PROCEEDINGS

The Ninth International Conference on Business Information Security

Belgrade Metropolitan University

Belgrade, 18th October 2017.

www.metropolitan.ac.rs

Publisher

Belgrade Metropolitan University
Tadeuša Košćuška 63, Belgrade, Serbia
http://www.metropolitan.ac.rs

For Publisher

Prof. dr Dragan Domazet

Editor

Doc. dr Igor Franc
Bojana Trebinjac
Sanja Kovačević

Chair of Organizing Committee

Doc. dr Igor Franc

Conference Secretariat:

Bojana Trebinjac
Sanja Kovačević

Printing

Copy Planet Beograd

Design

Mladen Radić
Petar Cvetković
Katarina Gobeljić

Circulation

60

CONTENT

ORGANIZERS

Belgrade Metropolitan University

Belgrade Metropolitan University
Tadeuša Košćuška 63
11000 Belgrade
Phone: +381 (11) 203 08 85
+381 (69) 203 08 85
Fax: +381 (11) 203 06 28
Email: info@metropolitan.ac.rs

www.metropolitan.ac.rs

Mathematical Institute of the Serbian Academy of Sciences and Arts

Mathematical Institute of the Serbian
Academy of Sciences and Arts
Kneza Mihaila 36
11001 Belgrade, p.p. 367
Republic of Serbia
Phone: 381-11-2630170
Fax: 381-11-2186105
Email: office@mi.sanu.ac.rs

www.esigurnost.org

SECURITY CHALLENGES TO DIGITAL ECOSYSTEMS DYNAMIC TRANSFORMATION

ZLATOGOR MINCHEV
Institute of ICT, Bulgarian Academy of Sciences, zlatogor@bas.bg

Abstract: *The present and future dynamic transformation of digital reality is inevitably establishing a heterogeneous ecosystem of smart gadgets, communication environment and software solutions for meeting modern human factor necessities. The rather rich data context, resulting from this innovative mixture, is practically transcending technological services much closer to human factor and is becoming both smart and challenging, due to multiple security demands that are expected to appear at live in the near future. The paper will outline some recent security trends in digital ecosystems transformation jointly with real practical illustrations for their analytical meeting and prognostic understanding. This will hopefully produce a relevant security support to the expected digital evolution and progressive human factor response in the new, dynamic and fast changing digital world.*

Keywords: *Digital Ecosystems, Security Challenges, Dynamic Transformation, Analytical Support*

1. INTRODUCTION

Digital technologies and fast evolving multiplatform ultra-connectivity have already entered a large majority of our lifestyle areas, to note: education, training, mass media, working environment, everyday life interactions, protesting, friends finding, shopping or even sports, health facilitation and entertainment.

This new change is part of the holistic transformation of the Fourth Digital Revolution [1], [2] that in fact is morphing the modern 21st century people towards new, 'digitised' ones – 'transhumans', having advanced, mixed with technologies: senses, memories, habits, behaviour and emotions [3].

Meanwhile, this dynamic transformation is outlining also another challenge – gradual cultivation of expectations and mind models in the new 'digitised thinking'.

The upcoming digital society is going to produce a rather complex but fully transformed ecosystem of both technologies and humans that will have a new, different joint living & co-existence [4].

Artificial Intelligence (AI) is one of the key factors, progressively extending the capabilities for human-machine interactive feedbacks and perturbations handling. These will probably emerge possibilities for simplified digital ecosystem control and influencing via the machine component but stays quite uncertain from the stability perspective due the huge evolutionary dynamics and constantly progressing scale towards singularity [5].

An important moment, to note here, is the new digital ecosystems multiple objectives that will be difficult for effective prioritizing from AI ethical perspective at the present devastating moment of digital transformation.

The multiple interconnected gadgets' Internet of Things (IoT) concept and numerous innovative cloud services, with big data technology implementation, are also giving an added value to the new digital ecosystem, providing multiple information handlings, strongly embedding at the same time – human factor with machine loop [2], [6]. An arguable moment here is the 'digital privacy' that still stays quite uncertain, regarding simultaneous sources information fusion and protection.

Thus, from one hand, the Fourth Digital Revolution is expected to create new opportunities and from another – to emerge new threats and risks towards the future digital ecosystem evolution.

The proper meeting of these digital progress security demands is a quite challenging task that requires an 'expert – technologies' joint analytical prognostic effort, combined with an appropriate validation & verification of the new digital ecosystems dynamic transformation.

Further in the paper, an exploration framework of this digital phenomena securing with some practical illustrations is outlined.

2. EXPLORATION FRAMEWORK

In brief, the presented framework (see Figure 1) takes a fourfold human-machine realisation of the scenario method application for future planning, implementing a matrix based 'plausible future' security context. It combines both – morphological and system analysis, keeping the holistic nature of the studied problem area for selected scenario families with common objectives on future security exploration [7].

Figure 1: Exploration framework for digital ecosystems future challenges securing

Next, trends forecasting of these scenario families is benefiting from an approach of complex dynamic system discrete representation [8], extensively disposed over the social cycles evolutionary paradigm [9], [10] in a mixed human-machine environment.

Additionally, due to multiple possibilities of the trends development a stochastic machine cyclic validation is added, extending the ideas in probabilistic sense [8].
Being a rather challenging and futuristic area of research, the obtained results for digital ecosystems securing – final verification is proposed with human factor interactive gaming simulation and results discussion.
More details on the proposed exploration framework four stages will be given further in the paper.

Context Matrix Establishing

The plausible future context definition is organized, implementing a matrix representation, combining wide survey data [11], [12], experts and users of different ages opinions, gathered during IFIP TC13 & TC14 Open Symposium [13] and 'Advanced Challenges on Digital Future Securing Training' of SRS' 2017 [14] with 2021 & 2027 time horizon. Final results generalization is conducted with morphological analysis in I-SCIP-MA environment [7]. The methodology is mainly used for solving unstructured and not quite certain big data clusterisation problems. Briefly, the approach is using a multidimensional, mutually exclusive alternatives definition over a cross-consistency multiple scenario matrix space. Resulting positive, neutral and negative scenarios (unordered and not-quantified in time) chains are produced, following an expert cumulative alternatives' interrelations assessment, using Relative Common Weight – RCW and coping the input uncertainties with fuzzy sets extensive implementation [15].
In the present study a five-dimensional matrix is established ('Tech Excesses', 'Social Issues', 'Interfacing Cycles', 'Mixed Resources', 'Security Challenges'), including different alternatives (cells) number (between 2 and 4) among it.
The total cross-consistency plausible future scenario matrix (see Figure 2) combinations number could be calculated as: N = 5 x 4 x 4 x 2 x 4 x 4, N = 2560. The morphological analysis support, selected 128 of them, defining: 111 active (RCW >0), 6 neutral (RCW=0) and 11 passive (RCW<0) scenario chains.

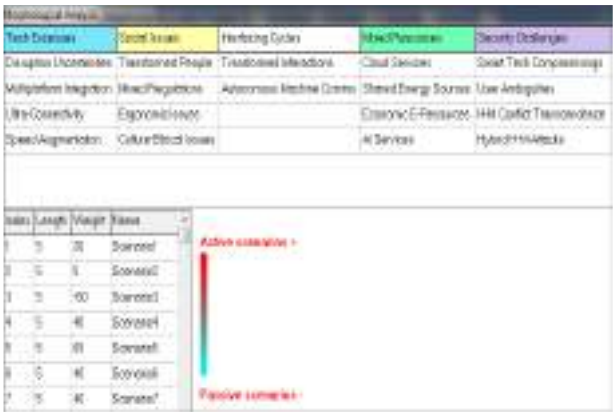

Figure 2: Context matrix screenshot for digital ecosystems securing in I-SCIP-MA environment

The summarized results from the performed analysis are defining notable passive (intangible) scenario future expectations towards security challenges like: 'H-M Conflict Transcendence', 'User Ambiguities' & 'Hybrid H-M Attacks', originating technologically from: 'AI Services' & 'Disruptive Uncertainties', that are socially influenced by: 'Transformed People' and 'Culture-Ethical Issues'. At the same time, the implementation of 'Economic E-Resources' will stay neutral as a source of possible unexpected threats, but still uncertain in the near future.
The rest of the scenario chains are expected to be tangible, which is a serious address towards securing of: 'Smart Tech Compromising' and technological exceeds related to: 'Multiplatform Integration', 'Ultra-Connectivity' & 'Speed Augmentation'.
Being somewhat aggregated and not quite certain, the established scenario matrix context is further studied, implementing risk analysis exploration for a selected scenario chains families, trying to understand and assess their possible interrelations in a system risk context.

Scenarios Holistic Risk Assessment

Proper understanding the complexity of this stage is a rather challenging task, as the objective here is to create system-of-systems interpretation [16], using the context matrix results. This provides a holistic view on the studied problem at hand, giving at the same time, tangible and intangible future risk assessment, taking the ideas of risk complex nature [17].
A suitable approach for solving this complex objective is performed with system modelling & analysis for selected labelled scenario families from the structural analysis matrix, using 'entity-relationship' representation of the idea (see Figure 3a) and defining risks expectations in I-SCIP-RA environment [18].

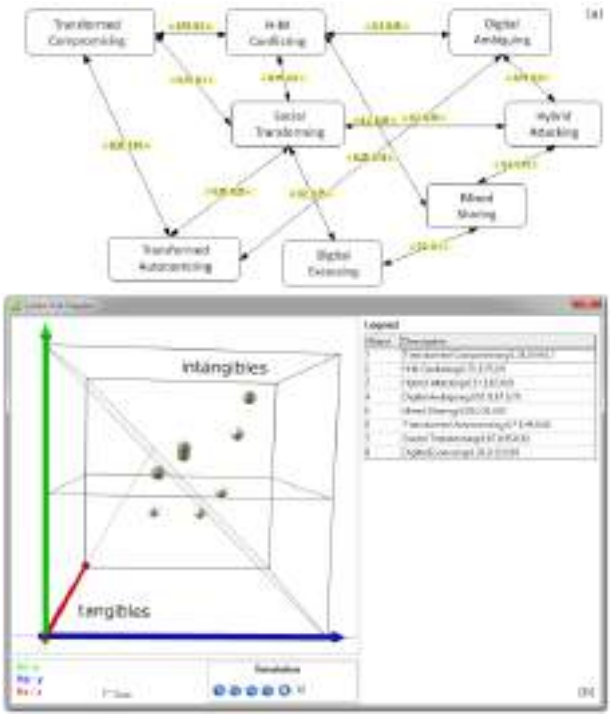

Figure 3: Digital ecosystems future tangibles & intangibles security risks' system modelling (a) and resulting assessment (b) in I-SCIP-RA environment

Further, obtained results are visualised in *3D System Risk Diagram* (see Figure 3b), noting the system risk – *Rs* dual representation (following a causality system modelling assumption of forward – *Rf* & backward – *Rb* risk values)

and finally – identifying tangibles and intangibles model entities (being either active – white or passive – grey) with a probabilistic a priori expert views.
The resulting holistic risk assessment (towards year 2027) of the aggregated eight scenario families (model entities) are giving the following a priori, static classification: tangible (non-critical) security risks are expected to emerge from: 'Mixed Sharing' – 5, 'Transformed Compromising' – 6, 'Digital Excessing' – 8 scenario families. Most of the identified entities in the model are however noted as intangibles (critical): 'Social Transforming' – 7, 'H-M Conflicting' – 2, 'Digital Ambiguing' – 4, 'Hybrid Attacking' – 3 & 'Transformed Autonomizing' – 6.
These results are opening clear, near future, horizon for active AI developments and innovative cloud services digital disruption and multiplatform integration. At the same time, the human-machine expected conflicts and hybrid attacks (implementing AI active role and IoT concept integration) are giving ambiguous (passive) security perspectives to these digital ecosystem elements certainty handling.
In the next section a dynamic outlook towards the presented ideas of system-of-systems future risk assessments validation will be given.

Trends Dynamic Validation

Proper exploration of future assessment trends is an arguable field of work due to the problematic results error evaluation. Being discrete by nature the modern digital world mixing with the continuous social component, is naturally producing a new type of mixed ecosystem.
The resulting, transformed system behaviour is expected to evolve in an innovative way that should vague the limits between live and artificial matter, keeping live matter properties and objectives (to note: self-existence, autonomy, self-regulation towards equilibrium, etc.). What however could be implemented here, regarding the future trends validation is the overall cyclic (pseudo-periodic) nature of this new holistic socio-digital mix co-existence. In this sense, it is important to note the assumptions of Kondratiev [9] (noting the fourfold cycling of evolutionary social progress: prosperity, recession, depression & improvement) that could be enriched from the Forester [10] results (concerning system of interest transitions growth and behaviour – being positive, negative and aiming equilibrium). These could be further extended both in multiple degrees of freedom system-of-systems risk – *Rs* dynamics context, using probabilistic piecewise approximation and adopting risk Beta distribution (representing a probabilistic distribution of probabilities) trends usage for both a priori & a posteriori assessment (working with unknown a priori probability and having reasonable assumptions for this) [19] but with real world entities' interaction epochs cyclic machine simulations (see Figure 4).
Coping this problematic in fact is facing the system of interest detailed interrelations (connections) exploration in a narrow modelling world. Usually, a transfer function is normally assumed between system's inputs and outputs that could be studied from different perspectives [20]. The machine discretisation however also is generating uncertainties and non-stationaries [21] that are difficult to be solved from system singularity and multidimensional perspectives.

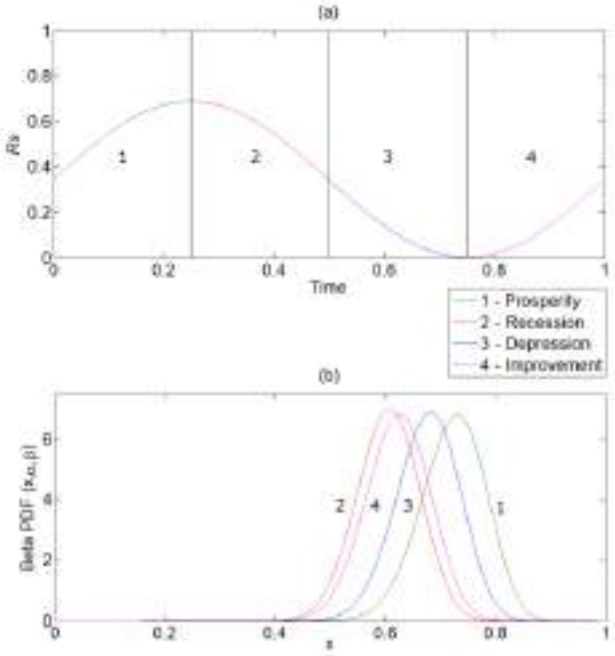

Figure 4: *Rs* a priori trends cyclic dynamics, after [9] (a) & resulting probabilistic a posteriori validation (b), implementing Beta PDF approach [17]

So, a problem oriented modelling approach instead of just simple transfer function (following system-of-systems idea and human-machine active support) could be assumed (see e.g. [7], [8]), taking both aggregated and detailed exploration of the digital future securing problem.
Finally, these trends dynamics concerns are also studied with the assumption of being periodic by nature, regarding system-of-systems interrelating.
An example, concerning the system risk model (see Figure 4a) for 'Transformed Autonomizing' *Rs* relations (see Figure 5) is provided below.

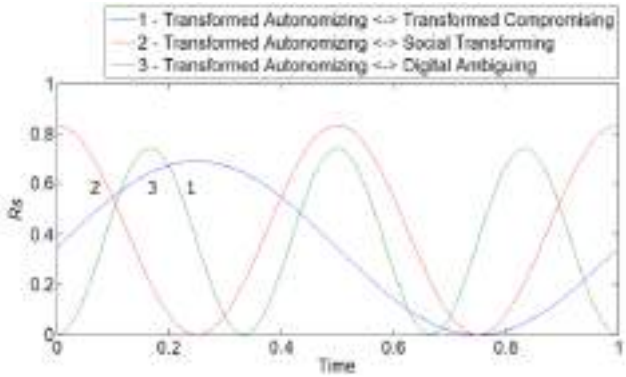

Figure 5: A probabilistic validation example for 'Transformed Autonomizing' a priori *Rs* values, concerning the dynamic evolution up to year 2027

Being somewhat limited from the modelling context and lacking a possibility for comprehensive unexpected future events handling (e.g. non-uniform periodicity changes of *Rs* phase, currently marked aggregated in the implemented probabilities) and multiple degrees of freedom computational expensive realisation (to note the Dirichlet multidimensional case [18] instead of the Beta ones [8]) the

proposed validation approach results are further tested with interactive human-in-the-loop verification.

