ICONS 2016

The Eleventh International Conference on Systems

ISBN: 978-1-61208-451-0

EMBEDDED 2016

International Symposium on Advances in Embedded Systems and Applications

February 21 - 25, 2016

Lisbon, Portugal

ICONS 2016 Editors

Gary Weckman, Ohio University, USA

Raimund Ege, Northern Illinois University, USA

ICONS 2016

Forward

The Eleventh International Conference on Systems (ICONS 2016), held between February 21-25, 2016 in Lisbon, Portugal, continued a series of events covering a broad spectrum of topics. The conference covered fundamentals on designing, implementing, testing, validating and maintaining various kinds of software and hardware systems. Several tracks were proposed to treat the topics from theory to practice, in terms of methodologies, design, implementation, testing, use cases, tools, and lessons learnt.

In the past years, new system concepts have been promoted and partially embedded in new deployments. Anticipative systems, autonomic and autonomous systems, self-adapting systems, or on-demand systems are systems exposing advanced features. These features demand special requirements specification mechanisms, advanced behavioral design patterns, special interaction protocols, and flexible implementation platforms. Additionally, they require new monitoring and management paradigms, as self-protection, self-diagnosing, self-maintenance become core design features.

The design of application-oriented systems is driven by application-specific requirements that have a very large spectrum. Despite the adoption of uniform frameworks and system design methodologies supported by appropriate models and system specification languages, the deployment of application-oriented systems raises critical problems. Specific requirements in terms of scalability, realtime, security, performance, accuracy, distribution, and user interaction drive the design decisions and implementations. This leads to the need for gathering application-specific knowledge and develop particular design and implementation skills that can be reused in developing similar systems.

Validation and verification of safety requirements for complex systems containing hardware, software and human subsystems must be considered from early design phases. There is a need for rigorous analysis on the role of people and process causing hazards within safety-related systems; however, these claims are often made without a rigorous analysis of the human factors involved. Accurate identification and implementation of safety requirements for all elements of a system, including people and procedures become crucial in complex and critical systems, especially in safety related projects from the civil aviation, defense health, and transport sectors.

Fundamentals on safety-related systems concern both positive (desired properties) and negative (undesired properties) aspects. Safety requirements are expressed at the individual equipment level and at the operational-environment level. However, ambiguity in safety requirements may lead to reliable unsafe systems. Additionally, the distribution of safety requirements between people and machines makes difficult automated proofs of system safety. This is somehow obscured by the difficulty of applying formal techniques (usually used for equipment-related safety requirements) to derivation and satisfaction of human-related safety requirements (usually, human factors techniques are used).

The conference had the following tracks:

- Specialized systems
- Security and protection systems
- Embedded and IT systems

The conference also featured the following symposium:

- **EMBEDDED 2016*, The International Symposium on Advances in Embedded Systems and Applications***

We take here the opportunity to warmly thank all the members of the ICONS 2016 technical program committee, as well as the numerous 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 ICONS 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 ICONS 2016 organizing committee for their help in handling the logistics and for their work that made this professional meeting a success.

We hope ICONS 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 systems. 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.

ICONS 2016 Advisory Committee
Mark Austin, University of Maryland, College Park, USA
Raimund Ege, Northern Illinois University, USA
Hermann Kaindl, Vienna University of Technology, Austria
Leszek Koszalka, Wroclaw University of Technology, Poland
Marko Jäntti, University of Eastern Finland, Finland

ICONS 2016 Committee Chair
Gary Weckman, Ohio University, USA

EMBEDDED 2016 Advisory Committee
Sabina Jeschke, RWTH Aachen University, Germany
I-Cheng Chang, National Dong Hwa University, Taiwan
Ralf-D. Kutsche, TU Berlin / Fraunhofer FOKUS institute, Germany
Albert M. K. Cheng, University of Houston, USA

ICONS 2016

Committee

ICONS 2016 Advisory Committee

Mark Austin, University of Maryland, College Park, USA
Raimund Ege, Northern Illinois University, USA
Hermann Kaindl, Vienna University of Technology, Austria
Leszek Koszalka, Wroclaw University of Technology, Poland
Marko Jäntti, University of Eastern Finland, Finland

ICONS 2016 Committee Chair

Gary Weckman, Ohio University, USA

ICONS 2016 Technical Program Committee

Mehmet Aksit (Akşit), University of Twente - Enschede, The Netherlands
Marco Aiello, University of Groningen, The Netherlands
Abdallah Al Sabbagh, University of Technology - Sydney (UTS), Australia
Cristina Alcaraz, University of Malaga, Spain
Eduardo Alonso, City University London, UK
Giner Alor Hernández, Instituto Tecnológico de Orizaba - Veracruz, México
César Andrés, Universidad Complutense de Madrid, España
Luis Anido-Rifon, University of Vigo, Spain
Mark Austin, University of Maryland, College Park, USA
Javier Bajo Pérez, Universidad Politécnica de Madrid, Spain
Lubomir Bakule, Institute of Information Theory and Automation of the ASCR, Czech Republic
Zbigniew Banaszak, Warsaw University of Technology | Koszalin University of Technology, Poland
Jacob Barhen, Oak Ridge National Laboratory, USA
Nicolas Belanger, Eurocopter Group, France
Ateet Bhalla, Independent Consultant, India
Jun Bi, Tsinghua University - Beijing, China
Francesco Bianconi, University of Perugia, Italy
Freimut Bodendorf, University of Erlangen-Nuremberg, Germany
Alisson Brito, Universidade Federal da Paraiba (UFPB), Brazil
Miroslav Bursa, CIIRC Institute - CTU in Prague, Czech Republic
Mario Cannataro, University "Magna Græcia" of Catanzaro - Germaneto, Italy
M. Emre Celebi, Louisiana State University in Shreveport, USA
Chi-Hua Chen, National Chiao Tung University, Taiwan , R.O.C.

Heinz-Dietrich Wuttke, Ilmenau University of Technology, Germany
Xiaodong Xu, Beijing University of Posts and Telecommunications, China
Linda Yang, University of Portsmouth, UK
Sameh Yassin, Cairo University, Egypt
Chang Wu Yu (James), Chung Hua University, Taiwan
Mudasser Wyne, National University, USA
Wai YuenSzeto, University of Hong Kong, Hong Kong
Sherali Zeadally, University of Kentucky, USA
Xiangmin Zhang, Wayne State University, USA
Wenjie Zhang, University of New South Wales - Sydney, Australia
Ying Zhang, University of New South Wales - Sydney, Australia
Ty Znati, University of Pittsburgh, USA
Dawid Zydek, Idaho State University, USA

EMBEDDED 2016

EMBEDDED 2016 Advisory Committee

Sabina Jeschke, RWTH Aachen University, Germany
I-Cheng Chang, National Dong Hwa University, Taiwan
Ralf-D. Kutsche, TU Berlin / Fraunhofer FOKUS institute, Germany
Albert M. K. Cheng, University of Houston, USA

EMBEDDED 2016 Technical Program Committee

Arnulfo Alanis, Instituto Tecnológico de Tijuana, Mexico
Cristina Alcaraz, Universidad de Malaga, Spain
Eric Armengaud, AVL List GmbH, Austria
Mohamed Bakhouya, International University of Rabat, Morocco
Juergen Becker, Karlsruhe Institute of Technology - KIT, Germany
Fadila Bentayeb, University of Lyon 2, France
Francois Bossard, CNES, France
Jalil Boudjadar, Linköpings University, Sweden
Patrick Brezillon, LIP6 - University Pierre and Marie Curie (UPMC), France
Juan Carlos Cano Escribá, Universitat Politècnica de València, Spain
I-Cheng Chang, National Dong Hwa University, Taiwan
Jiann-Liang Chen, National Taiwan University of Science and Technology, Taiwan
Albert M. K. Cheng, University of Houston, USA
Li-Der Chou, National Central University Taoyuan, Taiwan
Cesar A. Collazos, Universidad del Cauca, Columbia
Jianguo Ding, University of Skövde, Sweden
Manuel Filipe Santos, University of Minho, Portugal
Francesco Flammini, Ansaldo STS, Italy
Luis Javier García Villalba, Universidad Complutense de Madrid, Spain

Zhishan Guo, University of North Carolina at Chapel Hill, USA
J. Javier Gutiérrez, Universidad de Cantabria, Spain
Chia Hung Yeh, National Sun Yat-Sen University, Taiwan
Sabina Jeschke, RWTH Aachen University, Germany
Zhen Jiang, West Chester University of Pennsylvania, USA
Mehdi Khouja, University of Gabes, Tunisia
Brian (Byung-Gyu) Kim, SunMoon University, South Korea
Jeongchang Kim, Korea Maritime and Ocean University (KMOU), South Korea
Ki-Il Kim, Gyeongsang National University, Korea
Markus Kucera, Ostbayerische Technische Hochschule Regensburg (OTH Regensburg), Germany
Ralf-D. Kutsche, TU Berlin / Fraunhofer FOKUS institute, Germany
Chang-Gun Lee, Seoul National University, South Korea
Claudia Linnhoff-Popien, Ludwig-Maximilians-Universität München, Germany
Yasser M. Madany, Alexandria University, Egypt
Zoubir Mammeri, IRIT - Paul Sabatier University, France
Francisco Martins, University of Lisbon, Portugal
Piotr Matyasik, AGH University of Science and Technology, Poland
Michele Melchiori, Università degli Studi di Brescia, Italy
Young B. Moon, Syracuse University, USA
Maciej Ogorzalek, Jagiellonian University, Poland
George Perry, University of Texas at San Antonio, USA
Lee Pike, Galois Inc., USA
Loïc Plassart, JCP Connect, Rennes, France
Kostas Psannis, University of Macedonia, Greece
Grzegorz Redlarski, Gdansk University of Technology, Poland
Gunter Saake, Otto-von-Guericke-University Magdeburg, Germany
Lars Schnieder, Deutsches Zentrum für Luft- und Raumfahrt e.V. (DLR), Germany
Agusti Solanas, Rovira i Virgili University, Spain
Rafael Sotelo, Universidad de Montevideo, Uruguay
Marcin Szpyrka, AGH University of Science and Technology, Poland
Ronald Toegl, Siemens AG Österreich, Austria
Shin-Jer Yang, Soochow University, Taiwan
Shingchern D. You, National Taipei University of Technology, Taiwan
Gera Weiss, Ben Gurion University of the Negev, Israel
Chang Wu Yu, Chung Hua University, Taiwan
Thomas Canhao Xu, University of Turku, Finland
Olivier Zendra, INRIA, France

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 or 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 is 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

Peer-Sourced Media Management for Augmented Reality

Raimund K. Ege
Dept. of Computer Science
Northern Illinois University
DeKalb, IL, USA
email: ege@niu.edu

***Abstract*—Wearable devices with advanced recording devices enable the capture of current scenes and scenery in real time. It enables device holders to become media producers. Moreover, with its computing power and network connectivity, the wearable device can become a peer in a peer-to-peer based content delivery network. In this paper, we will describe a framework for allowing peers to join a content capture and delivery system that collects real-time media streams with virtual reality models in the cloud. The combined media pool represents an augmented reality world which is made available back to the peers and rendered onto their wearable devices. Issues arise, such as the combining and correlating of media streams in real time as well as capturing reference data from related streams and virtual reality models: we discuss our approach. We also outline an implementation in the Java programming language for Android and cloud platforms to justify and demonstrate the feasibility of our approach.**

Keywords-Android; augmented reality; virtual reality; peer-to-peer systems; multi-media content delivery

I. INTRODUCTION

Wearable devices with computing power, display and recording capabilities are becoming increasingly available and affordable. Moreover, these devices are network connected at high speeds, low latency and high bandwidth. In this paper, we will describe a framework for pairing these smart devices with computing power from the cloud to form a peer-sourced media content management system that enables realistic immersion into an augmented reality world.

A real-world application scenario could be a team of firefighters entering a burning building on a search and rescue mission. Each firefighter is equipped with network-connected sensors that capture video, audio, temperature, air quality data etc. The multiple streams of data are gathered and combined by cloud-based compute nodes with a virtual reality model of the building to form an augmented reality model. Each firefighter wears a network-connected display device in his/her helmet which receives a customized heads-up display of his/her forward view and situation. Even if smoke has filled the immediate surroundings of the firefighter the heads-up display might enable him/her to accomplish life-saving actions.

The framework we describe in the paper has many components, which we will describe in detail. Section 2 surveys the state of research as it relates to virtual and augmented reality, wearable devices and multi-media mediation. Section 3 explains how we gather multi-media sources from sensors and tag them suitable for adequate correlation and mediation. Section 4 details how peers are established, join the content delivery network and establish trust relationships with each other. Section 5 covers how tagged media streams are embedded into a reference virtual reality model and rendered into a geo-referenced attitudinal output media stream suitable for a heads-up display. Section 6 reports on the status of our prototype implementation in Java for cloud-based compute nodes and Android-based handheld and wearable devices. We conclude the paper with an assessment of our approach and future directions for our research.

II. BACKGROUND

In the computer and game console world, multi-player games are common: players throughout the Internet participate in a fantasy world and interact for a purpose, typically chasing and fighting enemies and each other in real time. Such multi-player games are part of the larger augmented virtual reality set of applications with a long research history [1]. These applications are being investigated for uses in telemedicine, manufacturing, etc. The basic idea is to combine a virtual model with actual sensor data to guide humans in their endeavor. The advent of wearable devices with a wide range of sensors is enabling a richer and deeper immersion in the given scenario [2] [3]. The ultimate goal is to increase the ratio of actual "real" content to virtual content. Reducing the virtual reality component to zero yields pure augmented reality.

The key capability of an augmented reality system must be the accurate recording from a real-time sensor. Multi-media I/O components include high-definition screens and video cameras, high-fidelity speakers and microphones. Plus components to determine device location, position, and attitude: GPS, accelerometers, compass, etc. Not only must the data be recorded and made available in real-time, but it must also be annotated and tagged with a variety of attributes

 ISBN: 978-1-61208-451-0

to enable correlation with other streams and embedding into a reference frame work [4]. For example, a video recording stream, in addition to capturing the sequence of video frames must also record exact time and location where the location must include direction and attitude. The richness of such metadata determines how precise and life-like the resulting augmented reality world will be [5].

Wearable connected computing devices, such as wristwatches and even eye glasses (Google GLASS) have reached the consumer market. The focus is shifting from computing and storage capabilities on these devices to connectivity and multi-media sensor and reproduction components. Connectivity capabilities are typically wireless and include high-bandwidth cellular (4G, LTE) and WLAN (IEEE 802.11) connections, plus lower-bandwidth near field connections (Bluetooth, NFC, etc.). Transmission rates in the multi megabits per second range and latency rates in the sub millisecond range are currently quite standard.

Who does the recording and contributes is another important aspect of sourcing sensor data to augment reality. Is the other player that appears in a shoot-first-ask-questions-later game a friend or foe? In peer-sourced augmented reality systems, the management of the multi-media source and establishment of trust is essential. In our prior work [6] [7], we investigated the authentication of participants in peer-to-peer networks, the establishment and management of trust, and the use of such media sources in building content management systems. An important lesson was that while modern mobile devices are compute-capable, cloud-based components add additional heft and authority to a seamless and smooth creation of a truly immersing virtual and augmented reality experience.

III. MEDIA CAPTURE

It all starts with recording something in real time using a sensor. The type of sensor can be a video or audio recorder, a location sensor, an attitude sensor, or even a sensor of biological data, and many more. In addition to the actual data sensed, it must be packaged with the exact time of recording. Multiple streams of sensor data are combined into a multi-media stream which interleaves its content streams plus provides meta-data to ensure their proper sequencing and correlation. It is important that the container format used to wrap the content streams is flexible enough to accommodate not only the stream data but also extensive amounts of reference information used to combine the streams. We are using an extension of the WebM project [8] format. The WebM container format is an open standard and allows us to collate an unlimited number of video, audio, pictures and subtitle tracks into one stream. We add the capability of identifying reference elements at identified points in time and at locations.

Video data is the key stream type captured via video sensors, i.e., cameras, available on the wearable devices carried by a peer. Video is captured as a sequence of video frames. Each frame carries a time stamp as major meta reference data. Equally important is the location of video capture, lens parameters and attitude, i.e., which way the camera points. Our container format allows us to group sequential video frames into video sequences that share a common location. We represent the location with a “Normal Vector”.

Figure 1. Normal Vector

Figure 1 shows how a normal vector captures not only the location of a video plane but also its relative position. The normal vector is represented via 2 points: its origin and extent points. Both points are captured in absolute latitude and longitude coordinates. While the distance between origin and extent point of the normal vector is not normally relevant, we use the length of the vector as a guide to the size of the video frames being referenced. A longer vector indicates a larger area shown in the video. We use the length of the normal vector when attaching multiple streams into a virtual reality frame.

Audio data is captured by microphones and sequences into frames that are referenced with a time stamp. While it would be possible to also capture and store directional information, which might be meaningful in the case of a directional microphone, we are able to deduce that information from the normal vector stored for video frames recorded at the same time on the same device. Of course, if the wearable device only records video, then such directional information is not available. We are considering this extension for future work. The audio data with its correlated reference metadata is also wrapped into the same container as the video data.

Any other data, such as gathered from biological data sensor, is equally framed and referenced. Examples of such data might be the heart rate of the person wearing the device, the temperature of the surroundings, or movement/acceleration data measured. Our container format

 ISBN: 978-1-61208-451-0

allows a free-form type designator that enables sensors of any kind, as long as their sensed data can be digitized and framed.

Our container format also allows the carrying of virtual reality model data. Actually, such data is similar in nature to "real" data, but is derived not from sensors but from virtual reality models of the surroundings that the wearers of the wearable devices inhabit. To allow clear distinction of sub-streams within the container, each sub-stream carries a unique stream identifier, which is correlated to a stream dictionary that holds relevant information about the sub-stream. The complete stream dictionary is embedded into the multi-media stream at regular intervals.

IV. PEER MANAGEMENT

The richness of the resulting augmented reality depends on the number of contributing peers. Users with wearable devices can join the network and become peers in the peer-to-peer content sharing network. In our prior work [5, 6] we investigated the authentication of participants in peer-to-peer networks, the establishment and management of trust, and the use of such media sources in building content management systems. In peer-sourced augmented reality systems management of multi-media source and establishment of trust is essential.

Our approach delegates peer identification to an OpenId provider, but maintains a shared understanding of how trust-worthy a peer is. Peers that have reached a trust threshold – initially just one bootstrap peer – maintain a database of trust information (called trust nuggets) per peer. The trust nuggets are encrypted with a private key that is shared by all trusted peers. The trust nugget stores information about a peer's past participation in the content sharing network and exposes the peer's trust rating. The public key to decrypt the trust nugget is shared among all peers that participate in the network, which enables any peer to evaluate another peer's trust worthiness.

Each peer can produce and consume media streams. A stream produced by a peer carries the peer's trust value. And a peer is only allowed to consume streams that match his/her trust value. Each successful participation of a peer in a construction of an augmented realty scenario adds to the peer's trust value. The necessary adjustment of the trust nugget has to be performed by a trusted peer, who will then reseal the trust nugget with the private key.

Each peer also generates a public/private key pair. The public key is also stored in the peer's trust nugget. The public/private key pair is used when peers exchange streams: the key pairs are used to establish a shared session key, which is used to encrypt and decrypt the streams.

V. MEDIA MEDIATION

Media streams that are collected by wearable devices worn by peers are embedded into a reference virtual reality model and rendered into a geo-referenced attitudinal output media stream suitable for a heads-up display. The key to an exact construction of a resulting stream is that input streams are mediated into a stable reference model. This model is provided by a virtual reality model of the surroundings of the scenario that the peers inhabit. We use the Virtual Reality Modeling Language (VRML). Geo-referenced meta and attitudinal data that accompanies the peer-gathered streams is used to create an augmented reality model. The resulting augmented reality model is rendered into a resulting media stream which is available to participating peers.

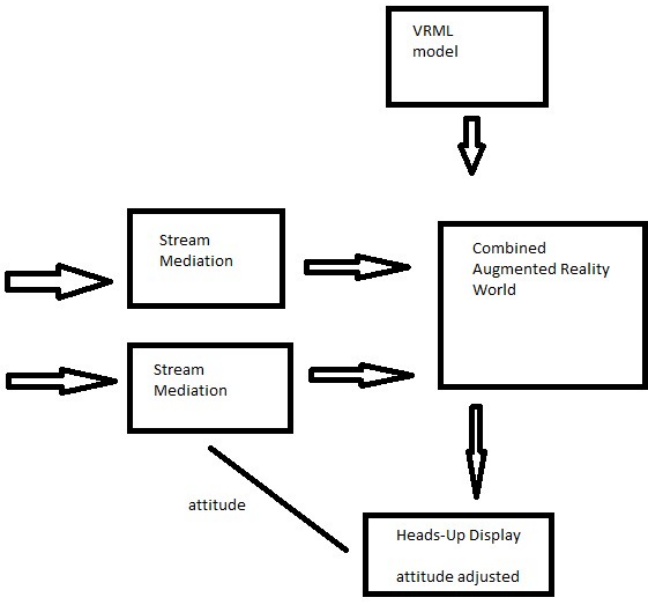

Figure 2. Cloud Stream Mediation

Figure 2 shows how peer-contributed media streams are combined and embedded into a reference world that is created from input using the VRML standard. The mediation of input streams can involve geometric conversions to adjust the spatial location and dimensions to produce a life-like presentation. The rendering of the resulting augmented reality can be adjusted to conform to the attitude of an input stream: this enables a peer to view the resulting stream in the same geo space as its own recording device.

 ISBN: 978-1-61208-451-0

The stream mediation is performed by a cloud-based general purpose peer. While any peer could perform this peer functionality, it requires significant computing power. Modern wearable devices are gaining computing power as technology progresses, but making this capability available in the cloud allows any peer to participate without risking to hamper the usability of its device.

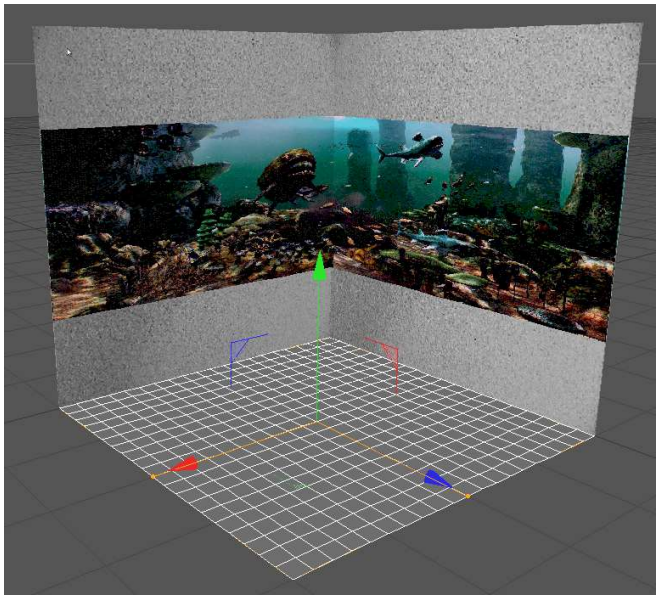

Figure 3. Augmented Reality World

Figure 3 shows a simple scenario of a cube-shaped world, where 2 input streams are embedded onto walls. The streams are contributed from two peers, the cube context is provided in VRML format.

VI. PROTOTYPE

Our effort includes the implementation of a prototype to demonstrate the feasibility of our approach. We provide peer capabilities in a modularized fashion: each module provides a feature that represents typical peer functionality. An actual peer can be constructed by assembly from one or more modules. All modules are programmed in the Java programming language suitable to be applied into an app for the Android platform. Peers can also live in the cloud, all they need is a Java virtual machine to run.

Four modules are part of our prototype implementation: (1) peer management: it implements peer authentication, key generation and heart beat connections to the trusted peer in the network; (2) source stream creation: it implements the selection of suitable input sensors, the recording a data from the sensor, the collection of meta data and the creating of the container stream to be output from this peer; (3) stream consumption: it implements the selection of an input stream and its display on an output devices; and (4) stream mediation: it implements the adaptation of input streams into a virtual reality world to form a augmented reality world. The peer management module also includes the capabilities of the trusted peer once the peer has reached the threshold required.

A typical peer app will include peer management, source stream creation, and stream consumption to enable a peer to be an active participant in an augmented reality scenario. The peer contributes the stream captured by its own sensor, e.g., video, and displays the combined and mediated resulting world from a cloud-based source.

VII. CONCLUSION

In this paper, we described a framework whose components produce a life-like augmented reality world. We gather multi-media sources from sensors and tag them suitable for adequate correlation and mediation. Peers join a content delivery network and establish trust relationships among each other. Cloud-based media streams – mediated based on their associated metadata - are embedded into a reference virtual reality model and rendered into a geo-referenced attitudinal output media stream suitable for a heads-up display. We are working on a prototype implementation in Java for cloud-based compute nodes and Android-based handheld and wearable devices.

Our goal for future research is to enhance the media mediation capabilities of our cloud-based peer components. We envision a rich augmented reality world that is populated by many dynamic input streams. While our current approach only provides one output stream that is adjusted from one input stream's geo attitude, we will work on better feedback from a given peer's input to the output that is consumed by the same peer.

REFERENCES

[1] R. Azuma, et al., "Recent Advances in Augmented Reality," IEEE Computer Graphics and Applications (CGA) 21(6):2001, pp. 34-47.

[2] D. Wagner, G. Reitmayr, A. Mulloni, T. Drummond and D. Schmalstieg, "Real-Time Detection and Tracking for Augmented Reality on Mobile Phones," IEEE Transactions on Visualization and Computer Graphics, 16(3), 2010, pp. 355-368.

[3] A. Morrison, et al., "Collaborative use of mobile augmented reality with paper maps," Journal on Computers & Graphics (Elsevier), 35(4), 2011, pp. 789-799.

[4] E. Macias, J. Lloret, A. Suarez and M. Garcia, "Architecture and Protocol of a Semantic System Designed for Video Tagging with Sensor Data in Mobile Devices," Sensors,vol.12, no. 2, 2012, pp. 2062–2087.

[5] Meenakshi Sundaram V., Shriram K. Vasudevan, A. Ritesh and C. Santhosh, "An Innovative App with for Location Finding with Augmented Reality using

 ISBN: 978-1-61208-451-0

CLOUD," 2nd International Symposium on Big Data and Cloud Computing (ISBCC'15), Procedia Computer Science 50, 2015, pp. 585 – 589.

[6] Raimund K. Ege, "Peer to Peer Media Management for Augmented Reality," International Conference on Networking and Services (ICNS 2015), Rome, Italy, May 2015, pp. 95-100.

[7] Raimund K. Ege, "Secure Trust Management for the Android Platform," International Conference on Systems (ICONS 2013), Seville, Spain, January 2013, pp. 98-103.

[8] The WebM Project - About WebM. http://www.webmproject.org. [accessed August 19, 2015]

Complexity-based Thinking in Systems Intelligence for Systems Resilience

Roberto Legaspi
Transdisciplinary Research Integration Center
Research Organization of Information and Systems
Tokyo, Japan
E-mail: legaspi.roberto@ism.co.jp

Hiroshi Maruyama
Department of Statistical Modeling
The Institute of Statistical Mathematics
Tokyo, Japan
E-mail: hm2@ism.co.jp

Abstract— **We posit that our models and approaches in systems resilience persistently demonstrate fragmented and dispersed knowledge because we fail to fully perceive the complexities of our systems and the situations that daunt them. We argue for a systems intelligence that has complexity-based thinking at its foundation. Complexity-based thinking involves methodological pluralism, law of requisite knowledge, and complexity absorption. The system integrates knowledge from heterogeneous sources, namely, massive information data points, expert and experiential knowledge, and perceptions of human sensors. As new facts are continuously derived with incoming evidence, the intelligent system self-improves its knowledge. With the synergism of heterogeneous knowledge, the emergence of new intelligence is possible. The integrated knowledge may expose unstated assumptions, reconcile inconsistencies and conflicts, and elucidate ambiguities in complex system behavior. The integrated knowledge is also aimed to influence the course of system vulnerabilities, destructive perturbations, and critical systemic changes.**

Keywords-complex systems; intelligent systems; complexity-based thinking; systems resilience.

I. INTRODUCTION

Although we have achieved significant advances in science and technology, human and economic losses due to disasters, accidents, social upheavals, and humanitarian crises remain significant. We believe that the reason for this is that we have yet to fully perceive the complexities of our world and life systems. Their nature is indeed complex, i.e., nonlinear, spanning multiple simultaneous temporal and spatial scales, and with large interdependencies among parts, which make it almost impossible to decompose them into independent processes [1]. Such behaviors may cause one situation, albeit due to a small stress or shock, to become critical and trigger other events in a cascading fashion such that the different situations within the propagation enhance themselves to criticality. Their heightened complexity can also pave the way for hazards that are extreme, unknown, unforeseen, or ill defined to impact the social, physical, environmental and technological dimensions simultaneously.

Consider for example the tortilla riots in Mexico in 2007 that was indirectly caused by a seemingly disconnected event – Hurricane Katrina in 2005. Zolli and Healey [2] recounted the events: Katrina disrupted 95% of oil production in the Gulf for several months that led to the price surge of American gasoline. This surge spurred investments in the alternative ethanol, which has corn as primary ingredient. Mexican farmers found themselves competing with the euphoria of ethanol investment bubble and with the US corn being dumped on Mexican markets at almost 20% less production cost. These interrelations and interdependencies between systems, namely, the oil rigs in the energy system, Katrina of the ecosystem, corn of the agricultural system, global trade system, and the social and political systems of US and Mexico, resulted to a rich world-energy (barrel of oil of nearly $140) being in direct competition with poor-world food (skyrocketed cost of corn). Each individual shock might have been felt by some sectors individually, but no one would have predicted the coincident crises in energy, finance, and food, as well as the confluence of shocks [3].

Another example, as recounted by McCracken [4], is the fragmented knowledge that if were connected could have helped predict the events of 9/11: At the time of 9/11, any clear predictive knowledge of an attack was absent – vital indicative information that in combination could have served as a warning were scattered in isolated stovepipes in the CIA, NSA, FBI, and State Department. Hence, analysts across these agencies had only fragments that did not help prevent the loss of nearly 3,000 lives and 6,000 more injured, tons of rubble and steel that obscured eight city blocks for almost nine months, and a shock to pertinent US financial markets. Only after did the intelligence community piece together the indicators of the large-scale attack on US soil.

In the presence of daunting complexities, our systems need to be resilient. *Resilience* is the ability of a system to persist, adapt, or transform in structure and function in the midst of even large shocks and stresses that come from a range of hazards [1]. With looming world crises, resilience has rapidly found itself at the top of the global development agenda [5]. It is the case, however, that our shortcoming to comprehend the complexities of our systems leads to our models of, and approaches to, systems resilience to persistently demonstrate linear, fragmented and dispersed knowledge. Such limited knowledge prohibits the acquisition of a complete and coherent view, which only adds to the uncertainty problem. Our models do not demonstrate the critical links and interdependencies that mesh our world and life systems. Our approaches are intimidated by the task of elucidating our hyper-connected systems. These lead to our shallow understanding of the nature of systems complexity.

Hence, we argue in this paper for a systems intelligence that has complexity-based thinking at its foundation.

 ISBN: 978-1-61208-451-0

Primarily, complexity-based thinking accepts the notion that our knowledge can only be limited and bounded. But instead of surrendering to this shortcoming, complexity-based thinking seeks new ways to redefine the limitation threshold and further enlarge the boundary of knowledge. Complexity-based thinking involves methodological pluralism, law of requisite knowledge, and complexity absorption. To embody this thinking, our system acquires (i.e., represents and infers) and integrates knowledge from heterogeneous sources, namely, massive information data points, expert and experiential knowledge, and perceptions of human sensors. With the synergism of diverse knowledge, the system opens itself for the emergence of new intelligence in the face of daunting complexities. The integrated knowledge is aimed to expose hidden assumptions, reconcile conflicts and inconsistencies, and elucidate ambiguities in complex system behavior. More importantly, the integrated knowledge may be used to sense, make sense of, and shape the course of impending, on-going or ensuing system vulnerabilities, destructive perturbations, and critical systemic changes.

Our paper is structured as follows. In Section II, we argue for resilience thinking to be complexity-based thinking, and elucidate in Section III the framework of our multi-dimensional intelligent system that embodies complexity-based thinking. We then discuss in Section IV the broad impact of the system's integrated knowledge. Finally, we conclude in Section V.

II. RESILIENCE SHOULD BE COMPLEXITY-BASED

Gilpin and Murphy [6], while citing Richardson and Cilliers [7], explained that in the midst of numerous schools of thought within the complexity sciences, three broad approaches have emerged, namely, reductionist, soft, and complexity-based. They differentiated these three as follows. Reductionist complexity science uses a limited number of universal laws to characterize natural reality. Its problem, however, is that in emphasizing universal commonalities, it hides in the abstraction the individual system idiosyncrasies. Second, soft complexity science distinguishes between social reality and the natural world, and therefore rejects the application of complexity, which is a theory that originated in nature, to the social realm unless done metaphorically. However, it has been known that society exhibits complexity [8]-[10], and that the social and the natural are linked [11][13]. Complexity-based thinking, however, distances itself from any pursuit of exact knowledge or universal absolutes and deals with the fact that our knowledge is bounded, and we can only seek this boundary in whatever way possible and suitable [6].

We believe that resilience thinking should be complexity-based thinking. Complexity-based thinking, as pointed out by Gilpin and Murphy [6], adheres to the notion that "complex matters demand a methodological pluralism" [9, p.12]. Because knowledge can only be partial and bounded, pluralism provides several elucidations of a phenomenon. This roots back to Ashby's law of requisite variety, which states that by having diverse response mechanisms available to the system, the system is able to compensate a larger variety of perturbations [13]. Gilpin and Murphy also echoed Cooksey's compelling point, i.e., complexity-based thinking seeks "diverse avenues for discovering what may end up being a multiplicity of answers that are differentially sensitive to and grounded in specific circumstances, conditions, people, times, and places" [14, p.84]. Similarly, in resilience thinking, by making available to the system a diversity of mechanisms to sense, make sense of, and respond to situations, the system is able to be resilient to a larger variety of shocks and stresses.

We believe that complexity-based thinking also includes the law of requisite knowledge, which states that a system must not only be dependent on a variety of available response mechanisms but must also know which one to select and how [15]. Otherwise, the system would have to try out actions blindly, which would consequently compromise its resilience. Therefore, in managing the resilience of a system against shocks and stresses, increasing the variety of its actions must be accompanied by the increase in the capacity to choose the best action.

Lastly, complexity-based thinking involves complexity absorption. As explained by Gilpin and Murphy [6], complexity-based thinking not only prefers the diversity of options, as well as tolerance for their possibly conflicting representations, there is also the ability to adapt, as well as self-organize, as fresh knowledge is captured, generated or re-created in order to modify an existing goal or drop it for a new goal. In other words, there is a paradigm shift from fragmented, individual and controlling views to a complex adaptive system view that enables capture, creation and refinement of knowledge.

III. SYSTEMS INTELLIGENCE FRAMEWORK

We detail in this section the various aspects of our intelligent system, as shown in Figure 1, with complexity-based thinking at its foundation. We then relate these aspects in the succeeding section to the processes of sensing, making sense, and shaping system-related complexities.

First, we need to distinguish between data, information, knowledge, and intelligence (which in our case refers to artificial intelligence) [6][16][17]. Data is a stream of facts about entities and events that are situated in a sequence. We can imagine everything in the world as represented by data points. We are surrounded by a massive amount of data that is "getting ever vaster ever more rapidly" [18, p.1]. What we have is a digital universe that is huge, and continuously increasing exponentially [19]. Information consists of data related to a given context [20]. It is descriptive as it demonstrates patterns that have sense [21]. Knowledge is the collection of information that has been proven useful in a given context [21]. Knowledge involves the appraisal of the information of its relative significance [20] using one's experience, values, insights and expertise [22][6]. Data, information, and knowledge lie along this continuum [20][6].

