



PESARO 2016

The Sixth International Conference on Performance, Safety and Robustness in
Complex Systems and Applications

ISBN: 978-1-61208-522-7

February 21 - 25, 2016

Lisbon, Portugal

PESARO 2016 Editors

Jaime Lloret Mauri, Polytechnic University of Valencia, Spain

Pascal Lorenz, University of Haute-Alsace, France

PESARO 2016

Forward

The Sixth International Conference on Performance, Safety and Robustness in Complex Systems and Applications (PESARO 2016), held between February 21-25, 2016 in Lisbon, Portugal, continued a series of events dedicated to fundamentals, techniques and experiments to specify, design, and deploy systems and applications under given constraints on performance, safety and robustness.

There is a relation between organizational, design and operational complexity of organization and systems and the degree of robustness and safety under given performance metrics. More complex systems and applications might not be necessarily more profitable, but are less robust. There are tradeoffs involved in designing and deploying distributed systems. Some designing technologies have a positive influence on safety and robustness, even operational performance is not optimized. Under constantly changing system infrastructure and user behaviors and needs, there is a challenge in designing complex systems and applications with a required level of performance, safety and robustness.

The conference had the following track:

- Performance, Safety and Robustness in Complex Systems and Applications

We take here the opportunity to warmly thank all the members of the PESARO 2016 technical program committee, as well as all the reviewers. The creation of such a high quality conference program would not have been possible without their involvement. We also kindly thank all the authors that dedicated much of their time and effort to contribute to PESARO 2016. We truly believe that, thanks to all these efforts, the final conference program consisted of top quality contributions.

Also, this event could not have been a reality without the support of many individuals, organizations and sponsors. We also gratefully thank the members of the PESARO 2016 organizing committee for their help in handling the logistics and for their work that made this professional meeting a success.

We hope PESARO 2016 was a successful international forum for the exchange of ideas and results between academia and industry and to promote further progress in the field of performance, safety and robustness in complex systems and applications. We also hope that Lisbon, Portugal provided a pleasant environment during the conference and everyone saved some time to enjoy the beauty of the city.

PESARO 2016 Advisory Committee

Piotr Zwierzykowski, Poznan University of Technology, Poland

Wolfgang Leister, Norsk Regnesentral (Norwegian Computing Center), Norway

Yulei Wu, University of Exeter, UK

Harold Liu, IBM Research, China

PESARO 2016

Committee

PESARO 2016 Advisory Committee

Piotr Zwierzykowski, Poznan University of Technology, Poland
Wolfgang Leister, Norsk Regnesentral (Norwegian Computing Center), Norway
Yulei Wu, University of Exeter, UK
Harold Liu, IBM Research, China

PESARO 2016 Technical Program Committee

Amr Aissani, USTHB - Algiers, Algeria
Morteza Biglari-Abhari, University of Auckland, New Zealand
Andrea Bondavalli, University of Florence, Italy
Andrea Ceccarelli, University of Florence, Italy
Salimur Choudhury, Queen's University - Kingston, Canada
Dieter Claeys, Ghent University, Belgium
Juan Antonio Cordero, École Polytechnique / INRIA, France
Raimundo J de A Macêdo, Federal University of Bahia (UFBA), Brazil
Mark Dillon, International Criminal Court, Netherlands
Nadir Farhi, IFSTTAR/COSYS/GRETTIA, France
Francesco Flammini, Ansaldo STS, Italy
John-Austen Francisco, Rutgers University - Piscataway, USA
Mina Giurguis, Texas State University - San Marcos, USA
Teresa Gomes, University of Coimbra, Portugal
Michael Hübner, Ruhr-University of Bochum, Germany
Sokratis K. Katsikas, University of Piraeus, Greece
Peter Kieseberg, SBA Research, Austria
Yih-Jiun Lee, Chinese Culture University, Taiwan
Wolfgang Leister, Norsk Regnesentral (Norwegian Computing Center), Norway
Harold Liu, IBM Research, China
Giovanni Livraga, Università degli Studi di Milano, Italy
Olaf Maennel, Tallinn University of Technology, Estonia
Stefano Marrone, Seconda Università di Napoli, Italy
Birgit Milius, Institut fuer Eisenbahnwesen und Verkehrssicherung (IfEV) /Technische
Universitaet – Braunschweig, Germany
Haralambos Mouratidis, University of Brighton, UK
Asoke Nandi, Brunel University, UK
Mohammad Rajabali Nejad, University of Twente, Netherlands

Maciej Piechowiak, Kazimierz Wielki University - Bydgoszcz, Poland
Francesco Quaglia, DIAG - Sapienza Universita' di Roma, Italy
M. Zubair Rafique, KU Leuven, Belgium
Roger S. Rivett, Land Rover - Gaydon, UK
Luis Enrique Sánchez Crespo, University of Castilla-la Mancha, Spain / University of the Armed Forces, Ecuador
Jean-Pierre Seifert, Technische Universität Berlin & Telekom Innovation Laboratories, Germany
Dhananjay Singh, Electronics and Telecommunications Research Institute (ETRI), South Korea
Mukesh Singhal, University of California, Merced, USA
Young-Joo Suh, Pohang University of Science and Technology (POSTECH), South Korea
Yuejin Tan, National University of Defense and Technology - Changsha, China
Yang Wang, Georgia State University, USA
Yun Wang, Bradley University - Peoria, USA
Hironori Washizaki, Waseda University, Japan
Jun Wu, National University of Defense and Technology, China
Yanwei Wu, Western Oregon University, USA
Yulei Wu, University of Exeter, UK
Gerhard Wunder, Fraunhofer Heinrich Hertz Institut - Technical University Berlin, Germany
Gaoxi Xiao, Nanyang Technological University, Singapore
Bin Xie, InfoBeyond Technology LLC - Louisville, USA
Nabila Zbiri, Université d'Evry Val d'Essonne, France
Xiaoyong Zhou, Indiana University, USA
Yanmin Zhu, Shanghai Jiao Tong University, China
Piotr Zwierzykowski, Poznan University of Technology, Poland

Copyright Information

For your reference, this is the text governing the copyright release for material published by IARIA.

The copyright release is a transfer of publication rights, which allows IARIA and its partners to drive the dissemination of the published material. This allows IARIA to give articles increased visibility via distribution, inclusion in libraries, and arrangements for submission to indexes.

I, the undersigned, declare that the article is original, and that I represent the authors of this article in the copyright release matters. If this work has been done as work-for-hire, I have obtained all necessary clearances to execute a copyright release. I hereby irrevocably transfer exclusive copyright for this material to IARIA. I give IARIA permission to reproduce the work in any media format such as, but not limited to, print, digital, or electronic. I give IARIA permission to distribute the materials without restriction to any institutions or individuals. I give IARIA permission to submit the work for inclusion in article repositories as IARIA sees fit.

I, the undersigned, declare that to the best of my knowledge, the article does not contain libelous or otherwise unlawful contents or invading the right of privacy or infringing on a proprietary right.

Following the copyright release, any circulated version of the article must bear the copyright notice and any header and footer information that IARIA applies to the published article.

IARIA grants royalty-free permission to the authors to disseminate the work, under the above provisions, for any academic, commercial, or industrial use. IARIA grants royalty-free permission to any individuals or institutions to make the article available electronically, online, or in print.

IARIA acknowledges that rights to any algorithm, process, procedure, apparatus, or articles of manufacture remain with the authors and their employers.

I, the undersigned, understand that IARIA will not be liable, in contract, tort (including, without limitation, negligence), pre-contract or other representations (other than fraudulent misrepresentations) or otherwise in connection with the publication of my work.

Exception to the above is made for work-for-hire performed while employed by the government. In that case, copyright to the material remains with the said government. The rightful owners (authors and government entity) grant unlimited and unrestricted permission to IARIA, IARIA's contractors, and IARIA's partners to further distribute the work.

Table of Contents

Evaluation of Selecting Cloud Services Approach for Data Storage using Secret Sharing Scheme <i>Shohei Ueno, Atsushi Kanai, Shigeaki Tanimoto, and Hiroyuki Sato</i>	1
Coping with System Hazards in Early Project Life Cycle; Identification and Prioritization <i>Mohammad Rajabalinejad</i>	7
Using XPath to Define Design Metrics <i>Jan Soderberg, Ali Shahrokni, and Bashar Nassar</i>	13

Evaluation of Selecting Cloud Services Approach for Data Storage using Secret Sharing Scheme

Shohei Ueno
Hosei University
Tokyo, Japan
shohei.ueno.4h@stu.hosei.ac.jp

Atsushi Kanai
Hosei University
Tokyo, Japan
yoikana@hosei.ac.jp

Shigeaki Tanimoto
Chiba Institute of Technology
Chiba, Japan
shigeaki.tanimoto@it-chiba.ac.jp

Hiroyuki Sato
The University of Tokyo
Tokyo, Japan
schuko@satolab.itc.u-tokyo.ac.jp

Abstract—Cloud services have become more popular because of their decreasing cost. However, it is difficult to select the optimal cloud service because there are many services whose service levels are different. We evaluate our proposed method for dynamically selecting the optimal cloud services to store data in a heterogeneous multi-cloud environment. The evaluation used the SLAs of actual cloud services and the results indicate it is possible to select a combination of cloud services.

Keywords—cloud computing; multi-cloud; hybrid cloud; secret sharing scheme; availability; confidentiality

I. INTRODUCTION

Cloud computing has recently become popular. Methods involving a combination of multiple cloud services have been proposed, [2]-[4], which provide users with more advantages (availability or confidentiality) than usage of single clouds.

These methods need to select the best combinations of cloud services. As there are many different types of cloud services with various service levels, a wide variety of service levels can be constructed in heterogeneous multi-cloud environments. Especially, multiple services are used at the same time.

We first describe the proposed method. Then, we are quantifying the evaluation using the developed prototype. Furthermore, we present a concrete case using actual cloud services.