Interactive Results Verification

The idea of this final stage is to provide experimental observations in a realistic, artificial, futuristic environment of real disruptive technologies users' live responses. The approach has been successfully implemented over the recent years as a future digital threats, risk & challenges verification procedure [8], [17].

CYREX 2017 [22] was an international training event, conducted for the third time in Bulgaria, and hosted by Plovdiv University 'Paisii Hilendarski' as part of the training course 'Security Foundations in Cyber Space' [23]. The exercise took 180 minutes, exploring a hybrid simulation [8], encompassing: industrial espionage, social engineering, malware and multiple targeted attacks. The event gave practical verification results towards near future cyberthreats and challenges user responses, related to transformed reality and IoT multiple integrations phenomena expectations.

Figure 6: Selected moments of *CYREX 2017* [22]

A futuristic scenario within year 2045, implementing disruptive entertainment VR gadget development from a start-up company that is trying to be controlled from both hacktivist group and a multinational corporation for providing bio-connectivity between both users and machines is studied. Additional NGO and public bodies' roles were added for the theatre of simulation comprehensiveness. Trainees have to use in real a lot of smart devices: phablets, tablets, smart watches and bands. Regular desktop and mobile computers were also included. Supportive open cloud services, privately hosted e-mail accounts, chat services, multimedia data, avatar messaging, encryption, brute-force decryption and QR codes were gathered within a closed social network group. The participants accessed is secured via a VPN network (with cable & wireless access) in order to experimentally create and study an ad-hoc transformed digital reality that, together with the participants, established a kind of innovative digital ecosystem.

Players' activities were monitored according to the scenario event script, following response time delays statistics, similar to [12] and some video supportive recording. Selected participants were further equipped with brain activity mobile recorders and smart QR stickers for temperature and galvanic skin response monitoring that provided more detailed physiological supportive information to their scenario responses.

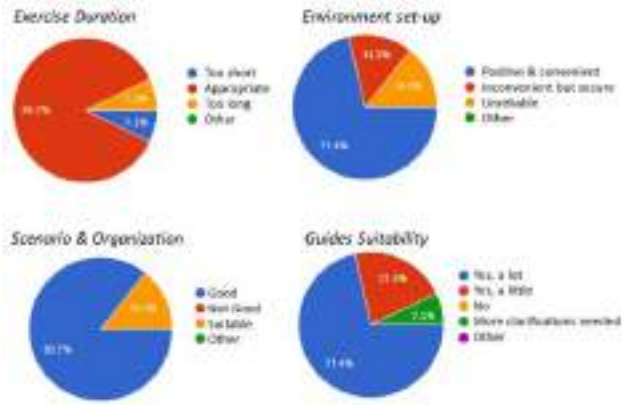

Figure 7: CYREX 2017 participants generalized feedbacks on organizational, content & tech issues

A final questionnaire e-survey in *Google Docs* among the participants is presented (see Figure 7) in order to guarantee comprehensiveness for the exercise evaluation process.

An overall positive assessment (above 85 %) is given to the organizational, duration & scenario content issues of *CYREX 2017*. Even rather interesting the training transformed reality and guidelines were found somewhat ambiguous for the users, though also widely accepted (more then 70%).

Possible improvements of these issues could be further given with a longer training duration and short education, concerning the embedded innovative technological apps and smart gadgets models, utilised with the different cyberattacks and monitoring handlings.

These results are however quite valuable from the digital disruptions fast adaptation perspective of the modern people in the new, futuristic mixed ecosystem of living.

3. DISCUSSION

Future disruptive mix of live and artificial, smart, highly integrated and ultra-connected matter will inevitably produce a new digital ecosystem with autonomous and advanced super-abilities towards the digitally transformed objectives of existence and evolution.

Exploring this phenomenon from the security perspectives is quite challenging and fascinating, especially in the futuristic plausible & non-plausible context.

The outlined framework, implementing 'system-of-systems' holistic exploration though rather comprehensive, could be further extended in three ways: (i) establishing a broader expert discussion forum for further security challenges exploration; (ii) implementing high-performance computational power cloud service in the validation cyclic detailed exploration; (iii) adding highly-integrated monitoring and behaviour stimulating bioimplants.

The extended work in this context is already in active progress with the new 'Securing Digital Future' initiative of Joint Training Simulation & Analysis Center, web portal development: www.securedfuture21.net, encompassing more than 60 experts throughout the world.

4. REFERENCES

[1] L. Floridi, “The Fourth Revolution (How the Infosphere is Reshaping Human Reality)”, 1st ed., Oxford University Press, 2014

[2] K. Schwab, “The Fourth Industrial Revolution: What It Means, How to Respond”, World Economic Forum, May 9, 2017, https://goo.gl/e1Kc3F

[3] R. Sirius & J. Cornell, “Transcendence: The Disinformation Encyclopedia of Transhumanism and the Singularity”, Disinformation Books, 2015

[4] T. Blanke, “Digital Asset Ecosystems: Rethinking Crowds and Cloud”, Elsevier, Chandos Publishing, 2014

[5] M. Shanahan, “The Technological Singularity”, MIT Press, 2015

[6] R. Buyya, & V. Dastjerdi, “Internet of Things: Principles and Paradigms”, Morgan Kaufman-Elsevier, 2016

[7] Z. Minchev, “Human Factor Role for Cyber Threats Resilience”, in Handbook of Research on Civil Society and National Security in the Era of Cyber Warfare, 1sted., M. Hadji-Janev & M. Bogdanoski, Eds., IGI Global, 2015, pp. 377–402

[8] Z. Minchev, “Cyber Threats Identification in the Evolving Digital Reality”, in Proc. of Ninth National Conference “Education and Research in the Information Society”, Plovdiv, Bulgaria, May 26-27, 2016, pp. 011–022

[9] K. Dark, “The Waves of Time: Long-Term Change and International Relations (History and Politics of the 20th Century)”, Bloomsbury Academic, 2016.

[10] D. Meadows, J. Randers & D. Meadows, “Limits to Growth: The 30-Year Update”, Chelsea Green Publishing Company, 2004.

[11] Z. Minchev, G. Dukov, “Emerging Hybrid Threats Modelling & Exploration in the New Mixed Cyber-Physical Reality”, in Proc. of BISEC 2016, Belgrade Metropolitan University, October 15, 2016, pp. 13–17

[12] Z. Minchev, L. Boyanov, “Predictive Identification Approach for Emerging IoT Hybrid Threats”, in Proc. of ICAICTSEE – 2016, Sofia, UNWE, December 2-3, 2016 (in press)

[13] IFIP TC13 & TC14 Open Symposium 2017 Web Page, http://ifip-tc13.org/ifip-tc13-tc14-open-symposium-2017-march-22th-2017/

[14] Advanced Challenges on Digital Future Securing, SRS’ 2017, https://goo.gl/1KKm1F

[15] Z. Minchev, L. Boyanov, & S. Georgiev. “Security of Future Smart Homes. Cyber-Physical Threats Identification Perspectives”, in Proc. of National Conference with International Participation in Realization of EU HOME/2010/CIPS/AG/019 project, Sofia, Bulgaria, June 4, 2013, pp. 165–169

[16] F. Vester, “The Art of Interconnected Thinking – Ideas and Tools for Dealing with Complexity”, München, MCB – Verlag, 2007

[17] Z. Minchev, “Analytical Challenges to Modern Digital Transformation”, in Proc. of Tenth National Conference “Education and Research in the Information Society”, Plovdiv, Bulgaria, June 22-23, 2017, pp. 38–47

[18] Z. Minchev, G. Dukov, D. Boyadzhiev & P. Mateev, “Future Cyber Attacks Modelling & Forecasting”, in ESGI 120 Problems & Final Reports Book, Fastumprint, 2017, pp.77-86

[19] A. Gupta & S. Nadarajah, “Handbook of Beta Distribution and Its Applications”, 1st ed., CRC Press, 2004

[20] S. Bhattacharya, “Control Systems Engineering”, 2nd ed., Dorling Kindersley, 2008

[21] G. Ossimitz, & M. Mrotzek, “The Basics of System Dynamics: Discrete vs. Continuous Modelling of Time”, in Proc. of 26-th International System Dynamics Conference, Athens, Greece, July 20-24, 2008, https://www.systemdynamics.org/conferences/2008/proceed/papers/OSSIM407.pdf

[22] Cyber Research Exercise – CYREX 2017 Web Page, https://goo.gl/sBvVWW197

[23] Z. Minchev, “Security Foundations in Cyber Space”, PU “Paisii Hilendarski” Training Course Selected Materials, http://dox.bg/files/dw?a=f42e63cffd

BLOCKCHAIN/CRYPTOCURRENCIES AND CYBERSECURITY, THREATS AND OPPORTUNITIES

ALEKSANDAR MATANOVIĆ
Master in Digital Currency, Founder of ecd.rs – Serbian Cryptocurrency Exchange, alex@ecd.rs

Abstract: *The document gives the short introduction of cryptocurrencies and blockchain technology and then aims to give an overview of how it relates to some of the cybersecurity challenges. The role of bitcoin in the recent ransomware plague is analysed, having in mind that it has been the currency of choice in most of the ransomware attacks that took place in the last couple of years. It is followed by the brief description of what valuable lessons we might learn from cryptocurrencies and apply elsewhere. Finally, the document analyses the potential of applying the blockchain technology for data storage, the benefits it can bring and the limitations it has, with examples of some projects focused specifically on that area.*

Keywords: *Cryptocurrencies, Bitcoin, Blockchain, Ransomware*

1. INTRODUCTION

Cryptocurrencies are digital, decentralized currencies that use cryptography to secure the transactions and control the creation of additional units of a currency. They exist only in digital form and they function independently of a central bank or any other central authority. Unlike traditional currencies, their supply is usually limited, predefined by the mathematical algorithm.

The oldest and the most famous cryptocurrency is bitcoin. Bitcoin was first presented, as *peer-to-peer electronic cash system*, in a white paper [1] published by Satoshi Nakamoto[1] on October 31st, 2008 and the first bitcoins were created on January 3rd, 2009. Bitcoin attracted little attention at the beginning as it took almost a year and a half until some value was attributed to it. The first time it was used as a payment method was on May 22nd, 2010, when 2 pizzas were bought for 10.000 bitcoins, valuing bitcoin at around $0.003. Bitcoin price has risen significantly since then, reaching all-time-high on Septembet 1st, 2017, when bitcoin was traded for slightly over $5000. Bitcoin price is very volatile and, although the price has mostly been rising since bitcoin was created, there have been some very sharps drops in its value. The last one happened at the beginning of this month, when the price fell around 40%, right after reaching $5000.

Although bitcoin code is open source, which makes it easy for anyone to create a similar cryprocurrency, no such thing happened until late 2011, when Litecoin[2] was created. It would soon be followed by other projects, with number of cryptocurrencies increasing rapidly and now exceeding 1.000.

Cryptocurrencies use Blockchain as the underlying technology. According to [2], *a blockchain is a distributed database that maintains a set of information bundles called blocks*. Blockchain applied in cryptocurrency represents a public ledger of all the transactions of a certain cryptocurrency that ever happened. There are some significant differences between blockchain and the other ways of storing the data.

As the name implies, data is packed in blocks and blocks are connected in a chain. Each block contains the hash of the previous one, as shown in Figure 1[3], making it impossible to change the data within a single block without changing all the blocks that come after it. Ledger is shared between thousands of independent participants (nodes), connected in a peer-to-peer network, with every single one of them having the full copy of the ledger, and they all verify every transaction and every block that is added to the blockchain.

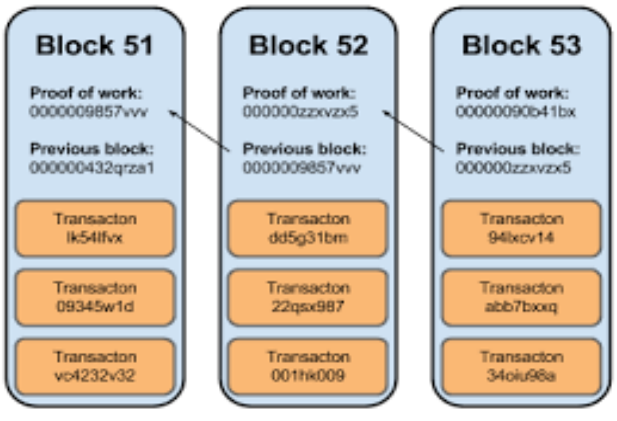

Figure 1: Connecting blocks in the blockchain

[1] Satoshi Nakamoto's true identity still remains a mystery. There were several attempts to reveal his/her true identity but they were unsuccessful and it is still unknown who is hidden behind the name.

[2] Cryptocurrency which resembles bitcoin in many ways, with slightly different mining algorithm, and blocks created every 2.5 minutes instead 10 minutes, which is the case with bitcoin.

[3] Source: https://www.ybrikman.com/writing/2014/04/24/bitcoin-by-analogy/

Since there is no central authority that governs the system, there also isn't any central server where the system is run. Instead, the system is maintained by so called "miners", people who dedicate their resources to maintaining the network and are being awarded for that with newly generated units of a cryptocurrency. There are different types of mining algorithms, with most dominant being proof-of-work[4] and proof-of-stake[5]. There are also cryptocurrencies that use combination of two different mining algorithms.

Networks of the top cryptocurrencies are being maintained by at least thousands of nodes and changing the data that has already been added to the blockchain would require consensus of the majority of nodes. It is a very unlikely event and it is probably safe to say that blockchain technology brings us an unprecedented level of data integrity. Changing the history is not, however, totally impossible. The most famous rollback, which caused a lot of controversy, is the one done on Ethereum[6] Blockchain on July 20th, 2016, when transaction history was rewriten in order to delete transactions connected with the DAO hack[7].

Blockchains can be public (e.g. Bitcoin, Ethereum, Litecoin...) and private (or permissioned). Public blockchain are accessible to everyone and they have their native cryptocurrency, most of all because miners need to be incentivized for their contribution to the network. Private blockchains are accessible only with a permission. They are newer type of blockchains and usually don't have a native currency or a token attached to it. Some argue that private blockchains are useless and that true power of blockchain technology lies in decentralization, transparency and accessibility for all.

Potential applications of the blockchain technology go far beyond cryptocurrency and even far beyond the whole financial industry. They include industries such as: insurance, media, healthcare, government, identity, asset titles, supply chain and many more. It is still a young technology and most of those applications are still in experimental phase.

2. BITCOIN AND RANSOMWARE

Because of its pseudo-anonimous[8] nature, bitcoin has often been linked to different illegal activities. In most cases, its supposed use in such activities was highly exacurated. However, it is hard to deny the link between ransomware and bitcoin, the currency that has been used lot of those attacks. Acording to [3], 75% of attackers demanded the ransom to be paid in bitcoins.

The most common pattern is:

- Hackers send the ransomware, usually as an e-mail attachment.
- When attachment is opened, the ransomware starts encrypting files on the computer and also on other computers that might be connected to the infected one through the internal network. More advanced ones target the recently used files in order to maximize the damage by making sure the most important files are encrypted first. In that case, even if the encryption is interupted at some moment, the most important files are already encrypted.
- In the Windows Desktop and in each folder that contains infected files, ransom notes are created. Notes provide a step-by-step guide on how to set up a bitcoin wallet, where and how to purchase bitcoins and where to send them in order to acquire a file decrypter.
- Amount of bitcoins that is required is usually doubled after certain time intervals (e.g. every week) urging the victim to pay as soon as possible, also threatening that every chance for data recovery will be lost forever if the payment is not completed until the deadline.
- After the payment is made, victims are given the link they could use to download the decrypter which decrypts the infected files.

While most fo the hackers do provide file decrypter after being paid for, there is no guarantee that they would actually do so. Some just collect the payment and never deliver anything. Others ask for additional payment after receiving the initial one. The problem in predicting hacker's behavior is the fact that the barrier to entry for hackers is extremely low. One can literally send the ransomware to potential victims without having any knowledge about hacking or even coding. On Dark Web[9], it is possible to purchase a ransomware kit and have your own ransomware within minutes.

Depending on the type of ransomware, in some cases it is possible to exchange messages with the hacker. In those cases, if the payment is the only solution, victim should try to negotiate the price because there were cases when attackers were willing to lower the price significantly. However, payment should always be the last option. Unfortunately, statistics [4] shows that corporate victims pay the ransom in 70% of the cases, with half of those paying over $10000. Table 1 shows the posibilities for bargaining depending on the type of the ransomware.

[4] Rewards that miners are getting are proportional to the processing power of the hardware they use to help maintain the network.

[5] Rewards that miners are getting are proportional to the amount of a currency they already hold.

