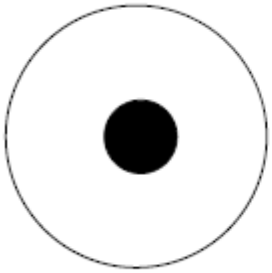
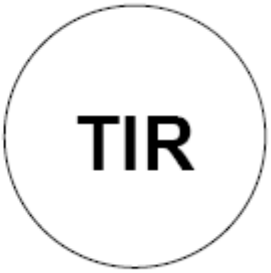




The IPSI BgD Transactions
on
Internet Research

Multi-, Inter-, and Trans-disciplinary Issues in Computer Science and Engineering

A publication of IPSI Bgd Internet Research Society, New York, Frankfurt, Tokyo, Belgrade
January 2018 Volume 14 Number 1 (ISSN 1820-4503)

Special issue: „Selected Student and Faculty Research in Information and Communication Systems and Applications“

Guest Editor: Anton Kos

Table of Contents:

Special Issue Articles:

A contribution from the regular submission:

The IPSI BgD Internet Research Society

The Internet Research Society is an association of people with professional interest in the field of the Internet.
All members will receive these TRANSACTIONS upon payment of the annual Society membership fee of €500
(air mail printed matters delivery).

Member copies of Transactions are for personal use only

IPSI BGD TRANSACTIONS ON ADVANCED RESEARCH

www.internetjournals.net

EDITORIAL
Selected Student and Faculty Research in Information and Communication Systems and Applications

Anton Kos, Guest editor

Faculty of Electrical Engineering, University of Ljubljana, Slovenia

This special issue includes cutting edge research papers in the area of information and communication technology (ICT) and in the research areas closely connected to it. The presented results are based on the research conducted by a number of post-graduate students, young researchers, international mobility students, visiting scholars, and faculty members at the Information and Communication Technology Department of the Faculty of Electrical Engineering (FE), University of Ljubljana, Slovenia.

Information and communication technologies are playing an increasingly important role in many parts of our lives. While the introduction of computers, broadband networks, Internet, smartphones, and other technologies has already changed our daily lives in many areas, similar trends are present in economy and industry. The fourth industrial revolution, described by the term Industry 4.0, introduces the automation and data exchange in manufacturing technologies. It comprises of Internet of Things (IoT), sensor technologies and networks, cyber-physical systems, mobile and cloud computing, advanced communication networks, and others. A similar role of ICT is perceived in the fields connected to human well-being, such as healthcare and sports. Here, sensor and communication technologies, coupled with cloud computing and big data analytics, are providing the means for better understanding of actions and processes of the human body.

The first four papers of this issue deal with the coming challenges for the ICT and the applications in the field of IoT. In the first paper, ***Tackling the Challenges of ICT Innovation and Talents for Industry 4.0*** (by Urban Sedlar, Andrej Kos, Matevž Pustišek, Janez Bešter, Matevž Pogačnik, Luka Mali, and Emilija Stojmenova Duh, all from FE) the authors investigate the innovative models of education and training, such as solution oriented design thinking and prototyping, which would provide all the competences needed for the challenges in the future. Also, a short review of ICT-based innovation activities and good practices at Faculty of Electrical Engineering, University of Ljubljana is given. The second paper, ***Security Risk Evaluation Methods in IoT*** (by Marjeta Pučko, Andrej Kos, and Matevž Pustišek, all from FE) addresses cybersecurity threats and security risk evaluations in the Internet of Things. The paper lists the set of existing IoT risk classification methods and presents an original risk classification method, combining the architectural and product views. It also gives some practical examples of use of their method in the IoT domains of energy production and distribution and in eHealth. In the third paper, ***Blockchain Support in IoT Platforms*** (by Matevž Pustišek, Leon Štefanič Južnič, and Andrej Kos, all from FE) the blockchain technology as a service is discussed. New scalable and trusted approaches are described that are based on fog computing and communication architectures. Currently the two viable blockchain candidates are the Ethereum and the Hyperledger Fabric. If the devices are central and payments are required, then Ethereum is the favorite. In case of business-to-business applications Hyperledger Fabric might be a preferable option. The fourth paper, ***Electric Switch with Ethereum Blockchain Support*** (by Matevž Pustišek, Nicolas Bremond, and Andrej Kos, from FE and France) presents a

prototype implementation of a device with Ethereum blockchain support for booking and payments. It provides an end-to-end solution comprised of a blockchain end-device, Web applications with blockchain clients, and corresponding smart contracts for the specific transaction logics. This approach could be extended to electric vehicle chargers, smart grid supply and demand management or other utility support systems.

The second group of papers focuses on the systems and applications for wellbeing that are based on sensor data acquisition, processing and interpretation. The fifth paper, ***Golf Swing Data Classification with Deep Convolutional Neural Network*** (by Libin Jiao, Rongfang Bie, Hao Wu, Yu Wei, Anton Kos, and Anton Umek, from China and Slovenia respectively) presents a machine learning approach to identification of golf swing shape errors. The developed system collects signals from three types of sensors attached to the golf club. It uses Deep Convolutional Neural Network to distinguish between the correctly performed swings and swings with errors from different players. The authors of the sixth paper, ***A Time-Dependent Multi-Class SVM Algorithm for Crowdsourced Mobility Prediction*** (Yuan Zhang, Anton Umek, Alex Adim Obinikpo, and Anton Kos, from China and Slovenia), discuss the algorithm for accurate prediction of the user's next location based on the crowdsourced data. This information can be useful in many situations, for example, alerting the taxi drivers and their passengers of the amount of environmental pollution at predicted locations. The proposed T-MSVM algorithm can achieve an accuracy of 90% over a week period and more than 95% accuracy over a month period. In the seventh paper, ***A Review on Methods for Assessing Driver's Cognitive Load*** (by Kristina Stojmenova, Emilija Stojmenova Duh, and Jaka Sodnik, all from FE) the authors discuss new possible methods for assessing the cognitive load of drivers. This is a quite demanding task, especially in a dynamic environment such as operating a vehicle. Methods can be divided into three groups: methods for subjective assessment, methods for indirect assessment, and methods based on psycho-physiological and neurological measures. Reliable data collection is possible mainly by using expensive, high-end equipment, which consequently makes these methods less widely assessable.

The eighth paper, ***Commutative Rotations in 3D Euclidean Space and Gimbal Spatial Angles*** (by Sašo Tomažič from FE) is closely connected to the previous paper because it proposes possible mathematical tools and solutions for an implementation of a driving simulator. It discusses commutative rotations in 3D Euclidean space and compares them to the Gimbal spatial angles that are also found to be commutative. Authors of the ninth paper, ***Automated Broadcast Video Quality Analysis System*** (Urban Burnik, Marko Meža, and Janez Zaletelj, all from FE), present the system capable of automatic evaluation of broadcast video quality. The system substitutes subjective quality monitoring run by human observers with an integrated, objective video quality evaluation system. It automatically provides validated MOS-correlated results of video quality from 5 simultaneously observed locations and is, as such, a valuable tool for quality comparison of broadcasts from several delivery providers. In the tenth and the final paper, ***An Overview of Fiber Fluorimeter Probes*** (by Ahmed Samir and Boštjan Batagelj, both from FE) the authors give an overview of fiber probes used for fluorimetry, an optical method that can quantitatively measure the fluorescence for chemical, biomedical, and clinical applications. The small size, light weight, flexibility, and non-toxicity of the optical fiber are attractive advantages that make it possible to monitor minute volumes and have the capability of remote monitoring.

The editor believes that this special issue contains a lot of interesting material and that it could serve as the guideline for those who would like to get a deeper insight into one or more of the presented topics. The editor also observes that the research in the field of ICT has become highly interdisciplinary as many of the presented papers refer to other research fields.

Tackling the Challenges of ICT Innovation and Talents for Industry 4.0

Sedlar, Urban; Kos, Andrej; Pustišek, Matevž; Bešter, Janez; Pogačnik, Matevž; Mali, Luka; and Stojmenova Duh, Emilija

Abstract: ***In the last years, the ICT innovation, research and development area has changed drastically. Nowadays, the ICT is part of the industry 4.0, health care, education, training etc., and has as such a crucial role in all parts of our lives. Developing products and services, conducting research and innovation activities, where ICT has a supporting role, means that in addition to the ICT domains, researchers and developers have to acquire domain specific knowledge from different application sectors, such as health, energy or education. This means that traditional ICT education and training are not providing all the competences needed and are therefore inefficient in meeting the needs of the industry and economy. New, innovative models of education and training are necessary and they should include solution oriented design thinking and prototyping. In addition, a short review of ICT-based innovation activities and good practices at UL FE is given.***

Index Terms: creativity, ICT, innovation, IoT, skills and competences, talents.

1. INTRODUCTION

INFORMATION and communication technologies (ICT) play an important role in today's economy, transforming practically every sector of the industry [3].

In the last years, the ICT innovation, research and development areas have changed drastically. Communication and information technologies and services have become a commodity in most parts of the world. Broadband networks, wired and wireless, as well as datasets, accessible via open APIs, are increasingly available as resources all the time and from everywhere, similarly as mains wall plugs and water from a tap.

There is still basic research within core ICT domains, however it is present in fewer areas, i.e. mobile, quality of experience, software defined networking/radio, cyber security etc. A great deal of research has moved to applicative, interdisciplinary areas, where ICT plays just a part, although an important one, of the solution.

As shown later on, nowadays ICT is a part of health care, automotive industry, education, training etc., and has, as such, a crucial role in all parts of our lives.

The trends in ICT go towards the Internet of Things (IoT). The number of connected devices already exceeds the world population; and, according to forecasts, it should increase to more than 30 billion by 2020. The development and expansion of the IoT is mainly due to three underlying factors: (i) a significant increase in the processing capacity of electronic devices (i.e., Moore's Law), (ii) proliferation of communication technologies that power the connectivity of a large number of devices into a common network, and (iii) cloud software that enables large-scale data processing and extraction of actionable information.

At the same time, the commoditization makes technology more affordable to everyone, which fosters tinkering and has the potential to level the playing field between established companies and start-ups. In such environments, the market differentiation is more than ever a function of creativity and innovation.

Developing products and services, conducting research and innovation activities, where ICT has a supporting role, means that in addition to ICT domains, researchers and developers have to acquire domain specific knowledge from different domains. This means that the traditional ICT education and training are not providing competences needed and are therefore inefficient in meeting the needs of the economy. Therefore, new, innovative models of education and training are necessary.

Manuscript received June 10, 2016; revised June 27, 2017; accepted June 28, 2017. The work was supported in part by the Ministry of Education, Science and Sport of Slovenia, the Slovenian Research Agency within the research program Algorithms and optimization methods in telecommunications, and the European Union within the ERUDITE Interreg Europe Central project (Corresponding author: Andrej Kos.)

U. Sedlar, A. Kos, M. Pustišek, J. Bešter, M. Pogačnik, L. Mali, E. Stojmenova are with the Faculty of Electrical Engineering, University of Ljubljana, Ljubljana SI-1000, Slovenia, e-mails: urban.sedlar@fe.uni-lj.si; andrej.kos@fe.uni-lj.si; matevz.pustisek@fe.uni-lj.si; janez.bester@fe.uni-lj.si; matevz.pogačnik@fe.uni-lj.si; luka.mali@fe.uni-lj.si; emilija.stojmenova@fe.uni-lj.si

2. ICT AS PART OF INNOVATION IN ALL DOMAINS

There is still basic research going on in the field of core ICT domains; however, it focuses on very specific topics, such as next-generation mobile technologies, quality of experience, software defined networking, software-defined radio, cyber security, etc.

These researches try to advance the state of the art and remove various inefficiencies, but on a larger scale, they do seem to be subject to the law of diminishing returns. At the same time, a great deal of ICT-related research has moved to applicative and interdisciplinary areas, where ICT serves just as a background infrastructure and enabler for the domain-specific solution.

ICT has become an indispensable part of all parts of our lives. As the Figure 1 shows, it is part of health care, industry, education and training, etc.

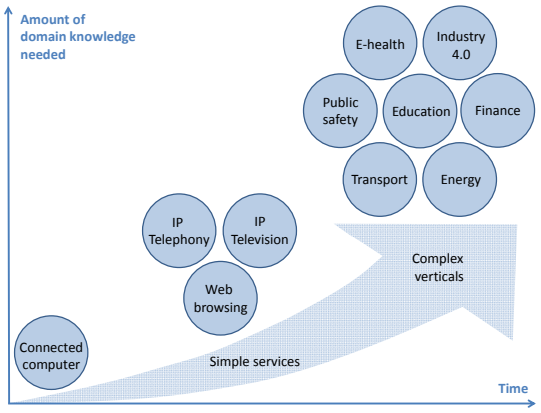


Figure 1: ICT in various economy sectors

Conducting research and innovation, where ICT has the role of supporting and enabling services means that researchers have to acquire specific knowledge from a number of different domains. Moreover, for this time, money and competences are needed.

Today, competitiveness and innovation depend on mastering a wide range of information and communication technologies. Firstly, it has become possible to deploy ubiquitous sensing solutions to gain awareness of many industrial processes, as well as to disrupt entire industries. There is a plethora of already available devices that can collect and share data, for example: (i) vehicles, home automation equipment, city infrastructure, (ii) wristbands and smart watches, smart textiles, and other wearables, (ii) industrial robots, cameras, sensors, etc.

However, data collection is only the first step on the way to obtain actionable information. Data needs to be transferred, stored and processed, for which many novel approaches have appeared, such as map-reduce, in-memory processing, stream mining, etc. In addition to that, knowledge of basic machine learning techniques, mathematics/statistics and visualization is needed for almost any work in that area; this has led to a new profession of data scientist – an expert versed in the above skills that can extract knowledge from the data.

However, before processing the data, one has to know what questions to ask. For this, a suitable domain knowledge is needed, which can be either provided by an expert in the field, or inferred by observing the processes, conducting interviews and performing requirements engineering.

Only once the whole chain is completed, resulting data and algorithms can be used to close the feedback loops and drive complex systems in an optimal manner.

3. INDUSTRY 4.0

Lately, the term that describes the above-mentioned ICT innovation used for growth in industry has a name – Industry 4.0. Industry 4.0 is a trend of automation and data exchange in manufacturing technologies. It includes cyber-physical systems, the Internet of things and cloud computing. Industry 4.0 creates what has been nicknamed a "smart factory". Cyber-physical systems monitor physical processes, create a virtual copy of the physical world and make decentralized decisions. Over the IoT, cyber-physical systems communicate and cooperate with humans and with each other in real time; and via the Internet of Services (IoS), both internal and cross-organizational services are offered and used by participants of the value chain [11,12].

4. HUMAN RESOURCES AND DEVELOPMENT

The field of human resources development is extremely important especially in the domains of ICT/IoT and Industry 4.0 for two reasons. (i) Not only Slovenia, but also the entire European Union is suffering from a long-term lack of ICT experts to fill the need for more than 500.000 ICT jobs [2]. This has a direct impact on the weakening of innovation and competitiveness of industry and economy. (ii) This area requires very dynamic change or adaptation of competences, since we can never talk about one branch, but a distinctive inter-branch or multidisciplinary approach in ensuring appropriate competences. This approach requires close engagement of the industry and economy, education as well as other governmental and non-governmental organizations in attaining both formal and non-formal knowledge and skills.

The Generation X was followed by the generation of the millennials. The millennials represent individuals born between the early 1980s and early 2000s. However, the Generation Z (post-millennials or digital natives) is different

and therefore they need different ways of motivation. Some of their most important values are togetherness and visual communication. What is very important for them is the need to make things, not just share them, as it was the case in the former generation.

The modes of work have changed radically over time, as is also shown in Figure 2. The figure is based on UK data and statistics [4]. Workers entering the labor market need different motivation compared to those already in the labor market. This also changes the work process and consequently the competences employees should have in order to get involved in the working environment as soon as possible. As it can be seen from Figure 3, the most important competences in 2015 were complex problem solving, coordinating with others, and people management. By 2020, these three will be replaced by complex problem solving, critical thinking, and creativity [4].

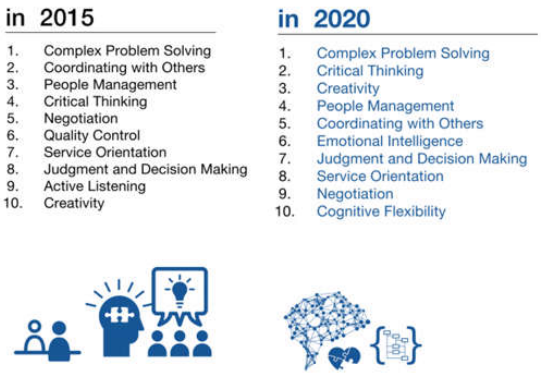


Figure 2: Top 10 skills in the fourth industrial revolution in 2015 and 2020 [5]

The learning process must adapt to new needs of today's generations and to new needs of the labor market. All these key processes, changes, and technologies (digital technology, Industry 4.0, IoT, ICT) are also strongly reflected in the field of personnel acquisition, education and training, which can follow and shape new research, development, and business opportunities. This also implies digital transformation of the school and research system.

Furthermore, the transformation of the school and research system will also have to bring, in addition to the new teaching methods, better collaboration of the school system and engagement with companies (co-curricular programs, interconnection of personnel, and integration into the pedagogical and work process) [6]. It is about connecting at all levels (kindergarten, primary and secondary school, apprenticeship, practice, scholarships, higher education, incubators, accelerators, technology parks, related organizations, and companies).

The key areas to be addressed in solving these issues are listed below:

a) Improving and extending ICT subjects in elementary and secondary schools, increase the number of hours and opportunities for practical learning of the ICT/IoT technical content.
b) Preparation of study programs for teachers in elementary and secondary schools.
c) Upgrading and extending university programs and courses in the field of the ICT/IoT and increasing the number of enrollment places.
d) Preparation of specialized educational programs for the ICT/IoT and Industry 4.0 companies.
e) Increasing the creative and fabrication literacy of people (f-literacy).
f) Planned and coordinated promotion of the ICT/IoT field among young people, with a special emphasis on the promotion of the ICT/IoT field among girls.
g) The IoT devices in schools: schools should be equipped with the IoT devices and lectures on them should be included in engineering, computing, and/or science.

5. *Innovation and Activities Belonging to Talents at LTFE and LMMFE*

In this chapter, we give a short review of ICT-based innovation activities and good practices at the Laboratory for Telecommunications (LTFE) and the Laboratory for Multimedia (LMMFE) at the Faculty of Electrical Engineering of the University of Ljubljana. In relation with the above-mentioned issues, we are focusing on:

a) Prototyping: MakerLab [7] and the national FabLab network [10]
b) Learning by doing: within the course Interdisciplinary projects [8]
c) Multimedia: new study program for next generation talents [9].

MakerLab is an open prototyping laboratory for students and creators, who organize work in the laboratory, prepare content, teach their peers and other creators, and participate in innovative projects. MakerLab follows basic principles for teaching science, technology, engineering and mathematics (STEM) courses to Millennial and Generation Z students, like learning by doing, teamwork, interdisciplinary projects, solving the real life problems, and collaborating with start-ups and well-established corporations.

In 2016, MakerLab offered support to over 150 students and other creators, both through preparation of practical seminars and final projects, and through self-initiative projects. In addition, more than 15 free of charge workshops with more than 250 participants were organized. Participants gained knowledge in the field of ICT/IoT product design and development.

The main goal of the MakerLab is to encourage young people to start exploring new technologies,

enroll into technology study courses as well as to actively involve themselves in participating in innovative educational and development projects.

Some of the most prominent and innovative projects of the MakerLab are Olympic Countdown Clock, SmartFroc, and T.A.F.R.

Olympic Countdown Clock is an interactive countdown clock that was counting down to the start of the Olympic Games in Rio de Janeiro in 2016 and is currently counting down to the start of the next games in Pyongyang. For this project, two custom developed embedded platform were used, each of them controlling a set of LED matrix displays. The clock is connected to the Internet, which allows us to replace the displayed content remotely and to monitor the operation of the clock through several sensors inside the sculpture. The project was realized in cooperation with the Slovenian Olympic Committee and the clock is installed in Ljubljana.

SmartFroc is an adjustable chair for children aged 1-10 years and features built-in weight sensors that allow adults to measure a child's weight through a smartphone application. The chair has four built-in load sensors that, in combination with a custom developed Bluetooth 4.0 electronics board, measure the child's weight and sends the data to a connected smartphone. The electronic board is neatly hidden inside the chair's legs. The project was developed in cooperation with Slovenian wood product manufacturer.

T.A.F.R. is an autonomous farming robot that monitors plants, applies fertilizers and pesticides and helps with everyday work at the farm. With a robotic helper, the chores on the field get cheaper, are done faster, and can be remotely managed.

Fabrication labs (or FabLabs) represent prototype environments for promoting innovations and inventions in the fields of modern digital technologies, ICT and IoT applications. They help increase creative literacy, which means that people can use new high-tech tools. They are dedicated for creators, students, researchers, and entrepreneurs who want to express their creativity in the form of development of innovative products with high benefit. In addition to the basic tools found in classical workshops, FabLabs have modern equipment such as 3D printers, CNC milling machines and laser cutters. Modernly furnished rooms represent only the first step; mentors that help creators overcome problems on their way and through education involve inexperienced creators in the FabLab form the second step. The third step represents linking of creators to groups that encourage formation of ideas and mutual motivation to stay on this difficult journey. Networks of related laboratories, exchanging knowledge flows, and equipment form the fourth step, which also opens up important opportunities for linking with the industry and financing the projects in the early stages of product development. FabLabs enable industry, and especially small and medium-sized enterprises, to test their ideas before entering the path of digitization.

Knowledge in engineering – most importantly communications, programming and media skills – upgraded by participatory team student work, learning by doing and design thinking with prototyping are essential parts of the Multimedia study program and Interdisciplinary projects course at the Faculty of Electrical Engineering, University of Ljubljana.

The students and youngsters are motivated to participate at different innovation activities by interesting topics and mostly by the ability to work on real and "cool" projects. For some of their activities they also earn ECTS credit points, but as we see from experience, this is not their main motivation.

For the innovation activities other than study programs, we currently do not have any systematical funding scheme in place. We are covering the costs from participation in Horizon 2020 and Interreg projects. Such kind of financing, from project to project, is not well suited for this kind of activities. Our goal is that National Slovenia FabLab network become recognized as an innovation priority in Slovenia and thus get national system funding. Similarly, our goal is to have the course Interdisciplinary projects more intensively backed by university in terms of logistics and finance.

6. CONCLUSION

ICT/IoT technologies and services have become a commodity and are transforming all and every sector of the industry and economy (digitalization). Therefore, competitiveness and innovation needed for Industry 4.0 depend on mastering a wide range of ICT/IoT skills and competences.

However, conducting research and innovation, in different domains in which the ICT/IoT has the role of a supporting and enabling service, requires from the researches to have specific knowledge from a number of different domains as well as a broad and open perspective of issues and possible solutions.

Matching the Industry 4.0 needs for talents and innovation with the needs and perspectives of the generations X, Y and Z is a challenging task.

Some of the most important values of the new generations are togetherness, communication, sharing and (very importantly) a need to create and make things by themselves.

Transformations of the industry and economy combined with the drive and motivations of young generations, implicates the need for transformation (how the educational and innovation processes run) at universities.

The steps to tackle the challenges of the ICT

innovation and the ICT talents for the Industry 4.0 that we, at the University of Ljubljana, Faculty of Electrical Engineering, have already taken are: (i) MakerLab (prototyping), (ii) the course Interdisciplinary projects (team work, design thinking, inter- and multidisciplinary solutions), and (iii) new Multimedia study program to prepare young talents for new workspace reality such as innovation, changing of the domains, shorter term commitments, solution design, etc.

Our ongoing efforts focus on involving many more students from the whole University of Ljubljana to participate in Interdisciplinary projects course, increasing the share of female students, and setting-up a national Slovenia FabLab network.

REFERENCES

[1] Nordrum Amy, "Popular Internet of Things Forecast of 50 Billion Devices by 2020 Is Outdated", IEEE Spectrum, http://spectrum.ieee.org/tech-talk/telecom/internet/popular-internet-of-things-forecast-of-50-billion-devices-by-2020-is-outdated, August 2016

[2] European Commission, "Digital Skills", https://ec.europa.eu/digital-single-market/en/policies/digital-skills, May 2017

[3] Roland Berger, "The Digital Transformation of Industry", https://www.rolandberger.com/en/Publications/pub_digital_transformation_industry.html, February 2015

[4] Fourhooks, "The Generation Guide - Millennials, Gen X, Y, Z and Baby Boomers", http://fourhooks.com/marketing/the-generation-guide-millennials-gen-x-y-z-and-baby-boomers-art5910718593/, June 2017

[5] World Economic Forum, "Future of Jobs Report, World Economic Forum", http://reports.weforum.org/future-of-jobs-2016/, June 2017

[6] C. Chun, K. Dudoit, S. Fujihara, M. Gerschenson, J. A. Burns, A. Kennedy, B. Koanui, V. Ogata, J. Stearns "Teaching Generation Z at the University of Hawai'i", https://www.hawaii.edu/ovppp/Leaders/files/2015-2016-Projects/PELP_GenZ_PaperV.6.0-5.4.16.pdf, June, 2017

[7] Makerlab Ljubljana, http://www.maker.si/index_en.php, June 2017

[8] University of Ljubljana, Faculty of Electrical Engineering: "Interdisciplinary projects", http://www.fe.uni-lj.si/en/education/2nd_cycle_postgraduate_study_programme/electrical_engineering_msc/subjects/20141229112 53029/, June 2017

[9] University of Ljubljana, Faculty of Electrical Engineering: "Multimedia study program", http://www.fe.uni-lj.si/en/education/1st_cycle_academic_study_programme/multimedia/presentation/, June 2017

[10] ERUDITE Interreg Europe Central project, "Enhancing Rural and Urban Digital Innovation Territories", https://www.interregeurope.eu/ERUDITE/, June 2017

[11] Wikipedia, "Industry 4.0", https://en.wikipedia.org/wiki/Industry_4.0, June 2017

[12] M. Hermann, T. Pentek, B. Otto, "Design Principles for Industrie 4.0 Scenarios", 49th Hawaii International Conference on System Sciences, January 2016, http://ieeexplore.ieee.org/document/7427673/?arnumber=7427673&newsearch=true&queryText=industrie%204.0%20design%20principles, June 2017

Urban Sedlar (M'07) received his Ph.D. in electrical engineering from the Faculty of Electrical Engineering, University of Ljubljana, Slovenia, in 2010. He is currently assistant professor at the Faculty of Electrical Engineering, University of Ljubljana. His work focuses on Internet technologies and protocols, quality of service, and quality of experience in fixed and wireless networks, and applications of distributed sensor networks in domains of infrastructure monitoring, e-health and emergency systems.

Andrej Kos (SM'98) received the Ph.D. degree in electrical engineering from the University of Ljubljana, Ljubljana, Slovenia, in 2003. He is a full professor at the Faculty of Electrical Engineering, University of Ljubljana, Slovenia, head of the Laboratory for Telecommunications, and the chair of University of Ljubljana Innovation Commission. He started working in the field of telecommunications in 1996. Since 1999, he has specialized in modeling and designing high-speed networks and services. Currently, he centers his work on broadband systems and applications of the Internet of things. Prof. Kos was part of the team that set up the MakerLab.

Matevž Pustišek received a Ph.D. degree in electrical engineering from the University of Ljubljana, Ljubljana, Slovenia, in 2009. He is a senior lecturer at the Faculty of Electrical Engineering, University of Ljubljana, Slovenia. His research is focused on Internet services and applications, including mobile, Web, and IoT. A special interest is oriented towards the IoT architectures and security aspects. Recently additional focus is set on the use of block-chain technologies in the IoT. At present, he is collaborating in the Ekosmart project (http://ekosmart.net/en/ekosmart-2/) on smart cities and communities.

Janez Bešter received a Ph.D. degree in electrical engineering from the University of Ljubljana, Ljubljana, Slovenia, in 1995. He is a full professor at the Faculty of Electrical Engineering, University of Ljubljana and the Head of Laboratory for Multimedia. His work focuses on implementation and application of multimedia technologies into education and economic opportunities for knowledge-based societies. He leads different projects, bridging the gap between industrial development and academic research. In 2014, Professor Bešter was part of the team that set up the MakerLab, the first open laboratory devoted to the talents at the University of Ljubljana.

Matevž Pogačnik received a Ph.D. degree in electrical engineering from the University of Ljubljana, Ljubljana, Slovenia, in 2004. He is presently employed as an associate professor with the Faculty of Electrical Engineering in Ljubljana. He led the preparation of the first (graduate) cycle University study program of Multimedia at the University of Ljubljana. His research and scientific work focuses on development of interactive multimedia services for different devices with a special emphasis on user experience and different interaction and presentation modalities.

Luka Mali received his Ph.M. in electrical engineering from the Faculty of Electrical Engineering, University of Ljubljana, Slovenia, in 2016. He is a research associate at the Faculty of Electrical Engineering, University of Ljubljana. His work focuses on Machine-to-Machine communications, Low Power Wireless Networks, Connected Devices, Internet of Things, Industry 4.0, and Smart City applications. He is the Head of MakerLab, the first open laboratory for the young innovators at the University of Ljubljana.

Emilija Stojmenova Duh (M'10) received a Ph.D. degree in electrical engineering from the University of Maribor, Maribor, Slovenia, in 2013. After graduation in 2009, she was employed as a user experience manager at a large multinational telecommunication company Iskratel, where she obtained valuable experience in telecommunication industry. She is currently an assistant professor at the Faculty of Electrical Engineering, University of Ljubljana. Her research work focuses mainly on user-centered design, design thinking, and open innovation and is putting a lot of effort in building the national FabLab Slovenia network.

Security Risk Evaluation Methods in IoT

Pučko, Marjeta; Kos, Andrej; and Pustišek, Matevž

Abstract: ***The paper addresses the field of cybersecurity threats and risk evaluation with focus on the Internet of Things (IoT) where, for the business and private users, it is extremely difficult to get a balanced picture about risk severity. The reasons are the amount of different data sources, lack of common methodology, and market orientation of the security reports. An important part of risk evaluation methodology is a risk classification. In the paper we overview a set of existing IoT risk classification methods regarding restrictions that they are either architecture or product oriented. We present an original risk classification method, combining the architectural and product views with the view of business risks on top of risk classification. Practical examples of use in the IoT domains of energy production and distribution and in eHealth are also given.***

Index Terms: ***cybersecurity* threats, *Internet of Things (IoT), information security risk classification, information security risk evaluation***

1. INTRODUCTION

It is evident that with the continued increase of the Internet use, particularly via mobile devices, the gap between the level of global cybersecurity threats and the ability to early detect the risks and prevent from cyberattacks increases continuously, too. According to Eurobarometer cybercrime research [1] just under a half (47%) of EU citizens feel well informed about the risks of cybercrime. At the same time, European internet users express a high level of concern about cyber security. For example, 89% agree that they avoid disclosing personal information online and 85% agree that the risk of becoming a victim of cybercrime is increasing. 73% agree they are concerned that their online personal information is not kept secure by websites. Around two in three Internet users in the EU are concerned about experiencing identity theft (68%) and about discovering malicious software on their devices (66%). The same source also reports that these levels of concern about specific types of cybercrimes are considerably higher than in previous years, with the largest increase in relation to the identity theft (up to 16 percentage points).

KPMG [2] shows that many organizations lack the insight, both in terms of the outside threats and in terms of what is at stake for their organizations: 48% say that employees are not sufficiently aware of cyber risk, 44% of executive boards are not sufficiently aware of the risks of cybercrime, 51% cannot detect ongoing attacks, and 59% are not convinced that their service providers know how to defend against cyberattacks. 75% of respondent organizations agree that the main driver for intensifying controls is the occurrence of an incident and 51% believe that cyberattacks cannot be prevented.

According to Cisco 2016 Annual Security Report [3] particularly small and medium businesses pay less attention to security risks and security threat defenses. Only about 40% of the organizations use mobile security, secured wireless, and vulnerability scanning. For all of mentioned defenses their use was reduced about 10% in comparison to the year 2014.

However, a high level of awareness and concern about information security can be hardly achieved for business and private users without providing a clear picture on risk severity based on common risk evaluation methodology. In this paper we address the field of cybersecurity threats and risk evaluation in the Internet of Things (IoT), with focus on the methodological level to provide more balanced risk classification in the process of risk evaluation and planning of defenses.

2. MOTIVATION AND OBJECTIVES

2.1 Background and Motivation

Despite the enormous number of available public and private information sources on the state of cybersecurity threats, it is extremely difficult to get a realistic insight into actual security threats and risks. The reasons are the amount of different data sources, lack of common methodology and market orientation of security reports, provided by leading security equipment vendors from the viewpoint of their product portfolio. For an average Internet user and even for an ICT security professional, the actual state

Manuscript received June 10, 2016; revised June 27, 2017; accepted June 28, 2017. The work was supported in part by the Ministry of Education, Science and Sport of Slovenia. T. C. Author is Marjeta Pučko is a private consultant and lecturer, Slovenia (e-mail: marjeta.pucko@guest.arnes.si). A. Kos and M. Pustišek are with the Faculty of Electrical Engineering, University of Ljubljana, Ljubljana SI-1000, Slovenia, e-mails: andrej.kos@fe.uni-lj.si; matevz.pustisek@fe.uni-lj.si.

of cyber security and severity degree of security risks in her/his sphere of use is in practice difficult to estimate. There are so many data sources available and various alerts dispatched that the current situation leads more to confusion of business and private users rather than to supportive feeling. What they would require is a realistic cyber risk understanding as the base for the effective security management. IoT specifics brings, in addition to the security threats related to information technology, additional threats related to building blocks of operational technology and smart objects, where each device can be at the same time a target, as well as a possible entry point of attack.

2.2 Objectives

Objectives of this research are particularly to:

- overview the relevant existing classifications of IoT risks and ability to cover different views,
- develop an integrated classification, wide and open enough for use in different IoT domains (energetics, industry production, automotive, health, etc.), business oriented, and based on international information security management standards,
- provide practical examples of use in different domains of IoT.

2.3 Terminology

Terms related to information security and the IoT are used in the paper with reference to the ISO/IEC 27001:2013 considering general terms of information security systems [4]. Terms considering information security risk management are used with reference to the ISO/IEC 27005:2011 [5] and in reference to the security risk assessment methodology with reference to IntelliGrid environments [6].

3. RELATED WORK

We analyzed a set of existing classifications to cover different views of technology and products relevant for inclusion in an integrated view.

Cvitić, Vunjić and Husnjak [7] presented an IoT technological layer-oriented classification made bottom up from the perception layer to the application layer. Threats and vulnerabilities and protections are systematically analyzed for each layer considering, also, particular technologies (as Bluetooth, 6LoWPAN, etc.). The work is focused on the technology aspect and provides a solid methodology base to cover the IoT architecture and technology view of risk classification.

Cisco classification [8] is based on types of the connected devices defined as an origin of potential vulnerabilities/security risks. The Internet of Things is defined as the convergence of IT networks, operational technology (OT), and smart objects where each device can be a target of a possible attack. The model considers both cyber security and physical security to protect the operational technology and information technology networks working together. However, different security policies and priorities can be applied when needed. The classification covers the technological and the product view of different IoT building blocks and provides a security model regarding visibility of security events, control over security policy, analytics of real-data data from network and end-devices, and decision support.

Another product oriented view to risk threat/risk classification is used in the Security intelligence and event management (SIEM). SIEM solutions include products designed to aggregate data from multiple sources to identify patterns of events that might signify attacks, intrusions, misuse, or failure. The SIEM products are currently in a transitional period, which stands at a crossroads between legacy SIEM solutions, and newer solutions focused on the integration of big data, network forensics, and User and Entity Behavior Analytics (UEBA) focused tools [9]. The products are being improved by threat intelligence [10], with the addition of behavior profiling and better analytics. A similar classification concept is used in the unified threat management (UTM) and the next generation firewall product view [11]. It is based on product capability and feature strengths. Devices combine multiple security functions under one roof, incorporating next-generation firewall, intrusion prevention system (IPS) functionality, antimalware, virtual private networks (VPN), application control and other threat detection mechanisms. For now the UTM products lack more granular features with strong analytical support.

3.1 Security Affecting Features of the IoT

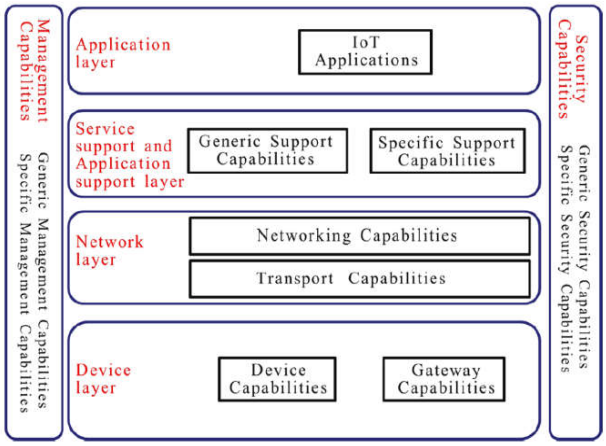


Figure 1: ITU-T Y. 4000 IoT reference model [12]

The IoT reference model [12] refers to security through the overall architecture (Figure 1).

There are several aspects of the use and operation of the IoT systems, which have a significant impact on their security evaluations.

In terms of IoT system architecture—as shown in Figure 1 - an IoT can be decomposed to smart (physical) devices, communication gateways and networks, and set of backend systems, usually provided as cloud-based services. Cloud backend systems collect, process, analyze, and act on data generated by connected devices, enable long term storage and (big data) analytics, and facilitate easy development of the IoT applications. Security capabilities stretch along all these layers.

The IoT devices can be numerous and very diverse. Often they have limited communication and computation resources, which can inhibit the use of the most advanced security algorithms. Moreover, cost reduction of the devices is often of key importance and life-cycles in IoT ecosystem can be longer than in e.g. use of mobile phones or computers. Along with increased users' expectation for ubiquitous and easy to use service, security in the IoT can be neglected for sake of utility, too.

4. CLASSIFICATION

4.1 Integrated Classification

The classifications described in Section 3 provide a detailed insight in different aspects of security risk evaluation and serve as the base of a multilevel integrated classification, taking into account also the business part of security risks.

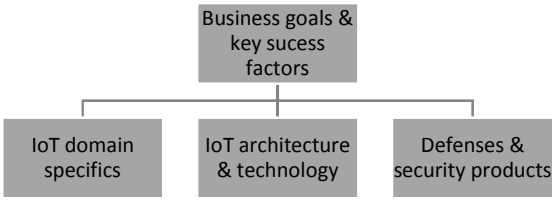


Figure 2: Levels of an integrated view to risk classification

Each risk of the view on higher level is decomposed into risks of the view on the lower level. The end result of this classification method is a tree of security risks, starting with the business view at the top and detailed technological view at the bottom. The classification level structure is presented in Figure 2. Risk classification starts with consideration of key business goals, success factors and identification of related security risks. At the technology level, security specifics of the IoT domain (data privacy requirements, regulation etc.) is analyzed in the next step, followed by the architecture and technological view to analyze and identify the IoT architecture layers with potential risks. In the final step of classification, actual defenses and security products are analyzed to secure the selected

Table 1: Security risks classification – an example of the smart grid IoT domain

View of risk evaluation	Outline	Risk description	Severity level
Business goals and key success factors	Increasing number of end customers	Low level of confidence in maturity and security of smart grid solutions	Medium
	Total operational cost reduction	Initially no cost reduction through investment, data breaches and opportunity costs	High
	Distributed sources of energy		
IoT domain specifics	Devices installed at every household	Dependent on existing energy grid infrastructure	High
	High manageability of network and end devices		
	Responsibilities for implementation depend on national regulation/policy	Service coverage dependent on geographical area	Medium
IoT architecture & technology	Use of smart meters	Architecture not designed for high level of security	High
	Use of smart gateways	Vulnerabilities in energy grid implementation, data breaches	High
	Use of SCADA systems		
Defenses and security products	Network protection devices – threat management	Insufficiently protected network, in particular wireless	Medium
	Analytic tools (SIEM)	Analytics not implemented or insufficient to detect security incidents	High
	IoT device management, hardware security		

architecture. The end result is a tree with the business view at the top (at the root node) and subsequently derived technological view of risks at the bottom (at the leaves).