The rest of this paper is organized as follows. We present the related work in Section 2. In Section 3, we describe the assumed environment and the proposed method. In Section 4, we show the overview of the evaluation system. In Section 5, we describe the evaluation using actual public cloud services implementing the prototype of this proposed method, and evaluate the communication speed. Finally, we conclude the paper in Section 6.

II. RELATED WORK

Approaches which use multiple cloud services have been proposed to improve availability and confidentiality, cost, performance, etc., when compared with single cloud services. For example, DepSky [3] improved the availability, integrity, and confidentiality of data stored in clouds. The high-availability and integrity layer (HAIL) [6], which accepts a set of servers to prove to clients that stored files are complete and recoverable, was developed on links between multiple cloud services.

Files that users want to manage in cloud storage have properties of various degrees of confidentiality and availability. Therefore, it is necessary to change the requirements per file. This means one has to reselect the best combination of cloud per file. Cardellini et al. demonstrated how to select the best services [10] in relation to the cost-effective use of such services. Tsai et al. proposed a cost-effective intelligent configuration model [11]. In addition, a file-distribution method using a secret sharing scheme was proposed and evaluated in a homogeneous multi-cloud environment [12]. A data management method in this environment was also proposed [13].

There are also security concerns about public clouds. Cloud security in terms of data management has also been discussed [15]-[18]. To solve one of these issues, a method in which a system automatically selects appropriate cloud services using a service-level agreement (SLA) written in extensible markup language (XML) has been proposed [19].

Currently, there is no way to select and evaluate optimal cloud services from many different clouds (heterogeneous multi-clouds) in using multiple clouds at the same time in the proposed environment [12][13].

III. PROPOSED METHOD

A. Assumed Environment

We assumed a multi-cloud environment with many cloud-storage services in a secret sharing scheme, and all of these services had machine readable SLAs written in XML [12][13][22]. In this section, we introduce the proposed method [22] which is evaluated.

Figures 1 and 2 outline the proposed method. A user selects a set of cloud services using their SLAs depending on the required availability, confidentiality, and cost. Then, all combinations of cloud services are calculated by the user requirements, and it is determined to store in cloud services. When a file is stored in cloud services, it is distributed using a (k, L, n) secret sharing scheme [19]. However, the user requirement is different per file. The best combination of cloud services is selected by calculating when a user stores the secret information.

B. (k, L, n) secret sharing scheme

The (k, L, n) secret sharing scheme was devised by Yamamoto [20] and is an extension of the (k, n) secret sharing scheme presented by Shamir [21]. It can reduce the amount of distributed information compared to the (k, n) secret sharing scheme.

By applying the (k, L, n) secret sharing scheme to secret information x , n pieces of distribution information are obtained. The restoration of the information is performed by collecting k pieces. Additionally, the data size of the distributed information becomes $1/L$ times that of the secret information. It is possible to identify part of the secret information from many k - L s that are less than the k s of distributed information. Fewer k - L s provide safety with regard to information theory, so it is not possible to obtain any secret information.

C. Matching user requests with cloud service levels

In the proposed method [22], user requirements are defined using four indicators.

- 1) Cost: Required cost per amount data (to store 1 MB [yen/MB])
- 2) Confidentiality: Risk of secret data being identified from data stored to cloud services
- 3) Availability: Total operating rate [%] of multi-cloud
- 4) Transfer time: Upload time and download time [s/MB]

We assumed these indicators are written in SLA of cloud services. Therefore, we calculate and select the best combinations of cloud services using the user requests.

D. Formulas that correspond to user requests

In the proposed method, the best combination of cloud services is selected by calculating [22].

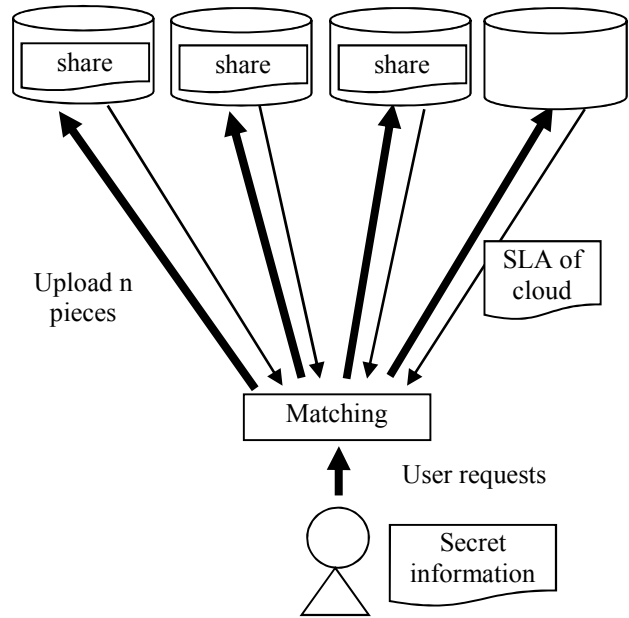


Figure 1. The image of uploading

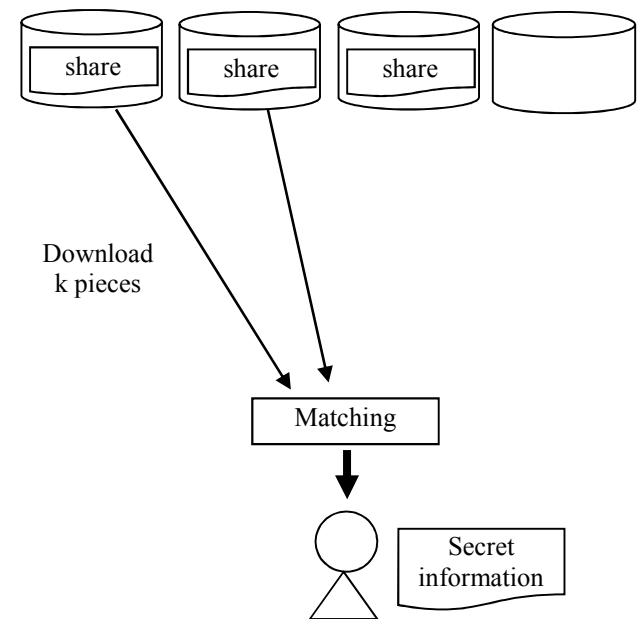


Figure 2. The image of downloading

a) Uploading

As it is necessary for users and cloud services to communicate during uploads, availability, transfer time, and costs are important as metrics.

a. Cost

Cost is the total expense of all cloud services and is expressed as

$$\text{Cost} = \frac{1}{L} \sum_{i \in n} \text{Cost of Cloud}_i. \quad (1)$$

b. Availability

Because it must be able to communicate with all cloud services to store shared information, availability becomes:

$$\text{Availability} = \prod_{i \in n} \text{Operating rate of Cloud}_i. \quad (2)$$

c. Transfer time

Transfer time is the total upload time to reach each service, and it becomes:

$$\text{Transfer time} = \frac{1}{L} \sum_{i \in n} \frac{1}{\text{Communication speed of Cloud}_i}. \quad (3)$$

b) Storing

As it is not necessary to communicate with cloud services, confidentiality of the secret data is very important.

a. Confidentiality

Confidentiality is related to the probability of information leakage from each cloud and the total disclosure level of the information. Here, $0 \leq x \leq n$.

Confidentiality =

$$\sum_{x=k}^n \left(\sum_{i \in \{y | y \in P(n), |y|=x\}} \prod LP(i) \prod_{j \in n-y} \{1 - LP(j)\} \right) * \text{Specific Level}(x). \quad (4)$$

where $P(n)$ is the power set of n , and $LP(i)$ is the leakage probability of cloud i .

The disclosure level is represented by the following formula depending on the parameters of the (k, L, n) secret sharing scheme.

$$\text{Specific Level}(x) = \begin{cases} 0 & (x \leq k - L) \\ 1 - \frac{k-x}{L} & (k - L < x < k) \\ 1 & (k \leq x) \end{cases} \quad (5)$$

c) Downloading

As it is necessary to communicate with clouds, the availability and transfer time is important.

a. Availability

The availability in a cloud service to upload distribution information is the probability that users can communicate

with all the cloud services necessary to restore the shared data in all services that have stored shared data.

The $A(i)$ in this equation is the operation ratio of cloud i .

$$\text{Availability} = \sum_{x=k}^n \left(\sum_{i \in \{y | y \in P(n), |y|=x\}} \prod A(i) \prod_{j \in n-y} \{1 - A(j)\} \right). \quad (6)$$

b. Transfer time

Transfer time is the time to communicate with the cloud and restore information.

$$\text{Transfer time} = \frac{k}{Ln} \sum_{i \in n} \frac{1}{\text{Communication Speed of Cloud}_i}. \quad (7)$$

E. Relationship between the indicators and (k, L, n) secret sharing scheme

In the proposed method [22], the combination of cloud services is calculated by the user requirements, and secret information is uploaded for the selected cloud services using (k, L, n) secret sharing scheme. Table I summarizes the relationships between the indicators and the actions.

The availability in uploading is worse when the value of n is increasing. The total operating rate is worse because of increasing the number of distributions. The cost in uploading is better when the value of L is increasing, but it is worse when the value of n is increasing. The smaller size of data can be stored in cloud services inexpensively, but the total cost is increasing because of increasing the number of distributions. The transfer time in uploading is better when the value of L is increasing, but it is worse when the value of n is increasing. The smaller size of data can be stored in cloud services quickly, but the total transfer time is increasing because of increasing the number of distributions.

The confidentiality in storing is better when the value of k is increasing, but it is worse when the value of L and n are increasing because of equation (5).