[6] The second biggest cryptocurrency in terms of market capitalization

[7] The DAO was Decentralized Autonomous Organization run on Ethereum blockchain network. On June 17th, 2016, the vulnerability in DAO code was exploited and about $50M was stolen

[8] All the bitcoin transactions there ever happened are visible to everyone. However, names and other personal data of the participants in transactions are not exposed.

[9] A part of the internet only accessible using special software. That software helps website operators and users to remain anonymous and untrecable

FAMILY	STARTING DEMAND	LOWEST DEMAND	%DISCOUNT
CERBER	530	530	0%
CRYPTOMIX	1900	635	67%
JIGSAW	150	125	17%
SHADE	400	280	30%
			AVERAGE: 29%

Table 1: Bargaining opportunities for several types of ransomware[10]

After suffering a ransomware attack, victims' attitude towards bitcoin is usually negative, some even go as far as blaming bitcoin for the attack. It is important to know that ransomware existed well before bitcoin and will continue to exist after hackers stop using bitcoin as a payment method. It is just the most convenient payment method for them at the moment with the right blend of anonymity and accessability for the victims. There are other cryptocurrencies (e.g. Monero[11]) that have advanced anonymity features, but they are more difficult to acquire[12] than bitcoin, which is probably why they still havent't replaced bitcoin as a preferred payment method for ransomware. However, since the number of users and the availability of those currencies is rising, it is just a matter of time when bitcoin will be replaced by more anonymous cryptocurrencies. It has already happened on Dark Web illegal markets, where bitcoin has lost ground to Monero.

Bitcoin is generally much better understood than it was before. It is not nearly as anonymous as some might think. All the transactions are recorded forever in a public ledger and visible to everyone. Although the names of the participants in a transaction are not shown, it is possible to discover the names behind almost every bitcoin transaction. The current problem is that the procedure is relatively complicated and time-consuming so it is not feasible to track every single transaction. Instead, those who track them are mostly focused on larger suspicious transactions. Besides, there is still a lack of personel skilled enough to conduct blockchain research and tools that would facilitate blockchain analysis. Once those obstacles are overcome, it is logical to expect that bitcoin will be avoided by most of the hackers and others that currently use it in their illegal activities.

3. WHAT CAN BE LEARNED FROM CRYPTOCURRENCIES?

With traditional currencies, people trust banks or other financial institutions to conduct the payments on their behalf and store money for them. The convenience of having some institution taking care of one's money is paid for with limited freedom in using of using that money.

The approach that cryptocurrencies offer is completely different. One of the main reasons for their relatively slow adoption is in fact their nature, which requires people to change the mindset and change the perception of money as a concept. For the first time, we have unlimited freedom to use our money and unlimited control over our funds. However, that comes with a price – unlimited responsibility and that is something people are having hard time getting used to.

Embracing cryptocurrencies is about embracing both the freedom in using the money and the responsibility that comes with it. The one who holds a private key to a cryptocurrency wallet has a total control over whatever is on that wallet. On the other hand, if the private key is lost and no backup has been done prior to that, the funds on the wallet are lost forever. That is the lesson many have learned the hard way.

Living in a highly centralized world, we have learned to trust different centralized organizations and trade freedom for convenience. What cryptocurrencies teach is that freedom can be great, but can also be very expensive if not handled approached properly. It is something that can be applied to the area of cybersecurity. The freedom of using the internet can be very dangerous and expensive if there is a lack of responsibility. It is important to always remember that we are the ones holding the "private key" and that the whole responsibility is always on us.

4. SHOULD BLOCKCHAIN BE USED TO STORE IMPORTANT DATA?

Whenever there is an important piece of data, it has to be stored in a secure way and it needs to have a backup. However, backups are sometimes not done often enough or even not done at all. Blockchain has an inherent backup mechanism within itself. As explained earlier, blockchain-based systems have a big number of nodes who hold the complete copy of the ledger. Those nodes communicate with each other and update their respective copies of the ledger in real-time. Having thousands of automatically updated copies of the database looks like a perfect way to store any kind of data. Is it really so?

[10] Source: "F-Secure State of Cybersecurity", https://www.f-secure.com/documents/996508/1030743/cyber-security-report-2017

[11] Monero uses ring signatures to make the transactions untraceable. Besides the names of the participants in a transaction, Monero addresses and the amounts are also invisible.

[12] Number of cryptocurrency exchanges and similar services that are listing Monero is still very limited.

If we are talking about public blockchains, they have higher number of nodes, they are more secure, more robust, they have a longer track record, but they also have one serious limitation – capacity. Those systems are maintained by individuals and are designed to attract as many of those individuals as possible. Therefore, designers of those systems had to be mindful about users' bandwidth and hard drive capacity. As a consequence of that, the rate of adding new data to the blockchain is relatively low. In bitcoin, up to 1MB of data is added to the ledger every 10 minutes[13]. Some public blockchains have bigger capacity, but not big enough.

What public blockchains do provide is the very high level of data integrity. Blockchains are almost immutable as changing the history requires a huge effort that can't be unnoticed. So, when we have some piece of information and it is critically important that the information cannot be altered, storing it in the blockchain makes a lot of sense, as long as the size of the data is relatively small.

Another issue with public blockchain is the transparency. If blockchain contains the data of interest to the public, it is a great feature. However, if the data is sensitive and shouldn't be revealed to third parties, public blockchain is far from perfect solution.

There is a way benefit from blockchain's immutability even when we have larger chunks of important data. We would have to store the data outside of the blockchain, then apply cryptographic hash function to the data and store only the hash of that data on the blockchain. If the data changes, the hash of that data changes too and it doesn't match the hash stored on the blockchain anymore.

Private blockchains don't have issues with the capacity. They are usually maintained by an organization or a group of organizations that can have full nodes hosted on big servers with high bandwidths. They are, however, less secure than the public ones, because they usually have significantly lower number of nodes.

Running a private blockchain within a single company, with a purpose of having the data stored in a more secure way doesn't make much sense. It could only benefit huge organizations that spread across different countries. Even then, there are cheaper and easier way to handle data security.

Having a private blockchain for a group of organizations, especially within the same industry, is something that has grabbed the attention of many companies and institutions all over the world. A simple example could be the group of insurance companies sharing the database of fraudulent customers to reduce the risk of frauds or banks sharing KYC data between themselves and saving a lot of time. It is still very questionable whether the blockchain-based solution is the more efficient one at this stage of the development of blockchain technology.

There are some blockchain-based projects specifically focused on storage. The idea behind most of those projects is to store the data on unused hard drive space of other users in the network instead on a centralized server. The latest big one that appeared and attracted a lot of attention was Filecoin. It has raised an astonishing $200M in only 1 hour through its ICO[14]. Other significant projects are Storj and Sia.

5. CONCLUSION

Cryptocurrencies and blockchain present a very significant innovation. Cryptocurrencies not only change the way we use money but also the way we think about money. By taking away the control over money from central banks and cutting off the intermediaries, cryptocurrencies manage to speed up the transaction, lower the transaction fees and make money accessible to anyone.

Blockchain technology revolutionizes the way data is sent and stored. It introduces the trust into a distributed system without central authority. Cryptocurrencies are the first application of the blockchain technology, but blockchain might be applied to many other different areas.

Ransomware has been an increasing threat in the recent years. Barrier for entry for hackers has been significantly lowered since ransomware-as-a-service is being offered on Dark Web. Besides, bitcoin has made it easier for hackers to collect their ransoms and stay relatively anonymous. Three out of four hackers use bitcoin as the payment method for ransomware. Around 70% of all the corporate victims end up paying the ransom. It implies that most of them probably don't do backups regularly. If payment is in fact the only option, bargaining should be tried whenever possible, because the ransom can be lowered as much as 67%.

Since bitcoin is not nearly as anonymous as many may think and tools for blockchain analysis are getting better and better, it is reasonable to expect that bitcoin will be replaced as a "ransomware currency" in the near future by one of the cryptocurrencies with advanced anonymity features. Judging by its penetration on Dark Web markets, Monero is probably the strongest candidate.

Blockchains as databases are almost immutable and have a big number of updated copies at any given moment. It is natural to think they would be ideal for storing data. However, not every type of data can take advantage of the blockchain technology. Public blockchains mostly suffer from limited bandwidth and limited storage capacity. They are also relatively transparent and accessible to anyone which makes storing sensitive data unfeasible.

Private blockchains are more flexible and don't have issues with the capacity. However, they are not as secure as public ones and, as young as the blockchain technology is, private blockchain as a concept is even younger. There is a lot of testing done in that area, but the number of usable solutions

[13] Maximum size of a block of data is 1MB, although some blocks are smaller. Blocks are not added exactly every 10 minutes, but in average. Time between 2 blocks may vary from 1 minute to couple of hours

[14] ICO is the short of Initial Coin Offering. It is a new method of raising the money for the project which has been heavily used (and abused) in 2017, with over 2 billion dollars raised since the beginning of the year.

is still very limited. There are blockchain-based projects focused specifically on data storage. Although some of them have attracted a lot of interest and investments, we are yet to see them being widely adopted.

REFERENCES

[1] S. Nakamoto, “Bitcoin: A Peer-to-Peer Electronic Cash System”, 2008

[2] D. Verma, N. Desai, A. Preece, A. Taylor, “A blockchain based architecture for asset management in coalition operations”, 2017

[3] Dr. L. Hadlington, “Exploring the Psychological Mechanisms used in Ransomware Splash Screens”, 2017, De Montfort University, Leicester

[4] IBM Study, “Businesses More likely to Pay Ransomware than Consumers”, 2016

THE ROLE OF SOFTWARE TESTING IN A SECURITY-ORIENTED IoT SOFTWARE DEVELOPMENT PROCESS

LJUBOMIR LAZIĆ
Belgrade Metropolitan University, Faculty of Information Technologies, ljubomir.lazic@metropolitan.ac.rs

Abstract: *The main aim of this paper is to enhance the security in Internet of things (IoT) and services, which have the potential scope and benefits not only for the end users but also for the service providers and adaptors. We look at the challenges of security testing IoT applications. When software gets deployed on components that can fly and accelerate, testing for safety and trustworthiness takes on new meaning. Poor security can lead to denial-of-service attacks, corporate espionage, theft and brand damage. As more devices become Internet-enabled, experts fear an embedded systems security worst-case scenario for enterprises, many of which are unaware of the risks or unable to mitigate them. This article discusses the role of software testing in a security-oriented software development process. It focuses on two related topics: functional security testing and risk-based security testing. As a case example we will focus on SCADA and PLC devices which are complex embedded systems often relying on some operating system. They are plagued by the same sorts of vulnerabilities and exploits as general purpose operating systems. One solution to prevent and design in secure IoT we propose a Software Assessments and Security Testing Framework.*

Keywords: *IoT, IoT security, IoT system, Smart devices, Smart services, SCADA, PLC security, Security Testing*

1. INTRODUCTION

Most likely the most demanding of requirements for the widespread realization of many IoT visions is security. IoT security has an exceptionally wide scope in at least four dimensions. In terms of security scope it includes rarely addressed tasks such as trusted sensing, computation, communication, privacy, and digital forgetting. It also asks for new and better techniques for the protection of hardware, software, and data that considers the possibility of physical access to IoT devices. Sensors and actuators are common components of IoT devices and pose several unique security challenges including the integrity of physical signals and actuating events [1]. Finally, during processing of collected data, one can envision many semantic attacks [2].[1]

There can be many loopholes in the security of IoT and to start with, these loopholes can be at the very basic level of IoT where the data is routed to the service provider. Usually the smart meters that forward data to the service providers do not do it directly but through a local hub, which is again another smart meter. The data is collected and stored in these local hubs and then it is forwarded to the service provider in bulk. This makes the data vulnerable to attacks as it is not being stored at just one place. But this flaw is never addressed or it is ignored and can be found in most of the internet of things. The reason behind this could be a compromise made in incorporating required technical features, or to have an infrastructure that could accommodate all the input devices or even to have a network that keeps the things connected all the time. This occurs when it becomes difficult to meet the expenses in maintaining a high-end infrastructure for the internet of things [4] [5] [6].

In the first stage of IoT evolution, objects are personified, i.e. they are given identities with the help of QR codes. The services are met when there is an interaction between the identities of the object and the intelligent systems – for example web services, smart devices etc.

In the 2nd stage, the internet of things develop the smartness of sensing things around. For example, locating the place the user is in, locating Bluetooth devices or WiFi networks etc. Let us consider a situation where a hacker takes advantage of this sensing capability of the smart devices. An air conditioner with a sensing capability of the owner's presence switches on/off automatically. So now the hacker can get control over the AC or even a thermostat and change the temperature to trouble the owner.

We're at a crisis point now with regard to the security of embedded systems, where computing is embedded into the hardware itself -- as with the Internet of Things. These embedded computers are riddled with vulnerabilities, and there's no good way to patch them. Typically, these systems are powered by specialized computer chips made by companies such as Broadcom, Qualcomm, and Marvell. These chips are cheap, and the profit margins slim. Aside from price, the way the manufacturers differentiate themselves from each other is by features and bandwidth. They typically put a version of the Linux operating system onto the chips, as well as a bunch of other open-source and proprietary components and drivers. They do as little engineering as possible before shipping, and there's little

[1] This work was supported in part by the Ministry of Science and Technological Development of the Republic of Serbia under Grant No. TR-35026.

incentive to update their "board support package" until absolutely necessary [6].

The problem with this process is that no one entity has any incentive, expertise, or even ability to patch the software once it's shipped. The chip manufacturer is busy shipping the next version of the chip, and the ODM (original device manufacturers) is busy upgrading its product to work with this next chip. Maintaining the older chips and products just isn't a priority. And the software is old, even when the device is new. For example, one survey of common home routers found that the software components were four to five years older than the device. The minimum age of the Linux operating system was four years. The minimum age of the Samba file system software: six years. They may have had all the security patches applied, but most likely not. No one has that job. Some of the components are so old that they're no longer being patched. This patching is especially important because security vulnerabilities are found "more easily" as systems age [6].

To make matters worse, it's often impossible to patch the software or upgrade the components to the latest version. Often, the complete source code isn't available. Yes, they'll have the source code to Linux and any other open-source components. But many of the device drivers and other components are just "binary blobs" -- no source code at all. That's the most pernicious part of the problem: No one can possibly patch code that's just binary. The result is hundreds of millions of devices that have been sitting on the Internet, unpatched and insecure, for the last five to ten years. Combine full function with lack of updates, add in a pernicious market dynamic that has inhibited updates and prevented anyone else from updating, and we have an incipient disaster in front of us. It's just a matter of when.

We simply have to fix this. We have to put pressure on embedded system vendors to design their systems better. We need open-source driver software -- no more binary blobs! -- so third-party vendors and ISPs can provide security tools and software updates for as long as the device is in use. We need automatic update mechanisms to ensure they get installed [6].

2. SECURE BY DESIGN - EXAMPLES

Vulnerabilities are introduced into our embedded systems during software design and development. Although incorporating security features such as encryption and password protection will help to safeguard access to devices and data, such features are insufficient when the application code contains defects that render it vulnerable. So while architects strive for more secure features and designs, the best approach for securing embedded software applications is to find and address coding issues at an early stage and then deliver high-quality, defect-free code. To do this, developer need tools that can help them ensure that the code they write is free from known weaknesses and follows proven guidelines and standards.

While our software line of defences will surely be less than perfect, we need to work on that line of defences with the immediate objective of reducing the size of the "attack windows" that exist in our software. The very first step in doing this is to try to think like an attacker. Ask how an attacker could exploit your system and your software in order to penetrate it. You might call this a threat analysis. Use the results to describe what your software should not do. You might call those abuse cases. Use them to plan how to make your software better resist, tolerate or recover from attacks.

Don't forget that our attackers have a big advantage when it comes to embedded systems: Most embedded software has severe execution time constraints, often a mixture of hard real-time and soft real-time tasks. This coaxes us to design application software that is "lean and mean," by reducing to a minimum intensive run-time limit checking and reasonableness checking (for example, invariant assertions) in order to meet timing requirements. Our attackers have no such execution time constraints: They are perfectly happy to spend perhaps weeks or months researching, preparing, and running their attacks--possibly trying the same attack millions of times in the hope that one of those times it might succeed, or possibly trying a different attack each day until one hits an open "attack window."

Many embedded devices use analog-to-digital-converters (ADCs) for data acquisition. These ADCs may be sampled on a regular timed basis, and the data samples stored by application software in an array. Application software later processes the array of data. But an attacker could view this in a totally different way: "What if I fed the ADC with electrical signals that, when sampled, would be exactly the hexadecimal representation of executable code of a nasty program I could write?" In that way, the attacker could inject some of his software into your computer. No network or Internet needed.

