table 1, it was defined $x_{ij}^{(g_{jt})}$, as the attribute regarding each alternative solution i, associated to a certain energy service j, stablished according to criteria t, associated to energy service j and problem dimension g considered (A - Economical; B - Environmental), i.e.:

$$g_{jt} \in \left\{ \left\{ A_{j1}, A_{j2}, ..A_{jn_{A_j}} \right\} \cup \left\{ B_{j1}, B_{j2}, ..B_{jn_{B_j}} \right\} \right\} \tag{1}$$

$$g = \{A, B\} \wedge j = \{1, 2, .., 7\} \wedge t = \left\{ \left\{ 1, 2, ..n_{A_j} \right\} \cup \left\{ 1, 2, ..n_{B_j} \right\} \right\} \wedge n_{A_j}, n_{B_j}, t, j \in \mathbb{N} \tag{2}$$

By following the notation described above and based on criteria stablished on Table 1, as well as the assumptions presented on Table A2 (ANNEX I) for the considered case study, it was defined the correspondent attribute behavior/pay-off $\left(x_{ij}^{(g_{jt})}\right)$, regarding each option i, belonging to energy service j. On Fig. 1(a), it's presented an example of this table, regarding the energy service "Lighting".

Therefore, and according to MAVT, there is a value $v_{ij}^{(g_{jt})}\left(x_{ij}^{(g_{jt})}\right)$, associated to the attribute $x_{ij}^{(g_{jt})}$, such as:

$$\boldsymbol{x}_{ij}^{(g_{jt})} \rightarrow \boldsymbol{v}_{ij}^{(g_{jt})}\left(\boldsymbol{x}_{ij}^{(g_{jt})}\right) \tag{3}$$

Table 1. Attributes used to define problem dimensions, regarding each energy service considered

	Dimension			
Energy service	A - Economics	Ref.	B - Environment	Ref
Ilu – lighting	Energy Efficiency Classification	Ilu.Al	CO_2e Emissions (Life cycle - Usage phase) [kg]	Ilu.B1
	Energy Cons. Savings (Life cycle - Usage phase) [m]	Ilu.A2	CO_2e Emissions (Life cycle - Production phase) [kg]	Ilu.B2
	⋮	⋮	⋮	⋮
AC – Air Conditioner	Energy Efficiency Classification (Heating)	AC.A2	CO_2e Emissions (Life cycle - Production phase) [kg]	AC.B2
	Energy Efficiency Classification (Cooling)	AC.A3	CO_2e Emissions (Life cycle - End use phase) [kg]	AC.B3
	⋮	⋮	⋮	⋮
MLR – Washing Machine	Energy Efficiency Classification	MLR.A.1	CO_2e Emissions (Life cycle - Usage phase) [kg]	MLR.B.1
	⋮	⋮	⋮	⋮
	Water Cons. Savinss (Life cycle - Usage phase) [€]	MLR.A.4	Water Consumption (Life cycle - Usage phase) [L]	MLR.B.4
MSR – Dryer Machine	⋮	⋮	⋮	⋮
	Investment Savings (Life cycle - Usage Phase) [€]	MSR.A.3	CO_2e Emissions (Life cycle - End use phase) [kg]	MSR.B3
FRIG – Refrig	⋮	⋮	⋮	⋮
FE – Oven	⋮	⋮	⋮	⋮
MLL – Dishwasher	⋮	⋮	⋮	⋮
	Water Cons. Savings (Life cycle - Usage phase) [€]	MLL.A.4	Water Consumption (Life cycle - Usage phase) [L]	MLL.C.4

Given the 2 objectives considered, different attributes are used to measure the performance in relation to that set of objectives. This attributes, are usually measured on different measurement scales. Therefore, in order to transform the criteria to follow the same scale and units, it was used an expression to stablish the relationship between the new and the previous value of $x_{ij}^{(g_{jt})}$, respective $\left(v_{ij}^{(2)}\left(x_{ij}^{(g_{jt})}\right)\right)$ and $\left(v_{ij}^{(1)}\left(x_{ij}^{(g_{jt})}\right)\right)$, by using as well, the correspondent worst and better results, for a given criteria g_{jt}, i.e.:

$$v_{ij}(x_{ij}^{(2)}) = \frac{\left(v_{ij}^{(1)}(x_{ij}^{(g_{jt})}) - v_{worst \cdot ij}(x_{ij}^{(g_{jt})})\right)}{\left(v_{better \cdot ij}(x_{ij}^{(g_{jt})}) - v_{worst \cdot ij}(x_{ij}^{(g_{jt})})\right)} \tag{4}$$

The new values of each $x_{ij}^{(g_{jt})}$ $\left(v_{ij}\left(x_{ij}^{(g_{jt})}\right)\right)$, fills a new evaluation table, belonging to each energy service j. On Fig. 2, it's shown an example for the two tables regarding energy service "Lighting".

Lighting	$A.1_{\cdot 1}$	...	$A.n_{\cdot 1}$	$B.1_{\cdot 1}$	...	$B.n_{\cdot 1}$
X_{11}	$x_{11}^{(A.1.)}$	$\ddots$	$x_{1n_1}^{(A.n_{A1}.)}$	$x_{11}^{(B.1)}$	$\ddots$	$x_{11}^{(B.n_{B1})}$
X_{21}	$x_{21}^{(A.1.)}$	$\ddots$	$x_{1n_1}^{(A.n_{A1}.)}$	$x_{21}^{(B.1)}$	$\ddots$	$x_{21}^{(B.n_{B1})}$
...	...	$\ddots$	...	...	$\ddots$	...
X_{101}	$x_{101}^{(A.1.)}$	...	$x_{1n_1}^{(A.n_{A1}.)}$	$x_{101}^{(B.1)}$	...	$x_{101}^{(B_{B1})}$

a)

Lighting	$A.1_{\cdot 1}$	...	$A.n_{\cdot 1}$	$B.1_{\cdot 1}$	...	$B.n_{\cdot 1}$
X_{11}	$v_{11}\left(x_{11}^{(A.1.)}\right)$	$\ddots$	$v_{11}\left(x_{1n_1}^{(A.n_{A1}.)}\right)$	$v_{11}\left(x_{11}^{(B.1)}\right)$	$\ddots$	$v_{11}\left(x_{11}^{(B.n_{B1})}\right)$
X_{21}	$v_{21}\left(x_{21}^{(A.1.)}\right)$	$\ddots$	$v_{21}\left(x_{2n_1}^{(A.n_{A1}.)}\right)$	$v_{21}\left(x_{21}^{(B.1)}\right)$	$\ddots$	$v_{21}\left(x_{21}^{(B.n_{B1})}\right)$
...	...	$\ddots$	...	...	$\ddots$	...
X_{101}	$v_{101}\left(x_{101}^{(A.1.)}\right)$	...	$v_{101}\left(x_{101}^{(A.n_{A1}.)}\right)$	$v_{101}\left(x_{101}^{(B.1)}\right)$	...	$v_{101}\left(x_{101}^{(B.n_{B1})}\right)$

b)

Fig. 2. Example of evaluation table (Lighting energy service): (a) $x_j^{(g_{jt})}$ (b) $v_{ij}\left(x_j^{(g_{jt})}\right)$

Based on the value attributes previously achieved, it was used the additive model to aggregate them, referred to each option i, regarding energy service j, which was further improved, by applying optimization techniques, by using NSGAII algorithm.

Given the trade-off, and the diversity of features, regarding each solution, the consumer will be confronted with a problem of combinatorial nature (Fig. 1), where the number of combinations is dependent on the number of options to be considered, regarding each dimensions. This number, although, potentially bigger (23 million of combinations approximately for the case study considered), can be reduced, by assuming that the consumer cannot perform any choices (x_{ij}), given his limited budget (Fig. 1).

Constraints like the air conditioner capacity and appliances noise minimal requirements, will also be accounted to suit consumer needs, to improve its social wellbeing.

After defining the value attributes of each potential solution, and by using the additive model, the problem presented here, can be formulated as follows:

$$\max\ V_m(x), \frac{c}{m} = A, B$$

$$subjecttox \in Xc/V_m(x) = [V_A(x), V_B(x)]^T \tag{5}$$

Where x is the decision variable vector, defined as:

$$x \in X : x \in \left\{ x_{ij}^{(A_{jt})}, x_{ij}^{(B_{jt})} \right\} \wedge t, i, j \in \mathbb{N} \tag{6}$$

$$w/j = \{1, .., 10\} \wedge j = \{1, .., 7\} \wedge t = \left\{ \left\{ 1, .., n_{A_j} \right\} \cup \left\{ 1, .., n_{B_j} \right\} \right\} \wedge n_{A_j}, n_{B_j} \in \mathbb{N} \tag{7}$$

Where $V_A(x)$ and $V_B(x)$, are the aggregate objective functions, regarding each dimension considered (A-Economics; B-Environment):

$$V_g(x) = \sum_{j=1}^{n_j} \sum_{t=1}^{n_{g_j}} v_j(x_j^{(g_{jt})}) w/g = \{A, B\} \wedge v_j(x_j^{(g_{jt})}) \wedge n_j, n_{g_j}, t, j \in \mathbb{N} \tag{8}$$

Therefore, and based on (8), the objective functions are:

$$Economic Well - being : max\ V_A(x) = \sum_{j=1}^{n_j} \sum_{t=1}^{n_{A_{jt}}} v_j(x_j^{(A_{jt})}) \tag{9}$$

$$Environment Well - being : max\ V_B(x) = \sum_{j=1}^{n_j} \sum_{t=1}^{n_{B_{jt}}} v_j(x_j^{(B_{jt})}) \tag{10}$$

The first objective function is based on the works of [14] and [22].

Using the additive model from MAVT, the aggregated function, results into a unique objective function, weighted by the DA relative importance $\left(\omega_g\right)$ as follows:

$$V(V_A(x), V_B(x)) = \omega_A . V_A(x) + \omega_B . V_B(x) = \sum_{j=1}^{n_j} \left(\omega_A \sum_{t=1}^{n_{A_j}} v_j(x_j^{(A_{jt})}) + \omega_B \sum_{t=1}^{n_{B_j}} v_j(x_j^{(B_{jt})}) \right) \tag{11}$$

The constraints, regarding economic and environment well-being/dimensions, are:
Economic – Budget:

$$r1 : \sum_{j=1}^{n_{dim}} I_j\left(x_j\right) \le available\ budget\left(\eta_{disp.}\right) \Leftrightarrow \sum_{j=1}^{n_{dim}} x_j^{\left(A_{jt}\right)} \le \eta_{disp.} \tag{12}$$

$$w/A_{jt} = \{A_{14}, A_{26}, A_{35}, A_{44}, A_{54}, A_{64}, A_{75}\} \wedge nt, j_{dim} \tag{13}$$

Environment – Noise:

$$r_j : Noise_j \le Max.Noise_j \Leftrightarrow x_j^{\left(B_{jt}\right)} \le Max.Noise_j \tag{14}$$

$$w/B_{jt} = \{B_{24}, B_{35}, B_{44}, B_{54}, B_{64}, B_{75}\} \wedge nt, j_{dim} \tag{15}$$

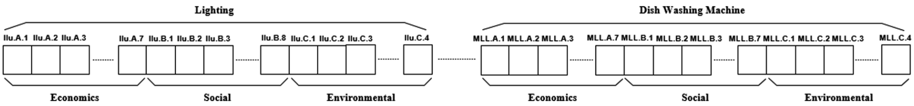


Fig. 3. NSGAII's individual framework

The NSGAII individual framework, is presented as follows on Fig. 3.

NSGAII's codification used, was real.

The model will be applied by using a case study, and considering a consumer (DA), who wishes to acquire 7 types of appliances. His building has four occupants (him included) and the relative importance's (ω_A and ω_B) are respectively 0,7 and 0,3. All the remaining assumptions, were made on Tables A1 and A2 (ANNEX I).

5 Results and Discussion

NSGAII was programmed in Matlab using the following parameters:

- Selection method: Tournament
- Crossover method: single point
- Mutation used: Normal Random Mutation

The remained NSGA-II parameters (initial population, crossover and mutation rate) were defined after several computational experiments.

The parameter of max generations was tested at first, where it was selected a maximum generation number of 90 to show that if 90 iterations were enough to find the Pareto frontier. Other parameters were also tested, such as the population size (100 individuals), the tournament size (10), the crossover rate (0.9), and the mutation rate (0.3). The obtained results, regarding 90^{th} and 100^{th} generations, are presented on Fig. 4(a), where it can be seen that both cases, have similar Pareto frontier. In this sense it was selected the max number iterations/generations of 90. Then, it was preformed several combinations of crossover and mutation rates of NSGA-II (Table 2).

Table 2. Combinations of crossover and mutation rates used

Experiment	Crossover rate	Mutation rate
1	0.8	0.1
2	0.8	0.3
3	0.9	0.1
4	0.9	0.3

The combinations, presented on Table 2, were performed by setting a max iteration of 90. The correspondent results, are shown on Fig. 4(b), where it's noted that the small change on parameters, has little effect on the results. Thus, it was used the following parameters of NSGA-II to show the results of the present case: population size (100), max iteration (90), tournament size (10), crossover rate (0.9) and mutation rate (0, 1).

NSGA-II is applied on resolution of multi-objective problems. Therefore, the correspondent solution, is a Pareto frontier, which is gradually formed through an iteration process, where is an increase in the number of nodes of Pareto frontier in the early generations. Once the frontier was formed, better results of each node were founded in further iterations.

After NSGAII calculations, the Pareto frontier in Fig. 4(a), is therefore obtained, where each node represents a potential solution of the problem, i.e., a set of sustainable solutions (appliances), regarding each energy service required.

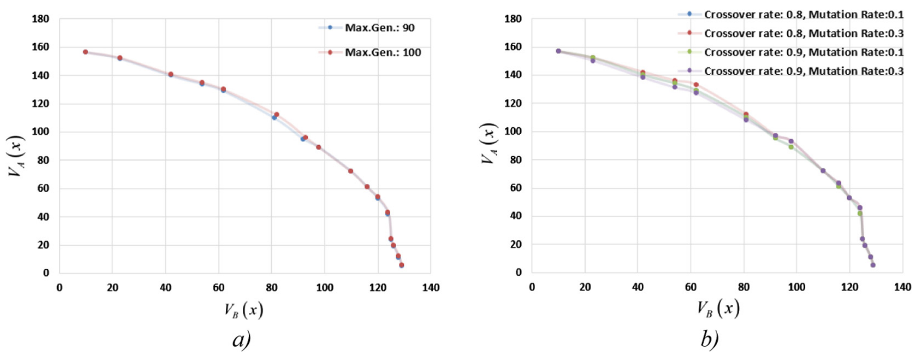


Fig. 4. Pareto frontier: (a) for 90th and 100th generations (b) different parameters (90th Generation)

Table 3. Example of a sustainable solution obtained from this approach

Dimension	Stand. solution total invest. (€)	Effic. sol. total invest (€)	Invest. saving (€)	Energy consum. savings (€)	Water consum. savings (l)	CO_2 savings (kg)	Brand	Model
Lighting	15,89	09,53	5,34	59,4	–	28,9	GE	EFL23 W
Air conditioning	368,0	299,0	69,0	1320,6	–	1315,7	Whirlpool	PACW9HP
Refrigerator	250,0	529,0	−279,0	708,1	–	8,7	Candy	CFET 6182 W
Dishwasher machine	310,0	349,0	−39,0	3,2	423	6,9	Bosch	SMS25AI00E
Washing machine	262,0	294,0	−32,0	6,9	317	94,8	Siemens	WI12A222ES
Oven	170,0	199,0	−29,0	1,7	–	2,2	Zanussi	ZZB21601XV
Clothes dryer	349,0	419,0	−70,00	12,3	–	1,7	Electrolux	EDP2074PDW
Total:	1724,9	2099,6	−374,7	2112,3	740	1458,9	–	–

One of these nodes are presented on Table 3, as an example of a feasible solution obtained, considering a budget of 2100 €.

It is also presented the CO_2 savings, regarding the choice of this solution, compared with the less efficient (standard) one (approx. values).

According to Table 3, if the consumer, opts for the solutions set, provided by this approach, he can save up to 2112,3 € per year, further contributing to a 1458,9 kg of CO_2 and 740 L of water, both savings/years, given the 10 years life cycle considered.

6 Conclusions and Further Work

In this work, it was presented an approach to provide sustainable household appliances from the market, to the decision-agent (consumer), by considering two objectives, regarding sustainability; environment and economic well-being.

Both solutions were pre-selected, according to a set of specific criteria and regarding each type of appliance, considered by the case study.

Criteria was used, to pre-select the appliances from the market, adjusting therefore, the method to the case study presented here. Other criteria, were combined with MAVT, to modelling the consumer preferences, according to the two problem dimensions.

The main objective, was to maximize the consumer well-being (environment and economics). The social wellbeing was also promoted, by suit the obtained solutions to the consumer needs.

The relative importance, given by the DA (consumer), was also considered, in order to weight the DA decision through both dimensions.

NSGAII, where then applied here, to get optimal solutions, by maximizing both dimensions, considering the environmental impact (CO_2 savings), as well as the economic rationality (initial investment and energy consumption savings) regarding the lifecycle of each appliance, during its usage phase.

The results show that this method performs well in this case, by providing optional and also, sustainable appliances, that attends the consumer needs.

We test different parameters (max number of iterations, crossover rate and mutation rate) and in the common range of values, these parameters are not sensitive to the results. Additionally, NSGA-II can also find the Pareto frontier of the solutions, providing therefore several alternative solutions to the consumer.

The achievements presented on this work, allows to proceed in a way of getting a more completed approach that maximizes both dimensions of sustainability (Economical, Environmental and Social) with the social dimension, being therefore integrated into the model developed here, in order to better express the consumer preferences, which will allow to answer the PhD research question presented here.

Appendix

ANNEX I – Criteria and consumer profile given the case study considered

Table A.1. Criteria used to pre-select the appliances from the market (applied to case study)

Dimension	Criteria used	Quantity
Air conditioner	• Types of air conditioner considered • Zone to be heated/cooled by the air conditioner • Minimum capacity required	•Wall (mono split) Wall (multi split) Portable • Living room • 9905,6 BTU
Washing machine	• Capacity, based on the number of household's occupants [3]	• 7 kg
Dishwasher	• Load capacity	• 12 cutlery
Oven	• Useful volume, available for cooking according to the nr. of occupants [3]	• 47 cm × 68 cm
Dryer machine	• Type of dryer machines • Load capacity from [3]	• By exhaust • 7 kg
Lighting	• Technology	• Halogen CFL Fluorescent
Refrigerator	• Capacity of the refrigerators [3] • Type of refrigerator according to the number of occupants [3]	• 150 L • Refrigerator combined type

Table A.2. Assumptions taken, regarding the consumer usage profile

Emission factor [gCO2/kWh]	675	Discount factor [%]		7
Life cycle (usage phase) [years]:	10	Annual factor		7,02
Electrical energy tariff [€/kWh]	0,162	Water tariff [€/m3]		1,19
Energy service	Usage profile (h)			
	Daily	Weekly	Monthly	Annual
Air conditioner	2	12	48	576
Washing machine	1	4	16	192
Dryer machine	1	4	16	192
Refrigerator	11	77	330	4015
Electric oven	1	2	8	96
Dish washing machine	1	4	16	192
Lighting	5	35	150	1825
Energy service	Usage profile (Frequency)			
	Daily	Weekly	Monthly	Annual
Air conditioner	1	6	24	288
Washing machine	1	4	16	192
Dryer machine	1	4	16	192
Refrigerator	1	7	30	365
Electric oven	1	2	8	96
Dish washing machine	1	4	16	192
Lighting	1	7	30	365

References

1. IEA: Energy Efficiency 2017 – Market Reports Series, OECD/IEA (2017)
2. Gul, M., Patidar, S.: Understanding the energy consumption and occupancy of a multi-purpose academic building. Energy Build. **87**, 155–165 (2015). ISSN 0378-7788
3. ADENE: Manual da Etiqueta Energética, ADENE, Lisboa ISBN: 978-972-8646-36-3 (2017)
4. DGEG: Eficiência Energética em Edifícios – Programa E4, Direção Geral de Energia e Geologia, Lisboa (2002)
5. Wong, I.L., Krüger, E.: Comparing energy efficiency labelling systems in the EU and Brazil: implications, challenges, barriers and opportunities. Energy Policy **109**, 310–323 (2017)
6. Fell, M.: Energy services: A conceptual review. Energy Res. Soc. Sci. **27**, 129–140 (2017)
7. Hoxha, E., Jusselme, T.: On the necessity of improving the environmental impacts of furniture and appliances in net-zero energy buildings. Sci. Total Environ. **596–597**, 405–416 (2017)
8. Ting, T.O., Rao, M.V., Loo, K.C.: A novel approach for unit commitment problem via an effective hybrid particle swarm optimization. IEEE Trans. Power Syst. **21**, 411–418 (2006)
9. Ko, M.J., Kim, Y.S., Chung, M.H., Jeon, H.C.: Multi-objective design for a hybrid energy system using genetic algorithm. Energies **8**, 2924–2949 (2015)
10. Randall, M., Rawlins, T., Lewis, A., Kipouros, T.: Performance comparison of evolutionary algorithms for airfoil design. Procedia Comput. Sci., vol. 51, pp. 2267–2276. Springer, Heidelberg (2015)
11. Goldberg, D.: Genetic Algorithms in Search Optimization and Machine Learning. Addison Wesley, Maryland (1989)
12. Chuah, J.W., Raghunathan, A., Jha, N.K.: ROBESim: a retrofitoriented building energy simulator based on EnergyPlus. Energy Build. **66**, 88–103 (2013)
13. Pombo, O., Allacker, K., Rivela, B., Neila, J.: Sustainability assessment of energy saving measures: a multi-criteria approach for residential buildings retrofittingda case study of the Spanish housing stock. Energy Build. **116**, 384–394 (2016)
14. Santos, R., Matias, J.C.O., Abreu, A.: Energy efficiency in buildings by using evolutionary algorithms: an approach to provide efficiency choices to the consumer, considering the rebound effect. In: Camarinha-Matos, L.M., Adu-Kankam, K.O., Julashokri, M. (eds.) DoCEIS 2018. IAICT, vol. 521, pp. 120–129. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-78574-5_12
15. Asadi, E., Silva, M.G., Antunes, C.H., Dias, L.: Multi-objective optimization for building retrofit strategies: a model and an application. Energy Build. **44**, 81–87 (2012)
16. Caccavelli, D., Gugerli, H.: Tobus: a European diagnosis and decision-making tool for office building upgrading. Energy Build. **34**(2), 113–119 (2002)
17. Kaklauskas, A., Zavadskas, E.K., Raslanas, S.: Multivariant design and multiple criteria analysis of building refurbishments. Energy Build. **37**(4), 361–372 (2005)
18. Pombo, O., Allacker, K., Rivela, B., Neila, J.: Sustainability assessment of energy saving measures: a multi-criteria approach for residential buildings retrofittingda case study of the Spanish housing stock. Energy Build. **116**, 384–394 (2016)
19. Jafari, A., Valentin, V.: An optimization framework for building energy retrofits decision-making. Build. Environ. **115**, 118–129 (2017). ISSN 0360-1323. https://doi.org/10.1016/j.buildenv.2017.01.020. (http://www.sciencedirect.com/science/article/pii/S0360132317300331)
20. Mauro, G.M., Hamdy, M., Vanoli, G.P., Bianco, N., Hensen, J.L.M.: A new methodology for investigating the cost-optimality of energy retrofitting a building category. Energy Build. **107**, 456–478 (2015)

21. Heo, Y., Augenbroe, G., Graziano, D., Muehleisen, R.T., Guzowski, L.: Scalable methodology for large scale building energy improvement: relevance of calibration in model-based retrofit analysis. Build. Environ. **87**, 342–350 (2015)
22. Santos, R.S., Matias, J.C.O., Abreu, A., Reis, F.: Evolutionary algorithms on reducing energy consumption in buildings: an approach to provide smart and efficiency choices, considering the rebound effect. Comput. Ind. Eng. **126**, 729–755 (2018). ISSN 0360 8352. https://doi.org/10.1016/j.cie.2018.09.050

Electric Vehicles Aggregation in Market Environment: A Stochastic Grid-to-Vehicle and Vehicle-to-Grid Management

Isaias L. R. Gomes[1,2], Rui Melicio[1,2(✉)], and Victor M. F. Mendes[1,3,4]

[1] Departamento de Física, ICT, Universidade de Évora, Escola de Ciências e Tecnologia, Évora, Portugal
ruimelicio@gmail.com
[2] IDMEC, Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal
[3] CISE, Electromechatronic Systems Research Centre, Universidade da Beira Interior, Covilhã, Portugal
[4] Instituto Superior of Engenharia de Lisboa, Lisbon, Portugal

Abstract. This paper addresses a development of a support management system for a power system aggregator managing a fleet of electric vehicles and bidding in a day-ahead electricity market. The support management system is modeled by stochastic mixed integer linear programming approach. The charge and discharge of the batteries of the fleet of vehicles are brought about to a convenient contribution for the maximization of the expected profit of the aggregator. The optimization takes into consideration the profiles of usage of the vehicle owners and the battery degradation of the vehicles. The vehicles are assumed as bidirectional energy flow units: allowing grid-to-vehicle or vehicle-to-grid operation modes. A strong interaction of information exchange is assumed between the aggregator and vehicle owners. A set of scenarios is created by a scenario generation method based on the Kernel Density Estimation technique and are subjected to a reduction by a K-means clustering technique. A case study with data of Electricity Market of Iberian Peninsula is presented to drive conclusion about the support management system developed.

Keywords: Electric vehicles aggregator · Day-ahead market · Scenario generation · Scenario reduction · Stochastic programming

1 Introduction

The anthropogenic greenhouse gas emission is mostly coming from fossil fuels usage, being the sector of transports in charge of approximately 15% of total emission [1] and said to be responsible for 28% of the total energy usage in most countries [2]. Although in the recent past, measures have taken to reduce greenhouse gas emissions from transports. For instance, measures to persuade for the heightening of the usage of an electric vehicle (EV), appealing to this usage instead of the conventional fossil fuels vehicles [3]. The desired rise on the EV market penetration has remained limited, due

Published by Springer Nature Switzerland AG 2019
L. M. Camarinha-Matos et al. (Eds.): DoCEIS 2019, IFIP AICT 553, pp. 343–352, 2019.
https://doi.org/10.1007/978-3-030-17771-3_30

to issues concerned with: EV acceptance; high costs; short driving range; high battery price; time to recharge the battery, battery degradation and the scarce concentration of geographic handy public charging stations. Although, integration in the electric power system of numerous EVs seems to be a menace for the grid in the sense that, if most vehicles are parked for charging during the night or the day at some periods, these periods are likely to be new peaks of consumption. At the same time, the penetration of EV is an opportunity of new business in energy markets. Hence, EV electric power management is an opportunity and an important line of research [4] to deliver knowledge for power system EV aggregators (EVA) as market agents with functionality of: (i) acting as an intermediary between EV owners and the market operator; (ii) preventing degradation of batteries by a convenient energy management; (iii) deciding for profitable bids in a day-ahead electricity market. Hence, an EVA must be supported with computational applications to convenient face the functionality [5].