The availability in downloading is better when the value

TABLE I. RELATIONSHIP BETWEEN PARAMETERS AND ACTIONS

	Uploading			Storing	Downloading	
	Availability	Cost	Transfer time	Confidentiality	Availability	Transfer time
k	-	-	-	Better	Worse	Worse
L	-	Better	Better	Worse	-	Better
n	Worse	Worse	Worse	Worse	Better	-

of n is increasing, but it is worse when the value of k is increasing. It is necessary to collect k pieces of distribution

information for restoring the secret information. The transfer time in downloading is better when the value of L is increasing, but it is worse when the value of k is increasing. The smaller size of data can be downloaded quickly, but the total transfer time is increasing because of increasing the number of distributions for restoring.

IV. OVERVIEW OF THE EVALUATION SYSTEM

In this section, we explain the evaluation of a method proposed in a previous study [22] using some of the metrics in a heterogeneous cloud environment. In the previous study, we assumed the value of SLA for private and public cloud services, and selected some combinations using the proposed method.

For the result, some combinations were calculated for some situations; highest availability, lowest cost or highest confidentiality.

In the current study, we investigated some actual SLAs of public cloud services. Specifically, we investigated the SLAs of Google Drive, CloudN, KDDI, BOX, Dropbox, and One Drive. Table II lists the SLA metrics for these cloud services. However, these name of cloud services were expressed from P0 to P5 in Table II for consideration to the cloud services. In Table II, all cloud services did not provide leakage probability and communication speed. Therefore, we assumed the value of leakage probability based on the description of confidentiality. We decided whether the acquisition of security standards and the policy of security are written in SLA of each cloud services or not.

For communication speed, we did not evaluate the communication speed because we cannot estimate the value. However, it is necessary to evaluate the communication speed. Then we developed a prototype in this proposed method and evaluated the communication speed between the cloud services and user. Here, we use five cloud services: Box, Dropbox, Google Drive, and One Drive. For the implementation, we use these cloud services, which provide API. The results will be described later.

Then, Figure 3 shows the image of the evaluation model. We selected the combination of cloud services, and

determined the parameter of a (k, L, n) secret sharing scheme using the proposed equation. In addition, we developed a prototype for uploading and downloading the distributed data using that parameter.

V. ACTUAL CLOUD EVALUATION

A. Actual SLAs description and setting

Table III lists the metrics of private and public cloud services that satisfy the actual SLAs. However, we assumed the same value as that of the private clouds in Table III because the actual value of private cloud is not written. Additionally, P0 is getting the ISO 27001[23] and written the policy of security in SLA, we assumed it is the better value of leakage probability than other public cloud services. On the other hand, P3 and P5 are written nothing about security. Then we assumed these are the worse value of leakage probability than other public cloud services.

Here, cost is defined as [yen/(month · GB)], and the user has already contracted for all the public cloud services.

$$\text{Cost} = \sum_{i \in n} \text{Cost of Cloud}_i. \tag{8}$$

Therefore, the costs of all combinations are fixed.

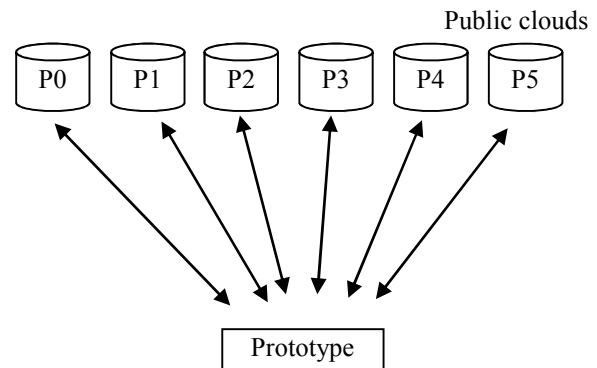


TABLE II. DESCRIPTION OF ACTUAL CLOUD SERVICES

Cloud	Operating rate	Cost	Leakage probability	Communication speed
P0	Written	Written	Not Written	Not Written
P1	Written	Written	Not Written	Not Written
P2	Written	Written	Not Written	Not Written
P3	Written	Written	Not Written	Not Written
P4	Written	Written	Not Written	Not Written
P5	Not Written	Written	Not Written	Not Written

TABLE III. PARAMETER SETTING FOR EVALUATION

Cloud	Operating rate	Cost	Leakage probability
Private Cloud	0.999	-	0.001
P0	0.999	16.6	0.01
P1	0.9999	8.6	0.1
P2	0.9999	30	0.1
P3	0.999	6.0	0.5
P4	0.999	0.54	0.1
P5	0.9999	0	0.5

All cloud services have sufficient communication speed; therefore, we did not evaluate transfer time. The L of all combinations was only one. Then, we calculated all combinations of cloud services, and Table IV lists the three unique combinations. As a result, the parameter of (k, L, n) secret sharing scheme is (k = 2, L = 1, n = 4) or (k = 4, L = 1, n = 4)

Combinations (p0, p1, p2, p4 in k = 2) and (p1, p2, p3, p4) have the best availability; Combination (p0, p1, p2, p4 in k = 2) has a lower cost than Combination (p1, p2, p3, p4). Combination (p0, p1, p2, p4 in k = 4) has the highest confidentiality, which is better than that of private clouds. However, the availability of Combination (p0, p1, p2, p4 in k=4) is the worst. This is caused by parameter k, that made availability in downloading worse.

B. Evaluation of communication speed

Table V lists all the combinations of these cloud services. Here, the value of n in (k, L, n) secret sharing scheme is four. Combinations (p0, p3, p5), (p3, p4, p5), and (p0, p4, p5) have better availability than Combination (p0, p3, p4).

Then, we measured the communication speed for each cloud services (Table VI) and all the combinations (Table VII). Here, the data size is 10 [MB], upload time is the average of 10 measurements, and download time is the average of 3 measurements. In addition, the download time is measured for all combinations.

In Table VII, Combination (p0, p3, p4) has the best

TABLE VII. COMMUNICATION SPEED FOR ALL COMBINATIONS

Combination	Upload time [ms]	Download time [ms]	Download Clouds
p0,p3,p5	10070	7499	p0,p5
		12856	p3,p5
		8986	p0,p3
p3,p4,p5	10829	10182	p4,p5
		8652	p3,p4
		12094	p3,p5
p0,p3,p4	6820	5445	p0,p4
		11487	p3,p4
		8594	p0,p3
p0,p4,p5	10616	8372	p4,p5
		4390	p0,p4
		8601	p0,p5

upload time, and the combination of cloud p0 and p4 has the best download time. Therefore, depending on the combination of clouds chosen, it is possible to have a better communication speed than using only one cloud service. However, all of the combinations are worse upload time than only each cloud, and cloud p3 is also worse download time. Thus, the communication speed of some combinations is worse than using each cloud service.

Additionally, Combination (p0, p3, p4) has the worst availability in Table V. Combination (p0, p4, p5) does not have a good upload time but has a good average download time compared to other combinations. We need the communication speed of the SLA to evaluate cloud services not only operating rate, and find the best cloud services.

However, this evaluation is one example. In the actual situation, the best combinations can be selected by the calculation taking into consideration the user requirements in this proposed method.

VI. CONCLUSION

We evaluated a method using multiple cloud storage services in a heterogeneous cloud environment by using concrete values of three metrics. We found that some combinations of cloud services were more useful compared to only one private cloud service. All combinations had both advantages and disadvantages. Therefore, we found that the communication speed is necessary for the new evaluation value. However, we only implemented the prototype. In the future, we need to implement the actual system.

ACKNOWLEDGEMENT

This work was supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI Grant Number 15H02783.

TABLE IV. THREE BEST COMBINATIONS EXTRACTED FROM RESULTS

Combination	k	L	n	Leakage probability	Availability when downloading
p0,p1,p2,p4	2	1	4	304.3	0.9999999978
p1,p2,p3,p4	2	1	4	1495	0.9999999978
p0,p1,p2,p4	4	1	4	0.1	0.99780141

TABLE V. ESTIMATED VALUE

Combination	k	L	n	Leakage probability	Availability in downloading
p0,p3,p5	2	1	3	0.255	0.9999988
p3,p4,p5	2	1	3	0.3	0.9999988
p0,p3,p4	2	1	3	0.055	0.99997002
p0,p4,p5	2	1	3	0.055	0.9999988

TABLE VI. COMMUNICATION SPEED BETWEEN CLOUD SERVICES AND USERS FOR SINGLE SERVICE

Cloud	Upload time [ms]	Download time [ms]
P0	4118	6858
P3	3680	2452
P4	4744	4642
P5	8815	7451