ADC Code Injector - example

Seems like a lot of work to build an "ADC Code Injector" device just for this purpose. But the attacker might not be just a high-school kid. He might be a big industrial espionage lab, or a large, well-funded team working at the national laboratory of a foreign government.

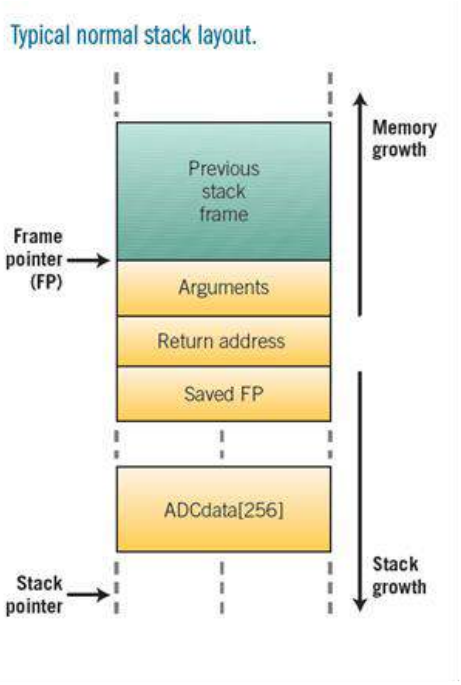

Image 1: Typical normal stack layout

Now, how could he get your processor to execute his program that he's injected? He might gamble that your software stores the ADC data array on a stack (perhaps using `alloca()` or `malloca()`). If his luck is good, he could cause an array overflow, possibly by toying with the hardware timer that controls the ADC data sampling. A typical normal stack layout is shown in Image 1.

If the attacker succeeds in causing an array overflow, the stack could become corrupted, as shown in Image 2. Note that "return address" was stored on the stack at a location beyond the end of the array.

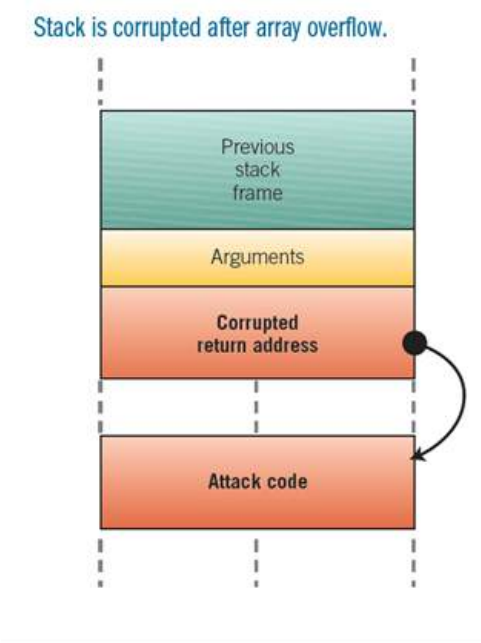

Image 2: Stack is corrupted after array overflow

If the attacker plans the corruption just right, the overflow will reach the location on the stack where the current return address was stored. This can be used to insert into this stack location a pointer to his own code. As a result, when "Return Address" is used by your code, control will pass to the attacker's code. Suddenly his code is executing on your processor, instead of your code.

This is called a **stack smashing attack**. Please note that it was done in this example without an Internet connection, and without a connection to any external communication line.

Of course, it could have been helpful for our attacker to have the source code for your embedded software--as a disgruntled ex-employee might. But think that a patient and resourceful attacker team could develop this kind of attack even without your source code.

Unique challenges of securing embedded applications

Developers have long been concerned with the quality of the software they create and have processes in place to detect and eliminate defects that adversely affect quality. Many organizations, however, have not yet adopted strategies for ensuring the security of the software they create. Because fixing issues in a deployed embedded device is both costly and difficult, addressing both quality and security problems in the early stages of development is imperative.

Compared to their counterparts who develop software for traditional devices, including computers and smartphones, embedded developers have far more variables to consider. Embedded developers face the unique and almost impossible challenge of gaining a deep understanding and proficiency in multiple combinations of operating system, platform, I/O interface, and language. Traditional developers work on a limited number of platforms, enabling them to become more familiar with specific security issues and the areas in which common software vulnerabilities can occur and be prevented. In contrast, embedded developers often work on a variety of platforms, each of which might handle data storage and memory usage in a different way. New platforms are introduced frequently, making it almost impossible for embedded developers to thoroughly understand the unique vulnerabilities of each OS/platform/language/interface combination.

“No matter what happens, don’t panic,” were the words used by hackers just before they hacked a 2014 Jeep Cherokee. It wasn’t your typical hack, where credit card information is stolen, or a denial of service attack is propagated, or a website is taken down. This incident involved disabling the transmission and brakes of a vehicle driving 70 mph. In other words, this is the kind of hack that could take someone’s life.

SCADA and PLC example

Supervisory Control and Data Acquisition (SCADA) systems have evolved over the past 40 years, from standalone, compartmentalized operations into networked architectures that communicate across large distances. In addition, their implementations have migrated from custom hardware and software to standard hardware and software platforms. These changes have led to reduced development, operational, and maintenance costs as well as providing executive management with real-time information that can be used to support planning, supervision, and decision making. These benefits, however, come with a cost. The once semi-isolated industrial control systems (ICS) using proprietary hardware and software are now vulnerable to intrusions through external networks, including the Internet, as well as from internal personnel. These attacks take advantage of vulnerabilities in standard platforms, such as Windows, and PCs that have been adopted for use in SCADA systems [6].

The control components of SCADA systems are optimized to provide deterministic, real-time performance at a reasonable cost. Thus, there are little computing resources available for executing other functions not considered necessary for the basic SCADA mission. As a result, SCADA system manufacturers view additional computing tasks, including information system security, as burdens on the computing capacity that could interfere with the proper operation of the system. Information system security was not inherent in SCADA protocols because, when the protocols were developed, SCADA systems were usually operating in closed environments with no vulnerable connections to the outside world. In today’s SCADA applications, the opposite is true.

SCADA systems are connected to corporate IT networks and use protocols and computing platforms that are under attack in the conventional IT world.

SCADA system based cyber security attacks have the very real possibility of impacting life safety, the environment and organizational survival. In a worst case scenario, say a SCADA system cyber-attack successfully penetrates a refinery system. In this scenario, the attacker alters some critical data to reflect a safe condition while blocking the ability to generate essential safety control commands. In this situation, the process could easily exceed a safe limit, an explosion and fire could occur which not only costs the loss of life but also destroys the firms basic process infrastructure. The refinery could go out of business. From this example it could be easily seen that SCADA system cyber security attacks can have a much greater impact on the organization than an IT cyber security attack.

Another reason IT cyber security processes cannot be directly applied to the SCADA system is associated with how the systems must operate, i.e. system availability. SCADA systems must operate non-stop where system outages and interruptions are not tolerated. This is a different environment than IT systems where planned system outages or unavailable times can be planned and do occur. A prime example of how the different availability technology requirements impact cyber security approach is highlighted with operating system updates.

SCADA systems cyber security challenges are also slightly different than enterprise systems in the areas of vendor certifications, anti-virus software verifications and password rules as well. Vendor supplied SCADA applications function within the operating system. The vendors provide extensive testing and validation that their SCADA system will perform as designed with a specific computer operating system. The difference comes about in how fast, if ever, that the software vendor provides certification that its SCADA system will operate correctly with the latest set of updates or the next operating system version. It is not uncommon to find some SCADA vendors are extremely slow in providing validation or that they will never validate that their older systems are capable of operating correctly with a newly released operating system.

The SCADA/ICS world is facing a situation where there will be a massive number of unpatched and vulnerable computers running on critical systems for the next years. And that is not good news. There are many examples of how the entire strategy of patching for SCADA and ICS security is broken.

Similar security issue we recognized with Programmable Logic Controllers (PLCs), which were designed to eliminate the higher cost of complicated, relay-based control systems. Today, almost every PLC, DCS, Remote Terminal Unit, or Safety Integrated System (SIS) controller on the market has a commercial operating system in it. Microsoft Windows vulnerabilities are abound and reported in various resources on the Internet. Similar is with Linux and QNX. Here we will emphasize the Allen Bradley Logix family, which is the most full featured programmable controllers in the line of Rockwell Automation.

The ControlLogix is the flagship product of the Logix family. It consists of a chassis with controller, power supply and I/O modules that can be used as both a controller and a gateway. The number and type of modules is determined based on the size and type of system being controlled, network topologies and protocols, and redundancy requirements. Its configurations can vary greatly with the large number of modules and ability to mix and match to meet requirements. The 1756-ENBT and 1756-EWEB (with web server) modules provide an Ethernet connection to the ControlLogix and warrant special attention from an information security perspective.

A wide range of control system protocols are supported on the ControlLogix platform. For communication from a server, HMI or other controllers, the ControlLogix supports EtherNet/IP, ControlNet and Data Highway as well as other standard protocols from third party modules such as Modbus TCP.

1756-ENBT/A brings Ethernet connectivity to the controller, thus opening up the door to a whole range of remote attack vectors. For example, it could be easily seen via nmap:

```
snmp-netstat:
TCP 0.0.0.0:80  0.0.0.0 ; http(GoAhead)
TCP 0.0.0.0:111   0.0.0.0 ; rpcbind
TCP0.0.0.0:44818 0.0.0.0 ; EtherNet/IP
UDP 0.0.0.0:68 *:*  ; dhcp(if enabled)
UDP 0.0.0.0:111 *:* ; rpcbind
UDP 0.0.0.0:161 *:* ; snmp
UDP 0.0.0.0:2222 *:* ; EtherNet/IP
UDP 0.0.0.0:44818 *:* ; EtherNet/IP
```

Port 44818 is used by the Rockwell Automation software (RSLogix, RSLink…) drivers to communicate with those ControlLogix controllers which have EtherNet/IP modules enabled. EtherNet/IP is an application layer protocol treating devices on the network as a series of "objects". It is built on the Common Industrial Protocol (CIP), for access to objects from ControlNet and DeviceNet networks. RSLogix, RSLinks and other Rockwell Software can be easily downloaded from Rockwell's support website. By interacting with this software while monitoring the network traffic we can easily analyze and extract the packets needed to monitor and control the PLC i.e. obtain information about the processes running on the CPU or update the firmware. With the little help from Shodan search engine it is easy to find ControlLogix devices on the web. The first site we have found was www.scrapmetal.net (American Iron & Metal Co. Inc.). We get there immediately when we enter http://204.101.14.75/index.html in our browser. It is an 1756-ENBT/A web page with completely operational menu on the left side, including the full diagnostics and refreshing rate every 15 seconds. It could be easily seen that the firmware date is Jan, 7 2005. This is valuable information for someone who wants to prepare an attack to the device. ControlLogix uses GoAhead web server, which is a simple, portable and compact web server for embedded devices and applications. It is one of the most widely

deployed web servers and is embedded in hundreds of thousands of devices. Unfortunately, this web server contains vulnerabilities that may allow an attacker to view source files containing sensitive information or bypass authentication. The information disclosure vulnerability was published in [10].

In our work [5], [7] we provided one possible solution, implement **Security design principles** that can be organized into logical groups, which are illustrated in Fig. 3. The logical groupings for the principles are in shaded boxes whereas the principles appear in clear boxes. For example, Least Privilege is a principle and appears grouped under Structure/Trust. In the case of "Secure System Evolution," the principle is in its own group. There are few Enterprise-wide software security improvement program initiatives to **establish secure SDLC - SecSDLC** (presented in Image 4) which consists of:

• Strategic approach to assure software quality
• Goal is to increase systematic approach
• Focus on security functionality and security hygiene

Organizations with a proper SDLC will experience an 80 percent decrease in critical vulnerabilities. In each initiatives (Gary McGraw Touch-Point Model [3], SEI Team Software Process for Secure Software Development, etc.) building Security Into the Software Life Cycle (**SecSDLC**) include strong security testing as in our OptimalSQM framework [8], [9].

3. SOFTWARE ASSESSMENTS AND SECURITY TESTING FRAMEWORK

The method by which security assessment and testing is carried out depends on the perspective of the tester relative to the software component. We developed **OptimalSQM** Test Framework Architecture we call BISA (**B**usiness **I**ntelligence **S**imulation **A**rchitecture). BISA is an SDLC framework architecture that provides a Software Testing Centre of Excellence for SMEs with scalability, objectivity, consistency and constant improvement features capable to support full lifecycle software assurance following software engineering standards and to provide better and cheaper software products on time. The concept and functional requirements of OptimalSQM have been developed since several years ago, resulting into cost effective software test metrics [9], economic model of software quality, and finally matured by SDLC framework deploying Software-as-a-Service (SaaS) and Testing-as-a-Service (TaaS) models.

It consists of several mutually connected subsystems as shown in Image 5. The software engineering heart of BISA is an SDLC-engine consisting of OptimalSQM expert tools [9] and SQA wheels of BISA, actually service-enabled SDLC engine is able to communicate with external environment either via portal or a service bus, software engineering-enabled ESB (Enterprise Service Bus). The last subsystem of the framework is e-invoicing. OptimalSQM framework focuses on a significant part of SPI (Software Process Improvement) effort, the deployment of which may result with ROI of 100:1, comparing with existing SDLC infrastructure of software companies that have achieved CMM and TMM maturity of the 1 and 2 level. The current research aims at achieving two goals: build an SDLC framework that is capable to support companies that have already reached higher levels of CMM, and inject service-oriented capabilities into this framework to allow customers to invoke a particular service on demand.

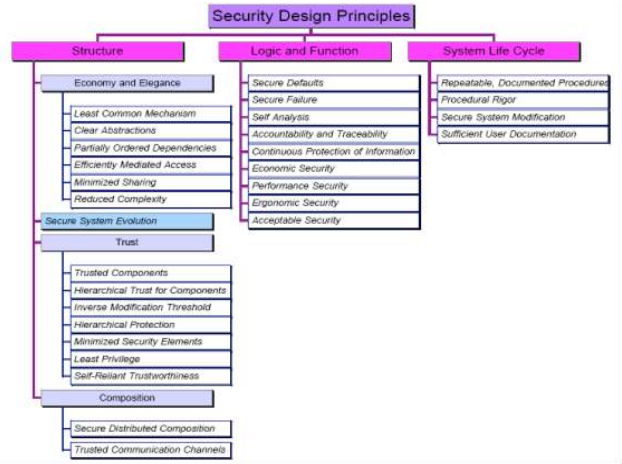

Image 3: Taxonomy of security design principles

Image 4: Building Security Into the Software Life Cycle (SecSDLC) initiatives

OptimalSQM framework deploy Testing-as-a-Service (TaaS) models. Testing as a Service is an outsourcing model, in which testing activities are outsourced to a third party that specializes in simulating real world testing environments as per client requirements. It is also abbreviated as TaaS.

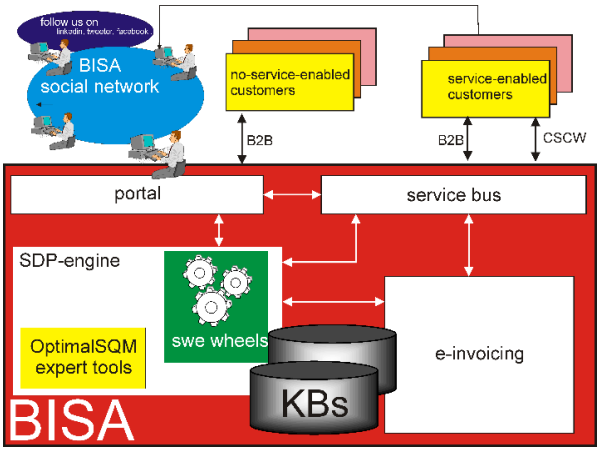

Image 5: The overall architecture of BISA

Types of TaaS

• **Functional Testing as a Service**: TaaS functional testing may include UI/GUI testing, regression, integration and

automated user acceptance testing (UAT) but not necessary to be part of functional testing

• **Performance Testing as a Service**: Multiple users are accessing the application at the same time. TaaS mimic as a real world users environment by creating virtual users and performing the load and stress test, and

• **Security Testing as a Service**: TaaS scans the applications and websites for any vulnerability.

Security Testing is a variant of Software Testing which ensures, that system and applications in an organization, are free from any loopholes that may cause a big loss. Security testing of any system is about finding all possible loopholes and weaknesses of the system which might result into a loss of information at the hands of the employees or outsiders of the Organization. The goal of security testing is to identify the threats in the system and measure its potential vulnerabilities. It also helps in detecting all possible security risks in the system and help developers in fixing these problems through coding. **Types of Security Testing:** There are seven main types of security testing as per Open Source Security Testing methodology manual. They are explained as follows:

• **Vulnerability Scanning**: This is done through automated software to scan a system against known vulnerability signatures.