Although the problem of bidding in electricity markets has already a significant state-of-art, the participation of an EVA in electricity markets justifies a specific research approach due to the functionality to be faced. EVA management has been researched according to the consideration of flow of electric energy relatively to the grid: unidirectional [6], i.e., grid-to-vehicle (G2V); bidirectional [7], i.e., G2V and vehicle-to-grid (V2G). Two lines of research address this management and have been developing computational application for market environments. The first line considers that EVs although random units can be modelled as deterministic ones [8, 9], implying a lack of consideration of uncertainty. Following this line: [8] proposes an optimization algorithm to manage optimal charging and discharging of EVs controlled by aggregators, offering, also, ancillary services in electricity markets; [9] proposes an algorithm for EVs providing unidirectional vehicle-to-grid regulation. The second line considers uncertainty in the EVs [10–16]. Following this line: [11] proposes a stochastic dynamic programming model in order to optimize the EV charging periods and the frequency regulation capacity bids; [12] proposes a stochastic optimization methodology for management of EV and considers the participation of the EVA in both day-ahead market and regulation market; [13] proposes a methodology based in stochastic optimization for aggregators having demand-side resources, e.g., plug-in electric vehicles and distributed generation, including in the formalization the risk management; [14] proposes a stochastic programming model for EVA participation in both day-ahead market and ancillary service markets, including in the formalization equilibrium constraints; [15] proposes a methodology based in robust optimization for the charging strategy of an EVA to reduce the charging costs and taking into account renewable generation; [16] proposes a stochastic scheduling of EVA for participation in energy and ancillary service markets, including in the formalization the consideration of uncertainty in EVs energy requirements and risk management. The state-of-the-art mainly addresses the simulation of EV owners and EVA or market participants managing EV in coordination with other power sources, e.g., renewable energy sources or even conventional power sources based on fossil fuels. This paper follows the second line of research and has contributions to the knowledge of this line on the modeling of: the uncertainty, regarding not only market prices, charging and discharging, but also on driving requirements of the owners of the EVs; the G2V and V2G incorporation for the support information management. The problem is treated as stochastic optimization problem reformulated as a mixed-integer linear programming (MILP) approach in order to benefit from well-established and successful commercial solvers.

2 Relationship to Industrial and Service Systems

The 18th century industrial revolution, said to be the First Industrial Revolution, brought into force the historically never seen before paradigm of industrialization and urban society. Subsequent research and development have delivered scientific evolution triggering technological innovation for industrial and service systems with significative impact into the society. After a said industrial revolution, a new set of technological innovation with significative impact into the society, i.e., expressively changing the society, is categorized as a consequent born age said to be the following industrial revolution. In nowadays a view of a new set of technological innovation for industrial and service systems points into what is said to be the paradigm of a fourth industrial revolution, i.e., Industry 4.0 [17]. Industry 4.0 indorses the interconnection of physical items such as sensors, devices and enterprise assets, interfacing with one each other directly or by the Internet [17–21], allowing Cyber-Physical Systems (CPS) processes. The main characteristics of these processes are decentralization and autonomous behavior [17]. An EVA having EVs G2V and V2G in the new set of innovation is in the scope of a CPS interfaced with information systems with real-time digital platform services. This EVA must be interfaced with urban infrastructures, allowing capabilities of smart grids to schedule the batteries in an environment of smart cities [22]. Demand side involvement in a power system is a functionality linked with smart grids and EV usage can enable active demand side involvement in the operation of power systems. The communication and computation requirements adopted by an EVA must include advanced metering infrastructure (AMI), supervisory control and data acquisition (SCADA), cellular communications and other wireless communication technology [23]. Technological innovation software for the EVA is needed for taking full advantage of EV G2V and V2G, persuading customer for benefits of being with an EVA and consequently favoring the desired rise on the EV market penetration [22].

3 EVA Problem Formulation

Electricity markets normally impose to the participation a requirement of minimum limit of energy for bidding, either affecting bids for selling or for purchasing energy [24, 25]. This requirement implies that power sources or loads that do not satisfy the requirement are not allowed to go into the market, e.g., a singular EV. Then an EVA aggregating EVs is a feasible alternative, acting as a manger intermediator between EV owners and the electricity market. A strong interaction is assumed between EV owners and the respectively EVA to have in due time the information accessibility for taking the most convenient decisions. Hence, the EV owners are assumed to communicate to the EVA the driving requirements of owners and the availability of energy in a time horizon of 24 h. The communication identify the EV owner requirements and flexibility for charge or discharge energy during the 24 h. Then the main objective of the EVA is to optimize the management of the fleet of EVs, taking convenient decisions of buying energy to charge the EVs or of selling energy by discharging the EVs and considering battery degradation due to cycling. This objective is modeled in the context of presenting profitable bids in the day-ahead electricity market by deciding the optimal charging and discharging periods, subjected to requirements of the owners of the EVs. Despite the fact that most of the current handy public charging stations only allows G2V mode of operation, in this paper is assumed that the charging stations also allow V2G operation as the possible future setting to be expected in a smart city context. The objective function for this management can be stated as a cost or as in this paper by the expected profit associated with the bidding in the day-ahead market stated as follows:

$$\sum_{s=1}^{N_S}\sum_{t=1}^{N_T}\sum_{e=1}^{N_{EV}}\frac{1}{N_S}\left(\lambda_{st}^{DA}P_{ste}^{D}-\lambda_{st}^{DA*}P_{ste}^{C}+\zeta E_{ste}^{R}-C_{ste}^{Deg}\right) \tag{1}$$

In (1) $\lambda_{st}^{DA}P_{ste}^{D}$, $\lambda_{st}^{DA*}P_{ste}^{C}$, ζE_{ste}^{R} and C_{ste}^{Deg} are the revenue associated with the selling offer, the cost associated with the purchasing offer, the cost of driving requirements, cost of battery degradation of each EV at scenario s and period t, respectively. N_S, N_T, N_{EV} are the number of scenarios, the number of periods and the number of EVs, respectively. The scenarios are equiprobable ones, i.e., a probability of $1/N_S$ is assumed for each scenario in (1). Considering the driving requirements and the battery size equivalent for the whole EV in the fleet, the EVA can be seen as a manager of one equivalent bulky load/battery. The objective function in (1) is reformulated for convenience and the mathematical programming problem is stated as follows:

$$\max \sum\nolimits_{s=1}^{N_S}\sum\nolimits_{t=1}^{N_T}\frac{1}{N_S}\left(\lambda_{st}^{DA}P_{st}^{D}-\lambda_{st}^{DA*}P_{st}^{C}+\zeta E_{st}^{R}-C_{st}^{Deg}\right) \tag{2}$$

subject to:

$$\underline{P^D}\sigma_{st}^{D}\le P_{st}^{D}\le\overline{P^D}\sigma_{st}^{D}\quad\forall s,\ \forall t\ \text{ and }\ \underline{P^C}\sigma_{st}^{C}\le P_{st}^{C}\le\overline{P^C}\sigma_{st}^{C}\quad\forall s,\ \forall t \tag{3}$$

$$0\le\sigma_{st}^{D}\le\sigma_{st}^{A}\quad\forall s,\ \forall t\ \text{ and }\ 0\le\sigma_{st}^{C}\le\sigma_{st}^{A}\quad\forall s,\ \forall t \tag{4}$$

$$\sigma_{st}^{D}+\sigma_{st}^{C}\le\sigma_{st}^{A}\quad\forall s,\ \forall t \tag{5}$$

$$SoC_{st}=SoC_{st-1}+\frac{\eta^{C}P_{st}^{C}}{\bar{E}}-\frac{P_{st}^{D}}{\bar{E}\eta^{D}}-\frac{E_{st}^{R}}{\bar{E}}\ \text{ and }\ \underline{SoC}\le SoC_{st}\le\overline{SoC} \tag{6}$$

In (2), $\lambda_{st}^{DA}P_{st}^{D}$, $\lambda_{st}^{DA*}P_{st}^{C}$, ζE_{st}^{R} and C_{st}^{Deg} are the scenario s and period t revenue associated with the selling offer, cost associated with an eventual reduced tariff for the purchasing offer, revenue associated with the equivalent total driving requirement and the cost associated with degradation of the batteries of the EV fleet, respectively. In $\lambda_{st}^{DA}P_{st}^{D}$, λ_{st}^{DA} is the day-ahead market price and P_{st}^{D} is the selling offer, i.e., the total energy discharged by the EV fleet. In $\lambda_{st}^{DA*}P_{st}^{C}$, λ_{st}^{DA*} is the eventual reduced tariff in order to favor of the operation in V2G mode and P_{st}^{C} is the purchasing offer associated with the charging of the batteries of the fleet of EVs. In ζE_{st}^{R}, ζ is a fixed tariff between the EVA and the EV owners for the energy due to driving requirements E_{st}^{R}. A compensation to the EV owners for the eventual degradation of the batteries if called to discharge is modeled by the cost C_{st}^{Deg} stated as in [26] as follows:

$$C_{st}^{Deg} = \left|\frac{m}{100}\right|\left(\frac{P_{st}^{C} + P_{st}^{D} - E_{st}^{R}}{\bar{E}}\right)C^{B} \tag{7}$$

In (7) m is the linear approximated slope of the battery life as function of the number of cycles [26], $\overline{E}$ is the EV fleet battery capacity, and C^{B} is the total cost of the EV fleet battery. The EVA because of energy driving requirements reduces the cost of the degradation of the batteries by the subtraction of E_{st}^{R} shown in (7). In (3) and (4), σ_{st}^{D} and σ_{st}^{C} are the binary variables to control the discharging and charging cycles and setting lower and upper bounds $\underline{P^{D}}/\underline{P^{D}}$ and $\overline{P^{D}}/\overline{P^{C}}$ for discharging and charging continuous variables, respectively. In (4) and (5) is constrained the discharging and charging cycles accordingly to the availability variable σ_{st}^{A} stated as follows:

$$\sigma_{st}^{A} = \begin{cases} 1, & EV\ fleet\ \ available \\ 0, & EV\ fleet\ \ unavailable \end{cases} \tag{8}$$

In (4) is set the lower and upper bounds of σ_{st}^{D} and σ_{st}^{C}. Three events can occur to the EV fleet: event 1, discharging ($\sigma_{st}^{A} = 1;\ \ \sigma_{st}^{D} = 1,\ \ \sigma_{st}^{C} = 0$); event 2, charging ($\sigma_{st}^{A} = 1;\ \ \sigma_{st}^{D} = 0,\ \ \sigma_{st}^{C} = 1$); event 3, unavailable ($\sigma_{st}^{A} = 0;\ \ \sigma_{st}^{D} = 0,\ \ \sigma_{st}^{C} = 0$). In (6) is defined the equation of balance of the virtual battery, where SoC_{st}, η^{D} and η^{C} are the state of charge, the discharging efficiency and the charging efficiency, respectively. Also, in (6) is set the lower and upper bounds of SoC_{st}, $\underline{SoC}$ and $\overline{SoC}$, respectively.

4 Scenario Generation and Reduction

Kernel density estimation (KDE) is employed for estimating the probability density curve of day-ahead market prices, availability and driving requirements of the EV fleet managed of the EVA. KDE allows the nonparametric estimation of a density f for a set of observations of a random variable [27]. The kernel density estimator can be stated as follows:

$$\widehat{f_{h}}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - x_{i}}{h}\right) \tag{9}$$

In (9), n is the sample size, x_{i} are random samples from the unknown distribution, h is the bandwidth, and $K(\cdot)$ is the kernel smoothing function. As result of the KDE employment a set of random numbers, the scenarios, are generated from the fitted kernel distribution. In the case study in this paper, a generation of the 1000 scenarios is considered and this number of scenarios is submitted to a K-means clustering algorithm for a convenient reduction, grouping scenarios based on an identification of similarities. The objective is to find groups of scenarios having significant similarity and associate with a group one scenario to be considered as the equivalent scenario in what regards the information needed for taking the convenient decisions [28]. The advantage of the K-means is the simplicity, efficiency and scalability [28]. Let $S = \{S_1 \cdots, S_I\}$ be a set of clusters S_i given by a set of scenarios belonging by clustering to S_i, I the number of clusters and *dist* a chosen metric, providing a way to measure how close two scenarios are. Let the centroid of a cluster S_i be the mean vector, $c_i = \frac{1}{|S_i|}\sum_{y \in S_i} y$, the K-means employs iterative refinement to deliver the set of clusters stated as in [28] as follows:

$$\underset{S}{argmin} \sum_{i=1}^{I} \sum_{y \in S_i} dist(c_i, y)^2 \tag{10}$$

In (10) the final set of clusters is identified by the minimization of the metric, for instance, giving the distance between the day-ahead market prices profiles y and the centroid of the cluster c_i of the day-ahead market prices.

5 Case Study

A bidding strategy of the EVA is accessed with data of Electricity Market of Iberian Peninsula given in [29] and of driving and parking patterns of European drivers given in [30]. The EVA manages 1000 EVs of 25 kWh each one, i.e., 25 MWh. For simplicity, the driving periods of the EVs are assumed to be the same. The cost of each battery is 250 €/kWh and the respective linear approximated slope m is of −0.0013 [26]. The EV distance/consumption ratio is of 6 km/kWh. The maximum charge and discharge power of the EV fleet is of 10 MW. The minimum and maximum SoC of the bulky battery are 20% and 80%, respectively. Considering the uncertainties regarding market prices and EV fleet driving requirements and availabilities, the 1000 scenarios are generated by the KDE. Then, the scenarios are reduced to a set of 10 scenarios for each uncertainty parameter. The original scenarios and the 10 reduced final scenarios of day-ahead market and the driving requirements final scenarios of the EV fleet are shown in Fig. 1.

In Fig. 1 left, the price scenarios show a tendency to have unfavorable values between hour 3 and hour 6 and a tendency to have favorable ones between hour 21 and hour 23. Note that these prices are generated applying the KDE and reduced by K-means clustering. In Fig. 1 right, when the energy requirement is different of 0, means that the EV fleet is unavailable to charge (G2V) or discharge (V2G). When the energy requirement is 0 the EV fleet is not driving, then the EV fleet can operate both in G2V or V2G modes. The result of the EVA management of the fleet of EVs, i.e., the optimal charging/purchasing offers and discharging/selling offers, are shown in Fig. 2.

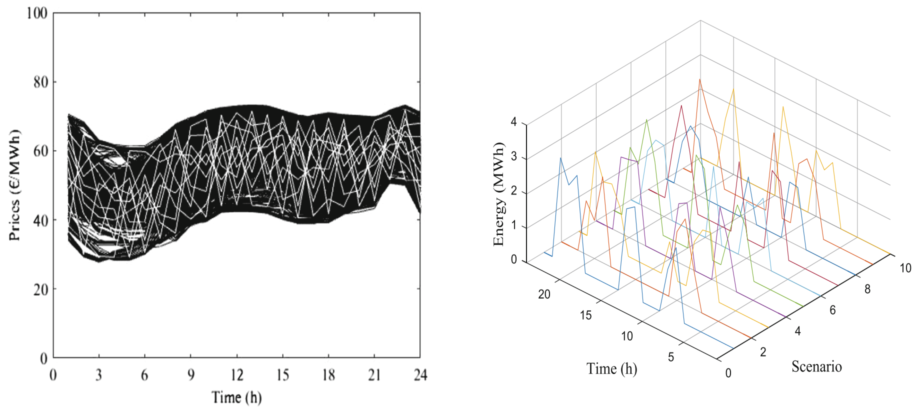


Fig. 1. Left: Day-ahead market prices: black – scenarios, white – reduced scenarios; right: scenarios of driving requirements of the EV fleet.

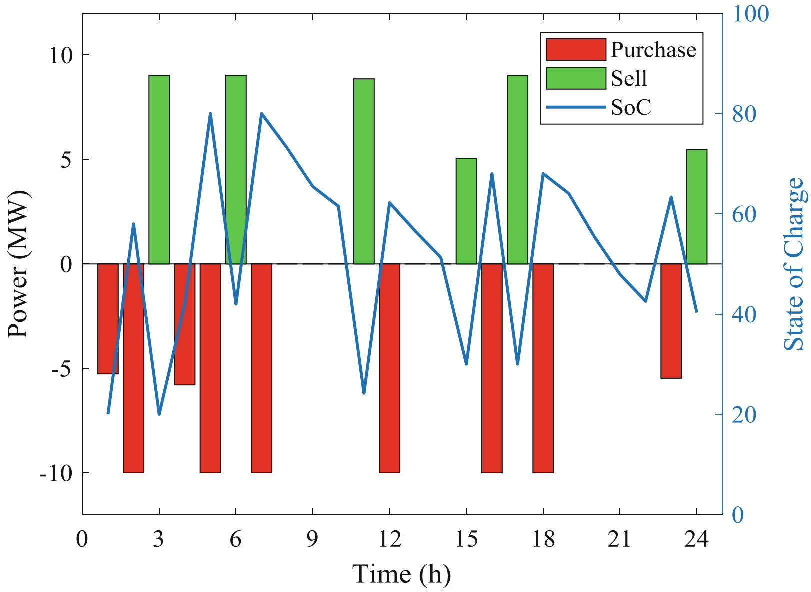


Fig. 2. Day-ahead market bids and SoC of the EV fleet. (Color figure online)

In Fig. 2 the optimal offers to purchase energy, to charge the fleet of EVs, the optimal offers to sell energy by discharging the fleet of EVs, and the SoC of the EV fleet are shown in red, green and blue, respectively. The SoC starts with a value of 20%, i.e., starts with the mandatory minimum value, and varies during the time horizon due to the charge of the batteries, energy consumption and energy sold to the market. The proposed management system in this paper allows the EV owners to be inflexible in what regards the driving patterns. So, the EVA only has power over the periods in which the EV owners are not driving to contribute to the purchase/sell of energy in the day-ahead market. The EVA management is as expected highly affected by the driving periods, which in some cases are periods of high market prices, where selling of energy is advantage, but inflexible requirements do not allow selling energy. This case study allows an EVA instance with an expected profit of € 338 in the day-ahead market, i.e., covering the costs: of charge of the batteries of the fleet, of driving requirements and of battery degradation. But notice that this EVA instance is with an expected profit having a non-null value, if and only if, the EVA is rewarded for discharge energy to the grid.

6 Conclusion

A support management system for an EVA bidding in day-ahead market is proposed in this paper to contribute with technological innovation in a context of smart grids for smart cities, consenting EV G2V and V2G. The problem is formulated as a stochastic mixed-integer linear programming approach. The scenarios of market prices, driving requirements and availabilities of EV are modelled applying a Kernel density estimation. The scenarios are reduced applying the K-means algorithm. A strong interaction between the EV owners and EVA is assumed, but the EVA is subjected to the requirement of taking only control over the periods in which the EV owners are not driving. This requirement is normal imposed by an owner of an EV, i.e., the driving patterns are valued as inflexible ones. The EVA management has possibility to be with an expected profit having a non-null value, since the difference between selling energy in the day-ahead market can be higher than the total cost incurred by the fleet of EV's: costs of charging batteries, of driving requirements and of battery degradation. But this is possible in an environment of an EV fleet operation in G2V or V2G modes not only due to convenient management of the EVA, but also due to the amplitude of the variation on the prices of energy in the day-a-head market.

Acknowledgments. This work is funded by Camões, I.P./Millennium BCP Foundation through the Programa Empresa Promotora da Língua Portuguesa and funded by: European Union through the European Regional Development Fund, included in the COMPETE 2020 (Operational Program Competitiveness and Internationalization) through the ICT project (UID/GEO/04683/2013) with the reference POCI010145FEDER007690; Portuguese Funds through the Foundation for Science and Technology-FCT under the project LAETA 2015–2020, reference UID/EMS/50022/2019; Portuguese Foundation for Science and Technology (FCT) under Project UID/EEA/04131/2013.

References

1. OECD: Reducing transport greenhouse emissions: trends & data. Organization for Economic Co-operation and Development (2010)
2. Lane, B.W., Dumortier, J., Carley, S., Siddiki, S., Clark-Sutton, K., Graham, J.D.: All plug-in electric vehicles are not the same: predictors of preference for a plug-in hybrid versus a battery electric vehicle. Transp. Res. Part D Transp. Environ. **65**, 1–13 (2018)

3. Bakker, S., Trip, J.J.: Policy options to support the adoption of electric vehicles in the urban environment. Transp. Res. Part D **25**, 18–23 (2013)
4. Rigas, E.S., Ramchurn, S.D., Bassilades, N.: Managing electric vehicles in the smart grid using artificial intelligence: a survey. IEEE Trans. Intell. Transport Syst. **16**(4), 1619–1635 (2015)
5. Kempton, W., Tomic, J., Letendre, S., Brooks, A., Lipman, T.: Vehicle-to-grid power: battery, hybrid, and fuel cell vehicles as resources for distributed electric power in California. UCD-ITS-RR-01-03, June 2001
6. Vandael, S., Claessens, B., Hommelberg, M., Holvoet, T., Deconinck, G.: A scalable three-step approach for demand side management of plug-in hybrid vehicles. IEEE Trans. Smart Grid **4**(2), 720–728 (2013)
7. Vaya, M.G., et al.: EV aggregation models for different charging scenarios. In: Proceedings of International Conference on Electricity Distribution (2015)
8. Jain, P., Das, A., Jain, T.: Aggregated electric vehicle resource modelling for regulation services commitment in power grid. Sustain. Cities Soc. **45**, 439–450 (2019)
9. Sortomme, E., El-Sharkawi, M.A.: Optimal charging strategies for unidirectional vehicle-to-grid. IEEE Trans. Smart Grid **2**(10), 131–138 (2011)
10. Baringo, L., Amaro, R.S.: A stochastic robust optimization approach for the bidding strategy of an electric vehicle aggregator. Electr. Power Syst. Res. **146**, 362–370 (2017)
11. Donadee, J., Ilic, M.D.: Stochastic optimization of grid to vehicle frequency regulation capacity bids. IEEE Trans. Smart Grid **5**(2), 1061–1069 (2014)
12. Vagropoulos, S.I., Bakirtzis, A.G.: Optimal bidding strategy for electric vehicle aggregators in electricity markets. IEEE Trans. Power Syst. **28**(4), 4031–4041 (2013)
13. Xu, Z., Hu, Z., Song, Y., Wang, J.: Risk-averse optimal bidding strategy for demand-side resource aggregators in day-ahead electricity markets under uncertainty. IEEE Trans. Smart Grid **8**(1), 96–105 (2017)
14. Wu, H., Shahidehpour, M., Alabdulwahab, A., Abusorrach, A.: A game theoric approach to risk-based optimal bidding strategies for electric vehicle aggregators in electricity markets with variable wind energy resources. IEEE Trans. Sustain. Energy **7**(1), 374–385 (2016)
15. Wei, W., Liu, F., Mei, S.: Charging strategies of EV aggregator under renewable energy generation and congestion: a normalized nash equilibrium approach. IEEE Trans. Smart Grid **7**(3), 1630–1641 (2016)
16. Manijeh, A., Mohammadi, B.-I., Moradi, M.-D., Zare, K.: Stochastic scheduling of aggregators of plug-in electric vehicles for participation in energy and ancillary service markets. Energy **118**, 1168–1179 (2017)
17. Batista, N.C., Melicio, R., Mendes, V.M.F.: Services enabler architecture for smart grid and smart living services providers under industry 4.0. Energy Build. **141**, 16–27 (2017)
18. Batista, N.C., Melicio, R., Matias, J.C.O, Catalão, J.P.S.: Zigbee standard in the creation of wireless networks for advanced metering infrastructures. In: Proceedings of 16th IEEE Melecon (2012)
19. Batista, N.C., Melicio, R., Matias, J.C.O, Catalão, J.P.S.: ZigBee wireless area network for home automation and energy management: field trials and installation approaches. In: Proceedings of 3rd IEEE ISGT Europe, pp. 1–5 (2012)
20. Batista, N.C., Melicio, R., Matias, J.C.O., Catalão, J.P.S.: Photovoltaic and wind energy systems monitoring and building energy management using ZigBee devices within a smart grid. Energy **49**(1), 306–315 (2013)
21. Batista, N.C., Melicio, R., Mendes, V.M.F.: Layered smart grid architecture approach and field tests by ZigBee technology. Energy Convers. Manag. **88**, 49–59 (2014)

22. Ensslen, A., Gnann, T., Jochem, P., Plotz, P., Dustchke, E., Fichtner, W.: Can product service systems support electric vehicle adoption? Transp. Res. Part A (2018). https://doi.org/10.1016/j.tra.2018.04.028
23. Shaukat, N., et al.: A survey on electric vehicle transportation within smart grid system. Renew. Sustain. Energy Rev. **81**, 1329–1349 (2018)
24. Gomes, I.L.R., Melicio, R., Mendes, V.M.F., Pousinho, H.M.I.: Decision making for sustainable aggregation of clean energy in day-ahead market: uncertainty and risk. Renew. Energy **133**, 602–702 (2019)
25. Laia, R., Pousinho, H.M.I., Melicio, R., Mendes, V.M.F.: Bidding strategy of wind-thermal energy producers. Renew. Energy **99**, 673–681 (2016)
26. Sarker, M.R., Dvorkin, Y., Ortega-Vazquez, M.A.: Optimal participation of an electric vehicle aggregator in day-ahead energy and reserve markets. IEEE Trans. Power Syst. **31**(5), 3506–3515 (2016)
27. Arora, S., Taylor, J.W.: Forecasting electricity smart meter data using conditional kernel density estimation. Omega **59**, 47–59 (2016)
28. Viegas, J.L., Vieira, S.M., Melicio, R., Mendes, V.M.F., Sousa, J.M.C.: Classification of new electricity costumers based on surveys and smart metering data. Energy **107**, 804–817 (2016)
29. REE-Red Eléctrica de España (2018). http://www.esios.ree.es/web-publica/
30. Pasaoglu, G., et al.: Driving and parking patterns of European car drivers – a mobility survey. Joint Research Centre, European Commission, European Union (2012)

Energy Control

Integrated System for Traction and Battery Charging of Electric Vehicles with Universal Interface to the Power Grid

Tiago J. C. Sousa(✉), Vítor Monteiro, and João L. Afonso

Centro ALGORITMI, University of Minho,
Campus de Azurém, Guimarães, Portugal
{tsousa, vmonteiro, jla}@dei.uminho.pt

Abstract. This paper proposes an integrated system for traction and battery charging of electric vehicles (EVs) with universal interface to the power grid. In the proposed system, the power electronics converters comprising the traction drive system are also used for the battery charging system, reducing the required hardware, meaning the integrated characteristic of the system. Besides, this interface is universal, since it can be performed with the three main types of power grids, namely: (1) Single-phase AC power grids; (2) Three-phase AC power grids; (3) DC power grids. In these three types of interfaces with the power grid, as well as in the traction drive operation mode, bidirectional operation is possible, framing the integration of this system into an EV in the context of smart grids. Moreover, the proposed system endows an EV with an on-board fast battery charger, whose operation allows either fast or slow battery charging. The main contributes of the proposed system are detailed in the paper, and simulation results are presented in order to attain the feasibility of the proposed system.

Keywords: Electric vehicles · Traction drive systems · Battery charging systems · Smart grids

1 Introduction

Electrical energy is an essential asset for the well-being of the present modern society, whose consumption has been increasing from decade to decade. This increase led to a fossil fuels shortage, as well as augmented levels of atmospheric pollution. In order to diminish the referred consequences, electric vehicles (EVs) appear as a sustainable alternative at the utilization level, which is proved by the growing stringency on the environmental laws concerning the emission of pollutant particles by internal combustion engine vehicles [1, 2].

The suitability of EVs is not only related to the mobility paradigm, but is also related to the energy management in smart grids. Besides performing the normal battery charging operation (grid-to-vehicle – G2V), which should be performed in a rationalized manner to avoid power grid congestion [3–8], an EV can also be employed to deliver energy to the power grid, performing support functionalities (vehicle-to-grid – V2G), which can be accomplished with a bidirectional battery charger, either

Published by Springer Nature Switzerland AG 2019
L. M. Camarinha-Matos et al. (Eds.): DoCEIS 2019, IFIP AICT 553, pp. 355–366, 2019.
https://doi.org/10.1007/978-3-030-17771-3_31

on-board or off-board [9]. Besides V2G, several operation modes for EVs in a context of smart grids can be found in the literature, such as: vehicle-to-home (V2H), where the EV battery charger operates as an isolated voltage source; home-to-vehicle (H2V), where the EV battery charger adapts its charging current in order to level the current consumed in the home where it is plugged-in; and vehicle-for-grid (V4G), where the EV battery charger operates as an active power conditioner, producing reactive power and/or harmonic currents in order to maintain high levels of power quality from the home point of view [10–14]. Moreover, EVs can be applied to interface renewable energy sources with the power grid, which is in accordance with the distributed generation paradigm implicit in smart grids [15, 16].