REFERENCES

- [1] NIST [Online] Available from: <http://www.nist.gov/itl/cloud/2016.01.11>
- [2] M. Vukolic, "The Byzantine empire in the intercloud," *ACM SIGACT News*, 41, 2010, pp. 105–111.
- [3] A. Bessani, M. Correia, B. Quaresma, F. André, and P. Sousa, "DEPSKY: dependable and secure storage in a cloud-of-clouds," *EuroSys'11: Proc. of 6th Conf. on Computer Systems*, 2011, pp. 31–46.
- [4] H. Abu-Libdeh, L. Princehouse, and H. Weatherspoon, "RACS: A case for cloud storage diversity," *SoCC'10: Proc. of 1st ACM Symposium on Cloud Computing*, 2010, pp. 229–240.
- [5] K. D. Bowers, A. Juels, and A. Oprea, "HAIL: A high-availability and integrity layer for cloud storage," *CCS'09: Proc. of 16th ACM Conf. on Computer and Communications Security*, 2009, pp. 187–198.
- [6] H. Abu-Libdeh, L. Princehouse, and H. Weatherspoon, "RACS: A case for cloud storage diversity," *SoCC'10: Proc. of 1st ACM Symposium on Cloud Computing*, 2010, pp. 229–240.
- [7] NRI Secure Technologies. [Online]. Available from: <http://www.nri-secure.co.jp/service/global/gss.html> 2016.01.11
- [8] T. Matsumoto, T. Seito, A. Kamoshita, T. Shingai, and A. Sato, "High-Speed Secret Sharing System for Secure Data Storage Service," *SCIS2012. The 29th Symposium on Cryptography and Information Security*, 2012.
- [9] V. Cardellini, V. Valerio, V. Grassi, S. Iannucci, and F. Presti, "A New Approach to QoS Driven Service Selection in Service Oriented Architectures," *SOSE*, 2011, pp. 102–113.
- [10] W. Tsai, G. Qi, and Y. Chen, "A Cost-Effective Intelligent Configuration Model in Cloud Computing," *ICDCSW*, 2012, pp. 400–408
- [11] Y. Kajiura, A. Kanai, S. Tanimoto, and H. Sato, "A File-distribution Approach to Achieve High Availability and Confidentiality for Data Storage on Multi-cloud," *SAPSE2013*, 2013.
- [12] A. Kanai, S. Tanimoto, and H. Sato, "Data Management Approach for Multiple Clouds Using Secret Sharing Scheme," *8th International Workshop on Advanced Distributed and Parallel Network Applications (ADPNA-2014)*, 2014, pp. 432–437.
- [13] Cloud Security Alliance, "Cloud Control Matrix Version 3.0", 2013.
- [14] H. Sato, A. Kanai, and S. Tanimoto, "A Cloud Trust Model in a Security Aware Cloud," *Proc. of IEEE/IPSJ International Symposium on Applications and the Internet (SAINT2010)*, 2010, pp. 121–124.
- [15] S. Tanimoto, S. Matsui, H. Sato, A. Kanai, and et al, "A Study of Risk Management in Hybrid Cloud Configuration," *Computer Information Science*, Springer, vol. 493, 2013, pp. 247–257.
- [16] S. Tanimoto, H. Sato, A. Kanai, and et al, "A Study of Data Management in Hybrid Cloud Configuration," *14th IEEE/ACIS, SNPD2013*, 2013, pp. 381–386.
- [17] H. Sato, A. Kanai, and S. Tanimoto, "Building a Security Aware Cloud by Extending Internal Control to Cloud," *Proc. of 10th Int'l Symposium on Autonomous Decentralized Systems (ISADS 2011)*, 2011, pp. 323–326..
- [18] H. Sato, S. Tanimoto, and A. Kanai, "A Policy Consumption Architecture that Enables Dynamic and Fine Policy Management," *Proc. of 3rd ASE International Conf. on Cybersecurity*, 2014, pp. 1–11.
- [19] H. Yamamoto, "Secret Sharing System Using (k, L, n) Threshold Scheme," *Electron. Commun. Jpn. (Part I: Commun.)*, 1986, vol. 69, no. 9, pp. 46–54.
- [20] A. Shamir, "How to share a secret," *Communications of the ACM*, 22(11), 1979, pp. 612–613.
- [22] Y. Kajiura, S. Ueno, A. Kanai, S. Tanimoto, and H. Sato, "Approach to Selecting Cloud Services for Data Storage in Heterogeneous Multi-cloud Environment with High Availability and Confidentiality," *The First International Workshop on Service Assurance in System Wide Information Management (SASWIM2015)*, 2015, pp. 205–210.
- [23] ISO/IEC 27001:2005, "Information technology – Security techniques – Information security management systems – Requirements," 2013

Coping with System Hazards in Early Project Life Cycle

Identification and Prioritization

Mohammad Rajabalinejad

Faculty of Engineering Technology, University of Twente, Enschede, the Netherlands
M.Rajabalinejad@utwente.nl

Abstract--The rising complexity of product and systems demands further attention to potential hazards. While researchers explore tools and methods to identify hazards, their prioritization remains a challenging task in a multi-stakeholder environment. A reason for this is that the hazards are hardly quantifiable. While the accurate quantification remains a challenge, a flexible and pluralistic approach can bring the important ones on top of the list. This paper offers a methodology for ranking hazards in early phases of design with presence of a high level of uncertainty. It uses a pluralistic approach for prioritization of hazards. It adapts probability theory to embed flexibly in communication with stakeholders and process the available information. A graphical tool facilitates this communication and probabilistically utilize available information about system hazards. It introduces the “degree of consensus” as a metric to rank the identified hazards. This metric represents the consent of stakeholders on the system of interest (SoI) concerns used for example in its architecture, design decisions, or alternative evaluation. The paper explains the mathematical formulation and presents an application example for this.

consensus; hazards; uncertainty; prioritization; ranking.

I. INTRODUCTION

A. Hazard Identification

Hazards are the risk sources, and their proper recognition and prioritization leads to a better understanding of risk and their management. The rising complexity and cross-disciplinary nature of systems demands further development for identification of hazards [1]. Hazard is the potential source of harm [2], and this creates a direct link between hazard and risk. If a hazard is not identified, risks remain unattended.

The European norm on risk assessment [3] summaries the tools and methods applicable to hazard identification in categories of strongly applicable and applicable. The strongly applicable methods for risk identifications are brainstorming, Delphi, Check-lists, Primary hazard analysis, Hazard and operability studies (HAZOP), Environmental risk assessment, SWIFT, Scenario analysis (SA), Failure mode and effect analysis (FMEA), Cause-and-effect analysis, Human reliability analysis (HRA), Reliability centered maintenance (RCM), Consequence/probability matrix. The applicable methods for hazard identifications

are Business impact analysis (BIA), Fault tree analysis (FTA), Event tree analysis (ETA), Cause and consequence analysis (CCA), Layer protection analysis, Sneak circuit analysis, Markov analysis, FN curves, Risk indices, cost/benefit analysis, and Multi-criteria decision analysis.

B. Early life-cycle

Designers can effectively impact a system in early design phases. In this phase, changes are often less costly and design decisions can profoundly influence the system of interest. In early design phases, proper information reduces uncertainties, increases utilities, and creates value for the system as shown in Figure 1. This is because proper information for a designer leads to better design choices that ultimately influence the rest of design including concept, detail, services, and etcetera.

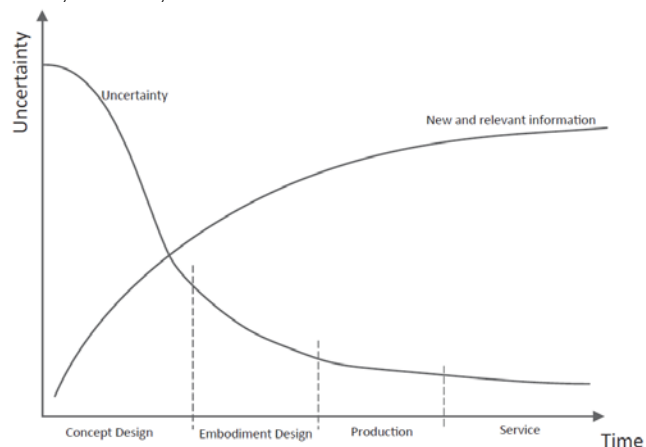


Figure 1. The information concern in the design process [4].

Yet information in the beginning of design can also be overwhelming. A design team may be exposed to a lot of information that hinders focusing on the key aspects of design. In system design with the multi-stakeholder nature of systems, divergent expectations of stakeholders can prevent a designer to focus on the key drivers for a system design.

In an interdisciplinary system, there are a lot of mono- or multi-disciplinary hazards that are hard to quantify or prioritize. Quantification of hazards in the form of frequency or severity comes after its realization. Furthermore, this quantification may be subject to change over time.

Lack of proper hazard identification or prioritization leads to rising complexity in the risk analysis and management. Most of the currently applied hazard identification methods result in a hazard pool. In such a view, a larger system results a larger hazard pool which makes the prioritization more complex. The next section discusses this in further details.

C. Hazards, risks and requirements

A good understanding of hazards and risks helps to develop a proper list of requirements. The importance of requirements have been discussed in design literatures, see e.g. [5-7]. This study adapts a pluralistic approach for highlighting system hazards, risks or requirements.

Literatures have discussed that many engineering design methods pay attention to system risks when there is already a concept for the system. Yet proper view of main hazards helps forming an architecture that fits better to them [8, 9]. Recognition of system hazards is indeed a pluralistic approach, and the design team/ architecture need to approach different system stakeholders and explore their concerns about the system risks and hazards. Stakeholder in this paper is used as a general term that includes system shareholders, users, designers, experts and etcetera, and the concern refers to a stakeholder concern including the specific hazard.

Literatures confirm that an incomplete set of stakeholders may lead to incomplete results since there are problems arising from the scope, understanding and validation of needs, concerns or concern [10, 11] in the course of communication with stakeholders. Therefore, identification of stakeholders and elicitation of information are considered as prerequisites for understanding the system hazards. Systems often involves a large number of stakeholders [12]. Figure 2 presents the functional diagram for identifying stakeholders and communicating with them. This results in a pool of concerns with a lot of information. Ranking of this information helps the designer to keep her focus on the key aspects. Recognition of key hazards is likely to be seen subjectively as different stakeholders tend

to focus on their areas of interest and pay more attention to the hazards that influence their interest.

This study assumes that key hazards are recognized by the stakeholders and that those key hazards can be determined through a pluralistic approach. It therefore focuses to offer a pluralistic approach that communicates well with stakeholders, provides freedom for presenting the opinions, and embraces doubts or uncertainties in their information.

D. System hazards

This study builds on the assumption that key hazards in design are recognized by the consensus of stakeholders, and they can be rated systematically through a ranking process. In general, ranking of parameters (hazards) based on their importance is well discussed in decision models. The use of multi criteria decision models typically involves a systematic ranking process as for instance indicated in [13, 14]. The influence of the ranking process on final decisions is for example explained in [15]. A review of subjective ranking methods shows that different methods cannot guarantee accurate results. This inconsistency in judgment explains difficulty of assigning reliable and subjective weights to the requirements. A systematic approach for ranking is described in [16] that is a generalization of Saaty’s pairwise structure [17]. Given the presence of subjectivity in the ranking process, sensitivity analysis of the design criteria is used to study the influence of variation and the ranking process on the decisions made [18]. Furthermore, some approaches e.g. the task-oriented weighing approach is effectively used. This approach is meant to limit the subjectivity of criteria weighting [19]. It suggests an algorithm to rank criteria objectively while considering the uncertainty in criteria weight [20]. The approach is based on introducing fuzzy numbers that imposes specified membership functions, which has been also used in [21, 22].