• **Security Scanning**: It involves identifying network and system weaknesses, and later provides solutions for reducing these risks. This scanning can be performed for both Manual and Automated scanning.

• **Penetration testing**: This kind of testing simulates an attack from a malicious hacker. This testing involves analysis of a particular system to check for potential vulnerabilities to an external hacking attempt.

• **Risk Assessment**: This testing involves analysis of security risks observed in the organization. Risks are classified as Low, Medium and High. This testing recommends controls and measures to reduce the risk.

• **Security Auditing**: This is an internal inspection of Applications and Operating systems for security flaws. Audit can also be done via line by line inspection of code

• **Ethical hacking**: It's hacking an Organization Software systems. Unlike malicious hackers,who steal for their own gains , the intent is to expose security flaws in the system.

• **Posture Assessment**: This combines Security scanning, Ethical Hacking and Risk Assessments to show an overall security posture of an organization.

4. CONCLUSION

This article provided a distillation, synthesis and organization of key security systems design principles, describes each principle, and provides examples where needed for clarity. Although others have described various principles and techniques for the development of secure systems, e.g. [3], it was felt that a concise articulation of the principles as they are applied to the development of the most elemental components of a basic security system would be useful.

High assurance is required for the embedded operating systems that control today's, and tomorrow's, safety and security critical systems. The security policies of operating system components must not fail; high assurance is the only path to this goal. A software security practitioner should perform the following to manage system security risks:

- creating security abuse/misuse cases,
- listing normative security requirements,
- performing architectural risk analysis,
- building risk-based security test plans,
- wielding static analysis tools,
- performing security tests,
- performing penetration testing in the final environment,
- cleaning up after security breaches.

REFERENCES

[1] Dong Chen, et al, “A novel secure architecture for the Internet of Things”, IGEC, 5th conference, IEEE Xplore, 2011.

[2] S. Milinković, Lj. Lazić „ON CYBER SECURITY OF INDUSTRIAL MEASUREMENT AND CONTROL SYSTEMS“, INFOTEH-JAHORINA Vol. 13, 19. mart - 21. mart 2014., Jahorina, hotel Bistrica, ZBORNIK RADOVA, ISBN 978-99955-763-3-2, March 2014. pp.903-908.

[3] J. Viega and G. McGraw, “Building Secure Software”, Addison-Wesley, New York, 2001.

[4] T. V. Benzel, C. E. Irvine et. al.: “Design Principles for Security”, NS-CS-05-010, Naval Postgraduate School, California, USA, 2005.

[5] Lj. Lazic, S. Obradović and A. Donchev." Access Control For E-Business:Structure Or Architectural Security Principles", UNITECH'08 - INTERNATIONAL SCIENTIFIC CONFERENCE, 21 – 22 November 2008, GABROVO, 2008, Proceedings pp.I324-I329.

[6] S. Milinkovic, Lj. Lazic: Some Facts about Industrial Software Security, Proc. XI International Conference on Systems, Automatic Control and Measurements (SAUM 2012), Nis, Serbia, November 14-16 (2012) 232-235.

[7] Lj. Lazić, Dž. Avdić, A. Pljasković, „Software Security Analysis, Metrics, and Test Design Considerations“, Proceedings of the 6th WSEAS EUROPEAN COMPUTING CONFERENCE (ECC '12), Prague, Szeh Republic, September 24-26, ISBN: 978-1-61804-126-5, 2012, pp.355-367.

[8] E. Kajan, L. Lazic, and Z. Maamar, “Software Testing as a Service (TaaS): The BISA Approach”, in Proceedings of IEEE TELSIKS’11, Nis, Serbia. October 2011.

[9] Lj. Lazić, "OptimalSQM: Optimal Software Quality Management Repository is a Software Testing Center of Excellence", Plenary Lecture 1 in the WSEAS The 6th EUROPEAN COMPUTING CONFERENCE (ECC '12), Prague, Czech Republic, September 24-26, 2012.

[10] US-CERT Vulnerability Note #975041: GoAhead Web Server discloses source code of ASP files via crafted URL, 11 Jan 2010.

SECURITY MECHANISMS IN IOT

IVAN TOT
University of Defence, Military Academy, Serbia, ivan.tot@va.mod.gov.rs

DUŠAN BOGIĆEVIĆ
University of Nis, Faculty of Electronic Engineering, dusan.bogicevic@gmail.com

KOMLEN LALOVIĆ
Faculty of applied management, economy and finance - MEF, komlen@mef.edu.rs

MIODRAG BRZAKOVIĆ
Faculty of applied management, economy and finance - MEF, miodrag.brzakovic@mef.edu.rs

IVANA OGNJANOVIĆ
University Donja Gorica, Humanistic studies, Montenegro, ivana.ognjanovic.edu@gmail.com

Abstract: *The paper deals with the security of the services used in IoT. The paper presents theoretical foundations and the IoT architecture. It describes in detail the architecture and types of IoT services in IoT, as well as the protocols used to communicate with the services in order to review possible security issues and suggest possible improvements regarding the security of IoT services. The work includes IoT devices, which are the basis of IoT, and their importance in the safe operation of IoT services is presented.*

Keywords: *IoT security, IoT services*

1. INTRODUCTION

The term Internet of Things (IoT) was created in 1999. It was conceived as a world of objects that exchange data. Data exchange is not only between man and machine, but communication between machines (M2M) is also introduced. Kevin Ashton, in his paper published in 2002, under the title IoT, said: "We need an internet for things, and a standardized way for computers to understand the real world."

The International Organization for Standardization (ISO), in 2012 founded the group ISO/IEC JTC 1/SWG 5[1], which will deal with standardization in the area of Internet of Things (IoT). This group defined IoT as: "An infrastructure of interconnected objects, people, systems and information resources together with intelligent services to allow them to process information of physical and virtual world and react." [1][2]

Nikola Tesla had a vision of IoT almost 100 years ago „when wireless is perfectly applied, the whole earth will be converted into a huge brain ... ". Researchers Caceres and Friday identified two critical infrastructures that will impact ubiquitous computing - Cloud Computing and the Internet of Things. [3]

IoT can be viewed from two perspectives. The first, where IoT is viewed from the perspective of the Internet where attention is paid to Internet services, while the second perspective focuses attention on smart things. [4]

2. IOT ARHITECTURE

The IoT architecture is not based on one device. It is about sets of devices that collect information in different ways. When talking about IoT, the most considered topic are environments. The prefix "smart" is often found like e.g. smart homes, smart streets, smart parking lots, smart garbage cans, smart cities etc. Smart environments can be defined as sets (federations) of sensors and actuators designed for house, building, city, transport etc. [5] Mark Weiser, who is considered as the founder of ubiquitous computing, has defined smart environments as a world of physical objects that are connected with sensors, actuators, displays, other environments over a network that allows interlaced connectivity. [4]

From the highest level, IoT consists of three parts (Figure 1): [5]

- **Part of devices:** Devices or actuators with their communication components integrated with them.
- **Middleware part:** The most complex part that implements data processing logic, stores data and provides access to data to users in such a way that they do not care about the architecture of individual devices (actuators). This layer is made up of several parts.
- **Presentation part:** This part adjusts to a specific application and performs data display, data management, etc.

[1] https://en.wikipedia.org/wiki/ISO/IEC_JTC_1/SWG_5

Figure 1: IoT parts

Integration and functioning of these three parts is possible by the following three components (Figure 2):

- **Cloud Platform:** This component is responsible for providing functionalities such as real-time data processing, data storage, data scalability, global data access and the provision of other functionalities such as machine learning etc.
- **Cloud Infrastructures:** The hardware on which the platforms are executed, the memory needed to store the data and the network resources needed for communication.
- **Middleware component:** The component that provides physical device abstraction, as well as the interaction between the network platform and the devices. [6]

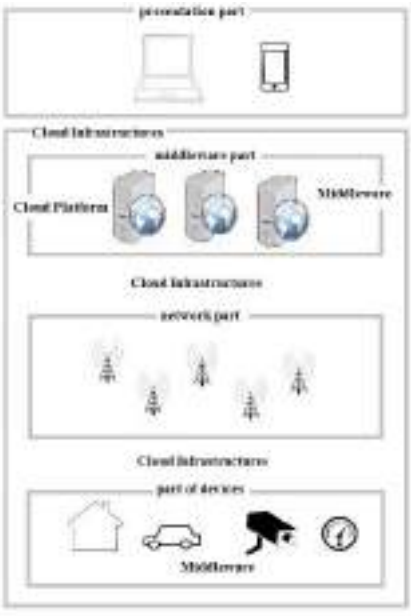

Figure 2: IoT components

3. IOT DEVICE

The definition of things in the IoT term changed as the technology changed. The radical change is reflected in the changing of the Internet into network-linked objects that not only produce information in their environments and interact with the physical world, but objects that now provide standardized Internet services that allow information exchange, analysis, communication and development of different applications. [4]

In order to achieve the idea of connecting objects, it is necessary that objects (devices) combine hardware and software components on their layers to achieve the functionality of physical objects. The development of technology in microelectromechanical systems (MEMS) has led to the creation of small digital devices that provide wireless communication as well as minimum dimensions, with the ability to measure values, calculate, and communicate at shorter distances. Such devices are called nodes and are connected to networks called sensor networks and find application in environments such as traffic monitoring etc. [4]

The purpose of the device layer is to process collected mostly analogue data and to send it in digital form over the network layer to the server layer.

The number of connected devices exceeds the number of human population. In 2010 the number of devices was almost twice the number of human population.

The architecture of the device (object) layer should consist of three parts (Figure 3): [7]

- **Middleware component**, parts of software that will allow device management,
- **IoT component**, which will collect data such as sensors (actuators), and the components through which communication with the server will occur (communication modules such as Ethernet, Wi-Fi, Bluetooth, ZigBee etc.),
- **Hardware part**, on which the software will be executed and to which the sensors (actuators) will be connected.

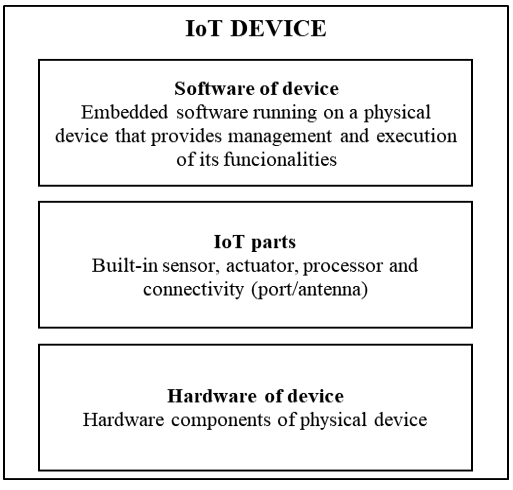

Figure 3: Device layer architecture

4. IOT SERVICES

The middle part of the IoT is the most complex and contains all three components (Cloud Platform, Cloud Infrastructures, Middleware component). The central part in the middle layer are services, so this layer is often called a service layer.

Services perform more functions, so depending on the function they perform, the service layer can be split into:

1. **Service composition:** This part allows the development of smaller parts for specific applications. Parts are often made as APIs. This layer does not provide the devices to the application part, but the set of services of a particular part. On this layer, it is possible to view all connected service instances and to represent complex processes as one sequence that coordinates multiple actions over one component.
2. **Service management:** This part provides the main functionality expected from the service, namely: finding, managing, monitoring, configuring. This layer should also allow the addition of new services, during run-time, to meet the needs of applications.
3. **Object abstraction:** Since IoT is based on a large number of versatile devices, in order to manage the objects themselves it is necessary to perform their virtualization, with common language and procedures for all objects.

Services used in IoT (as well as in Cloud computing) can be based on multiple models, but are most often based on one of the following three service models (Figure 4): [8][9]

- **Infrastructure as a Service (IaaS):** This model provides a service for using hardware and software components. The basic concept behind this model is virtualization, which allows users to use their operating system and not to worry about maintenance. Some of the IaaS examples are: Amazon Web Service (AWS), Rackspace, Windows Azure etc.
- **Platform as a Service (PaaS):** This model provides resources such as operating system, programming language, database etc. This platform serves as the basis for developing applications using the APIs, so developers develop applications for specific environments. The developer takes care of the application's functionality, but remains bound to the platform that it uses. The platform as a service reduces the complexity of applications that are developing, by choosing the necessary hardware and software that needs to be purchased. An example of the platform as a service is the Google App Engine.
- **Software as a Service (SaaS):** In this model, applications are developed and executed on the server. Applications are executed on the server and shared between users. They can be accessed through browsers that are connected to the Internet, while the user selects the functionalities of the software. The advantage of the software as a service model is that it does not require the installation of software, nor the possession of hardware on which the service would be executed.

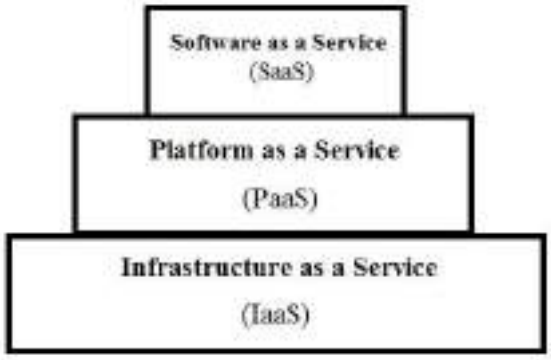

Figure 4: Service models

5. IOT PROTOCOLS

With the development of IoT, it was possible to define different protocols that could be used in communication between machines (M2M). Some of the limiting factors that these protocols should face are:

- limited communication channel width,
- small hardware resources of the device,
- different media for data transmission,
- a large number of participants,
- usage of wireless communications.

Some of the protocols designed for M2M communication and IoT are:

1. Advanced Message Queuing Protocol (AMQP),
2. Message Queuing Telemetry Transport (MQTT),
3. Constrained Application Protocol (CoAP).

5.1 Advanced Message Queuing Protocol (AMQP)

AMQP is an application protocol developed in 2003 by John O'Hara, for the needs of banks, and it was officially adopted in 2012 by OASIS (Organization for the Advancement of Structured Information Standards).

AMQP uses the TCP protocol on the transport layer, and has an overhead of 8 bytes. It is based on message exchange on the principle of an advertiser/subscriber. This protocol achieves great reliability and ensures that the message is delivered even when the network breaks down. In terms of security, the SSL protocol is used. [10]

5.2 Message Queuing Telemetry Transport (MQTT)

MQTT is an application protocol designed in 1999 by IBM and standardized in 2013, which has a relatively small overhead, thus providing a possible application on devices with limited resources (memory, processor etc.) such as IoT devices. This protocol, like the HTTP protocol, uses TCP on the transport layer, but has smaller overhead of 2 or 4 bytes.

The protocol uses the principle of advertisers and subscribers. Facebook Messenger application uses this protocol.

In terms of security, this protocol uses TLS. Brokers may require a username and password. [11]

5.3 Constrained Application Protocol (CoAP)

The main goal of this protocol is to reduce the overhead to a minimum and provide a mechanism that would be used on a large number of devices that have limitations in terms of power, resources and network considerations (low-range networks such as IEEE 802.15.4, Bluetooth, Low Power Wi-Fi). HTTP protocol was uses as a model for development. It is important to note that CoAP is not reduced HTTP, but it is a protocol optimized for M2M communication, which supports basic REST functionalities, common with HTTP protocol. Also, CoAP in some things represents a step forward in comparison to HTTP. It supports multicast, asynchronous messaging and has a mechanism for finding resources. [12] [13]

CoAP is an application protocol that uses two messaging models. It supports the request/response model, as well as the advertiser/subscriber model. Unlike HTTP, it relies on the UDP protocol.

The default transport protocol for transmitting messages with the CoAP is UDP, but the DTPLS protocol can also be used to increase the security above UDP. In terms of security, it relies on the functionality of other protocols. In addition to UDP, other protocols such as SMS, TCP or SCTP can be used.

6. SECURITY MECHANISIMS IN IOT

IoT core represents platforms that are specialized services for collecting data from IoT devices, data processing, and connection with other services for decision-making about further activities such as activating actuators, etc. [14]

Developing ones' own platform requires a larger team of people and a time of several years. It is estimated that there are over 300 of them on the market and that their number is growing. Existing platforms have been developed over the years, and behind them is the work of large teams that have been working on the development of the platform for nearly 20 years. [15]

Most of the developed platforms use the existing infrastructure (IaaS, PaaS). In this way, part of the responsibilities related to the physical execution of the service or the execution of the operating system software is transferred to the others, thereby reducing the part of the responsibility of the team developing its own platform.