Intelligence, however, is a process. It involves learning and applying knowledge in order to respond to changing contexts. Our system is intelligent since it performs the acquisition, integration, and application of knowledge. Furthermore, as new data and information come in, the system's knowledge should improve based on new evidence,

 ISBN: 978-1-61208-451-0

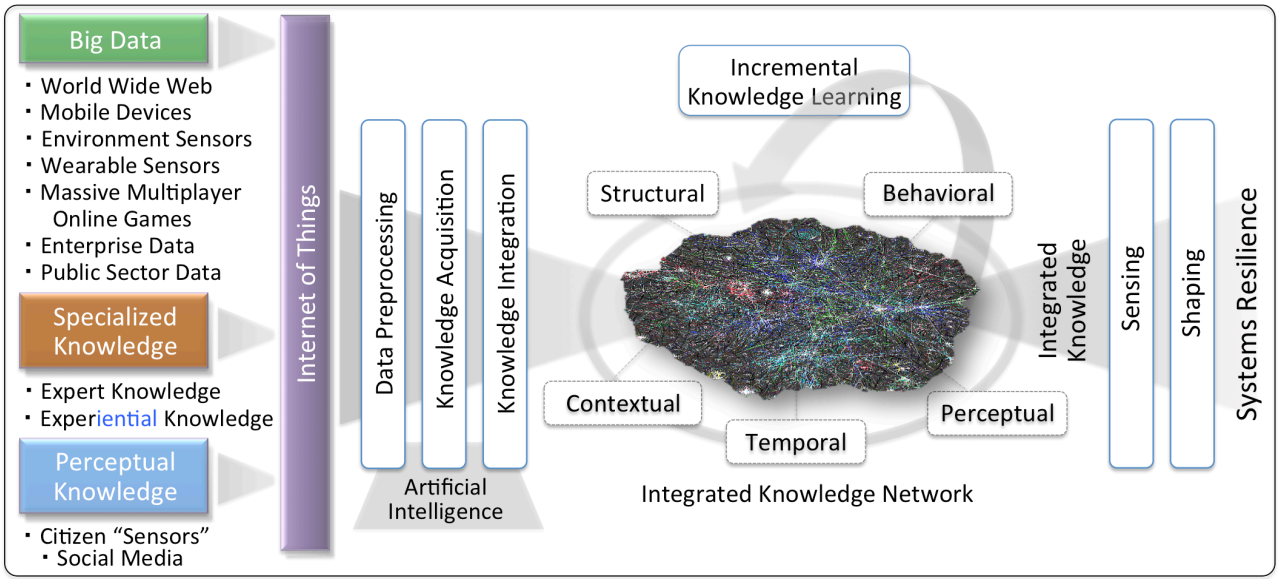

Figure 1. Complexity-based systems intelligence framework for systems resilience

hence, it learns new knowledge and does it incrementally. This makes our system suited for complexity – as Friedrich Engels wrote, "that the world is not to be comprehended as a complex of ready-made *things*, but as a complex of *processes* in which the things… go through an uninterrupted change of coming into being and passing away." [23, p. 44, italics in source]. Our system is also intelligent since it involves sensing (acquiring), making sense (integrating) of, and shaping (applying knowledge) itself or its situation. Sensing involves physical and social sensors obtaining data and information. The process of making sense involves predictive modeling, situation analysis and awareness, anticipation, as well as providing actionable information. Shaping is influencing and changing the course of the situation and the way the system adapts to its environment.

A. Complexity Thinking Is Methodological Pluralism

In light of the above, our goal is to realize an intelligent system that can infer and integrate heterogeneous knowledge from three varied sources, namely, massive mostly non-structured data or Big Data, expert and experiential knowledge, and public perceptual knowledge sourced by citizen sensors. While the last two are contextualized knowledge, the intended knowledge has yet to be derived from the first. The intelligent system must cull through massive data points, theoretical and empirical evidences, and human perceptions that are made available through the Internet of Things (IoT) where data and technology are democratized, i.e., available to as many people as possible.

IoT extends Internet connectivity beyond desktop and mobile computers to a diverse range of sensors (e.g., physical ambient sensors that include UAVs and satellites), machines (e.g., from coffee makers and washing machines to jet engine components), and devices (e.g., mobile and wearable devices) that we can think of that communicate and interact with the external environment and to each other. According to Intel, the IoT world is growing in a very fast pace, i.e., from two billion in 2006, 15 billion in 2015, to a projected 200 billion by 2020! [24].

1) Big Data

The digital universe is ever expanding as millions of data points from diverse sources are created every second from heterogeneous sources. The World Wide Web is basically an open world where information of various kinds are sent, uploaded, downloaded, and received. Web contents are created and duplicated rapidly and continuously. Crawlers or scrapers can be written to extract data stored deep in the Web Our mobile devices have powerful computing sensors and software that allow us to collect data about our physiology and mobility. It also allows us to log our daily activities that include fitness routines, web searches, online transactions, and interactions in social media platforms and micro-blogging sites, among others. Furthermore, ubiquitous ambient sensors [25] can also offer a wide range of possibilities for gathering large volumes of data that are human-related (e.g., displacements of millions of refugees), environmental (e.g., earthquakes, tsunamis, climate and weather changes, and changing landscapes). There are also the massive multiplayer online games that have become unprecedented tools to create and validate theories and models of social and behavioral dynamics of individuals, groups, and networks within large communities [26]. Enterprises may collect billions of real-time data points on products, resources, services, and stakeholders, which can provide insights on collective perceptions and behaviors, as well as resource and service utilizations. There are also data that public bodies produce or collect, which include geographical information, statistics, environmental data, power and energy grids, health and education, water and sanitation, and transport. There are the systematically acquired and recorded census data about households and the

 ISBN: 978-1-61208-451-0

services (e.g., health and medical, education, water, waste disposal, electricity, evacuation, and daily living-related programs) made available to them.

2) Specialized Knowledge

Carpenter et al. [3] stated that we have the tendency to wrap our minds around the computable and ignore the non-computable aspects of systems complexity. To account for the non-computable, they admonished considering a wide range of perspectives. They clarified why our society should put value not only on expert models, since they also emphasize narrow, segregated and domain-dependent views, but also look at instances where perceptions of experience-filled individuals, although lacked in formal education, led to breakthroughs. They cited several cases: crucial information from hunters and loggers prompted new approaches that saved the giant jumping rat in Madagascar from their sudden demise, and information from indigenous fishermen saved endangered bumphead parrotfish.

Complex problems may have many solutions that may differ in the required execution to obtain the quality of the desired outcome [3]. Hence, a diverse team of experienced individuals is more suited than a team of expert solvers [27].

3) Perceptual Knowledge of Human Sensors

We adopt the definition of perception as the process in which we actively and purposefully [28] acquire, organize, recognize, and interpret the sensory information we receive in order to make sense of what we perceive [29][30]. Perception therefore allows us to take in all, or part of, the sensory information presented to us and transform them into something meaningful [31]. Perception also includes how we respond to the information we receive [31]. An input to our perception triggers in us a psychological (i.e., cognitive and affective) response, and on the basis of this response, we perform an action [32]. Our individual differences dictate the difference in our perceptions, which can explain why individuals and communities respond differently despite being presented with the same facts, conditions, support, assurance, resources, and other forms of stimuli [33].

Perceptual knowledge on resilience can be provided by, or obtained from, individuals who are physically on the ground, i.e., those who are directly experiencing the situation (e.g., war, refugee migration, pandemic, radiation leaks, etc.) such as the members of the affected community, local government, law enforcers, first responders, and disaster managers, and from those who are virtually on it, i.e., those who are not in the affected area but have a good view of the situation as seen on TV, Internet, and social media. Armed with their mobile devices and sensors, these individuals will relay all information they perceive, as well as add their own analyses, sentiments, and judgments as they deem necessary.

B. Law of Requisite Knowledge

1) Knowledge Acquisition

Our focus is on data that are made available publicly and online through the IoT, and accessed freely for the common good. For example, if satellite imagery credibly shows troops massing outside a village, the people of that village can be informed so as to flee, or if necessary, respond with violence to save their lives [34]. There have been advocacies for data and technology to be *democratized*, i.e., to be made available to as many people as possible, including the grassroots population, and not just to those who are inherently, or were made to assume, to be in a position to use them (e.g., governments, international organizations, multi-national companies, and Internet giants, among others) [34]. History has demonstrated how ordinary people achieved resilience when empowered by technology and big data [4][35]. Furthermore, having more data openly available will encourage public participation to achieve novel and innovative solutions to societal challenges.

From structured data stored in spreadsheets and relational databases, Internet-based technologies have allowed the collection of unstructured data. For example, the company Digital Reasoning (www.digitalreasoning.com) estimates the unstructured data created per day to be 2.5 quintillion. Unstructured data does not follow the traditional database field formats nor adheres to a formal data model. This includes video, audio, images, graphics, and sensor signals, as well as partially formatted or semi-structured data, including text-based documents (e.g., word processing, PDFs, emails, blogs, wikis, tweets, web pages, and web components) and documents with self-describing elements such as tags and markers (e.g., XML and HTML). While structured data is organized in a database format and are readily accessible by search algorithms, the irregularities and ambiguities in unstructured data make its representation, let alone its comprehension by the machine, difficult.

Expert knowledge, which is organized and scientifically validated, mostly appears in scholarly publications. Experiential knowledge, however, is commonly unpublished in the literature since they are hard to express, specify, and scientifically validate. Consider tacit knowledge, for example, which highly depends on personal character traits that it cannot be subject to accurate communication [6]. Social computing, however, can help design platforms for diverse problem solvers to articulate and collaborate their perceptions, and incorporate them in a repository of evidences of what may or may not work in a situation [33]. Similarly, perceptions of citizen sensors are unstructured knowledge that populates traditional and social media in the form of meaningful texts, videos, photos, and audio.

The difficulty with data acquired from various sources is that they tend to be heterogeneous in terms of their spatial and temporal aspects, data collection modalities, structure type (structured, semi-structured or unstructured), data type (hard physical data vs. soft data), and in sensor outputs with different resolutions and sampling rates. Data preprocessing should therefore be carried out, the result of which will be a set of features that can characterize the various entities of interest. This is certainly a non-trivial task. If the varied data are commensurate, then raw signal data can be easily combined (e.g., using Kalman filtering) [36]. If not, extracting the common features may involve further data transformation, such as filtering out noises and outliers, data alignment (remove any positional or sensing geometry and timing effects from the various data), common spatio-temporal referencing, and data association (determine which

 ISBN: 978-1-61208-451-0

object is associated to which event) [37]. Metadata may also be generated to describe the heterogeneous data [36].

Artificial intelligence is needed to transform data to knowledge. The first step is knowledge acquisition (and not information acquisition as it is only a means to this end), which involves a lot of processes to transition from bytes to usable patterns and meanings. We use the term "acquisition" to refer to various approaches to obtain knowledge from data. Knowledge representation [38] is one of the central and most important concepts in artificial intelligence that constructs formalisms (e.g., semantic nets, ontologies, systems architecture, logic rules) that will make complex systems easier to design and implement. It represents information about the world in a machine-understandable form that a computer system can utilize for complex problem-solving tasks. Knowledge inference [38] refers to acquiring new knowledge from existing facts based on certain rules and constraints that are made understandable through knowledge representation. Automated reasoning (e.g., inference engines, theorem provers, and classifiers) is the mechanism behind inferring new knowledge by applying existing facts and logic rules [39]. A similar concept is knowledge discovery, which was borne out of data mining and describes the process of automatically searching through large volumes of data for patterns that can be considered knowledge.

2) Knowledge Integration

Once knowledge is inferred from these varied sources, the next step is to weave them together. *Knowledge integration* will involve inferring knowledge relationships among hugely varying domains into a coherent structure while revealing hidden assumptions and reconciling areas of conflicts, inconsistencies, and uncertainties. It should describe how domain-specific concepts are interrelated for transdisciplinary problem and solution formulation. It must be able to synthesize micro-level, individualized and domain-dependent knowledge to contextual systemic knowledge. The integration of knowledge from diverse sources should not lead to vague generalities, but rather to become effective in enhancing knowledge. This task is difficult and remains an open research area.

Knowledge integration involves weaving the diverse knowledge into coherent networks, hence, a *knowledge network*. Paperin et al. [40] provide an excellent survey of previous works that demonstrated how complex systems are isomorphic to networks and how many complex properties emerge from network structure rather than from individual constituents. Representing the integrated knowledge into coherent networks can be accomplished by using network and dynamic graphs theories and models.

Another aspect of knowledge integration is to incorporate new incoming knowledge into existing prior knowledge [41]. This allows the body of knowledge to grow incrementally. This is also a non-trivial task as new and existing knowledge may interact in unanticipated ways that may demand significant changes to the developing knowledge [41].

3) What Kind of Knowledge?

We aim for our system to infer knowledge that can be used to describe the complexity of our hyper-connected systems, the endogenous and exogenous forces that influence them, and their interactions. We believe that the Five Aspects Taxonomy [42] provides a good coverage of the essential aspects of the complexity, which our intelligent system needs to be knowledgeable of.

The taxonomy is conceived for the engineering of socio-technical systems that exhibit complexities in multiple levels (i.e., components, subsystems, systems, and system of composite systems) and dimensions (aspects). The five aspects include: (a) *structural* – elaborate hierarchical or layered network arrangement of system components that demonstrate couplings and dependencies in multiple scales; (b) *behavioral* – variances in system responses to different stimuli; (c) *contextual* – environmental circumstances in which the system exists; (d) *temporal* – various system properties, dimensions and needs may change over time together with the dynamic environment in which it exists; and (e) *perceptual* – stakeholder perceptions of the system and the environment that embeds it, which may change with context shifts and cognitive constraints and biases.

We can imagine the knowledge network as consisting of nodes that represent entities, such as system components (and subcomponents), as well as endogenous and exogenous factors. The edges depict their interactions as per the Five Aspect Taxonomy.

C. Complexity Absorption

As per above, knowledge integration can be defined as the process of incorporating new information into existing knowledge, which may require modifying the existing knowledge to accommodate the new knowledge *and*/*or* modifying the new in light of the existing knowledge [41][43]. Integration of knowledge from various knowledge sources can result in novel knowledge on how to solve a problem. Knowledge integration can also help unmask the uncertainty created by the multiple sources of knowledge.

The argument for knowledge integration is also present in resilience as evidenced by Bohensky and Maru [43]. Complexity and uncertainty management in socio-ecological systems can benefit from integrating diverse types of knowledge. Also, collaborations that facilitate integration of diverse perceptions lead to socio-ecological resilience [3].

Our intelligent system is aimed to integrate and preserve heterogeneous knowledge for triangulating for the truth, continuously track incoming and on-going information as well as evolving circumstances and conditions, and aid the system to better self-organize as it acquires new knowledge, adapts with new functions, and transforms to new goals. New facts should be continuously derived, and incoming evidence should be used, to improve current knowledge repositories. Hence, knowledge will be learned incrementally by our system. Our system, with its synergistic integration of knowledge, may lead to the emergence of increasing intelligence in the midst of complexity.

 ISBN: 978-1-61208-451-0

IV. SENSE, MAKE SENSE OF, AND SHAPE COMPLEXITY

With the network of connected and evolving knowledge about systems and their complexities in terms of structure, behavior and context, and how they are perceived, to be changing over time as derived from heterogeneous sources, what then can we use this knowledge for? Again, to sense, make sense of, and shape system behaviors when faced with complexities that can threaten the existence of the system.

Sensing involves detection, whereas making sense involves recognition and identification. Detection is similar to when an alarm goes off at home; we know that something has occurred but we may not recognize what it is. Recognition happens when the alarm is matched with a known reference, already known pattern, or learned category, e.g., the alarm matches either a gas leak or possible burglary. Identification is when we go down to the kitchen and finds out that the gas leak indicator is also blinking.

There were cases, however, that the alarm did not go off when it should. Indicators leading to 9/11 were pointing to an imminent large-scale attack on US soil [4]. The alarm did not go off because the consequence then of these indicators were not clear, the US intelligence community had only fragments, and there was no actionable information that was presented [4]. The increase of the price of tortilla by 400% easily set off the alarm of a food riot in the streets of Mexico, but no one could have predicted that Hurricane Katrina indirectly, but significantly, caused it [2]. Other examples are provided by Robertson and Olson [34]: When the social web during Iran's postelection crisis in 2009 was datamined, shifting perceptions in terms of awareness, advocacy, organization, mobilization, and eventual action and reaction were unmasked; the data visualization at that time of the Iranian blogosphere revealed a dramatic increase of user population with religious orientation; and the examination of microblogs related to Arab Spring revealed that socio-economic terms (e.g., housing, income and minimum salary) were most relevant in 2010, but in 2011, tweets were related to corruption, revolution, and freedom. The alarm did not go off in these instances because it was not set to detect the proxy indicators of upcoming dramatic system changes and there was no inter-related set of knowledge regarding a multiplicity of factors.

How then can the knowledge network be used for sensing and making sense? The following can only be possible with the connected multi-dimensional knowledge present in the integrated knowledge network:

- *Describe the emergence of system-wide properties*. In the sciences and arts, emergence is a process whereby systemic entities, patterns, and regularities arise through interactions among smaller or simpler entities that themselves do not exhibit such properties. Individual idiosyncrasies get lost as the components become tightly coupled and dependent. However, since our system will track the historical transitions of bytes to integrated knowledge, we theorize that it can also describe the evolution and emergence of system properties.
- *Perform anomaly detection*. After using intelligent algorithms to determine systemic "habitual" (i.e., normal, routinely behavior) patterns, anomaly detection techniques can then be used to detect what is out of the normal, which can include proxy indicators or digital alarms of upcoming changes [34].
- *Resolve conflicting information from same or different sources*. When platforms are made open for humans (and intelligent artifacts alike) to contribute information, it is possible that conflicting information are received due to cognitive biases, perceptual errors, or communication differences. For example, social media can generate a ton of interests in seconds, but can also warp and disperse the true information (at times intentional, e.g., tampering data, spreading rumors) into thousand fragmented pieces. With our methodological pluralism, it is possible to perform multi-dimensional corrections and validations to eliminate the false positives.
- *Perform "unsaid-knowledge" analytics*. We introduce this term to refer conceptually to mining for knowledge that cannot be explicitly stated since it depicts intuition, common sense, wisdom, and culture-based assumptions – those that are hard to quantify and measure but have proved essential to identifying anomalies. It also includes tacit knowledge that is abstracted by our understanding and difficult, if not impossible, to codify or transmit [6].
- *Perform descriptive and predictive analyses* [44]. Descriptive analysis is to mine past knowledge that are connected and related physically, semantically and conceptually to explain what has already happened and why it happened. Predictive analysis is forecasting future outcomes across various scenarios or situations.

Sensing technologies can provide support, but, unlike shaping, they do not necessarily change the system state. Early warning systems, for example, can help people get out of the way of an inevitable disaster, whether or not they change the course of events. The knowledge network is aimed to provide actionable strategies that will convert sensing to shaping, which can be achieved as follows:

- *Perform prescriptive analytics*. The aim of prescriptive analysis is to identify which decision and response will lead to the optimal or most effective result given a specific set of objectives and constraints [44]. The challenge for a truly strong prescriptive capacity is great. One that is formidable, for example, is determining the optimal path to the desirable regime where the potential paths are possibly thousands, each with its own set of multiple candidate divergences. Without the algorithms to find the optimal path efficiently, the required computing resources can become detrimental [1].
- *Guide the planning and implementation of a creative chaos* [45]. The idea is to use the knowledge network to simulate system shocks that can propel the system into the vortex of change. It is efficient and effective to scan for situations that can force latent problems to surface than design the system to not fail, which, paradoxically, only makes the system more vulnerable and less resilient. Furthermore, by introducing chaos into the system, not only do we prepare the system to be adaptive to failures, but also to bring out opportunities for innovation since

 ISBN: 978-1-61208-451-0

chaos would break tight couplings only to give way to new and previously unknown effective connections. Incidentally, the Five Aspect Taxonomy is a frame to comprehend facets of innovation strategies and communicate emerging technologies [42].

- *Infer a theory of lever point* [46]. A lever point is known as that critical point within the system where applying a little change can make a big difference and a small shift a big change. At that point the behavior of the complex system changes fundamentally.
- *Infer theories of system openness and modularity and their trade-offs* [47]. Modularity can help contain ensuing disasters by compartmentalizing. However, too much compartmentalization can prevent aid from moving in and out of the system from various sources. Also, too much openness can transmit harmful shocks, as in financial collapse, pandemics, and invasive species migrating easily across similar and connected landscapes.
- *Perform complexity mapping*. Provide real-time mapping of the events and feedback loops occurring during complex situations. The ability to monitor the behaviors of social, physical, environmental and technological systems in real time make it possible to understand where models, plans, policies and programs are failing and to make necessary adaptations.

V. CONCLUSION

Resilience can be enhanced and sustained by integrating, rather than fragmenting and dispersing, knowledge about the complexities of our systems. Mastering a more holistic understanding of our systems will shed light on their couplings, temporal and spatial boundaries, interaction behaviours, and emerging irregularities or inaccuracies, as well as proven or plausible alternative resilient strategies.

We argued that the fundamental difficulty in managing resilience is the complexity that characterizes our systems. We then argued for managing resilience by realizing an intelligent system whose intelligence framework is based on complexity thinking. The result is an integrated knowledge of systems complexity, which is automatically inferred from heterogeneous data about the nature and contextual interaction behaviours of our systems. With this integrated knowledge realized, our systems can better meet head-on the so-called unknown unknowns or uncertain uncertainties.

REFERENCES

[1] R. Legaspi and H. Maruyama, “Meta-theory and machine-intelligent modeling of systemic changes for the resilience of a complex system,” in Proc. Tenth International Conference on Systems (ICONS 2015), IARIA, L. Koszalka and P. Lorenz (eds.), ThinkMind Digital Library, 2015, pp. 102-111.

[2] A. Zolli and A. M. Healy, Resilience: Why Things Bounce Back. New York, NY: Free Press, July 2012.

[3] S. R. Carpenter, C. Folke, M. Scheffer, and F. Westley, “Resilience: Accounting for the noncomputable,” Ecology and Society, vol. 14, no. 1, article 13, 2009.

[4] J. McCracken, “In the shadow of 9/11,” Rare Events: Can We Model the Unforeseen? Sigma, vol. 10, no. 1, T. B. Fowler and M. J. Fischer (eds.), Noblis, 2010, pp. 30-35.

[5] L. Jones and T. Tanner, “Measuring 'subjective resilience': Using people's perceptions to quantify household resilience,” Working paper 423, Overseas Development Institute, July 2015. Available online via ODI: http://www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/9753.pdf (accessed on 2016.01.11).

[6] D. R. Gilpin and P. J. Murphy, Crisis Management in a Complex World. Oxford, NY: Oxford University Press, 2008.

[7] K. A. Richardson and P. Cilliers, “What is complexity science? A view from different directions,” Emergence, vol. 3, no. 1, 2001, pp. 5-22.

[8] J. A. Tainter, The Collapse of Complex Soceities: New Studies in Archaeology. Cambride, UK: Cambride University Press, 1988.

[9] P. J. Richerson and R. Boyd, “Institutional evolution in the holocene: The rise of complex societies,” in The Origin of Human Social Institutions, W. G. Runciman (ed.), Proceedings of the British Academy, vol. 110, 2001, pp. 197-204.

[10] P. Ball, Why Society is a Complex Matter: Meeting Twenty-first Century Challenges with a New Kind of Science. Berlin Heidelberg, Springer-Verlag, 2012.

[11] A. Jentsch, H. Wittmer, K. Jax, I. Ring, and K. Henle, “Biodiversity: Emerging issues for linking natural and social sciences,” GAIA - Ecological Perspectives for Science and Society, vol. 12, no. 2, June 2003, pp. 121-128.

[12] P. Dickens, “Linking the social and natural sciences: Is capital modifying human biology in its own image?” Sociology, vol. 35, no. 1, February 2001, pp 93-110.

[13] R. W. Ashby, An Introduction to Cybernetics. London: Methuen, 1956.

[14] R. W. Cooksey, “What is complexity science? A contextually grounded tapestry of systematic dynamism, paradigm diversity, theoretical electicism,” Emergence, vol. 3, no. 1, 2001, pp. 77-103.

[15] F. Heylighen, “Principles of systems and cybernetics: An evolutionary perspective,” in Cybernetics and System ’92, R. Trappl (ed.), World Science, Singapore, 1992, pp. 3-10.

[16] T. R. Kuhn, “Knowledge and knowing in organizational communication,” in The SAGE Handbook of Organizational Organization: Advances in Theory, Research, and Methods, Third Edition, Section IV, L. L. Putnam and D. K. Mumby, and K. J. Krone (eds.), SAGE, 2014, pp. 481-502.

[17] A. G. J. MacFarlane, “Information, knowledge and the future of machines,” Philosophical Transactions of the Royal Society A, vol. 361, 2003, pp. 1581-1616.

[18] The Economist, “Data, data, everywhere: A special report on managing information,” February 27, 2010. Available online via EMC: https://www.emc.com/collateral/analyst-reports/ar-the-economist-data-data-everywhere.pdf (accessed on 2016.01.11).

[19] V. Turner, D. Reinsel, J. F. Gantz, and S. Minton, “The digital universe of opportunities: Rich data and the increasing value of the Internet of Things,” IDC iView, April 2014. Available online via IDC: http://idcdocserv.com/1678 (accessed on 2016.01.11).

[20] H. Tsoukas and E. Vladimirou, “What is organizational knowledge?” Journal of Management Studies, vol. 38, no. 7, November 2001, pp. 973-993.

[21] A. MacFarlane, “Information, knowledge and intelligence,” Philosophy Now, Oct/Nov 2015. Available online: https://philosophynow.org/issues/98/Information_Knowledge_and_Intelligence (accessed on 2016.01.11).

[22] T. H. Davenport and L. Prusak, Working Knowledge: How Organizations Manage What They Know. Boston, MA: Harvard Business School Press, 1998.

 ISBN: 978-1-61208-451-0

[23] F. Engels, Ludwig Feuerbach and the End of Classical German Philosophy, 1946. First published in 1886 in Die Neue Seit, vol. 4 and 5.

[24] Intel, A Guide to the Internet of Things Infographic. Available online via Intel: http://www.intel.com/content/dam/www/public/us/en/images/iot/guide-to-iot-infographic.png (accessed on 2016.01.11).

[25] S. Poslad, Ubiquitous Computing: Smart Devices, Environments and Interactions. Wiley, 2009.

[26] K. J. Shim, N. Pathak, M. A. Ahmad, C. DeLong, Z. Borbora, A. Mahapatra, and J. Srivastava, "Analyzing human behaviour from multiplayer online game logs: A knowledge discovery approach," in Trends and Discoveries vol. 26, no. 01, H. Chen and Y. Zhang (eds.), 2011, pp. 85-89.

[27] S. E. Page, Diversity and Complexity. Princeton, NJ: Princeton University Press, 2011.

[28] P. J. Lang, "The varieties of emotional experience: A meditation on James–Lange theory," Psychological Review, vol. 101, no. 2, 1994, pp. 211-221.

[29] W. H. Ittelson, H. M. Proshansky, L. G. Rivlin, and G. H. Winkel, An Introduction to Environmental Psychology. New York: Holt, Rinehart and Winston, 1974.

[30] D. L. Schacter, D. T. Gilbert, and D. M. Wegner, Psychology (2nd Edition). New York: Worth, 2011.

[31] T. Gärling and R. G. Golledge, "Environmental perception and cognition," in Advances in Environment, Behavior, and Design, vol. 2, E. H. Zube and G. T. Moore (eds.), 1989, pp 203-236.

[32] R. G. Golledge, "Environmental cognition," in D. Stokols, I. Altman, (eds.) Handbook of Environmental Psychology. New York: Wiley, 1987.

[33] R. Legaspi, H. Maruyama, R. Nararatwong, and H. Okada, "Perception-based Resilience: Accounting for the impact of human perception on resilience thinking," Proc. 2014 IEEE Fourth International Conference on Big Data and Cloud Computing, December 2014, pp. 547-554.

[34] A. Robertson and S. Olson, Sensing and Shaping Emerging Conflicts: Report of a Workshop by the National Academy of Engineering and United States Institute of Peace Roundtable on Technology, Science, and Peacebulding. Washington, DC: The National Academies Press, 2013. Available via USIP: http://www.usip.org/sites/default/files/Sensing_And_Shaping_Emerging_Conflicts.pdf (accessed on 2016.01.11).

[35] P. Meier, "How to create resilience through big data," in iRevolutions: From Innovation to Revolution. Available online: http://irevolution.net/2013/01/11/disaster-resilience-2-0/ (accessed on 2016.01.11).

[36] D. L. Hall and J. M. Jordan, Human-Centered Information Fusion. Norwood, MA: Artech House, 2010.

[37] J. Roy, "From data fusion to situation analysis," Proc. 5th International Conference on Information Fusion, 2001.

[38] R. Davis, H. Shrobe, and P. Szolovits, "What is knowledge representation?" AI Magazine, vol. 14, no. 1, 1993, pp. 17-33.

[39] L. Tari, "Knowledge inference," in Encyclopedia of Systems Biology, W. Dubitzky, O. Wolkenhauer, K.-H. Cho, and H. Yokota (eds.), 2013, pp. 1074-1078.

[40] G. Paperin, D. G. Green, and S. Sadedin, "Dual-phase evolution in complex adaptive systems," Journal of the The Royal Society Interface, vol. 8, no. 58, 2011, pp. 609-629.

[41] K. Murray, Learning as Knowledge Integration, Ph.D. Dissertation, Technical Report TR-95-41. Department of Computer Sciences, University of Texas, Austin, November 1995.

[42] D. H. Rhodes and A. M. Ross, "Shaping socio-technical system innovation strategies using a Five Aspects Taxonomy," 2010. Available online: http://seari.mit.edu/documents/preprints/RHODES_EUSEC10.pdf 2015.12.31.

[43] E. L. Bohensky and Y. Maru, "Indigenous knowledge, science, and resilience: What have we learned from a decade of international literature on "Integration"?" Ecology and Society, vol. 16, no. 4, article 6, 2011. Available online: http://www.ecologyandsociety.org/vol16/iss4/art6/ (accessed on 2016.01.11).

[44] Ernst and Young Global Limited, "Big Data: Changing the way businesses compete and operate – Insights on governance, risk and compliance," April 2014. Available via EY: http://www.ey.com/Publication/vwLUAssets/EY_-_Big_data:_changing_the_way_businesses_operate/$FILE/EY-Insights-on-GRC-Big-data.pdf (accessed on 2016.01.11).

[45] H. Maruyama, R. Legaspi, K. Minami, and Y. Yamagata, "General Resilience: Taxonomy and strategies," Proc. IEEE 2014 International Conference and Utility Exhibition on Green Energy for Sustainable Development (ICUE), IEEE Press, March 19-21 2014, pp. 1-8.

[46] J. H. Holland, Hidden Order: How Adaptation Builds Complexity. USA: Perseus Book, 1995.

[47] S. R. Carpenter et al., "General resilience to cope with extreme events," Sustainability, vol. 4, 2012, pp. 3248-3259.

 ISBN: 978-1-61208-451-0

On the Riding Aids Recognition System for Horse Riding Simulators

Sangseung Kang Kyekyung Kim Suyoung Chi
Intelligent Cognitive Technology Research Department,
Electronics and Telecommunications Research Institute
Daejeon, Korea
e-mail: {kss, kyekyung, chisy}@etri.re.kr

***Abstract*—Riding aids are instruction signals from a rider to a horse in the horse riding. In this paper, we present a study on the riding aids recognition system for horse riding simulators using multimodal sensing data. The riding aids recognition system provides an effective interactive function based on the riding intentions. The system has three types of sensors including vision, voice, and touch sensors. It has adopted certain schemes for multimodal data acquisition, feature extraction, data fusion, template matching, and intention classification. The system can provide recognition means for a riding aid so as to provide sense of reality such as actual horse riding to the user.**

Keywords-riding aids; riding intention; horse riding; sports simulator.

I. INTRODUCTION

The horse riding simulator is a machine that simulates the riding motion through an imitation of the basic movements of a real horse [1][2]. A current simulator is developed to move like an actual horse so that similar exercise effects to actual riding are provided to a user. The effects of exercise and recreational elements are also taken into account. However, it is not easy to provide continuous interest to users when operating a simulator. To solve this problem, it is necessary to develop hardware and software platform. Our approach infers the intention recognition using multimodal sensing data in a horse riding simulation environment.

In this paper, we propose a riding aids recognition system with multimodal sensing in horse riding simulator environments. The proposed system consists of sensor data acquisition modules, a feature extraction and fusion module, and a class generation and template matching module for intention classification. It is possible to increase the sense of the real for the user by enabling the horse riding simulator user to perform similar interaction to actual horse riding, and to increase effects of horse riding training using the horse riding simulator. Particularly, it is possible to provide an effective method for recognition of the intention of the rider in the horse riding simulation environment.

II. HORESE RIDING SIMULATOR

We developed a horse riding simulation system for replicating the mechanistic movements of realistic riding motions [3]. The interaction control module performs a controlling function for the interaction between the rider and several different functional components. The simulation hardware control module drives the movements of the hardware apparatus including horse body and the peripheral devices, and detects the movement of the rider using mounted multi-modal sensor devices. The riding content display module presents the relative contents on a large interactive screen that enables the user to actually ride the device, and returns feedback notifications generated within the content environment. The logging and coaching database manages the user's historical and specialized coaching information for personalized exercise mode and systematic training.

This paper is organized as follows. Section 2 presents our developed system related to the horse riding simulator. Section 3 describes the structure of the proposed riding aids recognition in the simulation environment and functional components of that. Conclusions are given in Section 4.

III. RIDING AIDS RECOGNITION

A riding aid which is an instruction signal from a rider to a horse in the horse riding means enabling the horse to

TABLE I. NATURAL RIDING AIDS

	Applying methods	Reactions
Seat	collection, balance, steering, forward movement	stop, go forward, turn
Leg	applying equal pressure against the horse's sides	increase in speed, upward transition
	one leg in a neutral position when applied	turn on the haunches or turn on the forehand, pirouette
	one leg farther back, with the other leg in a neutral position	actively encouraging the horse forward
Hand	direct rein (one rein pulls straight back)	turn in the direction of pressure
	indirect rein or bearing rein (pulls back inward)	correct straightness, lateral movements
	opening rein (does not pull back, moves his hands away)	turn in the air when jumping a fence
	neck rein (laying the rein)	turn a horse without bit contact
Voice	calming tone	slow down
	upbeat voice	move forward

 ISBN: 978-1-61208-451-0

TABLE II. ARTIFICIAL RIDING AIDS

	Applying methods	Reactions
Whip	training tool, using light taps	collect gaits or perform movements
Spur	brief, light touch, sharp jab	encourage more impulsion, go forward

recognize user intention [4][5]. Riding aids are the cues a rider gives to a horse to communicate what they want the animal to do [6][7]. Riding aids are broken into the natural aids and the artificial aids. The natural aids that the rider uses are the hand, leg, seat and voice aids, as shown in Table I. The artificial aids are used to backup natural aids using equipment such as a whip or spur, as shown in Table II.

Our horse riding simulation system has three types of sensors. The touch sensors detect movements of the rider. The touch sensors are mounted on the artificial body of the simulator. They include six photo interrupter sensors for bridle rein and two film-type pressure sensors for detection of a spur. The voice sensor detects voice intentions of the rider. It is possible to provide a speech-based control function for the simulator. The voice sensor is mounted on a riding helmet using a condenser microphone. The vision sensors detect postural intentions of the rider using two depth sensing devices. The vision sensors are installed in positions such as the left and the reverse side.

The data acquisition module uses multiple combinations of touch, voice, vision and other sensors to sense rider's commands or movements. The key data of user's movements are balanced sitting, drawing or pulling reins, spur, whip, and the like. The verbal commands are results of speech detection with minimum pulse length and endpoint detection. For the postural intentions, the module collects image and depth data for riding postures of the rider. Feature extraction and fusion module extracts each feature data from the relevant collected sensing information. It generates a feature data set through data fusion of the extracted feature information. The class generation module combines the extracted feature information from the feature data set, and generates a combination class depending on a predefined template type. The intention recognition module performed the recognition function based on the above generated class as the riding aids.

The riding action of the user is predefined and converted into an object model. The action is stored in the database as the riding aids class. The class includes the following:

- start - spurring (if stop state)
- acceleration - continuously spurring
- left turn - pulling a left portion of the bridle
- right turn - pulling a right portion of the bridle
- deceleration/stop - simultaneously pulling the bridle
- stop - pulling back an upper portion (while pulling the bridle)
- balancing - bending the upper portion forward (if walking)
- propulsive force increase - leg pressure
- exercise maintenance - leg release or bridle release

The classes include strength information expressed through the action of the user as a parameter. The parameter includes gait level, driving velocity, pulling degree of the bridle, spurring strength, acoustic score, and rider's joint degree. The template class matching function compares the generated combination class with the intention class stored in the intention database. Based on a result of the comparison, the system recognizes the intention of the action of the user. Using information on the recognized user action intention, the system sends the feedback, including changes for the gait level and driving velocity, and the simulation hardware control module controls the artificial body. We performed an experiment on the six classes such as start, acceleration, left turn, right turn, deceleration, and stop. In the initial experiment, the average recognition rate of ten subjects was around 90.5 percent overall.

IV. CONCLUSION

A substantial interactive process between human and horse riding simulator system must be provided for natural interaction. The process usually requires continuous riding aids recognition. In this paper, we present a riding aids recognition system using multimodal sensing data in horse riding simulator environments. The system can provide an effective interactive function based on the riding intention recognition.

ACKNOWLEDGMENT

This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIP) (10041627, Development of the 5-senses convergence sports simulator based on multi-axis motion platform)

REFERENCES

[1] G. Chen, et al., "Biofeedback Control of Horseback Riding Simulator," In Proceedings of the First International Conference on Machine Learning and Cybernetics, 2002, pp. 1905-1908.

[2] R. Eskola and H. Handroos, "Novel horseback riding simulator based on 6-DOF motion measurement, a motion base, and interactive control of gaits," Advanced Robotics, Vol. 27, No. 16, 2013, pp. 1249-1257.