4.2 Use in the Smart Energy IoT Domain

Electric power grid is facing rapid changes that reflect gradual introduction of distributed power sources (e.g. photovoltaic, wind), increase in number of electric vehicles, automatic meter reading, demand side management and the overall ambition for a highly reliable and manageable electricity production and grid operation.

The key IoT devices in smart grid are smart meters. These are numerous, unattended devices, installed typically in every household. Their communication capabilities are limited, since they usually rely on power-line communications or long-range low-power wireless technologies. The smart meters can be upgraded to smart gateways, to support the control of demand (e.g. by switching off/on specific loads). There is a clear financial motivation for potential fraudulent actions in smart meters, with known examples of successful breaches.

In terms of IoT backend systems, smart grid and energy productions strongly relies on established Industrial control systems (ICS), including Supervisory control and data acquisition (SCADA). SCADA systems were not originally designed to be connected to the Internet, so their role in smart grid presents enormous security risk, with unfortunately many successful breaches.

A simplified example classification is presented in Table 1. Business goals of smart grid deployment are focused on the increasing area coverage/number of end-users and reduction of total operational cost. The top risks related to the mentioned goals considering information security are a low level of confidence in maturity and security of smart grid solutions, and no initial cost reduction through investment, data breaches and opportunity costs. These risks are then decomposed through the underlying structure into a subset of most relevant risks related for each particular view. As a risk case, insufficiently protected network is not designed for a high level security, can be dependent on existing grid infrastructure and leads to data breaches.

4.3 Use in the eHealth IoT Domain

Use of eHealth worldwide strongly influences medical services market in the last years. Rapidly growing market of telemedicine services—such as teleconsultations, telemonitoring, different forms of teleinterventions—sets up high level of security requirements for service implementation.

As stated in [13] networked medical devices are vulnerable to more than just criminal intent. Like any other technology, they are prone to failure. Should any high-profile failures take place, societies could easily turn their backs on networked medical devices, delaying their deployment for years or decades. A second immediate concern is protecting patient privacy and the sensitive health data inside these

Table 2: Security risks classification – an example of the eHealth IoT domain

View of risk evaluation	Outline	Risk description	Severity level
Business goals and key success factors	To reach the target population of patients and health personnel	Slowly increasing number of end-users	High
	Total cost reduction of health services	Initially no cost reduction through investment, data breaches and opportunity costs	Medium
IoT domain specifics	Medical safety and information security of services	Improper implementation, loss of confidentiality	Medium
	Depending on population size/age for particular diseases	Too low awareness of end-users about importance of information security	High
	Importance of best customer experience	Inability to manage the devices properly (especially by elderly)	High
IoT architecture & technology	Use of telemedicine platforms and personal sensor devices	Unsafe and unsecure technology selected, unsecure services implementation	Medium
	Use of smart phone applications		
Defenses and security products	Network protection equipment – threat management	Insufficiently protected network, in particular wireless	Low
	Cryptography on different levels (data, communications)	Too poor level of cryptography	High
	Analytic tools (SIEM)	Analytics not implemented or insufficient to detect security incidents	High

devices.

A typical example eHealth service architecture is composed of telemedicine platform, smart phone (or other mobile device with data hub) application and set of end-user medical sensor devices, such as ECG, glucometers, blood pressure meters etc. Mobile and home-based devices connect via the Internet to clinicians to reduce hospitalization through early detection of critical medical conditions.

In the example presented in Table 2, top business goals of reaching the target population of patients and health personnel, and total cost reduction of health services can be affected by similar top risks from the business view, but quite different underlying subsets of risks, depicting the specifics of eHealth services deployment. Too low awareness of end-users about importance of information security for personal medical data and inability to manage the devices properly, especially by elderly, are typical risks origination from the IoT use and application. Unsafe and unsecure technology selected, with poor defenses on the lowest level, lead to improper implementation and loss of confidentiality.

5. CONCLUSION

The presented risk classification method, where we enhanced the architectural and product views with the view of business risks at the top of risk classification, has been developed. It provides an improved methodological tool starting the risk evaluation process with focus on business goals and key success factors and thus leading to the highest risks and consequentially information security costs. Its use was demonstrated by the practical examples in the IoT domains of energy production and distribution and eHealth.

For the future, we plan to continue our research by development of a full risk evaluation methodology and supporting tools based on big data analytics.

REFERENCES

[1] European Commission, Special Eurobarometer 423, "Cyber Security Report 2014", February, 2015, http://ec.europa.eu/public_opinion/archives/ebs/ebs_423_en.pdf

[2] KPMG, "Clarity on Cyber Security", *KPMGKPMG International Cooperative*, Switzerland, 2015, http://www.kpmg.com/CH/en/Library/Articles-Publications/Documents/Advisory/pub-20150526-clarity-on-cyber-security-en.pdf

[3] Cisco, "Cisco 2016 Annual Security Report", *Cisco*, USA, January, 2016.

[4] ISO, International standard ISO/IEC 27001:2013, "Information technology -- Security techniques -- Information security management systems – Requirements", *International Organization for Standardization,* Switzerland, 2013.

[5] ISO, ISO/IEC 27005:2011, "Information technology -- Security techniques -- Information security risk management", *International Organization for Standardization*, Switzerland, 2011.

[6] IEEE, "Security Risk Assessment Methodology Using IntelliGrid Environments", IEEE, USA, IEEE P1649 Draft ver 1, October, 2005.

[7] Cvitić, Ivan, Vujić, Miroslav, and Husnjak, Siniša, "Classification of Security Risks in the IoT Environment", Proceedings of the 26th DAAAM Symposium on Intelligent Manufacturing and Automation, B. Katalinic (Ed.), *DAAAM International*, ISBN 978-3-902734-07-5, ISSN 1726-9679, Austria, 2016, pp.0731-0740.

[8] Cisco, The Internet of Things: Reduce Security Risks with Automated Policies, White paper, *Cisco*, USA, 2016.

[9] Gartner, "2016 Gartner Magic Quadrant for Security Information and Event Management (SIEM)", *Gartner publications*, USA, 2016.

[10] McMillan, Rob. "Definition: Threat Intelligence". *Gartner Publications*, USA, May, 2013.

[11] Gartner, "Magic Quadrant for Enterprise Network Firewalls", *Gartner Publications*, USA, 2016.

[12] ITU-T, "Y.2060: Overview of the Internet of things," *ITU-T*, Y.2060, June, 2012.

[13] Healey, Jason, Pollard, Neal, and Woods, Beau, "The Healthcare Internet of Things - Rewards and Risks", *Atlantic Council of the United States*, USA, ISBN: 978-1-61977-981-5, 2016.

Marjeta Pučko Marjeta Pučko received the B.S., M.S. and Ph. D. degrees in computer science from the University of Ljubljana, Slovenia, in 1988, 1991, and 1995, respectively. In 1988 she joined Jožef Stefan Institute, Department of digital communications and networks in Ljubljana as a researcher. From 1998 to 2009 she was with IskraTEL, Telecommunications Systems, Ltd., initially as an expert for telco systems design and testing, and later held different management positions in research, development and business improvement. In 2010 she joined Vzajemna health mutual insurance as head of IT department, information security manager and CIO deputy. Currently, at private consultancy, lecturing and managing different research and applicative projects, her interests concern ICT, business intelligence and data, information security, eHealth, e-learning systems and process management.

Andrej Kos (SM'98) received the Ph.D. degree in electrical engineering from the University of Ljubljana, Ljubljana, Slovenia, in 2003. He is a full professor at the Faculty of Electrical Engineering, University of Ljubljana, Slovenia, head of the Laboratory for Telecommunications and the chair of University of Ljubljana Innovation Commission. He started working in the field of telecommunications in 1996. Since 1999 he has specialized in modeling and design of high-speed networks and services. Currently, at the center of his work are broadband systems and applications of the Internet of things. Prof. Kos was part of the team that set up the MakerLab.

Matevž Pustišek (VM'01) received the Ph.D. degree in electrical engineering from the University of Ljubljana, Ljubljana, Slovenia, in 2009. He is a senior lecturer at the Faculty of Electrical Engineering, University of Ljubljana, Slovenia. His research is focused on the Internet services and applications, including mobile, Web, and IoT. A special interest is oriented towards the IoT architectures and security aspects. Recently additional focus is set on use of block-chain technologies in the IoT. At present he is collaborating in the Ekosmart project (http://ekosmart.net/en/ekosmart-2/) on smart cities and communities.

Blockchain Support in IoT Platforms

Pustišek, Matevž; Štefanič Južnič, Leon; and Kos, Andrej

Abstract: ***Blockchain (BC) technologies have a potential to blend with the existing Internet of things (IoT) platforms. BC enabled IoT platforms can be offered as a service (BIoTaaS) to provide scalable and trusted new approaches in e.g. IoT device authentication and management, trading with IoT data or in providing reliable and trusted interfaces between Web and smart contracts. At the same time this can lead to a gradual decentralization of highly centralized traditional cloud platforms – a needed change in IoT that can be anticipated from the fog computation and communication architectures, too. Currently the two viable BC candidates for BIoTaaS are the Ethereum (ETH) and the Hyperledger Fabric (HLF). Very diverse applications of BIoTaaS are possible, so it is unlikely that one platform approach or architecture will be meeting all these needs. Two differentiators have the key impact on selection of BC technology for particular BIoTaaS: existence of need for instant and independent on-chain payments and where the dominant focus is set – on the IoT devices or on the business-to-business (B2B) applications. If the devices are central and payments are required, then ETH BC is the favorite. In case of B2B HLF might be a preferable option due to security features beyond trust, derived from the permissioned network model. Beside the existence of BC, other requirements have to be met for efficient BIoTaaS. We defined a set of such common requirements, which include Web/HTTP/REST and other acknowledged application programming interfaces (API) for entire IoT and BC service access, on-chain smart contracts, low transaction confirmation delays for instant payments and near real-time operation, and smart oracles for interfacing the off-chain "real-world" objects and systems.***

Index Terms: ***API, blockchain, Ethereum, Hyperledger Fabric, Internet of things, platform***

1. INTRODUCTION

RECENT advancements in the Internet of things have brought it to the level of productivity. The Internet of things (IoT) solutions are being applied in industry, smart grids, health, mobility, wellbeing and other application domains. The IoT ecosystems are characterized by big number of heterogeneous and often constrained IoT devices, emerging user requirements and complex use-cases, and omnipresent demanding business and security requirements [1]. Traditional IoT architectures [2] are highly centralized, with cloud platforms providing services for collection, storage, analysis and use of vast quantities of data, provided by the IoT devices. However, with 5G network- and fog computation and communication architectures the traditional centralized IoT model has started evolving towards a more decentralized one. Decentralization is the key principle of the blockchain (BC) protocols and networks, too. BC enables trusted exchange of transactions in a system without trusted centralized authorities. Providing a native cryptocurrencies, autonomous machine-to-machine transactions including micropayments or distributed applications, BCs seem to be a valuable addition to IoT, too.

The key objective of this paper is to investigate the role of the BC in the IoT cloud platforms. In Section 2 we present cloud-based IoT platforms, which are along with communication gateways and IoT devices, the key building part of the IoT systems. Cloud APIs are presented as means for integration and use of cloud IoT services. Possible impact and decentralization of fog computation and communication architectures for IoT is discussed. In Section 3 a brief presentation of the blockchain and the distributed application concepts is given. Two key BC technologies for the IoT are exposed - the Ethereum and the Hyperledger Fabric. In Section 4 we summarize some of the initial blendings of established IoT platforms and blockchain technologies. Finally, based on our findings we present a set of general requirements for an IoT platform with blockchain support for IoT devices and for other cloud-based systems.

2. CLOUD BASED IOT PLATFORMS

IoT appears to be a playground for the most advanced new business and technological developments. 5G systems, machine-to-machine communications, fog computing and alike are constantly reshaping the established

Manuscript received June 10, 2016; revised June 27, 2017; accepted June 28, 2017. The work was supported in part by the Ministry of Education, Science and Sport of Slovenia. Authors are with the Faculty of Electrical Engineering, University of Ljubljana, Ljubljana SI-1000, Slovenia (matevz.pustisek@fe.uni-lj.si; leon.juznic@ltfe.org; andrej.kos@fe.uni-lj.si).

architectures of the IoT systems. But in spite of these new impacts, the IoT is reaching the productivity level and is being successfully applied in various application domains. All the key components to create IoT solutions are available as proven commercial (industry grade) products and services.

In a very simplified form, an IoT system—Figure 1—is comprised of devices, communication gateways and networks, and cloud-based backend systems. Devices are numerous and heterogeneous and have at least basic computation and communication capabilities. They incorporate sensors and/or actuators to face the real world environment.

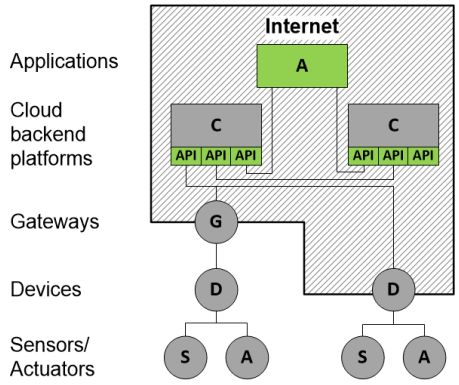


Figure 1: IoT architecture

Gateways add communication capabilities to connect devices to the Internet, where such a functionality is not already incorporated in the device, e.g. due to the lack of a complete network stack required for Internet connectivity, energy consumption constraints or limited computational capabilities. A gateway acts as a proxy, receiving data from devices and packaging it for transmission over Internet [3].

Backend systems/platforms collect, process, analyze and act on data generated by connected devices, enable long term storage and (big data) analysis, and facilitate easy development of the IoT applications that interact with the devices. The IoT platforms are frequently provided as a service (PaaS). Note: with term *IoT platform* in this paper we refer to backend cloud platforms and not to the IoT device platforms as for example single-board-computers (e.g. *RPi*) or microcontroller platforms (e.g. *Arduino*).

The leading IoT platforms are actually sets of products and services, which jointly provide the required functionalities. PaaS providers e.g. adapt their storage, big data or machine learning platforms, which are being used for non-IoT applications, as well. Apart from these more commonly or generally required functionalities, the IoT cloud platforms provide IoT specific services. These services shield applications from specific features of the IoT devices. They assure scalability (numerous IoT devices), security (access management, computation constraints) and management (registration, deployment, operation). IoT specific services also enable seamless interoperability among IoT specific parts and various other platforms that are combined in an IoT cloud backend.

There are numerous examples of IoT cloud platforms being in use. The largest share [4] of maker project is backed up with *MS Azure for IoT* [5] and *Amazon AWS for IoT* [6], closely followed by *Google Cloud IoT* [7]. These all are solutions from leading cloud service providers. They blend their existing cloud platforms with IoT specifics to provide a mash-up for IoT cloud services.

Google Cloud IoT is a rich set of various non-specific and IoT specific cloud platforms that are combined for IoT. Among the general services there are storage and databases, big data, machine learning and alike. The IoT specifics are arranged with *Cloud IoT Core* [8] which is a fully managed service that allows you to easily and securely connect, manage, and ingest data from numerous globally dispersed and heterogeneous devices.

Similar approach is taken in *Amazon AWS for IoT* [6]. IoT specific services are provided in the *AWS IoT Platform*, primarily focusing on secure and efficient communication between devices and other AWS. *AWS Greengrass* being more an IoT device- then a cloud-platform extends AWS to devices so they can act locally on the data they generate, while still using the cloud for management, analytics, and durable storage.

The *Microsoft Azure IoT Suite* [5] is an enterprise-grade cloud solution that enables you to get started quickly through a set of extensible preconfigured solutions. The services offer a broad range of capabilities. These enterprise grade services enable you to: collect data from devices, analyze data streams in-motion, store and query large data sets, visualize both real-time and historical data, integrate with back-office systems and manage your devices [9].

IBM Bluemix [10] is a cloud platform as a service (PaaS) developed by IBM. *Bluemix* is based on Cloud Foundry open technology and runs on SoftLayer infrastructure. It supports access to over 120 IBM cloud services (machine learning, storage, application services, blockchain, etc.). It also includes IBM's IoT Platform, which provides services for connecting and managing IoT devices, and analyzing the data they produce. It supports connecting to the cloud using open, lightweight MQTT messaging protocol or HTTP [11].

Some IoT cloud platforms operate at a smaller scale. They are not a part of an integral scope of

cloud services (IoT and non-IoT), but have been developed specifically for the IoT. *Thingspeak* [12] for example supports collection, storage, analysis and visualization of IoT data. For the collection, dominant IoT device platforms (*RPi*, *Arduino*, *and BeagleBone*) are directly supported. Analysis and visualization are made with *Matlab*. Therefore *Thingspeak* gained a lot of interest in academic and research communities. The *Open Source Elastic Stack* [13] is specialized in real time data analysis and visualization. Their key products - *Elasticsearch* and *Kibana* - enable you to reliably and securely collect data from any source, in any format, and search, analyze, and visualize it in real time. These two systems are available as integral modules also in the *AWS for IoT*.

Opensensors [14] is oriented towards acquisition and interchange of open IoT data and provides interfaces to efficiently exchange data among various IoT backend platforms.

2.1. Server Application Programming Interfaces

IoT cloud solutions are distinguished by various APIs for interoperability of their building components, for communication with the devices and for the applications based on these cloud solutions. IoT cloud platform APIs reflect the specifics and constraints of IoT: numerous and heterogeneous devices, low power devices, limited communication and computation capabilities. The key is, of course:

- HTTP/Web and real time APIs: to reliably and securely interact with cloud applications and other devices.

Apart from support for data and message exchange, IoT APIs may include features for:

- Virtual representations of devices: device shadows – implement an always available REST API for offline operation. Even if the actual device is temporarily offline, applications retain possibility to communicate with the device.
- Device management: registration, provisioning, deployment, updates and operation of devices at scale.
- Security and access management: for authentication and authorization of devices and platform users in form of API keys, JSON Web Tokens (JWS).
- Rule engines: gather, process, analyze and act on data from the connected devices and route the messages to other PaaS or their components.

In terms of implementation cloud APIs may rely on WebSocets, HTTPS REST, general-purpose RPC (gRPC), server-sent events (SSE) and others. The variety of implementation options reflects different needs of application developers, as well as different characteristics of messages and data streams passed over APIs.

IoT cloud platform providers frequently publish client libraries for various IoT device platforms and programming languages. These libraries facilitate the use of their cloud APIs and make application development easier.

2.2. Fog Architecture

Lately another architecture related to the IoT has been widely discussed. This is the fog computation and communication [15]. It reflects changes which are anticipated in mobile edge networks as envisaged in future 5G and partially outlined in current Evolved packet system (EPS) with LTE-A. The fog doesn't exclude cloud services and systems. It merely redistributes the location of computation, storage and control to decentralized elements in the architecture. In a unified end-to-end fog-cloud platform, cloud services continue to have an indispensable role. But the IoT system architecture is no longer limited to a device (full of constraints), transparent (dumb) communication networks and the smart cloud. Integrated fog nodes combine computation and communication.

The reasons for decentralization towards the fog are multiple: security, having applications closer to the end user, agility in application development (changing client application without a need to have the change implemented in cloud backend first) and efficiency. But the primary benefit of the fog computing is its ability to reduce latency and delay. There are additional features required in the fog-cloud systems: new service discovery, request and delivery mechanisms; different data management, taking into account local processing and storage; and service orchestration. In fog not only vertical interactions between the users/devices, edge nodes and cloud are foreseen. There are interactions among instances at same level, too [16].

Decentralized and distributed architecture of the fog computing and networking has, therefore, several similarities with decentralized blockchains, discussed in Section 3.

3. BLOCKCHAIN PROTOCOLS AND NETWORKS

Blockchains and distributed ledgers are listed among top strategic technology trends in 2017 [17]. They provide a decentralized framework for trusted transactions. The blockchain technologies are well known fundament of cryptocurrencies, but offer many other possible applications areas, too. In terms of IoT, two not mutually exclusive roles of a blockchain can be pointed out:

- As a distributed, scalable and trusted database, where the act of inserting/reading a parameter value is called a transaction, which is verified by a distributed community. The blockchain technology does not (necessarily) provide privacy of this data.
- Decentralized application environment for distributed deployment of applications.

Various specifications and implementations of blockchain technologies are available, but in our opinion at the moment two have relevant prospects for IoT. These are the Ethereum (ETH) [18] and the Hyperledger Fabric (HLF) [19]. Although the Bitcoin [20] is probably the most prominent BC technology, which gained reputation mostly due to the popular cryptocurrency Bitcoin [21], its potential role in IoT is extremely limited and is not a viable candidate for an IoT BC solution. Bitcoin protocol is namely lacking the distributed on-chain smart applications. Its role is thus limited more or less just to supporting a cryptocurrency.

In terms of application development for BC two approaches can be combined—off-chain and on-chain—as seen in Figure 2:

- Off-chain applications are Web, mobile and other applications, which use BC via client APIs that are exposed by BC clients. The BC client is responsible for the entire communication with BC and the application part for business logic (relaying on BC operation).
- On-chain business logic refers to smart contracts (i.e. chaincode in HLF), which are programming code written in Solidity – ETH - or in Go (or Java) - HLF, compiled and deployed in the BC network. Executions of smart contract are validated in the BC. BC thus provides a decentralized and trusted virtual machine for smart contract executions.

In Table 1 we compare the three blockchain protocols from the IoT perspective. Although different in their approaches (programming languages, etc), ETH and HLF both enable on-chain applications. The two protocols differ importantly in their consensus algorithms, too. For ETH public network the proof of work (PoW) is used currently – as it is in Bitcoin. In HLF the Practical Byzantine fault tolerance requires permissioned validating nodes (See Section 3.1. for details).

3.1. Ethereum

The Ethereum protocol [18] and corresponding networks are the basis for trusted, decentralized applications – *Dapps*. Apart from enabling a relevant cryptocurrency – ether [22] – Ethereum protocol is distinguished by a highly generalized programming language. With it one can code a smart contract which is deployed to the ETH network. It forms a contract account which is controlled by its contract code. The code is executed every time such an account receives a message/transaction from another account. This is the fundament for various IoT related blockchain applications.

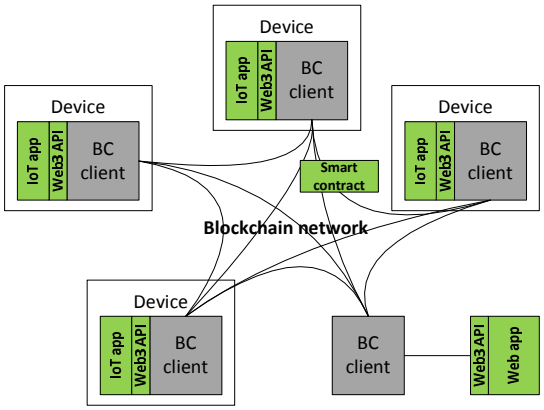


Figure 2: Ethereum BC architecture

3.1.1. Ethereum Client API

The entire functionality of ETH blockchain is available through the web3.js (and/or RPC/JSON) API [23] which geth – the ETH client - is exposing. Geth [24] is responsible for running ETH protocols and thus the entire communication with the blockchain. There are other ETH client implementations available, but geth usually serves as the reference, because it is being developed by Ethereum Foundation developers. Through this client API the entire functionality of ETH node can be exploited, including management of blocks and transactions, management of peers and network, monitoring of chain status, managing ETH accounts or mining ETH blocks. Application development in case of ETH and geth thus relies on using this client API, as depicted in Figure 2. Applications and geth can run on the same device or on two separate devices. In the latter case, HTTP protocol can be used in application to reach the distant geth node.

3.1.2. Smart Contract API

Smart contracts are on-chain business logic that is executed within the blockchain network. The execution can be verified by any network participant and thus trusted in the same way as any other transaction in BC network is.

A smart contract exposes functions, which are used by blockchain account. These functions represent a kind of an on-chain API for other BC accounts, and are accessible via blockchain.

Table 1: Comparison of Blockchain networks

	Bitcoin	Ethereum	Hyperledger Fabric
Native cryptocurrency	Yes	Yes	No
Distributed applications	No (very limited)	Yes – smart contracts	Yes - chaincode
Smart contracts	-	Solidity	Go, Java - (executed in containers)
Consensus algorithm	Proof of Work (PoW)	Proof of Work (PoW) Proof of stake (PoS) foreseen	PBFT - Practical Byzantine fault tolerance
Anonymous accounts	Yes	Yes	No (permissioned network)
Network	Public	Public or permissioned	Permissioned
Suitable for IoT	No	Yes	Yes
State channels	Yes (Lightning)	Yes (Raiden)	Not required

3.1.3. Decentralized Data Feeds with Smart Oracles

Smart contracts in ETH environment operate in a rather isolated space. The Solidity [25] smart contract programming language has e.g. no means to request data from URLs and thus to interface with "real" Internet world – e.g. Web sites or IoT devices - because these external information cannot be trustworthily verified by the contract. This shortcoming can be outdone by oracles [26]. These serve as intermediaries, providing data feeds along with an authenticity proof to the blockchain form/to external software (e.g. Web sites) or hardware entities.

3.1.4. State Channels

Another challenge in BC networks with PoW consensus is the transaction validation period. Due to the nature of the blockchain, this can take from several tens of seconds (in ETH about 20 s) up to several minutes (in Bitcoin 10 min) or more. Besides, distributed application developers do not have an influence on these times. In fact, the delays may become even longer due to higher transaction rates in the network or if several consecutive block verifications are required for security reasons. This limits, at least to some extent, the feasibility of scalable instant payments and is not suitable for near real-time IoT applications (e.g. door lock controlled by BC). A solution to this is being sought in state channels. This architecture combines off- and on-chain transactions to contribute to additional scalability, privacy and reduction of confirmation delays, compared to the current BC architecture. In ETH this approach is manifested in the Raiden [27] and in Bitcoin in the Lightning network [28].

3.2. Hyperledger Fabric

The Hyperledger Project [11] is a collaborative effort to create an enterprise-grade, open-source distributed ledger framework and code base. Established as a project of the Linux Foundation in early 2016, the Hyperledger Project currently has more than 130 members, including leaders in finance, banking, in the internet of things, supply chain, manufacturing and technology.

The Hyperledger Fabric [19], one of multiple projects currently in incubation under the Hyperledger Project, is a permissioned blockchain platform aimed at business use. It is open-source and based on standards, runs arbitrary smart contracts (called chaincode), supports strong security, identity features, basic REST APIs, CLIs and uses a modular architecture with pluggable consensus protocols (currently an implementation of Byzantine fault-tolerant consensus using the PBFT protocol [29] is supported).

The distributed ledger protocol of the fabric is run by peers. The fabric distinguishes between two kinds of peers: (i) *validating peer* is a node on the network responsible for running consensus, validating transactions, and maintaining the ledger and (ii) a *non-validating peer* which is a node that functions as a proxy to connect to validating peers [30].

4. IoT Platforms with BC Support

The IoT is facing several challenges that need to be addressed to continue with its successful practical deployments: centralized ecosystem, the cost of the connectivity, disrupted business models, security and trust and lack of functional value. A decentralized approach to IoT networking would solve many of the issues above. Blockchain technology is the missing link to cope with some of the future challenges in the IoT [31]. BC can:

- reduce costs - track billions of connected devices, enabling the processing of transactions and coordination between devices, managing updates [32],
- build trust - cryptographic algorithms used by blockchains would make consumer data more private, man-in-the-middle attacks

Table 2: Comparison of architectures: Cloud based Internet of Things PaaS vs. Blockchain

	Cloud based Internet of Things	Blockchain
Topology	• Centralized • (decentralization only being introduced in fog)	• Decentralized, fully distributed (P2P like)
APIs	• HTTP/Web and real time server APIs	• API at every particular BC client • Smart contract functions in form of backend API
Device libraries	• To use server APIs	• To use client API
Security focus	• Authentication and authorization of devices • Security and privacy of cloud services • Communication security • Availability	• Trust
Latency	• Low to moderate, near real time operation is possible • (fog architecture additionally reduces latency)	• High, due to the nature of transaction validation
(Micro)payments	• Not part of common IoT platforms	• Essential part of technology
Application logic in the platform	• In platform modules (big data, queries, etc). • Web applications accessing PaaS through APIs	• Smart contracts

cannot be staged, ledger cannot be manipulated,

- accelerate transactions - decentralized approach would eliminate single points of failure,
- keep an immutable record of the history of smart devices - no need for a centralized authority, and
- provide machine-to-machine transactions and micropayments.

Beside this, BC can easily facilitate:

- decentralized data feeds (Schelling coin [33]), where a vast amount of numerous concurrent (low fidelity) measurements from IoT devices is summarized into e.g. the most possible value of the temperature.

In spite of all its benefits, the blockchain model is not without its flaws and shortcomings. This is not surprising, because blockchain technologies are being relatively new and not as mature as e.g. IoT technologies.

4.1. Existing Examples

Big IT companies are already exploring the opportunities of blockchain in IoT. They are usually integrating blockchain as a service (BaaS) in their existing IoT platforms. So BaaS is provided along with the existing IoT PaaS, which were discussed in Section 1.

IBM is the leader in open-source blockchain solutions built for the enterprise. Their blockchain ecosystem brings together a range of people and organizations interested in building and leveraging blockchain solutions. IBM Watson IoT™ platform [34] enables IoT devices to send data to private blockchain ledgers for inclusion in shared transactions with tamper-resistant records. Its BaaS service is based on Hyperledger Fabric [19], which is one of the five frameworks hosted with Hyperledger. IBM contributed more than a half of the code used in HLF. This demonstrates IBM's strong commitment to provide open governance for the development of blockchain.

HLF in IBM Watson IoT is predominantly suitable for private blockchains in enterprise settings, because it is using a different consensus algorithm than e.g. ETH. It is distinguished by a well-documented HTTPS REST API [35] for all blockchain related functions. Web developers can thus benefit from BC features, but continue to utilize API technologies they are already familiar with. The Watson IoT API enables management of blocks and transactions as well as peers and networks, monitoring of chain status, and registrations and management of BC users.

Microsoft entered a partnership to create Ethereum blockchain as a service (EBaaS) on Microsoft Azure [36]. The service will allow users to efficiently create private, public and consortium based Blockchain environments using industry leading frameworks. Surrounding capabilities like Cortana Analytics (machine learning), Power BI, Azure Active Directory, can be integrated into apps launching a new generation of decentralized cross platform applications.

SAP Leonardo [37] is a digital innovation system, which integrates IoT, machine learning,

analytics, big data and blockchain, and runs them seamlessly in the cloud. The blockchain element is based on the Hyperledger open source blockchain platform, using its standards and protocols. SAP joined Hyperledger as a premier member early in 2017.

4.2. Selection Criteria and Common Requirements for Iot Platform with BC Support

Numerous and very diverse applications which combine IoT and BC are possible, so it is unlikely that one blockchain IoT platform (BIoTaaS) or platform architecture will be meeting all these needs. Two differentiators have key impact on the selection of BC technology for particular BIoTaaS: (i) whether there is a need for instant and independent on-chain payments and (ii) where the dominant focus is set – on devices or on the business-to-business (B2B) applications. In developing applications based on BIoTaaS, BC can be integrated in the IoT devices as well as in the backend (cloud) part.

ETH and HLF are the two viable BC candidates for BIoTaaS. If the devices are central and payments are required, then ETH BC is the favorite. In case of ETH, nodes can benefit from reliable scalable public network, recognized cryptocurrency, and existing examples of client and application deployments in computers, mobile devices and embedded systems. In case of B2B, HLF might be a preferable option due to security features beyond trust, derived from the permissioned network model. Despite several different features, ETH and HLF share many common requirements. These requirements have to be met also in any alternative BC technologies considered for IoT:

- Web/HTTP/REST and other acknowledged APIs to access the entire set of IoT and BC services.
- On-chain smart contracts.
- Low transaction confirmation delays for instant payments and near real-time operation. This can be achieved either by the consensus algorithm or through the availability of state channels.
- Smart oracles for interfacing “real-world”, which can be an integrated function of the BIoTaaS.

Beside the technical and functional features, other strategic decisions may determine selection of technologies for BIoTaaS. With BC being relatively new technology many parts remain in early development stages. So maturity of available solutions, size and support of the involved development community and successful use cases should be considered in selection, too.

5. CONCLUSIONS

Rapid progress in BC and lower maturity compared to the existing cloud services require a thoughtful positioning of BC in BIoTaaS. Many of current BC developments focus on creating yet another alternative coin along with corresponding smart contracts, to support vaguely defined potential use cases. Their motivation is in prospects of a successful initial coin offering (ICO). Not that many initiatives bring BC to the real world, including IoT.

We anticipate an important role of IoT platforms with highly integrated BC support in e.g. IoT device authentication and management, trading with IoT data or in providing a reliable and trusted interface between Web and smart contracts.

Our future research is oriented towards the architectures and position of BIoTaaS in smart city ecosystem.

ACKNOWLEDGMENT

The authors wish to acknowledge the support of the research program “Algorithms and Optimization Procedures in Telecommunications”, financed by the Slovenian Research Agency.

REFERENCES

[1] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, “Internet of Things (IoT): A vision, architectural elements, and future directions,” Future Generation Computer Systems, vol. 29, no. 7, pp. 1645–1660, Sep. 2013.

[2] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, “Internet of Things: A Survey on Enabling Technologies, Protocols, and Applications,” IEEE Communications Surveys Tutorials, vol. 17, no. 4, pp. 2347–2376, Fourthquarter 2015.

[3] “Overview of Internet of Things,” Google Cloud Platform, 19-Apr-2017. [Online]. Available: https://cloud.google.com/solutions/iot-overview. [Accessed: 05-Jun-2017].

[4] S. Palmer, “10 Best Internet of Things (IoT) Cloud Platforms – DevTeamSpace Blog,” 01-Mar-2017. [Online]. Available: https://www.devteam.space/blog/10-best-internet-of-things-iot-cloud-platforms/. [Accessed: 05-Jun-2017].

[5] “Azure IoT Suite | Microsoft Azure.” [Online]. Available: https://azure.microsoft.com/en-us/suites/iot-suite/. [Accessed: 19-Jun-2017].

[6] “AWS IoT,” Amazon Web Services. [Online]. Available: //aws.amazon.com/iot/. [Accessed: 20-Jun-2017].

[7] “Google Cloud IoT - Fully managed IoT services from Google,” Google Cloud Platform. [Online]. Available: https://cloud.google.com/solutions/iot/. [Accessed: 20-Jun-2017].

[8] “Cloud IoT Core.” [Online]. Available: https://cloud.google.com/iot-core/. [Accessed: 20-Jun-2017].

[9] D. Betts, “Microsoft Azure IoT Suite overview.” [Online]. Available: https://docs.microsoft.com/en-us/azure/iot-suite/iot-suite-overview. [Accessed: 19-Jun-2017].

[10] “Cloud Infrastructure, Storage, Security &, More” IBM Bluemix. [Online]. Available: https://www.ibm.com/cloud-computing/bluemix/. [Accessed: 19-Jun-2017].

[11] “Hyperledger – Blockchain Technologies for Business.” [Online]. Available: https://www.hyperledger.org/. [Accessed: 19-Jun-2017].
[12] “Learn More about ThingSpeak.” [Online]. Available: https://thingspeak.com/pages/learn_more. [Accessed: 20-Jun-2017].
[13] “The Open Source Elastic Stack.” [Online]. Available: https://www.elastic.co/products. [Accessed: 20-Jun-2017].
[14] “OpenSensors.” [Online]. Available: https://www.opensensors.io/. [Accessed: 20-Jun-2017].
[15] F. Bonomi, R. Milito, P. Natarajan, and J. Zhu, “Fog Computing: A Platform for Internet of Things and Analytics,” in Big Data and Internet of Things: A Roadmap for Smart Environments, N. Bessis and C. Dobre, Eds. Springer International Publishing, 2014, pp. 169–186.
[16] M. Chiang, S. Ha, C. L. I, F. Risso, and T. Zhang, “Clarifying Fog Computing and Networking: 10 Questions and Answers,” IEEE Communications Magazine, vol. 55, no. 4, pp. 18–20, Apr. 2017.
[17] Kasey Panetta, “Gartner’s Top 10 Strategic Technology Trends for 2017 - Smarter with Gartner,” 18-Oct-2016. [Online]. Available: http://www.gartner.com/smarterwithgartner/gartners-top-10-technology-trends-2017/. [Accessed: 05-May-2017].
[18] V. Trón, “Ethereum Specification,” 23-Jul-2015. [Online]. Available: https://github.com/ethereum/go-ethereum/wiki/Ethereum-Specification. [Accessed: 05-May-2017].
[19] “IBM Blockchain - The Hyperledger Project,” 16-Jun-2017. [Online]. Available: https://www.ibm.com/blockchain/hyperledger.html. [Accessed: 19-Jun-2017].
[20] “Protocol documentation - Bitcoin Wiki.” [Online]. Available: https://en.bitcoin.it/wiki/Protocol_documentation. [Accessed: 02-Aug-2016].
[21] “Bitcoin - Open source P2P money.” [Online]. Available: https://bitcoin.org/en/. [Accessed: 02-Aug-2016].
[22] “ETH/USDT Market - Poloniex Bitcoin/Cryptocurrency Exchange.” [Online]. Available: https://poloniex.com/exchange#usdt_eth. [Accessed: 05-May-2017].
[23] “JavaScript API,” ethereum/wiki Wiki · GitHub. [Online]. Available: https://github.com/ethereum/wiki/wiki/JavaScript-API. [Accessed: 05-May-2017].
[24] V. Trón, “Geth,” ethereum/go-ethereum Wiki · GitHub. [Online]. Available: https://github.com/ethereum/go-ethereum/wiki/geth. [Accessed: 08-May-2017].
[25] “Solidity — Solidity 0.2.0 documentation.” [Online]. Available: http://solidity.readthedocs.io/en/latest/index.html. [Accessed: 08-May-2017].
[26] “Oraclize Documentation,” Overview. [Online]. Available: http://docs.oraclize.it/#overview. [Accessed: 05-May-2017].
[27] “Raiden Network,” High speed asset transfers for Ethereum, 20-Dec-2016. [Online]. Available: http://raiden.network/. [Accessed: 05-May-2017].
[28] “Lightning Network,” Scalable, Instant Bitcoin/Blockchain Transactions. [Online]. Available: http://lightning.network/. [Accessed: 05-May-2017].
[29] “Byzantine fault tolerance,” Wikipedia. 10-Jun-2017.
[30] C. Cachin, “Architecture of the Hyperledger Blockchain Fabric,” in *Workshop on Distributed Cryptocurrencies and Consensus Ledgers*, Chicago, Illinois, USA, 2016.
[31] “IoT and Blockchain Convergence: Benefits and Challenges - IEEE Internet of Things.” [Online]. Available: http://iot.ieee.org/newsletter/january-2017/iot-and-blockchain-convergence-benefits-and-challenges.html. [Accessed: 19-Jun-2017].
[32] K. Christidis and M. Devetsikiotis, “Blockchains and Smart Contracts for the Internet of Things,” IEEE Access, vol. 4, pp. 2292–2303, 2016.
[33] V. Buterin, “SchellingCoin: A Minimal-Trust Universal Data Feed,” Ethereum Blog, 28-Mar-2014. .
[34] “IBM Watson Internet of Things (IoT),” 09-Jun-2017. [Online]. Available: https://www.ibm.com/internet-of-things/. [Accessed: 19-Jun-2017].
[35] “Core API - Hyperledger Fabric.” [Online]. Available: http://hyperledger-fabric.readthedocs.io/en/stable/API/CoreAPI/#rest-api. [Accessed: 20-Jun-2017].
[36] “Blockchain as a Service (BaaS),” Microsoft Azure. [Online]. Available: https://azure.microsoft.com/en-us/solutions/blockchain/. [Accessed: 19-Jun-2017].
[37] “SAP Leonardo | Digital Innovation System,” SAP. [Online]. Available: https://www.sap.com/products/leonardo.html. [Accessed: 19-Jun-2017].