Besides the battery charging operation, any EV must perform a traction operation in order to control its electric motor speed or torque. Since both consist of bidirectional AC-DC and DC-DC power electronics converters, these two systems can be combined into a single set of power electronics converters, establishing an integrated system for traction and battery charging [17–19]. This approach allows the EV to be equipped with an on-board fast battery charger, capable of operating with power levels that are usually found only in off-board battery chargers, besides reducing the hardware of the EV [20]. Many examples of integrated systems for traction and battery charging can be found in the literature [21–29], either using external inductors or the motor stator windings as the coupling inductors with the power grid. Some approaches use an auxiliary motor instead of external inductors, which is not desirable, since it increases the weight and volume of the system. Besides, the proposed approaches are limited in terms of operating modes and their interface is restricted to a single type of power grid. Taking into account these limitations, this paper proposes an integrated system for traction and battery charging of EVs with a universal interface with the power grid, which means that the battery charging operation can be accomplished with the connection to the three main types of power grids, namely: (1) Single-phase AC power grids; (2) Three-phase AC power grids; (3) DC power grids. Bidirectional operation is possible in these three types of interfaces with the power grid, i.e., the G2V and V2G operation modes are possible in the three power grid types. The traction drive operation mode also comprises bidirectional operation, i.e., the system is able to perform regenerative braking when the EV motor acts as a generator, returning the generated energy back to the batteries.

The rest of the paper is structured as follows. Section 2 describes the relationship of the proposed system with industrial and service systems. Section 3 describes the proposed integrated system for traction and battery charging of EVs with universal interface to the power grid, with simulation results for the referred operation modes. Finally, Sect. 4 outlines the main conclusions of the proposed work.

2 Relationship to Industrial and Service Systems

When equipped with an on-board slow battery charger that comprises bidirectional operation, an EV is capable of operating as an active power conditioner, such as a shunt active power filter, and as an isolated voltage source, such as an uninterruptible power supply (UPS). However, for the typical on-board battery chargers, these functionalities are only valid in low power installations (e.g., domestic installations). As aforementioned, the proposed integrated system is able to operate as an on-board fast battery charger for EVs. Besides, it can operate in bidirectional mode in single-phase AC power grids, three-phase AC power grids and DC power grids. Therefore, since the nominal power of the on-board battery charger is higher than the typical values found in traditional on-board battery chargers, an EV equipped with the proposed system is able to operate as a shunt active power filter or UPS in industrial facilities. These functionalities can provide support to an industrial facility without the need of acquiring dedicated hardware, i.e., a shunt active power filter or a UPS, since the EV can perform the same functionalities when parked in the facility. The compensation of power quality problems related with currents (e.g., harmonic distortion, unbalance and reactive power) is relevant in the sense that industries can suffer large economic losses due to poor levels of power quality, as referred in several studies [30, 31]. Moreover, the operation as UPS is also relevant in an industrial context, since power outages disrupt completely the production of the facility, which can be reflected in economic losses even higher than the previous case. In this way, the proposed integrated system has a relevant role in industrial systems, despite being an electric mobility solution.

3 Proposed System

Fig. 1 shows a block diagram of the proposed integrated system for traction and battery charging of EVs with universal interface to the power grid. As it can be seen, the proposed system is able to interface the EV batteries with the EV motor, a single-phase

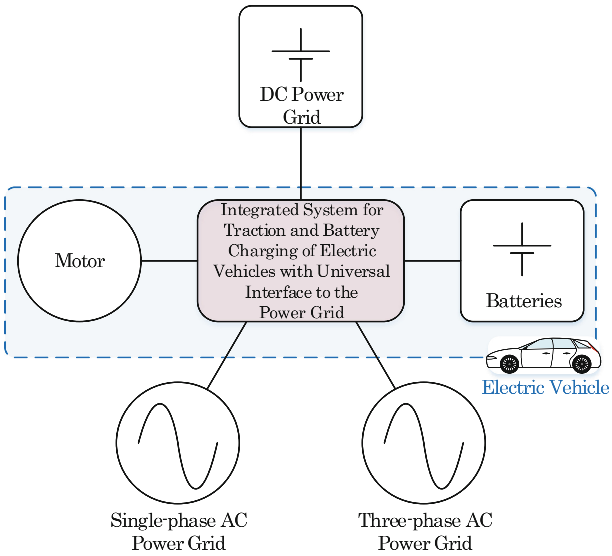


Fig. 1. Block diagram of the proposed integrated system for traction and battery charging of EVs with universal interface to the power grid.

AC power grid, a three-phase AC power grid and a DC power grid, keeping in consideration that a single set of power electronics converters is used for all the operation modes. In order to validate the feasibility of the proposed system, simulation results were obtained in the software PSIM v9.1 and are analyzed in this section, which is divided in four subsections: (1) Traction drive system; (2) Interface with a three-phase AC power grid; (3) Interface with a single-phase AC power grid; (4) Interface with a DC power grid. The bidirectional operation of the system is analyzed in all four cases. In the scope of this paper, only the G2V and V2G operation modes are analyzed regarding the interface of the system with the different types of power grid. The battery voltage was established as 200 V for all cases, with an internal resistance of 0.1 Ω.

3.1 Traction Drive System

Figure 2 shows a block diagram of the integrated system in the traction drive operation mode. In this operation mode, the integrated system establishes an interface between the EV batteries and the EV motor. During the normal traction operation, the integrated system uses the energy stored in the batteries and transfers it to the motor, controlling its speed or torque to perform the EV movement. On the other hand, when the EV experiences a downhill, or when it decelerates, the electric motor acts as a generator. The integrated system detects this transition and starts the regenerative braking operation mode, which consists of transferring the energy generated by the motor back to the batteries, extending the EV range.

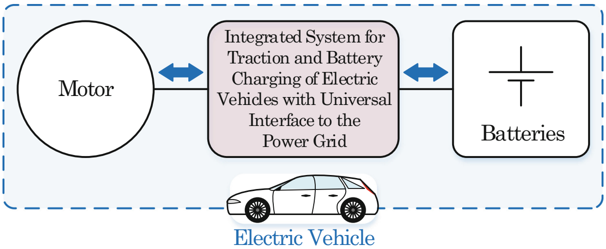


Fig. 2. Block diagram of the proposed system in traction drive operation mode.

Figure 3 shows the operation of the integrated system in the traction drive operation mode, illustrating the two aforementioned cases. This figure shows the two cases in terms of currents in the motor stator windings (i_a, i_b, i_c), the rotor mechanical speed (n_m), the developed torque (T_m) and the battery current (i_{BAT}). It should be mentioned that although the motor used in the simulation is an induction motor, the proposed system is not confined to a certain motor type. The implemented controller for this operation mode was the field-oriented control (FOC). In order to control the battery current, a deadbeat predictive control was used.

The simulation results of the normal traction operation of the integrated system can be seen in Fig. 3(a). For this operation, a reference mechanical speed of 500 rpm was established for the motor, connected to a constant torque load of 100 Nm. It can be seen that the mechanical speed increases steadily until the desired reference value, with the motor torque matching the load torque in steady-state. Also, the current i_{BAT} has a negative value, meaning that the batteries are discharging in order to supply energy to the motor.

Figure 3(b) shows the system operation during regenerative braking for the same reference speed of 500 rpm. In this case, a constant torque load of −100 Nm was used, with the negative sign meaning that the load drives the motor. It can be seen that the mechanical speed presents a similar behavior than the previous case. On the other hand, the torque developed by the motor is positive in the beginning, i.e., while the motor speed does not reach its reference, being negative in steady-state due to the selected value of −100 Nm, i.e., operation as a generator. Accordingly, the current i_{BAT} is positive in this case, meaning that the batteries are being charged with the energy generated by the motor. Moreover, in both modes is guaranteed the operation with sinusoidal, balanced stator currents, as well as low ripple in the mechanical speed, torque and battery current.

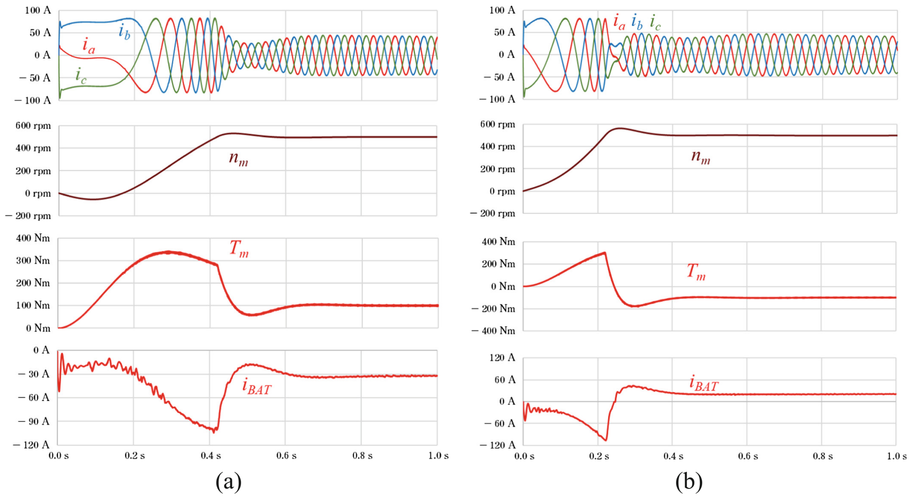


Fig. 3. Simulation results of the proposed system in the traction drive operation mode: (a) Normal traction operation; (b) Regenerative braking operation.

3.2 Interface with a Three-Phase AC Power Grid

Figure 4 shows a block diagram of the integrated system during the interface with a three-phase AC power grid. In this operation mode, the integrated system establishes an interface between the EV batteries and a three-phase AC power grid, enabling the charging or discharging of the batteries. During the battery charging operation (G2V), the integrated system absorbs energy from the power grid, with sinusoidal, balanced currents and high power factor, and transfers it to the batteries, with controlled current or voltage. On the other hand, during the battery discharging operation (V2G), the integrated system uses the energy stored in the batteries, with controlled current or power, and injects it into the power grid, with sinusoidal, balanced currents and high power factor, preserving the power quality of the electrical installation.

Figure 5 depicts the operation of the integrated system interfacing a 400 V, 50 Hz three-phase AC power grid, illustrating the two aforementioned operation modes. This figure shows the two cases in terms of voltages (v_a, v_b, v_c) and currents in the power grid (i_a, i_b, i_c) and the battery current (i_{BAT}). A deadbeat predictive control was used for the current control both at the grid side and at the batteries side.

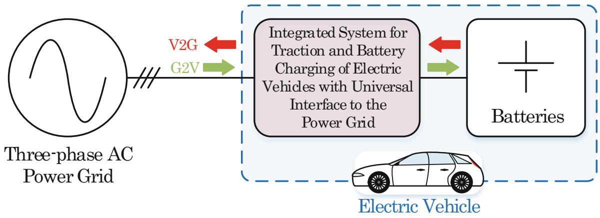


Fig. 4. Block diagram of the proposed system interfacing with a three-phase AC power grid.

The simulation results of the battery charging (G2V) operation of the integrated system can be seen in Fig. 5(a). For this operation, a battery reference current of 100 A was established, whose high value corresponds to a fast battery charging operation (20 kW). In this case, the current i_{BAT} has a positive value, meaning that the batteries are being charged. Furthermore, the grid currents are sinusoidal independently of the

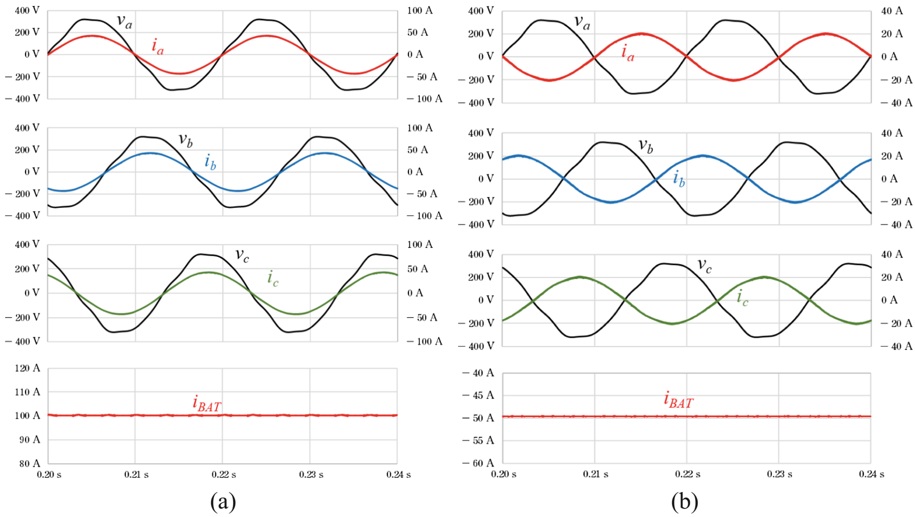


Fig. 5. Simulation results of the proposed system during the interface with a three-phase AC power grid: (a) G2V operation; (b) V2G operation.

grid voltages, presenting a total harmonic distortion (THD) of 0.8% even with a THD of 3.7% in the grid voltages. Besides a low THD, these currents are balanced and each of them is in phase with the respective grid voltage, achieving a unitary power factor.

Figure 5(b) shows the system operation for the battery discharging (V2G) mode. In this case, a battery reference current of −50 A was used, with the negative sign meaning that the batteries are being discharged. As it can be seen, the current i_{BAT} follows the mentioned reference value. On the other hand, the grid currents are in phase opposition with the respective voltages, meaning that the integrated system is operating as a power source, injecting energy into the power grid. As in the previous case, the grid currents are sinusoidal, with a THD of 1.8%, and balanced. Therefore, both the G2V and V2G operation modes are performed assuring high levels of power quality for the three-phase AC power grid.

3.3 Interface with a Single-Phase AC Power Grid

Figure 6 shows a block diagram of the integrated system during the interface with a single-phase AC power grid. In this case, the integrated system establishes an interface between the EV batteries and a single-phase AC power grid, enabling the charging or discharging of the batteries. Similarly to the previous operation mode, during the battery charging operation (G2V), the integrated system absorbs energy from the power grid with a sinusoidal current in phase with the grid voltage, and transfers it to the batteries, with controlled current or voltage. On the other hand, during the battery discharging operation (V2G), the integrated system uses the energy stored in the batteries with controlled current or power, and injects it into the power grid via a sinusoidal current in phase opposition with the grid voltage, preserving the power quality of the electrical installation.

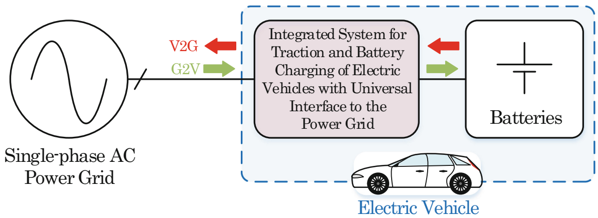


Fig. 6. Block diagram of the proposed system interfacing with a single-phase AC power grid.

Figure 7 depicts the operation of the integrated system interfacing a 230 V, 50 Hz single-phase AC power grid, illustrating the two aforementioned operation modes. This figure shows the two cases in terms of voltage (v_g) and current in the power grid (i_g) and the battery current (i_{BAT}). Similarly to the three-phase AC power grid interface, a deadbeat predictive current control was used both at the grid side and at the batteries side in order to control these currents.

The simulation results of the battery charging (G2V) operation of the integrated system can be seen in Fig. 7(a). For this operation, a battery reference current of 30 A was established, corresponding to a slow battery charging operation (6 kW). Accordingly, the current i_{BAT} has a positive value, meaning that the batteries are being charged. Moreover, the current i_g is sinusoidal independently of the voltage v_g and is in phase with the latter, achieving a practically unitary power factor.

Figure 7(b) shows the system operation for the battery discharging (V2G) mode. In this case, a battery reference current of −10 A was used, with the negative sign meaning that the batteries are being discharged. As it can be seen, the current i_{BAT} follows the mentioned reference value. On the other hand, the current i_g is sinusoidal and in phase opposition with the voltage v_g, meaning that the integrated system is operating as a power source, injecting energy into the power grid and, once again, with a unitary power factor. Therefore, as verified in the interface with a three-phase AC power grid, both the G2V and V2G operation modes are performed assuring high levels of power quality for the single-phase AC power grid.

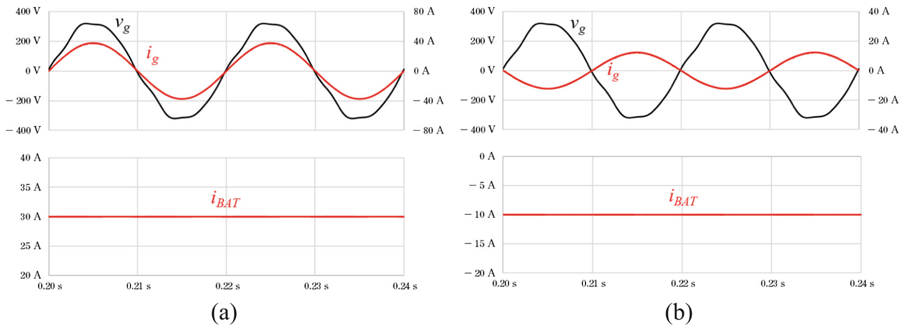


Fig. 7. Simulation results of the proposed system during the interface with a single-phase AC power grid: (a) G2V operation; (b) V2G operation.

3.4 Interface with a DC Power Grid

Figure 8 shows a block diagram of the integrated system during the interface with a DC power grid. In this case, the integrated system establishes an interface between the EV batteries and a DC power grid, enabling the charging or discharging of the batteries. Similarly to the previous operation modes, during the battery charging operation (G2V), the integrated system absorbs energy from the power grid and transfers it to the batteries, with controlled current or voltage. On the other hand, during the battery discharging operation (V2G), the integrated system uses the energy stored in the batteries, with controlled current or power, and injects it into the power grid, with controlled current. The main difference between this interface and the previous ones is the power quality requirements, provided that there is no reactive power in DC. Therefore, the main challenge of the system is to assure low levels of current ripple in both the grid current and the battery current. The current controllers are also of the deadbeat predictive type at both sides.

Figure 9 depicts the operation of the integrated system interfacing a 48 V DC power grid, illustrating the two aforementioned operation modes. This figure shows the grid current (i_{DC}) and the battery current (i_{BAT}). The possibility of connecting the proposed system to a low-voltage DC power grid is attractive in the future scenario of DC homes, being possible to charge the EV batteries from a power grid with a lower voltage than the batteries voltage.

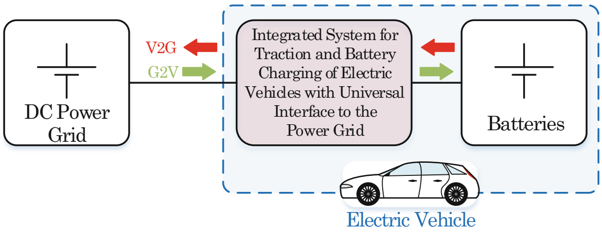


Fig. 8. Block diagram of the proposed system interfacing with a DC power grid.

The simulation results of the battery charging (G2V) operation of the integrated system can be seen in Fig. 9(a). For this operation, a battery reference current of 100 A was established, corresponding to a fast battery charging operation (20 kW). Accordingly, the current i_{BAT} has a positive value, meaning that the batteries are being charged. The current i_{DC} has also a positive value, since the DC power grid provides energy to the batteries. The value of the current i_{DC} is relatively high (437 A) due to the voltage step-up operation (from 48 V to 200 V), so that the power is practically the same both in the grid side and in the batteries side.

Figure 9(b) shows the system operation for the battery discharging (V2G) mode. In this case, a battery reference current of −10 A was used, with the negative sign meaning that the batteries are being discharged. As it can be seen, the current i_{BAT} follows the mentioned reference value. Similarly, the current i_{DC} is also negative, meaning that the DC power grid is absorbing energy from the batteries. As in the previous case, the current i_{DC} is higher than the current i_{BAT} due to the voltage step-up operation of the integrated system.

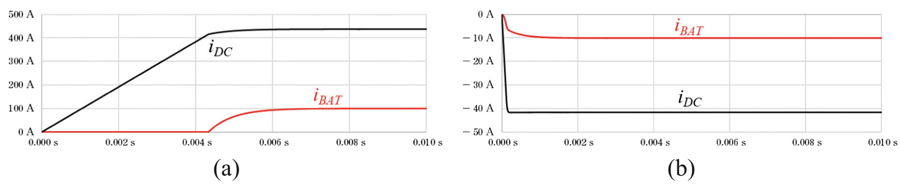


Fig. 9. Simulation results of the proposed system during the interface with a DC power grid: (a) G2V operation; (b) V2G operation.

4 Conclusions

This paper proposed an integrated system for traction and battery charging of electric vehicles (EVs) with a universal interface with the power grid. The proposed system can perform EV battery charging, as well as energy injection into the power grid, through the connection with three main types of power grids (hence the designation of universal), namely: (1) Single-phase AC power grids; (2) Three-phase AC power grids; (3) DC power grids. The traction drive operation mode also comprises bidirectional operation, i.e., the system is able to perform regenerative braking when the EV motor acts as a generator, returning energy back to the EV batteries. The proposed integrated system was validated through computer simulations for the traction drive system and for the three distinct types of interface with the power grid. In all the referred operation modes, bidirectional operation was verified, which is an important contribution to the smart grids, and both slow and fast battery charging modes were addressed. Besides, the operation as a shunt active power filter and as an uninterruptible power supply (UPS) is also possible. Hence, besides having advantages in terms of multi-functionalities, the proposed system offers additional operation modes that are not considered in a traditional EV battery charging system.

Regarding future work, besides validating the other possible operation modes at a simulation level (shunt active power filter and UPS), the design phase will be a critical aspect of the developed work, since there are size and weight constraints for a system that has to be installed in an EV. Provided that external inductors will be needed to interface the system with the power grid, several aspects should be taken into consideration, such as the power semiconductor approach, switching frequency and inductance values, in order to accomplish a relatively compact solution for power levels of 50 kW or higher, which are the typical power values for an EV.

Acknowledgement. This work has been supported by COMPETE: POCI-01-0145-FEDER-007043 and FCT – Fundação para a Ciência e Tecnologia within the Project Scope: UID/CEC/00319/2013. This work has been supported by FCT within the Project Scope DAIPESEV – Development of Advanced Integrated Power Electronic Systems for Electric Vehicles: PTDC/EEI-EEE/30382/2017. Mr. Tiago Sousa is supported by the doctoral scholarship SFRH/BD/134353/2017 granted by the Portuguese FCT agency. This work is part of the FCT project 0302836 NORTE-01-0145-FEDER-030283.

References

1. Milberg, J., Schlenker, A.: Plug into the future. IEEE Power Energy Mag. **9**(1), 56–65 (2011)
2. Gearhart, C., Breitenbach, A.: Connectivity and convergence: transportation for the 21st century. IEEE Electrification Mag. **2**(2), 6–13 (2014)
3. Torres-Sanz, V., Sanguesa, J.A., Martinez, F.J., Garrido, P., Marquez-Barja, J.M.: Enhancing the charging process of electric vehicles at residential homes. IEEE Access **6**, 22875–22888 (2018)

4. Teng, J.H., Liao, S.H., Wen, C.K.: Design of a fully decentralized controlled electric vehicle charger for mitigating charging impact on power grids. IEEE Trans. Ind. Appl. **53**(2), 1497–1505 (2017)
5. Knezovic, K., Martinenas, S., Andersen, P.B., Zecchino, A., Marinelli, M.: Enhancing the role of electric vehicles in the power grid: field validation of multiple ancillary services. IEEE Trans. Transp. Electrification **3**(1), 201–209 (2017)
6. Abousleiman, R., Scholer, R.: Smart charging: system design and implementation for interaction between plug-in electric vehicles and the power grid. IEEE Trans. Transp. Electrification **1**(1), 18–25 (2015)
7. de Hoog, J., Alpcan, T., Brazil, M., Thomas, D.A., Mareels, I.: Optimal charging of electric vehicles taking distribution network constraints into account. IEEE Trans. Power Syst. **30**(1), 365–375 (2015)
8. Vagropoulos, S.I., Kyriazidis, D.K., Bakirtzis, A.G.: Real-time charging management framework for electric vehicle aggregators in a market environment. IEEE Trans. Smart Grid **7**(2), 948–957 (2015)
9. Kempton, W., Tomić, J.: Vehicle-to-grid power implementation: from stabilizing the grid to supporting large-scale renewable energy. J. Power Sources **144**(1), 280–294 (2005)
10. Monteiro, V., Pinto, J.G., Afonso, J.L.: Operation modes for the electric vehicle in smart grids and smart homes: present and proposed modes. IEEE Trans. Veh. Technol. **65**(3), 1007–1020 (2016)
11. Liu, C., Chau, K.T., Wu, D., Gao, S.: Opportunities and challenges of vehicle-to-home, vehicle-to-vehicle, and vehicle-to-grid technologies. Proc. IEEE **101**(11), 2409–2427 (2013)
12. Monteiro, V., Exposto, B., Ferreira, J.C., Afonso, J.L.: Improved Vehicle-to-Home (iV2H) operation mode: experimental analysis of the electric vehicle as off-line UPS. IEEE Trans. Smart Grid **8**(6), 2702–2711 (2017)
13. Kesler, M., Kisacikoglu, M.C., Tolbert, L.M.: Vehicle-to-grid reactive power operation using plug-in electric vehicle bidirectional offboard charger. IEEE Trans. Industr. Electron. **61**(12), 6778–6784 (2014)
14. Hou, R., Emadi, A.: Applied integrated active filter auxiliary power module for electrified vehicles with single-phase onboard chargers. IEEE Trans. Power Electron. **32**(3), 1860–1871 (2017)
15. Monteiro, V., Pinto, G., Afonso, J.L.: Experimental validation of a three-port integrated topology to interface electric vehicles and renewables with the electrical grid. IEEE Trans. Ind. Inform. **14**(6), 2364–2374 (2018)
16. Ansari, J., Gholami, A., Kazemi, A., Jamei, M.: Environmental/economic dispatch incorporating renewable energy sources and plug-in vehicles. IET Gener. Transm. Distrib. **8**(12), 2183–2198 (2014)
17. Rippel, W.: Integrated traction inverter and battery charger apparatus, US4920475A (1990)
18. Rippel, W., Cocconi, A.: Integrated motor drive and recharge system, US5099186A (1992)
19. Cocconi, A.: Combined motor drive and battery charger system, US5341075A (1994)
20. Yilmaz, M., Krein, P.T.: Review of battery charger topologies, charging power levels, and infrastructure for plug-in electric and hybrid vehicles. IEEE Trans. Power Electron. **28**(5), 2151–2169 (2013)
21. Thimmesch, D.: An SCR inverter with an integral battery charger for electric vehicles. IEEE Trans. Ind. Appl. **IA-21**(4), 1023–1029 (1985)
22. Solero, L.: Nonconventional on-board charger for electric vehicle propulsion batteries. IEEE Trans. Veh. Technol. **50**(1), 144–149 (2001)
23. Tang, L., Su, G.-J.: A low-cost, digitally-controlled charger for plug-in hybrid electric vehicles. In: 2009 IEEE Energy Conversion Congress and Exposition, pp. 3923–3929 (2009)

24. Haghbin, S., Lundmark, S., Alakula, M., Carlson, O.: Grid-connected integrated battery chargers in vehicle applications: review and new solution. IEEE Trans. Ind. Electron. **60**(2), 459–473 (2013)
25. Chang, H.C., Liaw, C.M.: Development of a compact switched-reluctance motor drive for EV propulsion with voltage-boosting and PFC charging capabilities. IEEE Trans. Veh. Technol. **58**(7), 3198–3215 (2009)
26. Subotic, I., Bodo, N., Levi, E.: An EV drive-train with integrated fast charging capability. IEEE Trans. Power Electron. **31**(2), 1461–1471 (2016)
27. Liaw, C., Chang, H.: An integrated driving/charging switched reluctance motor drive using three-phase power module. IEEE Trans. Industr. Electron. **58**(5), 1763–1775 (2010)
28. Kim, D.H., Kim, M.J., Lee, B.K.: An integrated battery charger with high power density and efficiency for electric vehicles. IEEE Trans. Power Electron. **32**(6), 4553–4565 (2017)
29. Subotic, I., Bodo, N., Levi, E.: Single-phase on-board integrated battery chargers for EVs based on multiphase machines. IEEE Trans. Power Electron. **31**(9), 6511–6523 (2016)
30. Chapman, D.: The Cost of Poor Power Quality. Power Quality Application Guide, no. 0b, p. 8 (2001). http://www.cda.org.uk
31. McGranaghan, M., Roettger, B.: Economic evaluation of power quality. IEEE Power Eng. Rev. **22**(2), 8–12 (2002)

Auxiliary Digital Control Unit for Capacitor Banks

Patrícia Nunes[1,4], Mafalda Seixas[1,4], Rui Melicio[1,2](✉), and Victor M. F. Mendes[1,3,4]

[1] Escola de Ciências e Tecnologia, Departamento de Física, ICT, Universidade de Évora, Évora, Portugal
ruimelicio@gmail.com

[2] IDMEC, Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal

[3] CISE, Electromechatronic Systems Research Centre, Universidade da Beira Interior, Covilhã, Portugal

[4] Instituto Superior of Engenharia de Lisboa, Lisbon, Portugal

Abstract. This paper addresses the conception of a prototype based in an implementation of a microcontroller proposed as an alternative solution to programmable logic controllers for controlling a capacitor bank. The control considered is in accordance with the Portuguese Energy Services Regulatory Authority. The prototype uses information concerned with the date and time for the accomplishment of industrial schedules. The equipment for the prototype is implemented on a printed circuit board designed and tested for industrial application and is in accordance to current standards for harmonic disturbances.