[3] S. Kang, K. Kim, S. Chi, and J. Kim, "Interaction Control for Postural Correction on a Riding Simulation System," In Proceedings of the 9th ACM / IEEE International Conference on Human-Robot Interaction, HRI 2014, 2014, pp. 195-196.

[4] S. Qu, J. and Y. Chai, "Beyond Attention: The Role of Deictic Gesture in Intention Recognition in Multimodal Conversational Interfaces," In Proceedings of the 13th international conference on Intelligent user interfaces, IUI'08, 2008, pp. 237-246.

[5] K. A. Tahboub, "Intelligent Human-Machine Interaction Based on Dynamic Bayesian Networks Probabilistic Intention Recognition," Journal of Intelligent and Robotic Systems, Vol. 45, No. 1, 2006, pp. 31-52.

[6] Wikipedia, Riding aids, https://en.wikipedia.org/wiki/Riding_aids. [retrieved: October, 2015]

[7] A. M. Swinker, New Hampshire 4-H Horse Project Member's Manual, University of New Hampshire, 2004.

 ISBN: 978-1-61208-451-0

Safe Operation of Autonomous Machines

Raj Aggarwal
Adjunct Professor, ECpE
Iowa State University
Ames, Iowa, USA
rka@iastate.edu

Haoyuan Lin
Graduate Student, ECpE
Iowa State University
Ames, Iowa, USA
linhaoyuan@gmail.com

***Abstract*—Reliable detection of foreign objects is a key requirement for safe operation of autonomous machines. In farming applications, an object is considered a foreign object if it can damage the machine or be damaged by it. The traditional machine vision approaches rely on detecting and classifying each type as a separate class of thereby increasing the computational load and compounding the machine vision problem since these machines have to operate in real-time and the foreign objects can appear in any orientation thereby increasing complexity. Moreover, it is an over kill since the safe operation of an autonomous machine does not require that we classify these objects as long as we can reliably detect them and direct the machine to take appropriate maneuvering action. In our application, the object of interest is bales and everything else is a foreign object. The foreign objects include humans, animals, vehicles and standing crop. We use disparity information from the two cameras in a stereo configuration and use the camera model to calculate the distance to the object. This object detection framework based on distance and size has proved to be more efficient and robust compared to traditional machine vision approaches.**

Keywords-machine vision, autonomous operation, safe operation, stereo configuration, disparity

I. INTRODUCTION

Most object detection research focuses on how to design algorithms which are both accurate and fast and treat each type of object as a separate class [1, 2, 3]. However, in our application, the task is to detect foreign objects with arbitrary shape and size and reliably differentiate them from the bales. A straight forward method is to simply divide the foreign object into multiple categories and use traditional multi-class object detection & classification algorithms. However, this is not computationally efficient or robust because of infinite variations in size and shape. Moreover, the deformation of the object may also degrade the performance of this approach. Since our problem is to reliably detect in real-time and not classify, we used two cameras in a stereo configuration to generate a distance map and find blocks of certain size. The merit of this method is to consider a foreign object as a block and not care which category it belongs to as long as its size is within the range. We designed a detection framework and associated algorithms to detect foreign objects and their locations that are within certain volume at a given distance.

The rest of the paper is structured as follows. In Section II, we describe the framework and algorithm using depth map to detect foreign objects. In Section III, we present the experimental results. Finally, Section IV concludes the paper.

II. STEREO CONFIGURATION APPROACH

A. 3-D Reconstruction

The goal of stereo vision is mainly to recover the 3D structure of the scene using two or more images acquired by cameras in a stereo configuration. With known camera configuration, a disparity map can be generated by calculating disparities of all the pixels in an image. One method to calculate the disparity is using the feature matching [4, 5], such as edges. The edge features can be derived for both the left and right images by using Gaussian filters. The features are then matched by comparing their orientations and strength. In the disparity map, the value for each pixel is the distance between the pixels which has the highest match score.

The camera model is shown is Fig. 1 and its parameters are described in Table 1.

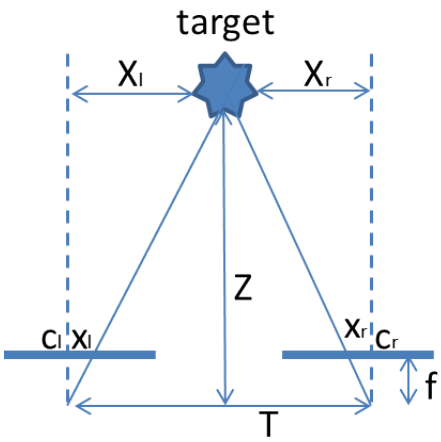

Fig. 1 The camera model

TABLE I. CAMERA MODEL PARAMETERS

Distance between the two cameras	T
Focus length of the cameras	f
Distance between target and camera	Z
Position of middle point of camera film	c
Position of object in camera file	x
Disparity of the target	d
Displacement between target and camera	X

Based on the triangulation principle, we have

$$\frac{x_l}{X_l} = \frac{f}{Z} \quad (1)$$

$$\frac{x_r}{X_r} = \frac{f}{Z} \quad (2)$$

This implies $X_l = \frac{Z}{f} x_l$ (3)

$$X_r = \frac{Z}{f} x_r \quad (4)$$

$$X_l + X_r = T \quad (5)$$

And $\frac{Z}{f}(x_l + x_r) = T$ (6)

Where $x_l + x_r$ is the disparity "d". So, the distance between the target and the camera is given by $Z = \frac{Tf}{d}$

B. Foreign Object Detection Framework

Disparity map is widely used in the computer vision applications to recover the 3D structure of the scene [6, 7]. The framework of the foreign object detection system using disparity map is shown is Fig. 2. The following sections will describe the flow chart in detail.

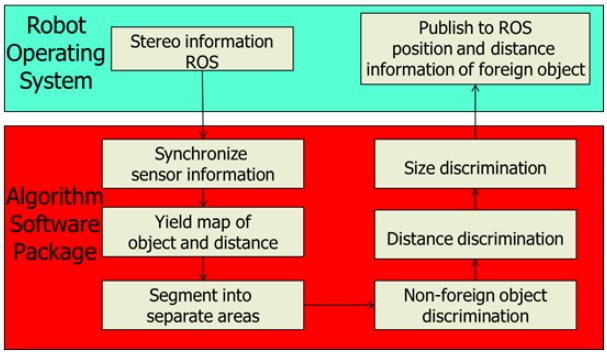

Fig. 2 The framework of the foreign object detection

The left camera and the right camera are of the same configuration and calibration. The pointing direction of both the cameras is the same and can use ground as a reference. The calibration step is done once offline before the system is used.

Since the distance similarity is the only feature used for detecting objects, the accuracy of the depth map [8, 9] is critical for the success of the algorithm. In order to make the distance calculation robust, the area of each frame is divided into sets of blocks instead of using each pixel. The edge feature within a block in the left frame can be used as the pattern to search in the right frame. For the purpose of the foreign object detection task instead of 3D construction, the depth calculation can be done coarsely. From our experiment, if the area of the block is too small, some holes will be shown is the depth map. Meanwhile, if the area of the block is too big, the object may not be detected, especially when the object is far away from the camera.

Fig. 3 shows the depth map when there is no foreign object using the 50*50 pixel block. In the figure below, the positive number means the distance between the background and camera in meters. The negative value means one of two things. One reason is that there is no matching block from the left camera frame to the right camera frame. A portion of the scene captured by the left camera may not be captured by the right camera. Since we use feature block in the left camera to find a match in the right camera, the distance along the left vertical line is negative. The second reason is the distance is too far and outside the range of interest. The disparity for such a block may be zero since the feature block is the same when watching from a long distance. The area in the sky is too far and the feature block looks the same. In either case, no special operation is required by the machine.

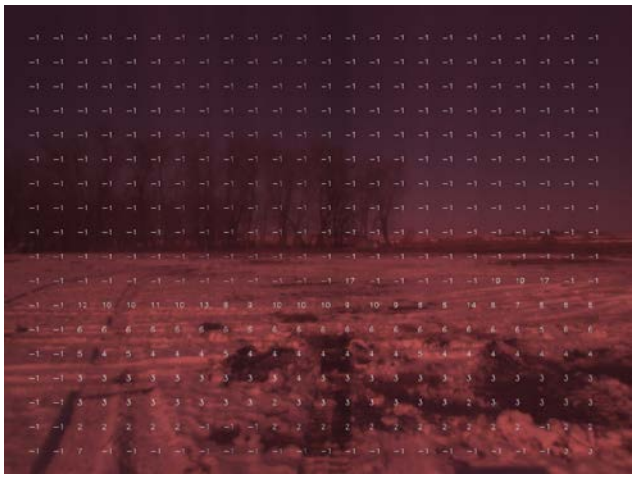

Fig. 3 The depth map when there is no foreign object

 ISBN: 978-1-61208-451-0

Once the depth map for the initial scene is generated, the system is ready for autonomous operation. The distance of the new frame is calculated and compared with the initial depth map. The distance filter us then used to separate the background object. Any feature block may be ignored if it is outside the critical range. A feature block is denoted if the distance is within the critical range. In our work, the ignored feature block is denoted as a negative sign and the blocks of interest are denoted with a positive sign, as shown in Fig. 4. After the processing of the filter, block fill algorithm is used to connect the neighbor blocks into one integrated block. For all blocks that are denoted by positive sign, the breadth first search algorithm with the neighbor rule is used to find all positive sign blocks and to mark these positions as a group. Each group represents one object. In our experiment setup, we use 8-neighbor rule to recognize the neighbor candidate around one block. We consider that the foreign object can be shown as any kind of shape; all 8 directions are considered as extension of the foreign object.

The size discrimination is used to create a decision table with the distance and object size information. When a small object is too close to the camera, the size of the pixel block is shown as a big block. By using the decision table we store the low-bound level of the size that has high confidence.

Fig. 4 shows the output of the filter based on the depth map when a person is walking in front of the camera.

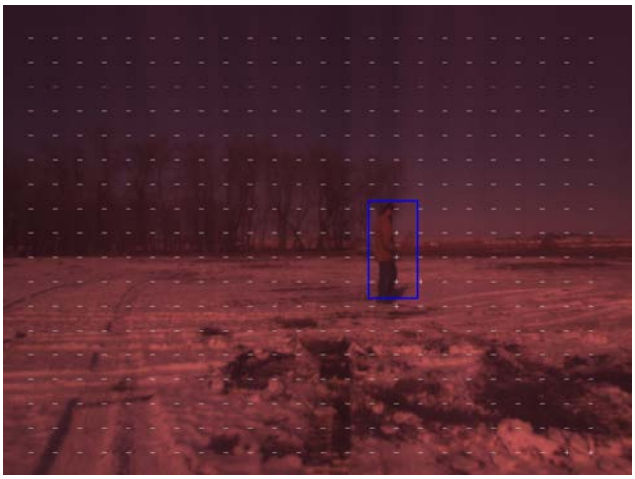

Fig. 4 The result after the block fill algorithm when one person is walking in front of the camera

III. EXPERIMENTAL RESULTS

In our experiment, we used nine data sets, shown in Table 2, to test the correctness of the algorithm. The data sets include two kinds of objects, a pedestrian and a vehicle. For each data set, the "# frames" means the number of frames in the video clips and the "# object" means the times the object appears in the video clip. We also note the moving direction of the object for the purpose of testing all corner cases.

The foreign object could move in any direction along a set route. It could move towards or away from the camera. It could move from left to right or right to left. The moving speed was slow to normal. The pose of the objects was changed from erect to leaning.

We compared our stereo-based algorithm based on the depth information with the previous work using the multi-classifier method based on the shape information [10]. From the results, the detection rate improved over the shape-based method. Besides that, the false alarm also decreased over the shape-based method. Sometimes, the foreground and background may mix together to make the frame area appear as a target of interest. However, the distance to the object is not always in the range. Such cases can be eliminated by using the depth information.

TABLE II. EXPERIMENT RESULT COMPARISON BETWEEN STEREO-BASED AND SHAPE-BASED METHOD

Datasets	**Stereo-based**			**Shape-based**		
	Detected	Miss	False alarm	Accuracy	Miss	False alarm
Dataset1 (42objects/ 116frames)	40	2	0	39	3	2
Dataset2 (15objects/ 34frames)	14	1	0	13	2	0
Dataset3 (12objects/ 56frames)	11	1	0	10	2	0
Dataset4 (22objects/ 60frames)	21	1	1	21	1	2
Dataset5 (12objects/ 46frames)	11	1	0	9	3	2
Dataset6 (27objects/ 51frames)	24	3	0	24	3	1
Dataset7 (8objects / 35frames)	7	1	0	6	2	0
Dataset8 (18objects/ 38frames)	13	5	0	7	11	1
Dataset9 (8objects / 37frames)	8	0	0	7	1	0

 ISBN: 978-1-61208-451-0

IV. CONCLUSION

In this paper, we describe a computer vision approach which is robust and efficient in detecting foreign objects with no pre-set shape, essential for safe operation of autonomous machines. We use a stereo configuration to generate a depth map. We then use a stereo matching algorithm to get the disparity information based on intensity images from stereo cameras and using the camera model to retrieve the distance information. From the result of our experiments, the proposed framework has a better performance with higher detection rate with lower false alarm.

The algorithm performed very well at short ranges (<10 meters); however, not as well object of interest is further away. One could investigate ways to improve the range accuracy by more rigorous modeling. The target classification accuracy can also be improved by incorporating shape information. The object tracking algorithm can also be improved by sequential frame processing.

References

[1] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, volume 1, pp 886–893, 2005.

[2] P. Dollar, C. Wojek, B. Schiele and P. Perona, "Pedestrian detection: An evaluation of the state of the art," IEEE Transactions on Pattern Analysis and Machine Intelligence, 34(4):pp 743–761, 2012.

[3] S.Tang, M. Andriluka and B. Schiele, "Detection and tracking of occluded people," International Journal of Computer Vision, pp 1–12, 2012.

[4] T.D. Sanger, "Stereo disparity computation using Gabor filters", Biological Cybernetics, vol 59, no. 6, pp 405-418, Oct. 1988, doi: 10.1007/BF00336114

[5] P. Kauff, N. Atzpadin, C. Fehn, M. Muler, O. Schreer, A. Smolic and R. Tanger, "Depth map creation and image based rendering for advanced 3DTV services providing interoperability and scalability", Signal Processing: Image Communication, Vol 22, No. 2, pp 217-234, Feb. 2007, doi: 10.1016/j.image.2006.11.013

[6] S. Izadi, et al., " Kinectfusion: real-time 3d reconstruction and interaction using a moving depth camera", Proc. of the 24th Annual Symposium on User Interface Software and Technology, pp 559-568, ACM 2011.

[7] M. Jenkin, A. D. Jepson and J. K. Tsotsos. "Techniques for disparity measurement." CVGIP: Image Understanding 53, no. 1 (1991): 14-30.

[8] S. Battiato, A. Capra, S. Curti, and Marco La Cascia. "3D stereoscopic image pairs by depth-map generation,". Proc. 2nd International Symposium on Data Processing, Visualization and Transmission, pp. 124-131. IEEE, 2004.

[9] K. Ikeuchi, "Constructing a depth map from images." In AI Memo AIM-744, Artificial Intelligence Laboratory, MIT. 1983.

[10] H. Lin, "Foreign object detection (FOD) using multi-class classifier with single camera vs. distance map with stereo configuration." (2015), Graduate thsis, Paper 14565.

 ISBN: 978-1-61208-451-0

Can I Let You In?
On Event-Centric Context-Aware Authorization

Philip Attfield*, Paul Chenard*, Marta Piekarska†, Julia Narvaez*, Mike Hendrick* and Simon Curry*

*Sequitur Labs
33404 Redmond Fall City Rd. SE,
Fall City, WA, 98024, USA,
E-mail: <name>.<last_name>@seqlabs.com

†Technische Universität Berlin
Security In Telecommunications,
Ernst-Reuter-Platz 7, 10587 Berlin, Germany,
E-mail: marta@sec.t-labs.tu-berlin.de

Abstract*—Current mechanisms for control and protection of computing resources were conceived decades ago. At that time constraints on power management, connectivity and the types of computing assets were far simpler. Today's mobile and distributed information systems are vulnerable to much wider and sophisticated threats. Thus, they require more flexible, extensive and powerful policy-based protection. This paper contributes a framework for policy-based authorization and illustrates its implementation. Details describing the architecture, methodology and tool flow for reliable synthesis of custom policy-based authorization are presented. The hypothesis is that access control applicable to a variety of devices should be event-centric and context-driven. The integrity and security of the authorization systems as well as the end-to-end trust that is guaranteed in the process used to create them are discussed. The applicability of the solution and its ability to mitigate the threats are discussed. A wide range of systems from simple to complex, including the emerging Internet of Things is covered. *(Abstract)

Keywords–Mobile Operating Systems, Access Control, Mobile Device Management, Internet of Things.

I. INTRODUCTION

Users are accustomed to a world full of choices and possibilities. Thus they are often surprised and frustrated by usage restrictions and constraints that computer and communication devices impose upon them [1]. The way that we manage devices is by policies/rules, that define access to, and the use of, the resources made available. There is a growing need, even an expectation, that these should cover the variety of ways in which people and organizations want to use the devices instead of reflecting the engineering limitations [2].

With good *authentication*, the system can validate the *bona fides* of the current operator of the device. However it cannot offer control over the way the assets available on those devices are being used, either deliberately or inadvertently. These include hardware components like the camera or microphone on a mobile phone, and software elements such as files, applications or network resource access. This paper elaborates on the problem of policy for the *authorization* of requests that arise in the everyday use of the device. It also proposes an implementation that solves that issue.

In a world where we can software-define almost anything, policies must be programmable. It is not important whether the computing asset is a mobile device, an Internet of Things (IoT) element, a Personal Computer (PC), a server, a Virtual Machine (VM) or a newly instantiated virtual network function (VN). A real-time, dynamic, contextually-aware policy model with a definable policy enforcement action (not just "allowed" or "not allowed") is more effective and more useful than the traditional static approach, as will be shown. The solution presented creates a hierarchy of policies, where multiple parties can author rules, which are then prioritized. This paper contributes to the field in several ways:

- It identifies lack of authorization mechanisms suited for mobile and distributed information systems.
- It presents a framework and an architecture that has general application to the management of many device types. This solution will find wide application with spread of Internet of Things approach.
- It shows how much more extensive and complex systems can be subject to powerful and flexible policy control of arbitrary granularity.

The rest of the paper is organized as follows. Section II reviews work that precedes our research. Section III presents an overview of the Framework architecture, explaining the general approach that was taken. Section IV discusses the components of the event centric portion of the Framework, while Section V discusses policy and policy administration for it. The work is summarized in Section VI.

II. RELATED WORK

On a more abstract level, the work done by van Thanh et al., presents the concept of Device Management Service (DMS) [3]. This is a "virtual terminal" that can be used to manage multiple phones, both mobile and stationary. Unlike the Universal Personal Telecommunications (UPT), DMS allows for parallel registration, collaboration and work. This work from 2001 can be seen as the precursor of actual Mobile Device Management (MDM) solutions.

Another attempt to solve the problem of device management is presented in [4]. Here, the authors present a secure device management framework that aims to securely deliver services to user devices, manage credentials and interact with service providers. However, it mostly focuses on access control within a single unit, usually for the security purposes and secret handling.

Mei et al. discuss a design of a remote device management framework that is crafted for personal devices in [5]. This tool allows access to information about operational status, maintenance functionalities and potential security issues of the device. The advantage of their work is that it includes an open

 ISBN: 978-1-61208-451-0

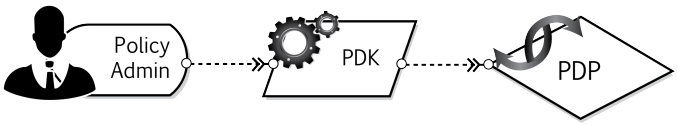

Figure 1. Framework Overview. The PDK captures the user's design intent at Level 1, and generates an instance of the PDP at Level 0.

implementation based on the SyncML Device Management (DM) specification.

Song et al. identify the problem of managing and controlling machine-to-machine (M2M) devices in [6]. This mostly state-of the-art work shows several architectures for M2M devices that are subject to standardization, with the focus on key functionalities that are used to manage and control them. On the other hand, Internet Protocol (IP) Multimedia Subsystem device management issues are presented in [7]. The paper also suggests an architecture for management by separating functions and providing service provisioning and tracing functionalities.

On the subject of security level analysis, Landman provides an in-depth analysis of the types and nature of threats to an organization from the use of smartphones as well as controls, available security software and tools, in [8]. He also shows the state of corporate security programs.

Another extensive work, where the authors suggest a new modeling methodology and present threats that MDM systems face is presented in [9]. They achieve this by an analysis of the agents, assets and ad-verse actions.

III. Framework

This section provides details of a novel, context-aware policy management solution that is applicable to a wide variety of systems, as will be shown. On a high level it comprises two layers:

a) Level 0:: Includes the operational components that communicate together to implement the policy authorization process in a live system. This level also includes compilers[1] that synthesize policy specifications in the Policy Object Language (POL) language into executable Policy Decision Point (PDP) instances for use in a live system.

b) Level 1:: Contains the Policy Design Kit (PDK), a GUI-based policy design environment that simplifies the authoring of complex policy sets. It also manages the end-to-end process of creating an instance of a Level 0 framework, ensuring its integrity, authenticity and correctness.

The data flow from the user's design intent, through Level 1, and Level 0 to produce an instance of the framework is shown in Figure 1.

IV. Framework Level 0 Architecture

As depicted in Figure 2, the operational architecture of the framework has four components: Policy Enforcement Point (PEP), Policy Decision Point (PDP), Policy Information Point (PIP) and Fabric.

[1]The discussion of POL compilers is deferred to the section on Level 1.

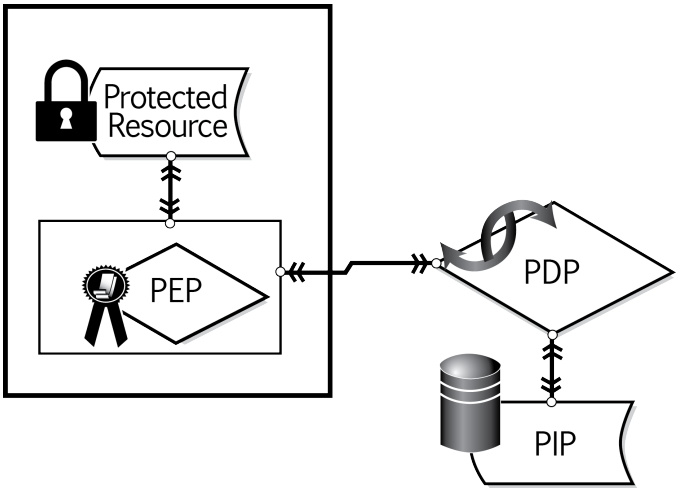

Figure 2. Framework Level 0 Architecture. The Level 0 architecture consists of four components: Policy Enforcement Point, Policy Decision Point, Policy Information Point and the Fabric (denoted by the double headed arrow).

A. Operational Elements

A PEP is an agent located on a device, such as a mobile phone. It monitors events on the device that represent requests to access resources under policy control. For each such request, the PEP creates a *query* message that contains details of the request and of the state of the device at the time of the request and transmits it to a PDP; it then waits for a response. When the PEP receives a response, it enforces the decision contained therein.

A PDP is a server that may be located remotely or, sometimes, co-located with the PEP in a device. The PDP listens for queries from PEPs and, using the data within the query and policies specific to the instance of the PDP, transmits a response to the requesting PEP.

A PIP is a data source, external to the PDP, containing information needed to evaluate policies. PIPs are typically directory services. The PDP has special features that exploit access to PIPs efficiently.

Depending upon the nature and disposition of the PEPs and PDPs and the application for which they provide an authorization service, PEP queries and PDP responses may be transmitted through the Fabric that may be any of a wide variety of external media. As will be detailed further, these may be a complex as User Datagram Protocol (UDP) over Universal Mobile Telecommunications System (UMTS) [10] or as simple as shared memory. The Fabric is depicted as a double headed arrow in Figure 2.

B. Structure of a query

From an abstract point of view, *queries* and *responses* consist of an array of elements and a fabric dependent header. Each array element represents a PEP state attribute or data pertinent to the request. For responses, each array element contains an attribute of the response, such as a verdict or information qualifying the verdict, for example a stipulation. Array items are identified by their indices, as agreed in a dictionary shared by the PEP and the PDP. The array is referred

 ISBN: 978-1-61208-451-0

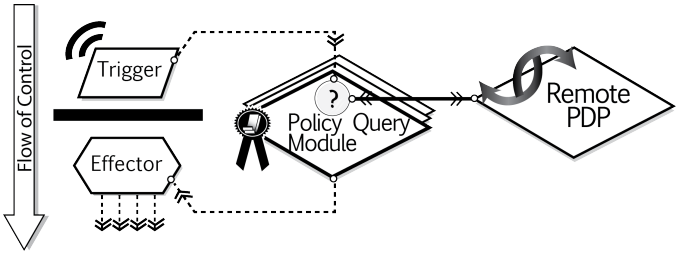

Figure 3. Policy enforcement. The Policy Enforcement Points (PEPs) are remote agents located on the managed device that are triggered by access requests. PEPs query the PDP for authorization and enforce the response.

to as the *dynamic array*, and each of its elements is referred to a *dynamic*.

C. Policy Enforcement Point (PEP)

Depending on the type of resource, the PEP can be located at the driver level or implemented as an application. The PEP consists of a Trigger, a Policy Query Module (PQM) and an Effector, as depicted in Figure 3.

The Trigger, interrupting the normal flow of an event, gathers the relevant dynamic content and sends it to the PQM.

The purpose of the PQM is to create queries and enforce the intent of responses on the operation of the device. Depending on the Trigger and any included dynamic attributes it may look for additional environmental variables to form a query. It then sends the query to a designated PDP (local or remote). Finally, it obtains a policy decision and provides the response to the appropriate Effector for that Trigger.

The process followed by the PQM varies according to the nature of the PEP. This includes consideration of the state of the device (online or offline), whether or not the device policy cache may be used and which PDPs may be consulted for a policy response.

The Effector enforces verdicts returned by the PQM. It provides a meaningful control path for each possible verdict/stipulation combination and ensures a sensible user experience for all outcomes.

Caching is shared by all PEPs on a given device because the wireless network can present an inconsistent and sporadic connection to all back-end services. Policies can direct caching of responses for a specified period of time. This reduces fabric bandwidth demand on the network and allows for near-instant response (to queries that match the same conditions) without recourse to an external connection. This is useful for devices that have a high frequency request cycle, such as a camera that may query as frequently as 60 times a second.

In the event that there is no cached decision and the PDP is not reachable, a simple set of decisions is stored on the device to provide a fail-safe response for all PEPs. If the network does not allow for a real-time decision to be received, the response is equivalent to that provided by traditional MDM solutions.

Policy decisions may include stipulations that require logging. There is a common service which collects log data and delivers it efficiently. For example, if multiple control points are logging and the network is unavailable, the common service will hold the data and deliver it as a bulk payload once the connection is restored. This relieves individual control points from tracking and performing these functions.

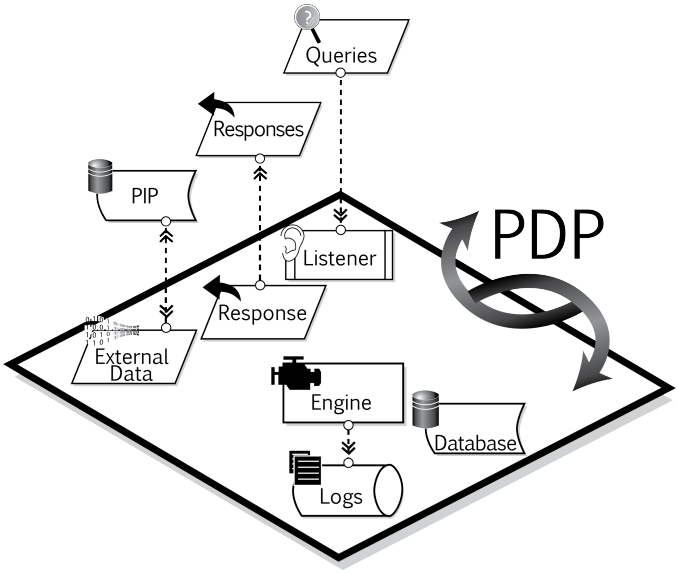

Figure 4. Policy Decision Point. The PDP is a server that issues that issues responses to queries from PEPs. Its responses are based on PDP-specific policies. PEPs and PDPs communicate through a two-way channel.

The heartbeat function maintains periodic communication between the PEP and the PDP confirming PEP presence and state. The PDP may respond to heartbeats with control messages.

The installation of PEPs to devices depends on the nature of the implementation and the protected resource. PEPs protecting hardware resources are implemented at the hardware resource driver level. For example, a camera resource needs to be implemented in the camera driver. PEPs protecting data files (for example), may be implemented as part of a file system driver or as part of an application.

The PEP ensures a sensible outcome for the end user or application when an access attempt is denied. Devices do not appear broken; applications do not crash.

D. Policy Decision Point (PDP)

A PDP is a server that issues responses to queries received from PEPs. The responses are based on policies that are specific to the PDP and that make use of the dynamic values in the query. Refer to Figure 4 for more detail.

1) Characteristics of a PDP: In order to ensure availability, a PDP always returns a response to a PEP even if no relevant policy is found. A PDP is defined as *permissive* or *restrictive* depending on whether it sends an ”allow” or ”deny”. A PDP can handle large collections of policies (thousands or more) without significant performance degradation. The PDP is stateless. This improves performance and simplifies interactions with PEPs and management of PDP farms. For audit and forensic analysis purposes, a PDP generates a log of all query transactions. A PDP may respond to PEP heartbeat signals with commands that affect the state of the PEP or host device. One of the operations that may be requested is maintenance for a PEP instance.

 ISBN: 978-1-61208-451-0

The decision engine of the PDP interfaces at the level of abstract query objects. This allows the use of various harnesses to adapt the PDP to a variety communication fabrics. Furthermore, this provides for the creation of systems with multiple, concurrent PDP instances for higher bandwidth needs.

2) Security Design of a PDP: PDP architecture minimizes the opportunities for malicious intervention. The code of the PDP itself, as well as the representation of policies is attestable (signed and authenticated) through a chain of trust. The process of creating PDPs and their policies and delivering them can be made verifiably secure.

The PDP is synthesized by a compiler which directly generates its executable, including the embedding harness, from policy *specifications* in the proprietary POL language. The PDP offers no programming API or facility for source code tampering or modification and the executable may be signed.

To limit the attack surface, the PDP has only one I/O channel on which encrypted packets are received (queries) and transmitted (responses). It has an output channel on which logging information is emitted. Moreover, in order to prevent unauthorized access and modification of the policy database and rules, the database is password-protected. Only the compiled PDP and the data-base server have the key, which is generated at compile time. Essential database fields may also be encrypted.

Lastly, when it comes to network fabric, queries and responses are transmitted using a protocol that is resilient to man-in-the-middle attacks and spoofing. The PDP drops incoming packets as soon as protocol discrepancies are detected.

E. Policy Information Point (PIP)

Corporate data, particularly personnel data, is often stored in active directories and databases which deliver fast data retrieval and consistency. Policy formulation often requires precisely this kind of data. The PDP has been designed to connect to directories. The POL language provides directory specification and search functionality to make the data available for policy evaluation by the PDP.

V. Policy and Policy Management

From the forgoing description it is clear that the behavior of devices governed by the system is dependent upon the policies embedded in the PDP. A policy is a rule that dictates what actions should be taken for a particular event under a given set of conditions. Individual policies are gathered together into policy sets which together address all of the events and circumstances of interest to the policy authors.

A. Characteristics of Policy Management

The objectives of policy management include the following:

- The method of expressing policies is rich enough to express policy author intent under a wide range of circumstances, some of which cannot be foreseen. It is succinct so that the resulting policy sets remain manageable in size and complexity.
- Policies are written only by those who have the required authority for the resource being governed.
- Policy sets created by many authors are combined with a clear order of precedence, with any conflicts or logical problems detected and addressed.
- Information referenced by the policies to arrive at decisions, in PIPs or otherwise, are only writable by properly authorized authors.
- Compilation and deployment of PDPs employs a controlled process and is done by authorized individuals.

B. Chain of Trust

A result of a stringent implementation is a chain of trust for the policy data, beginning with the creation of policy and extending through to the deployment of PDP. This chain of trust ensures that the integrity of policy intent is maintained.

The establishment of the chain of trust requires two types of policies: Level 0 policies which specify how PDPs respond to device requests, and Level 1 policies which govern access control for data used to create Level 0 policies. These also secure custody of the data.

1) Policy Specification, Level 0: Level 0 policies are specified in a formal policy language, the POL. POL is terse and declarative to facilitate synthesis and static verification of policy sets. It follows a pattern of: Subject, Agent, Object, Action and Environment (SAOAE). Each component has a corresponding clause in the model which consists of an arbitrary expression that can reference dynamic data from the query, static data from a PIP or data from the policy set itself. The clauses organize policy considerations along the following lines:

- Subject: the identity of the entity making the request, e.g. the user.
- Agent: the means by which the request will be carried out, e.g. the program that will make the access.
- Object: the elements and items affected by the request and being acted upon by the Agent.
- Action: the specific function that the Agent applies to the Object.
- Environment: information in the request that would be observable at the PEP but independent of any given event, such as time or location.

The POL language provides a mapping feature that allows query and PIP data to be tagged along arbitrary lines. Tags can be tagged themselves, forming chains which begin with query or PIP data, allowing large numbers of specific data elements to be aggregated into manageable categories for expressing policy intent. Support is provided for regular expression constructs, simple geolocation constructs and time intervals. For example, a tag chain could be constructed to map an identifier from a mobile device into an employee number, and then into a role or a location or an authorization level. Clauses in SAOAE statements can reference tag chains in their expressions. The number of tags and length of tag chains is not dictated by the language. It can be used to construct both policies with a wide scope and general application, or policies for specific situations with very fine granularity.

 ISBN: 978-1-61208-451-0

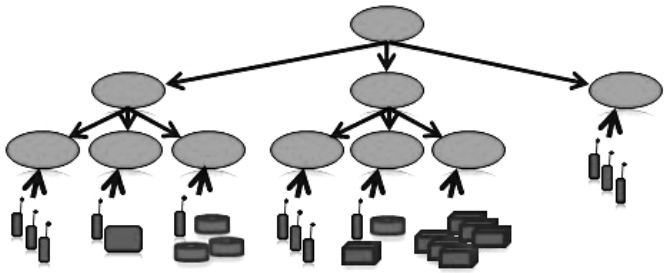

Figure 5. Policy Ownership Tree. Policy stakeholders are arranged in a hierarchy, each responsible for their own domain and allowed to delegate authority to those beneath them.

2) Policy Application, Hierarchy and Delegation: Designated authorities are permitted to write the policies for those areas within their domain. As shown in Figure 5, Policy stakeholders are arranged in a hierarchy and are allowed to delegate authority to those beneath them. This permits higher-ranking authorities to reserve for themselves the rights they need and to delegate policy decisions to others as appropriate. Policies may be marked as default, to permit policy set closure if no lower-ranking policy authors provide an applicable policy.

This policy ownership tree model can be applied in various ways to simplify the management of complex policy sets. For example, one approach might use an organization chart to map the hierarchical tree such that levels of policy authority correspond to levels of organizational authority. Another example might use a high-ranking policy set to specify some coarse-grained generic default policies and one or more lower-ranking policy sets to create categories of specific exceptions to the generic rules.

3) Administrative Policies, Level 1: The POL language and its constructs define Level 0 policies which address the manner in which the system responds to requests from devices. In order to implement the chain of trust as outlined above, another layer of administrative, or Level 1 policies is required to govern the authentication and authorization of stakeholders as they perform their duties within the system. These duties can consist of policy and data entry, configuring policy information points, compiling and deploying PDPs, introspection and debugging of policy sets, and administration and management of the administrative policies themselves. Administrative policies are defined and enforced in the PDK.

C. Policy Design Kit (PDK)

The PDK is a general purpose authentication and authorization platform that provides controlled access to the data and tools and a policy life-cycle framework. Its work is presented in Figure 6. After authenticating to the PDK, users receive a fine-grained set of abilities dictating allow/deny access to various resources. Data objects in the PDK have a strict chain of ownership back to a user. These abilities combine with the object ownership to enforce Level 1 policy. Relationships between users are established in the PDK to enforce rules for sharing data, and to establish relative positions in the policy ownership tree.

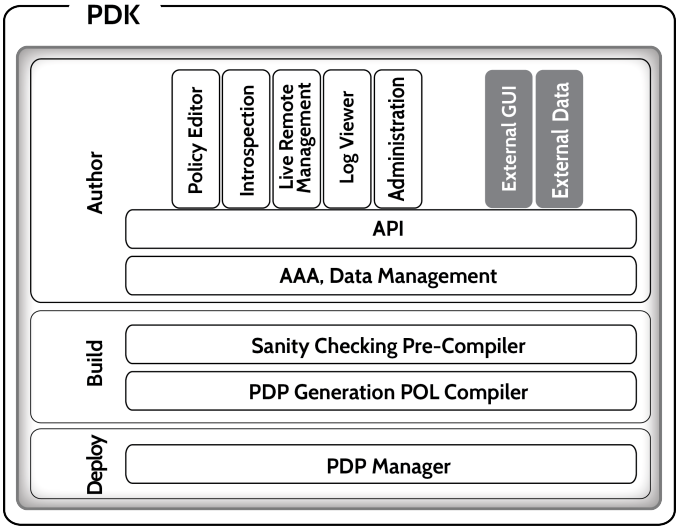

Figure 6. Policy Design Kit. The PDK is a policy life-cycle management framework with a set of tools for the policy administrator to conduct policy authoring, introspection of policy sets, PDP generation, sanity checking and deployment.