Electric Switch with Ethereum Blockchain Support

Pustišek, Matevž; Bremond, Nicolas; and Kos, Andrej

Abstract: ***This paper presents a prototype implementation of a USB charging device with Ethereum blockchain support for booking and payments. It outlines the design of the system and selection of hardware and software components for the end-to-end solution. Blockchain technologies are relatively new and promising development for decentralized trusted transactions, including micropayments. There are many possible application domains of blockchain technology envisaged; one of them is smart grid. Several initiatives are prototyping solutions in this domain; however, there is still a vast unexplored area related to technological and business aspects of blockchain in energetics. Our prototype provides an end-to-end solution that is comprised of a blockchain enabled end-device, Web applications with blockchain clients for users and administrators, and corresponding smart contracts for the specific transaction logics. The prototype was implemented as a DIN rail compatible device based on RPi and a user application as Javascript code embedded in HTML page that can be opened in Chrome browser with Metamask plugin installed. The system was successfully implemented and tested and it confirmed the viability of the concept. With modest modifications this approach could be extended to electric vehicle chargers, smart grid supply and demand management or other utility support systems.***

Index Terms: ***blockchain, charger, dapp, Ethereum, prototype, smart grid, switch.***

1. *INTRODUCTION*

TRUSTED decentralized applications based on blockchain (BC) technologies are raising immense interest in technical and business communities. However, the actual maturity of BC in terms of technology, business models and applications does not yet match this reputation. According to Gartner's hype cycle for emerging technologies in 2016 [1] blockchain is placed close to the peak of inflated expectations. A way to progress towards productive solutions is developing working prototypes, which apply combinations of technologies to demonstrate viable use-cases that further help reinforcing the role of novel technologies in an even broader scope of ICT applications. Besides, there are only scarce studies with solid scientific backgrounds that research application possibilities of blockchain technologies. This is also one of the motivating factors behind our work.

Energy and smart grid seems a natural application domain for blockchain technologies, so there are already some proof-of-concept developments being presented in this domain. Slock.it is a German company [2] specialized in blockchain applications with real-world objects. One of the potential use cases that they are addressing is energy and smart grid. The Share&Charge [3] is their PoC project in this vertical. Brainbot technologies AG [4] develops and investigates the role of Raiden (see Section 2 for details). It demonstrated in the Raiden Network IoT Demo [5] how an electric switch could be controlled via Raiden as well.

We designed, developed, implemented and tested a smart electric switch with charger, which is controlled through the Ethereum blockchain network [6]. It is comprised of end-user and administrator applications, a smart contract, as well as a BC aware IoT device. It is thus a prototype of an end-to-end IoT solution, based on blockchains — unlike many recent blockchain developments, which frequently focus on creating yet another alternative BC coin along with corresponding smart contracts, to support vaguely defined potential use cases. Not that many initiatives bring BC to the real world—including IoT. The main motivation for this work was gaining experience with blockchain technologies, and related application development and implementation requirements.

In Section 2 we briefly present the role of the Ethereum protocol, which was applied in our prototype. In Section 3 we elaborate the system design and in Section 4 we present the implementation and results. We conclude the paper with a brief outline of possible extensions of the current system.

Manuscript received June 1, 2017.

Author is with the Faculty of Electrical Engineering, University of Ljubljana, Slovenia (e-mail: matevz.pustisek@fe.uni-lj.si).

2. BLOCKCHAINS FOR IOT

Blockchains and distributed ledgers are listed among top strategic technology trends in 2017 [7]. They provide a decentralized framework for trusted transactions. The blockchain technologies are well known fundament of cryptocurrencies (e.g. Bitcoin), but offer many other possible applications areas, too.

The Ethereum protocol [6] and corresponding networks are the basis for trusted, decentralized applications – *Dapps*. Apart from enabling a relevant cryptocurrency – ether [8] – Ethereum protocol is distinguished by a highly generalized programming language. With it one can code a smart contract which is deployed to the Ethereum network. It forms a contract account which is controlled by its contract code. The code is executed every time such an account receives a message/transaction from another account. Executions of smart contracts are validated in the blockchain network. This is a fundament for various IoT related blockchain applications. Particular benefits from BC, relevant to IoT, include machine-to-machine transactions, micropayments, decentralized data feeds, and scalability. Ethereum protocol is deployed in various public Ethereum networks[9]. The two key ones are the *mainnet* and *testnet* (current version is called Ropsten Revival) and both apply the same Ethereum protocols. In *mainnet* the cryptocurrency (i.e. ether) has a real value and can be traded for fiat currencies. The ether in *testnet* has no real value, and the network is meant for testing purposes.

Many established IoT platforms announced or have already implemented support for BC [10], [11]. Different alliances are being built here and different BC protocols can be found in this role. But what they mostly have in common is an API that facilitates use of full scope BC functionality from the Web application development tools. However, while exploring integration of BC into actual IoT architectures, several shortcomings started appearing with implementation of first prototypes. Smart contracts in Ethereum environment operate in a rather isolated space. The Solidity [12] smart contract programming language has e.g. no means to request data form URLs and thus to interface with "real" Internet world — e.g. Web sites and IoT devices — because these external information cannot be trustworthily verified by the contract. This shortcoming can be outdone by oracles [13]. They serve as intermediaries, providing data feeds along with an authenticity proof to the blockchain form/to external software (e.g. Web sites) or hardware entities. Another challenge is transaction validation period. Due to the blockchain nature, this can take from several tens of seconds (in Ethereum about 20s) up to several minutes (in Bitcoin 10min or more). Besides, distributed application developers do not have an influence on these times, which in addition may become even longer due to higher transaction rates in the network. This limits, at least to some extent, the feasibility of scalable instant payments. A solution to this problem is being sought in state channels. This architecture combines off- and on-network transactions to contribute to additional scalability, privacy and reduction of confirmation delays, compared to the current BC architecture. In Ethereum this approach is manifested in the Raiden [14] and in Bitcoin the Lightning network [15].

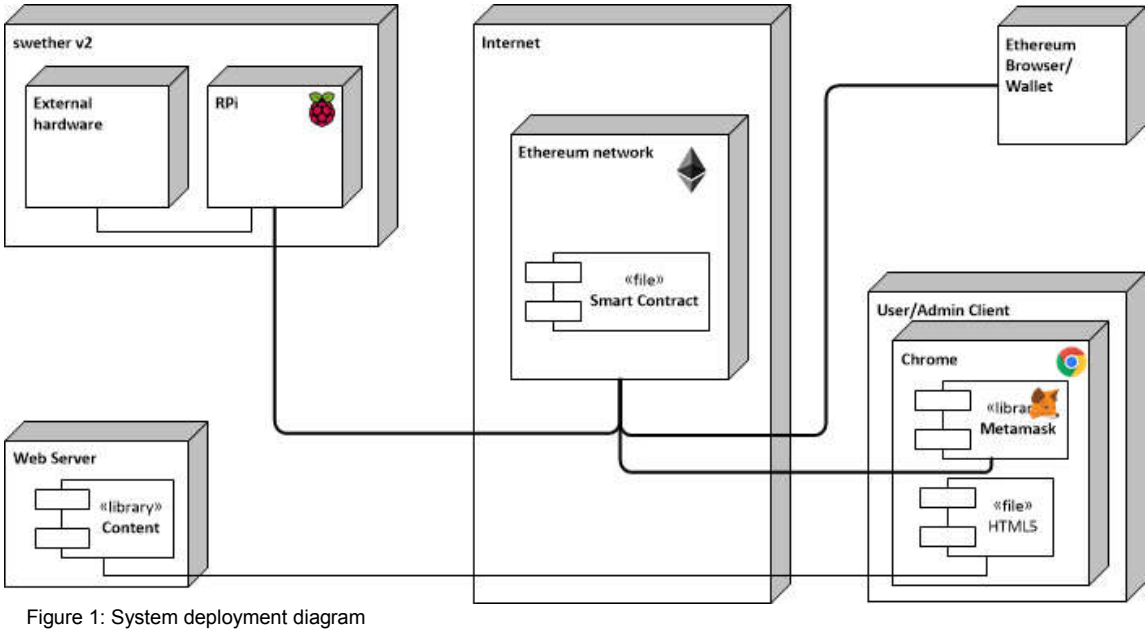


Figure 1: System deployment diagram

3. PROTOTYPE SYSTEM DESIGN

To demonstrate feasibility and to address the full scope of development activities for blockchain, we decided for a system that is comprised of a blockchain enabled IoT device (called Swether – switch with Ethereum support), as well as of customized Web applications for users and administrators to interact with the device via blockchain. The same principle can be easily extended e.g. to 230 V AC chargers or used to control access (and not to charge for it) to devices and services with control of their power supply.

3.1 DApp Architecture and System Model

Overall system architecture is depicted in Figure 1. Swether — switch with Ethereum support — is an IoT device controlled via a DApp and blockchain network. A smart contract implementing the system backend is deployed to the BC network during the system setup by the system owner. Users and administrator access the functionalities implemented in Web applications through Ethereum Web browsers. In our case, the default browsers are Chrome with Metamask plugin and Mist. Web server merely provides the access to user and administrator Web pages. All the actors in the system presented in Section 3.2 need valid Ethereum accounts and minimal sum of ether to compensate for charging costs and transaction fees. The system is currently deployed in Ethereum *testnet* (Ropsten), but can be in the same way deployed in *mainnet*, too.

3.2 System Outline – Actors and Activities

There are four actors foreseen in the system: device user, device administrator/owner, smart Web application, and the Swether device. A **user** can review current availability of free charging slots and book a desired quantity of time for charging. The booked time is instantly paid with Ether. Swether device monitors the Ethereum network and toggles the charging slots according to transaction status. An **administrator** can set the basic parameters of Swether, like the maximum number of charging slots in a device and current price per time unit. Administrator can transfer funds from smart contract address to an arbitrary Ethereum address.

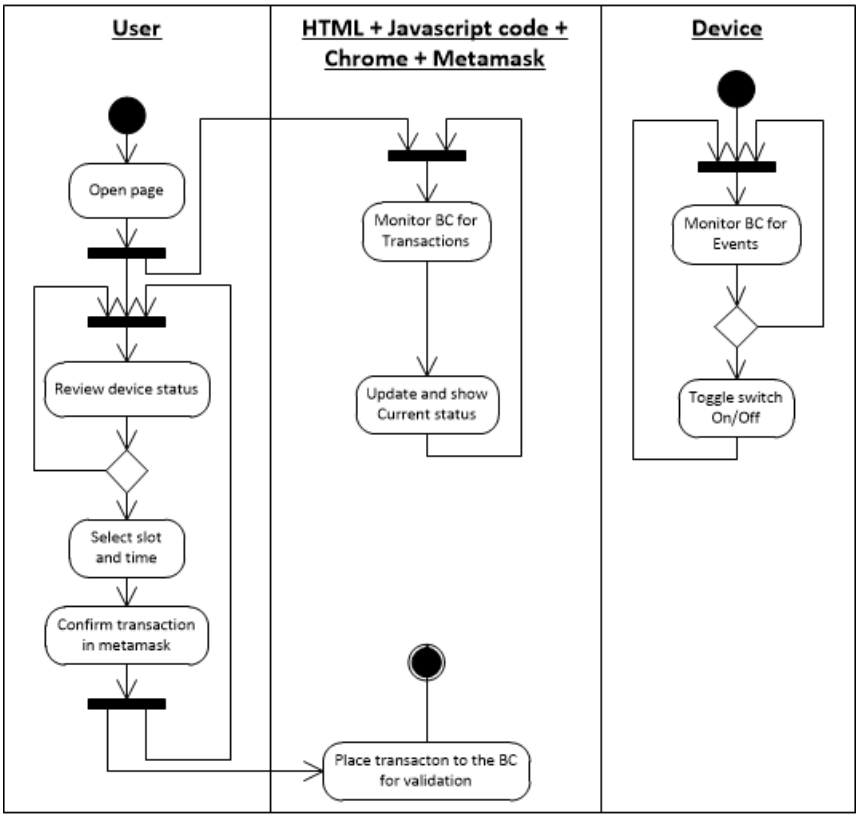


Figure 2: User activity diagram

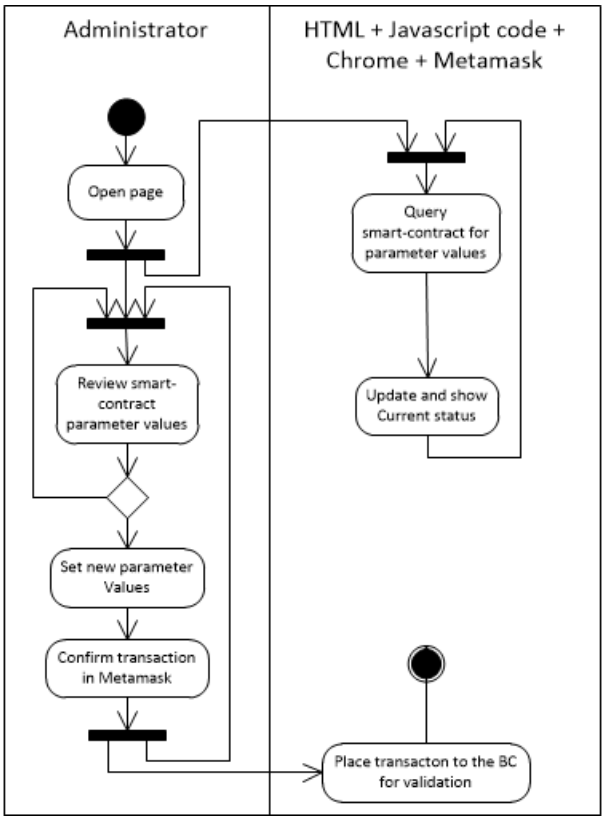


Figure 3: Administrator activity diagram

As depicted in Figure 2, two activities run in parallel. One is related to the operation of the device, which constantly monitors the blockchain for relevant events. An event is created when the transaction for a charging request is validated by the network. If a corresponding event is identified, the device toggles on the selected charging slot based on the parameter values in the transaction. When the charging period terminates, it automatically toggles off the slot.

The second activity involves user and his smart application. User downloads the HTML file and the Javascript logic for communication with blockchain. The page has to be opened in an Ethereum compliant Web browser. At the time of writing we relied on Google Chrome with Metamask plugin [16] as the Ethereum client and Mist. The Javascript application retrieves expected device and charging slot status from the smart contract in blockchain and visualizes it in the page. User can thus review the status and proceeds to the reservation of one of the slots. He selects the desired booking duration or the price in Ether he is willing pay for charging. The application creates the appropriate transaction. Final confirmation is left to the user and then the transaction is placed to the chain for validation. Once validated, the Swether device intercepts corresponding event and reacts on it. There is thus no direct communication between the user application and the device. The only interconnection is via a transaction validated in blockchain.

Administrator on the other hand, has more limited scope of activities. Once the system has been set up, including the deployment of the smart contract to the blockchain (this step is not a part of the runtime and is therefore not depicted in Figure 3), administrator reviews current contract parameter values (e.g. number of charging slots, energy price per minute) and sets new values. The modified values are placed in the smart contract via a BC transaction, too.

Retrieval of funds received by the smart contract is not depicted either. This is an activity which is inherent to smart contract accounts in Ethereum. It enables administrator to collect the revenue, by transferring the funds to another Ethereum account.

4. System Implementation and Results

4.1 Swether

Swether device deployment is depicted in Figure 4. The core of the device is a Raspberry Pi 3 Model B [17], a popular miniature computer and embedded platform, rich in communication capabilities (Ethernet, Wifi, BLE, USB and HDMI). It is able to run various operating systems; among others the Linux flavored

Raspbian [18]. As an embedded device it has 40 GPIO pins to interface external hardware. In our case these included a 5V 4-channel power relay module with optocouplers and LED for status indication.

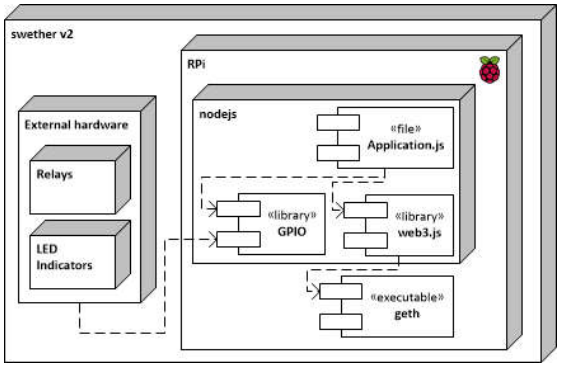


Figure 4: Deployment diagram of Swether v2

The application logic in Swether is implemented in Javascript. This is mostly due to the well documented web3.js API of the *geth* [19] Ethereum client. The *geth* is responsible for running Ethereum protocols and thus the entire communication with the blockchain. Through the web3.js API application can monitor status of the client and monitor or create transaction. As the application requirements for external hardware interfacing via GPIO were not complex, we implemented this part of the application in Javascript, too. In future version of the device we plan to implement hardware related part in Python, due to the better support of additional HW items foreseen in the device and richer programming capabilities.

Figure 5: Swether — the smart charging device

Once configured and deployed, the device requires no user interventions. There are just three LED indicators in the housing to indicate the status of the Ethereum network/client (Figure 5).

Swether device is not represented in the Ethereum network as an active entity and is thus not identified by its own Ethereum address. The *geth* client in the device constantly monitors the transactions in the network can catches the transaction events that were generated by the corresponding smart contract. The appropriate smart contract address is defined at the time of the device deployment.

4.2 End-user Web Application with Ethereum Support

End-user interface and application was implemented as HTML 5 page that has to run in an Ethereum compliant Web browser. This can be done in dedicated Ethereum wallets/browsers like Mist [20]. However, the application of Metamask plugin and Google Chrome browser assures a more transparent user experience. In this case user applies the same Web browser he is already accustomed to. The Metamask [21] is a light Ethereum client exposing the JSON-RPC and standard Ethereum web3 APIs [22]. The web3 library communicates with the Metamask client through JSON-RPC. A HTML page with application logic for Metamask has to be placed and retrieved from a HTTP server (and not from a local file) due to browser security restrictions.

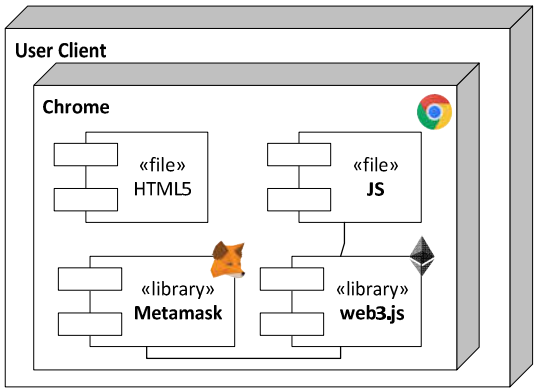


Figure 6: Deployment diagram of the end-user Web application

Upon the launch, the application verifies the Metamask client and Ethereum network status. If the status is OK, user proceeds to the overview of Swether device status, as seen in Figure 7. User can then select one of the unoccupied slots and the desired duration of the reservation or the total price she is willing to pay. The total price or time, which is indicated, is calculated from the current per minute rate that was retrieved from the smart contract. By pressing the “Book” button application code defines the transaction content. It passes it via web3 API to Metamask, which builds the transaction and sends it to the Ethereum network. The transaction is addressed

to the smart contract, where it triggers the function *bookAPlug* with parameters required for a booking. Within the transaction/execution of this function, an event is created. It serves later as an indication to the smart device about a relevant state-changing transaction. Once the transaction is validated by the blockchain, the act of booking is trustworthily recorded.

The key benefit of a custom Web user/admin application is twofold. First, it provides better user experience and customized application appearence. All the functionalities of the smart contract can be used from a general Ethereum compliant wallet, too. But such a use resembles very basic parameter settings. The Web based user interface in our solution can profit from all features of modern Web user interfaces design. Second, the Metamask is easy to be installed and used. It relies on remote geth nodes for better responsiveness (i.e. light-client mode). This again reduces the burden from user and facilitates promp start for using the system.

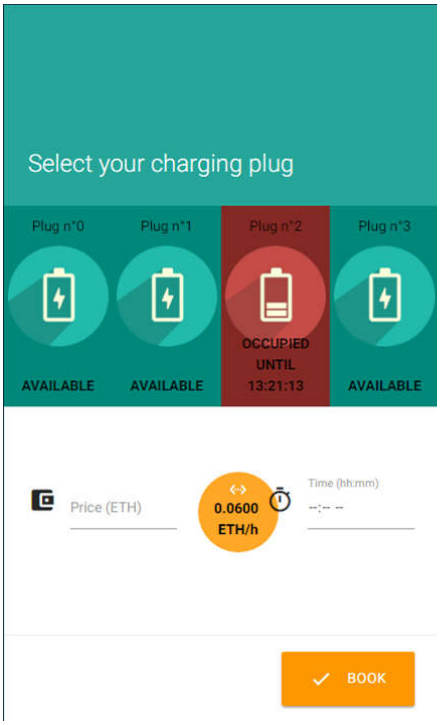


Figure 7: User-interface in Chrome browser

4.3 Administrator Web Application with Ethereum Support

Administrator interface and application was implemented in the same way as the user-interface. The difference is in functionality. On the page the administrator can monitor and change the parameters, which determine device setup and operation. They are listed and described in Section 4.4.

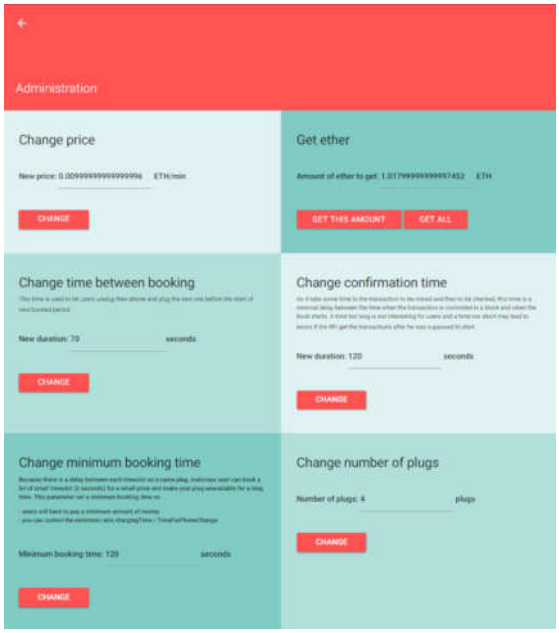


Figure 8: Administrator interface in Chrome browser

The page is publically accessible from a Web server. But only if the page user applies an Ethereum account in Metamask plug-in which is marked in the smart contract as an administrator account, he can set new values for the parameters. The page and the parameters are not hidden, because they can be derived from the Ethereum blockchain anytime. Granting the access to the functions changing the values is assured by a smart contract and not at the Web application level.

4.4 Application Backend — Smart Contract

System operation is supported by a dedicated smart contract, called *plugBooking*. It was implemented in Solidity and deployed to the Ethereum network from the administrator account. The *plugBooking* contract provides two sets of parameters and functions, one supporting the user application and the other the administrator Web interface.

Key parameters that determine the device setup and operation are:

- *plugNumber* – the number of charging plugs/ports that a device has, it must be set accordingly to physical implementation of the device;
- *minimumBookingTime* – minimal duration of plug reservation as a means to prevent malicious reservations with 0 duration and respective DOS of the Swether device;
- *price* – the current per second rate of charging in Wei[1];

[1] Ether (ETH) is the native cryptocurrency in the public Ethereum network and it can be traded for real currencies. $1Wei=10^{-18}$ ETH

- *userSwapTime* – a safety period between two consecutive chargings at the same plug, which allows two users to successfully swap their charged devices;
- *confirmationTime* – an estimation of time needed for the transaction to be considered as valid by Swether. In our case this includes time for 3 block depth for security, too.

The access to the functions for an administrator is limited to special accounts that are listed as admins. The owner account of the contract automatically serves as an administrator's, too. Besides, he can grant access privileges for up to 16 other accounts. The only action he cannot take is destructing the contract. Key administrator functions enable setting new values for the key parameters that determine the device setup and operation (see the list above). Besides, the following function is provided:

- *getEther* – to retrieve funds from the smart contract address to the administrator account address.

In the same contract there is a single function for users and their plug bookings. Access to this smart contract function is not limited and is thus available to any Ethereum network user (willing to book and pay for booking):

- *bookAPlug* – as parameters it receives a selected plug ID, expected duration and a time limit for maximal acceptable starting time (including current time of possible offsets due to transaction processing). There is an optional NCF tag ID parameter foreseen for future extensions of the smart-charging device, where additional user authentication via NFC will be provided.

There is a separate contract implemented for management of the list of *plugBooking* contract administrators.

The Swether device itself does not update status information in the smart contract. It merely applies the settings from the blockchain, to set its status accordingly.

4.5 Application Remarks

After the development, the Swether device was set for operation for a period of several weeks. During this time we experienced several issues in operation, mostly due to the blockchain and Ethereum clients. These included e.g. BC synchronization problems – node was connected to peers, but the chain was not syncing; problems with Ropsten testnet operation (they were solved by the Ethereum developers by forking the BC in testnet), and *geth* stability. The PRi 3 proved to be capable of running light Ethereum client, but we plan to conduct a deeper investigation of the performance of RPi with various client setups and networks. The times for a booking to become valid (transaction time plus time needed for 5 validations) are in our experience around one minute. This is too long for a quasi-real-time operation and this fact has to be considered in future use-case/activity definitions. There is some space to reduce this delay by reconfiguring the clients and setting a private Ethereum network. But even with all the efforts to do that, we do not expect delays to drop significantly without risking the stability of the blockchain.

Decision for user/administrator interfaces to be implemented with Web technologies proved to be the right one. It ensured nicely designed end-user interface with minimum efforts compared to e.g. native mobile app development. We are currently investigating options to develop mobile clients in comparably pragmatic manner.

5. CONCLUSION

The presented version of Swether is a prototype of an Ethereum blockchain controlled IoT device. We implemented the hardware and software for the device along with Ethereum compliant Web application for Swether control.

In the future Swether devices will be extended with three additional hardware components: LCD for status indication, NFC reader for user authentication and power meter to monitor actual energy consumption. The smart contract and device software will be modified as well. The Swether device will become an active Ethereum node capable of e.g. autonomously reporting actual energy consumption to the smart contract. This will enable us novel use cases, including scheduled reservations, refunding pre-booked but unspent energy, etc. We have also foreseeing a version of the system which applies the Raiden protocol to enable near real-time operation and to reduce the transaction validation costs.

ACKNOWLEDGMENT

The authors wish to acknowledge the support of the research program "Algorithms and Optimization Procedures in Telecommunications", financed by the Slovenian Research Agency.

REFERENCES

[1] "Gartner's 2016 Hype Cycle for Emerging Technologies Identifies Three Key Trends That Organizations Must Track to Gain Competitive Advantage," 16-Aug-2016. [Online]. Available: http://www.gartner.com/newsroom/id/3412017. [Accessed: 05-May-2017].

[2] "Slock.it - Solutions," Workshops, Projects and PoCs. [Online]. Available: https://slock.it/solutions.html. [Accessed: 05-May-2017].

[3] “Share&Charge - Charging Station Network - Become part of the Community!” [Online]. Available: https://shareandcharge.com/en/. [Accessed: 05-May-2017].
[4] “Brainbot Technologies AG,” Smart Contract and Blockchain Consulting for Enterprises. [Online]. Available: http://www.brainbot.com/. [Accessed: 05-May-2017].
[5] “Raiden Network IoT Demo,” *Brainbot Technologies*, 20-Dec-2016. [Online]. Available: https://www.youtube.com/watch?v=t6-rf68taTs. [Accessed: 05-May-2017].
[6] V. Trón, “Ethereum Specification,” 23-Jul-2015. [Online]. Available: https://github.com/ethereum/go-ethereum/wiki/Ethereum-Specification. [Accessed: 05-May-2017].
[7] K. Panetta, “Gartner’s Top 10 Strategic Technology Trends for 2017 - Smarter With Gartner,” 18-Oct-2016. [Online]. Available: http://www.gartner.com/smarterwithgartner/gartners-top-10-technology-trends-2017/. [Accessed: 05-May-2017].
[8] “ETH/USDT Market - Poloniex Bitcoin/Cryptocurrency Exchange.” [Online]. Available: https://poloniex.com/exchange#usdt_eth. [Accessed: 05-May-2017].
[9] “What’s the difference between the `testnet` and the production network technically?,” Ethereum Stack Exchange. [Online]. Available: https://ethereum.stackexchange.com/questions/6278/whats-the-difference-between-the-testnet-and-the-production-network-technical. [Accessed: 05-May-2017].
[10] “IBM Watson IoT - Private Blockchain.” [Online]. Available: https://www.ibm.com/internet-of-things/platform/private-blockchain/. [Accessed: 05-May-2017].
[11] “Blockchain as a Service (BaaS),” Microsoft Azure. [Online]. Available: https://azure.microsoft.com/en-us/solutions/blockchain/. [Accessed: 05-May-2017].
[12] “Solidity — Solidity 0.2.0 documentation.” [Online]. Available: http://solidity.readthedocs.io/en/latest/index.html. [Accessed: 08-May-2017].
[13] “Oraclize Documentation,” Overview. [Online]. Available: http://docs.oraclize.it/#overview. [Accessed: 05-May-2017].
[14] “Raiden Network,” High speed asset transfers for Ethereum. [Online]. Available: http://raiden.network/. [Accessed: 05-May-2017].
[15] “Lightning Network,” Scalable, Instant Bitcoin/Blockchain Transactions. [Online]. Available: http://lightning.network/. [Accessed: 05-May-2017].
[16] “MetaMask,” Brings Ethereum to your browser. [Online]. Available: https://metamask.io/. [Accessed: 05-May-2017].
[17] “Raspberry Pi 3 Model B.” [Online]. Available: https://www.raspberrypi.org/products/raspberry-pi-3-model-b/. [Accessed: 05-May-2017].
[18] “Raspbian.” [Online]. Available: https://www.raspbian.org/. [Accessed: 05-May-2017].
[19] V. Trón, “Geth,” ethereum/go-ethereum Wiki · GitHub. [Online]. Available: https://github.com/ethereum/go-ethereum/wiki/geth. [Accessed: 08-May-2017].
[20] “Ethereum/mist: Mist,” Browse and use Ðapps on the Ethereum network. [Online]. Available: https://github.com/ethereum/mist/#mist-browser. [Accessed: 05-May-2017].
[21] “MetaMask/faq.” [Online]. Available: https://github.com/MetaMask/faq/blob/master/DEVELOPERS.md. [Accessed: 05-May-2017].
[22] “JavaScript API,” ethereum/wiki Wiki · GitHub. [Online]. Available: https://github.com/ethereum/wiki/wiki/JavaScript-API. [Accessed: 05-May-2017].

Golf Swing Data Classification with Deep Convolutional Neural Network

Jiao, Libin; Bie, Rongfang; Wu, Hao; Wei, Yu; Kos, Anton; and Umek, Anton

Abstract: ***Smart sport equipment and body sensory systems are being gradually adopted in professional and amateur sports, so the problem of analyzing the surge of data from sensors used in sports is a novel topic and it is the focus of our research. In this article, we propose a procedure for golf swing data classification using deep convolutional neural network to distinguish between the correctly performed swings and swings with errors from different golf players. The devised convolutional neural network has as input a sequence of 13 signals in which each signal is composed of 1500 data samples. The output is the likelihood of to which golf player and to which swing shape the swing belongs. Based on the swing data sampled from the system integrating two orthogonally affixed strain gage sensors, 3-axis accelerometer and 3-axis gyroscope, we test the performance of the coherence of our network on the real-world dataset. The experimental results including accuracy, precision-recall, f1-scores, and confusion matrix show that our network performs well in the identification of swing shape errors from the professional and amateur golf players.***

Index Terms: ***golf data analysis, classification, convolutional neural network.***

1. INTRODUCTION

Science and technology are playing ever more important role in professional sport, amateur sport, and are penetrating the recreational sport. With the development of miniature, lightweight sensors, sensor networks, and communication technologies, the collection of sport performance data has become easier than ever before. Consequently, the need for processing and analyzing these data has become more demanding, both in volume and in timing constraints. Novel methods and approaches are needed to address this challenge.

Manuscript received Jun. 13, 2017. This work was supported in part by the Slovenian Research Agency within the research program Algorithms and Optimization Methods in Telecommunications.

Libin Jiao, Rongfang Bie, adn Hao Wu are with College of Information Science and Technology, Beijing Normal University. (email: 92xianshen@mail.bnu.edu.cn, {rfbie, 11132015314} @bnu.edu.cn). Yu Wei is with Computer Teaching and Research Section, Capital University of Physical Education and Sports (email: weiyu@cupes.edu.cn), Anton Kos and Anton Umek are with Faculty of Electrical Engineering, University of Ljubljana (email: {anton.kos, anton.umek}@fe.uni-lj.si).

The corresponding author is Rongfang Bie (email: rfbie@bnu.edu.cn)

Various sensors can be attached to the user's body and/or integrated into (smart) sport equipment. The motivation for processing and analyzing sensor data is many fold, from monitoring particular movements of an individual to overseeing the complete action in a group sport match. Our vision is to use sensors' data in biofeedback applications, particularly in biomechanical feedback systems with terminal and/or concurrent feedback [1]. By identification and prevention (interruption) of incorrectly performed actions, a speed up in proper action learning could be achieved [2]. The final goal is a real-time biofeedback system that notifies the user about the incorrect action during its duration or immediately after each period of a periodic activity.

As the state-of-the-art classification approach, convolutional neural network (CNN) has been extensively used in computer vision, pattern recognition, and data mining because of its automatic feature extraction, high accuracy, and high scalability in classification. Due to its benefits in data processing and classification, we propose a convolutional neural network approach for golf data classification. The goal of the classification is to distinguish between the correctly performed swings and swings with errors from different golf players. The input to our model is a sequence of 13 signals in which each signal is composed of 1500 data samples. The output of our model is the likelihood of to which golf player and to which swing it belongs. Our methodology and experimental results can be an inspiration to enhance the classification model in the signal processing to raise the accuracy in a real-time system [3].

The main contribution of this paper is a state-of-the-art convolutional neural network for the classification of 1D sequences of golf swing signals in order to identify the golfers and to detect the golf swing shape errors. Another contribution of our work is the simultaneous analysis of 13 signals each consisting of 1500 samples. We collected signals from the 2-axis strain gage sensor, 3-axis accelerometer, 3-axis gyroscope, and 3-axis magnetometer. The above signals are involved in the synthesized analysis to classify the distinct swing shape error and the golf player,

which demonstrates that our model is adequate to process the multiple-dimensional sequences from golf signals.

The paper is organized as follows: Section 2 presents related work concerning convolutional neural network and golf swing signal analysis. Section 3 describes the designed network model, some implementation details and an acceleration strategy for accelerating the training of the model. Section 4 presents the design of the experiment and experimental result for the validation of the model effectiveness. Section 5 concludes the paper and lists future work.

2. RELATED WORK

In recent years, the neural network application in computer vision has been attracting tremendous attention, especially in image classification, object detection, and image retrieval [4]. Since the AlexNet [5] achieved a successful and groundbreaking performance in ImageNet Large Scale Visual Recognition Challenge (ILSVRC) competitions [6], a surge of breakthroughs concerning the image classification occurred after 2012. Some representative convolutional neural network architectures, such as VGGNet [7], GoogLeNet [8], and ResNet [9], have reached extreme accuracy in image classification, even exceeding the human-level performance of 5.1% top-5 test error [10]. The state-or-the-art Generative Adversarial Network (GAN) [11] and Deep Convolutional Generative Adversarial Network (DCGAN) [12] has achieved that the network can be trained on its own with its specified discriminator & generator structure instead of the supervision of human being.

3. METHODOLOGY

In this section, we present the architecture of our CNN classification model, its implementation details, and the employed training acceleration strategy. Based on a general convolutional neural network, our model takes as input 1D signal sequences from 3 unique sensors, and outputs the likelihoods of player identity and swing shape. In addition, due to the large volume of signals, the model exploits the devised dual GPUs architecture to achieve the high-performance acceleration of training, which is presented in subsection 3.4.

3.1 Network Architecture

The model is composed of 3 types of layers, similarly to the general convolutional neural network: convolutional layers, pooling layers, and fully connected layers, as shown in Figure 1. Following the inherited designing of LeNet-5 [13], we alter and simplify the general convolutional neural network. We take sequences of 13-channel signals as input where each channel is composed of 1500 signal samples, rather than the common CNN that is bred with a batch of 3-channel images; i.e. input layer is bred with sequences of signal with the shape of (num batch)×13×1500. The following 1D convolutional layer convolves the 13-channel signals with the trainable kernels to extract features automatically, and forward propagates the activation of the brewed feature maps as presented in equation (1) [14].

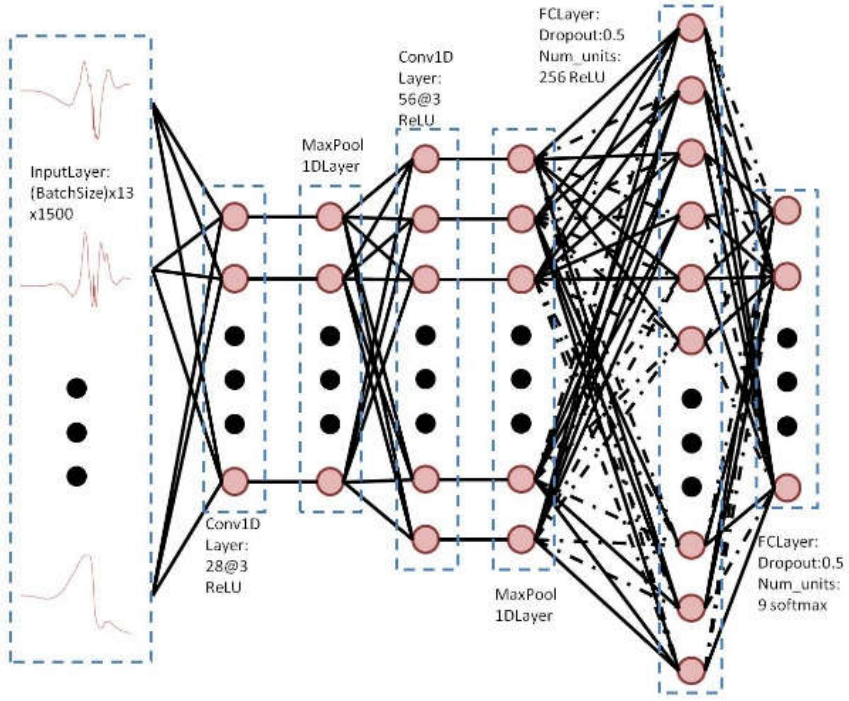


Figure 1 Convolutional neural network for golf swing classification.

$$x_j^\ell = \sigma(\sum_{i \in C_j} x_i^{\ell\ -1} * k_{ij}^\ell + b_j^\ell) \quad (1)$$

The operator * represents the convolution operation, the *i* and *j* represent the i^{th} and j^{th} channel of feature maps from layer *l*-1 and layer *l* the k_{ij} and b_{ij} represent the trainable kernel and the bias, and c_j represents the number of input channel or feature maps from the layer *l*-1.