Keywords: Power factor compensation · Capacitor bank · Prototype · ERSE schedules

1 Introduction

Industrial utilities loads are usually dominated by asynchronous electric motors requiring a magnetizing field, implying that a component of power, known as reactive power, must be supplied. The reactive power is associated with an oscillating energy flow from the electric grid to the motor and vice-versa. This flow carries energy to be supplied to the motor during a part of a cycle and returned to the electric grid during another part of the cycle. The reactive power is not converted into mechanical power in the motor but needed for the accomplishment of the desired conversation of energy. If excessive reactive power is supplied by electric grid, the grid is loaded with too much flow of energy unnecessary for delivering work. More losses of energy into the grid are to be expected, mainly due to the Joule effect, than the one strictly associated with the electric energy for conversion into mechanical energy. So, penalty costs are currently imposed to industrial utilities connected to the electric grid to promote the mitigation of the impact of the flow of reactive power into the grid [1]. The Energy Services of the electric grid in Portugal is an activity regulated by the Energy Services Regulatory Authority (ERSE), offering a range of tariff options, through energy levels and schedules for the use of Energy Services. This offer encompasses the purposes to

Published by Springer Nature Switzerland AG 2019
L. M. Camarinha-Matos et al. (Eds.): DoCEIS 2019, IFIP AICT 553, pp. 367–375, 2019.
https://doi.org/10.1007/978-3-030-17771-3_32

improve energy efficiency usage within the electric grid by the mitigation of the flow of reactive power. Power factor of an installation is defined as the ratio between the active power and the apparent power required by the loads connected to the installation. Power factor of an electrical installation can be seen as a measure of how efficiently the flow of electrical power supplied carries the delivering of the active power associated with electric energy need to be converted into the appliances [2, 3].

Power factor compensation must be considered by industrial utilities to comply with regulation to preserve within the established limits the reactive power supplied by the electric grid. So, contributing to improve energy efficiency usage within the electric grid means as much as possible avoiding penalty costs due to an excessive supply of reactive power. Several solutions are in used for power factor compensation systems, being a predominant one based on the installation of a capacitor bank (CB). Industrial timetables on the use of the Energy Services are accomplishment through equipment that contains internal clocks, allowing to perform actions in a suitable predetermined schedule. Some of the industrial equipment using Energy Services on timetable are hardware based in programmable controllers (PLCs), for instance, some of the power factor compensation systems. Manufacturers of these power factor compensation systems are not necessarily manufacturing the PLCs. So, for these manufacturers compatibility must be the considered in the design phase of the power compensation systems, for instance, trade-offs are needed due to the interfacing with commercial PLCs.

This paper proposes an alternative solution to control a CB, replacing the PLC by a microcontroller in the equipment of the power factor compensation systems, accordingly with ERSE. The alternative solution is a contribution for manufacturer of the power factor compensation system giving more autonomous design over the equipment and added proprietary on the final solution. The alternative solution is implemented in the developed prototype and tested in accordance with industrial use.

2 Relationship to Industrial and Service Systems

Industrial Revolutions have led to new human paths and a latest industrial revolution, i.e., the Industry 4.0, is believed to be on the way of driving a novel path. Industry 4.0 refers to a further evolving stage in the organization and management of the entire value chain process involved in manufacturing industry. Industry 4.0 promotes the connection of physical systems such as sensors, actuators infrastructures and enterprise assets, both to each other and to the Internet [4, 5], i.e., Cyber-Physical Systems (CPS). The main characteristics are decentralization and the autonomous behavior of the production process, towards smart factory and smart industry. A sustainable-oriented decentralized organization in a smart factory focuses on the efficient allocation of products, materials and energy. This concept towards holistic resource efficiency is being described as one of the essential advantages of Industry 4.0 [6–10]. Energy management and energy efficiency have a core roll Industry 4.0. The regulations and specific power needs have an increasing importance to drive the smart industry. Motivated by a mix of environmental factors, cost challenges and regulations the attention for pro-active energy consumption capabilities has never been higher. In the light of the current energy scenario, industrial clients who are required to follow the regulations have a major interesting in energy efficiency. The proposed solution in this paper provides an alternative design for power factor compensation, which is one of the concerns regarding energy efficiency. This solution based on microcontrollers is foreseen to progress in the sense of being equipping with communication devices that allow an easy integration into a smart industry, i.e., connected with the internet by CPS. This integration comes in the sense of providing a means for the intelligent energy allocation, which is a goal of the industry 4.0.

3 Prototype Conception

The prototype controls the switching on/off from a CB (from the power supply of the varmeter controller), and/or controls the switching of a step capacitor according to predetermined schedules. The prototype is equipped with a varmeter relay (VR) and satisfies the following requirements of: withstanding electromagnetic interference, harmonic disturbances, and thermal variations due to being placed in the neighborhood of the VR; withstanding voltages from 350 VAC to 450 VAC; withstanding environmental testing of avionics hardware [11]; having three relay outputs, 5 A and 400 VAC, that can be reconfigured to in landing the CB or to force a CB step into service; having two inputs, one time selector/viewer to allow the user to select schedules and a display input/output status and a communication port. A photo of the prototype final assemble is shown in Fig. 1.

Fig. 1. Prototype final assemble.

The hardware of the equipment is implemented in a designed electronic board in accordance with the architecture showed in Fig. 2.

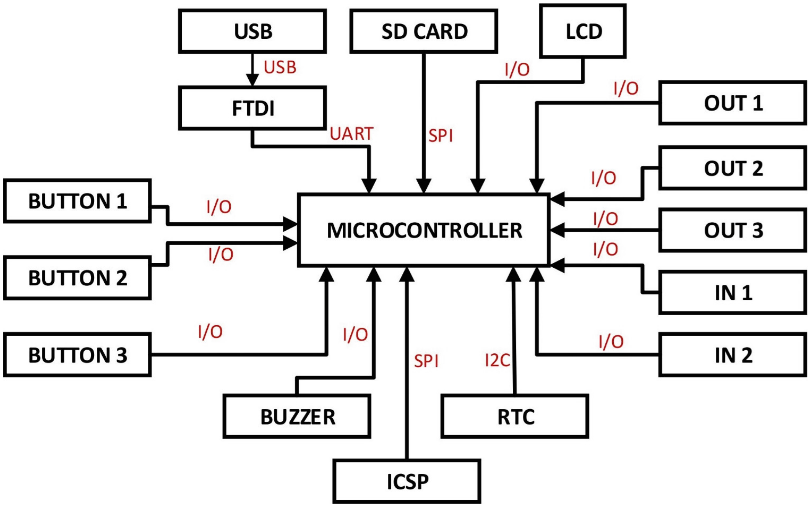


Fig. 2. Electronic board architecture.

The main purpose of the microcontroller is satisfaction of the ERSE requirements on reactive power, accordingly to time tables. A Real-Time Clock (RTC) is used instead of the internal clock of the microcontroller, due to having higher precision and the lack of needing an external power supply. The prototype also has: one Secure Digital (SD) card to save the user decisions; one Liquid Crystal Display (LCD); buttons to visualize and interact with the user; one buzzer for warnings; one Universal Serial Bus (USB) for input and output. Finally, the following protocols are required for:

- The Peripheral Interface (SPI) used by microcontrollers to communicate with one or more peripherals components over short distances in master-slave model. For the communication to coexist is normally required in all the devices four signal lines: Master In Slave Out (MISO), Master Out Slave In (MOSI), Serial Clock (SCK) and Chip Select (CS) [12];
- The Inter-Integrated Circuit (I2C) also working with the master-slave model. This protocol contains their own pins to establish communication, the Serial Data (SDA) and the Serial Clock (SCL). The SDA is the pin that transfers the data and receives it (bi-directional), while SCL controls the timing between the devices, to ensure a reliable communication from the SDA [12];
- The Universal Serial Bus (USB), a simple protocol that allows the transfer of data between computers and peripheral devices. On the computer is installed the host controller, while in the peripheral is installed the device controller. The host carries the polling to allow communication with all devices, authorizing communication only if the bandwidth has less than 90% occupation. The host uses the serial mode

communication with four lines, power, ground and a pair of twisted wires with differential voltage [13];

- The Universal Asynchronous Receiver/Transmitter (UART) protocol, having an asynchronous communication, works with two pins: the transmission (Tx) and the receiving (Rx) [14]. A Future Technology Devices International (FTDI) is used to convert the USB protocol into UART to allow the communication between the microcontroller and the USB peripheral.

The printed circuit board (PCB) with the prototype is showed in Fig. 3.

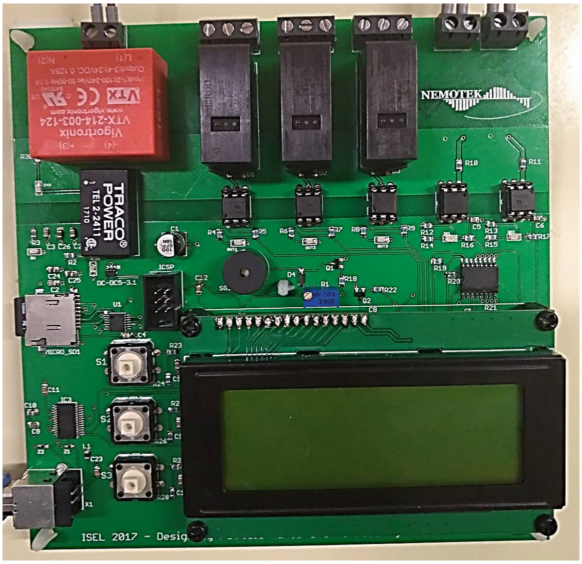


Fig. 3. Prototype in the PCB.

In Fig. 3 PCB design shown is made through the use of the software Eagle CAD and the components are implemented and the values computed to ensure the correct performance of the system, for instance, capacitors are added next to each power pin. The schematic for the bypass capacitors for the microcontroller is showed in Fig. 4.

The schematic for the low-pass filter for one input is showed in Fig. 5.

In Fig. 4 the capacitors implement the bypass operation used to reduce high current frequencies in a high impedance path, diverging the current into another path using a capacitor, discarding current noise that usually exists in electronic circuit feeds, preventing the injection of unwanted signals into the system. In Fig. 5 is shown the low-pass filters used at the input buttons to allow low frequencies and attenuating frequencies with higher than the cut frequency. The cut frequency (fc) is selected to be higher than the expected frequency at the input button and simultaneously lower than the lower frequency used in all the system. The lowest frequency present in the system is the PWM, which has a default value the 490 Hz [12]. Bypass capacitors of 0.1 μF are chosen in accordance with usual design practice [15, 16].

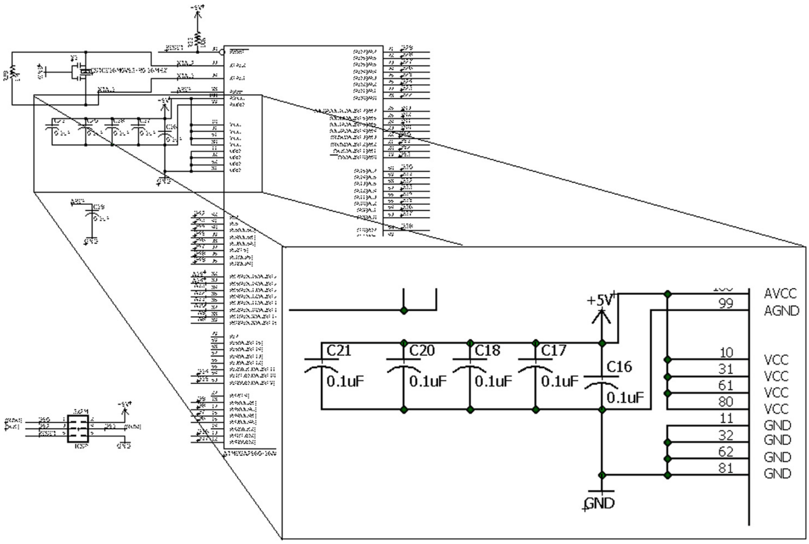


Fig. 4. Bypass capacitors for the microcontroller.

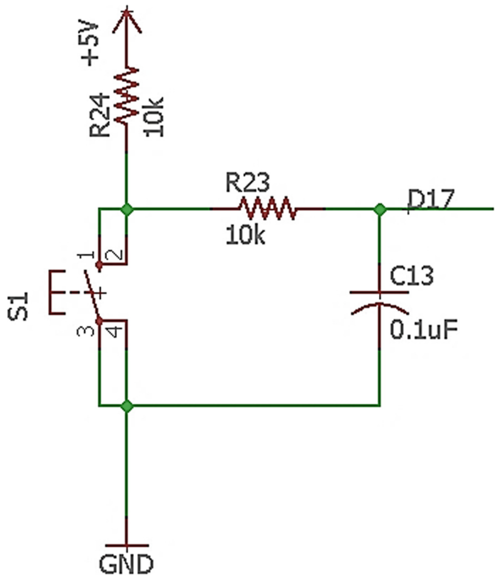
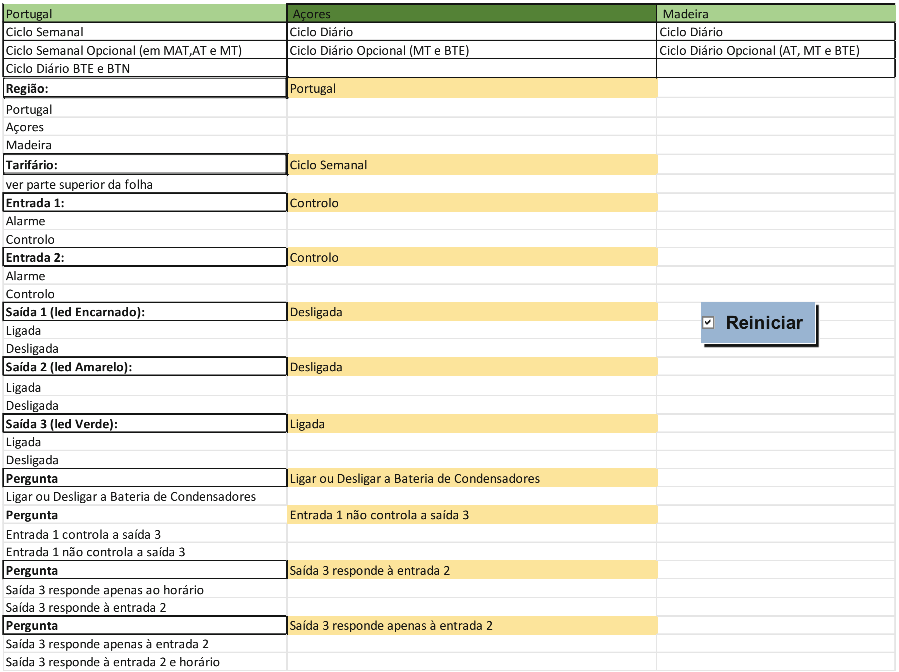


Fig. 5. Low-pass filter for an input.

The software algorithm uses the arduino platform, with C/C++ language and is organized in several routines, to perform specific tasks. An essential part for a feasible prototype is the approach used as an interface between the user and the prototype. A SD card is used for this approach to store the user information about the decisions to be carried out by the system. So, the user information is stored in the SD card in an Excel document that allows the user to input the information. This Excel is used to create several .txt files to be read by the equipment to implement the schedules. As a user alternative interface, and LCD and manipulated buttons are available. The information entered by the manipulated buttons is likewise stored in the SD card.

4 Excel Sheets and Buttons

The Excel document is design with three sheets as follows: the first is for the configuration of the user decisions; the second is for the holidays according to the geographical implementation of the equipment; the third is for the desired schedules. The sheet for ERSE configuration is showed in Fig. 6.



Portugal	Açores	Madeira
Ciclo Semanal	Ciclo Diário	Ciclo Diário
Ciclo Semanal Opcional (em MAT,AT e MT)	Ciclo Diário Opcional (MT e BTE)	Ciclo Diário Opcional (AT, MT e BTE)
Ciclo Diário BTE e BTN		
Região:	Portugal	
Portugal		
Açores		
Madeira		
Tarifário:	Ciclo Semanal	
ver parte superior da folha		
Entrada 1:	Controlo	
Alarme		
Controlo		
Entrada 2:	Controlo	
Alarme		
Controlo		
Saída 1 (led Encarnado):	Desligada	**Reiniciar**
Ligada		
Desligada		
Saída 2 (led Amarelo):	Desligada	
Ligada		
Desligada		
Saída 3 (led Verde):	Ligada	
Ligada		
Desligada		
Pergunta	Ligar ou Desligar a Bateria de Condensadores	
Ligar ou Desligar a Bateria de Condensadores		
Pergunta	Entrada 1 não controla a saída 3	
Entrada 1 controla a saída 3		
Entrada 1 não controla a saída 3		
Pergunta	Saída 3 responde à entrada 2	
Saída 3 responde apenas ao horário		
Saída 3 responde à entrada 2		
Pergunta	Saída 3 responde apenas à entrada 2	
Saída 3 responde apenas à entrada 2		
Saída 3 responde à entrada 2 e horário		

Fig. 6. Excel sheet for the ERSE schedules.

On alarm mode, when the input is not active, the output is functioning normally, when the input is active, all the outputs are turned off and a beep is trigged. On control mode the user can chose how the output works, only controlled by the input, by the input and the schedules or only by the schedules. The user makes decisions by clicking in an Excel sheet, depending on the previous decisions the options of the sheet fields change. On the sheet for the holidays the user types the necessary information about the holidays, setting the date and time. Under programming specifications, time is set always to zero hours, zero minutes and zero seconds. The final sheet is divided by legal winter/summer time and further subdivided into regions and tariffs. A change of information due to the ERSE entity is easily to be introduced in the Excel sheet and parsed by the system.

When the device is turned on it starts to verify the existence or the absence of the configuration document. If it does not exist, questions are asked through an LCD, to which the user responds by the interface manipulated by buttons. Eventually, new questions are arise depending on the previously given answers. The user makes the decision, selecting the button corresponding to option one, two or three. The buttons are arranged in order from top to bottom. So, the first button counting from top, selects the option 1 and so on.

5 Results

The device is placed on the outside of the capacitor bank cabinet as seen in Fig. 1. The device is not covered by any shield but is enough secured in what respects to possible electromagnetic interferences. Several tests on the prototype to verify the correct operation of the device are in accordance with the expected performance for industrial applications. The outputs such as the switching on/off the CB from the power supply of the varmeter controller and/or putting a step capacitance are rehearsed according to two types of inputs. The outputs could be set from a predetermined schedule or by buttons or even by both. Also, the check for correct operation of the device using the LCD, the visual feeding of the equipment (LEDs) and the sound signaling (buzzer) is in accordance with the practical application.

6 Conclusions

This paper presents the conception of a prototype implementation of an original study as an alternative solution to control a capacitor bank using a varmeter controller, according to the schedules provide by the Regulator of Energy Services in Portugal. The prototype uses a microcontroller on a developed circuit board, thus replacing the conventional use of a PLC. The prototype offers two different interface solutions between for the user. One of the solutions uses an Excel document to obtain information about the decisions to be carried out by the system. The other, uses an LCD and the buttons available, to obtain the user information. The information coming from both solutions are stored to be read and write through an SD card. The establish architecture for the device hardware considered all the procedures to build the PCB and protocols needed and the design the PCB is made through the use of the Eagle software. The experimental results shows that the hardware and software implemented is suitable for industrial application, assuring the performance characteristics and conveniently mitigating the use of reactive power. Also, this conception brings to the manufacturer of the power factor compensation system a full proprietary of the final solution, giving more autonomous control over equipment final solution. Future work is further concerned with innovation possibility in the software and a convenient adaptation for power system on conventional aircrafts.

Acknowledgement. This work is funded by: European Union through the European Regional Development Fund, included in the COMPETE 2020 (Operational Program Competitiveness and Internationalization) through the ICT project (UID/GEO/04683/2013) with the reference POCI010145FEDER007690; Portuguese Funds through the Foundation for Science and Technology-FCT under the project LAETA 2015–2020, reference UID/EMS/50022/2019; Portuguese Foundation for Science and Technology (FCT) under Project UID/EEA/04131/2013.

References

1. Nunes, P.: Desenvolvimento de unidade digital de controlo auxiliar para baterias de condensadores. MsC thesis, Instituto Superior de Engenharia de Lisboa (2017)
2. Brando, G., Dannier, A., Del Pizzo, A., Coppola, M.: An all-electric-aircraft tailored SiC-based power factor correction converter with adaptive DC-link regulator. Energies **10**, 1–14 (2017)
3. Ware, J.: Power factor correction. IET Wiring Matters **18**, 22–24 (2006)
4. Vaidya, S., Ambad, P., Bholse, S.: Industry 4.0- a glimpse. In: 2nd International Conference on Materials Manufacturing and Design Engineering, pp. 233–238. Procedia Manufacturing (2018)
5. Batista, N.C., Melicio, R., Mendes, V.M.F.: Services enabler architecture for smart grid and smart living services providers under Industry 4.0. Energy Build. **141**, 16–27 (2017)
6. Stock, T., Seliger, G.: Opportunities of sustainable manufacturing in Industry 4.0. Procedia CIRP **40**, 536–541 (2016). 13th Global Conference on Sustainable Manufacturing - Decoupling Growth from Resource Use
7. Batista, N.C., Melicio, R., Matias, J.C.O., Catalão, J.P.S.: Zigbee standard in the creation of wireless networks for advanced metering infrastructures. In: Proceedings of the 16th IEEE MELECON (2012)
8. Batista, N.C., Melicio, R., Mendes, V.M.F.: Layered smart grid architecture approach and field tests by ZigBee technology. Energy Convers. Manag. **88**, 49–59 (2014)
9. Batista, N.C., Melicio, R., Matias, J.C.O.: Photovoltaic and wind energy systems monitoring and building/home energy management using ZigBee devices within a smart grid. Energy **49**, 306–315 (2013)
10. Batista, N.C., Melicio, R., Matias, J.C.O., Catalão, J.P.S.: ZigBee wireless area network for home automation and energy management: field trials and installation approaches. In: Proceedings of the 3rd IEEE ISGT Europe, pp. 1–5 (2012)
11. RTCA: DO-160, environmental conditions and test procedures for airborne equipment. European Organisation for Civil Aviation Equipment (1975)
12. National Instruments: Data acquisition (DAQ) fundamentals. Application Note 007 (1999)
13. Tutoriais, Comparação Entre Protocolos de Comunicação Serial (2018). https://www.robocore.net/tutoriais/comparacao-entre-protocolos-de-comunicacao-serial.html
14. Computer Solutions Ltd - USB protocol (2018). http://www.computer-solutions.co.uk/info/Embedded_tutorials/usb_tutorial.htm
15. Data Acquisition Handbook: Measurement Computing (1997)
16. Kundert, K.: Power supply noise reduction. In: The Designer's Guide Community (2004)

Modeling and Simulation of PV Panel Under Different Internal and Environmental Conditions with Non-constant Load

Masud R. Rashel[1,2,3], Rui Melicio[1,2(✉)], Mouhaydine Tlemcani[1,2], and Teresa Goncalves[1]

[1] Departamento de Física, Universidade de Évora, Escola de Ciências E Tecnologia, Évora, Portugal
ruimelicio@gmail.pt
[2] Instituto de Ciências da Terra (ICT), Universidade de Évora, Évora, Portugal
[3] Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh

Abstract. This paper focuses on PV power conversion under different internal and environmental conditions with non-constant load, connected to a smart grid system. Due to environmental conditions, the PV system is a non-linear system and difficult to predict the power conversion. In the aspect of internal variables, it includes the five parameters of the single diode solar cell model identify their sensitivity through error function. It also identifies the relation between environmental conditions, mainly: irradiance, temperature and wind speed. The modeling and computational simulation with laboratory work identify the effects of internal and environmental effect on the system. The model gives details about the sensitivity of each environmental condition using error function. The work includes the decrease of energy conversion by the solar panel as a function of time due to the shadow effect that affects its performance. Besides these, a smart system is introduced as a DAQ system in laboratory environment to get in real time the power conversion value with the P-V and I-V characteristics of the PV panel.

Keywords: Internal and environmental variables · Solar panels · Error function · Shadow effect · DAQ system

1 Introduction

Due to increasing dependency on electric equipment requirement of electric energy also increases. For fulfilling the requirement, the energy supply sector is searching for source of energy. Renewable energy is a solution which ensures clean environment. Photovoltaics is one of the renewable energy source that get attention of the researchers, business men and policy makers. PV got the attention because of its simple mechanism and low cost production in last 10 years. Its behavior is random, depends on environment parameters [1–3]. It has great potential to supply electricity without producing CO_2 emissions in its working life cycle. PV reached up to 22.2% efficiency [4] and it has to go a long way to improve its efficiency and performance. Due to

Published by Springer Nature Switzerland AG 2019
L. M. Camarinha-Matos et al. (Eds.): DoCEIS 2019, IFIP AICT 553, pp. 376–392, 2019.
https://doi.org/10.1007/978-3-030-17771-3_33

different internal and external facts, it could not give the best performance when working in the active ground. There are different ways to improve the power generation rate of it and among them: (1) improvement of surrounding condition (remove dust, remove obstacles that decrease the performance); (2) optimum parameter value identification; (3) include the ambient temperature with the value of wind speed. These important facts have great impact on it. These are most important to get the best out of a solar PV plant. For estimating the plant value need these type of sophisticated calculation which is shown in this work. This work calculated the PV power generation more approximately than before and helped the smart grid (SG) system to estimate the power generation more accurately [1].

The solar irradiation is differed from place to place on the earth surface. But the source of solar energy is continuous and abundant in nature and in no need of refueling. But the intensity of the solar irradiation is varying place to place. The average distribution of solar radiation through the earth surface [5] is shown in Fig. 1.