Regarding the architecture of the service that is being used, the REST architecture is dominant, while in terms of security, some of the cryptographic algorithms are used such as SSL, TLS, AES, X.509 etc. [16]

It is important to emphasize that data is kept from others, but it should also be in mind that IoT represents the network of devices ("the social network of devices") that exchange data so the data should be available but in a controlled way. A large number of data will allow the development of ubiquitous computing as well as the development of context aware computing. [6]

Application developers also have a part of the responsibilities for the development of the application and its functionalities. In terms of security, access to the application and the data it has at its disposal is the most important. [17]

The platform is expected to have built-in security mechanisms. These mechanisms should also be implemented at the level of the protocols used in communication.

At the lowest level, it is necessary to take care of the safety of the IoT device. These devices are the most numerous members of the IoT architecture, and are the sources of information on which the functioning of the entire system is based. Security mechanisms used at the level of IoT devices should be adapted to their capabilities (limited hardware, software and communication capabilities).

7. CONCLUSION

The development of the IoT services in cases with limited development potentials should be aimed to usage of ready platforms that have implemented cryptographic and other mechanisms in order to reduce application development time and provide all the necessary functionalities.

REFERENCES

Articles from Conference Proceedings (published):

[1] A. Dorri, S. Kanhere, R. Jurdak, "Blockchain for IoT security and privacy: The case study of a smart home", IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops), 2017.

[3] R. Caceres, A. Friday, "Ubicomp Systems at 20: Progress, Opportunities, and Challenges", IEEE Pervasive Computing, 2012, pp. 14-21.

[4] C. Perera, A. Zaslavsky, P. Christen, D. Georgakopoulos, "Context Aware Computing for The Internet of Things: A Survey", IEEE Communications Surveys & Tutorials, 2013, pp. 414-454.

[5] J. Gubbi, R. Buyya, S. Marusic, M. Palaniswami, "Internet of Things (IoT): A vision, architectural elements, and future directions", Future Generation Computer Systems, Elsevier, 2013, pp. 1645-1660.

[6] M. Díaz, C. Martín, B. Rubio, "State-of-the-art, challenges, and open issues in the integration of Internet of things and cloud computing", Journal of Network and Computer Applications, Elsevier, 2016, pp. 99-117.

[7] S. Fachmedien, “Internet of ThingsTechnology and Value Added”, Business & Information Systems Engineering, 2015, pp. 221-224.

[8] I. Ashraf, “An Overview of Service Models of Cloud Computing”, International Journal of Multidisciplinary and Current Research, 2014, pp. 779-783.

[9] S. K. Sowmya, P. Deepika, J. Naren, “Layers of Cloud – IaaS, PaaS and SaaS: A Survey”, International Journal of Computer Science and Information Technologies, 2014, pp. 4477-4480.

[16] M. Sain, Y. Kang, H. Lee ,, Survey on security in Internet of Things: State of the art and challenges“,19th International Conference on Advanced Communication Technology (ICACT), 2017.

[17] D. Bogićević, I. Tot, R. Šendelj, “IoT Security Optimization“, The Eight International Conference on Business Information Security (BISEC), 2016.

Technical Reports:

[2] “Internet Of Things”, International Organization for Standardization, 2015.

[10] „AMQP Advanced Message Queuing Protocol Specification”, Cisco Systems, 2008.

[11] A. Stanford-Clark, H. L. Truong, “MQTT for Sensor Networks (MQTT-SN) Protocol”, 2013.

[12] V. Sharma, “Understanding Constrained Application Protocol”, Exilant Technologies Pvt, 2014.

[13] Z. Shelby, “The Constrained Application Protocol (CoAP)”, Internet Engineering Task Force (IETF), 2014.

[14] P. Mell, T. Grance, “The NIST Definition of Cloud Computing“, National Institute of Standards and Technology, USA, 2011.

[15] “IoT Platforms: The Central Backbone for the Internet of Things”, IoT Analytics Gmbh, 2015.

SECURE MODULAR AUTHENTICATION SYSTEMS BASED ON CONVENTIONAL XOR BIOMETRICS

NEMANJA MAČEK
School of Electrical and Computer Engineering of Applied Studies, Belgrade and eSigurnost Association, Belgrade, macek.nemanja@gmail.com

MILAN MILOSAVLJEVIĆ
Singidunum University, Faculty of Technical Sciences, mmilosavljevic@singidunum.ac.rs

IGOR FRANC
Belgrade Metropolitan University, Faculty of Information Technologies and SECIT Security Consulting, igor.franc@metropolitan.ac.rs

MITKO BOGDANOSKI
Military Academy General Mihailo Apostolski, Skoplje, Macedonia, mitko.bogdanoski@ugd.edu.mk

MILAN GNJATOVIĆ
University of Novi Sad, Faculty of Technical Sciences, milangnjatovic@uns.ac.rs

BRANIMIR TRENKIĆ
School of Electrical and Computer Engineering of Applied Studies, btrenkic@viser.edu.rs

Abstract: *This paper presents an approach to designing secure modular authentication system based on conventional XOR biometrics. System consists of one or more clients, an authentication server and a trusted storage. Client is a device used to capture biometrics, obtain auxiliary data and create encrypted cancelable templates during the enrolment and verification phases. Authentication server manages encryption keys and verifies cancelable templates, while the trusted storage, which can be either distributed or centralized, stores the encrypted templates. Two important characteristics of the proposed system are that it keeps biometric templates encrypted or cancelable during all stages of storage, transmission and verification, and that it does not suffer from severe computational costs and large sizes of encrypted templates like systems based on homomorphic encryption. Additionally, system is general (i.e., it does do not depend on specific cryptographic algorithms) and modular, which allows a user enrolled on one client to verify his identity on another client connected to the same authentication server. Finally, security of the system is compared with the requirements of a cryptographically secured biometric system that provides strong privacy protection*.

Keywords: *Biometrics, Authentication, Security, Cryptography*

1. INTRODUCTION

Biometric authentication is the process of establishing user identity based on physiological or behavioral qualities of the person [1, 2]. Biometrics can be addressed as an ultimate authentication solution: users do not need to remember passwords or carry tokens and biometric traits are distinctive and non-revocable in nature [3], thus offering non-repudiation [4]. However, like any personal information, biometric templates can be intercepted, stolen, replayed or altered if unsecured biometric device is connected to a network or if a skilled attacker gains physical access to a device. A brief surveys of attacks on biometric authentication systems, such as replaying old data, stored template modification and communication chanell interception are given in [5, 6]. Due to non-revocability of biometric data aforementioned attacks and misuses may lead to identity theft. Having that said, it becomes clear that biometric systems operate with sensitive personal information and that biometric template security and privacy are important issues while designing such authentication systems. To counterfeit identity theft, one should not rely on administrative countermeasures or misuse identification upon successful attack [7], followed by erradiction and recovery from damages caused by illegimite access to the resources. Identity theft should be prevented with technological countermeasures that provide sufficient level of security and privacy while downgrading the performance of the system (computational costs and storage requirements) to the reasonable level.

One approach to biometric template security and privacy is cancelable biometrics. Cancelable biometrics refer to intentional distortion of biometric features with non-invertible transforms [8]. In this scenario, while verifying the user the same transform is applied to a given sample as in enrolment phase. If template is considered to be compromised, it's revoked, as large number of transforms are available. If a non-invertible transform operates with a key, template is revoked and only the key is changed during template update. Examples of cancelable transforms are given in [9-11]. Non-invertible transforms are, however, not a fail-safe solution to a problem. They may

be computationally expensive, partially reversible and they degrade overall accuracy of the system. Additionally, system is vulnerable to substitution attack if an adversary who knows how the transform operates creates a masquerade sample.

Another approach to providing template privacy is the application of homomorphic encryption schemes [12, 13]. Homomorphic encryption refers to cryptographic algorithms that allow some computations to be performed in the encrypted domain. These schemes appears to be suitable for application in conventional XOR biometric systems (for example, iris based systems) as these systems use bitwise XOR to calculate Hamming distance during verification. Although applicable in theory, there are two reasons why homomorphic encryption is not actually practical: the encrypted template is large and the system is computationally expensive. According to [13], calculating the Hamming distance between two encrypted 1024 bit templates would take approximately 10 minutes on 2GHz processor.

The main contribution of this paper is a general secure modular authentication architecture based on conventional XOR biometrics applicable to a variety of real-life scenarios. An approach presented in this paper employs public key cryptography, pseudorandom number generators and cancelable biometrics. Non-invertible transform operates with the key stored on a token, thus reassembling two-factor authentication. The system does not suffer from the drawbacks of homomorphic encryption as cryptographic operations are not computationally expensive and no large templates are created. As stated before, biometric templates are encrypted or at least remain cancelable during all stages of operation (excluding feature extraction) resulting in a system prone to variety of attacks. Also, the system satisfies the requirements of a cryptographically secured biometric system that provides strong privacy protection listed in [7].

2. AN OVERVIEW OF ATTACKS ON BIOMETRIC SYSTEMS

Biometric systems, as all traditional systems are susceptible to variety of threats: Denial of Service, circumvention, repudiation, contamination, coercion and collusion [14]. Aforementioned threats are used to make attacks on biometric authentication systems. Eight different attack on unimodal biometric authentication systems consisting of sensor, feature extraction, matching and decision making modules have been identified in [15]. These include: sensor attack, replay attack (bypassing the sensor), attack on the feature extraction module, attack on the channel between feature extractor and matcher, compromising the database, attack on the communication channel between template database and the matcher and overriding the result declared by the matcher module. More on the protection from these attacks can be found in [16]. Attacks on biometric encryption systems (such as hill-climbing attack [17], non-randomness attacks [18], re-usability attack [19], blended substitution attack [20] and linkage attack [21]) are usually more complex when compared to traditional biometric authentication systems. The goal of an adversary is to reduce the search space, obtain the key or to create a masquerade version of biometrics [7].

3. MODULAR BIOMETRIC SYSTEMS AND SECURITY REQUIREMENTS

As mentioned before, biometric authentication systems consisting of four modules that reside in one device are vulnerable to variety of attacks [15]. To prevent execution of these attacks, entire system is split into three high-level modules (residing on at least two devices) and both cancelable biometrics and strong cryptographic protection are introduced to the system. The modular system now contains of: one or more clients (devices used to capture biometrics, obtain auxiliary data from the user and create encrypted cancelable templates), an authentication server (device that manages encryption keys and verifies cancelable templates) and a trusted storage that stores the encrypted templates. If two or more clients are used within the system, and a user enrolled on one client should be allowed to verify his identity on another client connected to the same authentication server, template storage must be centralized. As the proposed system deals with the XOR biometrics, a transform that reassembles the one-time-pad cypher is used.

Aside from cryptographic security, system is expected to provide strong privacy protection, resulting in the following set of requirements: (1) biometric templates remain encrypted or at least cancelable during all stages of storage, transmission and verification (e.g. authentication server should never obtains unencrypted biometric templates) and (2) no client is allowed to access private keys stored on authentication server as it may compromise the security of the templates. Further, resilience to a template substitution attack and all low level attacks is expected, the system should not suffer from severe computational costs and cryptographic countermeasures should not degrade the overall accuracy (i.e. they should not increase false acceptance or false rejection rates).

4. SYSTEMS WITH DISTRIBUTED STORAGE

Systems with distributed storage store encrypted templates on the clients. In this scenario, during the enrolment phase, the system operates as follows:

- User provides a token carrying numeric user ID and non-invertible transform key K_t to the client.
- Hash of the user ID is calculated on the client and sent to the authentication server. Authentication server generates a keypair (K_{priv}, K_{pub}), stores the private key with hash of user ID (H(id), K_{priv}) and sends public key to the client.
- Client obtains biometrics, creates a template b_0 and generates cancelable binary template $b = K_t \oplus b_0$.
- Client generates random seed s_0 and encrypts it with the public key: $s_E = E(s_0 , K_{pub})$. Client generates a keystream $s = PRNG (s_0)$ using pseudorandom number generator and given seed.
- Client calculates $s \oplus b$, stores (H(id), s_E, $s \oplus b$) and discards the rest of the data.

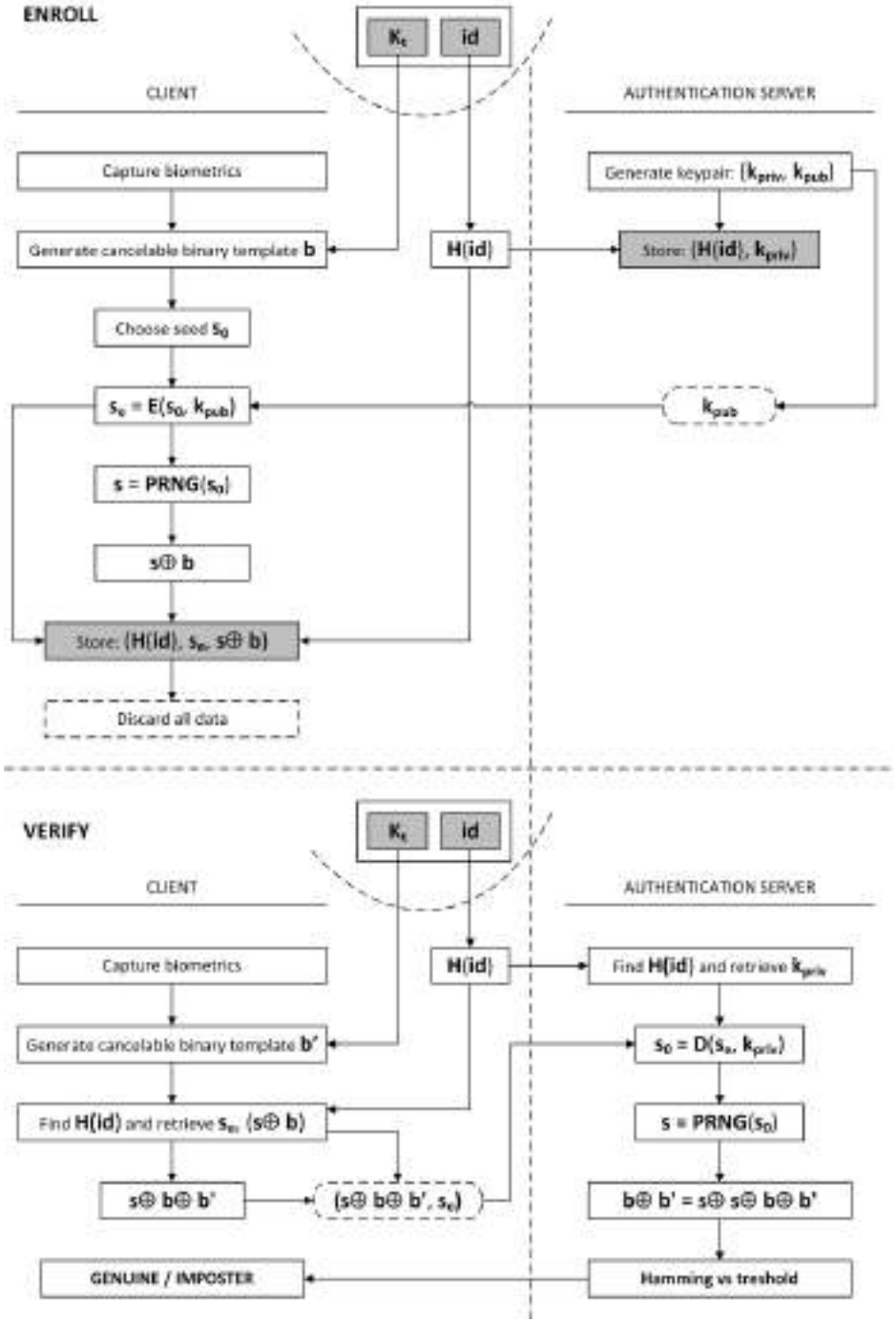

Image 1: Systems with distributed storage

During the verification phase, the system operates as follows:

- User provides a token carrying numeric user ID and non-invertible transform key K_t to the client.
- Client obtains biometrics, creates a template b_0' and generates cancelable binary template $b' = K_t \oplus b_0$'.
- Client calculates user ID hash and retrieves values s_E and ($s \oplus b$) from stored record (H(id), s_E, $s \oplus b$) with the corresponding user ID hash.
- Client calculates $s \oplus b \oplus b$' and sends it with the encrypted seed s_E to the authentication server.
- Hash of the user ID calculated on the client is sent to the authentication server. Authentication server retrieves private key from stored record (H(id), K_{priv}) with the corresponding user ID hash.
- Authentication server decrypts the seed with the private key s_0 = E(s_E , K_{priv}) and generates the keystream: s = PRNG (s_0).
- Server calculates $b \oplus b' = s \oplus s \oplus b \oplus b'$ and compares the Hamming distance between cancellable templates b and b' with the treshold. According to that result, the decision is made (user is genuine or imposter) and sent back to the client.