We refer to ReLU [15]-[16] as the activation function $\sigma(\cdot)$ in equation (2) to transfer the activation value to the following layers, since it can speed up the convergence of the network, and propagate backward more gradient differences to alleviate the diffusion of gradients [17]. The following max pooling layer *l* simplifies and compresses the feature maps from the convolutional layer *l-1*; it highlights the most significant features as well, as equation (3) defines.

$$\sigma(x) = \max(0, x) \quad (2)$$

$$x_i^\ell = \sigma(\max(x_{i_p}^{\ell-1}, x_{i_{p+1}}^{\ell-1})), p = 1, 3, \cdots \quad (3)$$

The operator $\max(\cdot,\cdot)$ chooses the maximum between the 2-neighbourhood of samples to intensify the features as well as simplify the representations of feature maps, which is a common strategy to speed up the feature extraction and the propagation cost, and *p* represents the p^{th} sampling point in a sequence. The last several layers are fully connected layers, defined in equation (4) that fits for the latent distribution among the signals, which play predominant roles of classification.

$$x^\ell = \sigma(W^{\ell-1} \cdot x^{\ell-1} + b^{\ell-1}) \quad (4)$$

Here *x* represents the whole data point that includes 13 channels in which each channel is composed of 1500 sampling points. The last layer consists of 9 softmax [18] activation neurons, which outputs a 0-1 vector representing the predicted category of the input sequence of signal.

3.2 Training and Testing

Our CNN model takes as input (batch_size) sequences of 13-channel 1500 sample signals, and forward propagates the sequences through convolutional layers, max pooling layers and fully connected layers to output the likelihood that who and which swing shape error the signals significantly belong to. The overall mean of category cross entropy [19] between each prediction o_i and its expectation t_i in a batch *B* is defined as the loss function to measure the error between the output and actual labels, which is defined by equation (5).

$$\bar{L} = -\frac{1}{|B|} \sum_{i \in B} \sum_j t_{i,j} \log(o_{i,j}) \quad (5)$$

The model propagates backward the loss from equation (4), and calibrates the parameters *W* and *b* to converge by the update function based on gradient descent methods [20]. The deterministic CNN model takes as input the data without labels to propagate forward, and the label l_i for each datum is determined by the output likelihood o_i that are calculated by equation (6).

$$l_i = \arg\max_j(o_{i,j}) \quad (6)$$

3.3 Implementation Details

We build up our model according to figure 1 that is composed of two convolutional layers, two max pooling layers and two fully connected layers. Layer 1 and layer 3 in our CNN architecture are convolutional layers that are formulated with 28 and 56 3-sample kernels, respectively. The max pooling layers follows each convolutional layer that selects the maximum from 2-neighbourhood of samples. The last two fully connected layers contain 256 and 9 neuron to classify, where the probability of dropout is set at 0.5 to alleviate the overfitting. We choose ReLU activation function since it is beneficial to back-propagate the errors from the last layer, and moderate the phenomenon of vanishing gradients. The softmax activation function is employed in the output layer to transmit the likelihood that the input signals significantly belong to. Some hyper parameters are listed here: the batch size is set to 5, and the update function is ADAM [21].

3.4 Acceleration Strategy

Inspired by the official implementation of multiprocessing in Python and multiple GPUs in Theano [22], we compose the multiple GPUs implementation for Lasagne [23] to accelerate training of the large volume of signals, which is shown in Figure 2. Given a large-volume dataset, CPU starts the batch iterative generator to output several mini-batches of data to breed the neural network in each GPU, and synchronizes the parameters and the gradients by averaging them from each GPU. The independent processes in GPUs execute the forward-propagation and backward-propagation to brew losses, parameters and gradients for each mini-batch from CPU, and transfer to CPU the update of losses, parameters and gradients. CPU builds the tunnel for data transferring by queues in the RAM, and exchanges the mini-batches GPUs take for shuffle

in the next epoch.

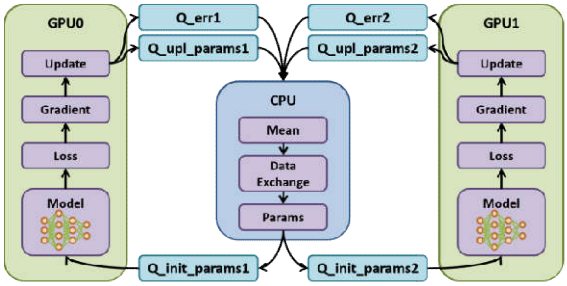


Figure 2 Multiple-GPUs acceleration for Theano [22]

4. *Experiment and Results*

We first introduce some meta-information concerning our real-world golf swing data that includes 13 channels involving 2-axis strain gage sensor data, 3-axis accelerometer sensor data, 3-axis gyroscope sensor data and 3-axis rotator data, shown in detail in Table 1.

Table 1 Meta-information of the 13-channel golf swing signal

#channel	Alias	Comment
1	SG1	2-axis strain gage sensor
2	SG2	
3	SGabs	Absolute value of SG
4	AccX	3-axis accelerometer sensor
5	AccY	
6	AccZ	
7	AccAbs	Absolute value of Acc
8	GyroX	3-axis gyroscope sensor
9	GyroY	
10	GyroZ	
11	RotX	3-axis rotator
12	RotY	
13	RotZ	

We train our model on the real-world dataset collected from one professional golf player and two amateur golf players - Player 1, Player 2 and Player 3, which consists of the correct swing and two categories of shape error: pull and slice. The real-world dataset is separated into 9 groups that cover all combinations of golf players and shape errors. Each of group is labeled by a numerical label from 0 to 8. Some details are presented in Table 2.

Table 2 Meta-information about real-world golf swing dataset

Player	Shape error	Error code	#Data	Data ID
Player 1	(correct)	6	22	21-33, 42-50
	Slice	7	5	34-38
	Pull	8	3	39-41
Player 2	(correct)	0	14	1-14
	Slice	1	0	
	Pull	2	0	
Player 3	(correct)	3	6	15-20
	Slice	4	0	
	pull	5	0	

The dataset is separated into a training set containing 80% of data and a test set containing 20% of data, each of which covers all the categories of the shape error from these three golfers. Note that the original dataset is not adequate to train a stable network; therefore, the data augmentation strategy is exploited to enrich the categories of the original dataset. The model is trained with the augmented dataset as input, and is enhanced by the multiple GPUs supported by the NVIDIA CUDA accelerators. We further validate the accuracy of our model with the augmented dataset derived from the 20% of data, and the experimental result is shown in Table 3.

Our model achieves the accuracy of 90.0% in simultaneous identification of golfers and swing shape errors, the accuracy of 100.0% in identification of golfers and the accuracy 90.0% in identification for swing shape errors. Even though the discrimination between correct swings and intentionally faulty swings is not as rigorous as that between golfers, it is demonstrated that our model achieves feasible results in identification of golfers and the categories of golfers and swing shape errors, simultaneously.

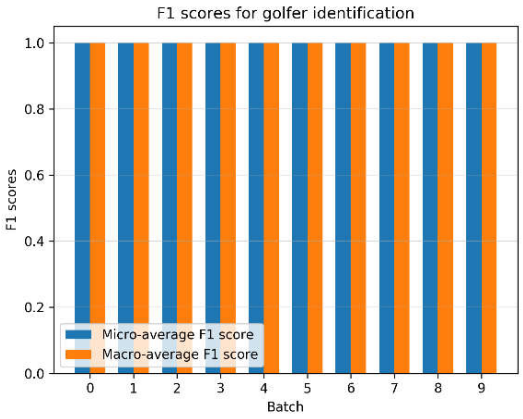


Figure 3 The micro-average and macro-average F1 score for golfer identification

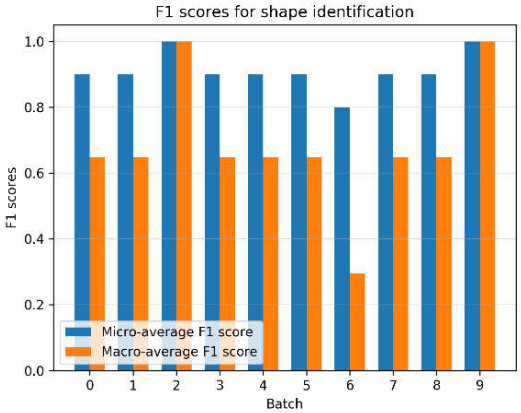


Figure 4 The micro-average and macro-average F1 score for shape identification

Table 3 Validation accuracy of golf player identification, swing shape error identification, and overall validation accuracy

Val acc	Batch number										average
	1	2	3	4	5	6	7	8	9	10	
acc (golfer)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.000
acc (sha err)	0.90	0.90	1.00	0.90	0.90	0.90	0.80	0.90	0.90	1.00	0.900
acc (overall)	0.90	0.90	1.00	0.90	0.90	0.90	0.80	0.90	0.90	1.00	0.900

The precision, recall and F1-scores from the classification reports of golfer and swing shape identification are presented in Table 4, Table 5, Figure 3, and Figure 4. Note that the average indicators including precision, recall, and F1-scores in classification report in Table 4 have reached their limits (1.0). It can be seen that our model can classify the signals from a distinct golfer non-linearly into proper groups on the test set. In addition, the significant precision and recall from each golfer also correspond to the aforementioned inference of effectiveness of our model. It is also shown that our model can adequately capture the unique latent feature representations from each golfer and accurately select all signals that belong to a specific golfer.

For shape identification, we can find some inspiring conclusions in terms of comparison of precision and recall. F1-scores are a synthesized indicator taking into account precision and recall together, which reflects the quality of classification balancing the tradeoff of retrieving as many relevant signals and as many accurate signals as the classifier is able to do. The F1-scores presented in Table 5 show that our model is acceptable in swing shape identification of "correct" signals. Higher precision and lower recall are an evidence that our model can be inferred to be "conservative" in classification; it should classify a signal into a group when it receives a strong probability and attempts to avoid classifying it into a wrong group. Generally speaking, our model is successful in classification of swing shape errors since it should be not "aggressive" in grouping signals, which means it should make fewer mistakes.

Table 4 Classification report of golf player identification

	precision	recall	f1-score	support
Player 2	1.00	1.00	1.00	30
Player 3	1.00	1.00	1.00	10
Player 1	1.00	1.00	1.00	60
Avg/total	1.00	1.00	1.00	100

Table 5 Classification report of swing shape identification

	precision	recall	f1-score	support
(correct)	0.90	1.00	0.95	80
slice	1.00	0.50	0.67	10
pull	1.00	0.60	0.75	10
Avg/total	0.92	0.91	0.90	100

The confusion matrices in Table 6 and Table 7 are adequate complement evidence manifesting the performance of misclassification. There is no error appearing in classification on our test set, which further supports our conclusion that our model can effectively capture the differences between golfers. The unbalanced test set should be the major reason for overfitting of our model on the test set, which leads to the misclassification of our model in shape identification. It is necessary to collect more negative sample to overcome the overfitting of positive samples of our model.

Table 6 Confusion matrix of golf player identification

		Predicted		
		Player 2	Player 3	Player 1
Actual	Player 2	30	0	0
	Player 3	0	10	0
	Player 1	0	0	60

Table 7 Confusion matrix of shape identification

		Predicted		
		(correct)	Slice	Pull
Actual	(correct)	80	0	0
	Slice	5	5	0
	Pull	4	0	6

5. CONCLUSION

In this paper, we propose a procedure for golf swing data classification using deep convolutional neural network in order to distinguish between the correct swings and intentionally faulty ones from different golf players. The state-of-the-art CNN model automatically extracts features from each channel, and identifies the pattern of unique signals from each golfer player and each category of swing shape errors, respectively. The multiple GPUs accelerating strategy exploits multiprocessing to activate acceleration from each GPU, and synchronizes losses, parameters, and gradients by average advocated by CPU. Multiple GPUs outperform the single GPU in time consumption, and maintain the property of fast convergence when training on a large-volume dataset. The experimental indicators including accuracy, precision-recall, F1-scores, and confusion matrix, show that our model achieves feasible accuracy and precision in distinguishing the real-world golf swings collected from the professional golfers and amateur golfers, and in identifying each category of shape error, simultaneously. In the future, we hope that our model can overcome the overfitting in detecting shape errors and help to coach golf players in real-time biofeedback scenarios.

ACKNOWLEDGMENT

This research is sponsored by National Natural Science Foundation of China (No.61571049, 61401029, 11401028, 61472044, 61472403, 61601033), the Fundamental Research Funds for the Central Universities (No.2014KJJCB32, 2013NT57), China Postdoctoral Science Foundation Funded Project (No.2016M590337),

and Graduate Student's Platform for Innovation and Entrepreneurship Training Program (No. 3122121F1) and by SRF for ROCS, SEM

REFERENCES

[1] Anton Umek, Sašo Tomažič, and Anton Kos. Wearable training system with real-time biofeedback and gesture user interface. Personal and Ubiquitous Computing, 19(7):989–998, 2015.

[2] Roland Sigrist, Georg Rauter, Robert Riener, and PeterWolf. Augmented visual, auditory, haptic, and multimodal feedback in motor learning: a review. Psychonomic bulletin & review, 20(1):21–53, 2013.

[3] Anton Umek and Anton Kos. The role of high performance computing and communication for real-time biofeedback in sport. Mathematical Problems in Engineering, 2016, 2016.

[4] Yanming Guo, Yu Liu, Ard Oerlemans, Songyang Lao, Song Wu, and Michael S Lew. Deep learning for visual understanding: A review. Neurocomputing, 187:27–48, 2016.

[5] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems, pages 1097–1105, 2012.

[6] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. International Journal of Computer Vision, 115(3):211–252, 2015.

[7] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.

[8] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2015.

[9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2016.

[10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In Proceedings of the IEEE international conference on computer vision, pages 1026–1034, 2015.

[11] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In Advances in neural information processing systems, pages 2672–2680, 2014.

[12] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434, 2015.

[13] Yann LeCun, L´eon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11):2278–2324, 1998.

[14] Jake Bouvrie. Notes on convolutional neural networks. 2006.

[15] Vinod Nair and Geoffrey E Hinton. Rectified linear units improve restricted boltzmann machines. In Proceedings of the 27th international conference on machine learning (ICML-10), pages 807–814, 2010.

[16] Andrew L Maas, Awni Y Hannun, and Andrew Y Ng. Rectifier nonlinearities improve neural network acoustic models. In Proc. ICML, volume 30, 2013.

[17] Sepp Hochreiter, Yoshua Bengio, Paolo Frasconi, and J¨urgen Schmidhuber. Gradient flow in recurrent nets: the difficulty of learning long-term dependencies, 2001.

[18] Yann A LeCun, L´eon Bottou, Genevieve B Orr, and Klaus-Robert M¨uller. Efficient backprop. In Neural networks: Tricks of the trade, pages 9–48. Springer, 2012.

[19] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Deep learning. MIT Press, 2016.

[20] Mordecai Avriel. Nonlinear programming: analysis and methods. Courier Corporation, 2003.

[21] Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014.

[22] Theano Development Team. Theano: A Python framework for fast computation of mathematical expressions. arXiv e-prints, abs/1605.02688, may 2016.

[23] Eric Battenberg, Sander Dieleman, Daniel Nouri, Eben Olson, Aaron van den Oord, Colin Raffel, Jan Schluter, and Soren Kaae Sonderby. Welcome to lasagne. http://lasagne.readthedocs.io/en/latest/index.html, 2015.

Jiao, Libin is a currently a PhD student in College of Information Science and Technology, Beijing Normal University. He received his Bachelor degree in June 2014 in College of Information Science and Technology. His current research interests include Data Mining, Machine Learning & Computer Vision.

Bie, Rongfang received her Ph.D. degree in 1996 from Beijing Normal University, where she is now a professor. She visited the Computer Laboratory at the University of Cambridge in 2003. Her current research interests include knowledge representation and acquisition for the Internet of Things, computational intelligence and model theory, and Cognitive Radio Networks.

Wu, Hao is currently a postdoctoral research fellow in College of Information Science and Technology, Beijing Normal University. He received the B.E. and Ph.D. degree from Beijing Jiaotong University, Beijing, China, in 2010 and 2015, respectively. From October 2013 to April 2015, he worked as a research associate in Lawrence Berkeley National Laboratory. His current research interests include image retrieval, image processing, and image completion.

Wei, Yu studied her Ph.D. degree in computer science at the College of Information Science and Technology of Beijing Normal University since Sep. 2012. She got her Ph.D. degree in June 2016. Her research interests included Data Mining and Machine Learning.

Kos, Anton received his Ph.D. in electrical engineering from University of Ljubljana, Slovenia, in 2006. He is an assistant professor at the Faculty of Electrical Engineering, University of Ljubljana. His teaching and research work includes communication networks and protocols, quality of service, dataflow computing and applications, usage of inertial sensors in biofeedback systems and applications, signal processing, and information systems.

Umek, Anton received his Ph.D. in electrical engineering from University of Ljubljana, Slovenia, in 1999. He is currently a Senior Lecturer at the Faculty of Electrical Engineering, University of Ljubljana. Since last year, he is the leader of industrial research and development projects in designing of sensor based smart sport equipment and sensor based forestry machinery. His teaching and research work includes signal processing, digital communication, secure communications , access network technologies and design of sensor supported sport training systems. He is a member of IEEE and since 2015 the Slovenian section ComSOC chapter chair.

A Time-Dependent Multi-Class SVM Algorithm for Crowdsourced Mobility Prediction

Zhang, Yuan; Umek, Anton; Obinikpo, Alex Adim; and Kos, Anton

Abstract: ***Accurately predicting a user's next location is quite beneficial for decision making purposes either to the individual or to authorities. With this view in mobile crowdsourcing, recently some researches have been carried out with various techniques employed including the use of repetitive nature of humans, recurrent neural networks, and mobility tracking based on public WLAN services. However, due to irregularities in human movements over a certain period of time, being able to get their precise predicted locations has rendered most of these techniques redundant in some aspects. In this paper, we propose a time dependent multiclass Support Vector Machine (T-MSVM) algorithm for mobility prediction. T-MSVM involves using a particular time to predict a user's future location based on his/her locations at that time in the past. This allows for a more accurate prediction result because it enables the next location of the user to be temporally narrowed down to a specific time frame. We build the proposed prediction algorithm on our T-MSVM model. The algorithm includes both the preprocessing and prediction stages. The preprocessing stage involves training the data while the prediction phase includes the prediction steps and this also concludes the T-MSVM algorithm. Through experiments, we showed that the proposed T-MSVM algorithm can achieve an accuracy of 90% over a week period and more than 95% accuracy over a month period in predicting the next location of a user.***

Index Terms: ***decision rule, mobile crowdsourcing, mobility prediction, multiclass SVM, time dependent***

Manuscript received July 1, 2017. This work was supported in part by the National Natural Science Foundation of China under Grant 61572231, in part by the Shandong Provincial Key Research & Development Project under Grant 2017GGX10141, and in part by the Slovenian Research Agency within the research program Algorithms and Optimization Methods in Telecommunications.

Y. Zhang is with the Shandong Provincial Key Laboratory of network Based Intelligent Computing, University of Jinan, 250022, China.
E-mail: yzhang@ujn.edu.cn
A. Umek is with the Faculty of Electrical Engineering, University of Ljubljana, Slovenia
A. A. Obinikpo is with the School of Electrical Engineering and Computer Science, University of Ottawa, Canada
A. Kos is with the Faculty of Electrical Engineering, University of Ljubljana, Slovenia
The corresponding author is Y. Zhang.

1. INTRODUCTION

The power of crowdsourcing cannot be overemphasized due to its ripple effects as evidenced in recent research and applications [15]. Thus, the need to get people involved in data gathering and information acquisition has been widely recognized and adopted in recent years [3]. This is supported by the diverse advancements in technological outputs involving the upgrade of smart devices, wearable sensors, and so on [14, 27]. Powered with these devices, people are able to participate in data acquisition and information gathering [12]. These generally have eased out the pressure on general data acquisition by reducing the time and energy when compared to conventional data acquisition methods. Gathering and processing the data acquired from these smart devices have become both convenient and pervasive, that is, the outcomes of these processes are being utilized in almost every aspect of life. Because, diverse data are being generated by these devices, there is a need for them to be utilized effectively. One of the possible ways of putting these sets of data to use is by predicting a user's next location. Mobility Prediction (MP) involves using patterns in users' movements from combined data to anticipate their next point of call [26].

Mobility is an inherent characteristic of users in mobile networks, however, it introduces considerable overhead in mobility management and forwarding services to ensure communication reliability [25]. Furthermore, MP uses certain users' information or spatial features such as location to determine a possible location that a user is likely to visit next [2, 11]. The ability of a user to know the conditions in his/her next location is an interesting problem since this information may provide a preventive service in the presence of risky circumstances. A good example for mobility prediction can be the mobility patterns of taxi drivers. Taxi drivers have certain time frames at which they habitually move to a certain location. Their movement patterns can be attained through a set of sensors since each of these taxis are connected to a particular network either through GPS or other location tracking devices like smartphones. These locations may lead to risky

consequences for these drivers and their passengers because of the amount of environmental pollution at these times at these locations. Thus, a taxi driver who has been informed about potential risks at a particular location thanks to the data gathered, may want to reconsider his options.

Despite the obvious usefulness of MP, it still faces a number of challenges which include: (1) the quality of data - too much noise in data gathered, (2) lack of (or inadequate) prediction techniques, and (3) anomalies in mobile user's movements and how to curtail these anomalies [16, 19]. These challenges are in part due to the kind of data gathered, since they are individual groups of data coming from different devices that were combined together, and as such, mining these data accurately is difficult. Data are scattered and using them in their raw state could lead to errors and nonviable research or application output. Training this data for our use is one action that is mandatory and doing this will require use of proper tools. In order to help solve all these challenges, we intend to use the support vector machine (SVM). SVMs are supervised learning models with associated learning algorithms that analyze data and recognize patterns, however it ideally works well under binary decision scenarios (i.e., 2-classes). As MP has multiple application scenarios, SVM is not a viable tool to be used in MP. However, a multi-class support vector machine (MSVM) is a good candidate that could be used for data mining as well as for MP [17, 24].

The main contributions of this research are:

1. The development of a time dependent MSVM (T-MSVM) algorithm for decision making in the Internet of Things. This is a novel approach to MP as we seek to use the user's timetable to predict his/her future locations in the relatively long run. This will help them make informed decision with respect to their intended location. Furthermore, the T-MSVM is a useful algorithm in pervasive computing, and it involves the use and processing of data gathered from pervasive devices for general application purposes e.g., MP.
2. We utilize a large volume of real data for experimental purpose to evaluate the performance of T-MSVM. Experimental results show that our proposed algorithm introduces 90% accuracy for a one week period and over 95% accuracy for a one month period predicting user's next location.

The rest of this paper is organized as follows. In section 2, we discuss the related works. In section 3, the problem formulation is introduced and the T-MSVM algorithm is described in detail. Section 4 focuses on the experimentation, and Section 5 concludes our work.

2. *Related Works*

Mobile phones and other smart devices, due to their being available everywhere, are convenient options for tracking and mining users' positions in daily life as they are usually placed in close proximity to the users. These devices help track user movement and give accurate locations a user visited. Although there have been some works on MP, long term MP still calls for new, effective, and efficient methods.

In [4], Chen et al employed the use of a multiclass SVM as a classifier for training data. This was done in order to be able to predict a user's next cell based on channel states. The authors formulated the prediction scheme as a classification problem based on information that is readily available in cellular networks. By using only Channel State Information (CSI) and handover history, they performed classification by embedding SVMs into an efficient pre-processing structure. However their method cannot be conveniently used for MP because it basically deals with the rate at which a user changes its point of attachment to the wireless communications infrastructure. This is also seen in [5, 6, 18].

According to the authors in [9], smartphones can learn a behavior model that can predict future activities and venues. Personalization is a key in some cases as the interest is in anticipating the movement of a single user. The authors developed a mobility prediction method based on the repetitive nature of human mobility. Using the probabilistic kernel method, location prediction consists of estimating the conditional distribution over the set of future locations and modeling those using discrete states. The probabilistic kernel method enables predicting the possibility of a user's presence at a given location at a given time. Another similar work was also found in [1].

In [8], the authors proposed a framework for predicting where users will go and which application they will use by exploiting the contextual information from smart phone sensors. They used current location context to predict the next location of user.

In [22], the authors introduced a social attractiveness factor of sub-regions for every user in a mobile crowdsensing setting, and proposed a probablistic model to predict the future location of a user. Similarly in [21], historical patterns have been used to predict the future location of a user in mobile crowdsensing setting through a triangulation method.

In [20], the authors reported that identifying the stable path helps improve routing by reducing the overhead and number of connecting interruption. The authors used stable paths because they do not degrade the routing quality of service, since route rediscovery phase involves a substantial overhead. The path stability estimation can be done by predicting the future locations of the nodes. The proposed method was based on using recurrent neural networks for MP. Location prediction is a

case of time series prediction as such time series was used to reach their conclusion.

In [10], the authors predicted user movements based on movements in public WLANs. Their proposed method involves using a combination of multiple similar users for mobility prediction. This however may have a shortcoming in terms of security of the mobile users since it involves the use of public WLANs.

Lv et al in [23] proposed the use of a spatiotemporal predictor and a next location predictor to predict a user's next point of call. These methods were combined with user's living habits for prediction to be made. However, this method might be challenged due to the fact that the living habits of human's tend to change over time as a result of certain factors. Therefore, being able to predict users' next locations based on their habits might be challenging in the long run.

In most of the previous works, the major challenges have been reported as accurate data gathering, combining such group of data and training the fused data. Furthermore, an emphasis was placed on using just the previous/present location to determine the next location of a user.

This research, however, focuses on the time dimension to predict the future location of a user after a few weeks. The idea is basically using the user's previous time table to ascertain their next location.

3. PROBLEM FORMULATION

User mobility results in location changes. These locations are monitored by built-in sensors of mobile devices, series of software or location specific GPS. At specific times, users move through certain locations based on their needs or jobs. The need to accurately get the next location based on these time frames has led to the development of the proposed algorithm, T-MSVM. Using time as a major input factor, we aim to predict user's next location. For instance, we can use the T-MSVM algorithm to predict where a user will probably be every Monday at 12:10:05pm. In order to achieve this, we adopt MSVM as a basis and incorporate the time factor into it.

3.1 Basic Definitions

3.1.1 Linear classifiers

The data for a two class learning problem consists of objects labeled with one of two labels corresponding to the two classes. For convenience and without loss of generality, we assume that the labels are +1(positive examples) or -1(negative examples). Given a line of separation, samples to the right are positive samples (or examples) while samples to the left are negative samples [7].

3.1.2 SVM model

SVMs as indicated in Section 1, are supervised learning methods whose main aim is to solve quadratic problems using optimization [7].
The general SVM model is given as

$$Max\ L = \sum \alpha_i - \frac{1}{2}\sum_i \sum_j \alpha_i \alpha_j\ y_i y_j K(x_i, x) \quad (1)$$

Whereas the decision rule is

$$D(x) = \sum \alpha_i y_i\ K(x_i, x) + b \quad (2)$$

If D(x) ≥ 0 the decision result is positive, otherwise it is negative.

A decision rule is the basis or function for accepting a result or making informed analysis.

Table 1 presents the summary of the basic notations and symbols.

Table 1. Basic Notations

Notations	Definition
$\boldsymbol{\alpha_i}$	Langragian multipliers
$\boldsymbol{K(x_i, x)}$	Kernel
$\boldsymbol{x_i}$	Vectors
$\boldsymbol{D(x)}$	Decision rule
b	Bias
t	Time
l	Location
$\bar{\boldsymbol{w}}$	Weighted vector
$\boldsymbol{\epsilon}$	Arbitrary variables

3.2 Model Development: T-MSVM

In this subsection, we describe our proposed T-MSVM model and t which represents time is the basis for our work. As stated in Section 2, other works were based on previous locations for future location prediction; this however might be limited due to high volume of randomness. But with time as the basis, location prediction gets narrowed down to time frames for efficiency. To develop this model, we intend to extend the 2-class SVM model from subsection 3.1.2 above to k class. Hence we have:

$$\emptyset(\bar{w}, \xi) = \frac{1}{2}\sum \|\bar{w}\| + c\sum_{i=1}^{L} \sum_{m=y_i} \xi_i^m \quad (3)$$

subject to

$$(\bar{w}_{y_i}.\bar{x}_i) + b_{y_i} \geq (\bar{w}_m.\bar{x}_i) + b_m + 2 - \xi_i^m \quad (4)$$

for all $\xi_i^m \geq 0\ ,\ i = 1,2,\ldots.L\ and\ m \in \{1,2,\ldots,k\}\backslash y_i$
Using Lagrangian multipliers, we have the following

$$L = \frac{1}{2}\sum_{m=1}^{k} \|\bar{w}\| + c\sum_{i=1}^{L}\sum_{m=1}^{k} \xi_i^m - \sum_{i=1}^{L}\sum_{m=1}^{k} \alpha_i^m\ [(\bar{w}_{y_i} - \bar{w}_m).\bar{x}_i + V \quad (5)$$

where $V = b_{y_i} - b_m - 2 + \xi_i^m$.

And the constraints are $\alpha_i^m \geq 0, \beta_i^m \geq 0, \xi_i^m \geq 0$
Taking partial derivatives of eqn (5), we have

$$L(\alpha) = 2\sum_{i,m}\alpha_i^m + \sum_{i,j,m}\left[-\frac{1}{2}c_i y_i A_i A_j + \alpha_i^m \alpha_j^{y_i} - \frac{1}{2}\alpha_i^m \alpha_j^m\right](\bar{x}_i.\bar{x}_j) \quad (6)$$

where $c_i^n = \begin{cases} 1 \ if y_1 = n \\ 0 \ if y_i \neq n \end{cases}$,

$A_i = \sum_{m=1}^{k}\alpha_i^m$ and $\sum_{i=1}^{L}\alpha_i = \sum_{i=1}^{L}c_i^m A_i \ ; n = 1, \dots, k$

Also, $0 \leq \alpha_i^m \leq c \ , \alpha_i^{y_i} = 0 \ . i = 1, \dots, L \ ; \ m \in \{1,2,\dots,k\} \backslash y_i .$

Taking the arguments of the maxima of equation (6), we have,

$$f(x,\alpha) = arg\ \max_n [\sum_{i=1}(c_i^n A_i - \alpha_i^m)(\bar{x}_i.\bar{x}) + b_n] \quad (7)$$

Equation (7) is the decision rule for the MSVM.

Therefore, given a set of user's locations with respect to a specific time, we can find a regular pattern between each time and each location.

Let $P(X = L\backslash t)$ be the probability that a user will be at a location at time t; then,

$$P(X = L\backslash t) = \frac{number\ of\ predicted\ location}{number\ of\ actual\ location} \quad (8)$$

$$\begin{cases} P(X = L\backslash t) > 0 \ If\ L\ is\ regular \\ P(X = L\backslash t) = 0 \ If\ L\ is\ irregular \end{cases} \quad (9)$$

The higher $P(X = L\backslash t)$ is, the more likely the user will be at the location. Now, for any t,

$$D(t) = arg\ \max_n \left[\sum_{i=1}^{k}(C_i^m A_i - \alpha_i^m).(t_i.l_i) + b_n\right] \quad (10)$$

Equation (10) is the decision rule for the T-MSVM. Since our input and output parameters are t and l, it is worth mentioning that that l is dependent on t. Thus, T-MSVM module is a one-way system. As time t is inputted into the system, it produces location l as the likely location the user intends to go to. For further confirmation, $P(X = L\backslash t)$ is taken and the value gives the indication of how accurate the prediction is.

3.3 T-MSVM Prediction Algorithm

In this section we describe the T-MSVM algorithm for mobility prediction. To this end, we use two parameters for our research purpose: Time, T_x (Input) and Location L_x (output).

Our goal is to optimally predict user's next location L_x at time T_x based on their previous locations L_p at time T_x. T_x is fixed because we are using it as our base parameter.

This work will be based on two assumptions. This is necessary because in the long run users tend to move periodically in a certain pattern towards a location as seen in the taxi example in Section 1. However certain deviations from their regular routines should be expected. To treat these deviations, we set a margin of deviation cap to check the level and rates of deviation. If the deviation occurs regularly and within the margin of deviation, we can add it to the frequent movement pattern, otherwise we discard it. The assumptions are:

a) Users movements are random, however in the long run they move in a certain (regular) pattern.
b) The margin of deviation is 0.05% whenever their movements become regular in the long run.

In order to predict, we need first to make the dataset useful. Because, the dataset, that could be a set of GPS coordinates or general information of users collected over a sensor network or via smart device, might have lots of not useful information in it. This naturally occurs because every user is associated with a spatial feature. For example home, school, work place etc, and anomalies in their access to these spatial features should be expected. As such, it is necessary to eliminate this unwanted information. This is done by setting parameters and time frame limits, directional limits or even spatial limits.

Next, we initialize $D(t)$ and t_x as this sets us on the path for prediction. Also, we let u represent a user in a pool of users and let l_x represent the respective locations of the user at respective time, t_x. It is quite possible for a user to deviate from his habitual schedule at time t_x, hence the reason for $P(X = L\backslash t)$. $P(X = L\backslash t)$ helps ensure the accuracy of predicted location. Below (Algorithm 1) is the pseudocode for the prediction stage.

Algorithm 1 Prediction Algorithm

```
Input: Time
Required time is the current time
Output: Location
Initiate data training
If data is clean then
  store as clean.data
else if Data is still noisy then
  Re – train
end if
Initialize D(t)
Initiate tx
for each u∈{U} do
  enter tx into D(t)
  for each tx get lx do
    for each lx do
      Compute P(l)
      If P(l)> 0.50 then
        lx is the next location
      else if P(l)≤ 0.5 then
        Discard lx
      end if
    end for
  end for
end for
End prediction process
```

For each user u in the set of users, inputting t_x in to $D(t)$ will give a set of possible locations the user is likely to be present at that particular time. While doing this, $p(l)$ is calculated and compared against the threshold. If it is greater than the threshold, then l_x is the next possible location.

3.4 Advising Algorithm

The T-MSVM system has useful applications in that a proper use of its outcome can help in effective decision making.

Applying the T-MSVM involves the use of information/data from the smart devices and also the prediction outcome. A user is advised to proceed if the intended location is deemed fit or good enough, or is advised to be cautious. The pseudocode is presented in Algorithm 2.

Algorithm 2 Advising Algorithm

***Input**: Time, location*
Output: *Advice*
  Get clean.data
  Get user.predict
  If *Location is safe* **then**
    advise user to proceed
  else if *Location is unsafe* **then**
    Advice user to be cautious
  end if

The location data is the input in this case, while the advice serves as the output. When we obtain clean.data, the already refined data is called up and made active. The get user.predict activates the already predicted mobility pattern and makes it available for use. Since both clean.data and user.predict run concurrently, the intended location is keyed in. The next line of command starts the loop. Should the intended location be safe, the user is advised to proceed otherwise, they are advised to go other way.

4. EXPERIMENTAL RESULTS

The dataset we used was from a southern city of China. Information in the dataset includes directions, latitudes and longitudes, vehicular speed and time. These are the attributes that we used with time being the main attribute. However, we used a time period of one month in our experimental setup. This is a large pool of data (over 30 Gigabyte) as adequate training is required so as to get quality information from the data. In the dataset, the number of instances is 40,216 and the number of attributes is 10. We thus reduced our sample size to 2000 (by random sampling) for quality rendition of the experiment. Another reason for using a sample size of 2000 is due to a limited time constraint to consider the entire set. However, we intend to expand the sample size in subsequent works. Meanwhile, we used 5 attributes out of the 10 attributes as these five were the required ones for the algorithm. These attributes are car id, rec time, latitude, longitude and direction.

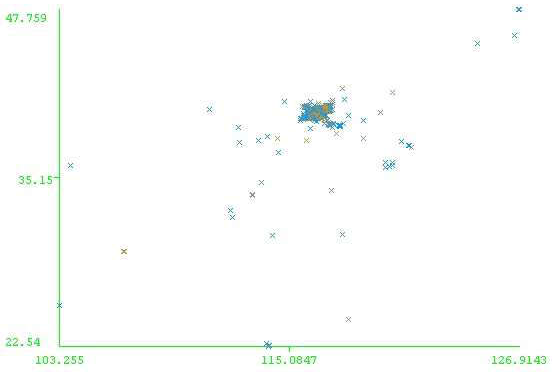


Fig. 1 Latitude vs. longitude: direction.

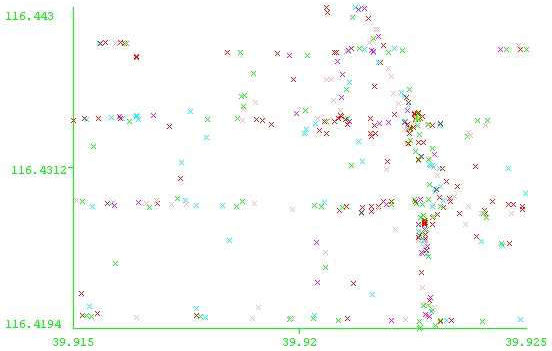


Fig. 2 Latitude vs. Longitude: CarId

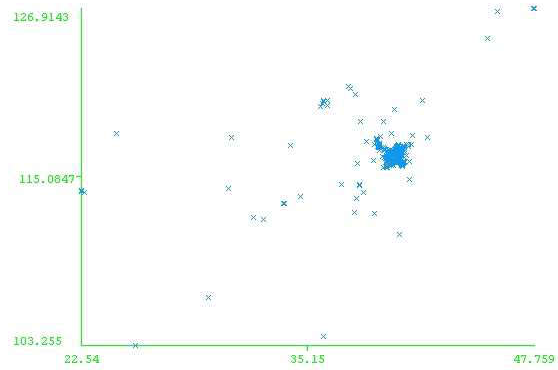


Fig. 3 Latitude vs. longitude: Rectime

4.1. Experiment

First, we preprocessed the data by discarding irrelevant information that was not helpful for our purpose. For example cars with static status during data generation and those that are immobile. Ambiguous date/time renditions were also eliminated. Next, using the ***libsvm*** package in ***WEKA*** and applying the 10 folds cross validation, we were able to reduce the data to a more usable and quality size. We later obtained a visualization of the data as seen in Figures 1, 2 and 3. These figures present the visualization of the data in terms of

latitude vs. longitude with the colored spots representing Direction, Car_Id and Rec_Time respectively. Figure 1 shows the location actualization in terms of direction, while Fig. 2 and Fig. 3 show location actualization in terms of Car_Id and RecTime, respectively. The clustered points in these figures show a more concentrated area of concentration of the taxis.

The prediction process is based on the preprocessed/trained data, and using time as our input attribute, we achieved the following results shown in 4.2 below.

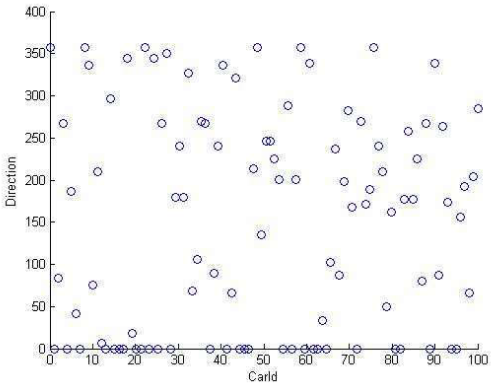


Fig. 4 Predicted Direction at time T.

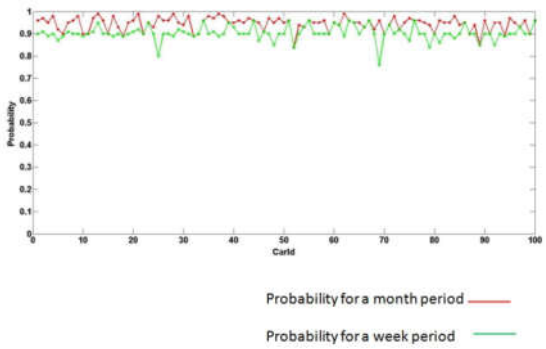


Fig. 5 Probability of future (for one week and a Month period) location at time T.