1) Policy Capture, Level 0: The PDK provides policy models that allow users to express their Level 0 policy intent at a level of abstraction suitable for a given application. For example, a policy model that governs a resource according to a time interval and a location might present an interface that captures just the interval and the location. No knowledge of POL is required to use the models. Data to populate a policy model's tags is also captured by the model at an appropriate level of abstraction. The PDK provides a number of policy models; however, a given user only has access to those for which they are authorized by Level 1 policy.

2) PIPs and Data Sharing: While PIPs usually contain common organization data, often additional data is required for policy authoring. For this reason, the PDK provides models that allow users to capture extra data and explicitly share the information with other selected policy authors by using the PDK's authorization mechanism. The shared data is available to a second user's policy models but cannot be viewed or changed. As an example, an employee in HR may be responsible for maintaining employee data, while an IT employee may dictate network policies using a role-based scheme bound to that employee data.

3) Configuration: The PDK provides a set of user administration functions that allow users authorized by Level 1 policy to manage other users in the PDK. These functions include the typical user life-cycle operations, as well as the assignment of abilities to users which grant them access to various functions, policy models and shared data in the PDK. Users and their policy sets can also be assigned to their positions in the policy ownership tree.

4) PDP Generation, Sanity Checking and Deployment: The PDK provides tools for the generation and compilation of PDPs from policy set trees. The first step in the process is to synthesize POL code from the policy model and captured data. The POL code is then compiled to produce C++ code and any associated database. The compiler carries out simple semantic

 ISBN: 978-1-61208-451-0

checks as well as a sanity check, which verifies that the policy set possesses certain properties. For example, it verifies that there is no ambiguity where two policies might apply to the same query, and it verifies that there is a set of query data capable of activating each policy in the set. The last step is to compile the C++ code to produce the executable PDP. The entire flow can be executed stepwise or automatically as a single step.

The PDK provides functionality to define server elements, to transfer the executable PDP to them and to set them running. Server elements can be designated as database servers, PDP servers or both. This allows flexibility in the number of PDPs deployed, and in load balancing the number of executing PDPs to the number of corresponding database servers.

D. Debugging and Introspection

The executable PDP produced by the compiler has minimal I/O. I/O is limited to receiving requests, emitting decisions, and providing a log file. The log file provides an indication of which policy was used to arrive at a decision; however, it doesn't provide any information on how that policy was selected. As policy set complexity increases, so does the need for tools to analyze and debug them, and examine them forensically. To address this, the PDK provides features for log file analysis and management as well as policy set introspection.

Authorized users of the PDK can collect PDP log files from various servers, combine and analyze them. The log entries for any given device may be distributed amongst a number of PDPs in a farm. Combining them allows the entire history for any one device to be examined. The combined log is elaborated to allow any sequence of events to be found quickly.

Introspection refers to the activity of studying policy set behavior by simulating the PDP look-up process, while displaying all intermediate results. For a given simulated request, each of the policies considered are displayed in rank order, along with the values of all policy clauses. Intermediate terms and tag expressions in the clauses are also evaluated and displayed, providing detailed information showing why any given policy was selected, rejected or not evaluated. This capability allows authors to evaluate different scenarios against policy sets and groups of policy sets to determine if desired outcomes are being produced.

VI. Conclusions

The paper postulates that as computing assets become commonplace, the need to manage data on an event-centric and context-aware basis is becoming urgent. Mobile Device Management has been adopted as the industry standard for mobile security. However, current MDM solutions are hard to manage, are inflexible and apply the rule "define once, run always". When it comes to mobile devices, Personal Computers, servers, health trackers, Virtual Machines or Internet of Thing devices, more power and flexibility is needed.

This paper presents a novel, context-driven policy definition and enforcement framework which addresses these shortcomings. Instead of deciding whether access is granted to a resource at configuration time, this solution takes into account the state of the system, time, location and any other definable factors at the time of the event. The solution addresses the entire policy life-cycle: formulation, management, verification, debug and analysis of policy behavior at time of definition as well as execution, policy server generation and secure delivery of the implementation.

The solution can be applied to many fields, not only those with a focus on the enterprise. The Framework elements that are common to all foreseeable applications are discussed: the Policy Design Kit, the Policy Decision Point, the Policy Enforcement Point and the Policy Information Point. The application of the solution to MDM is presented. It shows how extensive and complex systems can be subject to powerful and flexible policy control with appropriate granularity, while still remaining manageable. Evaluation is based on comparison with existing solutions.

The context-aware policy framework provides generalized resource access control based on a wide set of conditions. The wide range of factors that may be considered in policy, the unrestricted specification of policy, the hierarchical and delegation features, the real-time assessment of events and conditions, and applicability to other domains, such as the Internet of Things, are unavailable in current solutions. Most importantly, the combination of these features represents a significant step forward in the field of access control.

References

[1] B. J. Myers, "Student perceptions of computer anxiety: The relationship of computer attitude, computer experience, age, gender, and socioeconomic status," Ph.D. dissertation, Vermillion, SD, USA, 2006, aAI3255100.

[2] S. M. AlHaj, "Context-aware policy management platform: Based multi-agent systems mas," in *Proceedings of the 2011 Developments in E-systems Engineering*, ser. DESE '11. Washington, DC, USA: IEEE Computer Society, 2011, pp. 490–495. [Online]. Available: http://dx.doi.org/10.1109/DeSE.2011.89

[3] D. V. Thanh, T. Jonvik, E. Vanem, D. van Tran, and J. Audestad, "The device management service," in *Intelligent Network Workshop, 2001 IEEE*, May 2001, pp. 199–211.

[4] A. Leung and C. Mitchell, "A device management framework for secure ubiquitous service delivery," in *Information Assurance and Security, 2008. ISIAS '08. Fourth International Conference on*, Sept 2008, pp. 267–274.

[5] H. Mei and J. Lukkien, "A remote personal device management framework based on syncml dm specifications," in *Proceedings of the 6th International Conference on Mobile Data Management*, ser. MDM '05. New York, NY, USA: ACM, 2005, pp. 185–191. [Online]. Available: http://doi.acm.org/10.1145/1071246.1071275

[6] J. Song, A. Kunz, M. Schmidt, and P. Szczytowski, "Connecting and managing m2m devices in the future internet," *Mob. Netw. Appl.*, vol. 19, no. 1, pp. 4–17, Feb. 2014. [Online]. Available: http://dx.doi.org/10.1007/s11036-013-0480-9

[7] J. Ma, J. Liao, and X. Zhu, "Device management in the ims," *J. Netw. Syst. Manage.*, vol. 16, no. 1, pp. 46–62, Mar. 2008. [Online]. Available: http://dx.doi.org/10.1007/s10922-007-9092-7

[8] M. Landman, "Managing smart phone security risks," in *2010 Information Security Curriculum Development Conference*, ser. InfoSecCD '10. New York, NY, USA: ACM, 2010, pp. 145–155. [Online]. Available: http://doi.acm.org/10.1145/1940941.1940971

[9] K. Rhee, D. Won, S.-W. Jang, S. Chae, and S. Park, "Threat modeling of a mobile device management system for secure smart work," *Electronic Commerce Research*, vol. 13, no. 3, pp. 243–256, September 2013. [Online]. Available: http://dx.doi.org/10.1007/s10660-013-9121-4

[10] M. Sauter, *Communication Systems for the Mobile Information Society*. John Wiley, 2006.

 ISBN: 978-1-61208-451-0

Detection And Protection Against Unwanted Small UAVs

Igor Tchouchenkov, Florian Segor, Rainer Schönbein, Matthias Kollmann,
Fraunhofer Institute of Optronics, System Technologies and Image Exploitation IOSB
Karlsruhe, Germany
e-mail: {igor.tchouchenkov, florian.segor, rainer.schoenbein, matthias.kollmann}@iosb.fraunhofer.de

Thomas Bierhoff,
Atos IT Solutions and Services, Paderborn, Germany
e-mail: thomas.bierhoff@atos.net

Mark Herbold
Atos Nederland B.V., Utrecht, Netherlands
e-mail: mark.herbold@atos.net

***Abstract*—The market for small Unmanned Aerial Vehicles (UAVs) is growing fast. Based on increasing reliability, accessibility and decreasing operational restraints these systems become more and more used in different areas of application. Besides the benefits of this development, an increasing number of more or less hazardous incidents with small UAVs can be noticed. Accidents or the misuse of these technologies do lead to dangerous situations – from simple mishaps over illegal activities like spying or drug transportation up to the possibilities of terroristic actions. To deal with this development a concept of a new Low Altitude Air Surveillance Control (LASC) system is developed. Based on heterogeneous, distributed multi-sensor networks embedded into a capable and application oriented modular, adaptable and scalable backend system, the LASC concept addresses detection, localization, tracking and classification or identification of small UAVs providing interactive threat and risk assessment as well as suggestions for adequate counter measures.**

Keywords-UAV; air surveillance; multi-sensor network; threat protection; modular.

I. INTRODUCTION

Small Unmanned Aerial Vehicles (UAVs) are an emerging technology with a great potential to disruptively change our lives. They have by far exceed the capabilities of the niche products of remote controlled models with their ability in terms of payloads, flight duration and ranges, auto pilot capabilities, automated collision protection and video transmission capabilities. Leveraged with new technologies (e. g. high capacity battery packs, low energy consuming motors and small-sized high computing power) the small UAVs have started exceeding entertainment domain while entering more and more applications [1][2][3] (surveillance, reconnaissance and rescue, video production, logistics, etc.). This development has been boosted by a constantly rising commercial market for UAVs providing broad accessibility and diversity at a low cost scale.

As always, each technology comes along with drawbacks and potential for abuse and this is in particular true for small UAVs. With their inherent risk of crashing causing damage and harm to people, each business application has to be assessed seriously in terms of security issues and constraints. Furthermore, with the broad availability and low cost aspect of UAVs, a common and unforeseeable use of this technology is expected in the private sector. This will in turn demand for new legal laws, rules and enforcement technologies [4][5] to keep the low altitude air space save and controlled.

In this article, detection technologies for small UAVs are described in Section II, as well as the concept of a new Low Altitude Air Surveillance Control (LASC) system in Section III. The paper is closed with a final discussion in Section IV.

II. DETECTION TECHNOLOGIES FOR SMALL UAVS

Taking the established national and international airspace control as an archetype, the simplest solution for detecting and identifying a small UAV would be the mandatory introduction of a standardized identification friend or foe (IFF) system deployed in all UAVs. If this IFF provides the capability to broadcasts the current position, altitude, heading and speed together with a unique identifier, the integration of these systems into the lower airspace would be easy. But the standard IFF systems are not applicable on board of small UAVs, and the number of small UAVs is growing rapidly. Moreover, such a technology can only be a part of the solution as it provides no functionalities to deal with illegal activities. To establish a safe and secure lower airspace, additional technologies that do not rely on the cooperation of the traffic are of major importance.

Small, relatively slow and low flying UAVs as we know them today can be either detected by searching their electronic, acoustic, thermal or visual signature or by using active sensors like radar or lidar to search for anomalies that might be operating UAVs.

Video based airspace surveillance within the optical as well as the infrared spectrum is a promising approach. Existing commercial solutions rely mostly on easy change detection algorithms and basic data fusion [6]. Fraunhofer Institute of Optronics, System Technologies and Image Exploitation (IOSB) has done comprehensive research on robust computer vision algorithms allowing the detection and primary classification of different flying objects in real time [7]. But this approach comes with challenging requirements – especially for urban areas. The whole complicated low altitude airspace need to be continuously observed without knowing the number, the size and the distance of a small UAV to be safe detected and separated from other eventually moving objects.

 ISBN: 978-1-61208-451-0

Sound pattern emitted by UAVs are also representing a promising source for detection and classification. Microphones combined with digital signal processing using matched digital filters adjusted to the characteristic sound frequency spectrum can unveil an approaching UAV [6]. The approach vector can be identified by the alignment of directional microphones and lead to distance estimation by signal strength or triangulation.

Communication detection of UAVs is very promisingly and partially already in application [8]. Every remote controlled UAV needs up-/downlink to the ground control station. Control commands are sent to the UAV and sensor data like position, system state and in particular the payload data (video signals in most cases) are sent back. If the typical frequency bands are monitored, characteristic communication can be identified and transmission can be extracted. The information can be used to do a rough classification of the UAVs type. In conjunction with triangulation capabilities of cooperating sensors, position and heading of a detected UAV can be determined as well. But these technologies will fail if the UAV is operating autonomously. More sophisticated technologies detecting the electromagnetic background radiation emitted by the electronic equipment of the UAV could provide a promising approach, but such systems do not exist.

Radar technology is always tailored to its application scene (e.g., detection range, size of object, material, etc.). It is insensitive to environmental conditions (poor visibility, no light at night time, rain, fog, etc.) and therefore applicable in almost any environmental situation. Long range air surveillance radar operating at 1 to 2GHz (L-Band) for range <400km and large objects (>10m) are not suitable for the LASC system as UAV's small size, its low EM-wave reflecting composite material and low flying altitudes.

More suitable radar technologies are found in the K- and Ka-Band with frequencies between 20 and 40 GHz as far as in the W-Band (60 - 120 GHz). But a robust detection and recognition of small UAVs with radar are still a topic of research. Applied in urban areas, reflection by infrastructures and building are representing the major challenge for development of appropriate radar surveillance systems. In addition, a usage of many radar systems in urban areas can be problematic because of their radiation. Lidar seems to be a promising solution as it can provide a comprehensive three dimensional image of the surroundings allowing to precisely locate and in some cases to identify an airborne object, but the range of these sensors is too short.

III. LOW ALTITUDE AIR SURVELLANCE AND CONTROL CONCEPT

Most approaches or commercial available systems rely on a single sensor technology. Derived from the possible detection methods these systems are developed with the aim to get the best possible solution based on the capabilities of the selected single sensor type - in the best case aligned with an additional secondary sensor for a better detection rate. But all sensor technologies do have their advantages and disadvantages and a capable sensor can totally fail or become unsuitable if the situation changes.

The protection concept is normally centralized and focusing on the single sensor system itself. But systems for UAV detection and defence require solutions covering different spotting directions which often cannot be covered by a single deployed sensor. A multiple application of a sensor is not suitable if the technology is not using a control centre that is logically linked to the sensor system.

Additionally, the detecting is not sufficient if no adequate countermeasures are timely provided. On the other hand, some countermeasures can be unsuitable and even dangerous if a detected UAV is not classified or identified.

To cope with these aspects, the new LASC system concept was developed in a joint project between AToS SE and Fraunhofer IOSB.

The air space beyond the 500m is already been in control by conventional air traffic control mostly based on long-range radar technologies, so the major objective of a LASC system is the monitoring of the today uncontrolled air space below 500m with a special focus on urban areas. In order to guarantee seamless information exchange (e.g., some UAVs may enter the high altitude air space and endanger the air traffic), LASC system must be integrated into conventional air traffic control.

The common workflow of LASC is shown in Figure 1. The LASC system continuously monitors the air space with multiple sensors. Once a sensor detects a flying object, the system will try to locate it (e.g., by triangulation of signals from many sensors) and ensure tracking by orchestrating multiple sensors. Once the location and tracking is established, the backend system of LASC will carry out the (possible interactive) classification and if applicable the identification of the detected object. Once this is executed, the LASC backend system starts the evaluation of the drone's authorization to fly through the current air space corridor. If no authorization is given, system will prepare reaction options considering the current situation, estimated threat classification and risk assessment (disturbing, illegal or endangering). The LASC system provides decision support to a human operator who is in charge of selecting and initiating the counter measures. All activities that do not need necessarily the interaction with the operator must be automated to a high degree in order to guarantee a real time execution and to enable high scalability in terms of multiple events.

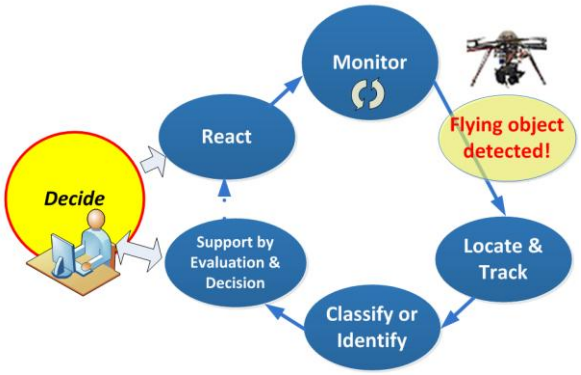

Figure 1. Common workflow of the LASC system

Depending on the application, the area of the surveillance is scalable by adding further sensor- and counter measure devices to the LASC systems or new sub-systems. Distributed LASC system can contain both mobile and stationary LASC sub-systems. This modular approach provides a broad field of application ranging from a single building to an entire suburb air surveillance. Therefore, a multimodal sensor network based on the technology stack recapitalized in chapter II must cover a certain part of the controlled air space providing day- and night time operations and coping with bad visibility conditions.

The next step after UAV detection is its classification based on sensor signal evaluation and its identification by IFF signals broadcasted by the UAV (if available). Once identified, the LASC system needs to reconcile the information with a LASC central UAV flight register to see if the UAV's flight is registered/allowed and further information is given covering payload, mission, flight path, destination and operator. If the UAV is registered, its flight register record must be updated by current position, altitude, speed and time stamp for consistent tracking. If there is no suitable record found in the LASC register, a new one must be created with all available information (e.g., ID with current position, altitude, speed and time stamp). In order to approve UAV flight authorization, its position must be mapped to pre-defined air space corridors, which will show permanent or temporarily valid restrictions or prohibitions. This is essential because different UAVs may have different air space transit rights (e.g., police UAV may enter the corridor which it prohibited for other participants). The last step of monitoring flight operations is the continuous tracking and updating the LASC register's records. Once a UAV is leaving the observation area of a sensor or a LASC system, it might enter another one. The hand-over of such tracking must be supported by the LASC intelligence, which provides analysing and predictions of flight path and identification of sensors which might detect the UAV soon.

The second objective of the LASC system is the downstream air space regulation enforcement for unauthorized UAV. Therefore, the violation of the air space must be unveiled, which must trigger a threat and risk analysis to determine suitable reaction. Based on the classification of risks and the availability of interception capabilities in reach, the LASC decision support must provide human operators with suitable enforcement and interception solutions. As the interception is near always related to potential collateral risks [9], it always needs to be initiated and controlled by a human operator. Therefore, the operator needs access to the sensors (e.g., video/IR camera) of the LASC system to leverage assessment of the situation and to gain a reliable decision base. A usage of geo-referenced situational awareness (e.g., showing locations of and distances to critical infrastructures in reach) facilitates the work of the operator considerably. All this needs to be integrated into the control centre of the LASC system to ensure a convenient use even in dangerous situations.

The major challenge of the LASC interception operation is the avoidance of collateral damages. Military interception solutions for low-altitude flying objects (e.g., lasers, shells or rocket defence systems) are not acceptable in urban areas. The "soft" interception techniques (like communication based mission distortion and interruption) need to be researched and integrated into the LASC solution. Another challenge is real time ability. In case of an abuse, the UAV might take off near its destination reducing reaction time drastically down to seconds. Therefore, most LASC system functionalities must be automated by delivering alerts and decision support to the operator within few seconds.

The common system design and major building blocks for a LASC system providing comprehensive monitoring, intelligence and interception functionalities are depicted in Figure 2.

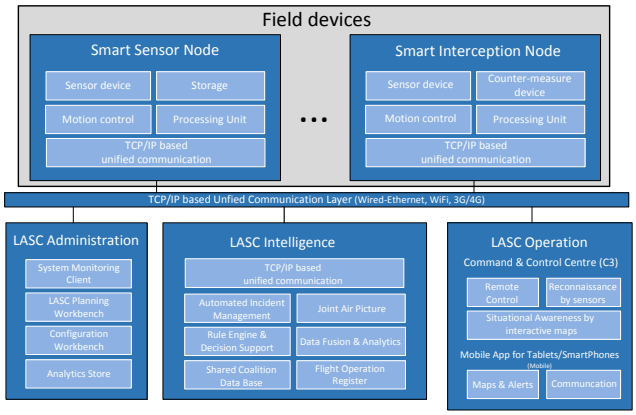

Figure 2. System concept with major building blocks

A field device of the LASC system contains the distributed network of smart sensor and interception nodes. The appropriate choice of the technology used by the nodes is strongly depending on the application and the environmental constraints (free space and the line of sight, electromagnetic or acoustic disturbances, realistic estimation of possible hazards, etc.). So deploying various sensors and interception technologies in a multimodal approach must be considered because different technologies can cooperate and extend their capabilities among each other during a parallel operation.

Smart sensor nodes are distributed on appropriate places across the observed area and equipped with data storage and processing units to pre-process sensor raw data using specific algorithms and embedded system technology. This will decrease the data volume to be transmitted to a ground station by far, as only derived data in case of an event needs to be transmitted. The derived information, analyzed data and detected events are provided via web services hosted in the sensor node and end-to-end secured (SSL, VPN) TCP/IP based communication channels. In addition, remote control and diagnose of components of each smart node is also enabled through the web services.

Smart interception nodes are also distributed across the observed area, placed stationary on buildings, infrastructures, balloons or mobile on vehicles (cars, UAVs, etc.) to provide appropriate interception capabilities. The interception nodes are based normally on the same subcomponents as the sensor nodes except that the data storage is replaced by a counter

 ISBN: 978-1-61208-451-0

measure device. The sensor device, deployed in the interception node, in conjunction with a processing unit will enable automatic interception process preparation and support – it is especially important by multiple simultaneous threats.

Beside the field devices, the LASC system will contain three further major building blocks covering the domains administration, intelligence and operation.

The LASC administration block covers all kind of tools to configure, monitor and maintain the LASC system infrastructure and its components. It uses automated system monitoring and observation clients, which enable the survey of the operational state of sensor and interception nodes, allow load balancing of the intelligence platform and support the configuration of the intelligence itself (e.g., setup of analytics chains). Another important part of the LASC administration is the planning workbench, which provides simulation-based design support for arranging the sensor- and interception node network across a defined area in order to achieve a certain degree of air surveillance performance.

The LASC intelligence is the core block of the entire system. Its major task is the fusion of information from the sensor network and appropriate fast data analytics and complex event processing to provide a real time joint air picture (JAP) of the observed area. The JAP uses the georeferenced visualisation of all detected and tracked UAVs including all restricted and/or prohibited air areas. Utilizing a coalition shared database and the flight operation register, the LASC tries to identify a detected UAV automatically. If this fails, the LASC intelligence classifies the UAV and provides all available information (first of all video streams) for manual identification by a human operator using decision support tools. Once an air space violation is unveiled, an automated incident management provides available options for the operator to react appropriately to the incident. This incident manager utilizes a rule engine including pre-defined decision trees, risk assessments and potentially applicable counter measure nodes in reach. Once the LASC intelligence platform is deployed on virtualized server infrastructure, its service-orientated architecture provides enough performance to handle big amount of sensors (>1000), process their data in real time and provide all kind of information to multiple authorized operators.

The last building block presents the operation clients of the LASC system, which can be distinguished into stationary or deployable command and control centres (C3) and mobile apps supporting mobile access to the LASC intelligence information system. It can be developed, e.g., based on AMFIS ground control station [10]. All clients are connecting to the LASC intelligence platform via secured end-to-end encrypted (VPN, SSL) TCP/IP based communication channels and utilize its web service provision to access data and control. Once an incident is detected, alerts are shown or sent to operator and other responsible persons. The operator can request not only decision support from the LASC intelligence, but also real time video streams from optical sensors in the reach. All counter measure activities can be initiated and controlled by the C3 client software and are sent via the LASC intelligence control proxy to the selected interception node.

For mobile solutions, client apps running on smart phones and tablets are provided to inform responsible persons about incidents in the near environment and/or transmit instruction for further action.

IV. CONCLUSION

The LASC system concept includes multi-sensor detection, localization, tracking and classification or identification of small UAVs and their payloads integrated in a scalable distributed system. LASC system provides fast interactive threat and risk assessment as well as selection possibilities for adequate counter measures supported by a user-friendly interface.

The scalable architecture of the distributed LASC system has open interfaces wherever it is possible and includes data analysis and fusion modules, coalition shared database as well as interactive visualization and decision support components. The LASC system must be integrated into conventional air traffic control to prevent possible incidents because of intersecting air spaces.

In the next steps, major components of LASC as well as the system framework will be developed.

REFERENCES

[1] J. Vasagar. Finanical Time. DHL to use ‘paracelcopter’ drones for delivery. [Online]. Available from: http://www.ft.com/cms/s/0/c00bd8e2-44ad-11e4-bce8-00144feabdc0.html#axzz3RtVgsK6d 2014.09

[2] Wikipedia. Amazon Prime Air. [Online]. Available from: http://en.wikipedia.org/wiki/Amazon_Prime_Air 2013.12

[3] P. Molina, et. al., ”Drones to the Rescue,” InsideGNSS Journal, pp. 37-47, July 2012.

[4] F. Lardinois. FAA proposed rules to open sky to some commercial drones. [Online]. Available from: http://techcrunch.com/2015/02/15/proposed-faa-rules-will-open-the-sky-for-some-commercial-drones-but-delivery-drones-remain-grounded 2015.02

[5] Federal Aviation Administration. Overview of small UAS Notice of Proposed Rulemaking. [Online]. Available from: http://www.faa.gov/regulations_policies/rulemaking/media/021515_suas_summary.pdf. 2015.02

[6] DeDrone. Multi-Sensor Drone Warning System. [Online]. Available from: http://www.dedrone.com/en/dronetracker/drone-detection-hardware 2015

[7] Fraunhofer IOSB. Experimental setup for object recognition and tracking. [Online]. Available from: http://www.iosb.fraunhofer.de/servlet/is/24431/ 2015.10

[8] DDC LLC. The Basic Drone Detection System. [Online]. Available from: http://www.ddcountermeasures.com/products.html 2015.05

[9] I. Tchouchenkov, F. Segor and R. Schönbein, “Einsatzmöglichkeiten und Abwehr kleiner unbemannter Fluggeräte,” POLIZEI-heute, 2012, Nr. 3, pp. 74-79.

[10] A. Bürkle, F. Segor, M. Kollmann, R. Schönbein, “Universal Ground Control Station for Heterogeneous Sensors,” In Journal On Advances in Telecommunications, IARIA, 2011, Volume 3, Numbers 3 & 4, pp. 152–161.

 ISBN: 978-1-61208-451-0

Privacy Issue in Federated IDMS for Cloud Computing

Yeongkwun Kim
School of Computer Sciences
Western Illinois University
Macomb IL, USA
Y-Kim2@wiu.edu

Injoo Kim
Department of Computer and Information Science
East-West University
Chicago IL, USA
injoo@eastwest.edu

Charlie Y. Shim
Department of Computer Science and Information Technology
Kutztown University of Pennsylvania
Kutztown PA, USA
shim@kutztown.edu

Abstract— **Federated cloud identity management systems (IDMS) enable users to use the same identity information to access all network services and resources in the trusted domain. Federated cloud IDMS has gained significant attention from the IT industry due to its support of cross organizational boundaries without creating additional user accounts. However, using the same identity information can disclose comprehensive user profile information, such as usage pattern, interests, or the behavior of the user. In this paper, we discuss possible security concerns, especially privacy issues, and ultimately propose a way to preserve privacy in the federated cloud IDMS by use of same single user identity information.**

Keywords - cloudcomputing; federal identity management; privacy.

I. INTRODUCTION

Due to the rapid development of computer and network communication technologies, cloud computing has achieved increased popularity. The National Institute of Standards and Technology [1] defines cloud computing as "a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources such as networks, servers, storage, applications, and services." In order to access network services or resources, users are required to present their personal identities for the purposes of authentication. This subsequently results in cloud computing posing potential a significant challenge to the user's information, security realization and privacy protection. Thus, in response, industries have introduced IDMS as a means of managing the identity information of different users [2]. According to deployment architecture, IDMS can be classified into the isolated cloud IDMS, the centralized cloud IDMS, and the federated cloud IDMS [3]. The federated cloud IDMS enables users to use the same identity information to acquire access to all network services and resources within any trusted group of enterprises [4]. For the purpose of authentication, network users typically maintain a set of user identity credentials such as a username/password combination with every service provider. Unfortunately, the number of interactions service provider typically engage in with the user has grown beyond the point of ordinary users' ability to memorize and recall the necessary access information. Thus, using the same identity credential to access network services or resources provided by any service provider has underscored the apparent ease and efficiency of the federated cloud IDMS and a possible solution is Single Sign-On (SSO). However, one of the main problems SSO poses is user privacy. User identity information can be shared with identity provider(s) and service providers, who can obtain the information from the identity provider. Thus, malicious identity and service provider can reveal users' identity information and activities. In this paper, we discuss a possible solution to preserve users' privacy in the federated cloud IDMS based on the SSO.

II. RELATED WORK

The SSO provides the ease of single authentication which subsequently allows users to become automatically logged into all other service providers within the same trust domain, thereby eliminating further manual interaction with the service provider. As such, SSO increases the overall usability of network. SSO system can be classified into pseudo-SSO and true SSO [5]. The pseudo-SSO component manages the service provider specific user authentication information. In the true SSO, the authentication service provider verifies the user. The authentication service provider must establish a trusted relationship with all service providers in order for the SSO to be achieved. Typically, the relationship may be established by a contractual arrangement. The federated cloud IDMS consists of a group of identity providers and service providers. Sharing user information and activities may result in further revealing by malicious providers and users have no control mechanism over disclosure of their identity information. Suriadi and et al. [6] have proposed a mechanism to provide user privacy by allowing users to enact some degree of control of their identity information. In the My Private Cloud project, Chadwick and et al. [7] proposed a trust based approach for federated access to cloud resources.

 ISBN: 978-1-61208-451-0

III. PROPOSED APPROACH

In our proposed approach, we assume that 1) there is a group of trusted third parties that provide identity services to users and service providers; 2) there are two levels of trusted identity providers; 3) identity service providers and actual service providers have a trusted relationship with each other; 4) proper encryption mechanisms are already in place to ensure secure transmission among users, identity providers and service providers. Exchanging security information must be based on the trustworthiness of users, identity providers, and actual service providers. In this paper, as a possible way of preserving users' privacy, we consider a group of identity service providers in a two-level hierarchical architecture. The higher layer (level 2) identity providers act as a central identity server and generate a security token (e.g., random number or nickname) based on the user's registration request. Only these providers know the user's actual identity information. This proposal also includes lower layer (level 1) identity providers for purposes of actual authentication of the authorized users based on the token provided by the user. When the level 2 identity provider receives registration, it generates a token that corresponds to the user's real identity, which in turn ensures the privacy of users. This token is then forwarded to the level 1 identity service provider to cooperate generating a certificate and security key(s) based on the token provided by the user. Thus, the level 1 identity provider does not have any information to identify the user's real identity. And the certificate itself does not reveal the user's real identification. The level 1 identity provider may send the generated certificate back to the level 2 identity provider that retains a record of the user's real identification if needed, while sending the corresponding certificate and security key(s) to the user without storing any relevant information. The level 2 identity providers should not be able to see the security key(s) and certificate. Furthermore, the level 1 identity provider only has information of the token, and the generated certificate and security key(s) based on the token. Government agencies need to be able to access both the level 2 and the level 1identity providers to obtain the user's real identity. Users will not be at risk of having their personal information revealed, even if they log on/off to access network services and resources from any service provider. Figure 1 illustrates the information flow in our proposed approach.

IV. CONCLUDING REMARKS

When we subscribe the new cloud service, it is important for users to firstly complete a registration to accrue a new set of credentials. Unfortunately, users tend to generate weak and easy to remember passwords, which can in turn cause users to easily become targeted by potential security attacks. Thus, one possible solution is to create single sign-on that allows users to authenticate just one time and then be automated for further authentication during the same login session. However, single sign-on caused another security issue, such as user privacy. This is because malicious identity service providers and actual service providers can cooperate to reveal and trace both the real identity and the activities of the user. Hence, in this paper, we proposed a way to mitigate the concerns regarding protecting the privacy of users in the federated cloud IDMS. We believe that the fruits of this work are evident, but remain in need of continued research.

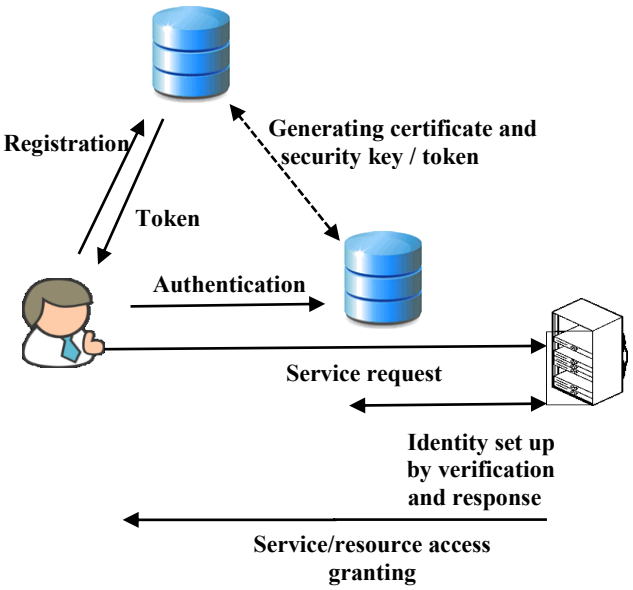

Figure 1. Information flow in privacy preserving SSO

REFERENCES

[1] Vince Lo Faso, "A practical view of NIST's cloud definition", *Global Knowledge*.

[2] M. S. Ferdous and R. Poet, " A comparative analysis of identity management systems", Proceedings of the International conference on high performance computing and simulation (Madrid, Spain, July 2-6, 2012), pp 454-461.

[3] U. Habiba, R. Masood, M. Shibli, M. Niazi, "Cloud identity management security issues and solution: a taxonomy", Complex Adaptive Systems Modeling. Vol. 2, No. 5.

[4] Y. Cao and L. Yang, "A survey of identity management technology", Proceedings of the IEEE International conference on information theory and information security (Beijing, China, December 17-19, 2010), pp 287-293.

[5] A. Pashalidis and C. Mitchell, "A taxonomy of single sign-on systems", Proceedings of the 8th Australasian conference on Information security and privacy (Berlin, Heidelberg: Springer, 2003), pp 249-264.

[6] S. Suriadi, E. Foo, and A. Josang, "A user-centric federated single sign-on system", Proceedings on IFIP International conference on network and parallel computing workshops (Liaoning , China, September 18-21), pp 99-106.

[7] D. Chadwick, M. Casenove, and K. Siu., "Security APIs for My Private Cloud – Granting access to anyone, from anywhere at any time", Proceedings of the IEEE conference on cloud computing technology and science (Athens, Greece, 2011), pp 792-798

 ISBN: 978-1-61208-451-0

Identifying Requirements for a Social Media-based Emergency Management System

Marko Jäntti, Taina Kurki
School of Computing
University of Eastern Finland
P.O.B 1627, 70211 Kuopio, Finland
Email: {marko.jantti, taina.kurki}@uef.fi

Laura Hokkanen
Emergency Services College
P.O. Box 1122, FI-70821 Kuopio, Finland
Email: laura.hokkanen@pelastusopisto.fi

***Abstract*—Social media tools provide public safety organizations with a new communication channel for the management of crises and emergencies. While existing studies have mainly focused on exploring how authorities use social media tools, citizens role in producing information for emergency management has not been studied adequately. The research problem of this study is: How social media tools support emergency management from a citizens perspective? The main contribution of this paper is to identify requirements for social media -based emergency management systems. Data on requirements were collected during an a emergency management exercise in Finland. Results indicate that citizens consider authorities presence in social media valuable and reassuring during emergency situations.**

***Keywords*—Emergency; social media; public safety.**

I. INTRODUCTION

Social media has rapidly become one of the daily used communication channels. At the same time citizens are more and more equipped with mobile technologies that enable them to be constantly interactive in online social media networks, anytime and almost anyplace. This online behavior is also present in emergencies, where an emerging trend of growing citizens participation through social media can be seen. Both citizens involved in and outside the emergency provide and seek information with implications for both the informal and the formal response effort.

Social media therefore provides opportunities for engaging citizens in response activities both by pushing information to the public and by pulling information from those involved in the emergency, enabling online exchange of information through conversation and interaction [1]. Public Safety Organizations have recently started to discuss how this citizen provided resource could be best utilized in emergency response. Large scale natural disasters, such as earthquakes trigger the need for more agile and scalable emergency communication channels [2].