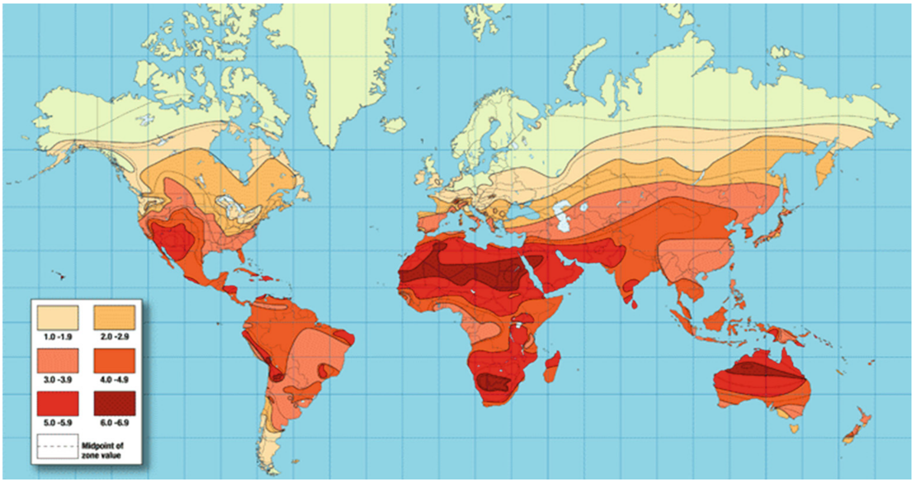


Fig. 1. Average distribution of solar radiation through the Earth surface [5].

In Fig. 1 on the earth surface, there are six zone, which is based on the average amount of solar irradiation in a year on an area of 1 m^2 (kJ) [6]. There are some places that have huge amount of solar irradiation all over the year and could generate huge amount of energy from solar power; Portugal is one of the places in Europe which gets one of the maximum amount of solar radiation. This place is impeccable to produce vast amount of electricity and heat from solar radiation; Portugal's industries are giving focus on this source of energy and huge investment is arranged to get more out of this. For making a smart grid system including the nano and macro grid solar plant with central grid, needs improvement to increase the efficiency of PV system.

2 Relationship to Industrial and Service Systems

The 18th century industrial revolution, said to be the First Industrial Revolution, brought into force the historically never seen before paradigm of industrialization and urban society. Subsequent research and development have delivered scientific evolution triggering technological innovation for industrial and service systems with significant impact into the society. After an industrial revolution, a new set of technological innovation with significant impact into the society, i.e., expressively changing the society, is categorized as a consequent born age said to be the following industrial revolution. In now-a-days a view of a new set of technological innovation for industrial and service systems points into what is said to be the paradigm of a fourth industrial revolution, i.e., Industry 4.0 [7]. Industry 4.0 indorses the interconnection of physical items such as sensors, devices and enterprise assets, interfacing with one each other directly or by the Internet [7, 8] allowing Cyber-Physical Systems (CPS) processes. The main characteristics of these processes are decentralization and autonomous behavior [7]. A PV power system is in the new set of innovation in the scope of a CPS interfaced with information systems with real-time digital platform services. This PV system must be interfaced with urban infrastructures, allowing capabilities of smart grids to schedule the batteries in an environment of smart cities [2, 9, 10] shown in Fig. 2

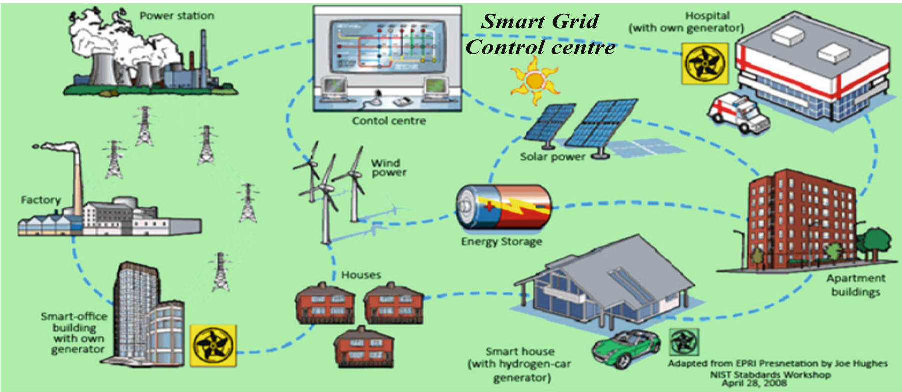


Fig. 2. Smart grid with renewable energy system [2, 10].

Demand side involvement in a power system is a functionality link with smart grids that is connected with PV power conversion system. The communication and computation requirements adopted by PV system user aggregator must include advanced metering infrastructure (AMI), bi-directional data communication using cellular communications and other wireless communication technology including internet of things (IoT), supervisory control and data acquisition (SCADA) [11]. Technological innovation software for the PV aggregator is needed for taking full advantage of it, persuading customer for benefits of being with an PV aggregator and consequently favoring the desired rise on the PV power generation market penetration [9]. This work is for increasing the efficiency of the PV system that is integrated with central smart grid system. Due to lack of maintenance and lack of knowledge, systems do not get the best performance. It needs deep analysis, both internal and external parameter that is accomplished through this research work. The work is consisting of four different parts. First part works with internal and external parameters sensitivity identification through error function to identify the optimal value identification point to get optimal value for the system to help PV panel to get optimal solution. For external variables like irradiance, ambient temperature and wind speed's get their sensitivity to the PV panel and how the effect the output of PV is identified. The effect of environmental parameters is shown graphically that introduce new equation which includes wind speed importance for getting more accurate approximate value of power generation, what makes the central grid smarter. In author part DAQ system is used to evaluate a PV panels output with different resistance to get the effect of non-constant load to the system. This work show different behavior that help to get idea of non-constant load over PV, also important for the energy service system. At the last part, the obstacles effect on PV panel is identified. Obstacles like shadow over panel's surface make PV more unstable in the sense of power production. It is important to identify this types of effect causes to get the idea of fault in the system it could be permanent or temporary. Identification of faults type is important to give the solution for the problem. One of the important identification is the characteristics of obstacles are categorized in two parts. This identification, could integrated with service system of PV [1].

This work knowledge is important to connect PV nano, micro grid system with center grid system; it has significant value in the sector of power industry. Accurate estimation of power from the small systems help the central grid to estimate the other energy source for generated desirable production.

3 State of the Art

In the year 2015 a historical decision had been taken in the Paris Climate Conference [12] to limit the global warming and help to protect the environment. In that conference 195 countries from all over the world agree in a decision to limit the carbon emission using available technology and science. Renewable energy sources and technologies have a great role in this decision and world is looking for developing a system that will help to build sustainable environment.

In [13] describes that it is not possible to make a risk free energy resource. In this work it suggests that to choose a source of energy, the priority should give to environmental effects and the cost issues. In this both section, renewable energy is the best solution. In [14, 15] state about renewable energy for sustainable development. Renewable energy sources named solar, biomass, geothermal or wind are analyzed from both environmental and economical view point. In [16] describes a detail overview about integrating renewable energy with central smart grid system. PV power conversion is not linear in nature and for connecting, them with SG need more improvement computational tools, smart devices.

In [17, 18] give details about the modeling and simulation for integrating wind farm and solar plant with the smart grid. The simulation model was developed in the Matlab/Simulink. It also gives details about nano grid with a smart house system with solar panels. It states that good prediction model is important for SG. In [19–24] states about different approaches to integrate PV system with SG. These works discuss about control methods and about DC-DC boost converter and two-level converter. And also define three level inverter. In [23] states the way how to integrate the poly-Si PV system with the central grid. This work introduces a fuzzy controller to control the converter and show the efficient way to connect with SG. In [25, 26] states different methods to get I-V and P-V characteristics from five parameters single diode PV model those are based on the datasheet values. These works include different test conditions under changing irradiation and temperature with those conditions PV power production. In [27] describes a five parameters model that included photocurrent, dark saturation current, series resistance, shunt resistance and diode ideality factor. Using datasheet values this model is created as like [26]. Under real environmental conditions, the model is tested and verified.

In [28] states the effect of wind speed on PV panels' power generation. It analysis the correlation between tilt angles and wind speed to get the maximum power generation angle in different season.

In [29] states and shows that the operating temperature of PV is an important variable that affect the maximum power generation from the PV. In this work it is stated that temperature has linear relation with PV power generation rate. In [30] states the effect of ambient temperature and wind speed about the performance of monocrystalline PV power conversion. In [31] states the enhanced condition for PV under varying environmental condition. Under varying environmental condition, load is also changing depending on it. In warm or cold condition, the use of different electronic devices to varying the temperature and the load also increase. This work show that load analysis is an important fact to make an optimum PV power conversion system.

In [32] describes and gives details about different dust condition in Libya on PV power conversion performance. It gives the comparison between PV power conversion performance under dust and clean environment. It shows that the dust on PV surface losses the power conversion performance. In [33] describe a simulator that can simulate the PV power conversion system under partial and dynamic shadow condition. In [34] states the dust effect on PV modules' surface. It shows the relation between dust deposition on PV surface and in that condition the sunlight way out to its surface.

After going through all these state of the art, it is important to get more accurate PV five parameter model that also should include wind speed effect in the equation. Load analysis is important and need more analysis with real time scenario. Shadow effect is important to take account to identify different loss from PV power conversion system. All these problems is analyzed and identified valuable result that help to make better approximate model that give better prediction result for PV power conversion system.

4 Research Contribution and Innovation

Four different analysis had done and generate knowledge for the PV power generation system to improve its performance in real working field. This information is important to integrate it with smart grid and identify fault in PV plant.

Single diode five parameters PV model is used to analysis different internal and external parameters with error function. Figure 2 is the equivalent circuit diagram for the PV cell model.

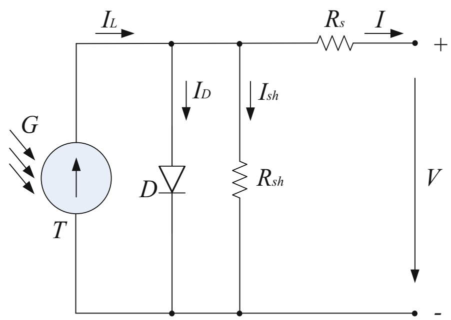


Fig. 3. Single diode five parameters PV cell model.

In Fig. 3 I_L is the photocurrent, which is the current source for the circuit, single diode D which have I_D current across it, a series resistance R_s which represents the resistance in the cell, shunt resistance R_{sh} is in parallel [1]. The I-V characteristic is formulated by an implicitly function is given by:

$$I = I_L - I_D - I_{sh} = I_L - I_0\left\{\exp\left[\left(\frac{q(V + IR_S)}{nkT}\right)\right] - 1\right\} - \left(\frac{V + IR_s}{R_{sh}}\right) \quad (1)$$

In (1) I_L is the photocurrent, I_D is diode current, and I_0 is the dark saturation current, R_s is the internal series resistance, R_{sh} is the shunt resistance, n is the diode ideality factor, q is the electron charge, k is the Boltzmann's constant, T is the cell temperature of the junction [1]. The data for the c-Si PV cell at STC is shown in Table 1 [1].

Table 1. Data for the c-Si PV cell at STC [1]

Technology	V_m	I_m	V_{oc}	I_{sc}	∞_{sc}
c-Si	0.55 V	1.98 A	0.64 V	2.1 A	1.7 mA/°C

In Table 1 the values are used to do the simulation under different internal and environmental conditions. I_{sc} is the value for short circuit current, V_{oc} is the open circuit voltage, and V_{mp} is voltage at maximum power point, I_{mp} is the current at maximum power [1].

For understanding, different internal and external parameters' effect to the PV power conversion is identified using the sensitivity test and that is using error function. For each parameter error function gives detailed output, how it is sensitive to total system. It also assists to get approximate value of the starting point to find an optimum value for that parameter. The error function is given by:

$$E(P)_{1.....n} = \frac{1}{n}\sqrt{\sum\nolimits_{i=1}^{n}\left(I_{iP} - I_{iM}\right)^2} \qquad (2)$$

Single parameter with its error function [2, 35, 36] is shown in Fig. 4.

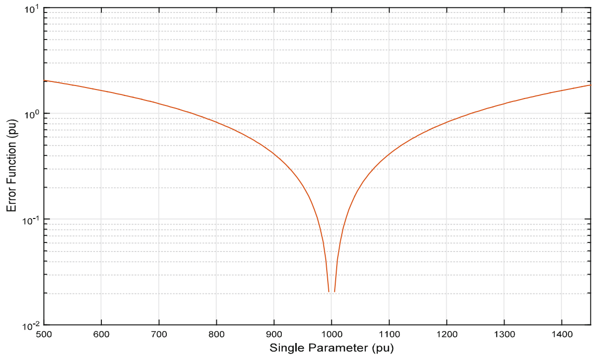


Fig. 4. Single parameter with error function.

Considering (2) the error function calculated for each internal and external environmental parameter. After getting the values, it builds a parameter vs error function graph like Fig. 3 and it gives the idea about the sensitivity [1]. The circuit diagram for the measurement system is shown in Fig. 5.

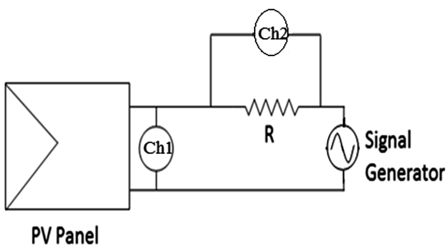


Fig. 5. Circuit diagram for the measurement system.

In Fig. 5 the DAQ system, channel 1 (Ch1) and channel 2 (Ch2) is connected with work station through USB cable and also with measurement system. GPIB cable is used to connect the work station with the signal generator. The measurement process is as, at the starting, measurement is done to get the open circuit voltage using (Ch1) and in that time the signal generator is disabled. After the measured value is passed to the workstation. This value is used to calculate the offset value and amplitude value for generating the RAMP signal. These values are send to the signal generator to generate accepted signals to initialize the measurement process, after it generates RAMP signal as commanded one. After this, the voltage is measured by (Ch2). The current is calculated by the value of additive resistance *R* and this measured voltages in different points. In this measurement process 10 Hz frequency is used for the signal. All these values are used to generate the I-V and P-V curves for the specific PV panel [1]. The Signal generator and DAQ connects with workstation is shown in Fig. 6.

Fig. 6. Signal generator and DAQ connects with workstation.

For analyzing the obstacles over the PV panel or fault in PV panel the simulation is done in this work and Simulink model is used to build the model. The model gives the value like real time scenario [1]. The Matlab/Simulink model for understanding different obstacles over PV panel is shown in Fig. 7.

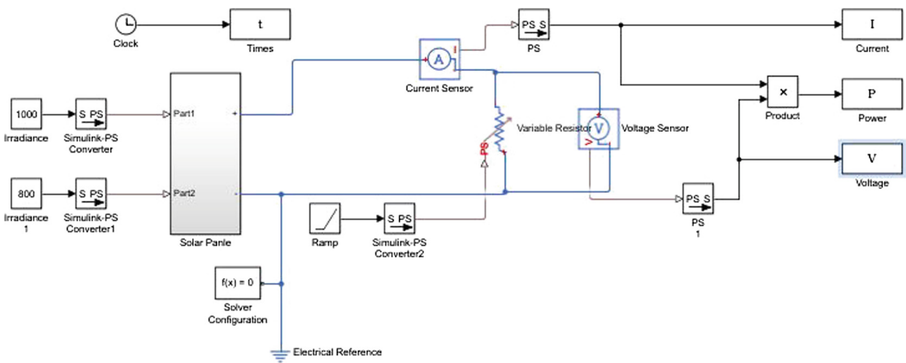


Fig. 7. Simulink model for understanding different obstacles over PV panel.

5 Result and Discussion

In this section, the result from four case studies are discussed. All these case studies are related to advance the quality of the PV power conversion that is connected with smart grid for better distribution.

5.1 Case Study 1

In this case study the sensitivity analysis for the internal and external parameters are analyzed through error function. Internal six parameters are analyzed through error function and those parameters are: photocurrent, cell temperature, diode ideality factor, series resistance, shunt resistance and diode saturation current.

Figure 8 shows the six internal parameters named as (a); the photocurrent as (b); the cell temperature as (c); the diode ideality factor as (d); the series resistance as (e); the shunt resistance as (f); diode saturation current.

All of the above curves show minimum points where the error becomes zero, there the best values are achieved by minimizing the error among measured values and optimal value. Photocurrent, cell temperature, saturation current and diode ideality

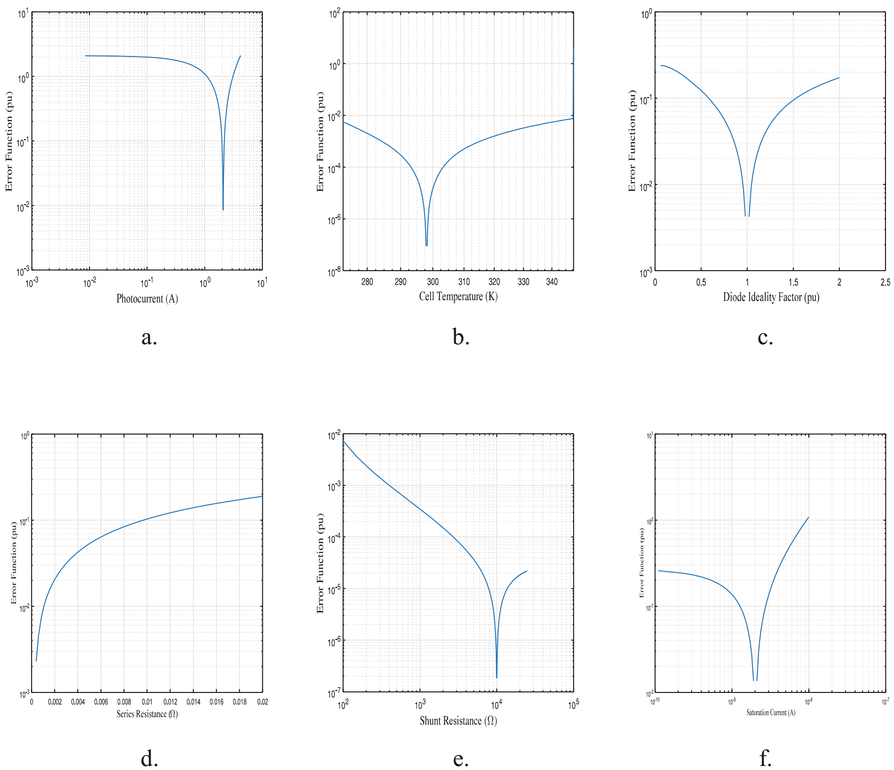
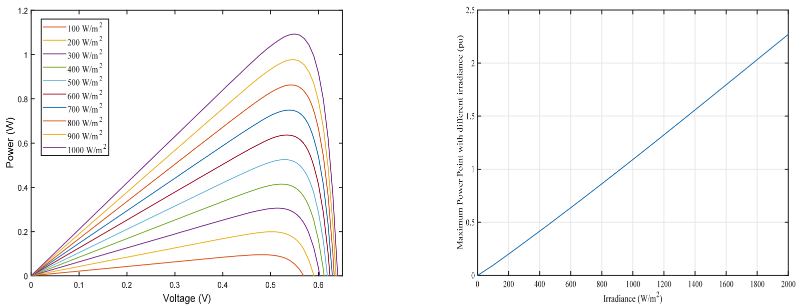


Fig. 8. Six internal parameters named as a. photocurrent; b. Cell temperature; c. Diode ideality factor; d. Series resistance; e. Shunt resistance; f. Diode saturation current, all of them with error function.

factor and series resistance are important parameters that show significant sensitiveness and on the other hand shunt resistance shows less sensitivity.

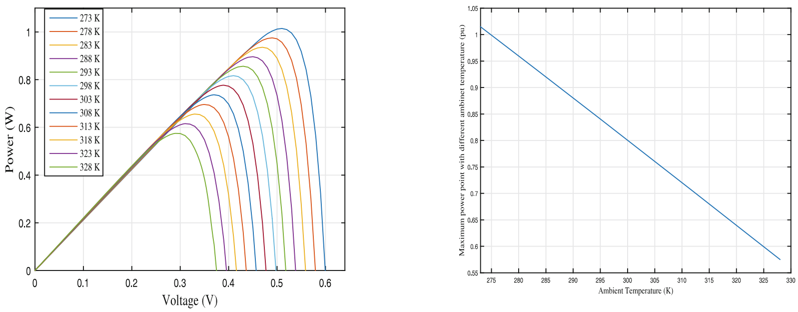
5.2 Case Study 2

In this case study environmental variables are important and they have direct and indirect relationship with PV power conversion rate. Figure 9 shows the changing value of irradiance as (a); the change of PV power, (b); increasing Irradiance increase the maximum power generation. This changing is happened nearly linear way. Figure 10 shows the changing ambient temperature as (a); the change of PV power conversion, (b); increasing ambient temperature decrease maximum power generation and it's also decreasing in linear fashion. For wind speed, the changing value of wind speed make change the value of PV power conversion shown in Fig. 11. Figure 11 shows the changing wind speed as (a); the change of PV power conversion, (b); increasing wind speed increase the power generation.



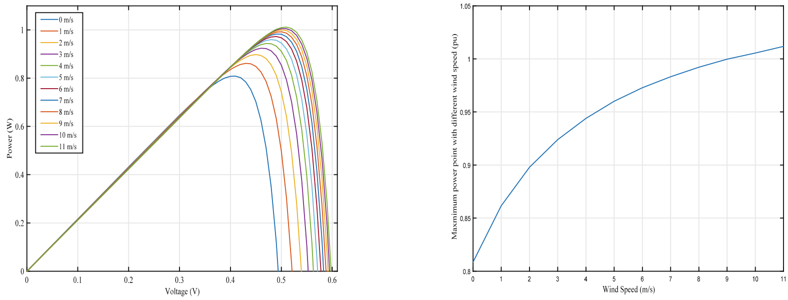
a. b.

Fig. 9. a. Changing value of Irradiance changes the value of PV power conversion; b. increasing Irradiance increase the maximum power point almost linearly.



a. b.

Fig. 10. a. Changing value of ambient temperature changes the value of PV power conversion rate; b. increasing ambient temperature decrease the maximum power point linearly.



a. b.

Fig. 11. a. Changing value of wind speed changes the value of PV power conversion; b. increasing wind speed increase the maximum power point as *ln x*.

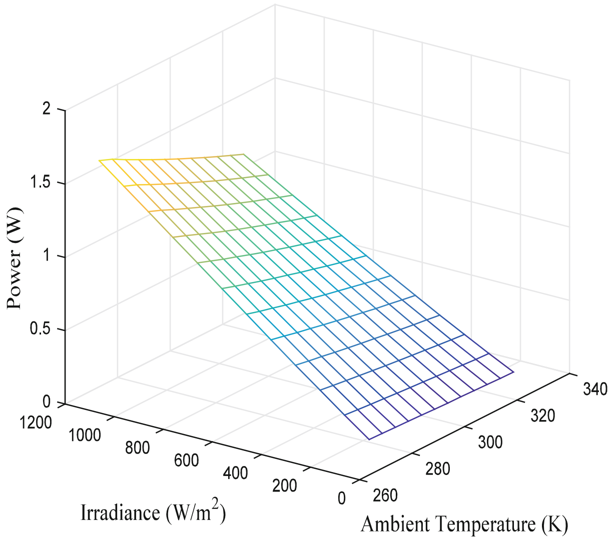


Fig. 12. PV power with changing irradiance and ambient temperature.

Two environmental, ambient temperature and irradiance are changing same time and their effect on PV power conversions is shown in the Fig. 12.

Two environmental, wind speed and ambient temperature are changing at the same time and their effect on PV power conversions is shown in Fig. 13.

Two environmental, wind speed and irradiance are changing at the same time and their effect on PV power conversions is shown in Fig. 14.

The results support of this section are in [29–31].

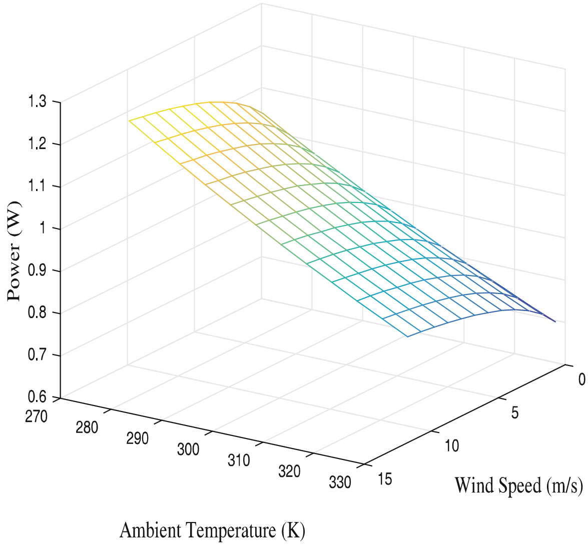


Fig. 13. PV power with changing ambient temperature and wind speed.

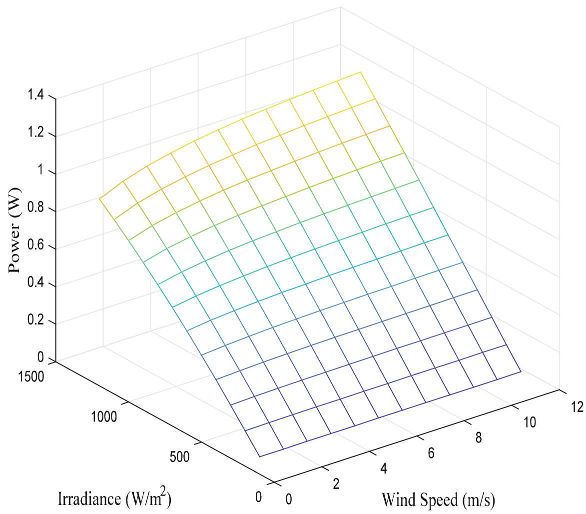


Fig. 14. PV power with changing wind speed and irradiance.

5.3 Case Study 3

Table 2 shows that the increasing load values are decreasing the maximum power point of specific PV panel. It is important to get an optimal load for a specific PV panel. It shows that the experiment that the current at maximum power (I_{mp}) and voltage at maximum power (V_{mp}) is changing. The value of load changing also change the

Table 2. Different additive load resistance with respective maximum power

Resistance	V_{oc}	I_{sc}	P_{max}
100 Ω	5.0909 V	0.0055 A	0.0281 W
220 Ω	5.0959 V	0.0057 A	0.0275 W
560 Ω	5.1214 V	0.0046 A	0.0204 W
1.5 kΩ	5.1048 V	0.0029 A	0.0149 W
4.7 kΩ	5.1188 V	0.0010 A	0.0053 W

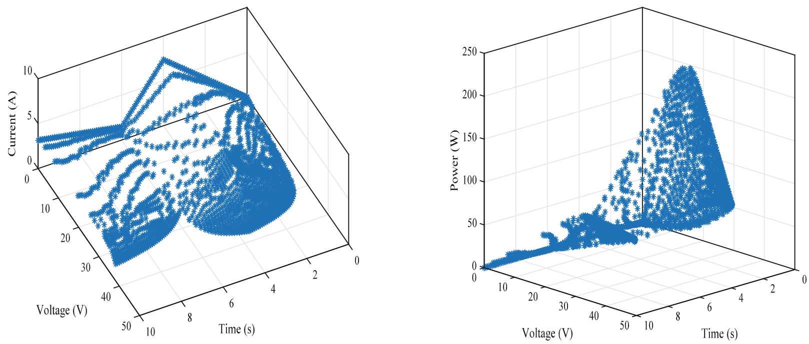
MPP. In the experiment, five different values of resistances are used to get the output from a specific PV panel in actual environmental state. Due to variation of load the PV power conversion rate become oscillatory.

5.4 Case Study 4

Shadow of different obstacles over a PV panel is simulated through Simulink model. In the real life scenario, different things create impediment for sunlight to reach the surface of PV panel. There are different types of obstacles named as shadow, dust, physical damage on PV panel. Due to obstacles over PV surface, the P-V curve gets multiple picks that introduce local maxima and it becomes difficult to get the exact maximum power point (MPP).