The methods used to identify the system hazards are mentioned earlier in this paper. The outlines of these methods are available elsewhere in for example [23]. The

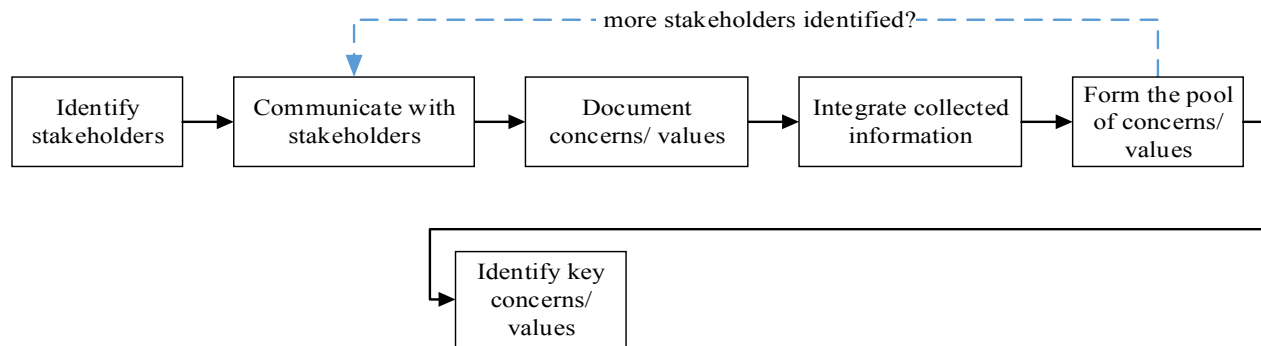


Figure 2. The process of identification of stakeholders and communication with them.

use of these methods results in a bank of information called a “pool of hazards”.

E. Pool of hazards

The so called pool of hazards integrates the identified hazards that threaten the system. This pool includes all the system hazards recognized by stakeholders. As the pool can become of enormous size, a method is required for listing them based on their priorities. Figure 3 schematically shows a set of hazards recognized for a system.

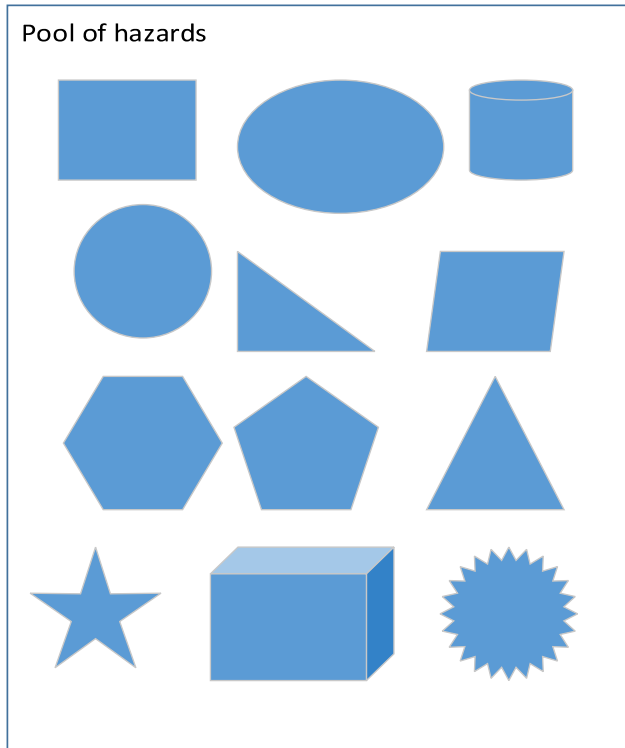


Figure 3. A schematic view for the pool of hazards.

II. COMMUNICATON OF HAZARDS

There are obstacles in communication with system stakeholders who can be individuals, corporations, organizations and authorities, with different fields/ levels of knowledge and experience [4]. They all have their interests and expectations. This study uses uncertainty to allow a human solution in terms of preferred alternatives [24, 25]. The uncertainty in importance of design concerns is also of human nature which should be reflected in the process [26]. The principle of the method is described elsewhere in [7] and discussed in further details through this next section.

A. Presentation

The method aims at a realistic and intuitive approach that can communicate to stakeholders with different fields of knowledge and expertise. The method must be transparent, easy to implement and readily adaptable by different users. For this purpose, graphs are used to effectively communicate with different users. The format presented in Figure 4 is used to identify and register the importance of a concern

according to a stakeholder. It shows that the linguistic scale may replace the numeric scale for the ease of communication, and one can assign a range of possible importance to a certain concern. For illustration, Figure 4(b) shows that the i-th concern, C_i , may have the importance somewhere from 0.6 to 0.8 according to one of the stakeholders in 0 to 1 grading scale, where 0 indicates no importance at all and 1 represents the absolute importance. Then, probability distribution function (PDF) is assigned to this recorded data. Symmetric opinions are assumed here in this paper as described in [27, 28] and the collected data is

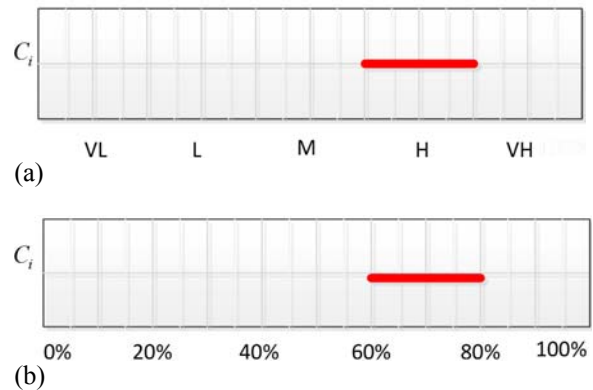


Figure 4. An example of a stakeholder’s opinion about the importance of the i-th concern C_i , treated as a random variable with a Gaussian distribution.

B. Formulation

Having m stakeholders, their opinions for the i-th design concern C_i is presented by stochastic variables $c_{i_1}, c_{i_2}, \dots, c_{i_m}$, where v_k presents the k-th stakeholder’s opinion over the importance of the i-th concern. The mean and standard deviation of these variables are respectively shown as $\mu_{i_1}, \mu_{i_2}, \dots, \mu_{i_m}$ and $\sigma_{i_1}, \sigma_{i_2}, \dots, \sigma_{i_m}$. As a result, the overall mean and standard deviation of opinions over the i-th concern are formulated by Equations (1) and (2), respectively.

$$\mu_i = \frac{1}{\sum_{k=1}^m \alpha_k} \sum_{k=1}^m \alpha_k \mu_{i_k} \tag{1}$$

$$\sigma_i^2 = \frac{\sum_{j=k}^m \alpha_k^2 \sigma_{i_k}^2}{\left(\sum_{k=1}^m \alpha_k\right)^2} \tag{2}$$

Where α_k represents the assigned weight to the k-th stakeholder. If the stakeholders are evenly graded (which is not very likely in the context of complex systems), Equations (1) and (2) transform to the following.

$$\mu_i = \frac{1}{m} \sum_{k=1}^m \mu_{i_k} \tag{3}$$

$$\sigma_i^2 = \frac{\sum_{k=1}^m \sigma_{i_k}^2}{m^2} \tag{4}$$

After normalization, the following equations are concluded.

$$\lambda_i = \frac{\mu_i}{\sum_{i=1}^n \mu_i} \tag{5}$$

$$\sigma_{\lambda_i}^2 = \left[\frac{\sigma_i}{\sum_{i=1}^n \mu_i} \right]^2 \tag{6}$$

$$\phi_i^2 = \frac{\sigma_i^2}{\sum_{i=1}^n \sigma_i^2} \tag{7}$$

Where λ_i , σ_{λ_i} and ϕ_i are respectively the weight factor, its standard deviation and the relative uncertainty for the i-th concern. Relative weight λ_i is often used as the criteria for ranking parameters or concerns. Under uncertain situation, however, λ_i is not the only parameter to rank data, and its uncertainty σ_{λ_i} can play an important role in the ranking process. High uncertainty can lead to high risk, and one may prefer a concern with more certainty but lower λ_i . On the basis of discussion above, we use “the reliability index” as an estimated measure of reliability of each concern. Therefore, the reliability index of each concern is estimated as

$$\beta_i = \frac{\lambda_i}{\sigma_{\lambda_i}} \tag{8}$$

The equation above indicates the relative standard error (RSE) for the importance of i-th estimated concern, which also can be referred to as reliability of the i-th concern [29]. It represents the degree of stakeholders’ consensus on the i-

th concern. The algorithm for applying this method is described next and an example application of it is presented in the next section.

C. Algorithm

Here, we describe the steps needed for ranking the requirements. A summary of this process is shown in Figure 5.

- List m stakeholders and n concerns for SoI. Determine the weight of stakeholders’ opinions if they are not evenly graded.
- Draw tables and list concerns (C_1, C_2, \dots, C_n) using the numeric or verbal format shown in Figure 4.
- Ask the stakeholders to fill the tables. This step concludes m series of tables. Use C_{i_k} format to label the collected information for each table, where k is the number of stakeholders.
- Calculate the expected concern and standard deviation (μ_{i_k} and σ_{i_k}) for each C_{i_k} .
- Calculate the mean and standard deviation for each concern (μ_i and σ_i) for the i-th concern. Use Equations (1) and (2). If the stakeholders are evenly graded, use Equations (3) and (4).
- Use Equations (5) to (7) to calculate the normalized weight of each concern, its standard deviation, and relative uncertainty.
- If new stakeholders or concerns are realized, reiterate from the first step. Otherwise use Equation (8) to calculate the degree of consensus on each concern and rank the concerns.