Public authorities in emergency management domain in Finland include: police, rescue services, municipalities and emergency response centre. In this study, we call these organizations Public Safety Organizations (PSOs). Public Safety Organizations in Finland now face the demand on being present online and there is also (more or less so) shared understanding in the field of emergency response that public authorities should be there where the citizens are, including social media [3], [4]. All of the 22 regional rescue departments in Finland have at least a Facebook account, some use also Twitter. Municipalities play an important communication-related role in Finnish emergency management model.

The usage of social media is, however, still taking its first steps. Communications in social media is mainly dissemination of safety information for citizens, and utilization of the information provided by citizens during crisis and emergencies, such as silent signals, first-hand reports, (geotagged) photos and video, or interactive communication between citizens and public safety organizations, is still meagre. Research on the use of social media in the field of emergency management in U.S. and UK have provided similar findings [5], [6], [7].

Studies on the theme show that one of the key barriers for the adoption of social media tools for emergency management is the lack of related policies and guidelines [8]. These studies have focused mainly on the adoption of social media enabled communication from the point of view of public authorities, focusing less on guidelines for citizens on the use of social media in emergencies and crises.

Additionally, there are studies that have explored social media tools in information sharing of disasters [9], [10] and how social media has been used in disaster management life cycle: prevention, preparedness, response, and recovery [11]. Carter [12] has studied how social media has been used in storm situations in US. Some of the studies have focused on challenges of using social media [13]. As a new mode of communication, its adoption shall likely cause change resistance among emergency management employees.

This study is related to EU Framework Programme funded research projects iSAR+ and SOTERIA that aim at producing recommendations to leverage the use of online and mobile social media in crises and emergency response efforts. In this paper, we focus on a citizen's viewpoint of emergency management. The study reports findings from a user showcase on social media in emergency management. The purpose of the showcase was to study the effectiveness of the interactive communication between the citizens and PSOs in crisis situations through three scenarios: hazardous material accident, aviation accident and a summer storm. These scenarios provided a fruitful playground for studying the bidirectional communication between the citizens and Public Protection and Disaster Relief organizations (PPDRs) in social media as well as testing the social media data analytics platform.

The remainder of the paper is organized as follows. In Sec-

 ISBN: 978-1-61208-451-0

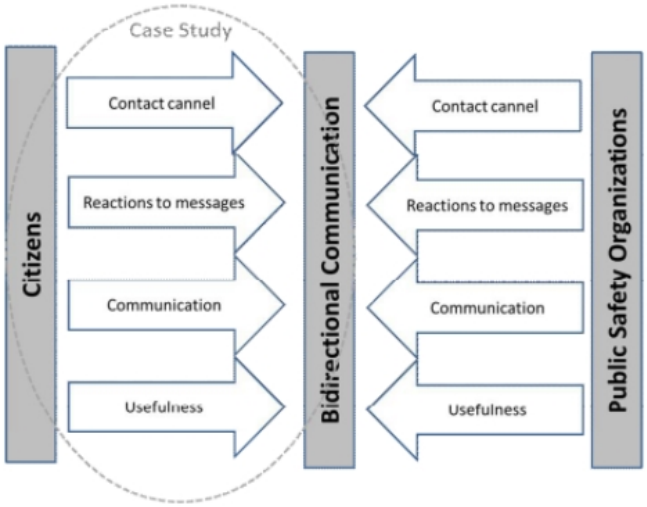

Fig. 1. The context of the study

tion 2, the research method and settings of this are described. In Section 3, the results of the study are presented. Section 4 includes analysis of results. Conclusions are given in Section 5.

II. Research Problem & Methodology

The case study research method was used to explore social media-based emergency management from a citizens perspective. Yin [14] defines a case study as an empirical inquiry that investigates a contemporary phenomenon within its real life context using multiple sources of evidence. The showcase was conducted in controlled environment in order to avoid misunderstanding and panic in social media. The closed environment also enabled protecting participants privacy and anonymity.

A. Research environment

The Finnish user showcase (on social media in emergency management) was organized along with Emergency Services Colleges Crises and Large Scale Emergencies exercise (Krisu exercise). This exercise is carried out twice a year and it is a part of studies of Rescue Activities Management of the graduating course of fire sub-officers and fire officer students.

The aim of the Krisu exercise is to practice management of crises and large scale emergency situations in which several public authorities take part, and in which rescue services is the authority responsible for the overall management of the situation. Several public safety authorities took part in the showcase. In the showcase, the settings of Krisu exercise were utilized bringing new tools for the players to support communication and crisis management. The context of the study is presented in Figure 1.

Showcase description: The showcase consisted of three simulated scenarios: Hazardous material accident in Finland nearby Finnish Russian border area, where a tank wagon of a freight train containing ammonium was leaking at the rail yard. The second scenario was an aviation accident in Helsinki-Vantaa airport: Airbus 320's landing fell short 200 meters before runway. The third scenario was a storm, during which few hundred tasks were reported to the emergency number 112, one of them a traffic accident with two cars and a tank truck. During a storm scenario a detection of an emergency without a call to 112 was also tested: lightning / damaged power started a fire in a building.

Fig. 2. A photo of an emergency situation tweeted by a citizen

Facilities: The showcase was organized in two different locations: Command and Control Center positioned in facilities of Emergency Services College. Citizen players were located in Korvaharju training ground that is a 36-hectare wide area with over 100 training and testing facilities for fire fighters.

B. Data collection and analysis

Data collection was carried out during the showcase in February 2015. The following data collection sources were utilized: documentation (showcase reports from previous showcases, showcase plan), archival records (iSAR project survey for citizens, tweets sent by citizens during a showcase), participative observation (participation in organizing the showcase, observation reports by citizen players), interviews and coffee table discussions (collecting information during the showcase from citizen players), direct observation (observation of Twitter feed and citizens action), physical artefacts (airplane, train, tank truck simulators (for rescue services training purposes), and feedback survey (directly after showcase and another after showcase).

A within case analysis technique was used in this study [15]. The technique focuses on exploring the case as a standalone entity. The showcase was also part of validation of results. In the analysis, we utilized a pattern matching analysis technique that focuses on comparing empirically based patterns with predicted patterns. Predicted patterns (categories) were created based on previous studies on emergency management and evaluation reports of the previous showcase. Categories focused on citizens viewpoint in management of emergencies. Finally, for each category, we aimed to establish subcategories that described the findings.

Settings: The study has been conducted as a part of EU Seventh Framework projects SOTERIA (Online and Mobile Communications for Emergencies) and iSAR+ (Online and Mobile Communications for Crisis Response and Search and Rescue). Pictures and data were collected from the showcase of February 2015 in training area (see Figure 2).

 ISBN: 978-1-61208-451-0

TABLE I. FINDINGS OF THE STUDY (SOCIAL MEDIA AS A CONTACT CHANNEL)

Id	Citizens' comments	Subcategory
1.1	If I had an urgent need, I would primarily call 112. But, of course, the more communication channels the PSOs use, the more useful it would be in emergencies.	Social media as an add-on channel
1.2	Through social media, one can receive much more information on emergency situation, compared to phone calls.	Rich information source
1.3	It should go in this way also in real life! Immediate response to the tagged tweet.	Fast response
1.4	Contacting the PSOs via social media felt a bit difficult and unclear.	Challenges in contact
1.5	I sent a tweet with a picture and location data (GPS). The response came within a minute as a private message .	Private message as a contact channel
1.6	PSOs could send their contact details as a private message (a direct phone number).	Private message as a contact channel
1.7	I wonder why I cannot use SoMe as a contact channel at the moment.	Social media as an add-on channel

iSAR+ project aimed at research and development of set of guidelines that enable citizens using new mobile and online technologies to actively participate in the response effort, through the bidirectional provision, dissemination, sharing and retrieval of information essential for critical Public Protection and Disaster Relief (PPDR) intervention, in search and rescue, and medical assistance. On-line and Mobile Communications for Emergencies (SOTERIA) project will result to developing recommendations, to provide guidelines and courses of action for Public Safety Organizations and citizens on how social media tools are best utilized to benefit both citizens and public authorities in emergencies. The participants consisted of Finnish end-user community, such as national PPDRs entities (Rescue Services, Police, municipalities), and citizen players.

III. RESULTS

Results from the case study are presented according to the following categories: 1) Social media as a contact channel, 2) Reaction of citizens to authorities messages and other citizens messages (reliable and reassuring), 3) Communication (message content, reaction time, information clarity, right timing, unambiguity), and 4) Usefulness of social media tools.

Social media as a contact channel. Table I shows the main findings related to social media as a contact channel.

Citizens addressed the role of social media as an additional, value-adding communication channel and emphasized the diversity of social media channel in relation to the traditional contact channel, phone. Citizen players were not trained or told to use private Twitter messages during the exercise. However, several citizen players used this for private communication with authorities.

TABLE II. FINDINGS OF THE STUDY (REACTION OF CITIZENS)

Id	Citizens' comments	Subcategory
2.1	There was a large number of unnecessary tweets and thus, I started to think whether emergency contacts can be easily identified among junk tweets.	Identification of essential information
2.2	Yes. Decreases speculations. In real situations I would read messages but would only trust messages from PSOs.	Trust
2.3	Social media could surely be used also in real situations but mining the accurate information from the false and unnecessary information is relatively hard.	Identification of essential information
2.4	The citizens tweeted various information and I really could not connect with authorities. However, the announcements from the PSOs felt reassuring .	Authorities presence in social media
2.5	The PSOs were actively present in Twitter, and they published some information this was good.	Authorities presence in social media

Citizens comments also showed the perception of social media as a bidirectional communication channel by nature. Contact with the PSOs was expected. Nevertheless, even when a personal contact with PSOs did not occur, the presence of PSOs and their messages in social media was found important. Regarding PSOs viewpoint, we observed during the study that authorities proceed carefully in their social media initiatives. This might be due to unfamiliarity of social media use in PSO communication. Additionally, the persons who deal with emergencies are not communication specialists.

Reaction of citizens to the messages of authorities and other citizens. Table II shows the main findings related to citizens reactions to communication of authorities and other citizens during the showcase. Some of the authorities messages included too generic language. For example, if they refer to a press release, it would be nice to get the webpage where it appears. (IN) Bi-directional communication through social media is surely useful for public authorities.

Citizens' reactions related to social media usage by PSOs was positive. Information provided by the public authorities through social media was considered trustworthy. In this study, the impact of the PSO communication to the citizen behavior was not explored, but the trust in PSO messages indicates that e.g. giving instructions or codes of conduct in social media would lead to preferred action. Citizens reported that PSOs participation in social media discussions was active but also highlighted that information was inconsistent between different data sources.

Communication. Table III shows our findings related to communication between PSOs and citizens.

Our findings emphasized the importance of social media literacy skills (skills required in reading social media feeds and understanding common terminology used in tweets and updates). Capturing essential information) from the flow of tweets (some were true, some spam and some misleading)

 ISBN: 978-1-61208-451-0

TABLE III. FINDINGS OF THE STUDY (COMMUNICATION)

Id	Citizens' comments	Subcategory
3.1	It would be useful to advise people to avoid accident area as early as possible	Proactive communication
3.2	I think interactive communication between citizens and PPDRs is useful as long as the tweets from PPDRs are clear, open and fast enough.	Quality of communication
3.3	It was good that authorities reported only the essential information such as number of injured persons	Quality of communication
3.4	Does the tank truck have warning signs of Hazardous material? If yes what kind of signs?	Emergency details
3.5	I sent a tweet with a picture and location data (GPS). Now, authorities were well aware and the response came within a minute as a private message	Private message as a contact channel
3.6	PSOs could send their contact details as a private message (a direct phone number)	Private message as a contact channel
3.7	I wonder why I cannot use SoMe as a contact channel at the moment.	Social media as an add-on channel

TABLE IV. FINDINGS OF THE STUDY (USEFULNESS OF TOOLS)

Id	Findings	Subcategory
4.1	It would be good if the PSOs used these tools in real situations.	Need for improvement
4.2	The closed Twitter environment worked almost perfectly.	Retweeting
4.3	The main idea of hashtags was not clear to everybody	Hashtags
4.4	This system would work in real situations! I would use it and I would know who to inform. Might create spam.	Spam filtering
4.5	I received a good understanding of train accident only based on following Twitter feeds.	Progress of emergency handling

was seen to require skills on media literacy (for social media), especially when considering the reliability of the information provided on social networks.

PSO communication should be timely, fast and active. Some citizen players considered authorities response too slow in the exercise. Authorities explained that response takes time due to validation of information provided by citizens. In most cases the validation requires a field visit. One solution could be that authorities first message is that the situation is under investigation or some generic information.

Citizens would also have liked the PSOs to define hashtags for the ongoing situations. During the showcase, citizens communication involved tweets, private messages, video content, and images related to emergency situation. Authorities considered this information useful because it provided details of accident scene such as number and type of cars, color of smoke and warning signs in a tank truck.

Usefulness of social media tools for emergency management. During the showcase, the following tools were used by PSOs: the iSAR+ alerting services (My Public Alerts), iSAR+ fusion centre services (social media monitoring, SMM crawler, text analysis tool), and iSAR+ portal software. A closed Twitter-based platform was used to manage communication between citizens and PSOs. Three citizen players used Permiloc application that enabled locating the citizen and receiving real-time messages. (e.g., evacuation commands, progress notifications on emergency situations) from PSOs. The showcase scenarios included the following data:

- Real time messages (e.g., tweets, photos and video streams of the events, tweets with geolocation data) generated by the players.
- Photos from real emergency situations provided by Emergency Services College.
- Updates created based on previous, real life communications during emergencies
- Rumours and trolls produced intentionally by the citizen players.

Table IV describes the findings concerning how useful citizens considered social media tools in emergency management.

IV. ANALYSIS

According to our findings, social media is a very promising communication channel for emergency situations. However, emergency services field needs guidelines and policies to establish and maintain these channels. Social media might be used as an additional channel to provide richer information on accident scenes than a phone-based channel. We observed that Twitter-based social media tool was easy and fast to use from a citizens perspective, even with no previous experience.

The research problem of this study was: How social media tools support emergency management from a citizens perspective? As main contribution of the paper, we showed the citizens' perspective to the social media-based emergency management. The following **requirements for social media -based emergency management systems** can be derived from our results:

Fast response to citizens' social media updates and messages. Citizens require immediate and almost real time response to their messages in social media. During the showcase, we observed that citizens would require some kind of response on emergency event from authorities in early phase. Authorities' perspective seems to be to first validate the event before commenting anything. (Findings 1.2, 1.3)

Support for public and private messages. In addition to public communication (tweets), some of the citizens tested direct communication with PSOs and this type of tailored communication was considered very useful and effective. We

 ISBN: 978-1-61208-451-0

observed that before a citizen is able to send a direct message to PSOs, he/she has to be a follower of the PSO role. From a citizen's perspective, opening an emergency chat channel from a hyperlink could be perhaps simpler way to open a direct communication. (Findings 1.4, 1.6, 3.5, 3.6)

Hashtags. Hashtags enable tagging and linking tweets on the same subject together. Some citizen players suggested that social media based emergency system could recommend citizens which hashtags should be used during the emergency situation. We also observed that the role of hashtags remained unclear for some citizens. (Finding 4.3)

Effective and reliable location based services. During the showcase, location based alerting systems (citizen players were using a mobile application) were tested. We observed that in some cases, inside the main building (of training area) the location-based alerting system did not work reliably because the mobile device and application lost the GPS signal.

Sharing information in periodic intervals scheduling. There were several citizen players that were asked (according to a script) to request information from PSOs during the showcase. Often, citizens ask simple questions related to who they should take contact with or where to get medical treatment. A certain number of these questions could be avoided by scheduling functions of social media tools. For example, a tool could be set to automatically post contact details of a hospital every 15 minutes. (Findings 2.1, 2.4)

Validation and analysis of emergency events. Especially, management of trolls provides social media tools with challenges. Image and video analysis tools should be used to check whether images and videos are real and have been taken from a correct place. Validation and analysis should be conducted not only reactively, but also proactively, for example scanning guns from images and analysing landslides from satellite images. During our showcase, some citizen players complained about slow response of PSOs. However, there were also citizens that had received immediate response to their tweets. (Findings 2.4, 2.5)

Educating citizens on what type of information is needed from accidents. The social media -based emergency management system should inform citizens what type of information is most useful for PSOs. For example, in hazardous material accidents, images on warning signs of container trucks and the colour of smoke help PSOs to identify what type of chemical the container is leaking. (Findings 3.4, 3.5)

The showcase results showed that social media enables new kind of citizen engagement in emergencies and crises. Disseminating information, providing eye-witness accounts, sending and receiving alert messages, exchanging experiences of the emergency response and searching and publishing information related to the event were conducted during a showcase in a closed social media platform. We observed that social media platform established strong real-time collaboration mechanisms that led to improved situational awareness of emergency situations.

Social media has been successfully used in several natural disasters to connect people and inform authorities of the scope of disasters/emergency situations. Our findings are congruent with previous studies that emergency management organizations still lack consensus on how the social media can be utilized in sharing emergency-related information. Negative impacts of social media use could be partly overcome by educating the citizens how they should or can act in social media channels. Our results showed that citizens would be willing to help if they knew what type of information authorities need in managing emergencies. Our study also addressed the role of rumor (troll) management. More effective data analytics capabilities of emergency management systems might speed up the validation of emergencies and lead to faster identification and more accurate response to trolls.

As implications for theory, our study contributes to emergency management theory by focusing on citizens viewpoint and exploring how they consider the use of social media as a contact channel and means of communications. Our main categories (social media as a contact channel, reaction of citizens to authorities and other citizens messages, communication, and usefulness of tools) and identified subcategories can be used by future social media case studies on emergency management.

As implications for practice, we validated the social media based emergency management approach with three accident scenarios and collected information on the limitations of social media tools and user experiences.

V. CONCLUSION

This study aimed to answer the research problem: How social media tools support emergency management from a citizens perspective? The main findings showed that the social media-based platform used in the study was easy to use and enabled the PSOs to find the most essential information from the information sent by citizens in social media. According to the showcase observations, using social media tools enhances the situational awareness of PSOs by providing information (photos, videos and tweets from emergency sites) directly from the citizens involved in the crisis.

However, some of the feedbacks collected from citizens indicated that implementation of these kind of new practices requires time. Using common hashtags in updates about the ongoing situations would have been useful; in this exercise, the citizen players thought that the PSOs could have recommended some hashtags. Citizen participants were worried that the important updates of the PSO might not be noticed among the flood of tweets. Spam and unreliable information was seen as a disadvantage of using social media during crisis situations.

Following social media and in this case the flow of tweets of which some are true, some spam and some trolls requires skills on media literacy, especially when considering the reliability of the information provided on social networks. During the scenarios demonstrated there was genuine need of official information, as it there would be in real life situations.

Pictures were found useful information, also from the point of view that they are helpful for foreigners who do not know the language used in the tweets. Although social media tools are actively used among involved citizens, our observations and interviews showed that utilizing them in bidirectional emergency management situations is relatively new.Feedback from the showcase and the earlier survey indicated that citizens

 ISBN: 978-1-61208-451-0

trust authorities. This is not obvious in other countries. The citizens considered the presence of PSOs in social media and the bidirectional communication valuable and reassuring during and after a crisis. Our results indicate that social media could be an additional communication channel between public safety organizations and citizens.

This case study contains certain limitations. First, data were collected during a relatively short time period (2 days incl. a showcase and a feedback workshop). However, we managed to get a rich set of material for the analysis through using six sources of evidence recommended by Yin. Second, we used a single case design in this study. We agree that this causes difculties in generalizing results. However, case studies should not aim at generalization but extending the theory. Third, regarding the construct validity, most of the players were relatively young (students) that use social media more actively on their free time than elderly people. Finally, the fact that the showcase was a simulated exercise certainly affects the results.

Further research will continue in Soteria project. Barents Rescue Exercise 2015 will be a starting point for campaigns of experimentations, where the role and impact of social media and related tools will be tried, experimented, analyzed, tested and evaluated in real and realistic simulated environments. This will be instrumental to the production of recommendations on the understanding of the impact and role of social media in emergencies and efficient and effective ways to incorporate mobile technology and social media into emergency response efforts.

Further research efforts could focus on exploring the options how social media based emergency platforms should be organized and provided as a viable business model. Rescue services could apply service management approach for managing social media based emergency services.

ACKNOWLEDGMENT

This study is funded by the European Union Seventh Framework Programme (FP7/2007-2013) under the Grant Agreement n 606796 — Soteria. We would like to thank SOTERIA project partners for their valuable contribution to emergency management research and the Finnish showcase.

REFERENCES

[1] R. Merchant, S. Elmer, and N. Lurie, "Integrating social media into emergency-preparedness efforts," *New England Journal of Medicine*, vol. 365, no. 4, pp. 289–291, 2011.

[2] J. L. Wybo, and M. Latiers, "Exploring complex emergency situations dynamic: Theoretical, epistemological and methodological proposals tarkista vuosi," *International Journal of Emergency Management*, vol. 3, no. 1, pp. 40 – 51, 2015.

[3] L. Hokkanen, K. Pylväs, T. Kankaanranta, N. Päivinen, and T. Kurki, "Using social media in emergency and crisis situations viewpoints of authorities and citizens (in finnish)," Poliisiammattikorkeakoulun katsauksia 1/2014, Poliisiammattikorkeakoulu, Tech. Rep., 2014.

[4] K. Pylväs, L. Hokkanen, and T. Kankaanranta, "Social media and mobile technology in communication between public safety authorities (in finnish)," Project final report, Tech. Rep., 2015.

[5] Y. San Su, C. Wardell III and Z. Thorkildsen, *Social Media in the Emergency Management Field. 2012 Survey Result*. CNA Analysis and Solutions, 2013.

[6] S. R. Hiltz, J. Kushma and L. Plotnick, "Use of social media by u.s, public sector emergency managers: Barriers and wish list," in *Proceedings of the 11th International ISCRAM Conference*. USA: Pennsylvania State University, 2014, p. 600609.

[7] C. Wendling, J. Radisch and S. Jacobzone, "The use of social media in risk and crisis communicationl," OECD Working Papers on Public Governance, No. 25, OECD Publishing, 2013.

[8] R. Beneito-Montagut, S. Anson, D. Shaw and C. Brewster , "Governmental social media use for emergency communication," in *Proceedings of the 10th International ISCRAM Conference*. Germany: Karlsruhe Institute of Technology, 2013, p. 15.

[9] T. Sakaki, F. Toriumi, K. Uchiyama, Y. Matsuo, K. Shinoda, K. Kazama, S. Kurihara, and I. Noda, "The possibility of social media analysis for disaster management," in *Proceedings of the Humanitarian Technology Conference (R10-HTC), 2013*. USA: IEEE, 2013, pp. 238,243.

[10] Y. Tyshchuk, and W. A Wallace, "Actionable information during extreme events – case study: Warnings and 2011 tohoku earthquake," in *Proceedings of the 2012 International Conference on Social Computing (SocialCom)*. USA: Academy of Science and Engineering, 2012, pp. 338–347.

[11] B. Abedin, A. Babar, and A. Abbasi, "Characterization of the use of social media in natural disasters: A systematic review," in *Proceedings of the 2014 IEEE Fourth International Conference on Big Data and Cloud Computing (BdCloud)*. USA: IEEE Computer Society, 2014, pp. 449–454.

[12] L. Carter, J. B. Thatcher, and R. Wright, "Social media and emergency management: Exploring state and local tweets," in *Proceedings of the 47th Hawaii International Conference on System Sciences (HICSS)*. USA: IEEE, 2014, pp. 1968–1977.

[13] S. Luna, and M. Pennock,, "Social media in emergency management advances, challenges and future directions," in *Proceedings of the 9th Annual IEEE International Systems Conference (SysCon)*. USA: IEEE, 2015, p. 792797.

[14] R. Yin, *Case Study Research: Design and Methods*. Beverly Hills, CA: Sage Publishing, 1994.

[15] K. Eisenhardt, "Building theories from case study research," *Academy of Management Review*, vol. 14, pp. 532–550, 1989.

 ISBN: 978-1-61208-451-0

Performance Evaluation of the new TUAK Mobile Authentication Algorithm

Keith Mayes
Information Security Group
Royal Holloway, University of London
Egham, UK
keith.mayes@rhul.ac.uk

Steve Babbage
Vodafone Group R&D
Vodafone Group Services Ltd.
Newbury, UK
steve.babbage@vodafone.com

Alexander Maximov
Ericsson Research
Ericsson
Lund, SE
alexander.maximov@ericsson.com

***Abstract*—TUAK is a new mutual authentication and key generation algorithm proposed by the Security Algorithm Group of Experts (SAGE) of the European Telecommunications Standards Institute (ETSI) and published by the Third Generation Partnership Project (3GPP). TUAK is based on the Keccak sponge function which has very different design principles to the pre-existing 3G MILENAGE algorithm and so promises a back-up/alternative in case algorithm vulnerabilities are discovered during long-term Machine-to-Machine (M2M) deployments. However, the practicality of implementing TUAK on currently deployed and/or future Subscriber Identity Module (SIM) cards is not well known. This paper describes the initial work and findings of a study in support of SAGE and GSMA to consider such implementation aspects.**

Keywords–3GPP; GSM; Keccak; SAGE; TUAK.

I. INTRODUCTION

The European Telecommunications Standards Institute (ETSI) [1] and later the Third Generation Partnership Project (3GPP) [2] standardised mobile networks so that Mobile Network Operators (MNO) were able to choose/design their own cryptographic algorithms for subscriber authentication and session key generation. In GSM, [3] there is a proliferation of algorithms, however for 3G most MNOs use the well-studied and openly published MILENAGE algorithm [4]. MILENAGE (AES [5] based) was designed and published by the ETSI Security Algorithms Group of Experts (SAGE), and more recently SAGE designed a second algorithm, called TUAK [6] based on the Keccak [7] sponge function. This was done for two main reasons. Firstly, although MILENAGE is currently considered strong, industry should have a proven alternative in case an advance in cryptanalysis exposes vulnerability. Secondly, machine-to-machine (M2M) devices will use "embedded SIMs", whereby a Subscriber identity Module (SIM) chip is fitted into a device, and the assignment (or re-assignment) to a MNO and the provisioning of security credentials is done later, over the air. Some devices may be deployed for at least twenty years, which is a considerable time in the life of a technical security solution. Having two strong algorithms (MILENAGE and TUAK) built into the hardware, and available for selection, should give good assurance that effective security can be maintained throughout the SIM lifetime.

TUAK inherits most of its security characteristics from Keccak, which is the winning SHA-3 design and has of course been extensively studied. See [8][9] for a closer analysis of TUAKs security. TUAK is fundamentally different from MILENAGE in its design, so that an advance in cryptanalysis affecting one algorithm is unlikely to affect the other. There are very few academic publications around TUAK as the standards are quite new, although a comprehensive security assessment [10] of the TUAK Algorithm Set was carried out by the University of Waterloo, Canada. It considered a wide range of cryptanalysis techniques, and finally concluded that TUAK can be used with confidence as message authentication functions and key derivation functions. However, industry acceptance and adoption of TUAK requires not just a secure design, but also confidence that it can be implemented on limited resource SIMs with sufficient performance.

- Is it possible to load the algorithm onto an existing deployed or stocked smart card platform?
- If so, will the algorithm run with acceptable performance?
- Will a new SIM require a crypto-coprocessor for adequate performance?
- Will a new SIM need to have a high performance processor (e.g., 32-bit type)?
- Will a new SIM require specialist low-level software for the algorithm?
- Will the algorithm benefit from hardware security protection?

There have been previous performance evaluation and comparisons [7][11][12], around the Keccak core for the SHA-3 competition [13], however these were aimed primarily at specialist hardware, or far more powerful and less memory limited processors than are typically found in SIMs. Therefore, at the request of SAGE, the evaluation described in this paper was undertaken, in which the entire TUAK algorithm performance was determined by experiment with the SAGE specified settings for Keccak, using their published source code as a starting point. The latter is important, as SIM vendors tend to base their implementations on the published security standards examples. In addressing the performance questions it was necessary to define a method of experimentation that would give relevant results yet would not be tied to a particular processor, platform or optimised for particular chip features. The work began with the PC example implementations, before forking to a parallel development suited for smart card evaluation. For the latter, simulation was originally considered, however it is difficult to map results to real card performance. The use of a multi-application card platform was included as a positive means of abstraction from any particular chip, and

 ISBN: 978-1-61208-451-0

could be representative of loading the algorithm onto existing/stock SIMs. However, the performance of such platforms (e.g., MULTOS [14]/Java Card [15]) is usually inferior to a native card implementation and so native mode was included as the principal benchmark.

In Section II an overview of TUAK is provided before describing the experimental setup and software development in Section III and Section IV. Results are presented in Section V and analysed in Section VI. Some comments on security defences and performance are discussed in Section VII and finally, conclusions and future work are presented in Section VIII.

II. TUAK Overview

In each of GSM/GPRS (2G), UMTS [16] (3G) and the Long Term Evolution (LTE 4G), a fundamental part of the security architecture is a set of authentication and key agreement functions [17][18]. The set of functions varies between generations, with 3G providing more security than 2G, and 4G adding some further refinements. These functions exist in the subscriber's SIM card (which is provided by their MNO), and in a network node called the Authentication Centre (AuC) that is run by the MNO. The 3G authentication and key agreement architecture requires seven cryptographic functions. MILENAGE [4] is a complete set of algorithms to fulfil these functions, built from a common cryptographic core (the AES block cipher) using a consistent construction. TUAK [6] is also a complete set of cryptographic functions for 3G authentication and key agreement. LTE security reuses the same set of functions, so both MILENAGE and TUAK can also be used for LTE. There is also a standardised method for using the 3G authentication and key agreement functions in GSM/GPRS.

A. Algorithm Inputs and Outputs

Whereas MILENAGE was designed only with 3G in mind, TUAK was also designed for LTE and so supports a 256 bit subscriber-unique secret key as well as the 128 bit key size used in 3G. Moreover, TUAK also allows for the possibility that certain other input or output parameters might increase in length in the future. The input and outputs of TUAK's seven cryptographic functions *f1, f1*, f2, f3, f4, f5 and f5** are defined in [6] and like MILENAGE, the TUAK algorithm-set expects one additional input parameter, an "Operator Variant Algorithm Configuration Field". In the case of TUAK, this field is called TOP and is 256 bits long; each mobile operator is expected to choose its own value for this, typically the same value for many SIMs. The 3GPP security architecture did not require this extra parameter, but it was included for two main purposes:

- SIMs for different MNOs are not interchangeable, either through trivial modification of inputs and outputs or by reprogramming of a blank SIM.
- By keeping some algorithm details secret, some attacks (such as side channel attacks like power analysis) become a *little* harder to carry out.

TUAK includes an algorithm to derive value TOP_c from TOP and the secret key K, and it is sufficient for the SIM card to be programmed with TOP_c rather than with TOP itself. This means that an attacker who is able to extract TOP_c from one card does not learn TOP or TOP_c for other cards.

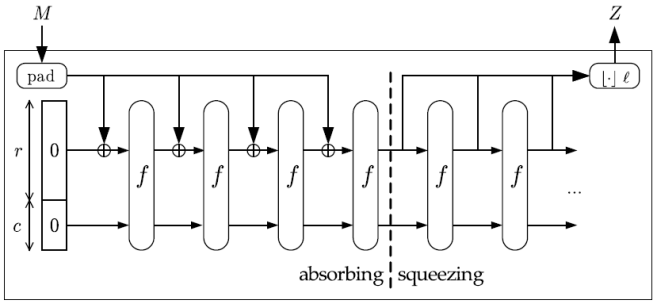

Figure 1. A Cryptographic Sponge Function

B. Algorithm Building Blocks

The main building block from which all of the TUAK algorithms are constructed is Keccak [7], the "cryptographic sponge function" which was selected by NIST as the winner of the SHA-3 hash function competition [13]. Sponge functions work by repeated application of a fixed length transformation or permutation *f*, as shown in Figure 1, which is copied from [19]. First the input bits are "absorbed", and then the output bits are "squeezed out".

TUAK uses the Keccak algorithm with permutation size n = 1600, capacity c = 512 and rate r = 1088. This rate value is big enough that each of the algorithms in the TUAK set needs only a single instance of the permutation f - repeated iteration of the permutation is not necessary.

Details of the TUAK algorithm can be found in [6], with test data in [20][21]. The TUAK algorithm functions are illustrated in Figure 2. In this diagram:

- The top picture shows how TOP_c is derived from TOP.
- The middle picture shows how MAC-A or MAC-S is computed (*f1* and *f1**)
- The bottom picture shows how RES, CK, IK and AK are computed (functions *f2, f3, f4, f5* and *f5**) - note that these functions all take exactly the same set of input parameters, so can be computed together
- INSTANCE is an 8-bit value that takes different values for different functions, for different input and output parameter sizes, and to distinguish between *f1* and *f1** and between *f5* and *f5**, providing cryptographic separation
- ALGONAME is a 56-bit ASCII representation of the string "TUAK1.0"
- The block labelled "Keccak" is the 1600-bit permutation, with the shaded part corresponding to the 512-bit "capacity" input; see Figure 2.

III. The Experimental Setup

Based on the arguments presented in the introduction, the goal was to use a combination of PC software, native smart card chip implementations and a secure platform for development and comparative testing. For the native implementations, we required two chips of comparable CPU power, yet different security protection to determine if the inherent protective measures impacted performance. Furthermore, to

 ISBN: 978-1-61208-451-0

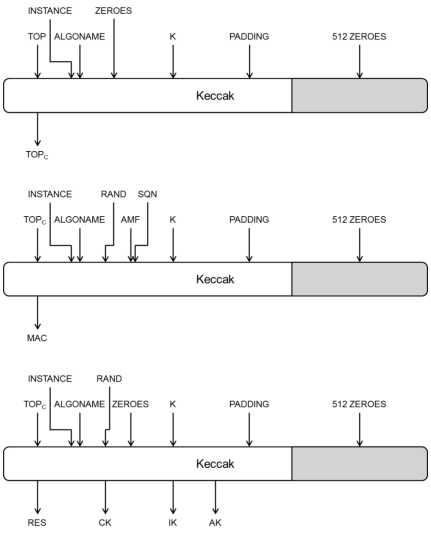

Figure 2. The TUAK Algorithm Functions

make useful comparisons with the secure platform implementations, we needed platforms based on similar chips. A solution presented itself based around native implementations on the Infineon SLE77 [22] and SLE78 [23]. The MULTOS platform was selected as the secure platform primarily because test cards (types M3 and M4) were available based on the same Infineon chips. The smart card experiments were preceded by measurements on a PC platform that used similar example C code. The code could in future also be ported to Java Card platforms, although the Java coding language would make comparisons less clear.

A. The PC Test Platform

The initial tests used an Intel Core i5-2540M CPU @ 2.60GHz, max turbo frequency 3.30GHz, 2 cores, 4 threads, Intel Smart Cache 3Mb, Instruction set 64-bit + AVX, 4Gb RAM, with a Windows 7 32-bit OS. The Keccak example implementations were written in C and compiled to optimise speed. Although the processor and the compiler supported 64-bit integers, the resulting assembly code was limited by the OS to 32-bit. Execution time was measured in CPU clock cycles, although multiple runs were necessary due to the multi-tasking OS interrupting execution. Various versions of the example code became available during development as shown in Table I. The smart card source code was originally modelled on version 1 and then developed in parallel.

In Keccak, f is a permutation. Keccak is a family of algorithms, from which a particular algorithm is selected by setting three security parameters:

- The permutation size n, which can be 25, 50, 100, 200, 400, 800 or 1600 bits.
- The "capacity" c, which is a security parameter (essentially, for a given capacity c; Keccak is claimed to stand any attack up to complexity $2^{c/2}$).
- The "rate" $r = n - c$, which determines how many input and output bits can be handled by each iteration of the permutation.

TABLE I. PC EXAMPLE CODE IMPLEMENTATION VERSIONS

Version	SupportedBits	ShortDescription
0	8/16/32/64	Size optimized, generic, use of % and more tables
1	8/16/32/64	Speed optimized, generic
2	64	Use of CPU 64-bit rotate instruction
3	8/32/64	Original from the specification
4	64	Similar to v2 but trying to combine more operations
5	32	Totally unrolled version, only C code
6	8/16/32	With bit-interleaving, generic, not optimized
7	32	Optimized bit-interleaving, part unrolled, 32-bit

B. The Smart Card Chips

The chips for experimentation both had 16-bit CPUs, which is a size representative of the majority of deployed SIMs (although there are still 8-bit CPUs around, as well as newer 32-bit CPUs). Whilst they are of similar family, horsepower and vintage they are quite different in security aspects.

1) SLE77: The SLE77 is a traditional style security controller intended for mid-range payment applications, and evaluated to Common Criteria [24] EAL5+. Its crypto-coprocessor does not support TUAK/Keccak so was not used in our tests. Details of the chip protection measures against physical, side-channel leakage and faults are not publicised, however in a traditional security chip one might expect protective shields, plus power smoothing and noise insertion to counter power analysis, and sensors/detectors to counter fault attacks. Some protection may arise from the application and OS software e.g., randomised/repeated operation and dummy cycles, although this may be optimised for the included algorithms. For a new algorithm running on this chip, we should expect some protection from the hardware, although the final algorithm code will need to improve this, which would likely degrade the performance measured in our experiments.