5.4.1 Time-Dependent Obstacle

Time-dependent obstacles are changing their shape with time over the PV surface. It is not permanent type but introduce local maxima due to shadow. They change the value of irradiance and due to this variation the PV power conversion rate also change. These types of obstacles are as; a sudden cloud, a flying bird over the PV etc. Time-dependent obstacles go over the panel surface and don't stay for long time over it [37]. Figure 15



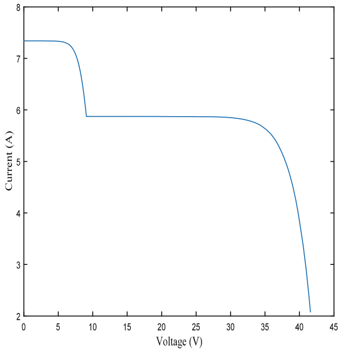
a. b.

Fig. 15. a. I-V curve with time under rapid changing obstacles; b. P-V curve with time under rapid changing obstacles.

shows (a); the I-V curve against time, under rapid changing obstacles, (b); P-V curve against time under rapid changing obstacles.

5.4.2 Time-Independent Obstacle

Time-independent obstacle is a type of permanent damage of PV panel or it may have to solve the problem manually. These types are as; few cells of PV panel are damaged, thick layer of dust, bird droppings etc. These types of problems have to solve manually. Figure 16 shows (a); the I-V curve which have 54 cells under time-independent obstacle and (b); P-V curve which have 54 cells under time-independent obstacle condition.



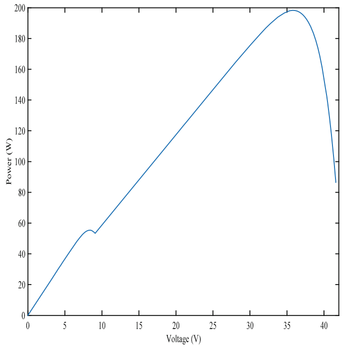


Fig. 16. a. I-V curve having 54 cells under time-independent obstacle; b. P-V curve with 54 cells under time-independent obstacle condition.

6 Conclusion and Future Work

For developing sustainable environment and to help the world for reducing the greenhouse gases (GHG) the Paris Agreement took important decision to save the world. To implement and fulfill the vision, making energy sources clean is needed. People around the world is giving emphasis on the renewable energy as it is a source of clean energy that can help to improve quality of the Ozone layer [38].

The simulation is done to identify the sensitivity of internal and environmental parameters through error function [1, 2, 35, 36]; Environmental variables, named irradiance, ambient temperature and wind speed variation influence on PV power conversion rate is clearly identified that is possible to include in prediction model [39]; Non-constant load influence on the PV panel is identified through the experiment is done [1]; Time-dependent and time-independent obstacles are categorized to identify different types of fault in PV panel, these are important for fault identification in real working ground of PV plant [3, 37].

For future direction of this work is followed as, making experimental workout in laboratory and real environmental condition using different types of PV panels for getting their specific information to model more precise PV system that permitting to integrate with smart grid. Also developing database based on different types of obstacles of PV targeting to classify faults. It improves the system's performance.

Acknowledgement. This work is funded by: European Union through the European Regional Development Fund, included in the COMPETE 2020 (Operational Program Competitiveness and Internationalization) through the ICT project (UID/GEO/04683/2013) with the reference POCI010145FEDER007690.

References

1. Rashel, M.R.: Modeling photovoltaic panels under variable internal and environmental conditions with non-constant load. PhD Thesis, Universidade de Évora (2018)
2. Rashel, M.R., Rifath, J., Gonçalves, T., Tlemçani, M., Melicio, R.: Sensitivity analysis through error function of Crystalline-Si photovoltaic cell model integrated in a smart grid. Int. J. Renew. Energy Res. **7**(4), 1926–1933 (2017)
3. Rashel, M.R., Albino, A., Rifath, J., Gonçalves, T., Tlemcani, M.: Matlab simulink modeling of photovoltaic cells for understanding shadow effect. In: Proceedings of IEEE International Conference on Renewable Energy Research and Applications, pp. 747–750 (2016)
4. Clean Energy Reviews. https://www.cleanenergyreviews.info/blog/best-quality-solar-panels-manufacturers. Accessed 12 Oct 2018
5. Alte Store. https://www.altestore.com/howto/solar-insolation-map-world-a43/. Accessed 07 Sept 2017
6. Chen, C.J.: Physics of Solar Energy. Wiley, Hoboken (2011)
7. Batista, N.C., Melicio, R., Mendes, V.M.F.: Services enabler architecture for smart grid and smart living services providers under industry 4.0. Energy Build. **141**, 16–27 (2017)
8. Batista, N.C., Melicio, R., Matias, J.C.O., Catalão, J.P.S.: Zigbee standard in the creation of wireless networks for advanced metering infrastructures. In: Proceedings 16th IEEE MELECON, pp. 220–223 (2012)
9. Gomes, I.L.R., Melicio, R., Mendes, V.M.F., Pousinho, H.M.I.: Decision making for sustainable aggregation of clean energy in day-ahead market: uncertainty and risk. Renew. Energy **133**, 602–702 (2019)
10. Transverter. http://www.transverter.com/smart.html. Accessed 13 Mar 2017
11. Shaukat, N., et al.: A survey on electric vehicle transportation within smart grid system. Renew. Sustain. Energy Rev. **81**, 1329–1349 (2018)
12. Paris Climate Change Conference: The global standard of globalization is not limited to the limitations of the system, XXI 2°C (2015)
13. Bozkurt, I.: Energy resources and their effects on environment. WSEAS Trans. Environ. Dev. **6**(5), 327–334 (2010)
14. Omer, A.M.: Energy and environment: applications and sustainable development. Br. J. Environ. Clim. Change **1**(4), 118–158 (2011)
15. Mitoula, R., Abeliotis, K., Vamvakari, M., Gratsani, A.: Sustainable regional development through the use of photovoltaic (PV) systems: the case of the Thessaly region. In: Proceedings of World Renewable Energy Congress, pp. 8–13 (2011)
16. Phuangpornpitaka, N., Tia, S.: Opportunities and challenges of integrating renewable energy in smart grid system. Energy Procedia **34**, 282–290 (2013)

17. Mekkaoui, A., Laouer, M., Mimoun, Y.: Modeling and simulation for smart grid integration of solar/wind energy. Leonardo J. Sci. **30**, 31–46 (2017)
18. Shafiullah, G.M., Amanullah, M.T.O., Ali, A.B.M.S., Wolfs, P.: Smart grid for a sustainable future. Smart Grid Renew. Energy **4**, 23–34 (2013)
19. Fialho, L.A.P., Melicio, R., Mendes, V.M.F., Viana, S., Rodrigues, C., Estanqueiro, A.: A simulation of integrated photovoltaic conversion into electric grid. Sol. Energy **110**, 578–594 (2014)
20. Fialho, L.A.P., Melicio, R., Mendes, V.M.F., Rodrigues, L., Viana, S., Estanqueiro, A.: Simulation of a-Si PV system linked to the grid by DC-DC boost and two-level converter. In: Proceedings of 16th International Power Electronics and Motion Control Conference and Exposition, pp. 934–938 (2014)
21. Fialho, L., Melício, R., Mendes, V.M.F., Collares-Pereira, M.: Simulation of a-Si PV system linked to the grid by DC boost and three-level inverter under cloud scope. In: Camarinha-Matos, Luis M., Baldissera, Thais A., Di Orio, G., Marques, F. (eds.) DoCEIS 2015. IAICT, vol. 450, pp. 423–430. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-16766-4_45
22. Fialho, L.A.P., Melicio, R., Mendes, V.M.F., Estanqueiro, A.: Simulation of a-Si PV system grid connected by boost and inverter. Int. J. Renew. Energy Res. **5**(2), 443–451 (2015)
23. Fialho, L.A.P., Melicio, R., Mendes, V.M.F.: Poly-Si PV system grid connected and fuzzy controlled. In: Proceedings 16th IEEE International Conference on Computer as a Tool, pp. 1–6 (2015)
24. Fialho, L.A.P., Melicio, R., Mendes, V.M.F., Estanqueiro, A., Pereira, M.I.C.: PV systems linked to the grid: parameter identification with a heuristic procedure. Sustain. Energy Technol. Assess. **10**, 29–39 (2015)
25. Tayyan, A.A.E.: PV system behavior based on datasheet. J. Electron Devices **9**, 335–341 (2011)
26. Tayyan, A.A.E.: A simple method to extract the parameters of the single-diode model of a PV system. Turk. J. Phys. **37**, 121–131 (2013)
27. Aoun, N., Chenni, R., Nahman, B., Bouchouicha, K.: Evaluation and validation of equivalent five-Parameter model performance for photovoltaic panels using only reference data. Energy Power Eng. **6**(9), 235–245 (2014)
28. Gokmen, N., Hu, W., Hou, P., Chen, Z., Sera, D., Spataru, S.: Investigation of wind speed cooling effect on PV panels in windy locations. Renew. Energy **90**, 283–290 (2016)
29. Dubey, S., Sarvaiya, J.N., Seshadri, B.: Temperature dependent photovoltaic (PV) efficiency and its effect on PV production in the world: a review. Energy Procedia **33**, 311–321 (2013)
30. Bhattacharya, T., Chakraborty, A.K., Pal, K.: Effects of ambient temperature and wind speed on performance of monocrystalline solar photovoltaic module in Tripura, India. J. Solar Energy **2014**, 1–5 (2014)
31. Zerhouni, Z.F., Zerhouni, M.H., Zegrar, M., Benmessaoud, M.T., Stambouli, A.B., Midoun, A.: Proposed methods to increase the output efficiency of a photovoltaic (PV) system. Acta Polytech. Hung. **7**(2), 55–70 (2010)
32. Mohamed, A.O., Hasan, A.: Effect of dust accumulation on performance of photovoltaic solar modules in Sahara environment. J. Basic Appl. Sci. Res. **2**(11), 11030–11036 (2012)
33. Piazza, M.C.D., Vitale, G.: Photovoltaic field emulation including dynamic and partial shadow conditions. Appl. Energy **87**, 814–823 (2010)
34. Wang, J., Gong, H., Zou, Z.: Modeling of dust deposition affecting transmittance of PV modules. J. Clean Energy Technol. **5**(3), 217–221 (2017)
35. Rashel, M.R., Gonçalves, T., Tlemçani, M., Melicio, R.: Photovoltaic cell performance analysis under different ambient temperature and wind speed for sustainable energy. Vietnam J. Sci. **3**(1), 1–16 (2017)

36. Rashel, M.R., Albino, A., Gonçalves, T., Tlemçani, M.: Sensitivity analysis of environmental and internal parameters of a photovoltaic cell. In: Proceedings of Energy for Sustainability International Conference - Designing Cities & Communities for the Future, Madeira (2017)
37. Rashel, M.R., Ahmed, T., Gonçalves, T., Tlemcani, M., Melicio, R.: Analysis of different types of obstacles for PV panel. In: Eradication Poverty Through Energy Innovation Workshop, Arizona State University (2018)
38. International Energy Agency: Energy and Climate Change (2015)
39. Rashel, M.R., Ahmed, T., Goncalves, T., Tlemcani, M., Melicio, R.: Analysis of environmental parameters sensitivity to improve modeling of a c-Si panel. Sens. Lett. **16**(3), 176–181 (2018)

Power Systems

Effect of Combined Stresses on the Electrical Properties of Low Voltage Nuclear Power Plant Cables

Ramy S. A. Afia[1,2(✉)], Tamus Zoltán Ádám[2], and Ehtasham Mustafa[2,3]

[1] Department of Electrical Power and Machines Engineering, Helwan University, Cairo, Egypt
ramysaad@h-eng.helwan.edu.eg

[2] Department of Electric Power Engineering, Faculty of Electrical Engineering and Informatics, Budapest University of Technology and Economics, Budapest, Hungary
{tamus.adam,mustafa.ehtasham}@vet.bme.hu

[3] Department of Electrical Engineering, Faculty of Engineering and Technology, Gomal University, Dera Ismail Khan, Pakistan

Abstract. Energy is the heart of any industry since each industry uses energy. One of the most critical sources for energy generation is the nuclear power plants. The plants building needs a lot of investments and a pivotal role from the financial services sector in providing the necessary finance for the capital works. Control, instrumentation and low voltage power cables have a vital role in the operation, reliability and lifetime of these plants. The performance of these cables has an effect on the energy industry and the industries which use this kind of energy source. During operation, these cables are subjected to different stresses such as thermal, environmental, electrical and mechanical stresses so, condition monitoring of these cables is a must. This paper discusses a combined aging mechanisms effect on the electrical parameters of these cables. The assessment was done by monitoring the capacitance and the loss factor of the cables.

Keywords: Nuclear power industry · Low voltage cables · Combined stresses · Condition monitoring · Capacitance · Loss factor

1 Introduction

Nuclear Power Plants (NPPs) are one of the most important energy sources that overcome the dramatic increase in the power demand. Some parts in these plants have a particular concern such as nuclear reactors, heat exchangers, concrete structure in addition to hundreds of kilometers of instrumentation and control (I&C) and low voltage power cables [1–5]. These cables have a vital role in the safety and operation of NPPs since they provide the link between the transducing, instrumentation and control systems that monitor these plants [1]. The functionality of these cables depends on its insulation integrity. Thanks to the harsh environmental conditions in NPPs

Published by Springer Nature Switzerland AG 2019
L. M. Camarinha-Matos et al. (Eds.): DoCEIS 2019, IFIP AICT 553, pp. 395–404, 2019.
https://doi.org/10.1007/978-3-030-17771-3_34

characterized by elevated temperatures and gamma- radiations, the insulating materials of these cables must withstand these stresses and other stresses such as electrical and mechanical stresses to ensure their safety functions over the long-term normal operation and during the design basis events (DBE) such as loss of coolant accident (LOCA). The insulating material of the cables which are particularly related to NPPs is made of polymers such as ethylene propylene rubber (EPR), cross-linked polyethylene (XLPE) and ethylene propylene diene monomer (EPDM) [4, 6, 7]. The effect of the stresses on the polymeric materials related to the nuclear environments affects its physical, chemical and electrical properties causing progressive degradation so, monitoring of these cables is a must. Monitoring of these cables can be done by different testing techniques such as visual, electrical, mechanical and chemical technique. The traditional method used for condition monitoring of these cables is the mechanical one which includes two types of testing, the 50% Elongation at Break (EaB) which has a destructive nature and the Indenter Modulus (IM) [7, 8]. So, many types of research based on non-destructive condition monitoring has been published recently especially there is a desire to extend the lifetime of NPPs however they are 30 to 40 years old. One of the non-destructive methods such as electrical technique to evaluate these cables is measuring the capacitance and the loss factor. In this research work, low voltage NPP power cables have been subjected to accelerated thermal aging followed by mechanical bending stress. The effect of combined stresses on the capacitance and the loss factor as a function of frequency was studied. Further analysis was done by correlating the loss factor and EaB results for the same aging conditions carried out at Laborelec lab, Belgium under the IAEA project [9].

2 Relationship to Industrial and Service Systems

Around the globe, the backbone of the industrial world is the electric power industry since it supplies the essential energy to industrial, manufacturing, commercial and residential customers. The investments in the mature power markets of the developing countries are driven to the transition of fuel and energy sources, increased environmental legislation and an ever-aging generation fleet and transmission-distribution infrastructure. In contrast, the expansion in the power bases in the developing economies is a must to meet the growing demand for electricity, especially in starved regions. For these reasons, the electric power industry continues to have the largest investments. Electric power can be generated by either renewable or nonrenewable resources. The nonrenewable resources are coal, oil, natural gas and nuclear. Nuclear power plants (NPPs) are relatively cheap to run but in contrast, they are very expensive to be built. In many places, in comparison with fossil fuels, nuclear energy is competitive as a means of electricity generation because waste disposal and decommissioning costs are usually fully included in the operating costs. If the social, health and environmental costs of fossil fuels are also taken into account, the competitiveness of nuclear power is improved. The NPPs contain enormous kilometers of I&C as well as low voltage power cables. The harsh environmental conditions in NPPs has a serious effect on these cables because of the high radiation levels and elevated temperature. High radiation and temperature stresses may be accompanied by other stresses such as electrical and mechanical. The operation, reliability and lifetime of NPPs are affected greatly with the performance of I&C in addition to low voltage power cables so these cables should be conditionally monitored to ensure their functionality. To ensure the continuity of energy supplied by NPPs, non-destructive Condition Monitoring (CM) techniques must be applied to these cables which are directly reflected in the industry and service systems.

3 Experimental Work

3.1 Sample Description

The measurements were carried out on samples of low voltage NPP power cable. The technical specifications of this cable are listed in Table 1. The cable consists of three main parts tin-coated copper conductor, XLPE as main insulation and outer jacket made of chlorosulfonated polyethylene (CSPE) as seen in Fig. 1.

Table 1. Cable specifications.

Parameter	Value
Nominal voltage	0.6 kV
Insulation material	XLPE
Insulation thickness	1.143 mm
Jacket material	CSPE
Jacket thickness	0.762 mm
Continuous bending radius	38.1 mm
Overall diameter	8.636 mm
Max. conductor temperature	90 °C

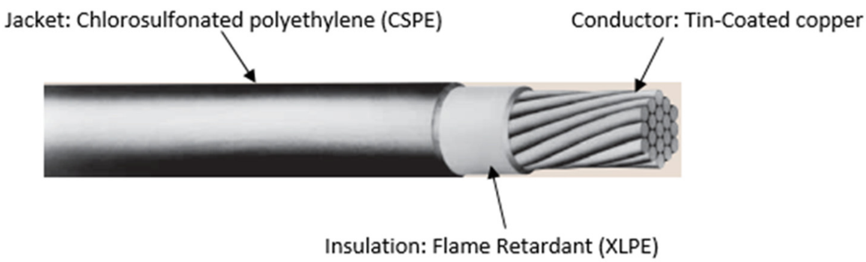


Fig. 1. Cable construction.

3.2 Accelerated Aging Procedure

The samples have been thermally accelerated aged at a temperature of 120 °C for different periods to simulate different service aging times. Based on the Arrhenius equation, Table 2 shows the accelerated aging periods and the equivalent service aging time at the service temperature of 55 °C and activation energy (Ea) of 1.13. The NPP cables should be installed according to the manufacturer's recommendations but in some situations, this does not happen as seen in Fig. 2 since exceeding the running bending radius makes the mechanical aging more significant [4, 10]. So, after the completion of the thermal accelerated aging, the samples have been subjected to mechanical bending stress. The samples were bent for two weeks with a bending radius of 50% (19.05 mm) lower than the nominal bending radius recommended by the cable manufacturer.

Table 2. Accelerated thermal aging and equivalent service aging time.

Cycle	Accelerated aging period (hours)	Equivalent service aging period (years)
1	176	15
2	342	29
3	516	44
4	793	67

Fig. 2. High mechanical bending stress on NPP power cable [4].

4 Results and Discussion

The measurements of the capacitance and tanδ have been carried out by precision component analyzer with a testing voltage of 5 V and a frequency range from 20 Hz to 500 kHz. Figures 3 and 4 shows the baseline measurements for the cable samples without aging. As shown in Fig. 3, the capacitance have a higher values at lower frequencies while it decreases moving to the higher frequencies. The tanδ values of

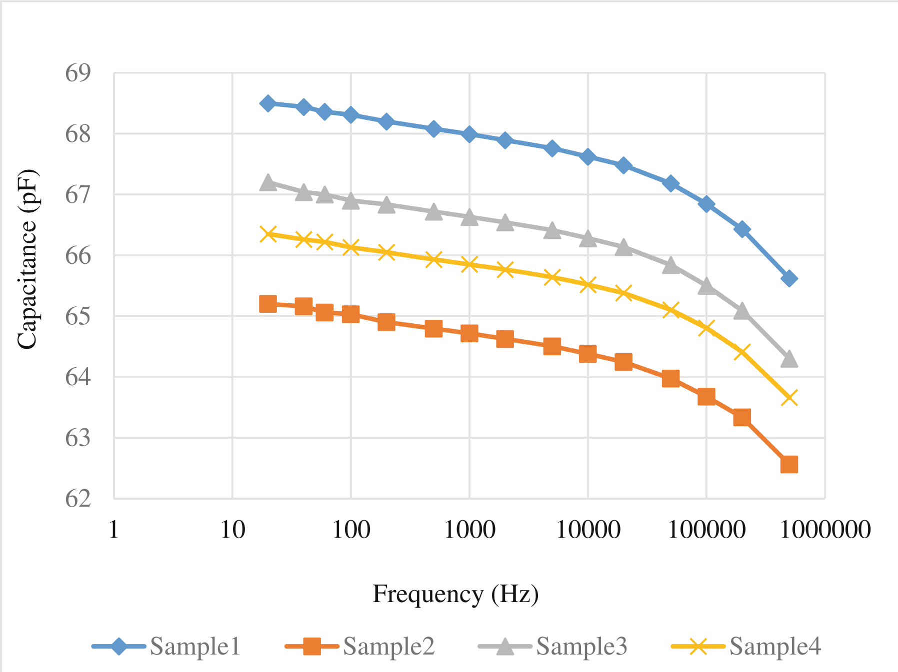


Fig. 3. The capacitance of unaged samples versus frequency.

unaged cable samples are shown in Fig. 4 since the higher values of tanδ shifted to the high-frequency range indicating how the leakage current dominates in the insulation [7].

The effect of each thermal aging cycle was investigated by measuring the capacitance and tanδ after each cycle. At all frequencies, the aged samples have higher capacitance than the unaged samples as shown in Fig. 5. Also increasing the aging time decreases the capacitance since the insulating material loses its capacitive property.

The plot of the tan δ values of the aged and unaged cable samples is shown in Fig. 6. The same result was obtained since the higher values of tanδ shifted to the high-frequency range in addition at each frequency the aged samples have higher values compared to the unaged samples. Also, Fig. 6 shows that the tanδ was increased with increasing the aging period, which means that the insulation leakage current has increased as reported in [7].

After the fourth thermal aging cycle (67-year), the cable samples have been subjected to mechanical bending stress for two weeks. The effect of combined (thermal and mechanical) aging is plotted in Figs. 7 and 8. As it is clearly shown in Fig. 7, the mechanical stress increased the capacitance compared to the thermal stress only and the high difference in the capacitance between the two different aging mechanisms was observed at lower frequencies decreasing when shifting to the high-frequency range. Figure 8 shows the variation in tanδ under different stresses where the combined aging mechanisms increased the tanδ the in comparison with thermal aging only.

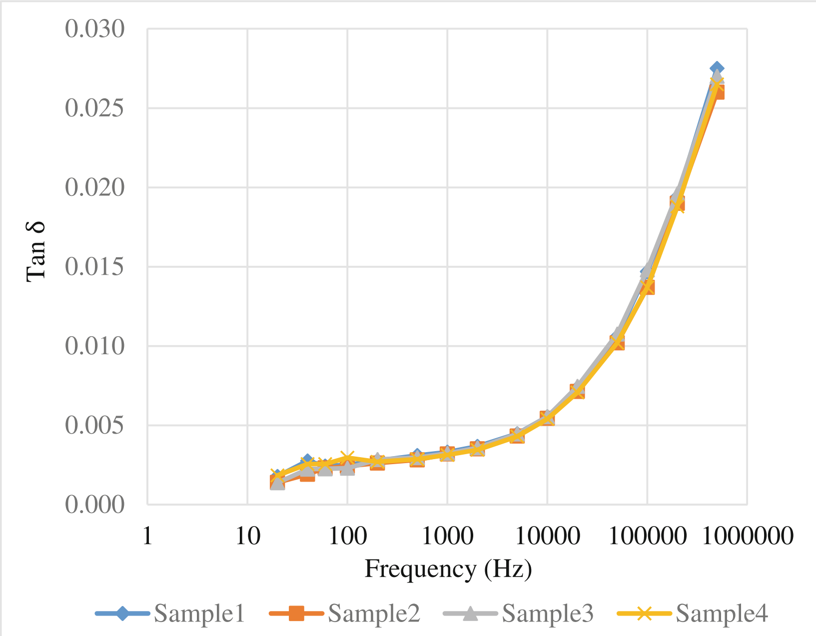


Fig. 4. Tanδ of unaged samples versus frequency.

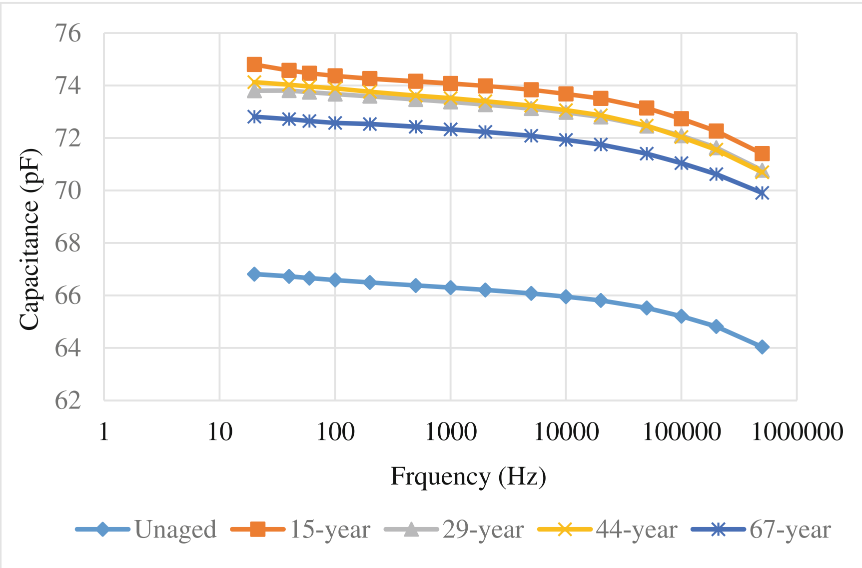


Fig. 5. Variation in the capacitance of aged samples versus frequency.

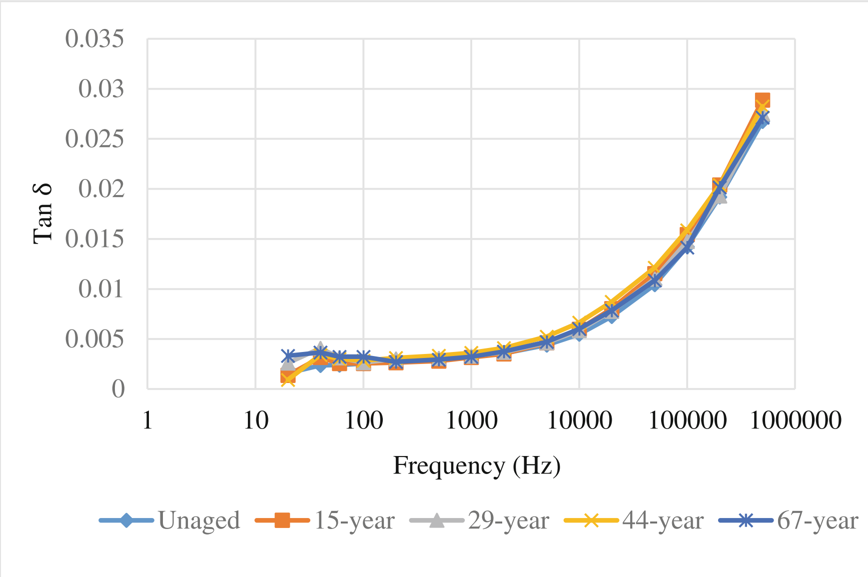


Fig. 6. Variation in the tanδ of aged samples versus frequency.