This process uses the collected information and sorts the system concerns based on the stakeholders’ opinion. The next section presents an example application for this.

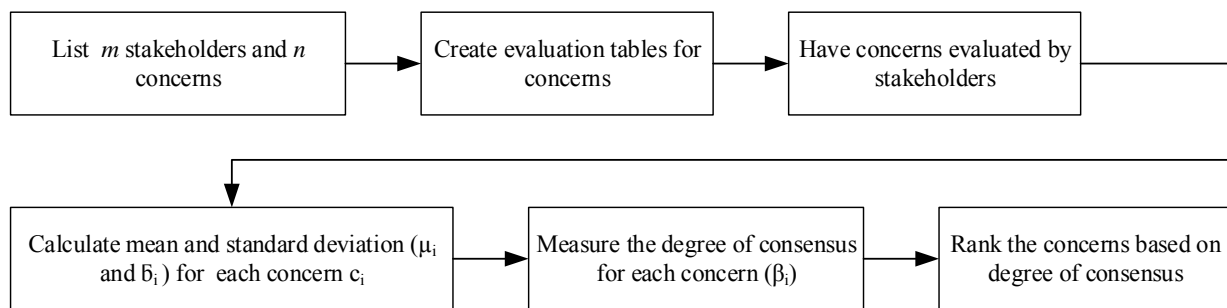


Figure 5. The process for ranking concerns.

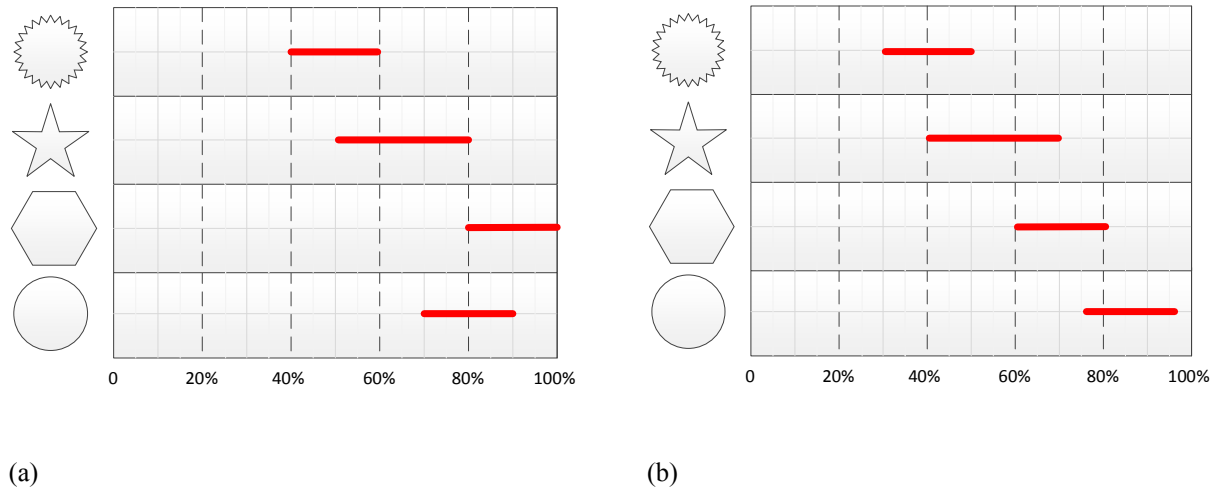


Figure 6. This figure presents the opinion of two stakeholders over the importance of four concerns shown by different figures. The numerical scale is used to present the importance of each concern.

III. EXAMPLE APPLICATION

To illustrate the application of the proposed method, a simple example is presented in this section. In the example, there are four concerns (hazards) in the pool of concerns (hazards). These concerns have typically been shown by different geometrical shapes (see Figure 6). Two stakeholders have ranked the concerns according to their views shown in this figure. The outcome of this ranking is presented in TABLE 1. The first column of this table shows a list of design concerns which are to be ranked. The rest of the columns respectively present the mean, standard deviation, relative weight, its uncertainty and relative uncertainty for each concern. The last column, which is highlighted, shows the degree of stakeholder’s consensus.

As seen in this table, there could be different results for ranking based on “relative weight” or “relative uncertainty”. Here the “degree of consensus” plays an important role to set the priority of concerns as it acts as a measure of the

reliability in each concern.





This example shows how the method is used to communicate with stakeholders, register their concerns, integrate the collected data and disclose the most important aspects. Similar results have been achieved through real-world case studies to prioritize the stakeholder consensus in terms of project requirements. See for example [6, 7].

IV. CONCLUSIONS

This study highlights the stakeholders’ concerns for identification of system hazards. Realization of key concerns and their ranking can be a challenging task due to a high number of stakeholders and their competing or conflicting interest.

The paper proposes an approach that uses a graphical

TABLE 1. THIS TABLE PRESENTS THE REQUIREMENTS AND THEIR WEIGHT FACTORS, STANDARD DEVIATIONS, RELATIVE WEIGHTS, UNCERTAINTIES IN RELATIVE WEIGHT, RELATIVE UNCERTAINTIES AND DEGREE OF CONSENSUS.

Concerns	Expected concern (μ_i %)	Standard deviation (σ_i %)	Relative weight (λ_i %)	Uncertainty in weight (σ_{λ_i} %)	Relative uncertainty (ϕ_i %)	Degree of Consensus (β_i %)
	45	5	10	1.1	11	9.1
	60	7.5	14	1.7	25	8.2
	80	5	18	1.1	11	16.4
	82.5	5	19	1.1	11	17.3

tool to communicate with stakeholders, collect the information and combine it in order to rank the concerns. The “degree of consensus” is used to rank concerns. The proposed approach is based on probability theory and promotes probabilistic thinking.

The use of this outcome for triangulation of hazard identification is the next step for this research.

REFERENCES

- [1] Beck, G. and C. Kropp, *Infrastructures of risk: A mapping approach towards controversies on risks*. Journal of risk research, 2011. 14(1): p. 1-16.
- [2] ISO(12100:2010), *Safety of machinery - general principles for design - risk assessment and risk reduction*. 2010.
- [3] EN(31010), *Risk management - risk assessment techniques*. 2010.
- [4] Rajabalinejad, M. and C. Spitas, *Incorporating uncertainty into the design management process*. Design Management Journal, 2012. 6(1): p. 52-67.
- [5] Engel, A. and T.R. Browning, *Designing systems for adaptability by means of architecture options*. Systems Engineering, 2008. 11(2): p. 125-146.
- [6] Rajabalinejad, M. and G.M. Bonnema. *Determination of stakeholders' consensus over values of system of systems*. in *Proceedings of the 9th International Conference on System of Systems Engineering: The Socio-Technical Perspective, SoSE 2014*. 2014.
- [7] Rajabalinejad, M. and G.M. Bonnema, *Probabilistic thinking to support early evaluation of system quality: Through requirement analysis*, in *5th International Conference on Complex Systems Design & Management (CSD&M) 2014, Paris, 12-14 November*. 2014: Paris.
- [8] Leveson, N., *Engineering a safer world*. 2012, Cambridge, Massachusetts, London, England: Massachusetts Institute of Technology.
- [9] Rajabali Nejad, M., G.M. Bonnema, and F.J.A.M.v. Houten, *An integral safety approach for design of high risk products and systems*, in *Safety and Reliability of Complex Engineered Systems* P.e. al., Editor. 2015, Taylor & Francis Group: Zurich, Switzerland.
- [10] Christel, M.G. and K.C. Kang, *Issues in requirements elicitation*. 1992, DTIC Document.
- [11] Heemels, W., L. Somers, P. van den Bosch, Z. Yuan, B. van der Wijst, A. van den Brand, and G. Muller, *The key driver method*. Boderc: Model-Based Design of High-Tech Systems, edited by W. Heemels and GJ Muller, 2006: p. 27-42.
- [12] Heemels, W., E. vd Waal, and G. Muller, *A multi-disciplinary and model-based design methodology for high-tech systems*. Proceedings of CSER, 2006.
- [13] Pahl, G., W. Beitz, and K. Wallace, *Engineering design: A systematic approach*. 1996: Springer Verlag.
- [14] Whitten, J.L., V.M. Barlow, and L. Bentley, *Systems analysis and design methods*. 1997: McGraw-Hill Professional.
- [15] Barron, F.H. and B.E. Barrett, *Decision quality using ranked attribute weights*. Management Science, 1996. 42(11): p. 1515-1523.
- [16] Takeda, E., K.O. Cogger, and P.L. Yu, *Estimating criterion weights using eigenvectors: A comparative study*. European Journal of Operational Research, 1987. 29(3): p. 360-369.
- [17] Saaty, T.L. and L.G. Vargas, *The logic of priorities: Applications in business, energy, health, and transportation*. 1982: Kluwer-Nijhoff.
- [18] Barzilai, J., *Deriving weights from pairwise comparison matrices*. Journal of the Operational Research Society, 1997. 48(12): p. 1226-1232.
- [19] Yeh, C.-H., R. J. Willis, H. Deng, and H. Pan, *Task oriented weighting in multi-criteria analysis*. European Journal of Operational Research, 1999. 119(1): p. 130-146.
- [20] Buckley, J.J., *Ranking alternatives using fuzzy numbers*. Fuzzy Sets and Systems, 1985. 15(1): p. 21-31.
- [21] Tsai, W.C., *A fuzzy ranking approach to performance evaluation of quality*. 2011. Vol. 18. 2011.
- [22] Mitchell, H.B., *Ranking-intuitionistic fuzzy numbers*. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 2004. 12(03): p. 377-386.
- [23] ISO(31000:2009), *Risk management — principles and guidelines*. 2009.
- [24] Zimmermann, H.J., *Fuzzy sets, decision making and expert systems*. Vol. 10. 1987: Springer.
- [25] Rajabalinejad, M. *Modelling dependencies and couplings in the design space of meshing gear sets*. 2012.
- [26] McManus, H. and D. Hastings, *A framework for understanding uncertainty and its mitigation and exploitation in complex systems*. IEEE Engineering Management Review, 2006. 34(3): p. 81-94.
- [27] Choy, S.L., R. O'Leary, and K. Mengersen, *Elicitation by design in ecology: Using expert opinion to inform priors for bayesian statistical models*. Ecology, 2009. 90(1): p. 265-277.
- [28] O'Hagan, A., J. Forster, and M.G. Kendall, *Bayesian inference*. 2004: Arnold London.
- [29] Melchers, R.E., *Structural reliability analysis and prediction*. 1999.