2) SLE78: The SLE78 is an innovative security controller intended for high security applications. Instead of relying mainly on shields and sensors it uses "Integrity Guard" [25], which exploits dual CPUs working in tandem. The claimed features include:

- Dual CPU implementation for fault detection
- Full CPU, memory, Bus and Cache encryption
- Error detection codes on all memories
- Error codes for cache protection
- Address and data scrambling of memories
- Side-channel leakage suppression
- Active Shield

Running the algorithm on the SLE78 offers a good deal of hardware protection with less reliance on added software countermeasures; so we would anticipate less performance degradation when compared with the SLE77.

IV. SOFTWARE DEVELOPMENT

The starting point for the smart card software development was the example code published in 3GPP TS 35.231 [6]. This went through several versions during the project, based on results/feedback and on-going optimisation work. The final versions should be regarded as optimised to the extent that was possible with a generic implementation avoiding chip specific enhancements. Referring to Table I the primary template for the smart card experiments was the generic speed optimised

 ISBN: 978-1-61208-451-0

version 1 that could be built for 8, 16, 32 and 64 bits, and made use of generic loops and macros. The 64 bit option was discounted as being unrepresentative of current smart cards and because legacy C compilers cannot easily cope with integer variables beyond 32 bits. Some minor modifications were made to the initial smart card code, but largely it remained true to the original generic code. Later, in order to understand performance issues relating to the algorithm running on the MULTOS platform, a 32-bit version of the code was part-optimised, which involved expanding the Macros and unrolling the inner loops within the main Keccak functions. The final MULTOS version also used fixed pointers for buffer manipulation. Note that in all versions of the code, the calculation of TOP_c was removed from each function. Within a smart card, this value would be pre-calculated and loaded into protected memory and so there is no need to recalculate it; and doing so could halve a TUAK function's speed.

A. Software Functional Testing

To test TUAK functionality, we used the six test data sets published in 3GPP TS 35.232 V12.0.1 [20]. The data sets were designed to vary all inputs and internal values, and assure correctness of an implementation; they thus also serve well for performance tests. To simplify testing the test data sets were included within the card application. This added an extra static data requirement, but meant that tests could be run by simply specifying the test set within the card test command, or by supplementing the test set with command data. Each command had an execution count so the targeted function could be run from 0 to 255 times (on the same input data). Typically the count would be '1', although '0' was useful for estimating round trip delays and higher counts improved measurement precision.

V. RESULTS

In this section, we present the experimental results, based on the 3GPP test data. The results were obtained via a scripting tool that would send a command message to the card in the form of an Application Protocol Data Unit (APDU) and then time the response. Although card processing time should be consistent and repeatable, scripting tools have tolerances. To compensate, the test commands instruct the card to execute a function multiple times before returning a result. A calibration was also carried out using a protocol analyser.

A. PC Results

The initial performance experiments used to refine the public example code were PC based, with results (in clock cycles) from the various versions (see Table I) summarised in Table II. Note that the cycle number includes pre, post data processing and overheads for a single run of Keccak-1600 (24 rounds).

Variation between minimum and average results arises from the OS. The minimum values are representative of the CPU capability. Generally, speed increased with the target build size.

B. Smart Card Performance

Native card performance was mainly measured on the SLE77; only the 32-bit algorithm was run on the SEL78. The MULTOS results used both chip types for all tests. The results are shown in Tables III and IV.

TABLE II. PC VERSION PERFORMANCE COMPARISON

Versions	Minimum Cycles (average cycles)		
	8-bit	16-bit	32-bit
0 (size opt)	168652(380066)	85988(215250)	
1 (speed opt)	49688(116200)	22496(55343)	7152(9024)
2 (N/A)			
3 (original)	202140(221564)		87350(193371)
4 (N/A)			
5 (unrolled)			6368(10391)
6 (bit-interl)	73120(185217)	59307(131112)	
7 (bit-interl opt)			10216(25570)

TABLE III. NATIVE MODE PERFORMANCE (ms)

Test	Mode/Chip	SLE77			SLE78		
Data		f1f1s	f2345	f5s	f1f1s	f2345	f5s
1	8-bit	18.11	18.17	18.11			
	16-bit	15.17	15.23	15.17			
	32-bit	19.58	19.64	19.51	19.58	19.70	19.51
2	8-bit	18.17	18.17	18.17			
	16-bit	15.23	15.23	15.17			
	32-bit	19.64	19.76	19.58	19.64	19.82	19.58
3	8-bit	18.23	18.17	18.17			
	16-bit	15.23	15.29	15.17			
	32-bit	19.70	19.82	19.58	19.70	19.88	19.58
4	8-bit	18.17	18.23	18.17			
	16-bit	15.17	15.23	15.17			
	32-bit	19.58	19.76	19.45	19.58	19.76	19.51
5	8-bit	18.17	18.23	18.17			
	16-bit	15.17	15.36	15.17			
	32-bit	19.58	20.01	19.58	19.58	20.00	19.58
6	8-bit	36.22	36.27	36.19			
	16-bit	30.16	30.28	30.10			
	32-bit	38.85	39.15	38.67	38.79	39.15	38.60

Normally, when the MULTOS organisation specifies a new function for the Virtual Machine (VM) it would be coded in low-level software and invoked from an Application Programming Interface (API). The API performance should be closer to that of Table III; however as this is currently not the case, the Table IV figures apply. All versions of the application benefit from a typical memory optimisation i.e., the Keccak main buffer (INOUT) was forced into a reserved section of RAM. Using non-volatile memory (NVM) instead made the 8-bit and 16-bit versions three times slower and the 32-bit version five times slower. The "32x" rows represent the

TABLE IV. MULTOS PERFORMANCE (ms)

Test	Mode/Chip	ML4 = SLE77			ML3 = SLE78		
Data		f1f1s	f2345	f5s	f1f1s	f2345	f5s
1	8-bit	19882	19952	19796	23837	23947	23962
	16-bit	10749	10826	10702	12824	12917	12838
	32-bit	6396	6505	6350	7239	7348	7192
	32x	3104	3214	3073	3432	3557	3400
	32p	1529	1575	1529	1623	1654	1622
2	32-bit	6474	6568	6396	7332	7441	7254
	32x	3198	3276	3120	3526	3619	3463
	32p	1544	1576	1529	1638	1669	1623
3	32-bit	6537	6615	6396	7379	7504	7254
	32x	3245	3339	3120	3603	3681	3463
	32p	1560	1592	1529	1654	1670	1623
4	32-bit	6427	6552	6349	7269	7410	7191
	32x	3151	3261	3089	3478	3603	3401
	32p	1544	1591	1529	1623	1669	1622
5	32-bit	6443	6708	6412	7301	7597	7254
	32x	3166	3432	3120	3494	3791	3463
	32p	1544	1622	1529	1638	1700	1622
6	32-bit	12543	12808	12402	14211	14492	14071
	32x	5990	6224	5866	6614	6879	6474
	32p	2980	3057	2949	3135	3198	3105

 ISBN: 978-1-61208-451-0

TABLE V. TEST DATA PARAMETER SIZES

Test Data	K	MAC	RES	CK	IK	Keccak Iterations
1	128	64	32	128	128	1
2	256	128	64	128	128	1
3	256	256	64	128	256	1
4	128	128	128	128	128	1
5	256	64	256	256	128	1
6	256	256	256	256	256	2

"unrolled" version of Keccak, which is a removal of inner loops and macros in the C code, and the "32p" version also uses fixed pointers rather than array index calculations. The smart card test results are further described and analysed in Section VI.

VI. Analysis of the Smart Card Results

To consider the experimental results, it is necessary to be aware of the parameter sizes (bits) inherent in the standardised test-sets, which are summarised in Table V. The test data parameters are designed to exercise TUAK in representative modes of use. Note that for the first five test sets (single iteration) the Keccak core has very similar execution time, with TUAK variations arising from the differing amounts of data to absorb or squeeze out of the sponge (working buffer).

Note that the common/fixed parameters sizes (bits) for the TUAK algorithm are: RAND = 128, SQN = 48, AK = 48, AMF = 16.

A. Performance Target

We need to define an appropriate performance target, so we can start by recalling the target used for the MILENAGE design [4].

> ..."The functions *f1-f5* and *f1** shall be designed so that they can be implemented on an IC card equipped with an 8-bit microprocessor running at 3.25 MHz with 8 kbyte ROM and 300byte RAM and produce AK, XMAC-A, RES, CK and IK in less than 500 ms execution time."...

Technology has advanced since this target was created and it might be difficult to find a SIM chip with these minimal capabilities, and indeed many do not have ROM. Furthermore, the target is ambiguous and could be interpreted that if you ran the functions in sequence each could take 500ms. It is also unclear how much of the ROM and RAM can be used. A more appropriate and modern target was defined during the study.

> ..."The functions *f1-f5* and *f1** shall be designed so that they can be implemented on a mid-range microprocessor IC card (typically 16-bit CPU), occupying no more than 8kbytes non-volatile-memory (NVM), reserving no more than 300bytes of RAM and producing AK, XMAC-A, RES, CK and IK in less than 500 ms total execution time."...

This revised target definition has been proposed to 3GPP for inclusion in future versions of the standard documents.

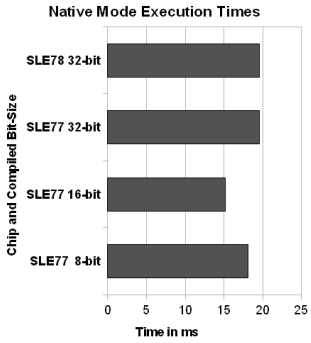

Figure 3. Comparison of Native Mode Execution Times

B. Native Mode

If we consider the results from the native implementation on the SLE77, the function execution times for the various test data sets are quite similar with the exception of test set 6. The latter uses a double iteration of Keccak, which roughly doubles the execution time. As can be seen from Figure 3, compiling the generic code for the different target bit widths affects the execution time, but not by an enormous margin. The most efficient version is the 16-bit target, which provides the best fit for the underlying processor.

Due to practical constraints we only have SLE78 measurements for the 32-bit target, which show similar speed to the SLE77 (native). The extra security features of the SLE78 seem not to penalise performance although there may be added financial cost. The striking observation is that native mode performance satisfies our target by a very comfortable margin. It is therefore reasonable to conclude that provided the algorithm is custom-coded on a typical (rather than highest performance) SIM chip there is no need for a crypto-coprocessor.

This study focussed more on performance than code-size minimisation, however, all native implementations fitted within our memory targets.

C. Platform Mode

Within the study we only considered the MULTOS platform; although a Java Card would make an interesting comparison. The results here were disappointing, although a significant overhead had been expected due to the operation of the secure Virtual Machine and the MULTOS Execution Language (MEL) [26] abstraction. In practice, the best results were around two orders of magnitude slower than native; see Figure 4. Furthermore, the performance improved with increasing compiled bit-size, which suggests that the compilation and MEL interpretation does not map closely to the underlying CPU size for the processing in TUAK.

On inspection of the generic Keccak function one saw extensive use of macros and loops. To determine if they were causing problems for MULTOS, an "unrolled" 32-bit version of Keccak was created, removing macros and inner loops. The results are in the Table IV rows marked "32x" and in Figure 4, showing a doubling of speed. A further improvement was to adapt the algorithm to use fixed location buffer pointers rather than indexed arrays; and the corresponding "32p" version shows a further speed doubling. However, a single function still takes around 1.5s.

 ISBN: 978-1-61208-451-0

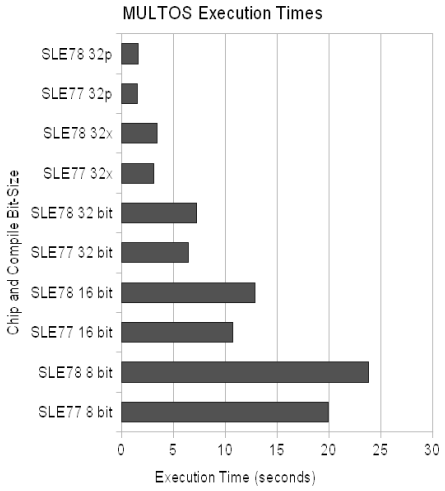

Figure 4. MULTOS f1() Execution Times

If we consider the unrolled Keccak there are many shifts on array contents, however MEL does not have a core shift instruction, but uses shift primitives. The unrolled Keccak is twice as fast as the generic version, partly due to the way that MULTOS handles shifts. The amount of shift on a buffer m can be known at compile time or run time, as shown below.

```
m = m << 3    or    m = m << n
```

The first example is handled as a single use of the shift primitive, whereas the second will loop n times shifting 1 bit at a time. This still leaves a big question mark over the efficiency of the primitive itself (and other bitwise operations).

If we consider the x2 speed-up from pre-computing TOP_c, the x2 from removing loops/macros and the x2 from using pointers, the application is x8 faster than the generic version. However the conclusion is still that the algorithm cannot meet the target performance if loaded as an application on a card platform (MULTOS at least). This suggets it is not practical to add the algorithm to deployed or existing stock cards. To use a card platform, an API would need to be added so that an efficient native implementation could be called.

VII. Security Defences and Performance

Modern SIM cards are normally based on tamper-resistant secure microcontrollers, which inherently have a range of defences against physical, side-channel and fault attacks. Therefore, a TUAK implementation on a SIM platform should be much better protected than an implementation on a general purpose microcontroller, with the latter incurring significant performance overhead to achieve modest attack resistance. If we consider the chips used in our tests then the SLE78 would be expected to offer significant protection against physical, side channel and fault attacks [24] due to the innovative underlying hardware; requiring less software countermeasures (and performance degradation) than a conventional secure microcontroller. The SLE77 would also offer hardware based protection, particularly against physical and fault attacks, but adequately preventing side-channel leakage will require additional measures in software. Fortunately, the SLE77 is quite fast and even if the performance was degraded by an order of magnitude, we could still run *f1, f2345* and *f5s* and meet the overall performance target. MULTOS platforms are known and marketed for their high security and had they been fast enough they would have been expected to offer added OS security to compliment the underlying chip hardware. However, the current view is that a new MULTOS primitive will be needed for the algorithm and so the issues are similar to the SLE77/78.

A. Fault Attack Defences and Performance Impact

The faults used in attacks are normally achieved by voltage glitches, radiation pulses and operating the target device beyond tolerance. The hardware sensors in tamper-resistant smart cards are intended to detect the likely means of fault insertion and prevent a response useful to the attacker; so there is no significant added overhead for the software. A very sophisticated and skillful attack might bypass the sensors, however by adopting TUAK as an openly published algorithm, with diversified card keys, we are avoiding proprietary secret algorithms that might motivate such effort. An added countermeasure could be to run the algorithm twice and only output a response if the results agree; this would counter attacks that analyse correct and faulty results from algorithms. The added countermeasure is probably unnecessary for the chips considered in this work, although halving the speed of operation would still keep it well within specification. Note that an attacker will seek to insert a fault at the most opportune moment, which may be determined from side-channel leakage.

B. Side-Channel Attack Defences and Performance Impact

Timing leakage attacks [27] can be possible when there are observable data dependent delays in the application; in which case added redundancy is needed in the implementation. Timing variations can be sufficiently large that they can be detected despite low level measures to disguise side-channel leakage that might be subject to power analysis. The leakage generation principle is quite simple, e.g., if a variable is true do something time-consuming else do something quick. The variable could represent a value that is tested at the application layer, or just a low-level bit test. A brief inspection of Keccak does not show obvious high-level timing leakage, as there are no conditional branches in the code. However, there could be lower level leakage if bit rotates are used. For example a processor may effect a rotate by shifting the contents of a register up one place and then testing the value that falls out of the register. If the value is '1' then this has to be added back in as the LSB, so unless the designer adds dummy operations, processing a '1' is going to take longer than a '0'.

The Keccak example code has macro names that imply rotate, but on inspection they are buffer shift operations rather than register rotates. However, there could be a timing effects when the compiled target size (8/16/32 bit) does not match the underlying register size. For example if we compile for 16-bits, but the CPU registers are 8-bits then our shift may need to modify the least significant bit of the upper byte based on the bit value shifted out of the lower byte. In the case of native code implementation, developers would be expected to take the CPU size/shift/rotate into account. In the platform approach the mapping between application variables and underlying registers is unclear.

We have assumed that the chips have hardware countermeasures to prevent bit-level side-channel leakage, as software

 ISBN: 978-1-61208-451-0

measures are inferior and significantly impact performance. For example, Hamming-weight equalisation is a technique that seeks to reduce leakage by ensuring that for each bit transition there is a complementary transition; so as a '1'changes to '0' there is also a '0' changing to '1'. In a practical implementation this could for example be a 16-bit processor where the lower 8-bits of a register handle the normal data and the upper 8-bits handle the complementary data. However, at the physical/electrical level, the register bits are unlikely to have equal contribution to the leakage and so Hamming-weight equalisation may not deliver a sufficient reduction. The impact on execution speed is also significant, as it is necessary to clear registers before and after use, and so a ten-fold rather two-fold reduction in performance should be anticipated.

VIII. Conclusions and Future Work

The main conclusion is that it is feasible to implement TUAK in software on typical smart card/SIM chips and meet the performance target for 3G/4G authentication algorithms, without the need for a cryptocoprocessor. Native mode implementation is required and so for a card platform (such as MULTOS) this should be supported via API calls. Processor and memory requirements are very modest suggesting that TUAK could meet performance targets even when implemented on simpler legacy CPUs. Although there is no high-level data dependent timing in TUAK, there is some potential for data dependent side-channel leakage due to shift operations, which will require countermeasures. Whilst high-end smart card chips (like the SLE78) may offer significant hardware-based resistance to side-channel analysis, other chips will require help from software countermeasures. Such measures may significantly impact performance; however the SLE77 results show that function execution time could be reduced by an order of magnitude and still satisfy the performance target. The primary impact of the work is that by showing TUAK to be a practical back-up or alterative to MILENAGE for typical SIM platforms, it will be adopted as a preferred public algorithm (initially in M2M systems); displacing proprietary solutions that are often the target and motivation for attack.

On-going work is considering a less advanced/legacy smart card chip (S3CCE9E4/8), side-channel leakage, and whether TUAK could be re-used in other applications. Preliminary results indicate that TUAK is sufficiently fast for use on more limited chip platforms, and this suggests it might also be a candidate for Internet of Things protocols. In fact re-using a 3G algorithm is not a new idea as MILENAGE has already been reused outside of mobile communications.

An initial collection of power traces from the original SLE77 TUAK implementation, shows the rounds and the round structure, although more traces and detailed analysis are still required to clearly determine data dependent leakage.

References

[1] (2016, Jan.) The European Telecommunications Standards Institute website, [Online]. Available: http://www.etsi.org/

[2] (2016, Jan.) The Third Generation Partnership Project website, [Online]. Available: http://www.3gpp.org/

[3] M. Mouly and M. Pautet, The GSM System for Mobile Communications, Cell and Sys (1992)

[4] 3GPP TS 35.206: 3G Security; Specification of the MILENAGE algorithm set: An example algorithm set for the 3GPP authentication and key generation functions f1, f1*, f2, f3, f4, f5 and f5*; Document 2: Algorithm specification (2014)

[5] Federal Information processing Standards, Advanced Encryption Standard (AES), FIPS publication 197 (2001)

[6] 3GPP, TS 35.231: 3G Security; Specification of the TUAK algorithm set: A second example algorithm set for the 3GPP authentication and key generation functions f1, f1*, f2, f3, f4, f5 and f5*; Document 1: Algorithm specification (2014)

[7] G. Bertoni, J. Daemen, M. Peeters, and G. van Aasche, The KECCAK Reference, version 3.0, 14 (2011)

[8] 3GPP TR 35.934: Specification of the TUAK algorithm set: A second example algorithm set for the 3GPP authentication and key generation functions f1, f1*, f2, f3, f4, f5 and f5*; Document 4: Report on the design and evaluation (2014)

[9] 3GPP TR 35.936: Specification of the TUAK algorithm set: A second example algorithm set for the 3GPP authentication and key generation functions f1, f1*, f2, f3, f4, f5 and f5*; Document 6: Security assessment (2015)

[10] G. Gong, K. Mandal, Y. Tan, and T.Wu, Security Assessment of TUAK Algorithm Set, [Online]. Available: http://www.3gpp.org/ftp/Specs/archive/35_series/35.935/SAGE_report/Secassesment.zip (2014)

[11] (2016, Jan.) eBACS: ECRYPT Benchmarking of Cryptographic Systems, [Online]. Available: http://bench.cr.yp.to/results-sha3.html

[12] Y. Jararweh, L. Tawalbeh, H. Tawalbeh, and A. Mohd, Hardware Performance Evaluation of SHA-3 Candidate Algorithms, Journal of Information Security, Vol 3, Number 2, p69-76 (2012)

[13] NIST, Announcing Draft Federal Information Processing Standard (FIPS) 202, SHA-3 Standard: Permutation-Based Hash and Extendable-Output Functions, and Draft Revision of the Applicability Clause of FIPS 180-4, Secure Hash Standard, and Request for Comments, (2004)

[14] (2016, Jan.) MULTOS website, [Online]. Available: http://www.multos.com/

[15] Oracle, Java Card Platform Specifications V3.04, (2011)

[16] F. Hillebrand, GSM and UMTS - The Creation of Global Mobile Communication - Wiley, (2002)

[17] 3GPP, TS 33.102: 3G Security; Security Architecture (1999)

[18] 3GPP, TS 33.401: Telecommunications Specification Group Services and System Aspects; 3GPP System Architecture Evolution (SAE); Security architecture (2012)

[19] G. Bertoni, J. Daemen, M. Peeters, and G. van Aasche, Cryptographic Sponge Functions, version 0.1, (2011)

[20] 3GPP, TS 35.232: 3G Security; Specification of the TUAK algorithm set: A second example algorithm set for the 3GPP authentication and key generation functions f1, f1*, f2, f3, f4, f5 and f5*; Document 2: Implementers' Test Data (2014)

[21] 3GPP TS 35.233: 3G Security; Specification of the TUAK algorithm set: A second example algorithm set for the 3GPP authentication and key generation functions f1, f1*, f2, f3, f4, f5 and f5*; Document 3: Design Conformance Test Data (2014)

[22] Infineon, SLE77CLFX2400P(M) Short Product Overview v11.11, (2012)

[23] Infineon, SLE78CAFX4000P(M) Short Product Overview v11.12, (2012)

[24] K. Mayes, and K. Markantonakis, Smart Cards, Tokens, Security and Applications, Springer (2008)

[25] (2016, Jan.) Infineon, Integrity Guard White Paper, [Online]. Available via: http://www.infineon.com/

[26] MULTOS, Developer's Reference Manual MAO-DOC-TEC-006 v1.49, (2013)

[27] P. Kocher, Timing Attacks on Implemenetationsof Diffie-Hellman, RSA, DSS, and other Systems, Advances in Cryptology CRYPTO 96 Volume 1109 LNCS p104-113 (1996)

Acknowledgment

The authors would like to thank members of ETSI SAGE for their expert advice.

Proving Transformation Correctness of Refactorings for Discrete and Continuous Simulink Models

Sebastian Schlesinger, Paula Herber, Thomas Göthel, and Sabine Glesner
Software Engineering for Embedded Systems
Technische Universität Berlin
Email: {Sebastian.Schlesinger, Paula.Herber, Thomas.Goethel, Sabine.Glesner}@tu-berlin.de

***Abstract*—MATLAB/Simulink is a state-of-the-art tool for model-driven engineering of embedded systems. Simulink enables engineers to model continuous and discrete parts of a system together in hybrid models. In such models, complexity reduction via refactoring plays an important role. However, formal verification of the equivalence between a hybrid Simulink model and its refactored counterpart is still an open problem. One challenge is that for many refactorings, equivalent behaviour can only be shown to be 'close' to each other rather than being the 'same'. To solve this problem, we propose a methodology to show behavioural equivalence based on approximate bisimulation. Our main contributions are a sound abstract representation for Simulink models that serves as a basis for proving equivalence, that we adapt the concept of approximate bisimulations to the operational Simulink semantics, and a methodology that enables the designer to prove transformation correctness. Our approach is applicable to both discrete and continuous Simulink models. With that, we provide an ideal starting point for the verification of hybrid models that integrate discrete and continuous model parts.**

Keywords–formal verification; transformation correctness; refactoring; MATLAB/Simulink; approximate bisimulation

I. INTRODUCTION

MATLAB/Simulink is a tool and de facto standard in model-driven engineering (MDE) in many industries, e.g., in the automotive and aerospace sectors. To control the complexity of Simulink models and to ensure their adherence to certain guidelines, *refactorings* are often used. Refactorings are model transformations that improve the inner structure of the model by keeping the behaviour of source and target model equivalent. Apart from reducing complexity and establishing compliance to defined guidelines for the models, refactorings can offer various perspectives of the model to the engineer - behaviourally equivalent, but different in other aspects. This helps the engineer to better understand the system and to find the proper refinement in forthcoming design phases to fulfil the requirements of the desired system. Other potential benefits of refactorings can be improved simulation performance or increased simulation precision after the transformation.

However, as Simulink is often used in safety-critical applications, it is crucial to ensure that refactorings preserve the intended behaviour of a Simulink model. While it is generally possible to validate the correctness of refactorings by simulation and testing, this does not provide any guarantees that the behaviour is preserved for all possible input scenarios. In this context, formal verification plays an important role. The ISO 26262 [1] states that it is highly recommended to formally prove system behaviour against the specification to cope with high Automotive Safety Integrity Levels (ASILs). However, formal verification of hybrid Simulink models, and in particular the formal verification of refactorings of hybrid models is a major challenge. The main difficulty is that for many refactorings, equivalent behaviour can only be shown to be 'close' to each other rather than being the 'same'.

In this paper, we present a methodology to enable the verification of behavioural equivalence of discrete or continuous Simulink models, in the sense that two models are close to each other. We are confident that our approach can be extended to actual hybrid models later in future work. The main challenge we address is that the standard concept for showing behavioural equivalence, namely bisimulation, is not applicable in cases where hybrid models are numerically approximated by Simulink. However, in such cases, we can prove that the values of the next simulation steps are *approximately* the same, i.e., they are located in an *'epsilon tube'* around the actual solution, which does not need to be known to the designer. To achieve this, we provide a new notion of equivalence for hybrid MATLAB/Simulink models by adapting the existing concepts for *approximate bisimulation* as described in, e.g., [2] to the operational Simulink semantics given in [3].

Our main contributions are the following.

1) We adapt the concept of approximate bisimulation for Simulink.
2) We provide an abstract and formally well-defined semantics for Simulink.
3) We propose a methodology for transformation correctness of hybrid Simulink models.

The main idea of our methodology to prove transformation correctness is that we provide an abstract representation for Simulink models and perform the verification on this abstract level using approximate bisimulation. We put the interpretation of the abstract representations on a firm footing by proving soundness with respect to an existing operational Simulink semantics [3]. To prove behavioural equivalence between a source model and its refactored counterpart, a designer can use our abstract representation to compute boundaries for the ranges in which the values are contained after each simulation step. From these boundaries, approximate bisimulation between the original Simulink models can be derived.

Note that in this paper, we focus on systems that are either discrete or continuous. It is part of future work to combine the findings to be able to handle fully hybrid models as well.

The rest of this paper is structured as follows: In Section II, we introduce the operational semantics of MATLAB/Simulink and the concept of approximate bisimulation. In Section III, we

 ISBN: 978-1-61208-451-0

discuss related work. Section IV provides a general overview over our approach. In Section V, we provide our adaptation of approximate bisimulation to Simulink. In Section VI, we present our abstract representation and discuss its soundness with respect to the operational semantics. In Section VII, we examine how behavioural equivalence between Simulink models can be established using our abstract representation together with our adapted concept of approximate bisimulation. We conclude in Section VIII.

II. BACKGROUND

In this section, we briefly summarise the necessary background to understand the remainder of this paper.

A. Simulink

Simulink is a widely used modeling language for dynamic systems. It is based on Matlab, and both products are developed by The MathWorks [4]. In Simulink, dynamic systems are modelled as block diagrams. They can then be simulated, and, with further software packages, it is also possible to automatically generate code. Simulink provides a large library of predefined blocks. Additionally, user-defined libraries and block types can be added.

In this paper, we consider five kinds of Simulink blocks: unsampled or direct feed-through blocks (e.g., arithmetic blocks), discrete time blocks (e.g., Unit Delay, Discrete Integrator), continuous time blocks (e.g., Integrator), sink blocks (e.g., Scope) and source blocks (e.g., Constant, Sine Wave, Ramp). Each block can have inports and outports. These are the interfaces via which the blocks are connected. We speak of *discrete* models if the model consists only of discrete and unsampled blocks and *continuous* models if the model consists only of continuous and unsampled blocks. Finally, a *hybrid* model is a model with all block types.

Simulink Operational Semantics

To capture the behaviour of Simulink models, we use the operational semantics described by Bouissou and Chapoutot [3]. There, a state is a mapping $\sigma : V \to \mathbb{R}$ where $V = \{l_i | 1 \leq i \leq n_b\} \cup \{d_i | 1 \leq i \leq n_d\} \cup \{x_i | 1 \leq i \leq n_x\} \cup \{t, h\}$. The numbers n_b, n_d and n_x express the amounts of variables for the output of blocks, discrete and continuous respectively. Every block has a unique ouput variable l_i, discrete blocks have an internal variable d_i, and continuous blocks an internal variable x_i in addition. The variable t is used for the simulation time and h stands for the simulation step size. Parameters that are provided by the user are denoted by a function π. For instance, $\pi(t_0)$ expresses the simulation start time, $\pi(t_{end})$ the simulation end time, $\pi(h_0)$ the initial simulation step size. Bouissou and Chapoutot then assign each block with a set of equations, examples are $l_1 = l_2 + l_3$, $l_i =_S d; \bar{d} =_S l_1$, $\dot{x} = l_1, x(0) = init$. These equations are evaluated by the operational semantics, which is defined as a set of inference rules. In the following, we provide a brief summary of the most important inference rules.

The global simulation rule states that if the simulation time is not expired (in which case nothing happens), the simulation consists of three steps that are consecutively evaluated:

1) The $M-$transition (*major step* transition) evaluates equations of the form $l = e$ and $l =_S e$, i.e., unsampled equations and outputs of discrete (sampled) blocks.
2) The $u-$transition (*update* transition) evaluates the internal variables of discrete blocks, i.e., equations of the form $\bar{d} =_S l$.
3) The $s-$transition (*solver* transition) evaluates the internal variables of the continuous blocks, i.e., equations of the form $\dot{x} = l$.

During the evaluation of these transitions, the following rules apply:

- The evaluation of expressions and predicates is performed by $o-$transitions: variables l evaluate to the value in the current state $\sigma(l)$, constants to the constant value, functions are evaluated by applying them on the values of their arguments, predicates analogously.
- Equations with an index S, e.g., $l_o =_S f(l_1, ..., l_n)$ only lead to a state change if $\sigma(t) \in S$, i.e., the sample time of the block is reached by the simulation time. If $\sigma(t) \in S$, the expression on the right hand side is evaluated by $o-$transitions to calculate the new value for $\sigma(l_o)$.

The solver transitions to calculate equations of the form $\dot{x} = l$ ($\dot{x}$ denotes the derivative of x) are more complex. Simulink approximates the value x_{n+1} at point $t_n + h_n$ starting from the value x_n at time index t_n by applying an approximation technique defined by the user, e.g., Euler method or one of the Runge-Kutta methods [5]. The solver transition finishes with the calculation of the subsequent simulation step size. Please note that all blocks are simultaneously evaluated to calculate the succeeding state according to the operational semantics, i.e., it is a synchronous semantics.

B. Approximate Bisimulation

Our goal is to prove equivalent behaviour of source model and refactored target model. As indicated in Section II, in case a model contains continuous blocks, Simulink does not compute the actual solution, but an approximation. Hence, although intuitively source and target model may express the same solution function, it is not possible to show exact equality. We therefore require a notion of equivalence that relaxes equality. To this end, we use approximate bisimulation, as defined in [2], [6].

Definition 1 (Approximate Bisimulation)**.** *Let* $T_1 = (Q_1, \Sigma, \to_1, Q_1^0, \Pi, \langle\langle . \rangle\rangle_1)$ *and* $T_2 = (Q_2, \Sigma, \to_2, Q_2^0, \Pi, \langle\langle . \rangle\rangle_2)$ *be two labelled transition systems (LTS), where* Q_i *are the sets of states,* $Q_i^0 \subseteq Q_i$ *the sets of initial states,* Σ *the sets of input alphabet (labels),* $\to_i \subseteq Q_i \times \Sigma \times Q_i$ *the transition relations,* Π *the set of observations and* $\langle\langle . \rangle\rangle_i : Q_i \to \Pi$ *mappings of states to observations (*$i \in \{1, 2\}$*; note that* Σ *and* Π *is in both LTS the same). Let furthermore* Π *be a metric space with metric* $d : \Pi \times \Pi \to \mathbb{R}$*. The metric allows us to measure distances in the set of observables.*

A bisimulation relation $B_\varepsilon \subseteq Q_1 \times Q_2$ *of precision* ε *is a relation fulfilling the following properties* $\forall (q_1, q_2) \in B_\varepsilon$.

1) $d(\langle\langle q_1 \rangle\rangle_1, \langle\langle q_2 \rangle\rangle_2) \leq \varepsilon$

2) $\exists q_1' \in Q_1 : q_1 \xrightarrow{\alpha} q_1' \Rightarrow \exists q_2' \in Q_2 : q_2 \xrightarrow{\alpha} q_2' \wedge (q_1', q_2') \in B_\varepsilon$
3) $\exists q_2' \in Q_2 : q_2 \xrightarrow{\alpha} q_2' \Rightarrow \exists q_1' \in Q_1 : q_1 \xrightarrow{\alpha} q_1' \wedge (q_1', q_2') \in B_\varepsilon$

The LTS are bisimilar with precision ε, denoted as $T_1 \sim_\varepsilon T_2$, if $B_\varepsilon \subseteq Q_1 \times Q_2$ exists such that

4) *B_ε is a bisimulation relation of precision ε,*
5) $\forall q_1 \in Q_1^0 \exists q_2 \in Q_2^0 : (q_1, q_2) \in B_\varepsilon$
6) $\forall q_2 \in Q_2^0 \exists q_1 \in Q_1^0 : (q_1, q_2) \in B_\varepsilon$

Before we come to the approach, we discuss related work in the next section.

III. Related Work

For the design and application of refactorings in Simulink, several approaches exist. In [7], a taxonomy of model mutations is defined. In [8], a normalisation of Simulink models is presented, which is used for clone detection in [9]. In [10], the authors present Simulink refactorings that allow subsystem and signal shifting in arbitrary layers. None of these approaches considers behavioural equivalence or transformation correctness.

There exists a broad variety of verification approaches for hybrid models, for a detailed introduction see for example [2], [6], [11]–[15]. In these papers, the notion of approximate bisimulation is introduced and utilised. However, none of them targets hybrid systems described in Simulink and it is a non-trivial challenge to transfer them to the specifics of the Simulink semantics and make use of its special characteristics such as deterministic behaviour. To reason about transformation correctness for Simulink models, a clear understanding of the Simulink semantics is required. However, the Mathworks documentation [4] defines the Simulink semantics only informally. Existing approaches for the formal verification of Simulink models typically overcome this problem by using transformations into some well-established formal language. For example, in [16], Simulink models are mapped to the verification system UCLID and the satisfiability modulo theories (SMT) solver UCLID is used for verification. In [17], this is done with the synchronous data flow language LUSTRE. In [18], an approach for contract based verification is presented. In this approach, the semantics is described via synchronous data flow graphs. In [19], the authors use Boogie, a verification framework developed at Microsoft Research, for verification. For this, the semantics of discrete Simulink models is again translated to the input language of a verification tool, in this case Boogie. In [20], Simulink models are translated to hybrid automata. This enables further investigation of hybrid automata semantics, e.g., in [21] or [22]. However, none of these approaches provide a formal semantics for Simulink, as they all use some transformation into existing formal languages. This also constrains the supported subset of Simulink models in all cases.

To the best of our knowledge, only Bouissou and Chapoutot [3] provide a direct formalization of the Simulink semantics. The authors consider a Simulink model as a set of equations evaluated following an operational semantics, which is defined using a set of inference rules. In our approach, we use this operational semantics as a formal foundation to capture the behaviour of Simulink models. In contrast to Bouissou and Chapoutot, we utilise the formal semantics to derive a set of syntactical equations as an abstract representation, which describes the changes of signals over time on an abstract level. Together with the concept of approximate bisimulation, this enables the verification of approximate behavioural equivalence between a source and a target model.