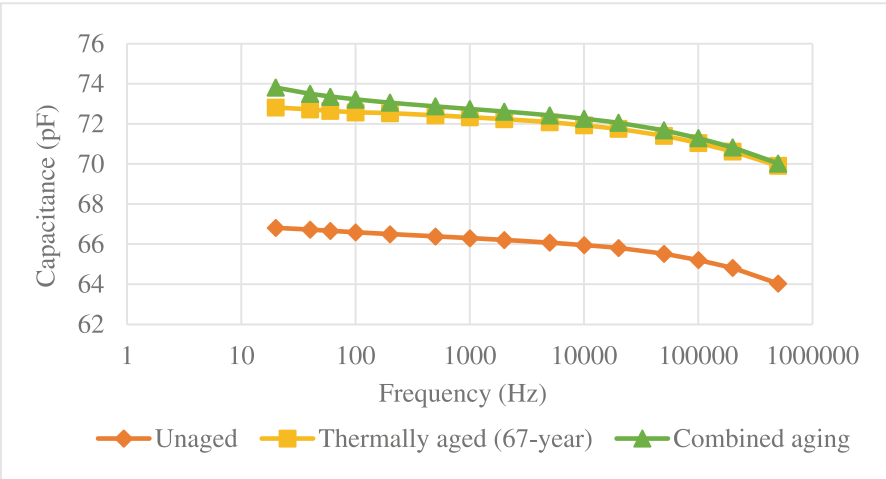


Fig. 7. Variation in capacitance under different aging mechanisms.

High difference between the two aging mechanisms was observed at lower frequencies so, at lower frequencies with combined stress, the insulating material greatly lost its capacitive property since the high increase in the capacitance was observed also the increase in the tanδ means high leakage current drawn by the insulating material.

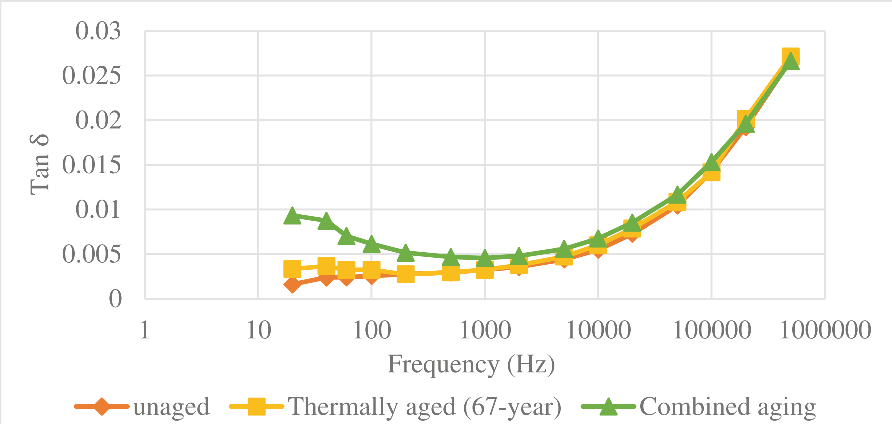


Fig. 8. Variation in tanδ under different aging mechanisms.

Laborelec Lab, Belgium created the EaB tests for the same cable samples under the same thermal aging conditions [9]. Correlating the electrical tests, tanδ values and the mechanical tests, EaB values is given in Fig. 9 showing that increasing the aging time decreases the EaB, which shows that the cable samples lost their mechanical properties while the higher values of the tanδ with the increase in the aging time shows that the insulation started to lose its electrical properties.

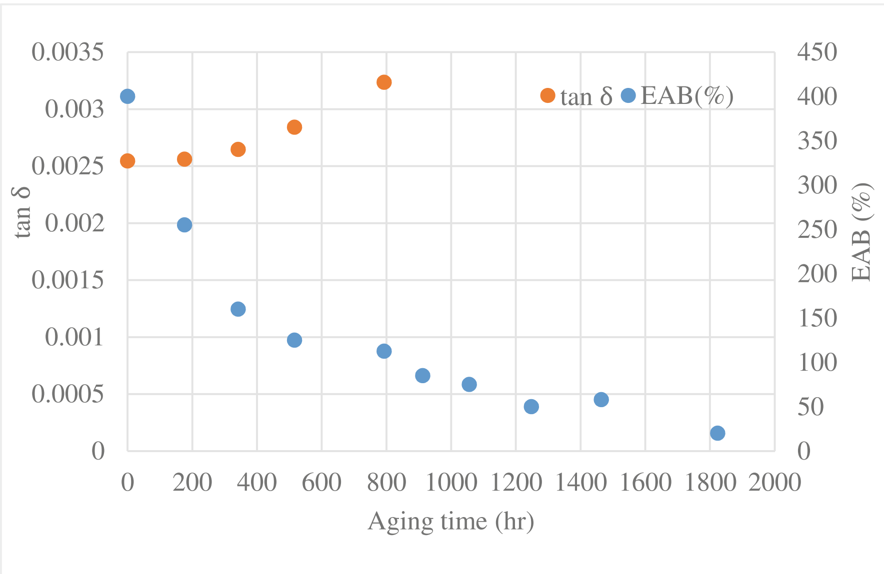


Fig. 9. Correlation between electrical and mechanical monitoring tests.

5 Conclusion

In this research work, low voltage NPP power cable samples were thermally aged under 120 °C for four different thermal cycles. The effect of each thermal cycle on the insulation integrity was done by investigating the capacitance and tanδ over a frequency range from 20 Hz to 500 kHz. It was observed that with increasing the aging time and frequency, the capacitance was decreased and tanδ was increased. Also, the cable samples were subjected to mechanical bending stress after the thermal aging and it was found that the capacitance and tanδ were increased compared to the thermally aged samples and a large difference was observed at lower frequencies showing that the insulation materials lost their capacitive function.

The correlations between tanδ values and the EaB values for the same cable samples at the same thermal aging conditions show that the cables EaB decreased while the tanδ increased with increasing aging time. Further chemical investigations are needed to better understand the presented behavior of tanδ.

Acknowledgments. This research work has been implemented under Project No. 123672 with the support provided from the National Research, Development and Innovation Fund of Hungary, financed under the KNN_16 funding scheme.

References

1. Banford, M., Fouracre, R.A.: Nuclear technology and ageing. IEEE Electr. Insul. Mag. **15**, 19–27 (1999)
2. Verardi, L., Fabiani, D., Montanari, G.C.: Electrical aging markers for EPR-based low-voltage cable insulation wiring of nuclear power plants. Radiat. Phys. Chem. **94**, 166–170 (2014)
3. Linde, E., Verardi, L., Fabiani, D., Gedde, U.W.: Dielectric spectroscopy as a condition monitoring technique for cable insulation based on crosslinked polyethylene. Polym. Test. **44**, 135–142 (2015)
4. Plaček, V., Kohout, T., Kábrt, J., Jiran, J.: The influence of mechanical stress on cable service life-time. Polym. Test. **30**, 709–715 (2011)
5. Linde, E., Verardi, L., Pourmand, P., Fabiani, D., Gedde, U.W.: Non-destructive condition monitoring of aged ethylene-propylene copolymer cable insulation samples using dielectric spectroscopy and NMR spectroscopy. Polym. Test. **46**, 72–78 (2015)
6. Seguchi, T., Tamura, K., Ohshima, T., Shimada, A., Kudoh, H.: Degradation mechanisms of cable insulation materials during radiation-thermal ageing in radiation environment. Radiat. Phys. Chem. **80**, 268–273 (2011)
7. Asipuela, A., Mustafa, E., Afia, R.S.A., Adam, T.Z., Khan, M.Y.A.: Electrical condition monitoring of low voltage nuclear power plant cables: tanδ and capacitance. In: IEEE International Conference on PGSRET, pp. 1–4 (2018 in press)
8. Mustafa, E., Afia, R.S.A., Adam, T.Z.: A review of methods and associated models used in return voltage measurement. In: IEEE International Conference on Diagnostics Electrical Engineering, pp. 69–72 (2018)

9. Series, I.T.: Benchmark Analysis for Condition Monitoring Test Techniques of Aged Low Voltage Cables in Nuclear Power Plants. Iaea-Tecdoc-1825, p. 179 (2017)
10. Afia, R.S.A., Mustafa, E., Adam, T.Z.: Mechanical stresses on polymer insulation materials. In: IEEE International Conference on Diagnostics Electrical Engineering, pp. 170–173 (2018)

Thermal Degradation and Condition Monitoring of Low Voltage Power Cables in Nuclear Power Industry

Ehtasham Mustafa[1,2(✉)], Tamus Zoltán Ádám[2], Ramy S. A. Afia[2,3], and Angel Asipuela[2]

[1] Department of Electrical Engineering, Faculty of Engineering & Technology, Gomal University, Dera Ismail Khan, Pakistan
[2] Department of Electric Power Engineering, Budapest University of Technology and Economics, Budapest, Hungary
{mustafa.ehtasham, tamus.adam}@vet.bme.hu, angel_electon_ppf@hotmail.com
[3] Department of Electrical Power and Machines Engineering, Helwan University, Cairo, Egypt
ramysaad@h-eng.helwan.edu.eg

Abstract. A strong electric power industry having diverse generation sources is a key for advances in the industry, agriculture, technology and standards of living for any nation. In the last few decades, the nuclear power industry has emerged as a strong competitor in the electric power production market due to its less carbon oxide emission, high capacity, and efficiency. The reliable operation of nuclear power plants is heavily dependent on the safe operation of the low voltage instrumentation and control and power cables, which are mostly installed in the containment area and are under thermal and radiation stresses which may decrease the expected life of the cable. In this research work, the behavior of Choloro-Sulfonated Polyethylene/Cross-Linked Polyethylene based low voltage nuclear power plant power cable samples under thermal stress have been studied with the help of non-destructive extended voltage response method. The effect of the thermal stress on the expected life of the cable are also discussed.

Keywords: Nuclear power industry · Condition monitoring · Low voltage cables · Thermal aging

1 Introduction

For any nation advancement in the industry, agriculture and technology and standards of living strongly depend on a strong electric power industry having a diverse generation sources. In the last few decades, out of the generation sources, nuclear power industry (NPI) has emerged as a strong competitor in the electric power production market due to its number of advantages such as less carbon oxide emission, high capacity, and efficiency. For any nuclear power plant (NPP) to operate reliably, it is important that the low voltage (LV) instrumentation and control (I&C) and power

Published by Springer Nature Switzerland AG 2019
L. M. Camarinha-Matos et al. (Eds.): DoCEIS 2019, IFIP AICT 553, pp. 405–413, 2019.
https://doi.org/10.1007/978-3-030-17771-3_35

cables must operate safely. Since, most of the LV cables are installed in the containment area hence, they are under a number of stresses such as environmental, electrical, mechanical and thermal [1]. These stresses affect the polymer insulation of the cables, electrical and mechanical properties, and hence may decrease the expected life of the cable. Practically, the lifetime of the cable is evaluated by subjecting the cables specimen under accelerated aging tests and then the integrity of them are monitored through the condition monitoring (CM) techniques. These tests provide either the critical properties or diagnostic properties by destructive or non-destructive tests, respectively. The International Atomic Energy Agency (IAEA) has narrated the desirable features of the CM techniques in its report [2], on the other hand, the CM techniques used in the recent times are also discussed in number of literature [3–8]. In spite of these research, still there is a search of an effective CM technique as the research results in regard to the effect of aging on the integral property i.e., electrical is still not sufficient. The main purpose of this paper is to search the electrical aging markers like specific and polarization conductivities and then to assess the condition of the LV cable insulation.

Out of the diagnostics tests, Extended Voltage Response (EVR) method has been successfully adopted for the CM of many insulations used in electric power equipment with the background to study the slow polarization processes in the insulation. The main advantage of the technique is its non-destructive nature. In this research work, the performance of LV NPP power cables based on Choloro-Sulfonated Polyethylene (CSPE)/Cross-Linked Polyethylene (XLPE) insulation under the thermal stress has been analyzed by the EVR method.

2 Relationship to Industrial and Service Systems

Electrical energy can be harvested either by burning energy sources such as natural gas, oil, coal (thermal energy) and nuclear or by taking advantage of renewable sources such as hydro, solar, wind and biomass. Out of the thermal energy sources, the NPI has an advantage of low operating cost with the drawback of high capital cost. The initial design of an NPP is 40 years but its reliable and safe operation is highly dependent on the services, structures, and components. In the NPP, there is about 1,000 km of long LV cables, which are exposed to thermal, radiation, mechanical and electrical stresses. With these stresses, the integrity of the cables decreases with the passage of time, aging. It is important to study the aging behavior of the cable during the operation. Also, in the case of extension in the service life of NPP, these cables must be capable of safe operation as the cable-replacement is cost-prohibitive and impractical [9]. Even though a substantial amount of effort has been expended on the degradation mechanism in the cable polymer insulation and the CM techniques, but with the challenge that each polymer has its own composition, geometry and passes through different design and fabrication processes, it is difficult to state the real cause of the cable aging in any particular polymer. Also, the effort to standardize the CM techniques is still an interesting task. Keeping in view the challenges, this research is focused on one such CM technique, EVR used to diagnose the thermally aged CSPE/XLPE based LV NPP power cable.

3 Extended Voltage Response

The electrical aging markers, stating the condition of an insulation can be obtained either by time domain analysis or frequency domain analysis. In the time domain analysis, measurement of return voltage is one such technique, with the main idea behind the technique to study the slow dielectric polarization processes, possessing high time constants, higher than 1 s [10]. Two new techniques emerged from this idea, Voltage Response (VR) and Return Voltage Measurement (RVM) method [11]. Both the methods have been adopted successfully for the condition monitoring of a wide range of the insulating materials. An advanced version of the VR method, EVR has been introduced in recent times with more charging time, 4000 s instead of 100–1000 s and discharging times to study a wide range of polarization spectrum. The method has shown its applicability on diverse high and low complex insulations materials [10, 12–14].

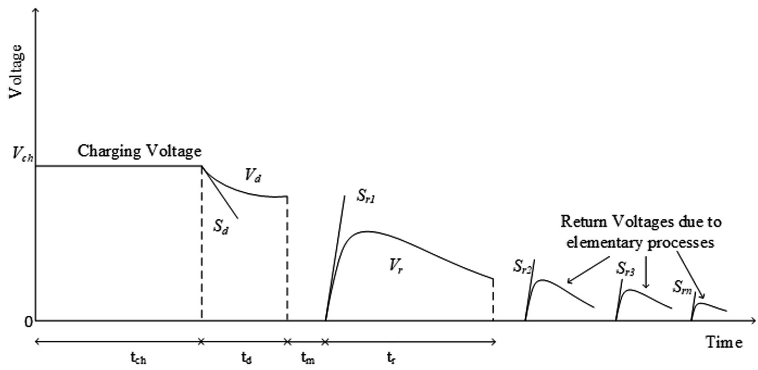


Fig. 1. Extended voltage response timing diagram.

In the EVR method, Fig. 1, the insulation is charged through a DC voltage, V_{ch} for a period, T_C. After a shortening period, the return voltage is measured across the insulation, representing the depolarization or relaxation processes inside the insulation material which did not relax during the short circuit period. Two tangents of both the decay voltage (S_d) and return voltage (S_r) are measured, showing the relationship between the specific conductivity (γ) of the insulation, Eq. (1) and the polarization conductivity (β) of the insulation, Eq. (2), respectively.

$$S_d = \left(V_{ch} / \in_o \right) \gamma \tag{1}$$

$$S_r = \left(V_{ch} / \in_o \right) \beta \tag{2}$$

Since there are a number of relaxation times in the insulation owing to the inhomogeneity of the material, so each relaxation time has its own time constant and correspondingly showing different return voltages [15] of the elementary processes.

4 Experimental Work

4.1 LV NPP Cable Specimen and Thermal Aging

The LV NPP power cable consists of the conductor, primary insulation of XLPE and outer insulation (jacket) of CSPE. Figure 2 shows the cross-sectional view of the cable sample.

Fig. 2. Cross-sectional view of the cable sample.

The CSPE is a special elastomer which is produced by the modification and functionalization of the polyethylene through simultaneous actions of chlorine (Cl) and Sulphur dioxide (SO_2) on PE dissolved in the CCl_4 in the presence of the radical initiators. The specification of the cable is given in Table 1.

Table 1. Specification of the cable.

Parameters	Values
Conductor	Stranded tin coated copper
Outer insulation/Jacket	CSPE
Outer insulation/Jacket thickness	0.762 mm
Primary insulation	XLPE
Primary insulation thickness	1.143 mm

The accelerated thermal aging of four cable samples was carried out in the oxygen controlled oven. The accelerated aging hours were calculated using the Arrhenius model. The model relates the degradation rate to temperature [3, 16] and is widely used for the thermal aging evaluation of the polymers in nuclear-qualified cables, it addition it is helpful in determining different ambient temperature influencing the polymer insulation of the cables. Table 2 shows the thermal exposure time (hours) of cable in the oven to the equivalent service time (years).

Table 2. Oven time (hours) and service time (years) for CSPE/XLPE cable samples.

Input aging parameters	Service time (years)	Oven time @120 °C (hours)
Service temperature: 60 °C	8	176
Oven temperature: 120 °C	16	342
Activation energy of CSPE/XLPE, $E_a = 1.13\,\text{eV}$	24	516
	36	793

4.2 EVR Measurements

For the EVR measurements, the cable sample was covered at the center by an aluminum foil and was connected to the ground terminal and the cable end was connected to the supply terminal, 1000 VDC. The equipment has connection with a static voltmeter and a computer for recording the charging, discharging times and the slope values. The S_r versus the shorting times for the unaged cable samples is shown in Fig. 3. The return voltages at the longer discharging times are a sign that there are intensive higher time constant polarizations in the insulation. Figure 4 shows the S_r for the samples after each thermal aging cycles. It is interesting to note that the values of S_r decreased as the thermal cycle time increased showing that the relaxation process has decreased.

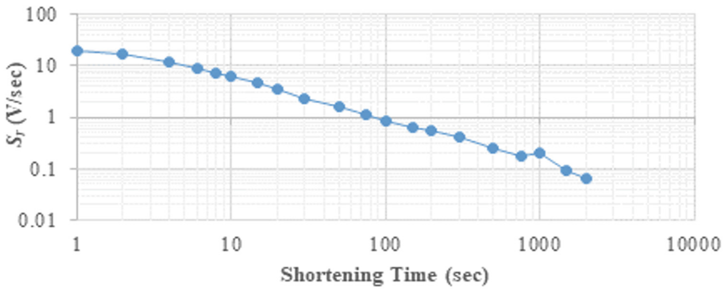


Fig. 3. S_r for unaged samples, existence of slow polarization process.

While the conductivity inside the material has been shown by the S_d plot, Fig. 5. The values of S_d first decreased then increased sharply and the end of the fourth thermal cycle the slope decreased.

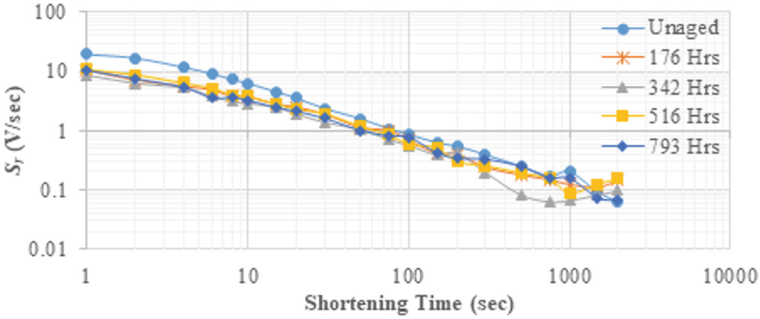


Fig. 4. S_r for aged and unaged cable samples.

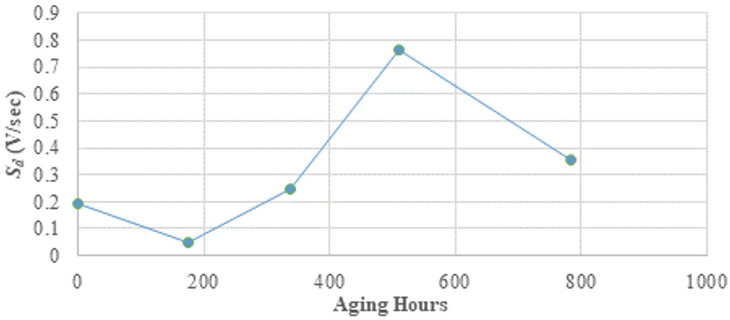


Fig. 5. S_d versus aging hours.

5 Results and Discussion

It is known that when a DC voltage is applied to an insulation material, three types of currents are generated in the material; leakage current due to the insulation resistance, absorption and charging current due to the capacitance of the insulation [17]. The leakage and absorption currents are more prominent in the case the AC voltage is applied while the charging current is more prominent under the DC voltage. Due to this charging current, there is the phenomenon of dielectric relaxation, which is related to the number of dipoles and their mobility. In CSPE, the dipoles are due to the C-Cl bond which is broken during the thermal aging and releases the chlorine, dehydrochlorination [18]. This, in turn, reduces the number of dipoles and hence the dielectric relaxation. In addition to this, since CSPE is a thermoset polymer, so due to the thermal aging there is a phenomenon of cross-linking, increasing with aging, which also affects the relaxation process. So, dehydrochlorination and cross-linking could be the reasons for the decrease values of S_r with aging, Fig. 6.

The profile of the S_d, Fig. 5, could be due to the presence of fillers, which are used in the polymer composites to increase the electrical and mechanical properties and hence effecting the conductivity of the material [19]. As shown in the Fig. 5, the

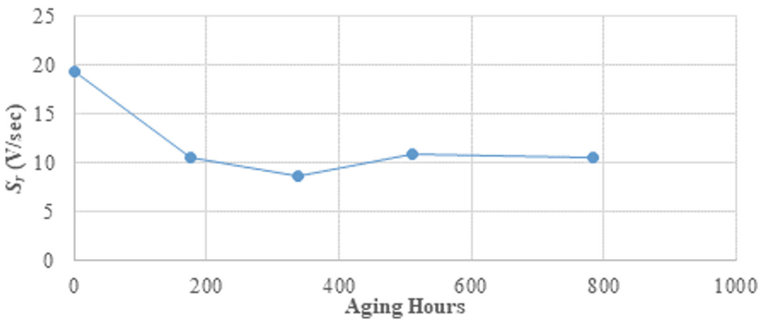


Fig. 6. S_r versus aging hours cable samples.

conductivity, S_d, decreased after the first thermal cycle, this could be attributed due to the polymer-filler interaction or the interface, which acts as the barrier to the conduction of the charge carrier in the conductor-polymer system. As the thermal stress increased, there is an increase in the electric charge which may affect the polarization of the fillers in the polymer matrix. This may result in the excess negative charge and is shown as a sharp increase in the value of S_d. After the fourth cycle, the conductivity decreased which may be due to the creation of the cross-links and dehydrochlorination, and evaporation or the reduction of the polar molecules. A further chemical investigation could be helpful in this regard [20, 21].

6 Conclusion

In this research work, CSPE/XLPE based LV NPP power cables were thermally aged under 120 °C for four different thermal cycles. The change in the polarization processes in the polymer insulation has been studied using the diagnostic method, EVR. It was observed that due to the formation of cross-linking and breakage of the C-Cl bond, resulting in a decrease of the number and mobility of dipoles the values of S_r decreased as the aging stress increased. While the values of S_d for the first cycle decreased owing to the formation of the filler-polymer bond, which may act resistive path to the charge carriers. While for the next two cycles the slope increased which may be attributed to the polarization of the filler matrix. The decreased in the slope at the end of the fourth cycle may be due to the decrease in the charge carriers and the formation of the cross-links.

From the curves of S_d and S_r, it could be concluded that in CSPE due to the thermal aging, the leakage and absorption current are more dominant as compared to the charging current as the number and mobility of the dipoles are reduced. A good chemical investigations will be helpful in backing these results.

Acknowledgments. This research work has been implemented under Project No. 123672 with the support provided from the National Research, Development and Innovation Fund of Hungary, financed under the KNN_16 funding scheme.

References

1. Afia, R.S.A., Mustafa, E., Tamus, Z.Á.: Mechanical Stresses in polymer insulating materials. In: IEEE International Conference on Diagnostics in Electrical Engineering (Diagnostika), pp. 170–173, IEEE Press, New York (2018)
2. International Atomic Energy Agency (IAEA): Pilot study on the management of aging of instrumentation and control cables, vol. 2 (1997)
3. International Atomic Energy Agency (IAEA): Assessing and managing cable aging in nuclear power plants, Vienna, Austria (2012)
4. IAEA TECDOC Series: Benchmark Analysis for Condition Monitoring Test Techniques of Aged Low Voltage Cables in Nuclear Power Plants, Vienna, Austria (2017)
5. Subudhi, M.: Literature Review of Environmental Qualification of Safety-Related Electric Cable, New York, USA (1996)
6. Initial Acceptance Criteria Concepts and Data for Assessing Longevity of Low-Voltage Cable Insulations and Jackets, California, USA
7. Blocker, E., Smith, S., Philpot, L., Conley, J.: Aging management guideline for commercial nuclear power plants-tanks and pools (1996)
8. Brookhaven National Laboratory: Condition Monitoring of Cables, Task 3 Report: Condition Monitoring Techniques for Electric Cables, BNL-90735-2009-IR (2009)
9. Fifield, L.S., Duckworth, R., Glass III, S.W.: Long-term operation issues for electrical cable systems in nuclear power plants. In: Proceedings of 24th International Conference Nuclear Engineering, vol. 2, pp. 1–5 (2016)
10. Tamus, Z.Á., Csányi, G.M.: Modeling of insulations by the results of voltage response measurement. In: The 19th International Symposium on High Voltage Engineering, pp. 6–10. Czech Republic, Pilsen (2015)
11. Mustafa, E., Afia, R.S.A., Tamus, Z.Á.: A review of methods and associated models used in return voltage measurement. In: IEEE International Conference on Diagnostics in Electrical Engineering (Diagnostika), pp. 69–72. IEEE Press, New York (2018)
12. Tamus, Z.Á., Csányi, G.M., Szirmai, Á., Nagy, A.: Insulation diagnostics of high voltage equipment by dielectric measurements – hungarian research and experience. In: International Scientific Symposium on Electrical Power Engineering, pp. 7–13 (2016)
13. Csányi, G.M., Tamus, Z.Á.: Investigation of dielectric properties of mixed PILC and XLPE cable insulation by the extended Voltage Response Method. In: 6th International Youth Conference on Energy, IYCE 2017, pp. 2–5 (2017)
14. Tamus, Z.Á., Szirmai, Á., Nemeth, B.: Comparison of voltage response and return voltage measurements of a transformer insulation model. In: The 19th International Symposium on High Voltage Engineering, ISH 2015 (2015)
15. Kao, K.C.: Dielectric Phenomena in Solids with Emphasis on Physical Concepts of Electronic Processes (2004)
16. Power, N., Committee, E.: IEEE Guide for Assessing, Monitoring, and Mitigating Aging Effects on Class 1E Equipment Used in Nuclear Power Generating Stations (2000)
17. Shin, Y.D.: Phase degree of response voltage and current versus excitation voltage of the accelerated thermally aged CSPE. Trans. Electr. Electron. Mater. **19**, 112–117 (2018)
18. Chailan, J.F., Boiteux, G., Chauchard, J., Pinel, B., Seytre, G.: Effects of thermal degradation on the viscoelastic and dielectric properties of chlorosulfonated polyethylene (CSPE) compounds. Polym. Degrad. Stab. **48**, 61–65 (1995)
19. Nanda, M., Chaudhary, R.N., Tripathy, D.K.: Dielectric relaxation of conductive carbon black reinforced chlorosulfonated polyethylene vulcanizates. Polym. Compos. **31**, 152–162 (2009)

20. Pinel, B.: Effects of irradiation dose rate and thermal aging on changes in the properties of elastomeric materials (EPR-CSPE) for cable insulation and sheathing. In: 7th International Symposium on Electrets, pp. 944–949 (1991)
21. Chailan, J.F., Chauchard, J., Seytre, G., Boiteux, G., Escoubes, M., Pinel, B.: Effects of radiation and thermal ageing on physico-chemical properties of elastomeric materials (EPR-CSPE) for cable insulation and sheathing. In: Sixth International Conference on Dielectric Materials, Measurements and Applications, pp. 418–421 (1992)

Novel Design of the Converter for an Active UPS Application Based on Marx Modulator Concept with Supercapacitors

Celso Pestana[1], Hiren Canacsinh[1,2,4(✉)], Miguel Chaves[1,4(✉)], Paulo Gamboa[1,4], Armando Cordeiro[1], Ricardo Luís[1,4], Ana Martins[1], Luís Encarnação[1,2,4(✉)], Fernando Pereira[1], José F. Silva[3,4], and Nuno Santos[1]

[1] Instituto Superior de Engenharia de Lisboa, Instituto Politécnico de Lisboa, Rua Conselheiro Emídio Navarro, 1, 1959-007 Lisboa, Portugal
A33533@alunos.isel.pt,
{hic, mchaves, pjgamboa, acordeiro, ricardo, anamartins, luisrocha, fjp, ngsanto}@deea.isel.ipl.pt,
[2] GIAAPP, Grupo de Investigação em Aplicações Avançadas de Potência Pulsada, Rua Conselheiro Emídio Navarro, 1, 1959-007 Lisboa, Portugal
[3] Instituto Superior Técnico, Universidade de Lisboa, Av. Rovisco Pais, 1049-001 Lisboa, Portugal
fernando.alves@tecnico.ulisboa.pt
[4] INESC-ID, Instituto de Engenharia de Sistemas e Computadores – Investigação e Desenvolvimento, Rua Alves Redol, 9, 1000-029 Lisboa, Portugal

Abstract. A new Marx modulator 4-leg multilevel converter is proposed to replace the well-known 4-leg inverter fed from batteries in Uninterruptible Power Supply (UPS) applications. The novel design concept for the Marx modulator based multilevel converter is presented and described. Each leg uses 2 Marx modulator modules, each one using an Electrochemical Double-Layer Capacitor (EDLC) (also known as supercapacitor (SC)) energy storage bank. The new topology concept enables multilevel operation with five voltage level per phase, allowing high quality voltage and current waveforms, distributed supercapacitor storage and reduction of supercapacitor bank voltage. The proposed converter can operate as UPS or as Active Power Filter (APF), powering non-linear or/and unbalanced loads, while balancing each leg supercapacitor voltage. A control strategy to choose the redundant vector configuration to balance the supercapacitor voltages is explained. Simulation results are presented for a Marx modulator based interactive 40 kVA UPS.