Using XPath to Define Design Metrics

Jan Söderberg, Ali Shahrokni

Systemite AB

Gothenburg, Sweden

e-mail: {jan.soderberg, ali.shahrokni}@systemite.se

Bashar Nassar

Chalmers University of Technology

Gothenburg, Sweden

e-mail: bashar.h.nassar@gmail.com

Abstract—Architecture description formats like EAST-ADL and automotive open system architecture (AUTOSAR) use an extensible markup language (XML) based file representation. The complexity of the systems based on these architecture description languages often call for metrics definitions for the purpose of complexity or completeness management. The Swedish research project *Synligare* deals with improved management of complex systems based on EAST-ADL. One result from the project was that XPath could be used as a basis for the definition of design metrics, offering several advantages. XPath has further been demonstrated in the project to offer sufficient expressiveness and usability for the purpose.

Keywords-Metrics; XPath; EAST-ADL; AUTOSAR; Exchange metrics.

I. INTRODUCTION

The evolution rate of automotive electric/electronic(E/E) systems has increased exponentially during the last decade, and the number of electronic control units now typically amounts to 50-100 [1]. New and complex functionalities and technologies are emerging, making the prospect of autonomous driving within reach [2]. A consequence of the higher complexity is that the classical document and file based methods are no longer sufficient to manage the product and process data. We have seen that the Software specification of a single Electronic Control Unit (ECU) can be in excess of 8,000 pages. Meanwhile, there is an increased demand for reduced development cycles and product costs.

Synligare¹ is a Swedish industrial research project that aims to improve methods and tool support for model-based development of automotive E/E systems within and between organizations [7]. The members of the Synligare project include Volvo AB, ArcCore AB, Autoliv AB, Semcon and Systemite AB. The parties represent the different roles in a typical E/E development project, including Volvo as a manufacturer and integrator ("OEM" in current automotive terminology), Autoliv as a Tier 1 supplier, ArcCore as a Tier 2 supplier, Systemite as a high level modeling tool supplier on high levels of abstraction, ArcCore as low level modeling tool supplier, and Semcon as a specialist engineering service supplier.

The project uses the EAST-ADL language [8] as a common specification for exchanging developed data within and between organizations. EAST-ADL is an adaptation of SysML[9] for automotive E/E systems. The language

includes support for high level specifications of the system, for instance, vehicle features, down to the implementation level, based on AUTOSAR[10]. The language includes optional packages for modeling of variability, timing, safety, and more.

One of the main objectives of the Synligare project is to enable exchange of functional safety data inside and across organizations. ISO 26262 is a standard for functional safety that challenges the automotive industry. The data is produced on different location by different companies. However, the progress needs to be measured, updated, and consolidated in different companies and exchanged between suppliers and OEMs. Many process and products metrics in the ISO 26262 standard are valid across organization boundaries. Many of the progress metrics can be extracted from product data. For instance, one such metric is the state of progress of the verification process for all technical safety requirements, or the state of fulfillment of safety goals on different levels of abstractions.

The Synligare project specifically addresses data exchange challenges between OEMs and suppliers. When the exchange is based on a single formalized representation like EAST-ADL the efficiency and quality of the exchange can be significantly improved, since handover of development, tracing impact of changes and analysis of data can be automated.

A remaining challenge when information is shared and exchanged is to assure that all involved parties can interpret the information in the same way. Although the XML based exchange format for EAST-ADL provides a formalization of the information, the way this information is viewed by different parties is not specified; specifically, when it comes to design metrics. For instance, EAST-ADL does not include progress measurements such as completeness or complexity of the design. In the Synligare project, these metrics were originally specified in natural language, with references to the constructs of the language. For specifying the metrics, we used a more formal alternative, inspired by XPath expressions[11], to express the metrics. These metrics could then be shared between different tools at the OEM and supplier sides to calculate the metrics in a unified way. Using common metrics enables the different groups and organizations to share a common view of the progress of the project. In this paper, we introduce this method of sharing metrics on model-based development data.

The remainder of this paper is organized as follows. In Section II, definition of the EAST-ADL Language, while metrics using path queries defined in Section III. Section IV presents the implementation aspects of XPath, while Section

¹Synligare means "more visible" in Swedish.

V dissection and conclusion, and Section VI gives a vision for future work.

II. THE EAST-ADL LANGUAGE

EAST-ADL is a domain specific architecture description language specialized for describing automotive E/E systems. The language supports the use of different levels of abstraction with traceability between the levels. The logical structure of an architecture expressed in EAST-ADL is according to a structural component model where components are connected through ports.

EAST-ADL defines an exchange format in XML, called EAXML [12]. The schema of the EAXML is the most precise definition of the language, although the underlying meta-model is defined in UML. The mapping between the meta-model and the XML schema is according to patterns defined in the AUTOSAR community. According to these patterns the schema becomes a reflection of the meta-model, and the schema will only include elements according to the meta-model.

Note that the principles behind the EAXML and ARXML (AUTOSAR xml) schemas differ from the schema of the XMI format, used for the representation of UML models; XMI is based on the more generic MOF (Meta Object Facility) framework [13]. This means that the schema of XMI will not reflect the used meta-model, but rather the meta-meta-model according to MOF. A consequence of importance to the use of XPath is that the element structure of an EAXML file is a direct reflection of the corresponding EAST-ADL model.

III. METRICS DEFINITIONS USING PATH QUERIES

XPath 2.0 became a W3C recommendation 2007. XPath is a specialized query language that can express selection criteria of nodes of an XML document, typically from within an XML style sheet. The selection criteria include the path to traverse in the structure of the document, and additional tests and predicates that must be fulfilled for the selected nodes.

The way XPath is used is by 1) selecting the sets of nodes in the XML document that are relevant for the specific metrics, and 2) performing arithmetic operations on the quantities defined by the sets.

In this section, we present two types of metrics that we have specified with path queries and shared between object model tools. The metrics are inspired by the XPath query language for XML files. The first type of metrics calculates the progress of the development process using the product data. The second type of metrics calculated the complexity of the product components.

A. Progress metrics

One type of the metrics that we defined and shared between tools extracts the state of the project from the development data specified in different tools. The underlying specification of the tools is EAST-ADL, which enables us to create generic metrics and share them between tools. One such metric describes the completeness of the allocation of requirements.

The metrics value was originally expressed in the Synligare project as: "Progress of requirement allocation is measured as the fraction of requirements allocated to architectural elements"

The two sets of elements involved in this calculation are 1) the set of all requirements, and 2) the set of allocated requirements.

The first set can be expressed as the path expression (1) below, which is assumed to start from a "EA-PACKAGE" context node of the EAXML document. Definition for different elements of the XML representation of the meta-model such as EA-PACKAGE is available on EAST-ADL's language specification documentation [8].

Note that since the EA-PACKAGE structure in an EAXML document is an arbitrary packaging structure, it is suitable to exclude this part from the definition, and define the part on a case to case basis.

$$\begin{aligned} & /ELEMENTS/REQUIREMENTS- \\ & MODEL/REQUIREMENTS/REQUIREMENT \end{aligned} \quad (1)$$

The set of allocated requirements is a subset of the set described above, with the additional constraint that the requirement must be included in a so called "Satisfy" relationship:

$$\begin{aligned} & /ELEMENTS/REQUIREMENTS-MODEL/OWNED- \\ & RELATIONSHIPS/SATISFY/SATISFIED- \\ & REQUIREMENT-REFS/SATISFIED-REQUIREMENT- \\ & REF \end{aligned} \quad (2)$$

The set of unallocated requirements can be defined as the difference between the two sets, using the "except" operation:

$$\begin{aligned} & /ELEMENTS/REQUIREMENTS- \\ & MODEL/REQUIREMENTS/REQUIREMENT \text{ except} \\ & /ELEMENTS/REQUIREMENTS- \\ & MODEL/REQUIREMENTS/REQUIREMENT \end{aligned} \quad (3)$$