4.2 Results and Analysis

During the experiment, we used rec.time as the major attribute since it is the input and through this attribute, we discovered that each taxi moved towards corresponding location and direction. Figure 4 shows a pictorial view of each car and their corresponding direction at time *T*. In fact, for each time value we discovered that more often than not, a user goes to the predicted location at least 3 out of 4 times. The blue dots in this case represent each taxi and their corresponding directional outlay.

This confirms that the wider the time frame, the higher the likelihood to be at the next location. The reason is that a much larger interval leads to a higher certainty of visitation. Because, if for 3 consecutive Mondays at time Time *tx*, a user visited same location, then there is a high certainty that he would visit the same location the next Monday at that same time. Also, error rate is quite low. The mean absolute error and root squared error are 0.008 and 0.0769, respectively. This is also expected as the difference between the predicted location and actual location was marginal. One may therefore be tempted to conclude that these taxi drivers were actually moving towards the predicted locations based on the spatial features (home, restaurant, client's home, etc) of these locations. While in some instances we noticed a varying degree of location actualization, but they were discarded as their resultant probability was below the required threshold.

In comparison with the Next Place Prediction using Mobility Markov Chains (n-MMC) developed by [13], Table 2 presents a comparative study between our proposed method and n-MMC. It can be seen that for a one-month time frame our method has an accuracy of 95% and it took a very little time for prediction to be done. While n-MMC has a varying accuracy of 70%-95% which is also dependent on the value of n. Also, there is no existing timing for prediction in n-MMC while ours has a time of 0.10s.

Table 2 T-MSVM vs. n-MMC

Methods	Prediction Accuracy	Time to predict(sec)
T-SVM	95%	0.10
n-MMC	70-95%	-

5. CONCLUSION

This paper proposed a time dependent multiclass SVM algorithm, T-MSVM for mobility prediction in decision making. This is based on the fact that users' movements are random; however they seem to follow a regular pattern in the long run. The algorithm has two phases incorporated into it: the data preprocessing phase and the prediction phase proper. In order to ascertain the accuracy of the T-MSVM, probabilities of the location based on the time were taken and checked with the criterion. It indeed showed that most taxis actually follow a usual pattern and turn up at the same location based on these times. For future work, we intend to further probe into the full use of T-MSVM as a tool for mobility prediction in other fields as well as develop a stand-alone application for this purpose.

REFERENCES

[1] Anjomshoa F, Catalfamo M, Hecker D, Helgeland N, Rasch A, Kantarci B, Erol-Kantarci M, Schuckers S. Mobile behaviometric framework for sociability assessment and identification of smartphone users. InComputers and Communication (ISCC), 2016 IEEE Symposium on 2016 Jun 27 (pp. 1084-1089). IEEE.

[2] Boc M, De Amorim MD, Fladenmuller A. Near-zero triangular location through time-slotted mobility prediction. Wireless Networks. 2011 Feb 1;17(2):465-78.

[3] Chatzimilioudis G, Konstantinidis A, Laoudias C, Zeinalipour-Yazti D. Crowdsourcing with smartphones. IEEE Internet Computing. 2012 Sep;16(5):36-44.
[4] Chen X, Mériaux F, Valentin S. Predicting a user's next cell with supervised learning based on channel states. InSignal Processing Advances in Wireless Communications (SPAWC), 2013 IEEE 14th Workshop on 2013 Jun 16 (pp. 36-40). IEEE.
[5] Chon Y, Talipov E, Shin H, Cha H. Mobility prediction-based smartphone energy optimization for everyday location monitoring. In Proceedings of the 9th ACM conference on embedded networked sensor systems 2011 Nov 1 (pp. 82-95). ACM.
[6] Chon Y, Talipov E, Shin H, Cha H. SmartDC: Mobility prediction-based adaptive duty cycling for everyday location monitoring. IEEE Transactions on Mobile Computing. 2014 Mar;13(3):512-25.
[7] Cortes C, Vapnik V. Support-vector networks. Machine learning. 1995 Sep 1;20(3):273-97.
[8] Do TM, Gatica-Perez D. Where and what: Using smartphones to predict next locations and applications in daily life. Pervasive and Mobile Computing. 2014 Jun 30;12:79-91.
[9] Do TM, Dousse O, Miettinen M, Gatica-Perez D. A probabilistic kernel method for human mobility prediction with smartphones. Pervasive and Mobile Computing. 2015 Jul 31;20:13-28.
[10] Duong TV, Tran DQ. Mobility prediction based on collective movement behaviors in public WLANs. InScience and Information Conference (SAI), 2015 2015 Jul 28 (pp. 1003-1010). IEEE.
[11] Etter V, Kafsi M, Kazemi E, Grossglauser M, Thiran P. Where to go from here? mobility prediction from instantaneous information. Pervasive and Mobile Computing. 2013 Dec 31;9(6):784-97.
[12] Frez J, Baloian N, Zurita G. SmartCity: Public Transportation Network Planning Based on Cloud Services, Crowd Sourcing and Spatial Decision Support Theory. In International Conference on Ubiquitous Computing and Ambient Intelligence 2014 Dec 2 (pp. 365-371). Springer, Cham.
[13] Gambs S, Killijian MO, del Prado Cortez MN. Next place prediction using mobility markov chains. InProceedings of the First Workshop on Measurement, Privacy, and Mobility 2012 Apr 10 (p. 3). ACM.
[14] Guo B, Chen H, Yu Z, Xie X, Huangfu S, Zhang D. FlierMeet: a mobile crowdsensing system for cross-space public information reposting, tagging, and sharing. IEEE Transactions on Mobile Computing. 2015 Oct 1;14(10):2020-33.
[15] Guo B, Liu Y, Wu W, Yu Z, Han Q. Activecrowd: A framework for optimized multitask allocation in mobile crowdsensing systems. IEEE Transactions on Human-Machine Systems. 2017 Jun;47(3):392-403.
[16] He H, Qiao Y, Gao S, Yang J, Guo J. Prediction of user mobility pattern on a network traffic analysis platform. InProceedings of the 10th International Workshop on Mobility in the Evolving Internet Architecture 2015 Sep 7 (pp. 39-44). ACM.
[17] Hsu CW, Lin CJ. A comparison of methods for multiclass support vector machines. IEEE transactions on Neural Networks. 2002 Mar;13(2):415-25.
[18] Javed U, Han D, Caceres R, Pang J, Seshan S, Varshavsky A. Predicting handoffs in 3g networks. InProceedings of the 3rd ACM SOSP Workshop on Networking, Systems, and Applications on Mobile Handhelds 2011 Oct 23 (p. 8). ACM.
[19] Jenssen R, Kloft M, Zien A, Sonnenburg S, Müller KR. A scatter-based prototype framework and multi-class extension of support vector machines. PloS one. 2012 Oct 30;7(10):e42947.
[20] Kaaniche H, Kamoun F. Mobility prediction in wireless ad hoc networks using neural networks. arXiv preprint arXiv:1004.4610. 2010 Apr 26.
[21] Kantarci B, Mouftah HT. Mobility-aware trustworthy crowdsourcing in cloud-centric internet of things. InComputers and Communication (ISCC), 2014 IEEE Symposium on 2014 Jun 23 (pp. 1-6). IEEE.
[22] Kantarci B, Mouftah HT. Trustworthy crowdsourcing via mobile social networks. InGlobal Communications Conference (GLOBECOM), 2014 IEEE 2014 Dec 8 (pp. 2905-2910). IEEE.
[23] Lv Q, Qiao Y, Ansari N, Liu J, Yang J. Big data driven hidden Markov model based individual mobility prediction at points of interest. IEEE Transactions on Vehicular Technology. 2017 Jun;66(6):5204-16.
[24] Sangeetha R, Kalpana B. Performance evaluation of kernels in multiclass support vector machines. training. 2011;2:2.
[25] Wang J. Exploiting mobility prediction for dependable service composition in wireless mobile ad hoc networks. IEEE Transactions on Services Computing. 2011 Jan;4(1):44-55.
[26] Yakoub F, Zein M, Yasser K, Adl A, Hassanien AE. Predicting personality traits and social context based on mining the smartphones SMS data. InIntelligent Data Analysis and Applications 2015 (pp. 511-521). Springer, Cham.
[27] Zhang Y, Sun L, Song H, Cao X. Ubiquitous WSN for healthcare: Recent advances and future prospects. IEEE Internet of Things Journal. 2014 Aug;1(4):311-8.

A Review on Methods for Assessing Driver's Cognitive Load

Stojmenova, Kristina; Stojmenova Duh, Emilija; and Sodnik, Jaka

Abstract: ***Performing additional tasks while driving distracts the driver and imposes additional cognitive load. Driver's cognitive distraction is a less known and explored form of distraction because it is not as obvious as visual or manual distraction. Assessing cognitive load is also quite demanding, especially in a dynamic environment such as operating a vehicle. In this paper, we review possible methods and their characteristics for the assessment of cognitive load. According to the type of data and the way it is collected, they can be divided into three groups: methods for subjective assessment, methods for indirect assessment, and methods based on psycho-physiological and neurological measures. Subjective assessment is considered as the easiest to implement; however, self-reported scores do not always give reliable results. Indirect assessment methods, on the other hand, observe task performance of an additional secondary task while driving, since research studies have shown it is highly correlated with changes in cognitive load. Secondary tasks, however, impose additional distractions, therefore, for safety, these methods should be performed primarily in simulated environments. The last group of methods uses driver's physiological and neurological signals as a reflection of changes in cognitive load. These signals provide most reliable data for changes in cognitive load, but only when collected properly. Reliable data collection is possible mainly by using expensive, high-end equipment, which consequently makes these methods less widely assessable.***

Index Terms: ***cognitive load, driving, human-computer interaction, distraction, injury prevention.***

1. INTRODUCTION

Human-computer interaction (HCI) or human-machine interaction (HMI) is a research field that explores possible ways of information exchange and interaction between a human and different types of machines. The interaction can be made using one or more, single or multimodal, senses (human) or sensors (computer or a machine). Multimodal interaction includes simultaneous use of a greater number of input-output devices and communication channels, and it occurs when the user has to perform more tasks simultaneously.

The usability of In-vehicle Information Systems (IVIS) is defined by their input and output components which compose the user interface. Most of today's IVIS are multimodal and present information in more than one form: visual, audio or even tactile, and can be operated manually or auditory (voice control). Multimodal in-vehicle interaction has become a necessity with the current lifestyle of multitasking and busy schedules, as a lot of work is done on the go, including in the car. However, driving is a dynamic task and requires constant attention and processing of surrounding information by the driver. Therefore, performing additional tasks can distract the driver from the primary task of driving and can impose additional cognitive load. In fact, using IVIS and different mobile devices while driving is one of the most common reasons for vehicle accidents among both novice and experienced drivers [1].

In order to preserve driver safety while using IVIS and keep or even increase the amount of information it presents to the driver, it is important to search for new innovative ways of in-vehicle human-computer interaction and, at the same time, develop reliable methods for assessment of driver distraction caused by these systems. Driver distraction is defined as anything that distracts the driver from his/hers primary task of driving, and can be found in three forms [2]:

- visual distraction (eyes off the road),
- manual distraction (hands off the wheel), and
- cognitive distraction (mind off the road).

1.1 Visual Distraction

Visual distraction or eyes off the road is any type of HMI interaction that requires driver's visual attention and diverts his/her visual focus away from the road. This kind of distraction is caused by IVIS that present information visually. Typical visual distraction is caused when the driver is presented with information such as, for

Manuscript received May 31, 2017. This work was supported by Slovenian Research Agency.

Kristina Stojmenova and Jaka Sodnik are with the Laboratory of Information Technologies, Faculty of electrical engineering, University of Ljubljana, Slovenia.
(Corresponding author: kristina.stojmenova@fe.uni-lj.si).

example, temperature levels of the inside or outside of the vehicle, road layout on the display of a navigation system, technical details of the operation vehicle in image or text, or even short messages, emails and written news (Figure 1).

Figure 1: Example of visual-manual IVIS where information is presented visually and input information is communicated manually [3].

Visual representation of information is widely used, as the visual perception channel can perceive the greatest amount of information at a time, compared to other perception channels. Most vehicles still present information on displays placed on the dashboard (head-down displays) and under the driver's visual field, which is otherwise focused on the road. In order to reduce the eyes-off-the-road effect, new innovative ways of information presentation have been introduced. One of them is the introduction of head-up displays, which project visual information to the driver's visual field (i.e., above the steering wheel) and allow his/her eyes to be focused on the road. Typically, they use the vehicle's windshield for projection of information, while some even project information virtually on the road, preventing not only the change of visual field but also the change of driver's visual focus (Figure 2) [4], [5], [6][7][8].

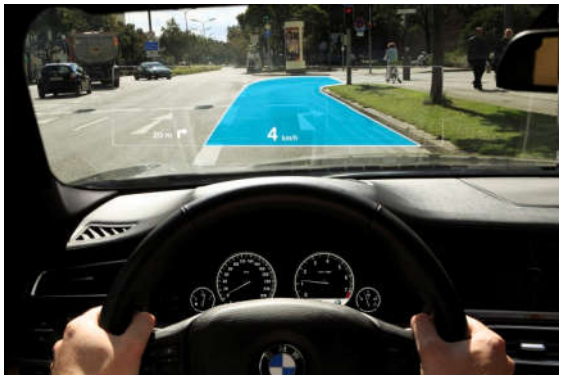


Figure 2: Head-up displays that project information on the road [9].

Visual distraction can be avoided by using audio IVIS, where information is offered auditorily instead of visually. Studies have shown that audio displays represent a safe way of in-vehicle information presentation [10], [11], [12], [13], [14].

1.2 Manual Distraction

Manual distraction is a type of interaction that requires the driver to let go one or both hands off the steering wheel. This kind of distraction is caused by IVIS that require manual input, as they are operated by hand gestures. Typical manual distractions are caused when the driver enters information/commands into IVIS, for example tuning the radio, turning on heating or cooling, entering address information into a navigation system, or even sending short messages. A great step toward the reduction of manual distraction was made when the input parts of the IVIS were separated from the information presentation part, and placed on the steering wheel. In most modern vehicles, drivers can operate almost all of the IVIS functions from the input buttons on the steering wheel [15], [16]. Change of modality is also possible in order to avoid manual distraction. Voice operated systems have also been introduced in vehicles and shown to be safer and well accepted among drivers [13][14].

1.3 Cognitive Distraction

Cognitive distraction or mind off the road is the most complex distraction as it distracts driver's thoughts from the happening on the road and operating the vehicle. Unlike the previous two distractions, cognitive distraction is caused by the content of the information and not from the way information is presented (visual, tactile or auditory). Cognitive distraction describes the mental state of the driver rather than the actions he/she performs. It takes place when the driver's thoughts are not "on the road" but are occupied with the processing of information irrelevant for operating the vehicle (for example, conversation on the telephone or reading an e-mail).

Cognitive load is defined as a multidimensional concept that represents the load a particular task imposes on the operator of the task [17]. The multidimensional concept consists of factors that affect cognitive load, and factors that are affected by cognitive load (factors that are being assessed). Casual factors represent tasks and subjects, their characteristics, environment and the interactions between them. Factors that are assessed are mental load, mental effort, and performance. Mental load is load imposed by the task; mental effort is the amount of effort that the subject allocates to solving the task; and performance describes the success rate of completing the task.

Cognitive load theory suggests that human cognitive processing skills are limited [18]. Therefore, it recommends the use of instructional

methods that effectively stimulate people to use their already acquired knowledge and skills (stored in their long-term memory, which has a greater capacity) for solving and dealing with new situations and information (in their sensory and short-term memories, which have less capacity). This theory can also be used when designing new IVIS in order that input and output information is communicated in a way that it is less cognitively demanding. Furthermore, it can be used to stimulate the use of long-term memory knowledge to process information in the short-term working memory.

In order to design less cognitively demanding IVIS, we must know how to estimate the amount of cognitive load that different types of tasks impose on the driver. In this paper, we present methods that have been proven to be sensitive to changes in cognitive load and could be used for assessing drivers' cognitive load.

2. COGNITIVE LOAD ASSESSMENT

Performing one or more highly cognitively demanding tasks simultaneously can cause cognitive overload. Attention wise, different tasks can use different attention resources or share them. Attention is defined as concentration on a specific source of information [19]. If several tasks performed simultaneously rely on the same attentional resource, they usually interfere with each other and compete for that resource. In such cases, humans use selective attention and concentrate only on one or a few tasks they consider most important at that given time. While high attention load can eliminate (with selective attention) the processing of less important tasks, high cognitive load increases the processing of irrelevant tasks as well [20] & [21]. For example, we can find and track our friend in a group of runners in a marathon, but cannot read and comprehend a complex research work at a rock concert.

In many daily situations not being able to process all of the perceived information at a time can go unnoticed, however when driving this can have serious consequences. In case of cognitive overload, the driver cannot consciously prioritize to perform only driving related tasks and ignore less important tasks such as operating a mobile device. It is therefore extremely important to be able to assess driver's cognitive load, especially the part imposed by the use of IVIS. This information can be later used for IVIS design that optimizes the use of short-term memory and can avoid causing driver's cognitive overload. When assessing driver's cognitive load, it is also very important to choose an appropriate and efficient method that takes into consideration the dynamics of the vehicle as testing environment and the characteristics of IVIS.

There are more than one possible ways to access driver's cognitive load. Most methods can be characterized in one of these three groups:

- methods for subjective assessment,
- methods for indirect assessment, and
- methods based on psycho-physiological and neurological measures.

3. SUBJECTIVE ASSESSMENT

Cognitive load can be assessed with self-evaluation questionnaires, where users report on the level of cognitive load they have experienced while performing a task or operating a system.

NASA Task Load (NASA TLX) is the most commonly used questionnaire for assessing cognitive effort when studying human-machine interaction. It is a multi-dimensional rating questionnaire (Figure 3) based on a weighted average of ratings of six different parameters (mental, physical and temporal demands, own performance, effort and frustration), which give an overall workload score [22]. It has been used in various research HMI fields, and has been shown to be efficient for assessing driver's cognitive load when interacting with new IVIS or touch screens [11] & [23]. This method consists of two parts. In the first part, participants rate each of the parameters based on how much they were present and important for the observed task (scores from 0 to 20, where 20 stands for most difficult). In the second part, estimations of weights are performed by comparing the parameters pair-wise, resulting in 15 comparisons. Whenever one parameter is selected as more important, its counter is increased by 1. The weight of each parameter is from 0 to 5, where 5 stands for most important. The overall workload is then calculated by summing the products of ratings and corresponding weights. Additionally, the sum is divided by 15 (15 paired weights comparisons) to normalize the final score. The overall workload is then calculated by summing the products of ratings and corresponding weights.

Although NASA TLX can be used for assessing driver's cognitive load, some of the subscales can be sometimes misunderstood or cannot be related to the evaluated task or system (for example, physical effort for evaluating audio-vocal IVIS). This was also recognized by Pauzié, who modified the NASA TLX and created an adapted version for driving – the Driving Activity Load Index (DALI) [24]. This method also evaluates six parameters which enable evaluation of driver related tasks and systems: attention, interface, situational stress, and visual, auditory and temporal demand.

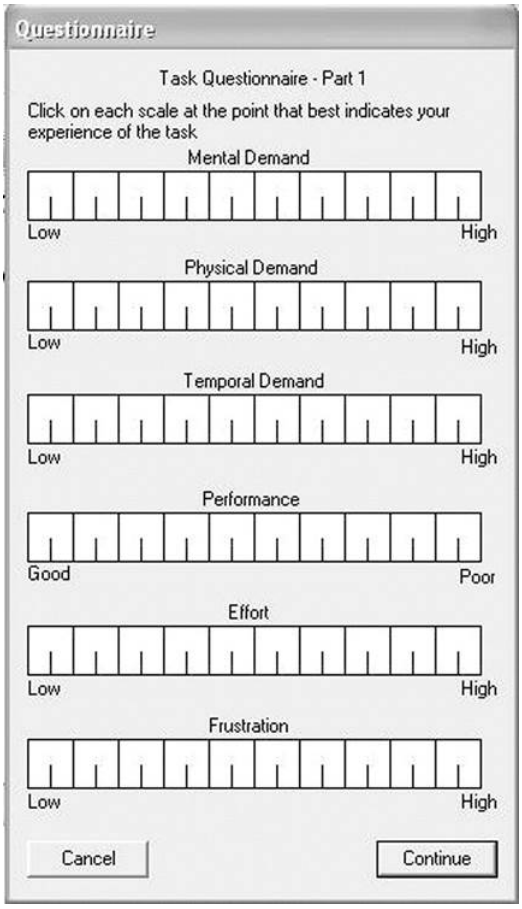


Figure 3: Part 1 of the NASA TLX questionnaire [22].

Research has shown that it is also possible to efficiently and unobtrusively access cognitive load by using single parameter questionnaires (for example, mental workload or task success rate [25]) with the use of 7- or 9-point Likert scales [26].

Although data collection with the use of self-evaluation questionnaires is relatively simple, it is not always reliable. The biggest disadvantage of questionnaires is the fact that they are filled in after a task has been completed. Because of this sequence of events, especially with longer tasks or shorter consecutive tasks, the answers in the questionnaire can be highly influenced by the last performed task and can reflect only that particular part and not the whole task. Another inconvenience of questionnaire tests is data processing and analysis. It is hard to compare absolute workload ratings because the maximum level of workload is not the same for every person. Therefore, the same rated score does not represent the same effort for two different people. Moreover, questionnaire answers can be influenced by people's expectations – test users often write down answers they think the experimenter is expecting them to write, which additionally reduces the reliability of the results.

4. INDIRECT ASSESSMENT

Due to the complexity of direct measurement of cognitive load, especially in a driving environment, indirect methods have been developed instead. For indirect methods, task performance of an additional secondary task, while driving is considered as an indicator of changes in cognitive load as research studies have shown that secondary task performance is highly correlated with changes in cognitive load.

One such method is the lane change test, presented in Figure 4 [27].

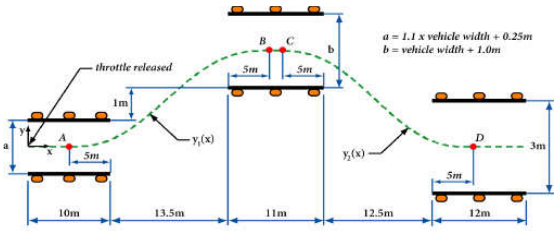


Figure 4: Example of a test track double lane change maneuver [27].

It is characterized by relatively simple implementation in different driving environments and by easy processing of the acquired data [28]. Driver's secondary task in this method is to change the lane in line with the indication of the road signs. Lane deviation, steering angle and percent of correct lane changes are considered as indicators of changes in cognitive load. The instructions for lane change are presented visually, which can influence the results when evaluating the level of cognitive load imposed by visual IVIS. However, an evaluation of how different tasks impact lane change results has shown that visual distractions affect path control, while cognitive tasks affect detection and sign recognition, which should be taken into consideration when practicing this task [30]. Another weakness of this method is the low ecological validity of the test (extent to which research findings would generalize to settings typical of everyday life), as the experimental settings do not give a highly realistic picture of a driving environment.

In October 2016, the Detection Response Task (DRT) was standardized as a secondary task for assessing the attentional effects of cognitive load by the International Standardization Organization (ISO) [31]. The DRT method instructs the driver to respond to 1 or 2 second long stimuli while driving and performing various secondary tasks that are a typical interaction with in-vehicle information systems – IVIS: visual-manual or cognitive tasks. The stimuli can be visual, tactile or auditory, and are presented to the driver in random time intervals that last from 2 to 5 seconds. The driver is asked to respond to the stimuli as soon as possible by pressing a button

(against the steering wheel) attached to his/her left hand index finger.

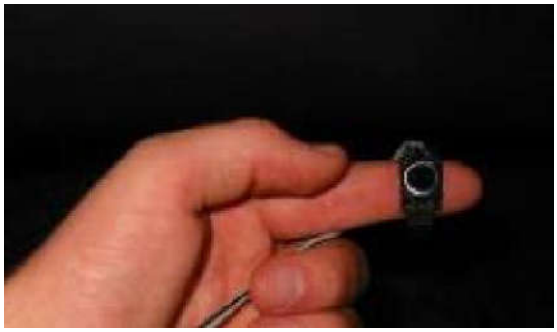


Figure 5: The response button and placement used in all versions of the Detection Response Task [31].

With this method, response times and hit rates are recorded and interpreted as indicators of changes in the driver's cognitive load. Response times are measured as the time from the stimulus presentation until the time the driver answers it, while hit rates are calculated as a percentage of correct responses (occurring from 100 ms to 2,500 ms after introducing a stimulus) out of all presented stimuli. One disadvantage of this method is that it requires high driver involvement, and can therefore be highly intrusive. It is, however, easy to use and gives straightforward results which allow simple assessment of the attentional effects IVIS have on driver's cognitive load.

Stojmenova et al. performed a comparative study of the three versions of the Detection response task and evaluated their sensitivity to the attentional effects that audio, visual and tactile tasks have on driver's cognitive load [32]. Their results indicate that all three versions of the DRT were sensitive to detecting changes in cognitive load; however, none of them showed a consistent advantage in sensitivity in differentiating multiple levels of cognitive load if all response time, hit rate, and secondary task performance are considered. Contrary to their expectations, no correlation was found between DRT modality and the stimuli modality used for the presentation of the secondary tasks. It is important to mention that the DRT ISO standard provides implementation instructions only for the visual and tactile DRT versions, but does not provide this information for the auditory version. In the mentioned study, the authors used a random sound stimulus, which may have influenced their results. Intrigued by this, Stojmenova and Sodnik performed another study with the purpose of determining the most appropriate sound stimulus for the auditory version of the Detection Response Task [33]. They compared white noise, four pure tones (1 kHz, 2 kHz, 4 kHz and 8 kHz) and two sums of pure tones (1 kHz+2 kHz+4 kHz and 2 kHz+4 kHz +8 kHz) as potential sound stimuli for the auditory DRT. The choice to test exactly these sound stimuli was based on previous psychoacoustic research and the fact the human hearing system is most sensitive to sounds between 2 kHz and 5 kHz. Their results showed highest differences in response times for 4 kHz and 8 kHz, indicating that these sound stimuli could be potentially used for the auditory version of the DRT. However, they suggest the use of a 4 kHz signal due to the fact, among others, that the upper frequency limit of hearing decreases with age so the 8 kHz signal could be more difficult to perceive for older drivers.

5. *Psycho-physiological and Neurological Measures*

Assessing cognitive load is also possible by observing a number of different psycho-physiological and neurological measures. Research has shown that features such as galvanic skin response, heart rate and pupillary activity are all correlated with cognitive activity, and that by monitoring the electrical activity of the brain it is possible to get a direct observation of changes in cognitive load.

It has been shown that cardiovascular activities (for example, heart rate and heart rate variability) are affected when the driver is exposed to cognitive load [34]. Reimer et al. performed a driving study in which they increased cognitive load gradually in three difficulty levels while observing heart rate among other psycho-physiological measures. The results showed that heart rate increased in a step-wise fashion through the first two increases in load and then showed a less marked increase at the highest task difficulty level [35].

Novak et al., for example, showed that it is possible to access heart rate data with a low cost wristband; however, they suggest that for more reliable results, a professional medical tool should be used [36]. On the other hand, they found out that the same low cost wristband (Microsoft band [37]) can also be used for galvanic skin response and skin temperature as cognitive load indicators. For more accurate measures, professional medical sensors should be used, such as the Empatica E4 wristband (Figure 6) [38]. This sensor device is equipped with an electro-dermal activity sensor (for assessing GSR), infrared thermopile (for measuring skin temperature), and a photoplethysmography sensor (for measuring blood volume pulse and consequently derivation of heart rate and heart variability values). Other researchers have also shown that these parameters can be used as accurate indicators of driver's cognitive distraction and stress [39].

Figure 6: Empatica E4 wristband [38].

Parameters such as pupil dilation, blink duration and frequency, fixation duration and frequency, and saccadic extent have all shown to be correlated to changes in cognitive load [40], [42], [43][44]. Task-evoked pupillary response or pupil dilation, for example, shows changes in cognitive and visual driver's distraction [40]. Blink frequency increases when the driver performs high cognitive load tasks as a function of time [41]. Blink duration, on the other hand, increases when the driver is tired or drowsed, and decreases at the right beginning of a new task [42]. The biggest advantage of these methods is data collection (remote or wireless glasses eye trackers), which is completely non-invasive and highly accurate. The only disadvantage is light sensitivity because of the speed changes of the driving environment, light traffic signs etc., which is, however, not the case with higher-end eye trackers (for example, Tobii [45], SMI [46] or Smart Eye [47]). The authors of this manuscript participated in a study (Čegovnik et al.) for low cost eye tracker evaluation and the preliminary results showed that in a light controlled environment it is possible to accurately observe eye parameters such as pupil size, blink rate and fixation duration. These parameters can be used for the efficient assessment of changes in cognitive load; this fact indicates that low cost devices can also be used for such kind of research when performed in a controlled environment that considers all of the limitations of the used measurement equipment.

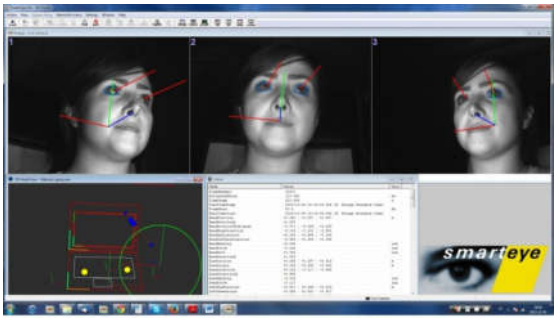


Figure 7: SMART EYE eye tracker pro for multi-camera remote 3D eye tracking [47].

Electroencephalogram or EEG is the most accurate method for assessing driver's cognitive load as it monitors the electrical activity of the brain, which is directly impacted by changes in cognitive load. Brain activity is acknowledged whenever a potential difference appears between the electrode with an active neural signal and the electrode that is placed on an inactive surface, which serves as a reference point [48]. Based on the type of the activity that provokes the signal (spontaneous or event-related), the obtained results are divided into two categories. In the first group, signals in the range of between 1 Hz and 100 Hz are detected. These signals correspond to spontaneous activities. The second group of signals corresponds to various sensory, cognitive or motor events, which can be detected as event-related potential (ERP) in the brain. Although EEG is used in driver related studies [49], [50][51], performing these measures in a moving or simulated vehicle is not a simple task. Moreover, the equipment for collecting quality EEG is very expensive and often available only to medical research institutions.

6. CONCLUSION

Assessing driver's cognitive load imposed by the use of IVIS is not a simple task. When choosing the right method, researchers should take into consideration the modalities of the IVIS (audio, tactile, visual or multimodal), the environment in which the study is performed (surrogate, simulated or real-life), and, understandably, the resources available for the study. However, regardless of the method that is chosen, it is important to include the assessment of driver's cognitive load and overall impact IVISs have on driver's and their performance in order to achieve better driver's experience, increase driver safety and reduce distraction related road accidents.

REFERENCES

[1] Klauer, S. G., Guo, F., Simons-Morton, B. G., Ouimet, M. C., Lee, S. E., & Dingus, T. A. Distracted driving and risk of road crashes among novice and experienced drivers. *New England journal of medicine*, 370(1), 54–59.

[2] Crash avoidance: Distraction. *National Highway Traffic Safety Administration*. Available at: https://one.nhtsa.gov/Research/Human-Factors/Distraction. Assessed on May 15th, 2017.

[3] Tesla. Model S. Available at: https://www.tesla.com/models. Assessed on May 30th, 2017.

[4] Charissis, V., Naef, M., & Patera, M. Calibration requirements of an automotive HUD Interface using a Virtual Environment: Methodology and Implementation. In Proceedings of: *International Conference in Graphics and Visualisation in Engineering,(GVE'07)*, Clearwater, Florida, USA, 2007.

[5] Park, H., & Kim, K. H. Efficient information representation method for driver-centered AR-HUD system. In Design, User Experience, and Usability. *User Experience in Novel Technological Environments* (pp. 393–400). Springer Berlin Heidelberg, 2013.

[6] Poitschke, T., Ablaßmeier, M., Rigoll, G., Bardins, S., Kohlbecher, S. and Schneider, E., Contact-analog information representation in an automotive head-up display. In *Proceedings of the 2008 symposium on Eye tracking research & applications* (pp. 119–122). ACM, 2008.

[7] Tran, C., Bark, K. and Ng-Thow-Hing, V.. A left-turn driving aid using projected oncoming vehicle paths with augmented reality. In Proceedings of the 5th International *Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 300–307). ACM, 2013.

[8] Weinberg, G., Harsham, B., & Medenica, Z. Evaluating the usability of a head-up display for selection from choice lists in cars. In Proceedings of the 3rd *International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 39–46). ACM, 2011.

[9] BMW. How can BMW's Head-Up Display support driving? Available at: http://www.bmwblog.com/2016/02/20/video-can-bmws-head-display-support-driving/. Assessed on May 30th, 2017.

[10] Dicke, C., Jakus, G., & Sodnik, J. Auditory and Head-Up Displays in Vehicles. *In Human-Computer Interaction. Applications and Services* (pp. 551–560). Springer Berlin Heidelberg, 2013.

[11] Jakus, G., Dicke, C., Sodnik, J. A user study of auditory, head-up and multi-modal displays in vehicles. In *Applied ergonomics*, vol. 46, pages 184–192., 2015.

[12] Sodnik, J., & Tomažič, S. Spatial Auditory Human-computer Interfaces. *Springer*. 2015.

[13] Sodnik, J., Tomazic, S., Dicke, C., & Billinghurst, M. Spatial auditory interface for an embedded communication device in a car. In *Advances in Computer-Human Interaction*, 2008 First International Conference on (pp. 69–76). IEEE. 2008.

[14] Sodnik, J., Dicke, C., Tomažič, S., Billinghurst, M. A user study of auditory versus visual interfaces for use while driving. In *International Journal of Human-Computer Studies*, vol. 66(5), pages 318–332. 2008.

[15] González, I. E., Wobbrock, J. O., Chau, D. H., Faulring, A., & Myers, B. A. Eyes on the road, hands on the wheel: thumb-based interaction techniques for input on steering wheels. In Proceedings of Graphics Interface 2007 (pp. 95–102). ACM.

[16] Fujimura, K., Xu, L., Tran, C., Bhandari, R., & Ng-Thow-Hing, V. Driver queries using wheel-constrained finger pointing and 3-D head-up display visual feedback. In Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (pp. 56–62). ACM. 2003.

[17] Paas, F. and van Merriënboer, J.J.G. Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review*, 6, 51-71. 1994.

[18] Sweller, J. "Cognitive load during problem solving: Effects on learning." *Cognitive science* 12, no. 2 (1988): 257-285.

[19] James W. The Principle of Psychology. (Holt) New York, 1890.

[20] Lavie, N., Hirst, A., De Fockert, J. W., & Viding, E. Load theory of selective attention and cognitive control. *Journal of Experimental Psychology: General*, 133(3), 339. 2004.

[21] Lee, Y. C., Lee, J. D., & Boyle, L. N. The interaction of cognitive load and attention-directing cues in driving. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 51(3), 271-280. 2009.

[22] NASA TASK LOAD INDEX (TLX) Paper and Pencil Manual. Available at: http://humansystems.arc.nasa.gov/. Assessed on May 30th, 2017.

[23] Jakus, Grega, Dicke, Christina, Sodnik, Jaka. Subjective evaluation of auditory and head-up displays in vehicles. In *Proceedings of the 4th International Conference World Usability Day Slovenia 2013*, Ljubljana, Slovenia, 25th November 2013, (International series on information systems and management in creative media). Tampere: IugYmedia, 2013, str. 29–32.

[24] Pauzié A. Evaluating driver mental workload using the driving activity load index (DALI). In *Procedings European Conference on Human Centred Design for Intelligent Transport Systems*, pp. 67–77.

[25] Paas, F. G., & Van Merriënboer, J. J. Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. Journal of educational psychology, 86(1), 122. 1994.

[26] Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. Cognitive load measurement as a means to advance cognitive load theory. Educational psychologist, 38(1), 63–71. 2003.

[27] ISO. Road vehicles -- Ergonomic aspects of transport information and control systems -- Simulated lane change test to assess in-vehicle secondary task demand. ISO 26022:2010.

[28] Mattes, S. The lane-change-task as a tool for driver distraction evaluation. In *Quality of work and products in enterprises of the future*, 57–60. 2003.

[29] Harbluk, J. L., Burns, P. C., Lochner, M., & Trbovich, P. L. Using the lane-change test (LCT) to assess distraction: Tests of visual-manual and speech-based operation of navigation system interfaces. In *Proceedings of the 4th international driving symposium on human factors in driver assessment, training, and vehicle design* (pp. 16–22). 2007.

[30] Young, K.L., Lenné, M.G. and Williamson, A.R. Sensitivity of the lane change test as a measure of in-vehicle system demand. Applied ergonomics, 42(4), pp.611-618. 2011.

[31] ISO. (2016). Road vehicles -- Transport information and control systems -- Detection-Response Task (DRT) for assessing attentional effects of cognitive load in driving. ISO 17488:2016.

[32] Stojmenova, K., Jakus, G. and Sodnik, J. Sensitivity evaluation of the visual, tactile, and auditory detection response task method while driving. *Traffic injury prevention*, pp. 431-436. 2016.

[33] Stojmenova, K., Policardi, F. and Sodnik, J., 2017. On the selection of stimulus for the Auditory Variant of the Detection Response Task Method for driving experiments. *Traffic Injury Prevention*, (just-accepted), pp.00-00.

[34] Mehler, B., Reimer, B. and Wang, Y. A comparison of heart rate and heart rate variability indices in distinguishing single-task driving and driving under secondary cognitive workload. In *Proceedings of the Sixth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design* (pp. 590-597). 2011

[35] Reimer, B., Mehler, B., Coughlin, J. F., Godfrey, K. M., & Tan, C. An on-road assessment of the impact of

cognitive workload on physiological arousal in young adult drivers. *In Proceedings of the 1st international conference on automotive user interfaces and interactive vehicular applications* (pp. 115–118). ACM. 2009.
[36] Novak, K., Stojmenova, K., Jakus, G., Sodnik, J. Assessment of cognitive load through biometric monitoring. In: Zdravković, M., Konjović, Z., Trajanović, M. (Eds.*) ICIST 2017 Proceedings* Vol.1, pp.303-306, 2017.
[37] Microsoft. Microsoft band features. Available at https://www.microsoft.com/microsoft-band/en-us/features. Assessed on May 30th, 2017.
[38] Empatica Inc. Empatica E4 wristband. Available at: https://www.empatica.com/e4-wristband. Assessed on May 30th, 2017.
[39] Healey, J. A., & Picard, R. W. (2005). Detecting stress during real-world driving tasks using physiological sensors. *Intelligent Transportation Systems, IEEE Transactions* on, 6(2), 156–166.
[40] Beatty J., "Task-evoked pupillary responses, processing load, and the structure of processing resources". In *Psychological Bulletin*, vol. 91 (2), 1982, pp. 276–292.
[41] Fakuda K., Stern J. A., Brown T. B. and Russo M. B. (2005). Cognition, Blinks, Eye-Movements, and Pupillary Movements During Performance of a Running Memory Task. In: *Aviat Space Environ Med*, vol. 76, pages 75–85.
[42] Siegle, G.J., Ichikawa, N. and Steinhauer, S. (2008). Blink before and after you think: blinks occur prior to and following cognitive load indexed by pupillary responses. *Psychophysiology*, 45(5), pp. 679–687.
[43] Palinko, O., Kun, A.L., Shyrokov, A. and Heeman, P., (2010). Estimating cognitive load using remote eye tracking in a driving simulator. In *Proceedings of the 2010 symposium on eye-tracking research & applications* (pp. 141-144). ACM.
[44] Heeman, P.A., Meshorer, T., Kun, A.L., Palinko, O. and Medenica, Z. (2013). Estimating cognitive load using pupil diameter during a spoken dialogue task. *In Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 242–245). ACM.
[45] Tobii. Tobii pro glasses 2. Available at https://www.tobiipro.com/product-listing/tobii-pro-glasses-2/. Assessed on May 30th, 2017.
[46] SensoMotoric instruments. SMI Eye Tracking Glasses. Available at https://www.smivision.com/eye-tracking/product/eye-tracking-glasses/. Assessed on May 30th, 2017.
[47] Smart eye. Smart eye pro. http://smarteye.se/technology/. Assessed on May 30th, 2017.
[48] Stojmenova, K., Sodnik, J. Methods for assessment of cognitive workload in driving tasks. In: Zdravković, M., Trajanović, M., Konjović, Z. (Eds.) *ICIST 2015 Proceedings* Vol.1, pp.229-234, 2015.
[49] Berka, C., Levendowski, D. J., Lumicao, M. N., Yau, A., Davis, G., Zivkovic, V. T. & Craven, P. L. EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviation, space, and environmental medicine*, 78(Supplement 1), B231–B244. 2007.
[50] Haapalainen, E., Kim, S., Forlizzi, J. F., & Dey, A. K. Psycho-physiological measures for assessing cognitive load. In *Proceedings of the 12th ACM international conference on Ubiquitous computing* (pp. 301–310). ACM. 2010.
[51] Borghinia G., Astolfia L., Vecchiato G., Mattiaa D., and Babiloni F. Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness, *Neuroscience and Biobehavioral Reviews* vol. 44 pp.58–75. 2014.