Keywords: Uninterruptible Power Supply · Supercapacitors · Bipolar Marx modulator multilevel converter

Published by Springer Nature Switzerland AG 2019
L. M. Camarinha-Matos et al. (Eds.): DoCEIS 2019, IFIP AICT 553, pp. 414–423, 2019.
https://doi.org/10.1007/978-3-030-17771-3_36

1 Introduction

Nowadays, electrical power quality is a concern for customers with loads that are sensitive to the quality of the voltage waveform and are responsible to immunizing their own facilities. The electrical distribution is often subject to grid distortion in terms of periods of variation and intensity (non-linear loads, etc.). However, some critical loads need a stable voltage supply and higher quality than the transmission and distribution network operators can satisfy. Although the less expensive methods to achieve protection against some power problems, none of them can reasonably isolate the system from power supply problems as well as good uninterruptible power supply (UPS).

UPSs are systems that are developed to supply sensitive loads such as telecommunication systems, computer sets, hospital equipment or production lines in the manufacturer industries [1], when there are disturbances in the electrical network for a limited period. UPS can also correct other network events such as overvoltage, voltage variations, noise, frequency instability or harmonic distortion. UPS differentiate themselves from other emergency systems by enabling instantaneous power supply when a power outage occurs in the network through electronic circuits associated with energy storage systems, such as flywheel for high power systems or batteries, and supercapacitors for low power systems, [2, 3]. For all these storage systems, it is necessary to use electronic power converters. In the case of a low-power UPS, the classic solution consists of the use of a DC-DC converter and a 4-leg inverter and a bank of batteries or supercapacitors (SC). In the case of supercapacitors, they have a high lifetime of more than 500 thousand cycles with discharge cycles in the order of 100%. However, its low working voltage requires the use of series of supercapacitors for the required voltage. In this way, the use of supercapacitors in series brings with it the need of voltage equalization, to prevent overvoltage and possible damage of the supercapacitors, which can become unbearable.

Thus, the investigation of the issues associated with the use of modular converters with capacity to mitigate the problem of voltage equalization in supercapacitor banks is relevant. An alternative solution is the multilevel converter, [4, 5]. The applications indicated for multilevel converters range from the interface between high-voltage transmission networks to variable-speed drives for use in medium-power drive systems. Different multilevel conversion topologies are described in the literature, being the most studied and most used: the multilevel converter with diodes of connection to the neutral point; the multilevel flying capacitor converter and the multilevel converters of half-bridges and complete cascade bridges [6–13]. Cascade converters are modular, however, they require independent DC voltage sources making them expensive. There is reference to the use of the cascade bridge converter in high power applications demanding from the point of view of the quality of energy [14]. Considering the first two converters, it is necessary to control the converters to ensure a balanced distribution of the DC bus voltage across all capacitors. Most of the techniques used to balance the voltage in the capacitors use the redundant configurations from the point of view of the output AC voltages, but they are different in relation to the current flows in the DC bus. Proper selection of the redundant configuration allows the equalization of the voltages in the capacitors to be achieved. This requirement increases the complexity of the control of the system, especially for a number of levels greater than 3, which may prevent the smooth functioning of the multilevel converter. Thus, it is current and fundamental to investigate new topologies of modulator converters, for issues associated with voltage equalization in supercapacitors bank, and in particular the topology of the Bipolar Marx Generator (BMG), [15–21].

The topology of the BMG is a modular circuit and it's largely used in the Pulsed Power field, which uses a switching system to charge a set *n* of capacitors in parallel from a DC voltage source, with a reduced amplitude and produces transiently a voltage whose amplitude approaches a multiple of the input DC voltage value. Thus, due to the characteristic of charging the capacitors in parallel, it is relevant to study the topology of the BMG to mitigate the voltage equalization problem in supercapacitors banks, under the UPS application. Due to the modularity characteristic of the BMG topology, the proper selection of the redundant configuration in terms of the output voltage can further contribute to achieve the equilibrium of supercapacitors bank voltages.

This paper presents the preliminary simulation results of a novel proposal for the Marx based converter for an active type UPS (Fig. 1), to minimize the impact of the voltage equalization between the set of supercapacitors bank.

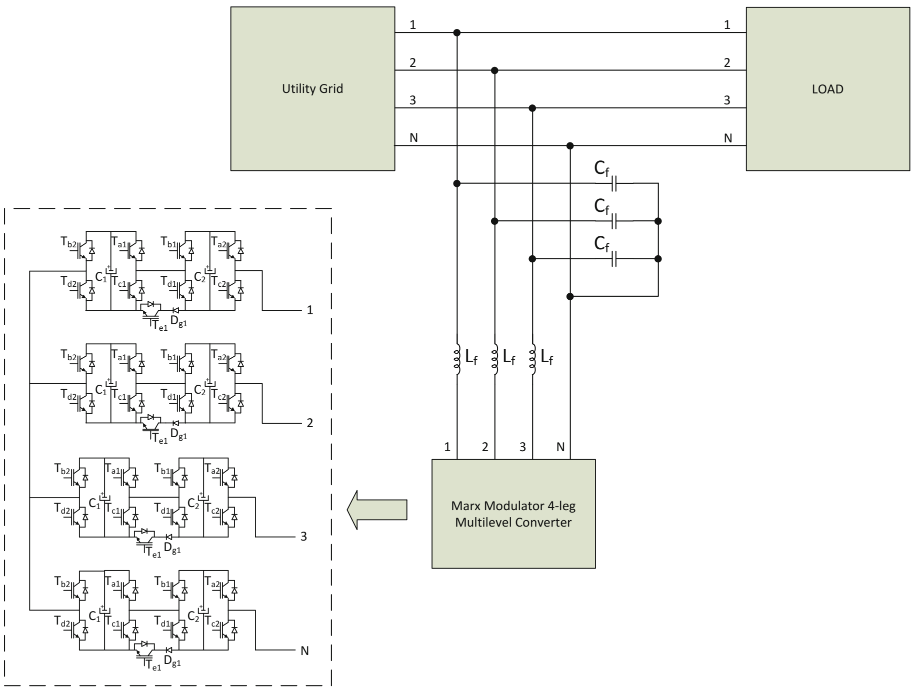


Fig. 1. General scheme of an active UPS.

Considering the general scheme of an active UPS represented in Fig. 1, the storage system/converter part consist in 4-leg Marx modulator multilevel converter as represented also in Fig. 1, and an output L_fC_f filter due to the distorted pulse-width-modulated (PWM) waveform generated by the converter.

2 Relationship to Industrial and Service Systems

In an increasingly competitive society focused on the value generation for customers through innovative business models instead of merely selling products, and namely for applications with sensitive loads needing UPS such as telecommunication systems, computer sets or hospital equipment's, the study of different type of approach in terms of the converter, is relevant.

In this paper we present a bipolar Marx multilevel converter with supercapacitors, which can operate as a UPS or as APF, powering non-linear or/and unbalanced loads, while generating 5 voltage level per phase and allowing high quality voltage and currents waveforms with distributed supercapacitor storage and reduction of supercapacitor bank voltage.

3 Circuit Operation

Considering the circuit of Fig. 1, each converter leg has one solid state bipolar Marx modulator with two stages. Thus, it is considered only one converter leg to explain the circuit operation as UPS and as an active filter, as represented in Fig. 2.

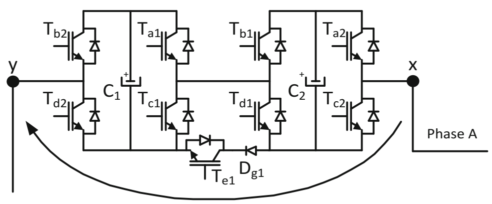


Fig. 2. Bipolar Marx modulator of the storage/converter system.

Figure 2 shows the proposal topology of the converter based on Marx concept with two modules, capable of delivering repetitive bipolar output voltage to different types of load, using insulated gate bipolar transistor (IGBTs) as ON–OFF switches. Each module consists in 4 switches T_{ai}, T_{bi}, T_{ci} and T_{di} and one supercapacitor bank C_i, which consist of supercapacitors in series. The modules are interconnected through the switch T_{e1} and diode D_{g1}. The operating modes of the circuit presented in Fig. 2 are described in the following sections both as UPS and as an active filter.

The circuit of Fig. 2 allows different paths to generate different levels of voltage. Considering that the supercapacitor banks C_1 and C_2 are equally charged and U_{C1} = U_{C2} = Us/2, and for instance, if a (+Us/2) voltage level is needed between XY, switches T_{d2}, T_{a1}, T_{d1} and T_{c2} (or switches T_{d2}, T_{a1}, and T_{b1} and T_{a2}) are turned ON and all others are OFF, using the supercapacitor bank C_1. Another path can be considered, using the supercapacitor bank C_2 voltage through the switches T_{a2}, T_{d1}, and T_{c1}, T_{d2} (or switches T_{a2}, T_{d1}, and T_{a1} and T_{b2}), as represented in Table 1. Other output voltage levels paths are summarized in Table 1.

Table 1. Semiconductors in On-state driven for different operating modes of the circuit of Fig. 2.

XY output voltage	Path	SC bank used
$+U_S$	T_{a1}, T_{d1}, T_{a2}, T_{d2}	C_1 and C_2
$+U_S/2$	T_{d2}, T_{a1} and [(T_{d1}, T_{c2}) or (T_{b1}, T_{a2})] or T_{a2}, T_{d1} and [(T_{c1}, T_{d2}) or (T_{a1}, T_{b2})]	C_1 or C_2
0	T_{c1}, T_{d2} and [(T_{d1}, T_{c2}) or (T_{b1}, T_{a2})] or T_{a1}, T_{b2} and [(T_{b1}, T_{a2}) or (T_{d1}, T_{c2})] or (T_{a1}, T_{d2}, T_{b1}, T_{c2}) or (T_{a1}, T_{d2}, T_{d1}, T_{a2})	- - (C_1-C_2) or ($-C_1+C_2$)
$-U_S/2$	T_{b2}, T_{c1} and [(T_{d1}, T_{c2}) or (T_{b1}, T_{a2})] or T_{c2}, T_{b1} and [(T_{c1}, T_{d2}) or (T_{a1}, T_{b2})]	C_1 or C_2
$-U_S$	T_{c1}, T_{b1}, T_{c2}, T_{b2}	C_1 and C_2

Besides the five levels of voltage described under Table 1, the topology of the Fig. 2 allows the equalization of the voltages between the supercapacitor banks C_1 and C_2, and only from C_1 to C_2 through the path of switches T_{a1} and T_{e1} and antiparallel diode of switch T_{b1} and diode D_{g1}.

3.1 Control Strategy

The converter is controlled in order to achieve the predefined technical goals: control AC load voltages, when in UPS operating mode and achieve sinusoidal grid currents, when in active filter operating mode. For this purpose, the AC current control is the first control loop. Figure 3 shows a block diagram of this control loop. The converter output phase currents are controlled using a sliding mode hysteretic controller, with reference signals given by the phase current dq components references.

Chosen the output voltage to be applied by the converter, the switching path is now selected between the available redundant configurations. The switching path is chosen, using a cost function which leads to the minimum error between the SC voltage and the reference that is previously stablished.

Upon current control loop, other controllers may be designed according to the system functions, such as: output AC voltage control for UPS standalone operation; active filtering and reactive power control and U_{dc} voltage control in respect to SC charge.

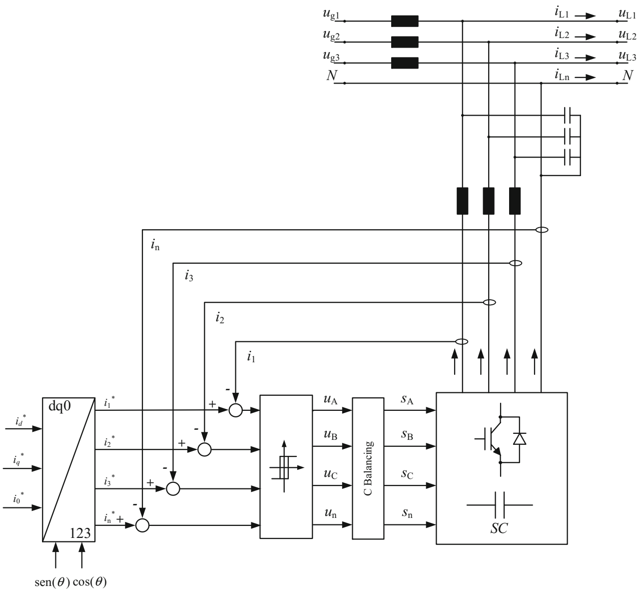


Fig. 3. General control diagram of the storage/converter system.

3.2 UPS Operating Mode

Under UPS operating mode, it is considered that all banks of supercapacitors are charged with Us/2. Thus, when a failure in the utility grid is detected, the grid is disconnected, and the system starts the operation of the BMG converters to supply the load in standalone, from the point the grid was interrupted once the controller is synchronous with the utility grid.

The converter output current components dq0 references are obtained from a predictive approach to control the load voltages [22], Fig. 4(a).

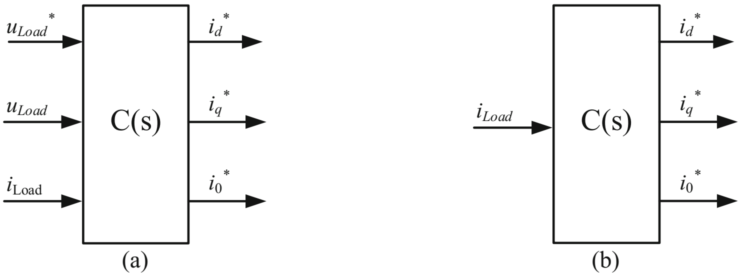


Fig. 4. Converter output current components dq0 references to control the: (a) load voltages in UPS operating mode; and (b) load currents in filter operating mode

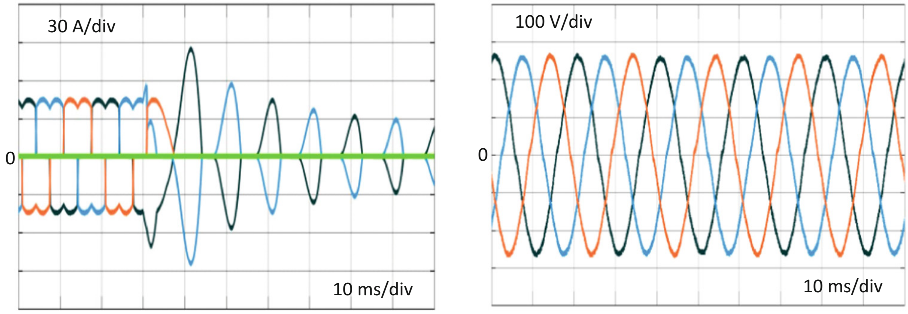
3.3 Active Power Filter Operating Mode

When in active filter operating mode, the main goal is to obtain sinusoidal current in the grid phases with almost unity input power factor. The converter supplies the total reactive and deformation power to the load and the grid provide only the fundamental component. In this operating mode, the converter output current components dq0 references are obtained extracting the average value of the fundamental component from the load currents [23, 24], Fig. 4(b).

4 Simulation Results

The circuit of Fig. 1 was simulated using Matlab/Simulink software. The equivalent capacitance of each supercapacitor bank is 2F. The parameters of the $L_f C_f$ filter are 600 mH and 5.6 μF respectively. The circuit of Fig. 1 was simulated as UPS (Fig. 5) and as APF (Fig. 7), powering non-linear or/and unbalanced loads. It was considered to show the simulation results for the worst load case, which is a non-linear and unbalanced load.

In Fig. 5 it is presented the simulations results for the current and voltage waveforms into a non-linear unbalanced load. Under this situation, the load was changed from linear to non-linear unbalanced load.



(a) (b)

Fig. 5. Simulation results of the (a) current and (b) voltage waveforms of the circuit of Fig. 1 operating as UPS and with a non-linear unbalanced load. The scales are 10 ms/div (horizontal) and (a) 30 A/div (vertical) and (b) 100 V/div (vertical).

Considering the simulations results as UPS operating mode (Fig. 5) the load voltage follows the three-phase voltages references for a non-linear and non-balanced load.

Figure 6 shows the voltages waveforms of two SC banks of one converter leg when operating as UPS and powering linear and non-linear or/and balanced and unbalanced loads. The results show that the controller strategy used can balance the SC banks voltage of each converter leg.

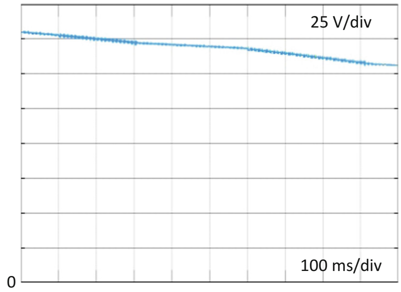


Fig. 6. Simulation results of the voltage waveforms in supercapacitor banks of the Marx modulator 4-leg multilevel converter, operating with linear and non-linear or/and balanced and unbalanced loads. The scales are 100 ms/div (horizontal) and 25 V/div (vertical).

Considering the active filter operating mode, Fig. 7 shows the current waveforms in the utility grid, the output currents of the Marx modulator 4-leg multilevel converter and the load currents, when operating with non-linear unbalanced load.

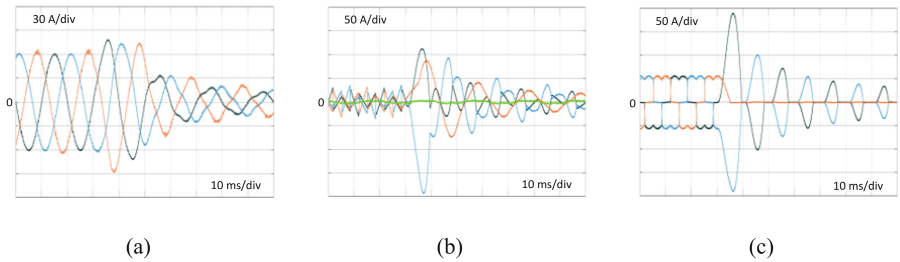


Fig. 7. Simulation results of the circuit of Fig. 1 operating as APF, for currents in utility grid (a), output currents of the Marx modulator 4-leg multilevel converter (b) and currents into non-linear unbalanced load (c). The scales are 10 ms/div (horizontal) and 30 A/div (vertical) and 50 A/div (vertical).

Considering the simulation results presented in Fig. 7, when the converter is in APF operation mode it can be seen that the converter imposes a three-phase balanced sinusoidal currents in the power grid, under non-linear unbalanced load condition.

5 Conclusion

A new design method for the converter for an active type UPS (Fig. 1) application, is proposed. The proposed concept allows the distribution of the energy necessary to supply the load along the converter legs. Yet, the modular characteristic of the Marx modulator converter allows to distribute the energy of each converter leg in equal parts as a function of the number of stages. Thus, the higher number of stages, the lower hold-off voltage of the supercapacitors bank. Also, the increase of the number of the stages in Marx modulator, allows to perform more levels which can allow to increase the electrical power quality.

Simulation results shows that the Marx based converter of the circuit of Fig. 1, with 2 stages in each converter leg, can produce a three-phase sinusoidal voltage into a linear and non-linear or/and balanced and unbalanced load. Also, the simulation results show that the waveforms of the voltages between supercapacitor banks of each converter leg have similar progress, due to the implemented controller that can choose the best redundant configuration that fits to a specific moment.

Under the active filter operating mode, the simulation results show that the system provides the correction of the currents of the utility grid.

Acknowledgments. This work was supported by national funds through Fundação para a Ciência e a Tecnologia (FCT) with reference UID/CEC/50021/2019, and through Instituto Politécnico Lisboa (IPL) with reference MIPREQUATE/0712036/2018.

References

1. Zaitsu, H., et al.: Uninterruptible power supply system utilizing electric double-layer capacitors. In: Power Conversion Conference, Nagoya, Japan, 2–5 April 2007
2. Hadjiaschalis, I.: Overview of current and future energy storage technologies for electric power applications in renewable and sustainable. Energy Rev. **13**, 1513–1522 (2009)
3. Gambôa, P., Pinto, S., Silva, J., Margato, E.: Predictive optimal matrix converter control for a dynamic voltage restorer with flywheel energy storage. In: 35th Annual Conference of IEEE IECON 2009, Porto, Portugal, 3–5 November 2009
4. Baker, R.H., Bannister, L.H.: Electric power converter. US Patent 3867643, February 1975
5. Nabae, A., Takahashi, I., Akagi, H.: A new-neutral-point-clamped PWM inverter. IEEE Trans. Ind. Appl. **17**(5), 518–523 (1981)
6. Rodríguez, J., Lai, J.S., Peng, F.Z.: Multilevel inverters: a survey of topologies, controls, and applications. IEEE Trans. Ind. Electron. **49**(4), 724–738 (2002)
7. Lesnicar, A., Marquardt, R.: An innovative modular multilevel converter topology suitable for a wide power range. In: IEEE Bologna Power Tech Conference Proceedings, 23–26 June 2003 (2003)
8. Khomfoi, S., Tolbert, L.M.: Chapter 31: multilevel power converters. In: Rashid, M.H. (ed.) Power Electronics Handbook. Elsevier, Amsterdam (2006)
9. Cordeiro, A., Silva, J.F., Pinto, S.F., Santana, J.E.: Fault-tolerant design for a three-level neutral-point-clamped multilevel inverter topology. In: EUROCON-International Conference on Computer as a Tool (EUROCON), IEEE, Lisbon, pp. 1–4 (2011)
10. Chaves, M., Margato, E., Silva, J., Pinto, S.: Generalized state-space modelling for m level diode-clamped multilevel converters. In: International Symposium on Mathematical Methods in Engineering, MME10, Coimbra, pp. 149–156, Outubro 2010
11. Chaves, M., Margato, E., Silva, J., Pinto, S., Santana, J.: HVDC transmission systems: bipolar back-to-back diode clamped multilevel converter with fast optimum-predictive control and capacitor balancing strategy. Electr. Power Syst. Res. **81**(7), 1436–1445 (2011). Elsevier

12. Rocha, L., Fernando, S.J., Redondo, L.: Multilevel high voltage pulse generation based on a new modular solid state switch. IEEE Trans. Plasma Sci. **42**(10), 2956–2961 (2014). https://doi.org/10.1109/tps.2013.2296141
13. Rocha, L., Fernando, S.J., Redondo, L.: Seven levels unipolar/bipolar pulsed power generator. IEEE Trans. Plasma Sci. **44**(10), 2060–2064 (2016). https://doi.org/10.1109/tps.2016.2519269
14. Leopoldo, F., Rodríguez, J., Leon, J., Kouro, S., Portillo, R., Prats, M.: The age of multilevel converters arrives. IEEE Ind. Electron. Mag. **2**(2), 28–39 (2008)
15. Redondo, L.M., Canacsinh, H., Fernando, S.J.: Generalized solid-state marx modulator topology. IEEE Trans. Dielectr. Electr. Insul. **16**(4), 1037–1042 (2009)
16. Redondo, L.M., et al.: Solid-state marx type circuit for ISOLDE voltage target modulator. In: 2009 IET European Pulsed Power Conference, CERN, Geneva, Swiss, 21–25 September 2009
17. Tastekin, D., Blank, F., Lunk, A., Roth-Stielow, J.: Power supply with bipolar pulsed output voltage and high repetition rate based on a solid state Marx topology. In: Pulsed Power Conference, pp. 1377–1381 June 2011
18. Gao, L., Wang, D., Qiu, J., Liu, K.: All-solid-state pulse adder with bipolar high voltage fast narrow pulses output. IEEE Trans. Dielectr. Electr. Insul. **18**(3), 775–782 (2011)
19. Sakamoto, T., Nami, A., Akiyama, M., Akiyama, H.: A repetitive solid state marx-type pulsed power generator using multistage switch-capacitor cells. IEEE Trans. Plasma Sci. **40**(10), 2316–2321 (2012)
20. Canacsinh, H., Redondo, L.M., Fernando Silva, J.: Marx type solid-state bipolar modulator topologies: performance comparison. IEEE Trans. Plasma Sci. **40**(10), 2603–2610 (2012)
21. Canacsinh, H., Redondo, L.M., Fernando Silva, J., Schamiloglu, E.: Solid-state bipolar marx modulator modeling. IEEE Trans. Plasma Sci. **42**(10), 3048–3056 (2014)
22. Emadi, A., Nasiri, A., Bekiarov, S.B.: Uninterruptible power supplies and active filters. CRC Press, 28 October 2004. ISBN 9780849330353
23. Akagi, H., et al.: Instantaneous Power Theory and Applications to Power Conditioning. IEEE Press/Wiley-Interscience (2007)
24. Soares, V., Verdelho, P., Marques, G.D.: An instantaneous active and reactive current component method for active filters. IEEE Trans. Power Electron. **15**(4), 660–669 (2000)

Correction to: Integration of Renewable Energy in Markets: Analysis of Key European and American Electricity Markets

Hugo Algarvio, Fernando Lopes, and João Santana

Correction to:
Chapter "Integration of Renewable Energy in Markets: Analysis of Key European and American Electricity Markets" in: L. M. Camarinha-Matos et al. (Eds.): *Technological Innovation for Industry and Service Systems*, IFIP AICT 553, https://doi.org/10.1007/978-3-030-17771-3_28

The original version of this chapter starting on p. 321 was revised. The year mentioned in the funding information was incorrect. The original chapter has been corrected.

The updated version of this chapter can be found at
https://doi.org/10.1007/978-3-030-17771-3_28

Published by Springer Nature Switzerland AG 2019
L. M. Camarinha-Matos et al. (Eds.): DoCEIS 2019, IFIP AICT 553, p. C1, 2019.
https://doi.org/10.1007/978-3-030-17771-3_37

Author Index