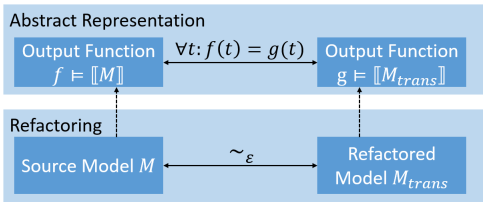

Figure 1. Overview over our approach

IV. Transformation Correctness Approach for Discrete and Continuous Simulink Models

In this paper, we present a methodology for proving transformation correctness of Simulink refactorings. Our approach is applicable to both discrete and continuous models. To keep the notation simple, we assume without loss of generality that there is only one output block (where the observation takes place) in the model. If there was more than one output block, our approach just needed to be repeated for each. One of the key ideas of our transformation correctness approach is that we define an abstract representation for Simulink models. The *Abstract Representation* (AR) links the outgoing with the incoming signals of a given Simulink model by describing the effect of the blocks via equations in a mathematical view. The resulting set of equations can be interpreted by an output function $f \models [\![M]\!]$ that fulfils the conjunction of all equations of the respective equation set. This function describes basically how the output (meaning the observation) is linked with the input. Using our abstract representation, we can show that two simulated models behave approximately equivalent up to a certain precision by proving that the output functions yield the same values for all possible simulation steps.

The overall concept is depicted in Figure 1. There, we assume that a source model M and its transformed refactoring M_{trans} are given. To prove approximate bisimulation ($\sim_\varepsilon$) between the source model and the refactoring, we compute abstract representations for both of them and then prove that $\forall t : f(t) = g(t)$ holds.

In the following sections, we first present our adaptation of approximate bisimulation for Simulink and our abstract representation together with its interpretation using the concept of an output function that captures the semantics of a given Simulink model. Then, we present our methodology for showing behavioural equivalence of two Simulink models using our abstract representation together with our adapted notion of approximate bisimulation. Note that in contrast to the operational semantics defined in [3], our abstract interpretation provides a mathematical interpretation, i.e., it describes the behaviour of a given Simulink model using an exact calculation rather than an approximation or simulation. Nevertheless, we can show that our abstract representation is sound with respect to the operational semantics defined in [3].

 ISBN: 978-1-61208-451-0

Limitations and Assumptions

As representatives for discrete and continuous blocks, we currently only support Unit Delay and Integrator respectively. All Unit Delays must have the same sample times because currently we do not support the rate transition block, which would be necessary otherwise. Feedback loops, i.e., cycles in the Simulink graph, are only permitted with at least one Unit Delay or Integrator in the cycle. This means that we do not support algebraic loops. Furthermore, the following Simulink items are currently not supported:

- buses (multiple signals on one signal line) and multi-dimensional signals
- atomic and virtual subsystems
- blocks that alter the control flow, e.g., Switches
- full hybrid systems, i.e., models containing both discrete and continuous parts.

With our current methodology, many industrially relevant models are already covered. Our approach can easily be extended to other blocks.

V. Approximate Bisimulation for Simulink

In this section, we adapt the notion of approximate bisimulation as introduced in Section II-B to Simulink. To this end, we define the semantics of Simulink as a labelled transition system (LTS) and provide a metric to measure observation distances.

First, we define the syntax of a Simulink model as follows.

Definition 2 (Simulink Syntax)**.** *A Simulink model is a tuple* (B, E, I, O)*.* B *is a finite set of blocks,* $E \subseteq B \times \mathbb{N} \times B$ *the signal lines together with the arity of the incoming signal (the arity is important because some functions defined by blocks are not commutative),* $I \subseteq B$ *the set of sources (for which* $\forall b' \in I \forall b \in B \forall n \in \mathbb{N} : (b, n, b') \notin E$ *applies),* $O \subseteq B$ *the set of sinks (for which* $\forall b \in O \forall b' \in B \forall n \in \mathbb{N} : (b, n, b') \notin E$ *applies).*

Please note that we often abbreviate signal lines $(b_i, n, b') \in E$ by variables l_i with $i \in \mathbb{N}$. Using our Simulink syntax, we can now define the semantics as an LTS.

Definition 3 (Labelled Transition System of Simulink)**.** *The* LTS *for a Simulink model* $M = (B, E, I, O)$ *is*

- $Q \subseteq \mathbb{R}^V$ *- the set of states (each variable is assigned with a value)*
- *We do not need a set of inputs* Σ*.*
- $\rightarrow \subseteq Q \times Q$ *is defined by the operational semantics*
- $Q^0 \subseteq Q$ *assigns the initial values to each variable (given as parameter in the Simulink model)*
- $\Pi \subseteq (\mathbb{R} \cup \{\perp\})^V$ *- the set of observations*
- $\langle\langle . \rangle\rangle : Q \rightarrow \Pi$*,* $\langle\langle \sigma \rangle\rangle(v) := \sigma(v)$ *if* v *is a variable associated with a sink block* $b \in O$*,* $\perp$ *otherwise. (The observation map provides the value at the sink of the model - and an undefined symbol* $\perp$ *otherwise)*

To keep notation simple, we denote the unique output variable of a block $b_i \in B$ with v_i from now on. The metric we use for approximate bisimulation is defined as $d_\infty : \Pi \times \Pi \rightarrow \mathbb{R}$, $d_\infty(\sigma_1, \sigma_2) := \|\sigma_1 - \sigma_2\|_\infty$ with $\|.\|_\infty : \Pi \rightarrow \mathbb{R}$, $\|\sigma\|_\infty := \max_{v \in \bigcup_{b \in O} : \text{v variable of b}} (\sigma(v))$. It takes the biggest distance of all observations.

TABLE I. EXAMPLE $eqExtract$ DEFINITIONS

Block type	$eqExtract(l)$
Input	$\{l(t) = in(t)\}$
Sine Wave	$\{l(t) = \sin(t)\}$
Math Operations	$\{l(t) = f(l_1(t), ..., l_n(t))\}$
Unit Delay	$\{l(t+h) = l_1(t), l(\pi(t_0)) = \pi(init_l)\}$
Integrator	$\{\frac{d}{dt} l(t) = l_1(t), l(\pi(t_0)) = \pi(init_l)\}$

VI. Abstract Representations and their Interpretation

In this section, we introduce the abstract representations (AR). Equations in the AR (we denote the set of AR equations as Eq_a) are of the form $l_i(t) = f(l_1(t), ..., l_n(t))$, $l_i(t + \lambda h) = l_1(t)$ or $\frac{d}{dt} l_i(t) = l_1(t)$, where f is an expression depending on other variables, h is the (synchronous) sample time, λ_i are integral coefficients.

Definition 4 (Abstract Representation)**.** *For a Simulink model* $M = (B, E, I, O)$*, the abstract representation is a function* $[\![.]\!]$ *that maps* M *to a set of equations* $[\![.]\!] : SM \rightarrow \mathcal{P}(Eq_a)$*,* $[\![M]\!] = \bigcup_{l \in E} eqExtract(l)$*, where* SM *is the set of Simulink models. The function* $eqExtract : E \rightarrow \mathcal{P}(Eq_a)$ *describes how the outgoing signal of a block is mathematically described depending on its incoming signals and must be defined according to the semantics of each block.*

Some examples for definitions of the function $eqExtract$ are shown in Table I. Note that Unit Delay and Integrator blocks both need an initial value, given as parameter. This is reflected by the second equations in these cases.

On the basis of our definition of an abstract representation of Simulink models, we can now define an output function that abstractly captures the semantics of a given Simulink model by interpreting our abstract representation.

Definition 5 (Interpretation of an Abstract Representations)**.** *Let* $M = (B, E, I, O)$ *a Simulink model with* $\#O = 1$ *(for simplicity) and simulation period* $\mathfrak{I} := [\pi(t_0), \pi(t_{end})]$*. Let furthermore* $[\![M]\!] := \{eq_1, ..., eq_n\}$ *the abstract representation. An interpretation of the abstract representation is a function* $f : \mathfrak{I} \rightarrow \mathbb{R}$ *that fulfils all equations of* $[\![M]\!]$ *for all* $t \in \mathfrak{I}$*. We denote with* $f \models [\![M]\!]$ *that* f *fulfils* $\forall t : eq_1 \wedge eq_2 \wedge ... \wedge eq_n$*.*

Please note that this function always exists if the input signals are continuous. For the uniqueness of the interpretation of a continuous model, Lipschitz-continuity must be guaranteed as well [23].

We now establish the soundness of the AR with the operational semantics. Due to space constraints, we only provide a proof sketch here.

Lemma 1 (Soundness of the Abstract Representation)**.** *Let* M *be a Simulink model that defines an LTS* $(Q, \rightarrow, Q^0, \Pi, \langle\langle . \rangle\rangle)$*. Then for all* $\sigma \in Q$*, the equations in* $[\![M]\!]$ *for unsampled and discrete blocks hold if we replace terms* $l_i(t)$ *by* $\sigma(v_i)$*,* $l_i(t + h)$ *by* $\sigma'(v_i)$ *(with* $\sigma \rightarrow \sigma'$*,* $\sigma'(t) \leq \sigma(t) + \sigma(h)$*) etc. For continuous blocks, we construct functions* $l_{h \rightarrow 0}$*. To define* $l_{h \rightarrow 0}$ *for a signal* l*, we define a function* $l_{(t_n)_n} : E \rightarrow \mathbb{R}^{\mathfrak{I}}$ *where* $\mathfrak{I}$ *is the simulation interval and* $(t_n)_n$ *is a sequence of simulation steps.* $l_{(t_n)_n}(\tau)$ *is defined as follows:*

$$l_{(t_n)_n}(\tau) = \frac{\sigma'(v) - \sigma(v)}{t_{i+1} - t_i}(\tau - t_i) + \sigma(v)$$

where $\sigma(t) = t_i, \sigma'(t) = t_{i+1}, t_i \leq \tau < t_{i+1}$.

We obtain $l_{h\to 0}$ *out of the family of functions* $l_{(t_n)_n}$ *with simulation step size* $t_{n+1} - t_n \to 0$*. Then the equations of the form* $\frac{d}{dt}l(t) = l_1(t)$ *hold if we replace all* $l(t)$ *by* $l_{h\to 0}(t)$ *etc.*

The function $l_{(t_n)_n}$ assigns a value to each $t \in \mathfrak{I}$: If t is a sample time in the sequence $(t_n)_n$, i.e., if an i exists with $t = t_i$, then $l_{(t_n)_n}(t)$ is the result of the execution of the Simulink semantics. If t is between two sample steps, the value is on the line connecting the points $(t_i, \sigma(v))$ and $(t_{i+1}, \sigma'(v))$.

Proof sketch: The unsampled and discrete cases follow from the operational semantics because unsampled blocks are directly evaluated, and discrete blocks only update if sample time is reached. For the continuous case, the function $l_{h\to 0}$ exists because the family of functions of the form $l_{(t_n)_n}$ is equicontinuous, bounded and all functions of this family are defined on the compact interval $\mathfrak{I}$. Consequently, according to the theorem of Arzela-Ascoli [24], a limit function $l_{h\to 0}$ exists towards which the functions uniformly converge. According to the semantics, the integrator block performs an approximation technique such as Euler. This yields immediately that $\frac{d}{dt}l_{h\to 0}(t) = l_{1,h\to 0}(t)$ holds. ■

We have now established that our AR is sound with respect to the operational semantics defined in [3]. Especially, we have put syntactical constructs that were already introduced in [3] on a firm footage regarding the interpretation with respect to an output function that mathematically captures the semantics of a given model.

VII. Behavioural Equivalences

We have established a sound abstract representation for Simulink models. Its interpretation yields a concise form of an output function, i.e., it expresses what the Simulink model actually does in the form of equations. The idea of our approach is that the designer verifies that the abstract representations of source and target model, which are either discrete or continuous, yield the same values on the simulation interval. From this, according to the next theorem, follows the approximately equivalent behaviour of both models.

Theorem 1 (Behavioural Equivalence of Simulink Models)**.** *Let* M *and* M_{trans} *be Simulink models (cf. Section IV for the limitations,* $\#O = 1$ *for simplicity),* $\mathfrak{I} = [\pi(t_0), \pi(t_{end})]$ *the common simulation interval. Let furthermore the interpretations of the abstract representations* $g \models [\![M]\!]$ *and* $h \models [\![M_{trans}]\!]$ *be equal during the simulation interval, i.e.,* $\forall t \in \mathfrak{I} : g(t) = h(t)$ *holds.*

Then $M \sim_0 M_{trans}$ *if the models are unsampled or discrete. If* M *is continuous,* $[\![M]\!]$ *expressing an ordinary differential equation (ODE; after re-arrangement of the equations)* $\frac{d}{dt}y(t) = f(t, y(t)), y(\pi(t_0)) = \xi$ *with* f *being Lipschitz-continuous in* y*, i.e.,* $||f(t, y_1) - f(t, y_2)|| \leq L||y_1 - y_2||$ *for all* t, y_1, y_2*. Then we have* $M \sim_\varepsilon M_{trans}$ *with*

$$\epsilon = \sup_{n=0,\ldots}(|t_n^t - t_n^s|) + \sup_{n=0,\ldots}(||\varphi(t_n^t) - \varphi(t_n^s)||) + \varepsilon_s + \varepsilon_t,$$

where t_n^s *is the sample time of* M *(source model) at the* $n-th$ *step,* t_n^t *analogously for* M_{trans} *(target model).* ε_s *and* ε_t *are the global errors between the approximations performed by Simulink for* M *and* M_{trans} *and the mathematical solution of the ODE, i.e., the unique function that actually solves the ODE rather than approximating it.*

If source and target model are sampled with the same fixed step size, then the error reduces to $\varepsilon = \varepsilon_s + \varepsilon_t$.

Proof sketch: For unsampled and discrete models, the precision ε is 0, i.e., the systems are bisimilar because the ARs are sound (cf. Lemma 1) and the operational semantics evaluates the expressions by providing exact and deterministic values. In the continuous case, let us abbreviate the approximated value at the $n-$th time step in the source model by $\psi_s(t_n^s)$ and for the target model by $\psi_t(t_n^t)$ respectively and the mathematical solution by φ. Then, with triangular inequality follows $\varepsilon \leq |t_n^t - t_n^s| + ||\psi_s(t_n^s) - \psi_t(t_n^t)|| \leq ||\psi_s(t_n^s) - \varphi(t_n^s)|| + ||\psi_t(t_n^t) - \varphi(t_n^t)|| + ||\varphi(t_n^s) - \varphi(t_n^t)|| = |t_n^t - t_n^s| + \epsilon_s + \epsilon_t + ||\varphi(t_n^t) - \varphi(t_n^s)||$. ■

Note that y and f in the continuous case can be multidimensional, e.g., $y = (y_1, y_2)$, $f = (f_1(t, y_1(t), y_2(t)), f_2(t, y_1(t), y_2(t)))$. This occurs if the system consists of more than one Integrator block.

Also note that the mathematical solution φ always exists and is unique under the assumptions of the theorem [23]. The ODEs do not need to be analytically solvable, i.e., our approach is applicable even though the solution may not be phraseable as a closed expression. The errors ε_s and ε_t depend on the approximation technique defined by the user in Simulink. For example, the Euler method yields

$$\varepsilon_{s/t} = \frac{1}{2}(\text{len}(\mathfrak{I}))e^{(\text{len}(\mathfrak{I})L)} \sup_{t\in\mathfrak{I}}\left(|\frac{d^2}{dt^2}\varphi(t)|\right) \sup_{n=0,\ldots} h_n^{s/t}$$

with the Lipschitz-constant L for f, $h_n^{s/t}$ the simulation step sizes at the $n-$th step, $\varphi : \mathfrak{I} \to \mathbb{R}$ the mathematical solution of the ODE and $\text{len}(\mathfrak{I}) = \pi(t_{end}) - \pi(t_0)$ the length of the simulation interval [23]. Euler has the worst approximation to the mathematical solution and consequently the precision ε can be improved by taking more advanced techniques available in Simulink into account. Also note that in case of fixed and equal sample step sizes on both models, the error reduces to just $\varepsilon = \varepsilon_s + \varepsilon_t$.

One important class of refactorings that immediately results from the theorem is if the ODE of M is analytically solvable and M_{trans} expresses the analytical solution directly. In this case, $\varepsilon_t = 0$ and M_{trans} is significantly less complex than M because usually feedback loops are removed. However, our approach is not restricted to this case. We also support refactorings that for instance keep the Integrator and transform something else mathematically equivalent.

It should also be mentioned that ε in the continuous case increases with the simulation interval size in the general case. However, we are confident to obtain certain classes of functions that allow to show that the global error of the approximation techniques is bounded for all $t \in [\pi(t_0), \infty[$.

Example

Let us go through a simple continuous example depicted in Figure 2 to demonstrate the approach. The model in the left is the source model M, on the right the target model M_{trans}. The labels and parameters are shown as well. Together with the models it is given that the expression for l_2 in M_{trans} solves the ODE in M. Suppose now that the designer wants to verify the behavioural equivalence of both models. In the first step, the abstract representations are extracted. $[\![M]\!] = \{\frac{d}{dt}l_1(t) =$

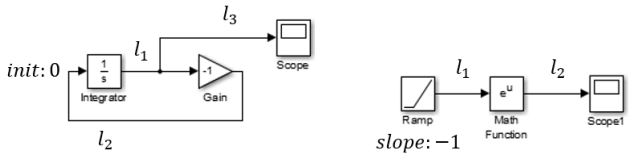

Figure 2. Refactoring Example

$l_2(t), l_1(0) = 1, l_2(t) = -l_1(t), l_3(t) = l_1(t)\}$, $M_{trans} = \{l_1(t) = -t, l_2(t) = e^t\}$. The observation at model M is $l_3(t)$, at model M_{trans} it is $l_2(t)$ because these signal lines are connected with an output block. The designer now has to establish that the interpretations $f \models [\![M]\!]$ and $g \models [\![M_{trans}]\!]$ fulfil $\forall t \in \mathfrak{I} : f(t) = g(t)$. To see that, the interpretations, i.e., the observations in both models must be equal. For M_{trans} we obtain this immediately by extracting $g(t) = e^{-t}$. For M, we need the solution of the ODE, which is given by M_{trans} as previously stated, i.e., we obtain $f(t) = e^{-t}$. This immediately yields $f(t) = g(t)$ for all t. However, to complete the proof, we finally need to establish that e^{-t} indeed solves the ODE in M (and the initial value problem $e^0 = 1$). This yields the final proof obligations $\frac{d}{dt}e^{-t} = -e^{-t}, e^0 = 1$, which hold. The designer consequently can apply our theorem and obtains the approximately equivalent behaviour. The precision ε depends on the chosen approximation technique.

VIII. Conclusion and Future Work

In this paper, we have presented a methodology for the verification of refactorings of discrete and continuous Simulink models. We have extended the operational semantics from [3] to an abstract representation of Simulink models, where the semantics is mathematically captured using a concise output function, and have shown its soundness with the operational semantics. The abstract representation is better suitable for automation than the operational semantics. Furthermore, we have adapted a suitable notion of behavioural equivalence, the approximate bisimulation, to Simulink models. This is especially useful for continuous models, since Simulink does not provide exactly the same values for the refactored model in comparison to the original model. Altogether, we have provided a methodology that enables a designer to verify the equivalence of discrete and continuous Simulink models and their refactored counterpart up to a certain precision.

Our approach offers many possibilities for further research. Currently we are working on the automation of our approach. This includes an algorithm that extracts expressions from Simulink models and sets up proof obligations that are then being verified with the assistance of a computer algebra system. In future work, we plan to extend our approach to hybrid models that integrate discrete and continuous model parts. To achieve this, we plan to investigate the behaviour 1) of models that contain both continuous and discrete blocks within the same subsystem and 2) of models that contain control flow blocks, e.g., switches. For the former, we plan to make use of delay differential equations theory. The challenge for the latter is that small changes at the switch input can yield diverging behaviour. Hence, additional proof obligations must be considered to establish the approximate behaviour in these cases.

References

[1] International Organization for Standardization, "ISO 26262 - Road vehicles – Functional safety."

[2] A. Girard and G. J. Pappas, "Approximate bisimulation: A bridge between computer science and control theory," European Journal of Control, vol. 17, no. 5, 2011, pp. 568–578.

[3] O. Bouissou and A. Chapoutot, "An operational semantics for simulink's simulation engine," ACM SIGPLAN Notices, vol. 47, no. 5, 2012, pp. 129–138.

[4] I. The Mathworks, "Simulink documentation website," retrieved: Jan 2016. [Online]. Available: http://de.mathworks.com/help/simulink/

[5] J. C. Butcher, Numerical methods for ordinary differential equations, 2nd ed. Wiley, 2008.

[6] A. Girard, "Approximate bisimulations for constrained linear systems," Automatica, 2005, pp. 1307–1317.

[7] S. Matthew, "Towards a taxonomy for simulink model mutations," in ICSTW 2014, IEEE International Conference. IEEE, 2014, pp. 206–215.

[8] B. Al-Batran, B. Schätz, and B. Hummel, "Semantic clone detection for model-based development of embedded systems," in Model Driven Engineering Languages and Systems. Springer, 2011, pp. 258–272.

[9] F. Deissenboeck, B. Hummel, E. Juergens, M. Pfaehler, and B. Schaetz, "Model clone detection in practice," in Proceedings of the 4th International Workshop on Software Clones. New York: ACM, 2010, pp. 57–64.

[10] Q. M. Tran, B. Wilmes, and C. Dziobek, "Refactoring of simulink diagrams via composition of transformation steps," in ICSEA 2013, The Eighth International Conference on Software Engineering Advances, 2013, pp. 140–145.

[11] G. J. Pappas, "Bisimilar linear systems," Automatica, vol. 39, no. 12, 2003, pp. 2035–2047.

[12] R. Alur, T. A. Henzinger, G. Lafferriere, and G. J. Pappas, "Discrete abstractions of hybrid systems," Proceedings of the IEEE, vol. 88, no. 7, 2000, pp. 971–984.

[13] A. Girard, "Approximate bisimulations for nonlinear dynamical systems," in European Control Conference 2005. IEEE, 2005, pp. 684–689.

[14] Van der Schaft, AJ, "Equivalence of dynamical systems by bisimulation," Automatic Control, IEEE Transactions on, vol. 49, no. 12, 2004, pp. 2160–2172.

[15] A. Tiwari, "Abstractions for hybrid systems," Formal Methods in System Design, vol. 32, no. 1, 2008, pp. 57–83.

[16] P. Herber, R. Reicherdt, and P. Bittner, "Bit-precise formal verification of discrete-time matlab/simulink models using smt solving," in Proceedings EMSOFT '13. IEEE Press, 2013.

[17] P. Caspi, "Translating discrete-time simulink to lustre," in ACM Transactions on Embedded Computing Systems, New York, 2005, pp. 779–818.

[18] P. Boström, "Contract-based verification of simulink models," in Formal Methods and Software Engineering. Springer, 2011, pp. 291–306.

[19] R. Reicherdt and S. Glesner, "Formal verification of discrete-time matlab/simulink models using boogie," in Software Engineering and Formal Methods. Springer, 2014, pp. 190–204.

[20] A. Agrawal, G. Simon, and G. Karsai, "Semantic translation of simulink/stateflow models to hybrid automata using graph transformations," Electronic Notes in Theoretical Computer Science, vol. 109, 2004, pp. 43–56.

[21] A. Edalat and D. Pattinson, "Denotational semantics of hybrid automata," in Foundations of Software Science and Computation Structures, ser. Lecture Notes in Computer Science, vol. 3921, 2006, pp. 231–245.

[22] E. A. Lee and H. Zheng, "Operational semantics of hybrid systems," in Hybrid Systems: Computation and Control. Springer, 2005, pp. 25–53.

[23] E. A. Coddington, An Introduction to Ordinary Differential Equations, ser. Dover Books on Mathematics. Dover Publications, 2012.

[24] J. B. Conway, A course in functional analysis, 2nd ed., ser. Graduate texts in mathematics. Springer, 2010, vol. 96.

 ISBN: 978-1-61208-451-0

Development of Real-time LCA System based on Automotive Radar

YoungSeok Jin, SangDong Kim, YoungHwan Ju, JongHun Lee*(corresponding author)
ART (Advanced Radar Technology) Lab.
Division of IoT and Robotics Convergence Research
DGIST
Daegu, Korea
Email: {ysjin, kimsd728, yhju, jhlee*}@dgist.ac.kr

***ABSTRACT*— In this paper, we developed a real-time lane change assist (LCA) system based on automotive radar. The existing blind spot detection system only makes detection in a very limited zone up to about 5m, thus responding more slowly to vehicles approaching at faster speeds. In order to overcome this limitation, a radar-based LCA system was developed to provide information before vehicles enter the blind spot detection (BSD) and LCA zone. The performance of the developed LCA system was verified in an anechoic chamber and driving environments.**

Keywords- *LCA*; *BSD*; *Intelligent Vehicle*; *Automotive Radar*; *Active Safety system*.

I. INTRODUCTION

Recently, many vehicles have been equipped with driver assisting systems to assist drivers in their judgement and to prevent preventable accidents. Statistics show that most car accidents are caused by carelessness or misjudgment on the part of the driver. As such, the United States and Europe have enforced regulations on the installation of driver assistance systems and safety devices.

Among the many safety systems, Blind Spot Detection (BSD) alerts the driver of possible collisions by detecting vehicles in the blind spot. The detection zone of the BSD system is shown in ISO 17387, which is presented in Figure 1 [1]. When a car approaches rapidly from behind while the driver is attempting to change lanes, the driver may not receive sufficient information in a short BSD zone [2]. A Lane Change Assist (LCA) system that provides information on target vehicles from longer distances is needed to alert the driver of vehicles approaching from behind.

Previously, BSD and LCA systems include infrared, vision, and ultrasonic sensors. However, these sensors are very sensitive to weather or have shorter detection distances. For vision sensor and ultrasonic sensors, the detectable ranges are about within 30m and 12m, respectively. Recently, to overcome these limitations, Radar based BSD and LCA systems is being developed having less sensitive to weather and with longer detection distances [3].

And, the main contribution is to extend the maximum detectable range up to 80m comparable to the conventional radar based LCA system. To do this, we enhanced SNR performance by accumulating the received beat signals using multiple chirp signals.

In this paper, we developed a real-time radar based automotive LCA system. Here, a real-time means the system operates the LCA algorithm fully every 200ms time according to the standard 17387. Then, developed LCA System is verified in a chamber and various driving test sites.

The rest of this paper is organized. In section II, we will illustrate development of a radar based automotive LCA system. In section III, to verify the feasibility of our developed LCA system, the environment of measurement and experimental results will be shown. Section IV concludes this paper.

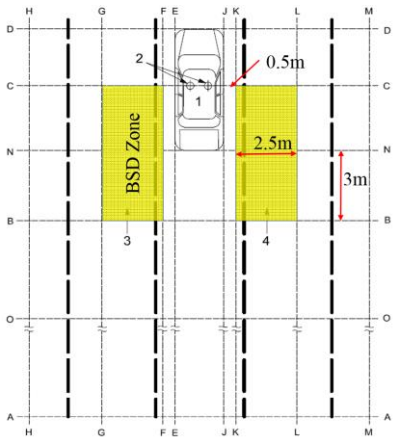

Figure 1. BSD detection zone [1]

II. DEVELOPMENT OF A RADAR BASED AUTOMOTIVE LCA SYSTEM

A. System parameter

In this section, we show the parameter of our considered LCA system in Table I. As shown in Table I, the center frequency is set to 24GHz, and the bandwidth is set to 200MHz. The Pulse Repetition Interval (PRI) is set to 80 μs, and 400 samples per PRI are sampled at a rate of 5MHz.

TABLE I. THE SPECIFICATIONS OF LCA SYSTEM

Parameter	Value
Center Frequency	24GHz
Bandwith	200MHz
Pulse Repetition Interval (PRI)	80 μs
Sampling Rate	5MHz

B. Front-End Module

We show the functional block diagram of the front end module (FEM) in Figure 2. The developed FEM operates for

 ISBN: 978-1-61208-451-0

a frequency modulated continuous wave (FMCW) radar [4] that outputs signals of 200 MHz in the range of 24.05GHz ~ 24.25GHz at 10dBm power in the transmitter.

Figure 3 shows the picture of antenna patterns in our developed LCA system. The transmission antenna is composed of 5 antenna element arrays. The transmission antenna arrays make beam shape cover the BSD and LCA zones simultaneously. There are two Rx channels having only output real signals for LCA detection, and a channel having I and Q output signals for BSD detection.

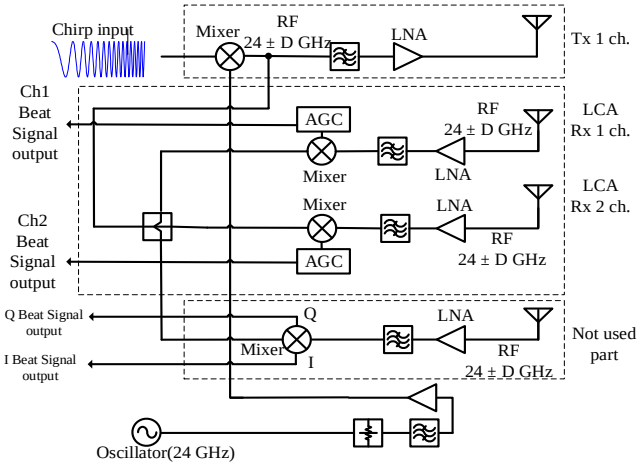

Figure 2. Fucntional block of FEM consisting of transmission antenna

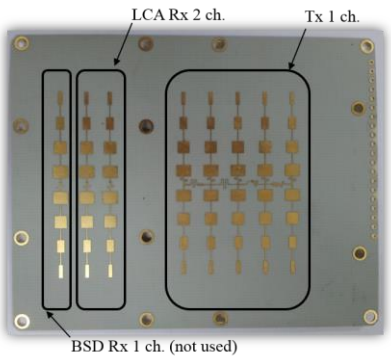

Figure 3. Antenna pattern

C. Back-End Module and Radar Signal Processing

We show the photo of back end module (BEM) in Figure 4. The BEM was developed on the TMS320 DSP processor that operates at a maximum of 150MHz. Because the DSP processor is a low-cost and high-efficiency processor, it makes our designed LCA system cost-effective. Also, it provides various interfaces, such as ADC, Direct Memory Access (DMA), and Controller Area Network (CAN), in addition to a high-performance signal processing library.

Figure 5 shows the block diagram of the developed LCA and BSD system. First, in the main task block of BEM, the RF_ON signal is transmitted to start the FEM. The FEM module receives the RF_ON signal to transmit FMCW radar signals, and outputs RF_Sync signals in a PRI period. The RF_Sync signals provide external interrupts to the BEM, which in turn leads to the external interrupt task and DMA interrupt task.

The external interrupt task triggers the ADC, and beat signals are converted into digitalized signals to be stored in the DMA buffer. The DMA interrupt task is synchronized with RF_Sync signals from the FEM, and stores the digitalized beat signals in the external RAM according to the predetermined size and number by the DMA buffer. Once storing signals is complete, the DMA interrupt task stops the external interrupt task and runs the main task. The main task runs the radar signal processing so-called detection and tracking algorithms.

Figure 6 presents the simplified radar algorithms of the radar signal processing in our developed system [6]. For distinguishing driving lane from another lane, digital beam forming [7] was applied to extract angles of target vehicles. The 2D-FFT algorithm was utilized to obtain the distances and velocities [7] and [8]. Due to the homodyne transceiver used for a cost-effective radar module, the DC–offset component is inherently included in the beat signals. For improving the performance of a near range detection, the DC offset was digitally calculated and removed. Next, a first FFT was processed the beat signals for range extraction, and a second FFT was processed the range bin in a PRI direction for Doppler extraction. Digital beam forming processing was performed the beat signals in an antenna direction for angle extraction. Finally, the target range, velocity, and angle were detected based on Constant False Alarm Rate (CFAR) and Peak Power Spectral Density (PSD) Detection. The final results provide various safety systems through a CAN communication. The total processing time of this system is within about 100ms.

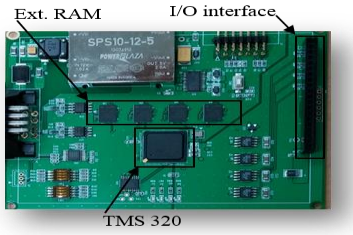

Figure 4. BEM consisting of a DSP processor

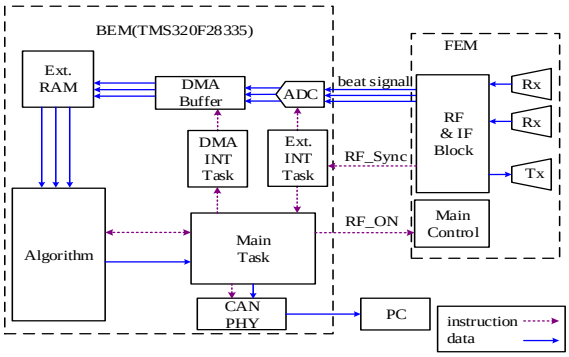

Figure 5. Entire functional block diagram of the developed LCA system

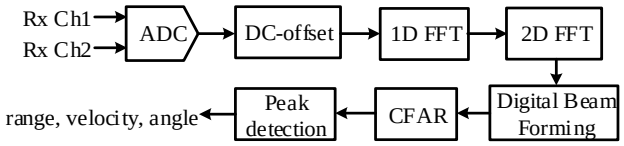

Figure 6. Simplified radar signal processing algorithm of the developed LCA and BSD system

III. MEASUREMENT RESULTS

This section illustrates the environment of measurement and measurement results. Figure 7 shows the experiment setup of our developed LCA system. The FEM and BEM are integrated using a single connection socket, and the signals are exchanged between the two. The output is passed through CAN, and the results are sent to a PC through a CAN to LAN convertor. The final results were displayed using a monitoring program.

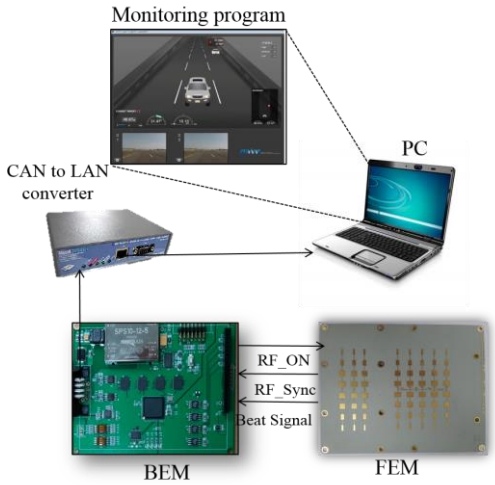

Figure 7. Verification Set-up of our developed LCA System

Table II shows the performance measurement of our developed LCA system. The 95% probability of detection is met to the OEM requirements and the range accuracy of 0.5m is good enough to meet the requirement according to the LCA standard ISO17387. The developed LCA system was evaluated in an anechoic chamber and real-road sites.

TABLE II. THE PERFORMANCE MEASUREMENTS OF OUR DEVELOPED LCA SYSTEM

Parameter	Our developed	Conventional
Maximum Detectable Range	< 80m	<50m
Probability of detection	< 95%	<95%
Range Accuracy	< 0.18m	<0.5m
Maximum Detectable Speed	250Km/h	250Km/h

Table III presents the used anechoic chamber for the performance measurement, which is a space with a length of 10m, a width of 5m, and a height of 4m. It is capable of measuring in a frequency range of 8GHz~110GHz, and has a shielding effectiveness of 60 dB at 8Ghz. Figure 8 shows the real picture of the used anechoic chamber.

TABLE III. THE USED ANECHOIC CHAMBER FOR THE PERFORMANCE MEASUREMENT

Chamber Spec.	Value
Chamber Style	Rectangular
Chamber Size	10m(L) × 5m(W) × 4m(H)
Shielding Effectiveness	60dB at 8GHz
Absorver Type	Microwave Absorber
Absorber Thickness	More than 8 inch
Absorber	More than -40dB at 8GHz

Figure 8. Used anechoic chamber for the performance measurement

First, the probability of detection and false alarm rate were verified. Measurements were taken in the chamber in cases of single and multiple targets. A total of 400 single targets were detected in 400 attempts, and 770 multiple targets were detected in 800 attempts. A detection rate of 97.5% was achieved over 1,200 attempts. Table IV shows the detection results.

TABLE IV. PROBABILITY OF DETECTION

	Target Number			
	Single Target	Multi Target	Total	percentage
Detection Number	400	770	1170	97.5%
Total	400	800	1200	

Range accuracy was measured 1000 times in the chamber with the target placed at a distance of 1m, 1.5m, 2m, and 2.5m. Table V shows the range accuracy. The range accuracy was 0.07 m at a measurement range of 1 m, 0.15 m at 1.5 m, 0.18 m at 2 m, and 0.14 m at 2.5 m. The minimum range accuracy is 0.18m.

TABLE V. RANGE ACCURACY

	Measurement Range			
	1 m	1.5 m	2 m	2.5 m
Average measurement	1.07 m	1.65 m	2.18 m	2.36 m

 ISBN: 978-1-61208-451-0

	Measurement Range			
	1 m	1.5 m	2 m	2.5 m
range				
Standard deviation	0 m	0.0001 m	0.05 m	0 m
Range accuracy	0.07 m	0.15 m	0.18 m	0.14 m

In order to evaluate the maximum detection performance, the developed LCA system was 20 times repeatedly tested on a moving target in a driving road site. The Figure 9 shows the experimental setup in the vehicle test site. Figure 9(a) shows the radar attached on the right side of the back bumper. A tilting device was installed to adjust the radar angle and direction. Figure 9(b) shows the radar-attached vehicle and the position of the target vehicle, which was set to move to the left of the radar. Figure 9(c) shows the scenario, while Figure 9(d) presents the GUI of the test results. The GUI outputs distance, velocity, angle, and camera images of the target vehicle. The target vehicle was detected at a maximum range of 80 m. The probability of detection was 97% measured.