The security of the system may be summarized as follows. Templates are encrypted or at least cancelable during all stages of storage, transmission and verification, and the client is not allowed to access private keys stored on authentication server, which satisfies the conditions set for an ideal biometric system. System employs two factor authentication thus making an imposter with auxiliary data virtually impossible to claim as genuine user. If templates stored on a client are somehow compromised, reenrolment with another transform key and encryption key-pair will remediate the situation. Substitution attacks cannot be performed, as the public key is discarded at the end of enrolment. As an adversary cannot recreate the keystream s from the encrypted seed s_E and the public key, system is resilient to most of the attacks on the biometric encryption systems. Regarding the usability of the system, the following conclusions can be made: system can be employed in one client – one server scenario. System can be employed in many clients – one server scenario only if users enrolled on one client are not expected to verify their identity on another. However, user may enrol on multiple clients, but this would require a client ID to be stored with the encryption keys and user ID on the server. In this case, user would have to re-enrol on each client if the transform key is lost or stolen. Another limitation to the usability is that system deals with conventional XOR biometrics, which is not applicable to all modalities.

5. SYSTEMS WITH CENTRALIZED STORAGE

Systems with centralized storage do not store encrypted templates on the clients. They are logical extension of distributed storage systems.

In this scenario, during the enrolment phase, the system operates similar to systems with the centralized storage, with two major differences (see image 2):

- Values (H(id), s_E, $s \oplus b$) are not stored on the client. After calculating these values, client asks authentication server to issue an request to storage to add a record containing (H(id), s_E, $s \oplus b$) into the database.
- Client discards all data, not just remaining ones (public keys, the unencrypted seed and original template). This means that no data is stored on a client.

The key point here is that encrypted seed should never be stored on the authentication server as the corresponding private key is stored on it. This enforces the usage of a database that is run on separate device which communicates with the authentication server via encrypted channel.

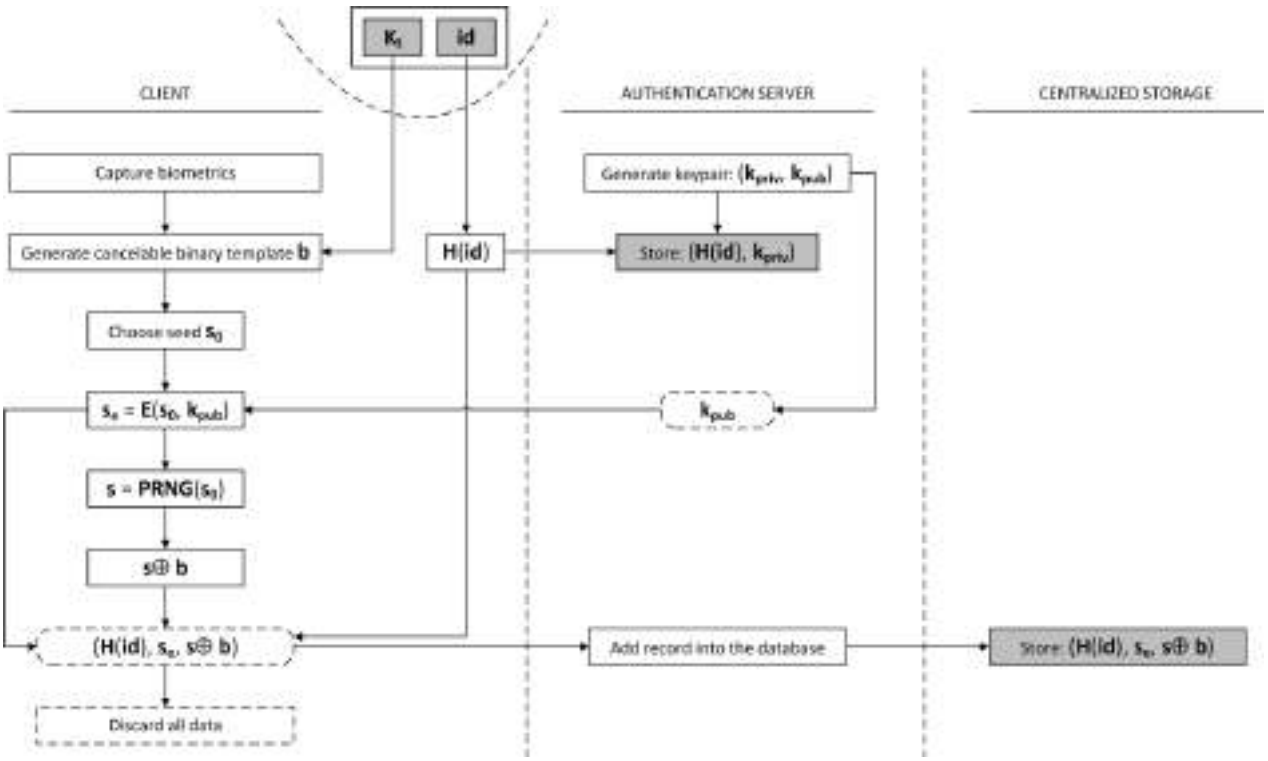

Image 2: Systems with centralized storage (enrolment phase)

During the verification phase, the system operates as follows:

- User provides a token carrying his numeric ID and non-invertible transform key K_t to the client.
- Client obtains biometrics, creates a template b_0' and generates cancelable binary template $b' = K_t \oplus b_0'$.
- Client calculates user ID hash and sends it to the authentication server.
- Authentication server contacts the centralized storage and retrieves values s_E and ($s \oplus b$) from corresponding (H(id), s_E, $s \oplus b$) stored on it.
- Authentication server sends value $s \oplus b$ to the client.

- Client calculates $s \oplus b \oplus b'$ and sends it back the authentication server.
- Authentication server retrieves private key from record (H(id), K_{priv}) with the corresponding user ID hash.
- Authentication server decrypts the seed with the private key $s_0 = E(s_E , K_{priv})$ and generates the keystream: $s = PRNG (s_0)$.
- Server calculates $b \oplus b' = s \oplus s \oplus b \oplus b'$ and compares the Hamming distance between cancellable templates b and b' with the treshold. According to that result, the decision is made (user is genuine or imposter) and sent back to the client.

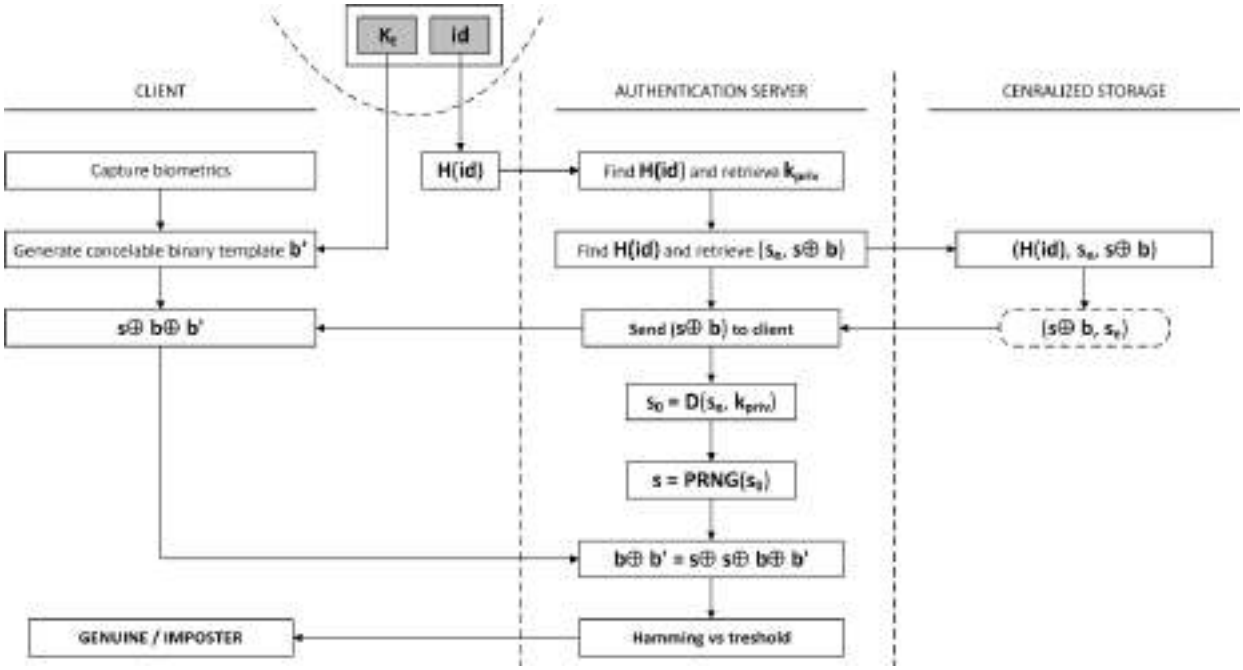

Image 3: Systems with centralized storage (verification)

The security of the system with centralized storage can be summarized as the security of the system with distributed one. However, additional cryptographic countermeasures are required to protect the communication channel between authentication server and centralized storage. The major difference between systems with distributed and centralized storage is the usability. One-to-many system does not require user to enrol on many clients as they share the stored templates on centralized storage. A user enrolled on one client can verify his identity on all clients connected to the same authentication server. These systems have a number of possible applications, ranging from facility entry control to securing mobile banking authentication.

6. CONCLUSION

This paper has introduced modular authentication systems architecture based on conventional XOR biometrics. The system keeps biometric templates encrypted or at least cancelable during all stages of storage, transmission and verification, and does not suffer from severe computational costs. Proposed architecture reassembles two factor authentication as the user who wants to verify identity must provide both biometrics and auxiliary data (non-invertible transform key). In further work we will explore the possible application of proposed authentication systems with centralized storage to secure mobile banking authentication.

REFERENCES

[1] A. K. Jain, A. Ross and S. Prabhakar, “An Introduction to Biometric Recognition”, IEEE Transactions on Circuits and Systems for Video Technology, Vol. 14, pp. 4-20, 2004.

[2] A. K. Jain and A. Ross, “Introduction to Biometrics”, in “Handbook of Biometrics”, A. Jain et al. (Eds), Springer, 2008.

[3] Y. C. Feng, P. C. Yuen and A. K. Jain, “A Hybrid Approach for Face Template Protection”, in Proceedings of SPIE Conference of Biometric Technology for Human Identification, Orlando, USA, Vol. 6944, pp. 325, 2008.

[4] P. Balakumar and R. Venkatesan, “A Survey on Biometrics-based Cryptographic Key Generation Schemes”, International Journal of Computer Science and Information Technology & Security, Vol. 2, No. 1, pp. 80-85, 2012.

[5] A. K. Jain, K. Nandakumar and A. Nagar, “Biometric Template Security”, EURASIP J. Adv. Signal Process, 2008:1-17, 2008.

[6] J. Galbally, C. McCool, J. Fierrez, S. Marcel and J. Ortega-Garcia, “On the Vulnerability of Face Verification Systems to Hill-Climbing Attacks”, Pattern Recognition, 43(3) pp. 1027-1038, 2010.

[7] A. Stoianov, “Cryptographically secure biometrics”, in SPIE Defense, Security, and Sensing, International Society for Optics and Photonics, 2010.

[8] N. Maček, B. Đorđević, J. Gavrilović and K. Lalović, “An Approach to Robust Biometric Key Generation System Design”, Acta Polytechnica Hungarica, Vol. 12, No. 8, pp. 43-60, 2015.

[9] N. K. Ratha, S. Chikkerur, J. H. Connell and R. M. Bolle, “Generating Cancelable Fingerprint Templates”, Pattern Analysis and Machine Intelligence, IEEE Transactions on, 29(4), pp. 561-572, 2007.

[10] J. Zuo, N. K. Ratha and J. H. Connell, “Cancelable iris biometric”, In Pattern Recognition, ICPR 2008, 19th International Conference on (pp. 1-4), IEEE, 2008.

[11] R. Ang, R. Safavi-Naini and L. McAven, “Cancelable Key-based Fingerprint Templates”, in C. Boyd & J. Gonzalez Nieto (Eds.), Australasian Conference on Information Security and Privacy, pp. 242-252, 2005.

[12] J. Bringer and H. Chabanne, “An authentication protocol with encrypted biometric data”, in International Conference on Cryptology in Africa, pp. 109-124. Springer Berlin Heidelberg, 2008.

[13] B. Schoenmakers and P. Tuyls, “Computationally secure authentication with noisy data”, in Security with Noisy Data, pp. 141-149. Springer London, 2007.

[14] A. K. Jain, A, Ross and U. Uludag, “Biometric template security: challenges and solutions”, in Proc. Europeon Signal processing conference, pp 1-4, September 2004.

[15] R. Jain and C. Kant, “Attacks on Biometric Systems: An Overview”, International Journal of Advances in Scientific Research, 1(07), pp. 283-288, 2015.

[16] B. Biggio, “Adversarial Pattern Classification”, Doctoral dissertation, University of Cagliari, Cagliari, Italy, 2010.

[17] A. Adler, “Vulnerabilities in Biometric Encryption Systems”, LNCS, Springer 3546, pp. 1100–1109, 2005.

[18] E.-C. Chang, R. Shen and F. W. Teo, “Finding the Original Point Set Hidden among Chaff” in Proc. ACM Symp. ASIACCS’06, Taipei, Taiwan, pp. 182–188, 2006.

[19] X. Boyen, “Reusable cryptographic fuzzy extractors”, in Proc. 11th ACM Conf. CCS, Washington, DC, pp. 82–91, 2004.

[20] W. J. Scheirer and T. E. Boult, “Cracking Fuzzy Vaults And Biometric Encryption”, Biometric Consortium Conference, Baltimore, September 2007.

[21] A. Cavoukian and A. Stoianov, “Biometric Encryption: The New Breed of Untraceable Biometrics,” in N.V Boulgouris et al., eds., “Biometrics: fundamentals, theory, and systems”, Wiley-IEEE Press, pp. 655-718, 2009.

The 9th International Conference on Business Information Security (BISEC-2017), 18th October 2017, Belgrade, Serbia

STRUCTURED APPROACH TO IOT SECURITY ANALYSIS

VITO LEGGIO, LYUDMILA ZHAROVA
University of Belgrade, FON, Belgrade, Serbia

ALEKSANDAR R. MIHAJLOVIĆ
Seven Bridges, Boston, MA, USA

RADOMIR A. MIHAJLOVIĆ
NYIT, New York, NY, USA

***Abstract**: Most of the papers on the very intensely researched topic of Internet of things or IoT announce that IoT technologies will profoundly change our lives. Our assessment is that advertised future is actually evolving now as we write or read these lines of text. IoT provides numerous benefits, and at the same time exposes container systems to a wide variety of security relevant problems in daily lives of systems users. In this work, we propose a structured approach to these sorts of problems proposing that all security issues may be related to the illegal access (access to data or services) or to the denial of services, (which we classify as soft and hard). In systems designs using IoT, both general classes of security relevant problems (we intentionally avoid the use of the terms such as threats or risks), being physical or logical are vastly expanded in comparison with the traditional stand-alone or networked systems. To avoid overextension, in this work, we maintain focus on the security problems of IoT protocols only, on their vulnerability, on possible defenses and specified use case distortions. In order to face a demand for IoT security relevant solutions capable of supporting a variety of application platforms with diverse device interactions, a structured approach to IoT security is more than justified. In this work we present a unique approach to IoT protocol classification, taking into account the most important security-relevant protocol parameters. Special attention has been devoted to the IoT protocols robustness correlation to the IoT user privacy.*

***Keywords:** IoT, Protocols, Security, Privacy, Robustness, Energy Footprint.*

1. INTRODUCTION

Current IoT platforms involve mostly technologies that originate from the traditional networks and embedded computing hardware and software.

In the application context when energy supply is not constrained IoT nodes may involve standard ISO-OSI seven-layer or five-layer TCP/IP-Internet protocols. In the deployment context when energy supply is limited or when other constraints impose relevant limitations IoT protocols must deviate from the traditional approaches. With such deviations security concerns, risks, attacks and defense mechanisms become different from the traditional too. A typical example of such IoT-specific protocol deviation from the traditional protocols is protocol 6LowPAN. The 6LowPAN protocol is used for L3 transmission of IPv6 packets over Low-power Wireless Personal Area Networks [1]. To become transmission power aware protocol, 6LowPAN was designed as a scaled down IPv6 protocol version whose use implied also modification/customization of other protocol stack members [2].