Kristina Stojmenova received her B.Sc. and M.Sc. degrees in industrial engineering from University of Maribor in 2011 and 2014, respectively. She is currently a PhD candidate and a junior researcher at the Information and communications technology department at the Faculty of electrical engineering, University of Ljubljana. Her field of research is human-computer interaction in in-vehicle information systems, mainly focusing on assessing their effect on driver's cognitive load.

Emilija Stojmenova Duh received her B.Sc. and PhD degrees in electrical engineering from University of Maribor in 2009 and 2013, respectively. Currently she is an Assistant Professor at the Faculty of Electrical Engineering, University of Ljubljana. Her research work focuses mainly on the fields of open innovation, co-creation, user-centered design and methodologies for evaluating user experience and usability for specific groups of users, such as: elderly people, children and people with disabilities. Previously, she was employed at Iskratel, Ltd. as a user experience manager, where she was responsible for the overall user experience in the company.

Jaka Sodnik received his B.Sc., M.Sc. and PhD in electrical engineering from the University of Ljubljana. Currently he is an Associate Professor for the field of Electrical Engineering, at the Faculty of Electrical Engineering, University of Ljubljana.
His early research focused on analysis and generation of spatial sound and its use in human-machine interaction. As a visiting researcher at the HIT Lab New Zealand, he was also involved in several research projects in the field of virtual and augmented reality. He also advises and supervises the R&D department of NERVteh d.o.o., a company that develops state-of-the-art motion driving simulators and offers various methods of evaluating the drivers' psychophysical state and their driving performance and abilities.

Commutative Rotations in 3D Euclidean Space and Gimbal Spatial Angles

Tomažič, Sašo

Abstract: ***Euler angles are often used to represent the orientation of an object in 3D Euclidean space. Euler angles are angles of three consecutive rotations around two or three axes of an orthogonal coordinate system and bring the object from its initial orientation to its final orientation. Because these rotations are not commutative, the order in which they are applied is important. There are 12 different sets of Euler angles. Because the rotations can be performed around axes of intrinsic or extrinsic coordinate systems, this yields a total of 24 different sequences of Euler angles. In this paper, we introduce gimbal angles as an alternative way of representing object orientation. Gimbal angles correspond to the angles of rotations around the axes of a gimbal. The first axis is extrinsic, the second axis is intermediate, and the third axis is intrinsic. Rotations around these axes are commutative; thus, the order in which they are applied does not matter. They always yield the same final orientation of the object.***

Index Terms: ***commutativity, Euler angles gimbal, rotation, rotation matrix, rotation vector, Tait-Bryan angles***

1. INTRODUCTION

THE orientation of a rigid object in 3D Euclidean space with respect to a fixed coordinate system can be represented with Euler angles. Euler angles are angles of three consecutive rotations that bring the object from its initial orientation to its final orientation. These rotations can be performed around two or three axes of the extrinsic (fixed) or intrinsic (mobile) coordinate system. The rotations are not commutative; thus, the order in which they are applied is important.

There are six sequences of three rotation angles around two axes (*x-y-x, x-y-x, y-z-y, z-y-z, x-z-x*, and *y-x-y*) called proper Euler angles and six sequences of three rotation angles around three axes (*x-y-z, y-z-x, z-x-y, x-z-y, z-y-x,* and *y-x-z*), called Tait-Bryan angles. Because these rotations can be performed either around the axes of the intrinsic coordinate system or the axes of the extrinsic coordinate system, there are 24 different sequences altogether, each yielding a different final orientation of the object.

Manuscript received June 3, 2017.

This work was supported in part by the Slovenian Research Agency within the research program Algorithms and Optimization Methods in Telecommunications.

The author is with the Faculty of electrical engineering, University of Ljubljana, Slovenia (e-mail: saso.tomazic@fe.uni-lj.si).

Tait-Bryan angles found their use in aircraft orientation, where rotations are performed around aircraft principal axes: pitch (lateral or axis x), roll (longitudinal or axis y), and yaw (normal or axis z) in the aircraft intrinsic coordinate system. Different orders of performing pitch, yaw, and roll result in different final orientations of the aircraft. When manipulating an aircraft during a flight, these rotations are mostly performed simultaneously, which yields yet another final orientation.

Another way to represent object orientation is with a single equivalent rotation of a certain angle around a specific axis. This rotation can be expressed by a rotation matrix or a rotation vector. The columns of a rotation matrix are the basis vectors of the rotated coordinate system initially aligned with the fixed coordinate system. The orientation of an object after the rotation is obtained simply by multiplying basic vectors of its intrinsic coordinate system with the rotation matrix.

Object orientation can also be represented by a rotation vector. The rotation vector is a vector oriented in the direction of the rotational axis of the equivalent rotation with the norm equal to the angle of this rotation. This representation is very convenient for use with 3D gyroscopes [1]. In this case, the rotation vector can also be called SORA (Simultaneous Orthogonal Rotation Angle), as its components are equal to the angles of simultaneous rotations around the three axes of the 3D gyroscope intrinsic coordinate system, i.e., the angles that can be calculated directly from 3D gyroscope angular velocity measurements.

None of the above representations are convenient for manipulating the object orientation. In this paper, we propose the set of gimbal angles Γ as an alternative way to represent object orientation in 3D Euclidean space. The set is composed of three angles of rotations around the three axes of a gimbal. We show that these rotations are commutative, and thus, the order in which they are applied is not important. This makes them convenient when used for the manipulation of object orientation.

The paper is organized as follows. In the next section, we introduce the gimbal angles and show that the underlying rotations are commutative. In Sec. 3, we derive the rotation matrix, and

in Sec. 4, we present the rotation vector of a rotation equivalent to the three rotations around the gimbal axes. The next two sections present equations for obtaining the gimbal angles from the rotation matrix and the rotation vector. In the last section, we give some concluding remarks.

2. GIMBAL SPATIAL ANGLE

Let γ_1, γ_2, and γ_3 be the angles of rotations around the axes g_1, g_2, and g_3 of a gimbal, as illustrated in Fig. 1. The object itself rotates around its intrinsic axis g_3, and axis g_3 rotates around the axis g_2, which in turn rotates around extrinsic (fixed) axis g_1. All these rotations can be performed sequentially or simultaneously, e.g., by built-in electric motors.

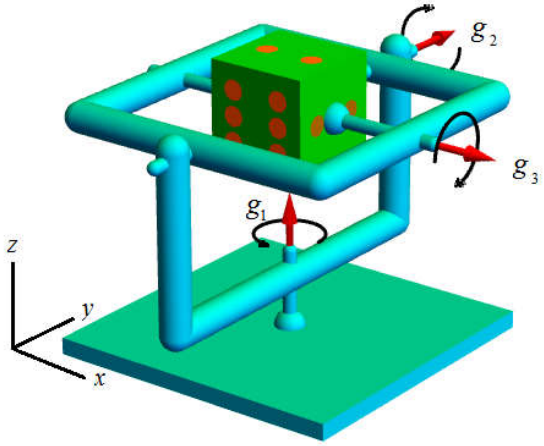


Figure 1a – Gimbal in the initial position.

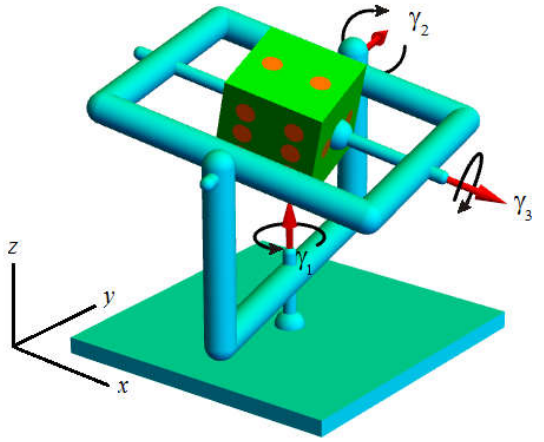


Figure 1b – Gimbal after rotation around gimbal axes. Axis g_1 is fixed in the extrinsic coordinate system. Axis g_2 is an intermediate axis. It rotates around axis g_2. Axis g_3 is fixed in the intrinsic coordinate system. It rotates around axes g_1 and g_2. Angles γ_1, γ_2, and γ_3 are measured in the right-handed direction of the rotations.

Fig. 1a shows the gimbal in its initial position when its axes are aligned with the axes of the extrinsic (fixed) coordinate system: g_1 to z, g_2 to y, and g_3 to x. Fig. 1b shows the same gimbal after rotations of γ_1 around g_1, γ_2 around g_2, and γ_3 around g_3 in the right-handed direction.

Because the rotating object (the dice in Fig. 1) is mechanically attached to the ground through the three gimbal axes, it is more than obvious that its final orientation is determined solely by the values of the angles γ_1, γ_2, and γ_3, regardless of how the object arrived at its final orientation and the order in which the rotations were performed. They can be performed sequentially, in any order, or simultaneously – the result will always be the same. This conclusion implies that the rotations around the gimbal axes are commutative.

Because gimbal angles can uniquely define the orientation of an object, we propose the set of gimbal angles Γ:

$$\Gamma = (\gamma_1, \gamma_2, \gamma_3) \tag{1}$$

as an alternative way of representing object orientation in 3D Euclidean space.

Due to the commutativity of the underlying rotations, such a representation is very convenient for object manipulation. To put an object into a specific orientation, all one has to do is set the correct gimbal angles γ_1, γ_2, and γ_3, regardless of the initial orientation of the object or the order in which rotations are performed.

3. GIMBAL ROTATION MATRIX

Gimbal rotations can be represented by a rotation matrix of a single rotation, which is equivalent to the sequence of the three rotations around the gimbal axes. Denoting the rotation matrix of rotation for the angle γ_n around the g_n gimbal axis, $n \in \{1,2,3\}$, with $\mathbf{R}_n$ and the rotation matrix equivalent to these three rotations with $\mathbf{G}$ we can, due to the commutativity of gimbal rotations, write the following:

$$\begin{aligned}\mathbf{G} &= \mathbf{R}_1 \cdot \mathbf{R}_2 \cdot \mathbf{R}_3 = \mathbf{R}_1 \cdot \mathbf{R}_3 \cdot \mathbf{R}_2 = \\ &= \mathbf{R}_2 \cdot \mathbf{R}_1 \cdot \mathbf{R}_3 = \mathbf{R}_2 \cdot \mathbf{R}_3 \cdot \mathbf{R}_1 = \\ &= \mathbf{R}_3 \cdot \mathbf{R}_1 \cdot \mathbf{R}_2 = \mathbf{R}_3 \cdot \mathbf{R}_2 \cdot \mathbf{R}_1\end{aligned} \tag{2}$$

It is important to note that $\mathbf{R}_1$ also rotates axes g_2 and g_3 and that rotation $\mathbf{R}_2$ also rotates axis g_3. Thus, if $\mathbf{R}_1$ is performed before $\mathbf{R}_2$ and $\mathbf{R}_3$, it modifies them. Similarly, the rotation $\mathbf{R}_2$ modifies the rotation $\mathbf{R}_3$ if performed before it.

Let us first examine the rotation sequence g_1-g_2-g_3. Initially, the gimbal axes are aligned with the axes of the fixed coordinate system:

$$\mathbf{g}_1 = \begin{bmatrix}0\\0\\1\end{bmatrix} \mathbf{g}_2 = \begin{bmatrix}0\\1\\0\end{bmatrix} \mathbf{g}_3 = \begin{bmatrix}1\\0\\0\end{bmatrix} \tag{3}$$

where $\mathbf{g}_1$, $\mathbf{g}_2$, and $\mathbf{g}_3$ denote unit vectors in the directions of the gimbal axes.

To obtain $\mathbf{G}$ we must first rotate the axes g_1 and g_2:

$$\begin{aligned}\mathbf{g}_2 &\leftarrow \mathbf{R}_1 \cdot \mathbf{g}_2 \\ \mathbf{g}_3 &\leftarrow \mathbf{R}_1 \cdot \mathbf{g}_3\end{aligned} \tag{4}$$

Then, we can calculate the intermediate matrix $\mathbf{R}_{21}$:

$$\mathbf{R}_{21} = \mathbf{R}_2 \cdot \mathbf{R}_1 \tag{5}$$

In the next step, we must rotate axis g_3:

$$\mathbf{g}_3 \leftarrow \mathbf{R}_2 \cdot \mathbf{g}_3 \tag{6}$$

Finally, we obtain:

$$\mathbf{G} = \mathbf{R}_3 \cdot \mathbf{R}_{21} \tag{7}$$

We derived $\mathbf{G}$ in accordance with the above expressions. The result was very complex and too long to be included in this paper. We unsuccessfully attempted to simplify it using Wolfram Mathematica [2].

However, because gimbal rotations are commutative, we can use any of the six possible sequences to derive the rotation matrix $\mathbf{G}$. The best choice is the sequence g_3-g_2-g_1 because, for this sequence, all rotations are performed around initial gimbal axes as they are set in (3). In this case, the gimbal rotation matrix is obtained simply by multiplying the rotation matrices of rotations around the axes *x, y,* and *z* of the extrinsic coordinate system:

$$\mathbf{G} = \mathbf{R}_1 \cdot \mathbf{R}_2 \cdot \mathbf{R}_3 = \mathbf{R}_z \cdot \mathbf{R}_y \cdot \mathbf{R}_x \tag{8}$$

where $\mathbf{R}_x$, $\mathbf{R}_y$, and $\mathbf{R}_z$ are rotation matrices of rotations around the axes *x, y,* and *z* of the extrinsic coordinate system, respectively.

After the derivation of the above expression, we obtain [2]:

$$\mathbf{G} = \begin{bmatrix} c_1c_2 & c_1s_2s_3 - c_3s_1 & c_1c_3s_2 + s_1s_3 \\ c_2s_1 & c_1c_3 + s_1s_2s_3 & c_3s_1s_2 - c_1s_3 \\ -s_2 & c_2s_3 & c_2c_3 \end{bmatrix} \tag{9}$$

where c_n denotes $\cos(\gamma_n)$ and s_n denotes $\sin(\gamma_n)$; $n \in \{1,2,3\}$.

Note that the gimbal angles γ_1, γ_2, and γ_3 are equivalent to the Tait-Bryan angles of the sequence *x*-*y*-*z*, with an important difference in that, for gimbal angles, the order of rotations can be altered without affecting the resulting orientation of the object, whereas, for Tait-Bryan angles, the order is important.

Although we were unable to reduce the expressions obtained from other sequences of gimbal rotations to the form in (9) and in this way prove commutativity, we successfully verified commutativity numerically by comparing the results of different gimbal rotation sequences for several randomly chosen gimbal angles.

4. GIMBAL ROTATION VECTOR

Sometimes, it may be convenient to represent object orientation with a rotation vector[1] of the equivalent rotation. Let $\mathbf{u} = [u_1\ u_2\ u_3]^T$ be the unit vector in the direction of the axis of the rotation equivalent to the three gimbal rotations, and let ϕ be the angle of this rotation.

Then, we define the gimbal rotation vector $\mathbf{\Phi}$ as:

$$\mathbf{\Phi} = \phi\mathbf{u} \tag{10}$$

The gimbal rotation matrix $\mathbf{G}$ can be expressed in terms of the gimbal rotation vector as [2]:

$$\mathbf{G} = \begin{bmatrix} u_1^2(1-c)+c & u_1u_2(1-c)-u_3s & u_1u_3(1-c)+u_2s \\ u_1u_2(1-c)+u_3s & u_2^2(1-c)+c & u_2u_3(1-c)-u_1s \\ u_1u_3(1-c)-u_2s & u_2u_3(1-c)+u_1s & u_3^2(1-c)+c \end{bmatrix} \tag{11}$$

where c and s denote $\cos(\phi)$ and $\sin(\phi)$, respectively.

To express the gimbal rotation vector in terms of gimbal angles, we first form vector $\mathbf{v}$:

$$\mathbf{v} = \begin{bmatrix} r_{32} - r_{23} \\ r_{13} - r_{31} \\ r_{21} - r_{12} \end{bmatrix} \tag{12}$$

where r_{jk} denotes the element in the *j*-th row and *k*-th column of the rotation matrix.

Substituting elements of rotation matrix $\mathbf{G}$ in (11)into (12) yields:

$$\mathbf{v} = \begin{bmatrix} 2u_1s \\ 2u_2s \\ 2u_3s \end{bmatrix} = 2\mathbf{u}\sin\phi \tag{13}$$

The vector $\mathbf{v}$ points in the direction of the axis of the equivalent rotation, and its magnitude is equal to $2\sin(\phi)$.

By summing the diagonal elements of the matrix $\mathbf{G}$ in (11), we can also see that:

$$\mathrm{tr}(\mathbf{G}) = 1 + 2\cos(\phi) \tag{14}$$

By substituting the elements of the matrix $\mathbf{G}$ in (9) into (13) and (14), we can express $\mathbf{v}$ and $\mathrm{tr}(\mathbf{G})$ in terms of the gimbal angles γ_1, γ_2, and γ_3:

$$\mathbf{v} = \begin{bmatrix} c_1s_3 + c_2s_3 - c_3s_1s_2 \\ s_2 + c_1c_3s_2 + s_1s_3 \\ c_2s_1 + c_3s_1 - c_1s_2s_3 \end{bmatrix} \tag{15}$$

and

$$\mathrm{tr}(\mathbf{G}) = c_1c_2 + c_1c_3 + c_2c_3 + s_1s_2s_3 \tag{16}$$

where c_n and s_n are defined as in (9).

Considering (13) and (14) we can now write:

[1] A rotation vector is sometimes called the Euler vector or Simultaneous Orthogonal Rotations Angle (SORA) in the literature.

$$\cos\phi = \frac{\mathrm{tr}(\mathbf{G})-1}{2}$$
$$\sin\phi = \frac{\|\mathbf{v}\|}{2} \qquad (17)$$

5. Determination of Gimbal Angles

A. From Rotation Matrix

We can determine the gimbal angles from the rotation matrix by examining elements of the gimbal matrix **G** in (9). Gimbal angle γ_2 can be obtained directly from r_{31} as follows:

$$r_{31} = -\sin(\gamma_2) \qquad (18)$$

The above equation has two solutions:

$$\gamma_{2a} = -\arcsin(r_{31}) \qquad (19)$$

and

$$\gamma_{2b} = \pi + \arcsin(r_{31}) \qquad (20)$$

Gimbal angles γ_1 and γ_3 can then be obtained by combining elements of the matrix **G** in (9), two of them for each angle:

$$r_{11} + i\,r_{21} = \cos(\gamma_2)(\cos\gamma_1 + i\sin\gamma_1) \qquad (21)$$

$$r_{33} + i\,r_{32} = \cos(\gamma_2)(\cos\gamma_3 + i\sin\gamma_3) \qquad (22)$$

We obtain two solutions for γ_1 and γ_3, where the first solution corresponds to γ_{2a}:

$$\gamma_{1a} = \arg\left(\frac{r_{11} + i r_{21}}{\cos(\gamma_{2a})}\right) \qquad (23)$$

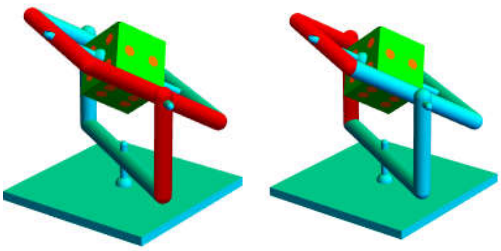
$$\gamma_{3a} = \arg\left(\frac{r_{33} + i r_{32}}{\cos(\gamma_{2a})}\right) \qquad (24)$$

and the second solution corresponds to γ_{2b}:

$$\gamma_{1b} = \arg\left(\frac{r_{11} + i r_{21}}{\cos(\gamma_{2b})}\right) \qquad (25)$$

$$\gamma_{3b} = \arg\left(\frac{r_{33} + i r_{32}}{\cos(\gamma_{2b})}\right) \qquad (26)$$

Note that two different sets of gimbal angles, $\mathbf{G}_a = (\gamma_{1a}, \gamma_{2a}, \gamma_{3a})$ and $\mathbf{G}_b = (\gamma_{1b}, \gamma_{2b}, \gamma_{3b})$ represent the same orientation of the object. In general, it holds that rotations for $\mathbf{\Gamma}_1 = (\gamma_1, \gamma_2, \gamma_3)$ and $\mathbf{\Gamma}_2 = (\pi+\gamma_1, \pi-\gamma_2, \pi+\gamma_3)$ around gimbal axes result in the same final orientation of the object, although the frames of the gimbal are in different positions, as illustrated in Fig. 2.



a b

Figure 2 – a: Gimbal after rotation for the gimbal spatial angle $\mathbf{\Gamma}_1 = (15°, 30°, 60°)$. b: Gimbal after rotation for the gimbal spatial angle $\mathbf{\Gamma}_2 = (195°, 150°, 240°)$.

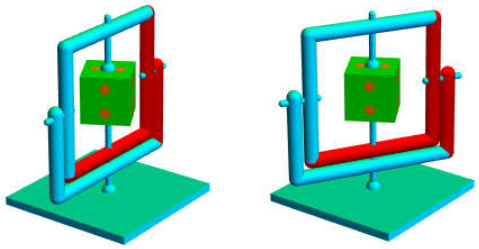
We still have to consider two special cases: $r_{31} = 1$ and $r_{31} = -1$. In both cases, the values of r_{11}, r_{21}, r_{32}, r_{33}, and $\cos(\gamma_2)$ are all equal to zero; thus, the gimbal angles in Equations (23) to (26) have indefinite values. When $r_{31} = 1$ ($\gamma_2 = -90°$), the gimbal rotation matrix **G** in (9) simplifies to:

$$\mathbf{G} = \begin{bmatrix} 0 & \sin(\gamma_1+\gamma_3) & -\cos(\gamma_1+\gamma_3) \\ 0 & \cos(\gamma_1+\gamma_3) & -\sin(\gamma_1+\gamma_3) \\ 1 & 0 & 0 \end{bmatrix} \qquad (27)$$

The sum of γ_1 and γ_3 can be determined from r_{22} and r_{23} as:

$$\gamma_1 + \gamma_3 = \arg(r_{22} - i\,r_{23}) \qquad (28)$$

When $\gamma_2 = -90°$, any pair of angles γ_1 and γ_3 with the same sum yields the same orientation of the object. This result is to be expected because, when $\gamma_2 = -90°$, the gimbal axes g_1 and g_3 are aligned with axis z of the fixed coordinate system; thus, the angles of both of these rotations are simply summed up. Fig. 3 shows gimbal positions for two different spatial angles with $\gamma_2 = -90°$, and $\gamma_1 + \gamma_3 = 60°$.



a b

Figure 3 – Special cases of gimbal angles:
a: $\mathbf{\Gamma} = (0°, -90°, 120°)$, b: $\mathbf{\Gamma} = (-30°, -90°, 150°)$

Similarly, these two axes are aligned when $r_{31} = -1$ ($\gamma_2 = 90°$), however, here, g_1 points in the direction of the axis z and g_2 in the opposite direction. For this reason, γ_3 is subtracted from γ_1. Thus, it holds that:

$$\gamma_1 - \gamma_3 = \arg(r_{22} - i\,r_{23}) \qquad (29)$$

B. From Rotation Vector

Gimbal angles can be expressed in terms of the

axes **u** and angle ϕ of the rotation vector simply by substituting elements of the gimbal matrix **G** into Equations (19), (23), and (24), or into (20), (25), and (26). We obtain:

$$\gamma_{2a} = \arcsin(u_2 \sin\phi - u_1 u_3 (1-\cos\phi))$$
$$\gamma_{1a} = \arg\left(\frac{(u_1^2(1-\cos\phi) + \cos\phi + i(u_1 u_2 (1-\cos\phi) + u_3 \sin\phi)}{\cos(\gamma_{2a})} \right)$$
$$\gamma_{3a} = \arg\left(\frac{u_3^2(1-\cos\phi) + \cos\phi + i(u_2 u_3 (1-\cos\phi) + u_1 \sin\phi)}{\cos(\gamma_{2a})} \right) \quad (30)$$

and

$$\gamma_{2a} = \pi + \gamma_{2a}$$
$$\gamma_{1a} = \pi - \gamma_{1a} \quad (31)$$
$$\gamma_{3a} = \pi + \gamma_{3a}$$

The special cases of $\gamma_2 = \pm 90°$ can be handled in the same way as in the previous section.

6. CONCLUSION

We have shown that rotations around gimbal axes are commutative and that the gimbal spatial angle $\mathbf{\Gamma} = (\gamma_1, \gamma_2, \gamma_3)$ uniquely defines the orientation of an object in 3D Euclidean space. The representation of spatial orientation by gimbal spatial angles can be attractive for object manipulation because rotations around the gimbal axes can be performed in an arbitrary order, therein resulting in the same final object orientation. We have also provided the equations for transformations between different spatial representations, namely, from gimbal spatial angles to the rotation matrix and rotation vector and from a rotation matrix and rotation vector to gimbal spatial angles.

REFERENCES

[1] Stančin, Sara, Tomažič Sašo, "Angle Estimation of Simultaneous Orthogonal Rotations from 3D Gyroscope Measurements," *MDPI*, Sensors, 2011, pp. 8536-8549.

[2] *Wolfram Research*, Mathematica 11.0, 2016.

Sašo Tomažič is a full professor at the University of Ljubljana. He is the head of the Laboratory of Information Technologies and the head of the Department of Information and Communication Technologies. He was an adviser for information and telecommunications system at Ministry of Educational System and, a member of Strategic Council at Ministry of Defiance from, and the national coordinator of research in the field of telecommunications. He has authored and coauthored seven textbooks, ten chapters in research monographs, and more than 200 journal and conference papers. He was leading researcher of 15 R&D projects, and he is the head of research program Algorithms and Optimization Methods in Telecommunications, which is one of two research programs every time named among the best research programs in Slovenia. Since 2006 his interest is in improving society for a better quality of life for all humanity.

Automated Broadcast Video Quality Analysis System

Burnik, Urban; Meža, Marko; and Zaletelj, Janez

Abstract: ***The paper presents an automated system for evaluation of broadcast video quality offered by different delivery service operators. The system has been designed to offer an objective evaluation of services offered by multiple broadcast delivery operators as required by the broadcast content provider. The system substitutes subjective quality monitoring run by human observers with an integrated, objective video quality evaluation system. The presented system fulfilled the requirements of the end user and serves as a tool for objective assessment of video quality.***

Keywords: perceptual video quality, MOS, ITU-T recommendations, broadcast delivery, TV delivery service evaluation

1. INTRODUCTION

WITH the advances in digital television (TV) distribution services, a broadcaster no longer manages the entire signal distribution chain from program production to the end consumer. Momentarily, a range of communication services and channels are being utilized, from satellite, terrestrial and cable digital video broadcast (DVB-x) to Internet protocol TV (IPTV) and mobile services. Typically, TV services are being offered in operator subscriber packages which obtain content from independent broadcast program providers. End users, however, still associate the quality of the service provided by the content provider while most quality issues are affected by the entire distribution chain.

National TV broadcast companies provide signal to service operators in a reference studio quality. It is therefore left to the operators to transcode the material according to specific parameters of their physical distribution network. At this point, the broadcaster loses control over the service quality, with many end users still trying to hold them responsible for the level of service offered.

Manuscript received June 13th, 2017. This work was supported in part by the RTV Slovenia, Contract JN-B0514.

U. Burnik (e-mail: urban.burnik@fe.uni-lj.si), M. Meža (marko.meza@fe.uni-lj.si) and J. Zaletelj (janez.zaletelj@fe.uni-lj.si) are with the University of Ljubljana, Faculty of Electrical Engineering, LUCAMI. Contact author is U. Burnik.

For the reasons mentioned, national broadcasters expressed a requirement for an automated video quality monitoring system. The system should provide a quantifiable measure of video quality which would produce grades close to human perception. The system should monitor how the quality of the signal is affected by the distribution chain of different service package operators in order to address the particular quality issues arising at identified critical locations. Real-time operation of the system was not required. The system introduced throughout the paper offers a broadcaster an automated monitoring facility which allows continuous and on-demand reliable measurement of service quality parameters at selected end user locations.

Evaluation of compressed video service quality is an ongoing challenge, as every modern video coding algorithm utilizes intentional data loss while minimizing perceptual loss arising in such a lossy coding mechanism. Yet, a generalized and reliable objective measure of perceptual quality does not exist. Therefore, a study of existing quality measures has been made and an evaluation system has been proposed that generates a grade of service at a given endpoint, with the objective to produce results comparable to subjective evaluation of video quality.

The paper is organized as follows. First, an overview of video degradation and quality measures is presented. Second, the methods suitable for automated video evaluation are selected. Based on the results, hardware and software components forming a processing chain are proposed capable of video sampling, time interval-based grading, and results archiving. Final conclusions are made based on the practical system behavior.

2. VIDEO QUALITY

Video content in the process of production, storage, delivery, and reproduction is subject to degradation of several origins. Most errors arise from lossy compression, transcoding and communication errors; they visually degrade signal quality and overall quality of experience. Several methods exist that give a quantitative

measure of degradation, however, the numerical value given does not always match the perceived level of quality. First, the typical degradation sources need to be identified, upon which a suitable measure of quality is to be derived.

A. Sources of Degradation

Typical sources of video degradation include the entire communication chain elements starting from production, coding, transfer to presentation and are presented in Figure 1. With respect to the domain of the present paper, most notable are errors due to lossy video coding, with a particular emphasis on video transcoding for distribution optimization, followed by communication channel errors and last, decoding for presentation.

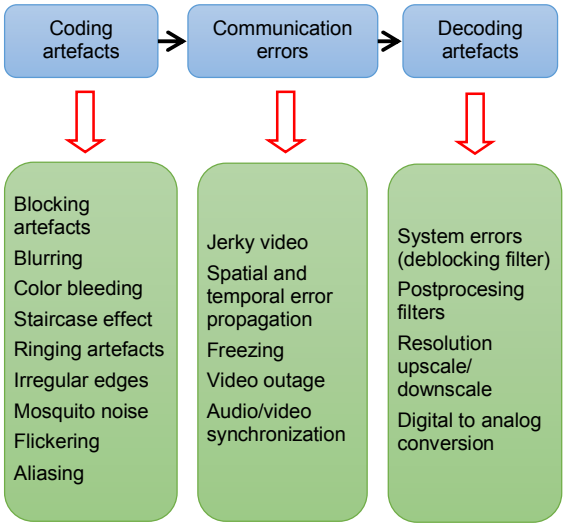


Figure 1: Typical sources of video degradation

Video compression algorithms utilize a common principle of lossy compression of images, based on motion compensation and block transform-based image compression utilizing a (heavy) quantization of coefficients that may lead to visible artefacts [1].

1) Transmission errors

Many errors arise in video transmission over the physical network, be it optical, wired or wireless. Additionally to misdetected bits, other errors arise due to lack of bandwidth, network overload, jitter etc. As compression methods remove redundancy from the encoded video, even a low percentage of errors may lead to notable visual degradation [1]. These errors include video freezing, video jerkiness [2], video blackout, and audio/video synchronization issues.

2) Presentation errors

In addition to errors caused by lossy compression and communication errors, post-processing of image at the receiver side can represent a source of notable signal degradation. These may be caused by digital-to-analog video signal conversion, undersampling of chrominance video component, framerate conversion [3], and de-interlacing. While it is wise to avoid unnecessary conversion steps using a proper configuration setup, many operators and end users ignore this final step and unnecessarily cause additional signal degradation; the reasons vary from bandwidth reduction (justifiable only in case of connection restrictions) to lack of technical knowledge (connection interfaces).

B. Quality Measures

An ideal video quality measure should provide instant real-time grades that match the level of error as perceived by a human observer. Currently, there are no methods fulfilling all of the mentioned requirements. Therefore a method that suits the particular case should always be individually selected.

The evaluation results may be severely affected by the nature of video content. For example, a static video utilizing little motion (e.g. a person talking in front of a camera) may produce acceptable quality at a bit rate, which proves not to be suitable for a dynamic sport coverage. Visual distortions tend to depend on video type and category (e.g. animated film, movie, sports broadcast, or studio shows). Therefore the evaluated video materials should be carefully selected to cover the statistics of several video categories. Standardized test sequences addressing the most typical video categories are produced by international organizations like Society of Motion Picture & Television Engineers (SMPTE), Institute of Electrical and Electronics Engineers (IEEE), and Video Quality Experts Group (VQEG).

1) Subjective methods

Human annotation of quality suffers from several subjective parameters which are not directly connected to video quality. These include individual interests (professional experience in video production, subjective expectations) as well as conditions of evaluation (screen type, environment). Subjective methods still represent the most reliable method for video quality evaluation. Most methods involve grades by multiple users with the final result interpreted on a well-known Mean-Opinion Score (MOS) scale. In reality, these methods are associated with a high evaluation cost as real observers need to be involved working under controlled conditions.

Another disadvantage is the time taken to produce final results; real-time analysis is not possible under any conditions. In modern evaluation, subjective methods often serve as a ground truth of video quality [5] and may present reference values for further evaluation of automated quality evaluation methods.

Based on the requirements given by VQEG, International Telecommunication Union (ITU) proposed a series of standardized methods of subjective video evaluation. In 2002, ITU-R Rec. BT.500 has been issued that addresses TV content. A new recommendation ITU-T Rec. P.910 has been issued in 2008 targeting multimedia content. The recommendations specify test environment and criteria for selection of observers and video content as well as other operational details.

Subjective evaluation can be performed using either double-stimulus or single-stimulus. Using a double-stimulus method, the observers are given video of a reference quality which is later compared to a test sequence; this method is known to be more accurate but also time-consuming. In a single-stimulus method, only test sequences are being evaluated for perceived quality; the method runs faster, therefore more observers can be involved, leading to a statistical enhancement of results.

Several scales are being used for subjective evaluation. A discrete scale 0 – 100 is defined by ITU-T recommendation BT.500 and usually associates with Double Stimulus Continuous Quality Scale (DSCQS) and Single Stimulus Continuous Quality Evaluation (SSCQE) methods. There are scales utilizing 11 (0-10) in 9 (1-9) values defined by ITU-T P.910 for Absolute Category Rating (ACR) method. Most widely deployed is a scale of 5 grades, which is typically used in Double Stimulus Impairment Scale (DSIS) [6], Degradation Category Rating (DCR) [7], and Absolute Category Rating (ACR) method. A discrete 5-point scale utilizes grade 1 as worst and grade 5 as best possible quality in terms of MOS. For more explanation on the listed evaluation methods, a reader is suggested to refer to the cited references.

For the reasons mentioned, modern objective methods of evaluation are being developed that produce results comparable to the subjective evaluation. Such objective methods would allow automated quality monitoring as well as equipment testing and parameter optimization. Recent advances in video technology call for objective standardized evaluation methods.

To automatically monitor video quality using a validated objective method, we suggest deployment of a full-reference evaluation system utilizing a reference and a test signal taken at pre-defined sample points. The reference signal is taken out of the production system before any lossy source channel coding is being deployed. The given uncompressed video stream is later being prepared for final distribution, which optionally includes lossy source channel coding. The distribution service (network) operator ensures network delivery of a signal, utilizing source coding or even transcoding thereof as required. At the users' premises signal gets decoded; this is the best point for test signal acquisition. Finally, the decoded signal is ready to be displayed.

2) Objective methods

There exist a wide range of objective image and video quality measures. Traditional measures are derived from calculation of noise power, where noise refers to an error, or a difference between reference image and the image tested for quality. It is known that these measures demonstrate bad correlation with subjective quality measures. Recent quality metrics utilize models of human vision in order to enhance correlation with subjective measures. Ideally, a good objective video quality metrics would generate a MOS scale matching subjective evaluation of the observed video cases.

Most objective quality measures are based on video content analysis; these may be categorized into data metrics (or energy-based metrics), based on signal values regardless of the underlying content, and visual metrics, utilizing the specifics of visual information.

Most commonly known image quality data metrics are mean squared error (MSE) and peak signal-to-noise ratio (PSNR), popular for its transparent and efficient computation. MSE is given as an average value of a squared difference between pixels in reference and test image [8]:

$$MSE = \frac{1}{n \cdot m} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} \|I(i,j) - \bar{I}(i,j)\|^2$$

where $I(i,j)$ and $\bar{I}(i,j)$denote a pixel value of a reference and test image, respectively. m and n refer to width and height of an image. PSNR [9] technically represents the ratio between maximum power of a signal and noise affecting the image. Its value is defined as

$$PSNR = 20 \cdot log_{10} \left(\frac{I_{max}}{\sqrt{MSE}} \right)$$

where I_{max} indicates a maximum pixel value within the image.

Unfortunately, the named methods do not correlate well with human perception [10]. For same amount of data in error, both methods demonstrate the same value regardless of the surrounding region the error occurs within.

Better results can be obtained if a quality

measure utilizes characteristic of human visual system and image perception. Quality metrics based on human visual system (HVS) utilize a model of HVS, and take into account the perception-affecting factors like color perception, contrast sensitivity, pattern masking and deploys experimentally obtained psychophysical models [11] [12]. Some examples include Visual Differences Predictor (VDP), Sarnoff JND (Sarnoff Just Noticeable Differences), Moving Picture Quality Metric (MPQM), and Perceptual distortion metric (PDM). As input, systems use well aligned reference and test images. In Step 1, a suitable color space is selected for analysis. Next, perceptual decomposition separates signal into more perception channels. Given channels are weighted upon contrast sensitivity. Stage 4 deploys visual masking. Both video sequences are prepared using the same procedure, and finally analyzed for differences. The results are merged into a quality scale value.

Further methods analyze and identify typical artefacts and objects in video sequence. Here, structure image elements (contours, edges …) and characteristic image artefacts (edge ringing, blocking …) are being identified and quantified. Final quality measure combines these values into a common score, taking psychophysical perception of human visual system into consideration [12]. Among these, best known methods are SSIM, VQM, PEVQ, and VQuad-HD.

Some quality evaluation metrics in IPTV directly evaluate the quality of a bitstream for errors. The advantage is no decoding. Their advantage is reduced processing demand and they may evaluate more data streams in parallel. The results are applicable only within comparable coding and transport protocols [13]. In hybrid mechanisms, these methods may complement traditional video quality metrics, utilizing packet statistics, bitstream analysis and decoded video signal to monitor quality of service in IPTV and mobile video networks. Hybrid methods are studied by the new working group »Hybrid Perceptual/Bitstream (HBS) Group« within VQEG.