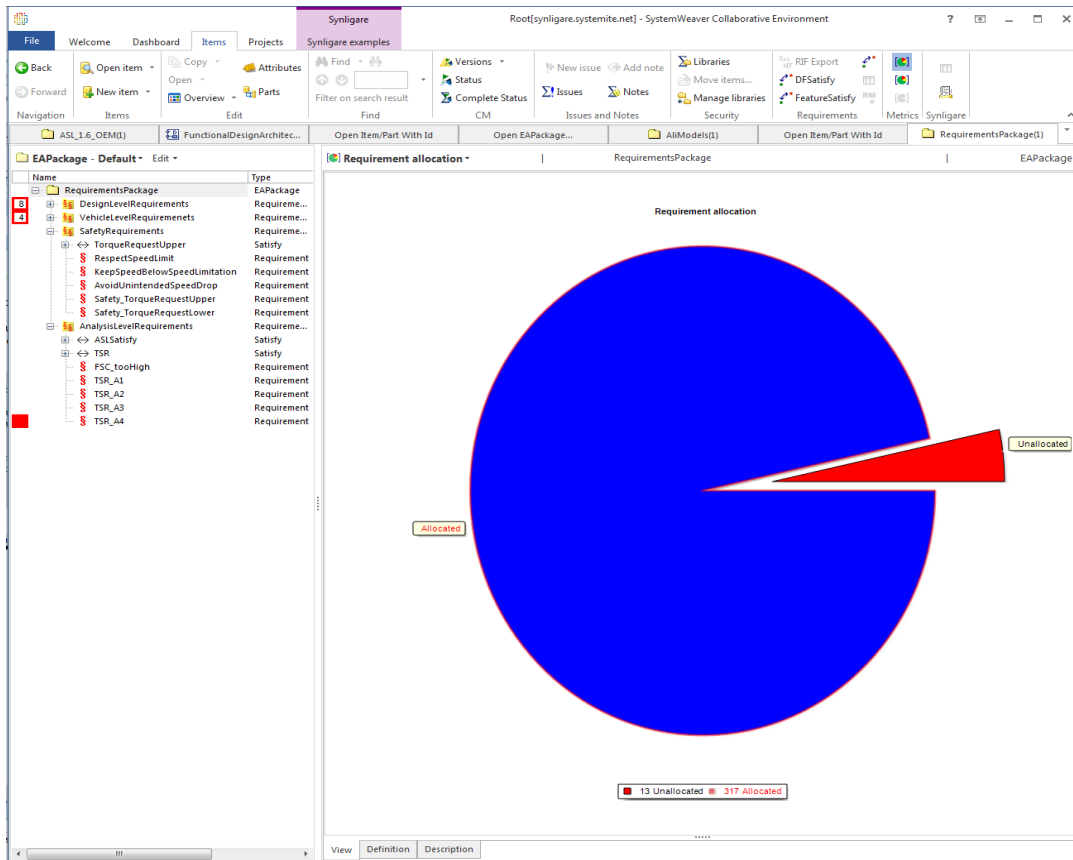


Figure 1 Completeness of allocated requirements

The fraction of the sets can be calculated using the XPath count function and div operator:

$$\frac{\text{count}(/ELEMENTS/REQUIREMENTS-MODEL/OWNED-RELATIONSHIPS/SATISFY/SATISFIED-REQUIREMENT-REFS/SATISFIED-REQUIREMENT-REF)}{\text{count}(/ELEMENTS/REQUIREMENTS-MODEL/REQUIREMENTS/REQUIREMENT)} \quad (4)$$

The real underlying need behind this metric is the need for traceability to the set of unallocated requirements. This traceability can be performed interactively using a pie chart representation of the set (3) in the SystemWeaver tool [14]. We see the evaluated system in the tree view to the left in Figure 1. The system is the reference system of the Synligare project, supplied by Volvo. The package "RequirementsPackage" has been selected, thereby selecting the context of the evaluation. The "Requirements allocation" view to the right displays a pie chart, where the two slices represent allocated requirements (in blue) and unallocated requirements (in red). By selecting the *Unallocated* slice, the set of model elements according to the XPath expression (3) become highlighted in the tree view.

B. Complexity of component models

Another type of metric that we investigated in this paper is the metrics concerning complexity of component models. One such complexity metric is cyclomatic complexity [5], calculated for a component model.

$$\frac{\text{count}(/CONNECTORS/FUNCTION-CONNECTOR) - \text{count}(/PARTS/DESIGN-FUNCTION-PROTOTYPE) + 2}{2} \quad (5)$$

Another component complexity metric uses couplings between objects [6]

$$\frac{\text{count}(/CONNECTORS/FUNCTION-CONNECTOR)}{\text{count}(/PARTS/DESIGN-FUNCTION-PROTOTYPE)} \quad (6)$$

IV. IMPLEMENTATION ASPECTS OF XPATH

In the Synligare project, support for metrics definitions expressed by the path query language was implemented in the SystemWeaver tool. SystemWeaver has a programmable meta-model and constitutes an internal database that can manage and integrate the content of multiple EAXML files. The constructs supported by the meta modeling framework in the tool supports the patterns used in EAST-ADL, like the type/prototype pattern. This means that the internal

representation in SystemWeaver to a high degree conforms to the EAXML file format. A database like the one in SystemWeaver is not limited to managing the content corresponding to a single system, but can manage any number of systems, and content shared between the systems.

SystemWeaver supports dimensions of data that is not supported by EAST-ADL, like versioning and management of contexts that go beyond the scope of a single system. Such dimensions correspond to additional axes of the XPath expressions that cannot be derived from the specific meta-model.

A specific challenge is the way references are expressed according to EAST-ADL and AUTOSAR. Instead of common XML ID/IDREF to express references, EAST-ADL and AUTOSAR uses element paths of the XML file to reference elements, e.g., "/DesignLevelElements/FCN/GlobalBrakeController/BrakeTorqueFL".

References like the one described above are common in the AUTOSAR/EAST-ADL models and means that the XPath expressions cannot be evaluated against a DOM (Document Object Model). Instead, the XML file has to be parsed and transformed into a custom object model where references have been replaced by object links. SystemWeaver for example represents the references as bi-directional object links. During an import of an EAXML file into SystemWeaver all path strings are parsed and replaced with object links.

It can be assumed that any tool that supports EAST-ADL or AUTOSAR will have an efficient internal representation of such references. We have seen that a real life AUTOSAR XML file can be of the size of 10 Mbyte or more, including more than 100,000 elements. A corresponding EAST-ADL model would include even more aspects, and thereby more elements. This means that efficiency becomes a real concern, especially when the evaluation of metrics is done interactively, or when the complexity of XPath expressions are $O(n^2)$ or higher, for instance, when set operations are used.

V. CONCLUSION AND DISCUSSION

In this paper, we presented a generic method to formalize metrics and share them between model-based data management tools. In the Synligare project, metrics originally expressed in natural language have been re-expressed in an XPath-like format and executed in different tools with identical results.

Being XML based, Xpath is intended for use with XML based representations. Since XPath is implementation independent it can work as a formal definition of the metrics, while also being executable.

Elwakil et al. [4] identified a number of advantages of using XQuery in metrics definitions for XMI based representations. These advantages have been found to hold also for XPath, being a subset of XQuery, for the case that data is represented in the more basic XML representations used for AUTOSAR or EAST-ADL:

- The XPath expressions can be expressed according to the meta-model of the used architecture language, meaning that the correctness of the expressions can be validated statically.
- The XPath language is standardized, technology independent, mature and wide spread.
- A tool implementation of the method may directly interpret and execute the XPath expressions. This makes it easy to try different metrics expressions in the tool implementation, without changing the tool itself.

In addition to these findings, the implementation of the support for XPath has taken benefit from the fact that XPath supports the selection of sets of elements, thus making it suitable for interactive analysis and traceability between the visualization of the metrics and the underlying data.

The solution has been demonstrated using industrial examples, with satisfactory performance.

There are some natural limitations and disadvantages of using XPath:

- The approach is likely feasible only for those cases where the language is expressed as XML; specifically, that the schema is a reflection of the used meta-model.
- Given the declarative characteristics of the language it is likely that not all types of metrics can be defined easily in the language. The use of XQuery as described in [3] has not been investigated for the type of representation used in the project, but may be an alternative for more complex types of metrics.

VI. FUTURE WORK

The evaluation of XPath for metrics definitions described in this paper was limited to the use cases of the Synligare project. It remains to evaluate the suitability of the approach for other types of metrics.

ACKNOWLEDGEMENTS

This work was supported by VINNOVA under the FFI Programme, Grant 2013-2196 Synligare.

REFERENCES

- [1] D. Goswami et al., "Challenges in automotive cyber-physical systems design." In *Embedded Computer Systems (SAMOS), International Conference on 2012*, pp. 346-354. IEEE.
- [2] R. Okuda, Y. Kajiwara, and K. Terashima, "A survey of technical trend of ADAS and autonomous driving." In *VLSI Technology, Systems and Application (VLSI-TSA), Proceedings of Technical Program-International Symposium on 2014*, pp. 1-4. IEEE.
- [3] M. Sharma, N. S. Gill, and S. Sikka, "Survey of object-oriented metrics: Focusing on validation and formal specification." *ACM SIGSOFT Software Engineering Notes*, 2012, 37.6: 1-5.
- [4] M. El-Wakil, A. El-Bastawisi, M. Riad, and A. Fahmy, "A novel approach to formalize and collect Object-Oriented Design-Metrics." *9th International Conference on Empirical Assessment in Software Engineering*, 2005.

- [5] Y.Jiang, B. Cuki, T. Menzies, and N. Bartlow, "Comparing design and code metrics for software quality prediction." In Proceedings of the 4th international workshop on Predictor models in software engineering, 2008, pp. 11-18. ACM.
- [6] S. S.Rathore and A. Gupta, (2012, September). "Investigating object-oriented design metrics to predict fault-proneness of software modules." In Software Engineering (CONSEG), CSI Sixth International Conference on 2012, pp. 1-10. IEEE.
- [7] Synligare Consortium: Synligare Project website. Available from: <http://www.synligare.eu/> [retrieved: Dec, 2015].
- [8] EAST-ADL Association: EAST-ADL web site. Available from: <http://www.east-adl.info/> [retrieved: Dec, 2015].
- [9] SysML. Available from: <http://www.omg.sysml.org/> [retrieved: Dec, 2015].
- [10] AUTOSAR Development Partnership: AUTOSAR web site. Available from: <http://www.autosar.org/> [retrieved: Dec, 2015].
- [11] XPath. Available from: <http://www.w3.org/TR/xpath/> [retrieved: Dec, 2015].
- [12] EAXML specification. Available from: http://www.east-adl.info/2.1.12/eastadl_2-1-12.xsd [retrieved: Dec, 2015].
- [13] MOF. Available from: <http://www.omg.org/mof/> [retrieved: Dec, 2015].
- [14] SystemWeaver. Available from: www.systemweaver.se [retrieved: Dec, 2015].