Figure 9. Vehicle driving test site environment: (a) mount view, (b) vehicle position, (c) scenario, (d) result GUI

IV. CONCLUSIONS

In this paper, we developed a real-time radar-based automotive LCA system up to 80m without increasing output power. The system was based on a low-cost high-efficiency processor. The radar signal processing was carried out using 2D-FFT and digital beamforming algorithms. The radar was tested in an anechoic chamber and real driving road test site. Further tests will be conducted in various actual driving environments and road conditions.

ACKNOWLEDGMENT

This research was supported by INNOPOLIS Daegu programs of INNOPOLIS Foundation (2015-01-3012) and DGIST R&D Program of the Ministry of Science, ICT and Future Planning, Korea (15-RS-01).

REFERENCES

[1] "Intelligent transport systems - Lane change decision aid systems (LCDAS) - Performance requirements and test procedures," BS ISO 17387, 2008.

[2] J. D. Lee, et al., "Collision warning timing, driver distraction, and driver response to imminent rear-end collisions in a high-fidelity driving simulator," Human Factors, vol. 44, no. 2, pp. 314-334, 2002.

[3] B. F. Wu, H. Y. Huang, C. J. Chen, Y. H. Chen, C.W. Chang, and Y. L. Chen, "A vision-based blind spot warning system for daytime and nighttime driver assistance," Comput. Electr. Eng. vol. 39, no. 3, pp. 846–862, Apr. 2013.

[4] A. G. Stove, "Linear FMCW radar techniques," Proc. IEE, Radar and Signal Processing, vol. 139, no. 5, pp. 343–350, Oct. 1992.

[5] M. Klotz, and H. Rohling, "24 GHz radar sensors for automotive applications," 13th International Conf. on Microwaves, Radar and Wireless Communications, MIKON 2000, Wroclaw, Poland, vol. 1, pp. 359-362, May 22-24, 2000.

[6] V. Winkler, "Range Doppler detection for automotive FMCW radars," Microwave Conference, 2007. European, Munich, IEEE, pp. 1445-1448, Oct. 2007.

[7] P.Barton, "Digital beam forming for radar," IEE Proceedings F (Communications, Radar and signal Processing), vol. 127, no. 4, pp.266-277, Aug. 1980.

[8] E. Hyun, S. Kim, J. Choi, D. Yeom, and J. Lee, "Parallel and pipelined hardware implementation of radar signal processing for an FMCW multi-channel radar," ELEKTRONIKA IR ELEKTROTECHNIKA, vol. 21, no. 2, pp.65-71, 2015.

 ISBN: 978-1-61208-451-0

The Anatomy of IT Service Incidents

Kari Saarelainen
IT advisory
KPMG Finland
Helsinki, Finland
e-mail: kari.saarelainen@kpmg.fi

Marko Jäntti
School of Computing
University of Eastern Finland
Kuopio, Finland
e-mail: marko.jantti@uef.fi

***Abstract*—An IT service is by definition "made up of a combination of information technology, people and processes". These elements, in addition to external factors, are also the key components of IT service incidents. This paper presents an integrated model of IT service incidents. This model extends the concept of root cause also to latent, contributing conditions to an incident. Additionally, the life cycle of an incident is presented in the model. Unlike incidents and accidents in other industries, an IT service incidents has a duration. The damage caused by an incident is proportional to this duration. In our study, we show that there are events and conditions during the incident, incidents within incidents, which cause delays in service restoration. The model is validated by a case study method using 15 incident descriptions to validate both the latent factors contributing to the direct root cause as well as the life cycle of an incident. The main contribution of this study is the incident model containing latent conditions and events contributing to the direct root cause and concept of incident within incident. The model improves the traditional root cause analysis and acts as a framework in IT service incident root cause categorization.**

***Keywords*—IT service management; ITIL; continual service improvement; root cause; categorization.**

I. INTRODUCTION

Most of the vital functions are dependent on the availability and quality of IT services. Critical IT systems are often committed to deliver 99.9% - 99.999% availability meaning monthly downtime from 43 minutes to 26 seconds. At the same time, the IT service production environment is growing more complex. The service quality and availability is controlled by a set of IT service management (ITSM), risk and security management processes and practices, as well as by processes related to organizational governance. IT infrastructure library (ITIL) [1] is the most common framework for ITSM processes. There are several processes in ITIL, which would benefit from a comprehensive model of IT service incidents:

- Incident trend analysis in proactive problem management. Incident trend analysis needs a solid basis for incident root cause categorization.
- Availability management focuses on reliability and on how to put in place alternative options to ensure the service continues. It is crucial to recognize the potential areas causing unavailability.
- Service level management focuses on delivering IT services with agreed quality.
- Continual Service Improvement (CSI) is a stage in the lifecycle of an IT service, which identifies and implements improvements.

A. *Swiss Cheese Model and latent conditions*

James Reason's Swiss Cheese Model (SCM) [2][3] explains the accident and incident as a result of long standing conditions and latent failures contributing to the unsafe act. The model is often described as a sequence of planes, cheese slices, which describe organizational levels, defenses and barriers of incidents. Failures (holes in cheese slices) can emerge at anyone of these levels. When the holes are at the same time in the same trajectory, the incident is likely to occur (Figure 1).

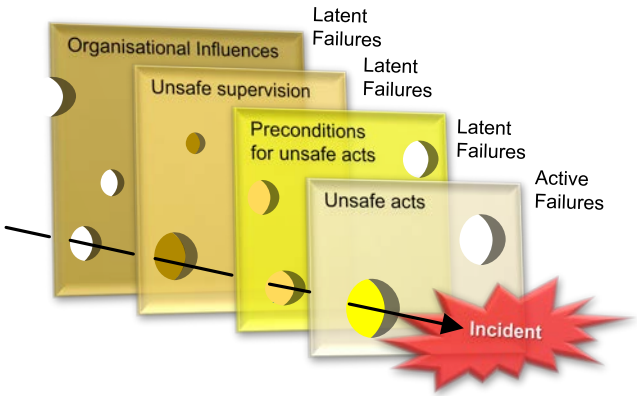

Figure 1. Swiss Cheese Model (SCM) of incident causation with HFACS taxonomy [3]

Wiegmann & Shappell [4] showed that many incidents have their roots high within the organization. Decisions made by top level management often influence the middle level management, as they supervise daily operations of the organization. The Human Factors Analysis and Classification System (HFACS) describes the taxonomy of human errors as well as the causal relationship between the unsafe act and the latent conditions behind it (Figure 2). HFACS has its origin in aviation, but it has adaptations in a broad range of other industries. Related to ITSM, HFACS is discussed in the studies [3][5][6]. HFACS, however, has some shortcomings in general and also when applied in ITSM.

- HFACS is a complex system. In aviation, where accident investigation may take weeks or months, complexity may be justified. In ITSM, service incidents are investigated in hours or days at most.
- HFACS has its roots in aviation, which is visible in its design. A model more adapted to IT work flows and ITSM environment is needed [6].
- HFACS covers only human factors. This may be justifiable in other industries, where human factors

 ISBN: 978-1-61208-451-0

cover 60-96% of incidents (Table I), but in IT the share of human errors is only 21-24%.

- The concept of incident in IT differs from the other industries. In ITIL, an incident is defined as "An unplanned interruption to an IT service or reduction in the quality of an IT service." When the service or service level is restored, the incident is over.

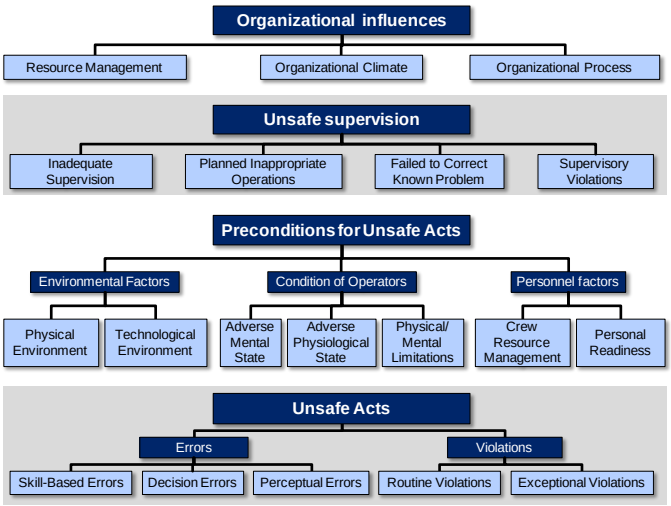

Figure 2. Hierarchical HFACS taxonomy of human factor contribution on incidents and accidents [6].

TABLE I. PROPORTION OF HUMAN ERRORS OF ROOT CAUSES OF INCIDENTS IN DIFFERENT INDUSTRIES

Industry	Human errors	Source
Aviation	70-80%	[7]
Maritime	75-96%	[8]
Railway	61 %	[9]
Healthcare	70-80%	[10][11]
Pharmaceutics	80 %	[12]
Nuclear energy	80 %	[13][14]
Chemical industry	60-90%	[15]
Telephony and Internet	19%	[16]
ITSM	18-24%	[17][18][19][22][23]

B. Other incident models and root cause analysis methods

There have been very few studies about IT service incidents, their causation, and incident models or frameworks.

Hinz [20] has studied causal modelling of end user computing. He has proposed some underlying factors, which increase probability of incidents with end users including hardware and software complexity, standardization, and maturity. The model, however, is limited to end user computing, the coverage of latent factors is rather limited, and leaves open the reasons to these underlying, latent factors.

Hazard and operability study (HAZOP) [20][29] is a process for identification of potential hazard & operability problems caused by deviations from the intended design. It was initially developed to investigate chemical production processes. HAZOP is also used for complex software systems [31]. HAZOP is, however, more a brain storming technique for system examination and risk assessment. It is a general purpose technique without ITSM specific parts. The technique finds best the direct risks and does not encourage to analyze the underlying conditions increasing the risk of incident.

ITIL presents 11 common root cause analysis (RCA) methods, which are given as examples [1]. Only two of them, namely 5-Whys and chronological analysis, address contributing causes to the direct root cause to some extent.

5-Whys RCA method works by starting out with a description of what event took place and then asking 'why this occurred'. The resulting answer is given, followed by another round of 'why this occurred'. Usually by the fifth iteration, a true root cause will have been found. 5-Whys does not, however, give a framework, where and how to look for these root causes. It is also a generic method with no adaptation to ITSM. It does not provide explanation, why to choose just the fifth root cause candidate, and omit the others.

Chronological analysis RCA method focuses on the timeline of events in order to see, which events may have been triggered by others. This method addresses clear, identified events, but misses the latent long lasting conditions. Also, it does not give a framework of formation of incidents.

The rest of the paper is organized as follows. The research problem and methods are described in Section 2. The creation and validation of the incident model is covered by Section 3. The analysis of the findings is covered in Section 4. The conclusion in Section 5 summarizes the study.

II. RESEARCH METHODOLOGY

The research problem in this study is: How the incidents in ITSM operating environment could be modelled in order to aid proactive problem management, root cause analysis and continual service improvement. The research problem was divided into the following research questions (RQ):

RQ1: HFACS brings an idea of latent conditions contributing to the incident related to human errors. Can this model be extended to technology and processes?

RQ2: What are the major differences in concept of incident in ITSM and in other industries?

RQ2: How these possible differences with the concept of incident should be taken into account in the model?

A. Data Collection Methods

Information of IT service incidents was collected in 2011 – 2014 from incident reports provided by an IT service provider organization. Incidents reflected issues from several customers and different types of environments. Multiple data collection methods proposed by [21] were used during the study and the following data sources were used:

- **Participant observation:** Meetings and discussions with managers, observation of service desk work.
- **Interviews**: Interviews of roles responsible of services offered to customers and interviews with service managers and experts involved in the incident
- **Documents:** Incident reports, process descriptions, work guides and guidelines
- **Records and archives**: The incident report pool included 215 incidents. From this pool, all the reports with three of more identified root causes or

 ISBN: 978-1-61208-451-0

contributing events or conditions were chosen. The number of these analysis units was 15.

- **Physical artifacts:** ITSM tool.

B. *Data analysis*

All the 215 incident reports were studied, and root causes, contributing events and conditions, and their sequences were extracted from the reports. In those reports, the direct root cause was clearly stated, if it was found. The other events and conditions were extracted from the narrative report text and other data sources described above. The model was build using the analyzed data and the fundamental ideas in Swiss cheese model and HFACS. 15 chosen incident reports were then applied to its model.

III. RESULTS

The main contribution of this paper is a comprehensive model of IT service incident for proactive and reactive problem management and risk management. In this section, we first present the model with its rationale and then apply it to the incident reports.

The model introduces latent factors and events contributing to the direct root cause. Additionally, it describes the lifecycle of incident containing the concept of *incident within incident.*

A. *The incident model*

1) *Top level root cause categories*

An IT service is by definition "made up of a combination of information technology, people and processes." [1]. These basic building elements of an IT service are also the major candidates for upper level root cause categories. Publicly available statistics of root causes usually omit processes, but add different external factors (e.g., forces of nature, cybercrime, etc.) to root cause categories (See Table II).

TABLE II. STUDIES OF ROOT CAUSES IN IT SERVICE INCIDENTS.

Source	Year	Technology	Human	External
Gartner/Dataquest [22]	1999	67 %	18 %	8 %
Enisa [16]	2014	66 %	20 %	14 %
Ponemon [18]	2013	58 %	22 %	30 %
Quorum [23]	2013	73 %	22 %	5 %

ITSM and other processes coordinate task flows performed by people and technology. If the process fails, the direct root cause is related to people or technology, not to process. The flaw in the process is a contributing factor in the background, and this is not usually gathered in statistics.

Figure 3 presents technology, human and external factors as direct root causes and processes in the background. Note, that the relevant process sets are different in external, technology and human factors. External factors are not controlled and managed in the same way as internal factors. One cannot set performance objectives or improve the quality of forces of nature or cyber criminals. The processes used to manage external factors include security, risk, availability and service continuity management. Architecture and procurement practices are unique to technology factors and leadership and HR processes are related to people. The incident root cause can be considered as a combination of several different factors (see Figure 3).

The idea of contributing factors has been presented by Reason [2] and later refined by Wiegmann and Shappel. [4], although restricted to human factors and originally in aviation industry.

Note that both direct causes and other contributing factors may be considered as root causes, which are by definition "the underlying causes of an incident or a problem, which if corrected would prevent or significantly reduce the likelihood of the incident's reoccurrence" [5].

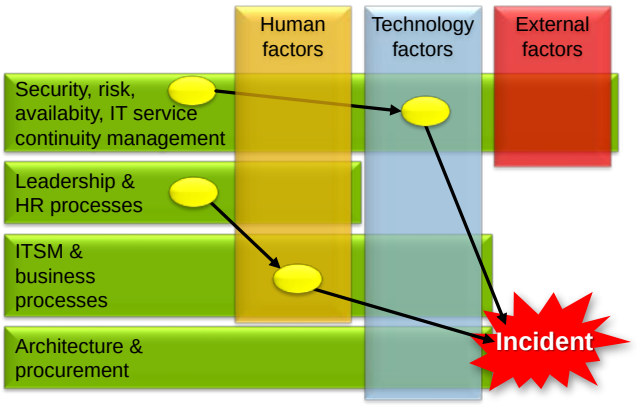

Figure 3. According to incident model presented in this paper, IT service incidents are caused by human, technology and external factors often contributed by failures in the background processes and practises.

2) *Causal levels*

The model consists of three incident root cause levels with causal relationship: Direct causes leading directly to incident, conditions to direct causes increasing probability of the incident, and organizational processes, policies, principles and practices contributing formation of conditions (Figure 4). Identifying these underlying factors helps identifying improvement possibilities in incident investigation and thus helping in making process improvements in CSI [3].

a) *Processes, policies, principles, practices (PPPP)*

Activities in an organization are guided by different processes, policies, principles and practices. In addition to ITSM processes, there are human resource and procurement policies, rewarding and motivations systems, architectural principles, cultural issues, etc. The category list of root causes in Figure 4 is not meant to be exhaustive: The difficulty of defining thorough category lists is visible already in ITIL: IT service management was handled with 10 processes by ITILv2 (2001) [24][25], while ITILv3 (2007, revised 2011) uses 30 processes and functions [26].

b) *Inadequate conditions*

Failures at PPPP causes inadequate design or behavior, which increases the probability of an incident.

Technology conditions: In technological environment failures in architecture may create complex, error prone systems. Flaws in risk calculations, and in availability management may cause non-resilient systems. Problems in procurement and IT operations & management may result in non-standard devices purchased from different sources and managed manually. Financial situation may result in savings in personnel, tools or system redundancy. In Figure 4, the maturity, standardization and complexity, design, maintenance and suitable tools are contributing technical conditions. Maturity, standardization and complexity are contributing technical conditions proposed by Hinz [20]. Other categories were extracted from the cases in this study.

 ISBN: 978-1-61208-451-0

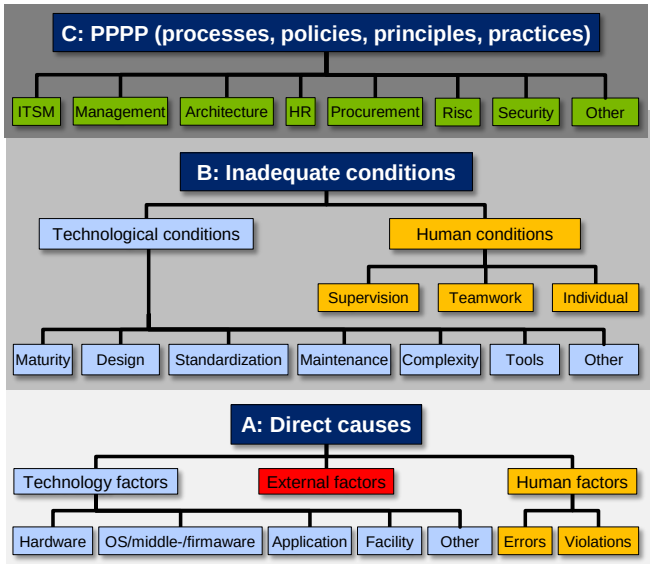

Figure 4. Causal levels of IT service incidents with indicative root cause categories in each level.

Human conditions: Saarelainen and Jäntti [3][6] have studied IT service incidents in matrix organization typical in IT service providers. According to these results, the human conditions in the model have dimensions reflecting line management (supervision) and processes (teamwork) in a matrix organization as well as individual readiness (training, cognitive factors, mental and physical state).

Note that not all the conditions are caused by PPPP. Technical devices have a measured and/or calculated mean time between failures (MTBF) that is one of the key reliability metrics within IT service availability management [27]. People may have mental or physical disabilities and behavioral features, which are beyond the control of PPPP.

c) Direct causes

Direct causes trigger the incident. They are usually reported as root causes in incident reports.

Technology factors: Technology factors are usually divided in hardware and software. Software related root causes are often further divided in operating system, firmware, middleware and application. Most ITSM root causes belong to technology factors (Table II).

Human factors: Wiegmann and Shappel [7] have made pioneering work in categorization of human errors and modelling contributing conditions. Their HFACS model as such seems to be too complex for ITSM. Having its roots in aviation it also needs adaptation to ITSM environment [3]. In this model the categories Human conditions and Human factors contain elements of HFACS simplified. At "direct causes" level human errors are categorized as unintentional (errors) or intentional (violations) by HFACS.

External factors: External factors is an umbrella term to all the factors that are beyond internal controls and process improvement efforts of the organization. Thus, they are managed only by security, risk, availability and IT service continuity management processes but not by other ITSM or business processes. If an external party, e.g., subcontractor participating in ITSM processes, it is considered non-external in this model. External factors include, e.g., vandalism, cybercrime, denial of service attacks, cable theft, flood, wind, snow/ice, lightning, other forces of nature, etc.

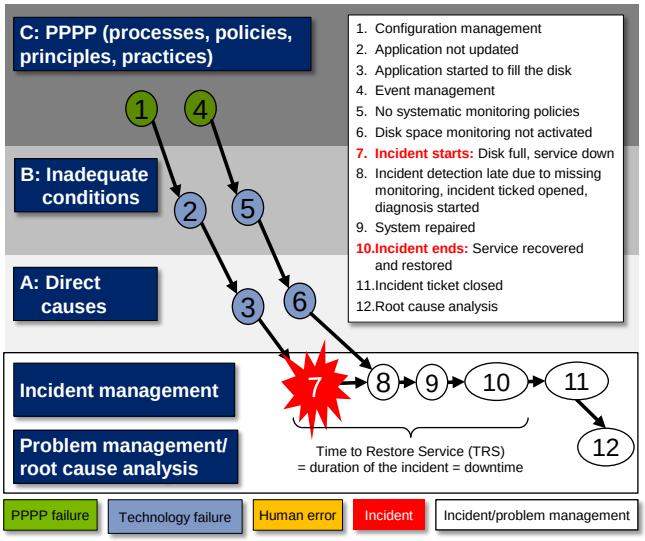

Figure 5. Extended life cycle of the incident nr. 9 in the text. The colors of background and events/conditions are as in Figure 4.

d) Incident life cycle

Traditionally, in other industries the incident or accident is over, when the damage has occurred. In ITSM, this is only the beginning of the incident (Figure 5) [27]. In point 7, the incident has taken place, in point 8 it is identified and the incident record is created (the ticket is opened) and diagnosis has started. In point 9, the system is repaired, and in 10 the service is restored. The time between points 7 and 10 is called Time to Restore Service (TRS). This is also the lifetime of the incident. If the incident causes service breaks, this period is called downtime.

3) Events before and during the incidents

The original hypothesis in this paper was related to the actual incident and events and conditions leading to it. Almost 50% of the cases under study, however, contained events and conditions that delayed the service recovery.

The core contribution of this paper is, what happens before and during the incident, and why it happens (points 1-6 in Figure 5). Figure 5 presents one case in this study, where the incident was directly caused by a software bug in applications, which started to fill the disc (point 3). This bug in turn was a known bug in an outdated version of the software (point 2). Leaving applications as outdated versions is one probable result of immature configuration management. Additionally, a failure occurred during the incident, which increased the incident duration. The detection of incident was delayed (time between points 7 and 8), because the disk was not under monitoring (point 6). Inadequate monitoring was a probable result of immature event management process (point 4), which is responsible for detection of events, including service failures. Usually, the IT service provider pays penalties stated in the service level agreement (SLA) according to cumulative service downtime. Thus, the damage to IT service provider caused by SLA penalties and also the damage to the customer caused by lost business is usually proportional to the length of the downtime or service degradation. By now, the focus in ITSM has been on direct root causes, not in contributing factors leading to this incident and not in factors increasing incident duration.

In the current ITSM practice, incident root cause analysis (point 12 in Figure 5) covers usually the direct cause of the

 ISBN: 978-1-61208-451-0

incident (point 3 in Figure 5, "Application started to fill the disk"). In this example, the usual root cause analysis approach leads to root cause "software bug". This gives a very moderate input to service improvement efforts. This study proposes that in IT service incidents root cause analysis should cover points 1-10. The latent, contributing factors leading to incident as well as contributing and direct factors slowing down service restoration should be covered in the analysis. In this example, a more extensive (points 1-10) root cause analysis approach would possibly lead to recommendation to revise configuration and event management practices.

B. *Validation of the model*

The model was validated using existing incident reports. According to our observations, the general quality of incident reports was poor. 27% (N=58) of the reports (N=215) had no identified root cause, 46% (N=99) had one root cause. 27% of reports identified two or more root causes/contributing factors. In this study, we selected all the cases (7%, N=15) with three or more identified root causes or contributing factors. Letters A, B and C in front of events/conditions refer to the causality levels in Figure 4.

1 **Unsuccessful disc space addition:** C: Configuration management -> B: Old SW version unable to repair file system -> A: File system got corrupted -> ***Incident***
2 **Network failure:** B: System was not designed redundant -> A: LAN Switch failure -> ***Incident***
3 **Unplanned service break:** C: Configuration management -> B: unsupported HW combination -> A: network failure -> ***Incident***
4 **Unsuccessful file transmission:** C: Poor instructions for testing in change management process description -> B: poor testing during the change -> A: file transfer did not work in different environment -> ***Incident*** -> A: no alarm of not successful file transfer -> ***Longer time to restore service (TRS)***
5 **Filled system log:** A: SW error -> A: System log getting full -> A: Wrong info is given to operator -> A: Log is full -> ***Incident*** -> A: Event is not identified by monitoring -> ***Longer TRS***
6 **Unplanned service break:** A: Disk space limitation of the database was not updated, when disk space was added -> ***Incident*** (Disc got full) -> B: Failure in monitoring agent -> A: no automatic alarm generated **-> *Longer TRS***
7 **Unsuccessful restoration of directories:** C: Configuration management -> B: Backup application did not support Windows version -> A: Long file names were not supported by OS but required by the applications -> ***Incident*** -> B: Access rights were not sufficient in troubleshooting activities -> ***Longer TRS***
8 **Problem with archive application**: C: Configuration management -> B: Components were not updated -> A: Old components caused performance degradation -> ***Incident***
9 **Unplanned service break:** C: Configuration management -> B: SW was not updated -> A: Old SW filled the disc -> ***Incident*** -> C: Event management -> B: No systematic monitoring policy -> A: Monitoring was not activated -> Incident is identified late -> ***Longer TRS***
10 **Network failure:** C: Capacity management -> B: Capacity of the redundant connections estimated incorrectly -> A: Router broken -> A: Overload in reserve connection -> ***Incident***
11 **Unplanned service break**: C: Capacity management -> B: Unexpectedly high amount of traffic -> A: Log files were filled -> ***Incident***
12 **Unplanned break in fax service**: C: No common change management with subcontractor -> B: Subcontractor changed the version of pdf file format -> A: Documents are not compatible with this version -> ***Incident***
13 **Network failure:** B: Broken fiber transmitter in redundant passive connection -> B: Lack of redundancy not communicated to the client -> A: Supervisory card failure in a switch on active connection -> ***Incident***
14 **Website down**: A: Network failure -> ***Incident*** -> B: Poor instructions -> A: Not all components were moved to another node in the first restoration attempt -> ***Longer TRS***
15 **Unplanned break in SAP service**: A: SW error (Application used all the memory) -> ***Incident*** -> B: Unclear text in the incident ticket -> A: Ticket routed to a wrong place -> ***Longer TRS*** **->** A: Poor documentation -> Not all the services were started -> ***Longer TRS***

IV. ANALYSIS

In the analysis stage, 15 IT service incidents were categorized according to the IT service incident model created in this study. The model was influenced by HFACS model, ITIL and the 215 incident reports used in this study. Regarding each incident, we analyzed how latent conditions had affected the formation of the incident, and how events during the incident affected restoration of service. The following five lessons learnt were identified in the analysis:

Lesson 1: *Latent, contributing factors should be investigated already in the original root cause analysis.* The quality of incident reports was poor for performing trend analysis. Only 27% of incident reports identified more than one root cause or contributing factor before or during the incident. Service improvement based on historical reports is challenging if the background of the incident is not opened. In order to identify service improvement opportunities one should go beyond the ordinary level of root cause "configuration error" or "network failure". This is in line with the findings of Saarelainen and Jäntti [6]. They also suggest, that in order to get more accurate results the principle of latent, contributing factors should be used already in root cause analysis phase.

Lesson 2: *The interface between event management and incident management needs to be clarified.* Event management focuses on managing automatic alerts created by IT infrastructure. In our case study, in four cases out of seven cases having incident within incident service restoration was delayed because of inadequate monitoring of events.

Lesson 3: *Configuration management.* There is a need for systematic configuration management. Four cases out of total 15 cases were affected by bugs and incompatibility caused by old software versions.

Lesson 4: *Duration of the incident.* Incident investigation should cover factors affecting the duration of incidents. Service downtime is one key parameter affecting the business impact of the incident and the SLA penalties. Until now the focus in incident investigation has been on the root cause, not on the downtime.

Lesson 5: *Role of tools, people, and processes.* The incident is often a mixture of conditions and events of all the above-mentioned elements added with external factors.

During our study, we identified some patterns (multiple levels of causality, "incident-within-incident") across the cases with causal relationships. In all of these cases there were at least one condition (level B in Figure 4) contributing to the direct cause (level A). Sometimes a poor process implementation (level C) was found with probable or apparent relationship to the condition.

 ISBN: 978-1-61208-451-0

In almost half of the cases, we identified a new phenomenon, namely, "incident-within-incident". This caused us to add time dimension in the model (Figure 5). This issue is addressed by ITIL indicating that change implementations may cause additional incidents. In our study we observed, that not only change implementations but also supporting tasks (incident detection, system repair, service restauration) during the incidents may trigger additional incidents.

In early phases in the analysis it was very clear, that HFACS model focusing mainly on human factors in incidents was not an adequate tool for our cases. Although previous studies and frameworks [28] have dealt with defect classification schemes, they have not provided classification that could be used successfully in the IT industry to manage a wide variety of IT service incidents. Additionally, the studies about distribution of top level root cause categories in Table II support this hypothesis. In our validation cases there were mixtures of process, technology and human related conditions and events in the very same causal link chain.

While RCA methods in general provides a process for conducting root cause analysis, 5-whys a thinking pattern, HAZOP brainstorming process, HFACS a human-centric error classification, our model aims at expanding error classification towards a more proactive and predictive model that would explain formation of incidents in an ITSM environment.

V. CONCLUSION

The research problem in this study was: How the incidents in ITSM operating environment could be modelled in order to aid proactive problem management, root cause analysis and continual service improvement. The research problem was divided into the following research questions:

Regarding the first research question (Can the idea in HFACS with latent factors be extended from human factors to technology and processes)? we found that extending the scope of top level incident root causes from human factors is a mandatory step. An incident often seems to be a mixture of all of these elements in different causal levels before and during the incident.

Regarding the second research question (Are there major differences in concept of incident in ITSM and in other industries?) we found that the IT service incident is already by definition different compared to the other industries. IT service incidents cover service failures as well as difficulties that service users experience while using IT services. An IT service incident has duration, which is a key component in the business impact of the incident.

Regarding the third research question (How these possible differences in the concept of incident should be taken into account in the model?) we found that in incident investigation more attention should be paid to events and conditions during the incident. Almost in half (7 cases of 15) of the incidents, we identified events or conditions that delayed restoring the service. As a general observation one can state, that business impact to the customer and SLA penalties by IT service provider are dependent on the duration of the incident.

There are, however, certain limitations in this study. The amount of sample cases was rather limited, only 15 cases. In future studies, we aim at conducting the study with a larger set of service incidents. The incident reports were prepared without knowing the incident model. The results would be more reliable if the incident investigator had studied systematically the latent conditions according the model. The case study results should not be used for statistical generalization but these have enabled us to extend the ITSM theory.

In the future work, adding probabilities to the model presented here may give predictive power to the model, especially if the probabilities and relationships are generated automatically or semi automatically from incident management records and configuration management database (CMDB). HFACS in the current format has limitations in ITSM environment, but an IT adaptation of HFACS may be useful related to human errors.

In the course of work, the poor quality of incident investigation and incident reports were observed. Lack of deeper analysis of contributing factors behind direct root causes gives poor starting point to proactive problem management and service improvement. This study has potential to increase incident awareness of persons working in related roles in ITSM. In order to effectively remedy the disease, one should know the mechanism.

The results in this study are useful for incident investigators and root cause analyst as well as to those involved in continual service improvement. This study gives tools to understand incidents more deeply and then fix the events and conditions increasing the probability of incidents.

ACKNOWLEDGMENT

We would like to thank the case organization's representatives for valuable feedback and responses that helped us to perform this study.

 ISBN: 978-1-61208-451-0

REFERENCES

[1] Office of Government Commerce, ITIL Service Operation. United Kingdom: The Stationery Office, 2011.

[2] J. Reason, The contribution of latent human failures to the breakdown of complex systems. Philosophical Transactions of the Royal Society of London B: Biological Sciences 327(1241), 1990, pp. 475-484.

[3] K. Saarelainen and M. Jäntti, "Quality and human errors in IT service infrastructures - Human error based root causes of incidents and their categorization." in Innovations in Information Technology (IIT), 2015 11th International Conference on. 2015.

[4] D. A. Wiegmann, Human error analysis of commercial aviation accidents: Application of the human factors analysis and classification system (HFACS). Aviation, Space and Environmental Medicine 72(11) , 2001, p. 1006.

[5] K. Saarelainen and M. Jäntti, "Human errors in IT services - HFACS model in root cause categorization, " in International Scholarly and Scientific Research & Innovation, 2015, pp. 340-345.

[6] K. Saarelainen and M. Jäntti, "A case study on improvement of incident investigation process, " in System, Software & Service Process Improvement & Innovation 22nd European & Asian Conference, EuroSPI 2015, 2015 .

[7] S. A. Shappell and D. A. Wiegmann, Applying reason: The human factors analysis and classification system (HFACS). Human Factors and Aerospace Safety 1(1), 2001, pp. 59-86.

[8] R. Hanzu-Pazara, E. Barsan., and P. Arsenie, L. Chiotoriou and G. Raicu, "Reducing of maritime accidents caused by human factors using simulators in training process, " Journal of Maritime Research, vol. V, 2008, pp. 3-18.

[9] F. Aguirre, M. Sallak, W. Schon, and F. Belmonte, Application of evidential networks in quantitative analysis of railway accidents. Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability, 2013, pp. 1748006X12475044.

[10] L. Kohn T.., J. Corrigan, and M. S. Donaldson, to Err is Human : Building a Safer Health System. Washington, D.C.: National Academy Press, 2000.

[11] S. Hunziker, A. C. Johansson, F. Tschan, N. K. Semmer, L. Rock, M. D. Howell, and S. Marsch, Teamwork and leadership in cardiopulmonary resuscitation. J. Am. Coll. Cardiol. 57(24), 2011, pp. 2381-2388.

[12] R. Cintron, "Human Factors Analysis and Classification System Interrater Reliability for Biopharmaceutical Manufacturing Investigations", Walden Univesity, Maryland, USA, 2015.

[13] Managing Human Performance to Improve Nuclear Facility Operation 2013, IAEA Nuclear Energy Series, No NG-T-2.7International Atomic Energy Agency, Vienna, Austria, 2013.

[14] S. Ziedelis, M. Noel, and M. Strucic, Human based roots of failures in nuclear events investigations. Internationale Zeitschrift Fuer Kernenergie 58(10), 2012, pp. 596-601.

[15] K. S. N. Raju, Chemical Process Industry Safety, Tata McGraw-Hill Education, New Delhi, India, 2014.

[16] C. Karsberg and C. Skouloudi, "Annual incident reports 2014," European Union Agency for Network and Information Security (ENISA), 2015.

[17] L. Shwartz, D. Rosu, D. Loewenstern, M. J. Buco, S. Guo, R. Lavrado, M. Gupta, P. De, V. Madduri, and J. K. Singh, Quality of IT service delivery — analysis and framework for human error prevention. Presented at Service-Oriented Computing and Applications (SOCA), 2010 IEEE International Conference on, 2010, pp. 1-8.

[18] Ponemon Institute Research Report, "2013 cost of data center outages," Ponemon Institute, December 2013.

[19] K. Saarelainen and M. Jäntti, Creating an ITIL-based multidimensional incident analytics method: A case study. Presented at The Tenth International Conference on Systems ICONS, 2015, pp. 24-29.

[20] D. Hinz and H. Gewald, The next wave in IT infrastructure risk management: A causal modeling approach with bayesian belief networks. Presented at Emerging Trends and Challenges in Information Technology Management. 2006.

[21] Robert Yin. Case Study Research: Design and Methods. Beverly Hills, CA:Sage Publishing, 1994.

[22] E. Marcus and H. Stern. Blueprints for High Availability: Designing Resilient Distributed Systems, John Wiley \& Sons, Inc, New York, NY, USA 2000.

[23] Quorum, "Quorum disaster recovery report Q1 2013", 2013.

[24] Office of Government Commerce, IT Infrastructure Library - Service Delivery. London: The Stationery Office, 2001.

[25] Office of Government Commerce, IT Infrastructure Library - Service Support. London: The Stationery Office, 2001.

[26] Office of Government Commerce, ITIL Continual Service Improvement. United Kingdom: The Stationery Office, 2011.

[27] Office of Government Commerce, ITIL Service Design. United Kingdom: The Stationery Office, 2011.

[28] W. Florac, Software quality measurement: A framework for counting problems and defects. Software Engineering Institute, Carnegie Mellon University. Pittsburgh, PA. 1992.

[29] T. A. Klet, Hazop and Hazan: Identifying and Assessing Process Industry Hazards, Institution of Chemical Engineers, Rugby, UK, 1999.

[30] BSI: BS IEC 61882:2001: Hazard and Operability Studies (HAZOP studies). Application Guide, British Standards Institute, UK, 2001

[31] J. A. Mcdermid, M. Nicholson, and D. J. Pumfrey, "Experience with the application of HAZOP to computer-based systems." Presented at Compass '95: 10th Annual Conference on Computer Assurance, pp. 37-48, 1995.

 ISBN: 978-1-61208-451-0