Regarding the availability of reference information, image quality can be evaluated using a full reference, no reference, or reduced reference principle. It is considered that a reference signal contains no errors. In case the reference image is being present, a full comparison between the reference and test image can be done. The compared images should be perfectly aligned. In case no reference is present, the quality of the image is evaluated for quality standalone, utilizing detection of typical artefacts. Here, a challenge is to correctly identify the artefacts among real video content. There exist a method of reduced reference signal that separately identifies some visual attributes of a reference image and compares them against the matching ones on a test image [5].

C.Experts Groups on Video Quality

VQEG is a leading expert group on video quality evaluation [14]. The group cooperates with International Telecommunication Union groups ITU-T and ITU-R) [15]. VQEG was established in 1997 to recommend suitable subjective grade methods and test sequences in order to compare several video communication systems but has further evolved to a type of video quality assessment. The group has proposed more recommendations for video quality evaluation of which most have been confirmed as ITU standards [14]:

- ITU-T Rec. J.247 (2008) "Objective perceptual multimedia video quality measurement in the presence of a full reference"
- ITU-T Rec. J.246 (2008) "Perceptual visual quality measurement techniques for multimedia services over digital cable television networks in the presence of a reduced bandwidth reference"
- ITU-R Rec. BT.1866, (2010) "Objective perceptual visual quality measurement techniques for broadcasting applications using low definition television in the presence of a full reference signal"
- ITU-R Rec. BT.1867, (2010) "Objective perceptual visual quality measurement techniques for broadcasting applications using low definition television in the presence of a reduced bandwidth reference"
- ITU-T Rec. J.341 (2011), "Objective perceptual multimedia video quality measurement of HDTV for digital cable television in the presence of a full reference"
- ITU-T Rec. J.342 (2011), "Objective multimedia video quality measurement of HDTV for digital cable television in the presence of a reduced reference signal"
- ITU-T Rec. J.249, (2010) "Perceptual video quality measurement techniques for digital cable television in the presence of a reduced reference"
- ITU-T Rec J.340 (2010) "Reference algorithm for computing peak signal to noise ratio (PSNR) of a processed video sequence with constant spatial shifts and a constant delay"

Momentarily VQEG is being active on hybrid metrics, and works on issues of 3DTV, HDR, JEG-Hybrid, MOAVI (Monitoring of Audio Visual Quality by Key Indicators), Multimedia (phase II), QART (Quality Recognition Tasks), and RICE

Parameter	MSE	PSNR	SSIM	VQM	PEVQ	VQuad HD	MPQM	CMPQM	NQM	NVFM
Full-reference	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes
MOS scale results	No	No	No	No	Yes	Yes	Yes	Yes	No	Yes
Computational complexity	1	1	3	5	3	3	3	3	3	3
Correlation with subjective methods	1	1	2	3	5	5	Acc.	Acc.	Acc.	Acc.
Commercial availability	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	No
HVS model	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Standardized				ANSI T1.801 .03-1996	ITU-T J.247	ITU-T J.341				

Table 1: Methods of video quality assessment

(Real-Time Interactive Communications Evaluation).

D.Selection of Evaluation Methodology

Based on the available methods of video quality assessment, we have decided to utilize a system that:

- Produces comparable assessment values in MOS scale, regardless of the assessment category.
- Provides an automated objective measurement system utilizing HVS perception models
- Is compliant to ITU-T standards
- Utilizes a full-reference system for exact results

Table 1 lists most commonly used methods for video quality assessment. Among these, we have favorised methods recommended by VQEG and certified by ITU.

Based on our review of methods, we have identified PEVQ as best possible method for video quality evaluation. It has been listed as one of the four possible assessment mechanisms by Recommendation J.247. The method correlates well with subjective assessments as it deploys a well-designed HVS model. A full reference is required to and the final result is given on a MOS scale 1-5. Validation of PEVQ by ITU covers only SD despite it can provide results for high resolutions. The only method for HDTV validated and listed by ITU is VQuad HD, specified and listed in ITU-T J.341. VQuad HD also correlates well with subjective assessments and is designed purely for HD resolutions of 1080i in 1080p, with an option of oversampling for 720p video.

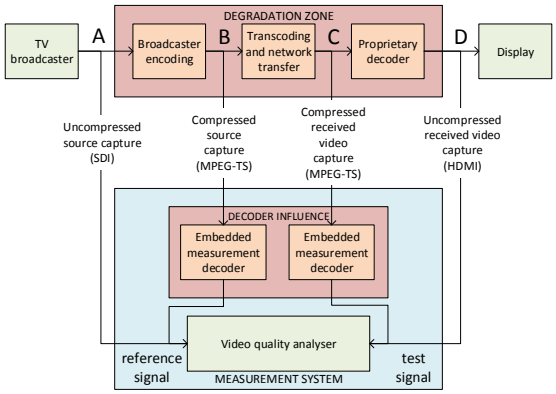


Figure 2: A simplified model of video distribution system indicating potential video acquisition points for quality evaluation

The recommended methods utilize parts of algorithms that are subject to patent protection; therefore only licensed software packages are available for commercial deployment. Using a verified commercial video assessment solution enhances reliability of the proposed system. The assessment modules are being integrated into evaluation software package developed by the authors that collects, selects, interpret, and archive the results automatically.

3. SYSTEM DESIGN

We propose an automated system for video quality analysis, which utilizes commonly recognized objective video quality evaluation methods. The core functionality of system components is to capture signals required for a full reference video quality evaluation system, to transfer the signals to a centralized objective video quality evaluation system and to evaluate the given results.

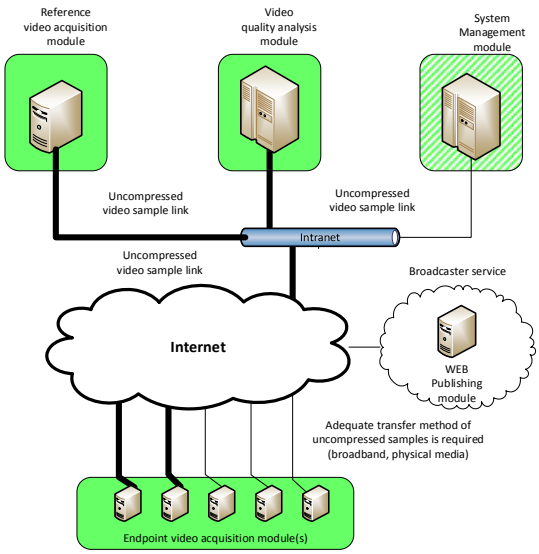


Figure 3: Quality evaluation system indicating key functional modules

A. Video Acquisition Requirements

In order to acquire video signals for full reference quality evaluation, the following acquisition points have been envisioned:

- Point A denotes a presence of uncompressed source signal as produced by the TV broadcaster;
- Point B denotes a potentially degraded user endpoint as induced by lossy compression and network transfer errors in MPEG-TS format, referred to as compressed source encoded video signal;
- Along the communication path the signal traverses the systems of delivery service providers, where additional degradation may occur due to communication errors, transcoding, resolution change and perceptual enhancement mechanisms; at point C it enters the end-user decoding system in MPEG-TS, referred to as compressed received video signal;
- End-user decoder decodes the compressed video signal using a proprietary decoder which produces an uncompressed received signal at point D ready for being displayed.

The signal from its source point traverses a number of alterations to the endpoint user. Some information may get lost and additional processing could be envisioned that affects the perceived end user quality; normally, the perceptual quality is reduced but in some cases it could even be enhanced due to various filtering and other video processing methods.

The full-reference video quality analyzing system requires a pair of uncompressed video signals as shown in Figure 2, namely a reference and a test signal where the latter is being evaluated for quality.

If a signal at the endpoint is available only in compressed form, an embedded video decoder is to be utilized prior to quality evaluation.

The goal of the system is to produce a quantifiable result of video quality as received by the final user; therefore Points C and D represent potential acquisition points of the test signal. For the purpose, the result at the endpoint D would include possible effects of a user decoder, which is usually provided by the distribution service provider and may affect the final video quality. Thus, a total effect of the video distribution system is being put under evaluation. The final selection of end user acquisition points depends on the availability of the testpoints and on evaluation criteria as determined by the evaluators.

B. System Architecture

Based on the user requirements as briefly listed in Introduction, we proposed a distributed video quality evaluation system featuring a set of functional modules supported by dedicated hardware and software equipment. There are no commercially available evaluation systems that would cover all aspects as defined by the user requirements in a single package. On the other hand, the evaluation system should comply with ITU recommendations which rule out a full-custom solution for patent and licensing reasons.

Therefore, an evaluation system has been proposed featuring the following modules:

- Endpoint video acquisition module
- Reference video acquisition module
- Management, archive, and processing module
- Video quality analysis module
- Web publishing module

Commercial modules should be used when required for the reasons of licencing and

optimization of costs.

1) Endpoint video acquisition module

An endpoint video acquisition module features video acquisition hardware, temporary sample storage and a system for signal transfer to a centralized signal archive and quality analysis system, controlled by a custom-made software module. The module is designed to acquire an uncompressed video signal as received by the end users with an option to operate with the end user decoder (set-top box) by mimicking infrared remote control commands. The module is portable and should in most cases be connected to a broadband internet connection of a sufficient capacity to upload the video signal samples offline in a reasonable amount of time. Coverage of other locations is also possible using a mobile internet connection for control and management of the system and utilizing physical media to transfer a vast amount of uncompressed samples. The operation of the modules is managed via HTTP/HTTPS communication protocol with data transfer utilizing FTP/FTPS.

2) Reference video acquisition module

For reference video acquisition, a similar system is used. Reference video is to be captured via SDI interface. The control software is capable of multiple source selection using an SDI matrix, which is controlled via IP terminal socket connection. The system is directly connected to high-bandwidth intranet connection for sample transfer.

3) System management module

The heart of the system is a management module, which controls the system operation, provides sample selection and scheduling and delegates the objective quality assessment tasks if individual samples to a dedicated video quality analysis module. The module is also responsible for data archiving and storage process of individual measurements which can later be statistically interpreted in order to make final assessment of distribution service providers.

4) Video quality analysis module

The video quality analysis module is a crucial part of automated video quality evaluation system. We were looking for a verfied objective quality evaluation method. For accuracy and sensitivity under low-degradation conditions, a full-reference method should be used. The method should produce results comparable to a MOS scale, allowing us to compare results against subjective evaluation. Real-time evaluation was not required by the end user, and we were free to use accurate, computational demanding methods. The system components have been selected upon feature analysis given in Table 1. We avoided costly and time-consuming validation of quality measures by using a method verified and recommended by VQEG and ITU. Our requirements narrowed the possible methods to modules PEVQ and VQuadHD, which are subject to licensing and were obtained commercially.

PEVQ (Perceptual Evaluation of Video Quality) is a software module capable of objective video quality evaluation based on human perception model [16] using a full-reference evaluation model. The system has been developed by OPTICOM [17] and is listed as one of the evaluation methods recommended by ITU in "Objective perceptual multimedia video quality measurement in the presence of a full reference" (J.247). The method produces good correlation with subjective analysis as it deploys a well-designed model of HVS and produces quality results on a MOS scale 1-5. It operates in 4 stages as follows:

- Temporal and spatial alignment and color equalization of sample videos.
- Determination of perceptual difference parameters, including video freezing and jerky video.
- Classification of distortion types based on indicators of perceptual difference
- Integration of distortion parameters into a common MOS score

The system also provides a list of individual quality indicators, which might serve in case source of distortion is to be identified. Despite the fact PEVQ can work ad SD and HD resolutions, its validation by ITU in J.247 covers only standard definition video.

VQuad-HD developed by SwissQual is the only full-reference HD video quality evaluation method validated by VQEG and ITU. The validation is covered by Recommendation "Objective perceptual multimedia video quality measurement of HDTV for digital cable television in the presence of a full reference" (J.341) [18]. The model consists of psychophysical and cognitive elements that mimic human perception. It may detect blocking artefacts, tilting, blurring, jerky video and perceptual difference. The evaluation process takes 6 stages, as follows:

- Pre-processing of video sequence (denoising and undersampling)

- Spatial and temporal video alignment
- Local properties similarity and difference analysis
- Calculation of global spatial distortion parameters
- Calculation of global temporal distortion parameters using movement estimates
- Non-linear integration of distortion parameters into a common MOS score

Figure 5: Remote configuration example of endpoint video acquisition module (part of image)

The method produces the results that are more robust compared to older methods. Apart from MOS, it produces an extensive list of individual distortion parameters for detailed analysis.

The system accepts a low-resolution video input through upscaling, however, through J.341 it is validated only for HD video.

5) Web publishing module

The results of final assessments can be published using a separate web publishing module. The module features graphical and empirical interpretation of key quality factors over an archive of results.

4. System Evaluation

The verification system has been designed and evaluated under laboratory conditions to identify any possible issued in module interaction and performance.

As an endpoint video acquisition module, a personal computer has been utilized using a Blackmagic Intensity PRO capture card. Uncompressed video signal capture could be performed via HDMI connection, providing the given signal is unprotected by means of HDCP. An alternative option is to capture a composite analog signal in case of low-budget set-top boxes not featuring HDMI output. A set-top box is controlled using IR blaster. The capture card was selected as it covers the necessary requirements at an acceptable budget. Its potential drawback is that it cannot capture HDCP-protected video stream which, under the observed case, was not required. Another issue of the capture system was identified in local storage bandwidth, as handling of a sustained datastream of over 200 Mbyte/s is required for Full HD (1080i) resolution. The performance requirements have been met using a desktop PC computer utilizing Intel Core i5 processor, 16 GB RAM, dual Gigabit Ethernet interfaces and a RAID0 dual disk array of a 4TB total capacity. As no end-user interaction is foreseen, the entire management of video

Figure 4: Video quality analysis using a modular PEXQ software package

Figure 6: Management interface with access to measurements results

acquisition module has been done via a central management interface. An example of capture device configuration setup done via centralized console is shown in Figure 4.

The same module configuration has been used for **reference video acquisition**, with an addition of Blackmagic Mini SDI/HDMI Converter and IP control of the existing SDI matrix.

A **video quality analysis module** has been set-up on a Windows PC configured with dual Intel Xeon processors 48 GB RAM, gigabit Ethernet interface, RAID1 dual SCSI disk array holding software and operating system and a RAID6 disk array of a total capacity 8 TB holding a maximum of 1 month temporal sample archive. The video analysis software packages Opticom PEXQ and VQUAD-HD have been installed directly onto the host operating system. The capabilities and results of the video analysis system are illustrated in Figure 5, which demonstrates the capabilities of PEXQ software package accessible via its graphical user interface. The results of MOS video quality estimation with a detailed quality parameter overview for a particular sample are visible in Figure 5, Part 1. Let us remind that for automated video assessment the necessary instruction and result flow have been controlled using API calls from a centralized system management software, developed by the authors.

The entire system is controlled via **management module**, which has been set-up in a virtual server environment hosted by the analysis module. The software has been written to control the entire system operation and to generate average quality scores for the evaluated service providers. Operation and scheduling of synchronized video acquisition can be controlled on multiple remote locations based on valid EPG data for balanced genre distribution. The control of offline sample transfer to a central archive has been fully automated. The video quality assessment over a given sample pair has been initiated automatically once the samples were fully copied to a local archive of the video quality analysis module. Based on multiple measures, the average results for multiple endpoints have been determined. The access to individual measurements and graphical results for selected operators is shown in Figure 6.

The automated dissemination of measurement results was not foreseen in the final implementation. The results could be manually prepared for web publication.

5. CONCLUSION

The paper presents development and configuration of an automated video quality assessment system. The main goals of the design have been reached as the system provides a valuable tool for quality comparison of a broadcast among several communication channels offered by a number of broadcast delivery providers. The system automatically provides validated MOS-correlated results of video quality from 5 simultaneously observed locations. Apart from its main goal, the results given by the system helpfully guided our national TV broadcaster in their decision to complement SD broadcast delivery using own DVB-T network with 2 additional HD channel versions over a single multiplex.

REFERENCES

[1] M. Yuen and H. R. Wu, "A survey of hybrid MC/DPCM/DCT video coding distortions," *Signal Processing,* vol. 70, pp. 247-278, 1998.

[2] T. S. Bonfim, M. M. Carvalho and M. C. Q. Farias, "Video Quality Evaluation for a Digital Television Broadcasting Scenario," in *International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM)*, Phoenix, 2010.

[3] S. Borer, "A model of jerkiness for temporal impairments in video transmission," in *Quality of Multimedia Experience (QoMEX), 2010 Second International Workshop on*, 2010.

[4] S.-H. Han, H.-K. Kim, Y.-H. Lee and S. Yang, "Converting the interlaced 3: 2 pulldown film to the NTSC video without motion artifacts," in *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, 2005.

[5] S. Winkler, "Video quality and beyond," in *Signal Processing Conference, 2007 15th European*, 2007.

[6] I. T. U. R. Assembly, Methodology for the subjective assessment of the quality of television pictures, International Telecommunication Union, 2003.

[7] I. T. U. T. Recommendation, "P. 910,"Subjective video quality assessment methods for multimedia applications,"," *International Telecommunication Union, Tech. Rep,* 2008.

[8] Z. Wang and A. C. Bovik, "Mean squared error: Love it or leave it? A new look at signal fidelity measures," *IEEE signal processing magazine,* vol. 26, pp. 98-117, 2009.

[9] S. Winkler, Digital video quality: vision models and metrics, John Wiley & Sons, 2005.

[10] S. Winkler and P. Mohandas, "The evolution of video quality measurement: From PSNR to hybrid metrics," *IEEE Transactions on Broadcasting,* vol. 54, pp. 660-668, 2008.

[11] Z. Wang and A. C. Bovik, "Modern image quality assessment," *Synthesis Lectures on Image, Video, and Multimedia Processing,* vol. 2, pp. 1-156, 2006.

[12] S. Winkler, "Perceptual Video Quality Metrics - A Review," in *Digital video image quality and perceptual coding*, CRC Press, 2005, pp. 155-179.

[13] S. Winkler, "Video quality measurement standards—Current status and trends," in *Information, Communications and Signal Processing, 2009. ICICS 2009. 7th International Conference on*, 2009.

[14] "Video Quality Experts Group (VQEG)," [Online]. Available: https://www.its.bldrdoc.gov/vqeg/vqeg-home.aspx. [Accessed June 2017].

[15] "VQEG meeting minutes, Krakow, Poland June 28 – July 2, 2010," [Online]. Available: https://www.its.bldrdoc.gov/media/5542/VQEG_Minutes_Krakow_Jun10.pdf. [Accessed June 2017].

[16] P. E. V. Q. OPTICOM, "Advanced perceptual evaluation of video quality," *OPTICOM GmbH, Germany: PEVQ Whitepaper,* 2005.

[17] I. T. U. T. S. Sector, "Objective Perceptual Multimedia Video Quality Measurement in the Presence of a Full Reference," *ITU-T Recommendation J,* vol. 247, 2008.

[18] I. T. U. T. S. Sector, "Objective perceptual multimedia video quality measurement of HDTV for digital cable television in the presence of a full reference," *ITU-T Recommendation J,* vol. 341, 2011.

[19] E. Wyckens, S. Borer and M. Leszczuk, "ITU-T J. 247 Objective perceptual multimedia video quality measurement in the presence of a full reference," *International Telecommunication Union,* p. 108, 2008.

U. Burnik (M'94–SM'17) a senior lecturer at the Faculty of Electrical Engineering, University of Ljubljana. His research is oriented towards audio, image and video processing, multimedia communications and services. He participated in several research projects in the fields of multimedia data management, interactive TV, personalized multimedia services and mobile multimedia devices. He is an active member of national standardization and education boards, and currently chairs IEEE Slovenia Section.

M. Meža, PhD, BSc, received his B. Sc. and Ph. D. degrees in Electrical Engineering from the University of Ljubljana in years 2001 and 2007. He is member of User-adapted communications and ambient intelligence lab, Faculty of Electrical Engineering, University of Ljubljana. His research interests cover various signal processing using machine learning and datamining.

J. Zaletelj is a senior lecturer at the Faculty of Electrical Engineering, University of Ljubljana. He participated in several international research projects funded under EU Framework Programmes, and he served as a workpackage leader in the EU FP6 project »Live Staging of Media Events«. His research interests include personalized interactive TV services, game-based learning, human behaviour analysis and image and video processing.

An Overview of Fiber Fluorimeter Probes

Samir, Ahmed and Batagelj, Bostjan

Abstract: ***Fluorimeter devices incorporate optical fiber probes as an essential component to deliver excitation light to a sample and collect the re-emitted fluorescence to a detector. The small size, light weight, flexibility, and non-toxicity of the optical fiber are attractive advantages that make it possible to monitor spectral signals originating from minute volumes and has the capability of remote monitoring. In this paper we present a series of approximation formulas for light-collection efficiency for different types of optical fiber fluorimeter probes.***

Index Terms: ***collection efficiency, fluorescence, fluorimeter, optical fiber, sensing probe***

1. INTRODUCTION

Fluorimetry is a widely used optical method that can quantitatively measure the fluorescence when monitoring environmental changes in samples for chemical, biomedical, and clinical applications [1], [2]. It characterizes the relationship between the absorbed and the emitted photons at specified wavelengths. Fluorescence phenomena occur when fluorescent molecules absorb photons from the ultraviolet, visible or near-infrared light spectra, the so-called excitation, and then return rapidly (in nanoseconds) to the ground state by emitting photons with a longer wavelength.

The emitted photon can be detected using two methods. In the straightforward method the emission is detected directly via simple coupling optics with a photon detector. This method needs a single photon detector with a wide photon-sensitive area (ideally larger than the photon-emission area), to achieve high-efficiency detection. In the alternative method, the emission is detected via an optical fiber coupled to a photon detector. Fluorescence-based fiber-measurement techniques are more convenient when compared to fluorescence-based, free-beam optics techniques due to their flexibility, cost-effectiveness, small size, and remote-monitoring capability, enabling their simple integration into existing structures.

Manuscript received on May 29th, 2017.

Samir. Ahmed. Author is with the Electrical Engineering Department, University of Ljubljana, Slovenia (e-mail: ahmed.samir@fe.uni-lj.si).

Bostjan. Batageli. Author is with the Electrical Engineering Department, University of Ljubljana, Slovenia (e-mail: bostjan. batagelj@fe.uni-lj.si).

Glass fibers are not only used for telecommunications [3], but also for delivering excitation and emission in biomedical optical spectroscopy [4-9]. Using fiber-optic probes in biomedical optical spectroscopy makes possible a local detection. The photon detector can be far from the sample and photon detectors with a small sensitive area can also be used. In the case of using fiber-optic probes, the photon-emission area is almost identical to the small core cross-section.

The most important parameter when designing a fiber fluorimeter probe is the ratio between optical power that is sent into the fiber probe, and the optical power coming back from the fiber probe, i.e., the fiber fluorimeter probe efficiency. There are many factors that can decrease the efficiency of the fluorescence detection. Some of them are Fresnel loss (a), background florescence, absorption losses (α), emission efficiency (Q), collection efficiency of the emitted fluorescence from the fluorophore to the optical fiber (η), and coupling efficiency between the other optical fiber end and the detector (ε).

The losses due to Fresnel reflection must be taken into account. It is well known that when the incident light passes from one medium to another a fraction of the light is reflected [10]. In general, it depends on the indices of the refraction and the ray angle of incidence. Since the numerical aperture of optical fiber is small, only rays with small incidence angles can be collected and therefore the influence of the angle on the Fresnel loss can be neglected, so we can assume that the light is perpendicular to the interface. In this case when light is propagating from the optical fiber to the fluorescent media the optical power loss is

$$a = 20\log\left|\frac{n_{\text{core}} - n_{\text{media}}}{n_{\text{core}} + n_{\text{media}}}\right| \qquad (1)$$

The same loss occurs when light is traveling from the fluorescent media into the optical fiber - when the re-emitted fluorescence is collected.

The background florescence is due to the auto-fluorescence generated by impurities in the fiber core [8]. The light beam passing through the fiber excites the fiber material and generates a back-propagating fluorescence light, referred to as the fluorescence background signal. In general, the fiber core induces a stronger fluorescence than the fiber cladding due to the doping material. It is

therefore necessary to use a pure-silica core fiber in applications using UV and shorter visible wavelengths.

The absorption losses are due to the primary absorption (where the excitation power is attenuated when traveling in the florescence material) and to the secondary absorption (where the emitted fluorescence is absorbed while travelling back to the fiber probe) [11].

The emission efficiency determines the efficiency of the fluorescence process. It is defined as the ratio of the number of photons emitted to the number of photons absorbed [12].

The collection efficiency is the ratio of the receiver (fiber) to the source (fluorophore molecule) phase space volumes [13]. Furthermore, we can define the collection efficiency as the ratio of the fluorescence power collected by the fiber core to the total fluorescent power emitted by the fluorophore.

Photomultiplier tubes (PMTs) have larger diameter of circular sensitive area (ϕ_{PD}=5.0 mm) than it is the diameter of fiber core. In this case the coupling efficiency is constant, and ideally it is 100 %, therefore, the optical fiber end is directly connected to the attachment of the PMT counting head without additional coupling optics. In the case of using avalanche photo diodes (APDs), the optimization of the fiber core diameter ($2r_c$), the fiber numerical aperture (NA), and a specific design of the coupling optics are necessary. Such optimizations of probe are necessary since the APD sensitive area is small and the coupling efficiency ε decreases as ($2r_c$) or as NA increases.

This paper reviews different types of optical fiber fluorimeter probes. In section 2, we present single-fiber probe in which the excitation and the emission radiation is transmitted. In sections 3 and 4, we present double parallel and collinear fiber probes in which one fiber guides the excitation to the sample and the other collects the emission radiation to the detector. In section 5, we present fiber-optic bundle probes, in which more than one fiber is used for the collection of the light. And, finally, in Section 6 multi-core fiber probes are presented.

2. SINGLE-FIBER PROBE

In fluorescence measurements with a single-fiber probe, the same fiber is used to transmit the excitation radiation to the sample and to guide the collected signal radiation to the detection system. The fluorescence light is emitted from the excited molecule within the excitation cone at 4π steradians [14], but only the fluorescence emitted within a cone similar to the excitation cone will be collected by a fiber. Let us assume a fluorescent molecule and a fiber with a core area of A_c. Out of the total fluorescence emitted by a fluorophore, only the fluorescence emitted within an optical fiber aperture can be collected, as in fig (1). Then the collection efficiency can be calculated by the ratio of power collected by the fiber (P_c) to the total power emitted by the fluorophore (P_s), which is equivalent to the ratio of the emitting source (fluorophore) (Ω_s) solid angle to the receiver (fiber) solid angle (Ω_c), as in eqn (2):

$$\eta = \frac{P_c}{P_s} = \frac{\Omega_c}{\Omega_s} \tag{2}$$

Since the fluorophore molecule emits isotropically. Then the solid angle is:

$$\Omega_s = 4\pi \tag{3}$$

The solid angle inside a cone of half-angle α_{max} can be determined as given in the Appendix:

$$\Omega_c \approx \frac{\pi NA^2}{n_{media}^2} \tag{4}$$

From eqn (2), eqn (3) and eqn (4) the collection efficiency is given by

$$\eta = \frac{NA^2}{4n_{media}^2} \tag{5}$$

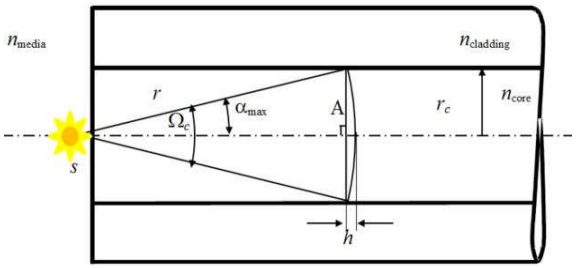


Fig. 1: The emission of light by the fluorophore molecule at the end of the optical fiber.

The collection efficiency calculated using eqn (5) can be used if the fluorophore is at the tip of the fiber, but if the sample has a thickness z then we should consider the attenuation of the travelling emission until it reaches the fiber. Then if P_o is the light power emerging from the fiber core, the excitation power reaching a molecule at a distance z from the fiber is

$$P_e = P_0 e^{(-\alpha_1 \cdot z)} \tag{6}$$

where α_1 is the absorptivity loss in units of m^{-1} and z is the thickness of the fluorescent sample.

If the emission efficiency is Q then the total emitted fluorescence power is

$$Q \cdot P_0 e^{(-\alpha_1 \cdot z)} \tag{7}$$

The fluorescence light at the wavelength λ_2 that is travelling back to the fiber is attenuated by another absorptivity loss α_2. Since the phase volume of the fluorescent source and the phase volume of the fiber can be calculated as follows

$$\Gamma_s = n_{media}^2 A_s \Omega_s \tag{8}$$

$$\Gamma_f = n_{media}^2 A_c \Omega_c \tag{9}$$

where A_s is the normal cross-sectional area of the emitting source at distance z, A_c is the cross-sectional fiber core area, Ω_s is the emitting source (fluorophore) solid angle and Ω_c is the receiver (fiber) solid angle, then the average collection efficiency for a single-fiber probe can be calculated as follows

$$\overline{\eta} = \frac{A_c \Omega_c \int_0^\infty Q \cdot P_0 e^{-\alpha_1 z} e^{-\alpha_2 z} dz}{A_s \Omega_s \int_0^\infty Q P_0 e^{-\alpha_1 z} dz} \tag{10}$$

From fig (2) the ratio of the cross-sectional fiber core area to the normal cross-sectional area of the emitting source at a distance z can be calculated using

$$\frac{A_c}{A_s} = \frac{\pi r_c^2}{\pi r_s^2}\left(1 + \frac{z}{b}\right)^2 \tag{11}$$

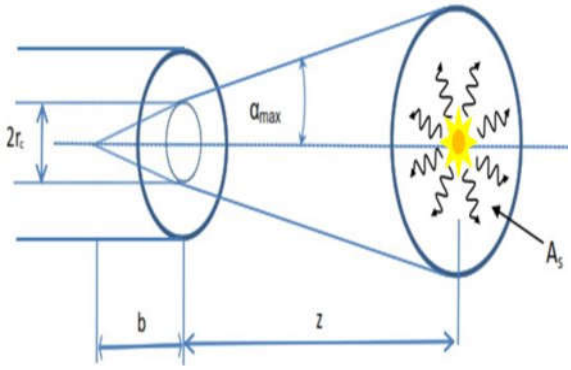


Fig. 2: Collection of fluorescent emitted light from fluorophore molecule at a distance *z* from the fiber tip.

From eqns (3), (4), (10) and (11) the average collection efficiency for a single-fiber probe can be calculated with

$$\overline{\eta} = \frac{NA^2}{4n_{media}^2}\left(\frac{b}{(b+z)}\right)^2 \frac{\int_0^\infty e^{-(\alpha_1+\alpha_2)z} dz}{\int_0^\infty e^{-\alpha_1 cz} dz} \tag{12}$$

The single-fiber probe has the advantages of a small probe diameter, large collection, the smallest possible beam spot size, and of a simple configuration. It is suitable to be used when signal-to-noise ratio is expected to be large and when very small volumes are to be measured [15]. It is limited by the difficulty in suppressing auto-fluorescence background signal induced by the fiber itself, and by the difficulty in reducing back-scattered excitation and illumination light at the fiber coupling site. Therefore, the single-fiber probe is a poor choice for many types of measurements, like measurements of weakly fluorescent samples, measurements requiring long fibers, or when the sample is heterogeneous [16].

3. *Double Parallel Fiber Probe*

In fluorescence measurements with separate fiber probes, one fiber is used to transmit the excitation radiation to the sample and a second is used to collect and guide the emission signal radiation to the detection system. Using separate fibers eliminates the need for fiber splitters, but the probability of capturing emission photons is significantly reduced because only a small portion of the excited fluorescence can be collected. This portion corresponds to the fluorescence emitted in the volume (s) defined by the overlap of the excitation and collection cones, as shown in fig (3). The collection efficiency of a double-fiber probe is therefore lower than the collection efficiency of a single-fiber probe, as the collecting fiber sees only the overlapping area beginning at z_0 [17]. The average collection efficiency for a double-fiber probe can be calculated using:

$$\overline{\eta}_{double} = S(z) * \eta_{\text{sin}gle} \tag{13}$$

$$S(z) = 2\left(R^2\left(\frac{\theta}{2}\right) - \frac{1}{2}(BD)(AE)\right) \tag{14}$$

where S(z) is the overlapping area between the excitation and the collection cone at a distance z from the fiber tip. The overlapping area in fig (3) can be calculated using S(z)= 2*(area of the circular sector (ADB) - area of the triangle(ADB)), which is given by eqn (15)

$$S(z) = 2\left((r_c + z\tan\alpha)^2 \cos^{-1}\left(\frac{(t+r_c)}{(r_c+z\tan\alpha)}\right) - \frac{1}{2}(2((r_c+z\tan\alpha)^2 + (t+r_c))^{\frac{1}{2}}((t+r_c))\right) \tag{15}$$

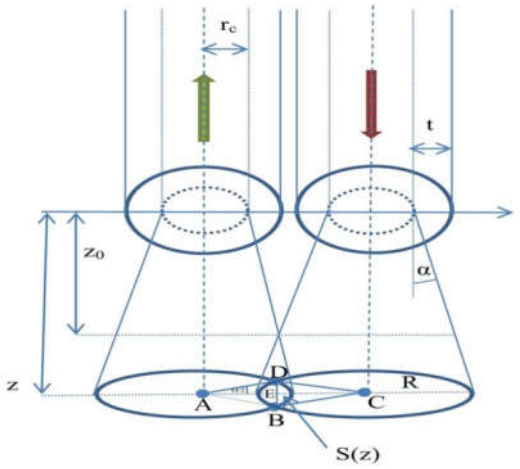


Fig. 3 : Double optical fiber system for calculating the fluorescence signal collected by a double-fiber probe.

4. *Double Collinear Fiber Probe*

In this case the excitation fiber and the collection fiber are collinear. This is assuming that the two fibers have the same radius (r_c), the same numerical aperture (NA) and the longitudinal separation distance (d) as in fig (4).

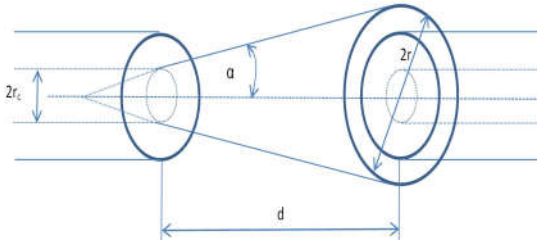


Fig. 4: Double optical fiber facing configuration.

Assuming that the input power to the first fiber is P_{in} and the detected power at the end of other fiber is P_{out}. The collection efficiency for this configuration is given by the ratio between the detected power and the input power

$$\eta = \frac{P_{out}}{P_{in}} = \frac{A_c}{A} = \frac{\pi r_c^2}{\pi r^2} \qquad (16)$$

where A_c and A are the area of the core and the total surface area of the illumination cone at a distance d, as shown in fig (4). The radius of the illuminated area can be calculated as follows

$$r = r_c + d\tan(\alpha) = r_c + d\,(NA) \qquad (17)$$

Then, substituting eqn(15) into eqn(14) we have

$$\eta \approx \frac{r_c^2}{(r_c + dNA)} \qquad (18)$$

Double collinear fiber probe is capable of detecting fluorescence from transparent samples but not from opaque ones. Collecting fluorescence in transmission with this configuration requires a bandpass filter in the excitation path and a long pass filter in front of the collection fiber in order to separate the fluorescence from the excitation light.

5. *Bundle-Fiber Probe*

A single-optic bundle probe including a central excitation fiber surrounded by six collection fibers increases the collection efficiency [18], as seen in fig (5). The effects of fiber parameters like the fiber diameter, the numerical aperture, the arrangement of collection fibers around the excitation fiber and the dead space between them, and the optical properties of the medium on the performance of probes are mathematically modeled by varying the fiber parameters and their geometrical arrangement using a Monte Carlo simulation technique [19]. The simulation results indicate that the fluorescence signal increases monotonically with an increasing collection fiber diameter. However, the inclusion of an effective space factor in the calculations does not yield the same monotonic rise in the collected fluorescence signal. The collection efficiency of a single-optic bundle probe, including a central excitation fiber surrounded by six collection fibers, is calculated in [11] as

$$\eta(r,z) = \begin{cases} \Gamma(r,z)\left(\arctan\left(\frac{r-2r_{cl}-r_c}{z}\right) - \arctan\left(\frac{r-2r_{cl}+r_c}{z}\right)\right), if\ r > 2r_{cl}+r_c \\ \Gamma(r,z)\left(\arctan\left(\frac{2r_{cl}-r_c-r}{z}\right) - \arctan\left(\frac{2r_{cl}+r_c-r}{z}\right)\right), if\ r < 2r_{cl}-r_c \\ \Gamma(r,z)\left(\arctan\left(\frac{2r_{cl}+r_c-r}{z}\right) + \arctan\left(\frac{r-2r_{cl}+r_c}{z}\right)\right), if\ 2r_{cl}-r_c < r < 2r_{cl}+r_c \end{cases} \qquad (19)$$

where Γ(r, z) is given by

$$\Gamma(r,z) = \frac{6}{\pi} arctan\left(\frac{\sqrt{16r_{cl}^2 r^2 - (4r_{cl}^2 - r_c^2 + r^2)^2}}{(4r_{cl}^2 - r_c^2 + r^2)}\right) \qquad (20)$$

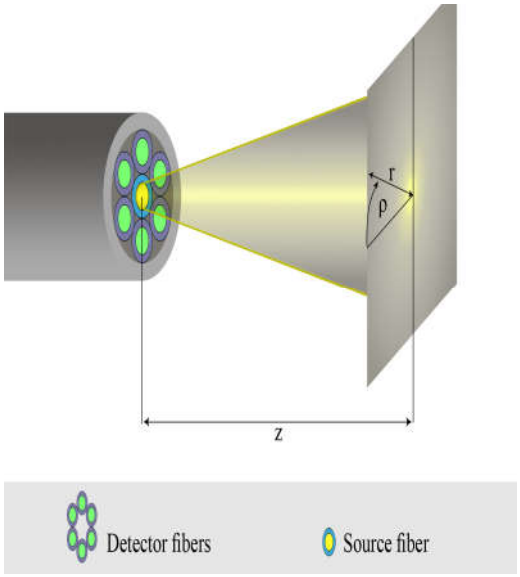


Fig. 5: A six-around-one fiber bundle probe.

Fiber bundles can be used for detecting weak fluorescence samples, but not from strongly absorbing samples. They are also expensive for long cable lengths for remote fluorescence spectroscopy in addition to ineffectively coupling the outputs of several fibers into a detector.

6. *Multi-Core Fiber*

Multi-core fibers (MCFs) are optical fibers, integrating more than one guiding core in one cladding. The overall size of the probe is reduced, because multiple cores can be designed in a fiber with the same width as a single-core fiber. Fiber probes based on a multi-core structure have been proposed to improve the collection efficiency [20], [21]. The core separations throughout the fiber are constant, compared to fiber bundles made by inserting multiple single-core fibers into a capillary.

MCFs integrate the optical paths (excitation path and collection path) within a single MCF probe as in fig (6). The central core is used to transmit and couple the excitation radiation to the outer six cores in order to excite the sample. Using all the cores for delivering the excitation light from the source to the sample reduces the risk of the sample being photo-chemically damaged when compared to the excitation using a single-core fiber. Fluorescence emission feedback radiation at a longer wavelength can be collected in the outer six cores, and then the fluorescence signal can be coupled from these cores to the central core.

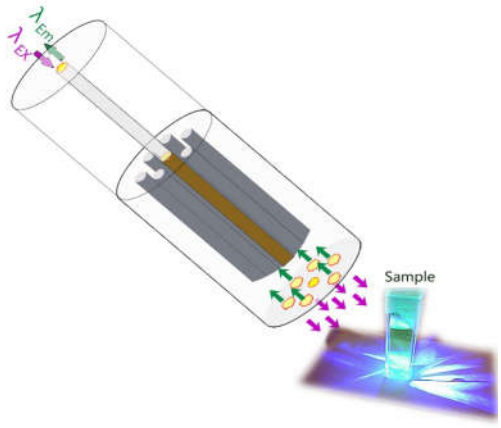


Fig. 6: Seven-core fiber for sending the excitation light to the fluorophore and collecting the emitted fluorescence light.

7. *Conclusion*

There has been a drive in fluorescence spectroscopic applications to push the sample volumes and detection limits to lower levels by designing integrated, multifunctional, and highly optimized fiber probes. Such probes enable optimal excitation and collection, developing highly sensitive detectors and decreasing the background (scattering light, auto fluorescence).

The type of the sample, concentration, sensitivity needed, and physical constraints of the fluorimeter system determine which probe configuration is the best for a particular study.

After the comparison of different fiber probe configurations, we can conclude that single-fiber probe configurations are the best choice for measurements that require short fibers; the signal-to-background ratio is expected to be large, and samples with very small volumes can be measured. However, it is a poor choice for measurements that require long fibers, or for weakly fluorescent samples.

A double parallel fiber probe configuration is capable of sampling over larger volumes and studying discrete luminescent particles in a suspension. Nevertheless, these probes have low fluorescence collection efficiency because they offer poor spatial overlapping between the excitation cone and the fluorescence collection cone.

Collecting fluorescence in transmission using a double collinear fiber probe requires the inclusion of high-quality filtering to separate the fluorescence from the excitation light.

Fiber-bundle probe configurations can be optimized for detecting weak fluorescence samples, but the dead space between the fibers limits the ability to measure the fluorescence from strongly absorbing samples.

The dead space in fiber-bundle probes can be reduced by using MCF probes, which integrate multiple cores in a single cladding.

APPENDIX

The solid angle inside a cone of half-angle α_{max} in fig (1) can be determined as

$$\Omega_c = \frac{A}{r^2}$$

$$= \frac{2\pi rh}{r^2}$$

$$= \frac{2\pi r(r - cos\alpha_{max})}{r^2}$$

$$= 2\pi(1 - cos\alpha_{max})$$

$$= 2\pi(1 - \sqrt{1 - sin^2\alpha_{max}})$$

$$= 2\pi\left(1 - \sqrt{1 - \frac{NA^2}{n_{media}^2}}\right)$$

$$\approx 2\pi\left(1 - 1 + \frac{NA^2}{2n_{media}^2}\right)$$

$$\approx \frac{\pi NA^2}{n_{media}^2}$$

REFERENCES

[1] Lakowicz, J. R., Principles of Fluorescence Spectroscopy, Springer, Third Edition, 2006.

[2] Kohen, E., Hirschberg, J. G., Kohen, C., Schachtschabel, D. O. Stanikunaite, R., Monti, M. "Fourier Interferometry as Applied to Micro spectro fluorimetry and Fluorescence Imaging of Living Cells," Fiber and Integrated Optics,2000, pp.411-425.

[3] Vidmar, M. "Optical-Fiber Communications: Components and System," Informacije MIDEM, 2001, pp.246-251.

[4] Schwab, S. D. and R. L. Mccreery. "Versatile, Efficient Raman sampling with Fiber Optics," Analytical Chemistry, 1984, pp.2199-2204.

[5] Andrade, J. D., R. A. VanWagenen, D. E. Gregonis, K. Newby, and J. N. LIN. "Remote Fiber-Optic Biosensors Based on Evanescent-Excited Fluori-immunoassay: Concept and Progress," IEEE Trans. Electron Devices ED-32: 1985, pp.1175-1179.

[6] Myrick, M. L., S. M. Angel and R. Desiderio, "comparison of some fiber optic configurations for measurement of luminescence and Raman scattering," Applied Optics, 1990, pp.1333-1343.

[7] Kharat, H. J., K. P. Kakde , D. J. Shirale , V. K. Gade , P. D. Gaikwad , P. A. Savale , and M. D. Shirsat, "Designing of Optical Fiber Sensing Probe," Fiber and Integrated Optics, 2006, pp.411-422.

[8] Utzinger, U., and R. R. Richards-Kortum, "Fiber optic probes for biomedical optical spectroscopy," Journal of Biomedical Optics, 2003, pp.121-147.

[9] Jason, T. M., M. Hunter, L. H. Galindo, J. A. Gardecki, J. R. Kramer, R. R. Dasari, and M. S. Feld, "Optical fiber probe for biomedical Raman spectroscopy," Applied Optics, 2004, pp.542-554.

[10] Judy, A. F., and H. E. S. Neysmith, "Reflections from polished single mode fiber ends," Fiber and Integrated Optics, 1988, pp.17-26.

[11] Munzke, D., J. Saunders, H. Omarani, O. Reich, H-P. Loock, "Modeling of fiber-optic fluorescence probes for strongly absorbing samples," Applied Optics, 2012, pp.6343-6351.

[12] Bernard, V., Molecular Fluorescence: Principles and Applications. Wiley-VCH Verlag GmbH. 2001.

[13] Hudson, M. C., "Calculation of the maximum optical coupling efficiency into multimode optical waveguides,", Applied Optics, 1974, pp.1029-1033.

[14] Kormvm, C., and J. S. Schultz, "fiber-optic fluorometer signal enhancement and application to biosensor design," Talanta, 1992, pp.429-441.

[15] Choi, H. Y., S. Y. Ryu, G. H. Kim, K. S. Chang, S. J. Park, and B. H. Lee, "Lensed Dual-Fiber Probe for the Effective Collection of Fluorescence Signals," IEEE photonics technology letters, 2011, pp. 359-361.

[16] Wang, L., H.Y. Choi, Y. Jung, B. Ha Lee, and K-T. Kim, "Optical probe based on double-clad optical fiber for fluorescence spectroscopy," Optics Express, 2007, pp.17681-17689.

[17] Park, W. H., Fluorescence lifetime sensor using optical fiber and optical signal processing, master thesis. University of Toronto. 1998.

[18] Ma, J., Y. Chinforooshan, W. Hao, W. J. Bock, and Z. Y. Wang., "Easily fabricated, robust fiber optic probe for weak fluorescence detection: modeling and initial experimental evaluation," Optics Express, 2012, pp.4805-4811.

[19] Bhowmic, G.K., N. Gautam, L.M. Gantayet., "Design optimization of fiber optic probes for remote fluorescence spectroscopy," Optics Communications, 2009, pp.2676-2684.

[20] Dochow, S. et. al, "multicore fiber with integrated fiber Bragg grating for background- free Raman sensing," Optics Express, 2012, pp.20156-20169.

[21] Samir, A., Batagelj, B., "A multicore fiber probe for fluorescence spectroscopy", Proceedings, 52nd International Conference on Microelectronics, Devices and Materials, MIDEM 2016.

Ahmed Samir graduated in 2006 from the Physics Department, Faculty of Science, Ain Shams University, Egypt, and was then employed in the Physics Department as a Teaching Assistant at the same faculty. He obtained an M.Sc. in Physics in 2012 at Aim Shams University, Egypt, entitled "Application of Laser Speckle Interferometry for Studying Nanoparticles". In 2013 he joined the Radiation and Optics Laboratory, Faculty of Electrical Engineering, University of Ljubljana, as a researcher. His area of research is in multi-core optical fibers for transmission and sensing. In 2017 he received Ph.D. from the University of Ljubljana for work on multi-core fiber fluorimeter probes.

Bostjan Batagelj received his Ph.D. from the University of Ljubljana in 2003 for work on optical-fiber non-linearity measurements. He is currently an assistant professor at the University of Ljubljana. His current research interests are in the areas of next-generation optical access networks, optical-technology-based timing systems and microwave photonics. He has authored or co-authored over 200 technical and scientific publications and is named as an inventor on ten patents.

High-velocity Fluid Impact on Flexible Structures

Irfanoglu, Ayhan

Abstract: ***High-velocity impact on structures has been studied for over half century. Most of the earlier work concentrated on stiff structures impacted by aircraft as safety of nuclear power plants near airfields was the primary focus. Following the attacks on September 11, 2001, several studies on response of flexible structures impacted by aircraft commenced. At Purdue University extensive experimental and numerical simulation-based research projects have been carried out in the last 15 years to build an understanding of response of flexible structural elements (beams, walls) as well as building structural systems (beam-column-slab systems) under high-velocity fluid impact. It has been found that the behavior of simple structural elements under such loads can be estimated reasonably well using engineering expressions or calibrated simulation tools. However, in the case of building structural systems, with World Trade Center Tower I and the Pentagon building as the examples studied extensively, estimating the response to high-velocity fluid impact reliably remains elusive.***

Index Terms: ***fluid impact, high performance computing, high velocity, impact experiments, simulation, structural response***

1. INTRODUCTION

RESPONSE of structures impacted by objects traveling at high speed has been studied for over half-century. Riera [22] proposed an approach to estimate the load on rigid planar elements, modeling the stiff containment shell structure of nuclear power plants, impacted by aircraft, modeled as an elastic-perfectly plastic solid, flying at high speed. Sugano et al. [24] reported results from a full-scale impact experiment in which a military aircraft (F4-Phantom) impacted on a massive rigid target. Since the September 11, 2001 attacks where large civilian aircrafts (Boeing 757 and 767) were crashed into buildings, various research groups have been studying the response of flexible targets (building structural elements or systems) impacted by aircraft flying at high-velocity. Here, experimental and numerical simulation based research on high-velocity fluid impact on flexible structures carried out at Purdue University over the last 15 years are summarized. Focus has been on the fluid-structure interaction as the massive amount of jet fuel present in commercial aircrafts is critical in the impact process (Fig. 1).

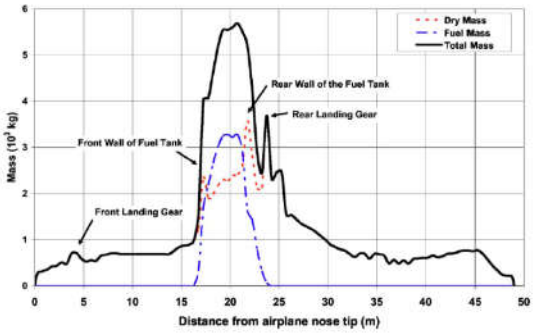


Fig. 1. Distribution of the solid, jet fuel, and total mass along the length of Boeing 767-200ER numerical model used in World Trade Center-I impact simulations at Purdue Un. [4].

2. SIMULATION OF AIRCRAFT IMPACT ON PENTAGON AND WTC-I BUILDINGS

2.1 Pentagon Building Simulation

Following the September 11, 2001 attacks, researchers from Purdue University were involved first in forensic investigation [15] and then computational simulation of the impact loading and response of the Pentagon building [18]. LS-DYNA [12] simulation platform which allows finite element (FE) representation for solid elements and smoothed particle hydrodynamics (SPH) for fluid element was used in the simulations (Figs. 2 and 3).

Output from the Pentagon simulation was processed and rendered into realistic visualizations of the impact and response process [3, 18, 19].

Manuscript received May 28, 2017. Support from the U.S. National Science Foundation (NSF-ITR DSC-0325227), the Scientific and Technological Research Council of Turkey (TUBITAK), and the Kettelhut grant at Purdue University are acknowledged.

Ayhan Irfanoglu is an Associate Professor in Lyles School of Civil Engineering at Purdue University, Indiana, USA. (e-mail: ayhan@purdue.edu).

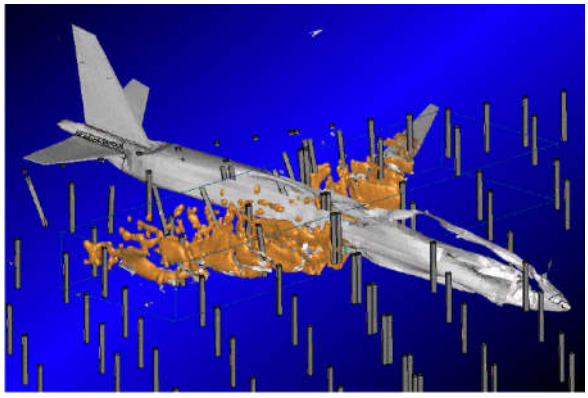


Fig. 2. Simulation of Boeing 757 impacting the reinforced concrete columns in the Pentagon building [17].

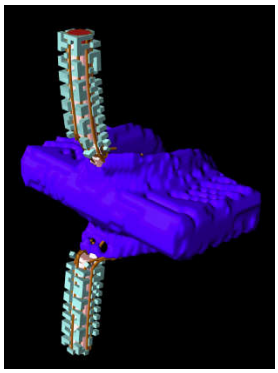


Fig. 3. Finite elements for concrete column and SPH for fluid are used in LS-DYNA impact simulations [18].

2.2 World Trade Center-I (WTC-I) Building Simulation

Extensive simulations of the response of WTC-I to the Boeing 767 impact were carried out at to investigate possible mechanisms that resulted in the total collapse of the building. Top twenty stories of the 110-story building as well as the aircraft were modeled in detail. Model parameters were calibrated so that the simulation-based estimate of damage to the exterior of the WTC-I matched closely with the observed damage (Fig. 4) [4,5]. Engineering simulation results were used to generate realistic visualizations of the impact and response [3,23].

Following the findings in an associated study at Purdue University on behavior of steel structural elements under thermal loads (fire) [13], the study focus was on estimating the damage to the structural system at the core of the building (Fig. 5). It was found that the estimated number of damaged core steel columns were very sensitive to material failure strain value used in the model. Purdue simulations resulted in a best estimate of 23 (out of 47) core columns having lost their ability to transfer loads from the portion of the building above the impacted floors.

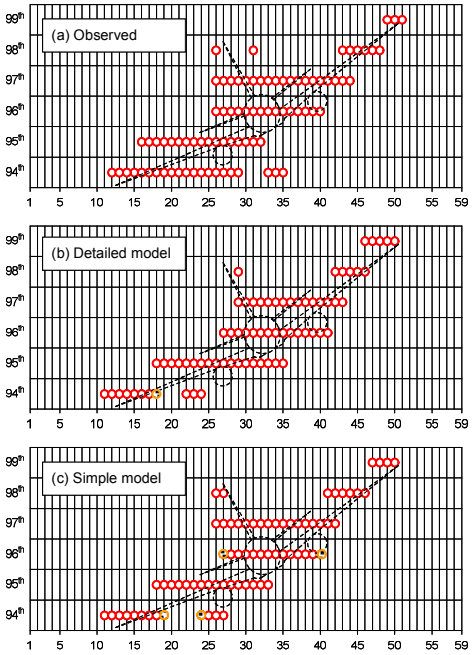


Fig.4.Observed and simulation-based estimated damage to the north (impacted) façade of WTC-I [2,5].

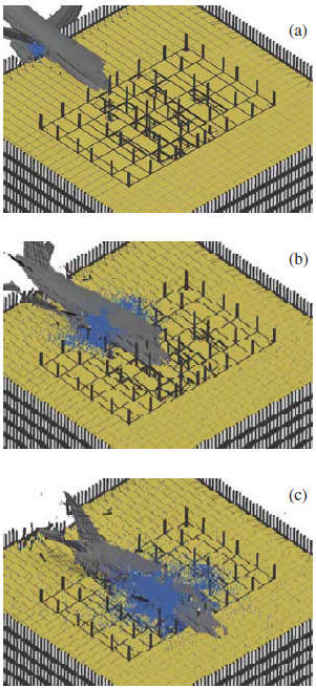


Fig. 5. Airplane and fuel penetrating the core region of the WTC-I in simulation [14].

This amounted to about 1/3 loss in the load carrying capacity. Due to a factor of safety of about 3 [13], however, the core structural system was able to provide sufficient support immediately after the aircraft impact. The story with the worst column damage (with 18 columns destroyed) was the 95th story [5]. Estimates by other researchers for the number of damaged core columns vary

between 7 and 21 [4, 6, 16, 17, 25], indicating presence of large uncertainty about the simulation results. Sophistication in analysis did not result in reduction in uncertainty. Using observations from the Pentagon building forensic investigation [15], through engineering reasoning it was postulated that the core columns in the impact zone had lost their thermal protection. Then, it did not matter whether or not the exposed columns had sustained structural damage; the core structural system would be prone to collapse once the temperature in the steel columns reached 600°C [14]. The decision process was a case of engineering reasoning overcoming irresolvable uncertainties [5].

Later, Brachmann [2] at Purdue University developed an approach for efficient modeling of high-velocity fluid solid impact, which cut the modeling time by up to two factors of magnitude and increased the increased the ratio of quality of results to labor required immensely. Brachmann compared his models and simulation results with data from Sugano F4-Phantom [2] test and from Purdue WTC-I simulations, and found the results to be satisfactory.

Several other fluid-structural element impact simulations have been done at Purdue as part of experimental research. These simulations will be discussed below as necessary.

3. LABORATORY EXPERIMENTS

Different sets of experiments have been carried at the Robert L. and Terry L. Bowen Laboratory for Large-Scale Civil Engineering Research at Purdue University, first, to identify primary parameters in fluid-structure interaction during high-velocity fluid impact and, second, to gather data for calibration and test of numerical simulation software.

During the first set of experiments, in 2007, Pujol carried out eight tests to develop a simple approach to estimate the energy transfer from impacting fluid to a stiff target [18]. Tests involved fluid filled cylinders ("missile") impacting stiff steel plates ("target") at speeds up to 90 m/s (320 km/hr). The fluid mass amounted to 98% of the total mass of the missile. It was found that the kinetic energy acquired by the target was equal to the kinetic energy of missile scaled by the ratio of the mass of the impacting missile and the mass of the target when the target had much higher mass than the missile (ratio of approximately 1:40 and 1:80 were considered). Brachmann [2] used these data to calibrate the LS-DYNA simulation parameters. Following these tests, responses of steel and concrete beams under high-velocity fluid impact were studied.

In 2008, Krahn [9] did a series of experiments in which 140 cm long, 2.5 cm deep slender steel beams (target) were impacted by fluid-filled cylinders (missile) 5 cm in diameter. The missiles were similar to those used by Pujol earlier. 1.25 cm and 1.90 cm wide beams were used to vary the missile-target contact area. Impact velocity range was between 74 m/s (265 km/hr) to 125 m/s (450 km/hr). The goal was to develop a single-degree-of-freedom (SDOF) model subjected to impact load representation of the process to estimate the mid-span peak displacement in the beam [11]. The impact load could be estimated using the velocity of the missile, density of the fluid, the head-on contact area between the missile and the beam. Ignoring the loss of speed in the missile during the impact, Krahn estimated the loading duration and back-calculated a coefficient (assumed to be constant) appearing in the load expression to vary between 0.68 and 0.94, with mean value of 0.74 and standard deviation of 0.07 (i.e., coefficient of variation, COV=0.09). Separately, numerical simulations were performed on LS-DYNA to estimate the mid-span peak displacements. These ratio of the simulation-estimated and measured peak displacements had a mean of 0.99 and a standard deviation of 0.07 (i.e., COV=0.07).

Ardila-Giraldo [1] studied the initial response of beams to high-velocity fluid impact and blast loads. In his experiments, Ardila-Giraldo used 2.5 cm deep prismatic beams, 20 to 30 cm long, with width ranging from 5 to 15 cm. The specimens had symmetric longitudinal reinforcement layout with 0.5% or 1% reinforcement ratio. 350 MPa and 830 MPa steel was used as longitudinal reinforcement. No transverse reinforcement was used. Fluid-filled cylinders, similar to those used by Pujol and Krahn, flying at speeds 47 m/s (170 km/hr) to 147 m/s (530 km/hr) were used to impact the beams at mid-span. The focus of the investigation by Ardila-Giraldo was to explain how the mode of failure for a beam that fails in a ductile manner (flexural failure) under slowly applied load changes to a brittle one (shear failure) when load is applied rapidly. Ardila demonstrated that the switch to shear failure occurs when the shear demand exceeds the shear strength. Calculating the shear strength and the shear demand under dynamic conditions requires accounting for, for the former, the loading rate as the strength of the materials vary with strain rate, and for the latter, the deflected shape during the initial phase of response as both the applied load and the inertial force play a role in shear demand. He concluded that shear strength of the specimens increased by 50% during impact loading.

Rezaei [21] did several additional experiments of high-velocity fluid impact on reinforced concrete beams. His test setup and parameters were similar to those of Ardila-Giraldo's with the

exception that Rezaei included transverse reinforcement in his beams to obtain a wider range of shear strength. Additionally, he compared SDOF model estimates with numerical simulation estimates (based on LS-DYNA) given by other researchers. He used field observations made in the Pentagon building following the September 11, 2001 attack as the testbed. Rezaei found that the ratio of the estimated and observed number of severely damaged columns in the Pentagon building varied between 1.2 and 1.5 when SDOF models were used and between 1.1 and 0.5 when sophisticated numerical simulations were used. He concluded that sophistication in analysis did not increase the accuracy of damage estimates in the Pentagon building.

Korucu [7] tested 32 prismatic reinforced concrete beams under static and high-velocity fluid impact loads. The specimens were 50 cm long beams with rectangular cross-sections (2.5 cm think with impacted face 5 cm wide) and hoop transverse reinforcement and beams with circular cross-sections (4 cm in diameter) and spiral transverse reinforcement. The fluid-filled cylinders were similar to those used in previous studies at Purdue University. The impact speed varied between 65 m/s (235 km/hr) and 208 m/s (750 km/hr). Korucu studied the effect of beam shape and transverse reinforcement configuration on the impact resistance of beams. He found that spiral transverse reinforcement provided better confinement than hoop transverse reinforcement. Spirally reinforced circular cross-section beams performed better than hoop reinforced rectangular cross-section beams, especially at high velocities (Figs. 6 and 8). Korucu also found that transverse reinforcement (tie) spacing had major influence on the failure mechanism as densely installed ties confined core concrete better and increased the strength during impact loads. Additionally, Korucu did simulations on LS-DYNA and tested the fidelity of the numerical simulations by comparing, first, the estimated failure mode with the observed one, and second, the estimated mid-span peak displacement to the observed one. The numerical simulations were successful in estimating the failure mode. In the case of mid-span peak displacement, for rectangular cross-section beams, the ratio of estimated to observed value varied between 0.25 to 0.78, with a mean of 0.53 and standard deviation of 0.17 (i.e., COV=0.32). For the circular cross-section beams the ratio of estimated to observed mid-span peak displacement varied between 1.00 and 1.67, with a mean of 1.29 and standard deviation of 0.18 (i.e., COV=0.14). The inability to obtain from laboratory material tests all of the material properties required in simulations and the use of steel reinforcement material properties per manufacturer's standards were suggested as the main sources for the discrepancy between simulation based displacement estimates and the observed ones.

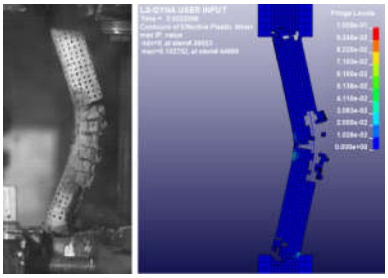


Fig. 6. A beam with rectangular cross-section and hoop transverse reinforcement (1.8% longitudinal and 0.6% transverse reinforcement ratio) after it was impacted by water filled cylindrical missile at 106 m/s. Experimental result compared with numerical simulation estimate.

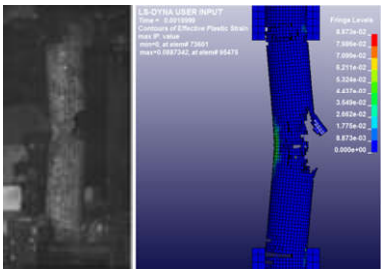


Fig. 7. Beam with circular cross-section and spiral transverse reinforcement (1.8% longitudinal and 0.7% transverse reinforcement ratio) after it was impacted by water filled cylindrical missile at 136 m/s. Test result is compared with estimate from numerical simulation.

Korucu also carried out high-velocity fluid impact tests on concrete plates [8,9]. Concrete plates had mesh reinforcement and between 0% and 3% by volume polypropylene fibers. He observed the effect of fiber ratio on the impact resistance. Twenty tests on 25 cm by 25 cm square, 2.5 cm thick specimens were conducted. Steel mesh reinforcement (reinforcement ratio of 0.43%) was used in the specimens. The polypropylene fiber ratios of 0%, 1%, 2%, and 3%, in volume, were used to form four different batches of specimens. Korucu also estimated the velocity of fluid that would cause scabbing, perforation or total disintegration in a given plate and found that perforation speed increased from 115 m/s (0% fiber ratio) to 135 m/s (1% fiber ratio) to 180 m/s (2% fiber ratio) before decreasing to 145 m/s (3% fiber ratio). He identified mesh reinforcement as very important for holding the overall integrity of the plates while polypropylene fibers preventing disintegration and spalling (i.e., reducing scabbing and formation of high-velocity projectiles on the back face). Korucu found that using numerical simulations, one could estimate the crack patterns in the impacted plates well. Using numerical simulation, he was able to estimate the planar dimensions of the scabbed

area in the rear face of the plates within 10% of the observed ones [9, 10].

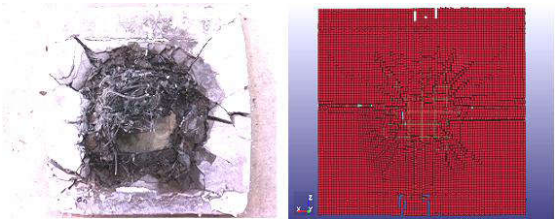


Fig. 8. Rear face of steel-mesh and 1% fiber-reinforced concrete plate impacted by fluid-filled cylinder traveling at 136 m/s. The plate perforated and had scabbing. Simulation result for the rear face, showing the crack patter, is given. [10]

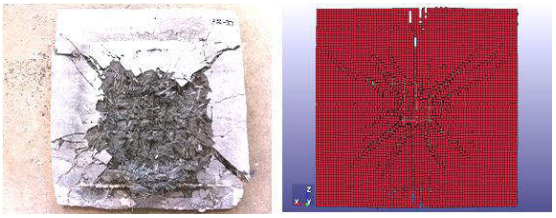


Fig. 9. Rear face of steel-mesh and 2% fiber-reinforced concrete plate impacted by fluid-filled cylinder traveling at 167 m/s. The plate sustained scabbing but did not perforate. Estimated crack pattern on the rear side is given. [10]

4. CONCLUSION

Over the last 15 years, numerous experimental and numerical simulation based investigations on high-velocity fluid impact on structures have been carried out at Purdue University. Always, the principle has been to use empirical data as much as possible, including calibrating the computational tools to the extent possible. Simple and efficient engineering expressions and representations were developed where possible. However, not in all cases empirical data were available and data from elsewhere along with engineering reasoning were used to make estimates or to overcome uncertainties. From the detailed study of the response of the Pentagon building and the World Trade Center Tower I on September 11, 2001, as well as studies on single structural elements subjected to fluid impact, the following conclusions can be reached: 1) sophistication in numerical simulation models do not necessarily reduce the uncertainties in simulation estimates; 2) efficient and simpler models that give results similar to those found using sophisticated models and analyses can be developed; and, 3) in certain cases, engineering reasoning could overcome uncertainties which are unavoidable.

ACKNOWLEDGMENT

The author is grateful for the support given by Prof. Mete Sozen, Distinguished Professor Emeritus of Structural Engineering, Prof. Santiago Pujol, Prof. Christoph Hoffmann, Prof. Voicu Popescu, Prof. Ahmed Sameh, and Prof. Ananth Grama at Purdue University. Dr. Hasan Korucu, formerly a postdoctoral scholar at Purdue University, is recognized for his close collaboration and dedication. Graduate researchers, now all PhD, Drs. Oscar Ardila-Giraldo, Ingo Brachmann, Konstantinos Miamis and Seyed Hamid Changiz Rezaei are recognized for their work and willingness to share their findings. Dr. Paul Rosen, now Prof., was instrumental in performing the WTC-I simulations and visualizations. Mr. Tyler Krahn is recognized for getting the experimental impact test setup ready, along with Prof. Pujol, and conducting the early experiments that opened a path for the rest.

REFERENCES

[1] Ardila-Giraldo, O. A., “Investigation on the initial response of beams to blast and fluid impact”, Doctoral thesis, Purdue University, West Lafayette, IN, 2010.

[2] Brachmann, I., “On efficient modeling of high-velocity aircraft impact”, Doctoral thesis, Purdue University, West Lafayette, IN, 2008.

[3] Hoffmann, C., “9/11 2001 attack simulations using LS-Dyna”, http://www.cs.purdue.edu/homes/cmh/simulation/

[4] Irfanoglu, A. and Hoffmann, C. M., “Engineering perspective of the collapse of WTC-I”, Journal of Performance of Constructed Facilities, 22(1), 2008, pp. 62-67.

[5] Irfanoglu, A., “Using numerical simulations and engineering reasoning under uncertainty: studying the collapse of WTC-1”, Computer-aided Civil and Infrastructure Engineering, 27(1), 2012, pp. 65-76.

[6] Karim, M. R., and Fatt, M. S. H., “Impact of the Boeing 767 aircraft into the World Trade Center.” J. Eng. Mech., 131(10), 2005, pp. 1066–1072.

[7] Korucu, H., Irfanoglu, A., and Sozen, M.A., “Behavior of RC beams under high-velocity fluid impact load”, Seventh Defense Technologies Congress, Ankara, Turkey, 2014.

[8] Korucu, H., "Polypropylene fiber reinforced concrete plates under fluid impact. Part I: experiments", *Techno-Press*, Structural Engineering and Mechanics, 60(2), 2016, pp. 211-223.

[9] Korucu, H., "Polypropylene fiber reinforced concrete plates under fluid impact. Part II: modeling and simulation", *Techno-Press*, Structural Engineering and Mechanics, 60(2), 2016, pp. 225-235.

[10] Krahn, T., “A simple method for determining the forces exerted on a structural element by an impacting liquid body.” Technical Report, Purdue University, West Lafayette, IN, 2008.

[11] Krahn, T. and Pujol, S., “A simple method for determining the forces exerted on a structural element by an impacting liquid body”, International Symposium on Interaction of the Effects of Munitions with structures (ISIEMS) 12.1, Orlando, FL, 2007.

[12] Livermore Software Technology Corporation, LS-DYNA Livermore, CA, USA. http://www.ls-dyna.com/

[13] Miamis, K., “A study of the effects of high temperature on structural steel framing”, Doctoral thesis, Purdue University, West Lafayette, IN, 2007.

[14] Miamis, K., Irfanoglu, A. and Sozen M. A., “Dominant factor in the collapse of WTC-I”, Journal of Performance of Constructed Facilities, 23(4), 2009, 23, pp. 203-208.

[15] Mlakar, P. F., Dusenberry, D. O., Harris, J. R., Haynes, G., Phan, L. T. and Sozen, M. A., “The Pentagon building performance report”, American Society of Civil Engineers, Structural Engineering Institute, Reston, VA, 2003.

[16] National Institute of Standards and Technology, “Final report of the National Construction Safety Team on the

collapses of the World Trade Center Towers", NIST NCSTAR 1, Gaithersburg, MD, 2005.
[17] Omika Y., Fukuzawa E., Koshika N., Morikawa H., and Fukuda R., "Structural responses of World Trade Center under aircraft attacks", J. Structural Engineering, 131(1), 2005, pp. 6-15.
[18] Popescu, V., Hoffmann, C., Kilic, S., Sozen, S., Meador, S., "Producing high-quality visualizations of large-scale simulations", Proc. Visualization 2003, Seattle, WA, 2003.
[19] Popescu, V., and Hoffmann, C., "Fidelity in visualizing large-scale simulations", Computer-Aided Design 37(1), 2005, pp. 99-107.
[20] Pujol, S., "Experimental and analytical study on the response of barriers to fluid impact", Purdue University, West Lafayette, IN, 2008.
[21] Rezaei, S.H.C, "Response of reinforced concrete elements to high-velocity impact load", Doctoral Thesis, Purdue University, West Lafayette, IN, 2011.
[22] Riera, J. D., "On stress analysis of structures subjected to aircraft impact forces", Nuclear Engineering and Design, 8, No.4, 1968, pp. 415-426.
[23] Rosen, P., Popescu, V., Hoffmann, C. and Irfanoglu, A., "A high-quality high-fidelity visualization of the September 11 attack on the World Trade Center", IEEE Transactions on Visualization and Computer Graphics, 14(4), 2008, pp. 937-947.
[24] Sugano, T., Tsubota, H., Kasai, Y., Koshika, N., Orui, S., von Riesemann, W. A., Bickel, D. C. and Parks, M. B., "Full-scale aircraft impact test for evaluation of impact force", Nuclear Engineering and Design, 140(3), 1993, pp. 373-385.
[25] Weidlinger Associates, "World Trade Center structural engineering investigation", ed. Levy M. and Abboud N., Hart-Weidlinger, New York, NY, 2002.

Ayhan Irfanoglu is an Associate Professor in the Lyles School of Civil Engineering a Purdue University, West Lafayette, Indiana, USA. His research interests include impact engineering, large-scale structural dynamic analysis and simulation, and earthquake engineering.

Reviewers:

Australia
Abramov, Vyacheslav; Monash University
Begg, Rezaul; Victoria University
Bem, Derek; University of Western Sydney
Betts, Christopher; Pegacat Computing Pty. Ltd.
Buyya, Rajkumar; The University of Melbourne
Chapman, Judith; Australian University Limited
Chen, Yi-Ping Phoebe; Deakin University
Hammond, Mark; Flinders University
Henman, Paul; University of Queensland
Palmisano, Stephen; University of Wollongong
Ristic, Branko; Science and Technology Organisation
Sajjanhar, Atul; Deakin University
Sidhu, Amandeep; University of Technology, Sydney
Sudweeks, Fay; Murdoch University

Austria
Derntl, Michael; University of Vienna
Hug, Theo; University of Innsbruck
Loidl, Susanne; Johannes Kepler University Linz
Stockinger, Heinz; University of Vienna
Sutter, Matthias; University of Innsbruck

Brazil
Parracho, Annibal; Universidade Federal Fluminense
Traina, Agma; University of Sao Paulo
Traina, Caetano; University of Sao Paulo
Vicari, Rosa; Federal University of Rio Grande

Belgium
Huang, Ping; European Commission

Canada
Fung, Benjamin; Simon Fraser University
Grayson, Paul; York University
Gray, Bette; Alberta Education
Memmi, Daniel; UQAM
Neti, Sangeeta; University of Victoria
Nickull, Duane; Adobe Systems, Inc.
Ollivier-Gooch, Carl; The University of British Columbia
Paulin, Michele; Concordia University
Plaisent, Michel; University of Quebec
Reid, Keith; Ontario Ministry og Agriculture
Shewchenko, Nicholas; Biokinetics and Associates
Steffan, Gregory; University of Toronto
Vandenberghe, Christian; HEC Montreal

Czech Republic
Kala, Zdenek; Brno University of Technology
Korab, Vojtech; Brno University of technology
Lhotska, Lenka; Czech Technical University

Finland
Lahdelma, Risto; University of Turku
Salminen, Pekka; University of Jyvaskyla

France
Cardey, Sylviane; University of Franche-Comte
Klinger, Evelyne; LTCI – ENST, Paris
Roche, Christophe; University of Savoie
Valette, Robert; LAAS - CNRS

Germany
Accorsi, Rafael; University of Freiburg
Glatzer, Wolfgang; Goethe-University
Gradmann, Stefan; Universitat Hamburg
Groll, Andre; University of Siegen
Klamma, Ralf; RWTH Aachen University
Wurtz, Rolf P.; Ruhr-Universitat Bochum

India
Pareek, Deepak; Technology4Development
Scaria, Vinod; Institute of Integrative Biology
Shah, Mugdha; Mansukhlal Svayam

Ireland
Eisenberg, Jacob; University College Dublin

Israel
Feintuch, Uri; Hadassah-Hebrew University

Italy
Badia, Leonardo; IMT Institute for Advanced Studies
Berrittella, Maria; University of Palermo
Carpaneto, Enrico; Politecnico di Torino

Japan
Hattori, Yasunao; Shimane University
Livingston, Paisley; Linghan University
Srinivas, Hari; Global Development Research Center
Obayashi, Shigeru; Institute of Fluid Science, Tohoku University

Netherlands
Mills, Melinda C.; University of Groningen
Pires, Luís Ferreira; University of Twente

New Zealand
Anderson, Tim; Van Der Veer Institute

Portugal
Cardoso, Jorge; University of Madeira
Natividade, Eduardo; Polytechnic Institute of Coimbra
Oliveira, Eugenio; University of Porto

Singapore
Tan, Fock-Lai; Nanyang Technological University

South Korea
Kwon, Wook Hyun; Seoul National University

Spain
Barrera, Juan Pablo Soto; University of Castilla
Gonzalez, Evelio J.; University of La Laguna
Perez, Juan Mendez; Universidad de La Laguna
Royuela, Vicente; Universidad de Barcelona
Vizcaino, Aurora; University of Castilla-La Mancha
Vilarrasa, Clelia Colombo; Open University of Catalonia

Sweden
Johansson, Mats; Royal Institute of Technology

Switzerland
Niinimaki, Marko; Helsinki Institute of Physics
Pletka, Roman; AdNovum Informatik AG
Rizzotti, Sven; University of Basel
Specht, Matthias; University of Zurich

Taiwan
Lin, Hsiung Cheng; Chienkuo Technology University
Shyu, Yuh-Huei; Tamkang University
Sue, Chuan-Ching; National Cheng Kung University

United Kingdom
Ariwa, Ezendu; London Metropolitan University
Biggam, John; Glasgow Caledonian University
Coleman, Shirley; University of Newcastle
Conole, Grainne; University of Southampton
Dorfler, Viktor; Strathclyde University
Engelmann, Dirk; University of London
Eze, Emmanuel; University of Hull
Forrester, John; Stockholm Environment Institute
Jensen, Jens; STFC Rutherford Appleton Laboratory
Kolovos, Dimitrios S.; The University of York
McBurney, Peter; University of Liverpool
Vetta, Atam; Oxford Brookes University
WHYTE, William Stewart; University of Leeds
Xie, Changwen; Wicks and Wilson Limited

USA
Bach, Eric; University of Wisconsin
Bolzendahl, Catherine; University of California
Bussler, Christoph; Cisco Systems, Inc.
Charpentier, Michel; University of New Hampshire
Chong, Stephen; Cornell University
Collison, George; The Concord Consortium
DeWeaver, Eric; University of Wisconsin - Madison
Gans, Eric; University of California
Gill, Sam; San Francisco State University
Hunter, Lynette; University of California Davis
Iceland, John; University of Maryland
Kaplan, Samantha W.; University of Wisconsin
Langou, Julien; The University of Tennessee
Liu, Yuliang; Southern Illinois University Edwardsville
Lok, Benjamin; University of Florida
Minh, Chi Cao; Stanford University
Morrissey, Robert; The University of Chicago
Mui, Lik; Google, Inc
Rizzo, Albert ; University of Southern California
Rosenberg, Jonathan M. ; University of Maryland
Shaffer, Cliff ; Virginia Tech
Sherman, Elaine; Hofstra University
Snyder, David F.; Texas State University
Song, Zhe; University of Iowa
Wei, Chen; Intelligent Automation, Inc.
Yu, Zhiyi; University of California

Welcome to IPSI BgD Conferences and Journals!

http://tir.ipsitransactions.org

http://www.ipsitransactions.org

CIP – Katalogizacija u publikaciji
Narodna biblioteka Srbije, Beograd
ISSN 1820 – 4503
The IPSI BGD Transactions
on Internet Research
COBISS.SR - ID 119127052