Among many proprietary IoT specific protocol solutions, many have become candidates for standardization. For instance protocols such as: CoAP, MQTT, MQTT, AMQP, XMPP, DDS, REST,HTML 5s WebSocket, etc. are used, offering IoT context and application matching characteristics.

In our research work, we have compared the context and application relevant capabilities of several commonly used protocols regarding their suitability for the IoT applications taking into an account protocol energy-footprint, (determined by the required computation and communication load), their reliability and their security attack sensitivity.

Due to the limited scope of this paper, we have focused on the security attack sensitivity presenting a structured classification of possible attacks targeted at IoT origin node device, interconnecting lines, and the operating context.

2. IOT NETWORKS AND SECURITY

IoT nodes are computation and communication enabled devices that appear as intelligent things. The most common IoT communications are node-to-node or device-to-device communications (D2D). IoT networks involve also data collection and transferred to designated application servers for data analysis analyzes (D2S), and finally communications between servers (S2S) which may involve service application architectures and relevant protocols.

Due to numerous constraints the most vulnerable IoT network communications are D2D and D2S. All functional and security issues of S2S communications are covered in literature related to traditional networking. This is reinforced by the agreement in the industry that almost every smart and networked IoT device is vulnerable to even simple attack [3,4].

Taking into an account the ecosystem of IoT devices, s smart device vulnerability can take two shapes. In the first case, computationally powerful devices with standard power supply present large attack surface and can be subjected to software-based traditional attacks ranging from L2 to L7 (Application layer), including attacks oo operating system with related services. In the second case, devices with battery power supplies, and reduced computation and communication capabilities, are easiest to attack in the physical L1 layer, (e.g., the signal layer of the wireless sensor and actuator networks [5]). One serious problem with such networks is a lack of feasibility of fixing exploit vector path by deploying device code updates, which is not the case with the previously mentioned powerful IoT device networks. In applications when the product lifetime of IoT devices is very short device code updates is realized as the deployment of the new devices. In applications when devices are expected to have long field life the code update problem becomes a serious security issue, The second problem with low computing, communicating and energy power devices, are reduced logging and forensic investigation options. One frequently neglected good security feature of these kinds of IoT devices are commonly found small dimensions (small physical footprints). For instance, small dimension wireless devices are very hard to spot, locate and physically attack.

Nonstandard proprietary protocols, not well documented or not published are sometimes good defense of the low computation power IoT device in the layers from L2 to L7, (Software layers).

I case of the successful attack on one of the IoT devices in the application network, a vulnerability of one device can cause attack propagation and automatic attacks against other devices in the network. As a countermeasure against the propagation of negative effects of the successful attack, require attack detection and prevention security dedicated meta-network (Out of band network) or application of innovative network monitoring, management, and control (sort of network micro-management). We address such an approach in [6] and propose an extension in a form of the Software Defined IoT Network (SDIoTN).

SDIoTN control application to IoT networks allows total minimization of the IoT node device functionality. With the minimized functionality IoT device obtains minimized attack surface and hardened security and robustness.

IoT devices are often built with firmware based code (Embedded software) which may reduce device system attack surface, requiring attacker-to-device physical contact [3]. To modify/infect device system firmware an attacker must physically replace device read-only memory media which is in many cases serious hardware intervention.

3. STRUCTURED IOT ATTACK MODELING

To better defend IoT networks, it is important to maintain the clear understanding of all possible attack exploit vector paths. Several groups of authors have simply itemized special cases of attacks in a form of specific class-objects itemizing while trying to identify some common attributes between recognized attacks [7,8].

Instead of presenting an unstructured taxonomy, in our paper we present a structured UML model of the “IoT Attack” class. As shown in Figure 1, IoT device or network can be attacked in all ISO-OSI layers. IoT device attack may involve physical destruction of a device, denial of service of the IoT device operating system or storage system, and denial of service of the IoT device applications operating in the L7 layer.

“L1 IoT Attack” subclass hierarchy represent all physical attacks on devices and network lines. The most common and the easiest attack to implement is a weak signal jamming attack which is Line DoS Attack.

In this particular model, we have proposed to classify L2-L7 and system DoS attacks as illegal access attacks.

Further expansion of the structured model of “IoT Attack” class as well as modelling of countermeasures are left to be presented elsewhere.

4. CONCLUSION

In no other domain of the cyberspace, security and productivity (doing more and better in shorter period of time at the lower cost) oppose each other more than in the IoT device networks. The most restrictive element of the confrontation is the cost of production and deployment.

In this paper we declare that seemingly paradoxical proposal to reduce cost while enhancing security is to deploy SDIoTN architecture. Additional details relevant to this IoT network design superpattern will be presented elsewhere.

REFERENCES

[1] Montenegro at al, "Transmission of IPv6 Packets over IEEE 802.15.4 Networks," Network Working Group, RFC-4944, September 2007. https://tools.ietf.org/pdf/rfc4944.pdf

[2] Nurul Halimatul, Asmak Ismail, Rosilah Hassan, Khadijah W. M. Ghzali, "A Study on Protocol Stack in 6LowPAN Model," Journal of Th. and Applied Info. Technology, 31st July 2012. Vol. 41 No.2, pp.220-229. http://eprints.utem.edu.my/5649/1/12Vol41No2.pdf

[3] Martin Unger, Bastian Bergmann, "The IoT Needs a Paradigm Shift from Security to Safety of Connected Devices," CicleID, Aug. 23, 2017.

http://www.circleid.com/posts/20170823_iot_needs_aradigm_shift_from_security_to_safety_of_devices/

[4] Mario Ballano Barcena, Candid Wueest, "Insecurity in the Internet of Things," Symantec White Paper, Version 1.0 – March 12, 2015.

[5] J. Sen, "A Survey on Wireless Sensor network Security", Int. Journal of Communications Network and Info. Security, vol. 1, no. 2, 2009 August, pp.59-82.

[6] Vito Leggio, Lyudmila Zharova, Sravanthi Dontu, Aleksandar Mihajlović, Radomir A. Mihajlović, "Software Development Problems of the SDN Internals Engineering," The 9th Int. Conf. on Business Info. Security (BISEC-2017), 18th October 2017, Belgrade, Serbia.

[7] Sachin Babar, Parikshit Mahalle, Antonietta Stango, Neeli Prasad, Ramjee Prasad, "Proposed Security Model and Threat Taxonomy for the Internet of Things (IoT)," N. Meghanathan et al. (Eds.): CNSA 2010, Springer, 2010, CCIS 89, pp.420-429.

[8] Bonnie Zhu, Anthony Joseph, Shankar Sastry, "A Taxonomy of Cyber Attacks on SCADA Systems," Internet of Things (iThings/CPSCom), 2011 International Conference on and 4th International Conference on Cyber, Physical and Social Computing, 19-22 Oct. 2011, http://ieeexplore.ieee.org/document/6142258/

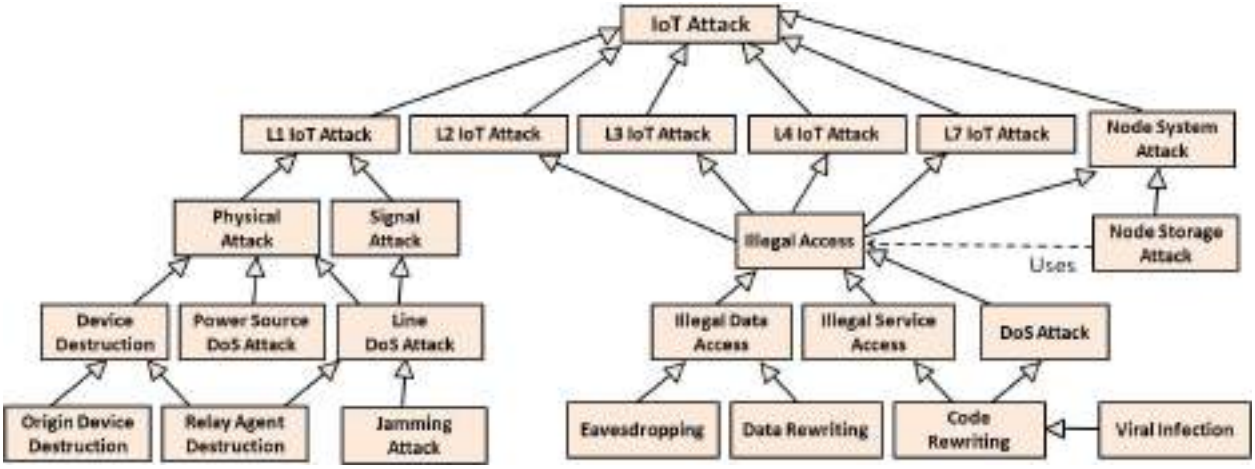

Figure 1: UML class diagram model of the "IoT Attack."

CONCEPTUAL MODELING OF AN INFORMATION SECURITY SYSTEM AND ITS VALIDATION THROUGH DLP SYSTEMS

IVAN GAYDARSKI
Institute of ICT, Bulgarian Academy of Sciences, i.gaidarski@isdip.bas.bg

ZLATOGOR MINCHEV
Institute of ICT, Bulgarian Academy of Sciences, zlatogor@bas.bg

Abstract: *The objective of the present study is to outline a conceptual reference model for defining meta architecture of information security system, implementing key holistic system principles of synergy, interconnectivity and object-, agent-oriented modelling approaches, keeping at the same time opportunities for: reusability, interoperability, easy deployment and scalability. During this process of architectural development the standards IEEE 1471 and IEEE 42010 are accomplished. The current idea aims to establish a universal in some sense, meta modelling of information security systems, addressing stakeholders' multiple views, context concerns and environment. The resulting model is further validated through COTS DLP (Data Leak Prevention) system environment..*

Keywords: *Information security, Conceptual reference model, Data Leak Prevention*

1. INTRODUCTION

Nowadays, the most important asset owned by modern companies of all industries and sizes is information. Its loss, leakage or unauthorized alteration can cause serious damages, loss of competitive advantage, and even take the company out of the market. Each asset must be adequately protected, depending on the risk associated with its loss.

One of the main goals of Information Security is the protection of the information in all possible forms. Past experience has shown that Information Security (IS) measures within an organization are usually incidental based or concern compliance with certain (single) legislative regulations necessary for its normal operation.

Although there are different standards for IS (Common Standarts as ISO 27000 [1] [2], NIST "800 series" [3], ISACA's COBIT [4] [5], sector-specific regulations as the Sarbanes-Oxley Act (SOX) [6] [7], Gramm-Leach-Bliley Act (GLBA) [8], Payment Card Industry Data Security Standard (PCI DSS) [9] and Health Insurance Portability and Accountability Act (HIPAA) [10]), dealing with different aspects of the IS protection of organizations, they in fact produce a set of good practices and guidelines for their achievement.

The examples when IS is approached methodically and all requirements of the standards observed are met is usually not that common. In practice, regularly some specific information security tasks are addressed, mainly related to various incidents (leakage, attack on infrastructure, loss of information, etc.), or newly challenged tasks (such as the already adopted regulation on Protection and Processing of EU citizens' personal data (General Data Protection Regulation) [11]).

Two approaches can be used to design an Information Security System (ISS) in an organization:

- Research (scientific) approach – top - down direction. It is based on the main goals to the organization's ISS, set by the management, the compliance with various legal provisions, regulations and standards, the requirements of the users, clearly defined roles and responsibilities of the participants, the appropriate technologies, which has to be used to protect assets, according to the risk they carry. This approach allows accurate planning of the different stages between the design and implementation of the ISS, the provision of adequate budgets, roles of the participants and technologies so that everything goes according to a plan, taking into account the conflict points and minimizing possible problems. An integral part of this approach is also the provision of units, tools and processes for feedback, business continuity, loss recovery, maintenance and continuous training of the various actors in the organization's information protection processes. The end result is approved organization security policy and an acurate implementation plan with clear deadlines.
- Reverse (engineering) approach – the direction is bottom-up. Starting from the day-to-day activities of protection of the information assets, gradually on the path of quantitative accumulation some degree of protection of the organization's assets is achieved. This approach does not use systematization and it is impossible to plan, to change the methods of protection in case of change of the internal / external conditions, it is impossible to adequately spend and allocate the efforts and tasks of the participants in the information protection process.

These disadvantages can be avoided by using reference models, which represent the structure of IS in the organization, i.e. constructing an architecture of an ISS. To use the approach, first a framework for architectural modeling must be created by determining conventions, principles and practices for the description of architectures of systems of interest, established within a specific domain of interest. The architectural description of ISS expresses the fundamental organization of a system by its components, their relationships to each other, and to the

environment, and the principles guiding its design and evolution [12]. The main purpose of the framework is to codify a common set of architecture practices within a community and to help the system developers to achieve interoperability, model independence when changing external or internal conditions, easy deployment and scalability of ISS across organizations. In defining of a framework for architectural modelling of ISS the main concepts related the conceptual model of the system architecture description (defined in the standards IEEE 1471 [13] and IEEE 42010 [14]) are accomplished the following:

- Stakeholders: Parties (individuals, groups and organizations) with interests in the system that hold concerns for the system of interest;
- Concerns: A concern space would be formed from the union of all stakeholder concerns, i.e. it has the characteristics of an ecosystem;
- Viewpoint: Capturing the conventions for constructing, interpreting and analyzing a type of view that is in relation to a specific Model Kind, which provides specific forms of representation, with its own mini-meta model (What) to address one or more identified concerns (Why), via associated methods and practices (How) [15]. Viewpoint modeling is a kind of meta-modeling, whose result is a way to make models (views) of a certain kind;
- View: A representation of the whole system from the perspective of a related set of concerns – each view corresponds to exactly one viewpoint and is addressed to identify system stakeholders and answers their identified concerns;
- Architecture Model: A view is comprised of Architecture Models – each model has a model orientation which declared the purpose, context, and a governing viewpoint. Models provide a means for sharing details between views and for the use of multiple notations within a view;
- Environment: Context determining the setting and circumstances of all influences upon a system. Environment is intended in the widest possible sense to include developmental, operational, technical, and all other influences, which can affect the architecture. These influences are categorized as concerns.

An architecture framework for modelling of the architecture of an ISS is shown on Figure 1.

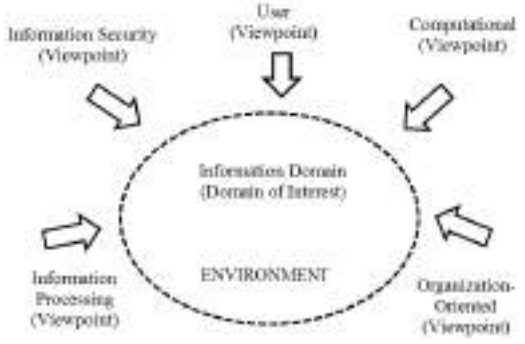

Figure1: Architecture framework for modelling of ISS

The objective of the present study is to outline a conceptual reference model for defining meta architecture of ISS, that is the first stage of their life cycle. Our approach suggests a two-layered conceptual model organized around the "Information Security" and "Information Processing" viewpoints, which are necessary for description of the architecture of an ISS. The resulting model is further validated through COTS DLP (Data Leak Prevention) system environment.

2. META MODEL OF INFORMATION SECURITY SYSTEM

There are two approaches for conceptual modelling of ISS of interest in the architectural framework sense [16]:

- Most of the frameworks provide a meta model of their intended subject matter. They attempt to cover the full range of entities in their domain of interest and to codify them once-and-for all and offer no provisions;
- An alternative approach is to construct focused and compassable meta models, organized around viewpoints or concerns.

We use the second approach and suggest a multi-layered conceptual model of ISS, organized around the viewpoints "Information Security", which will represent our first layer and "Information processing", which will present the second layer. The stakeholders that are related to the first viewpoint are developers and integrators. Some of their concerns are conceptual integrity, deployment, scalability, reusability, structure, system properties and others. The second viewpoint focuses on information processing, semantic of information and relationships between information objects.

The main questions to which an ISS has to answer from the viewpoint "Information Security" are "What", "How" and "Why", we suggest a meta model that consists of six components (see Figure 2): *Endpoint Protection*, *Communications Protection – (What)*; *Security Monitoring & Analysis*, *Security Management – (How)*; *Data Protection* and a system-wide *Security Model & Policy – (Why)*.

The *Endpoint Protection* provides defensive functionality to the endpoints (any element of ISS with computional and comunication capabilities), including identity, access control, cyber and physical security.

The *Communications Protection* component is responsible for protecting the communication between the endpoints via implementing different methods as authentication and authorization of the traffic, cryptographic techniques for integrity and confidentiality and information flow control techniques.

The purpose of the next two components – *Security Monitoring & Analysis* and *Security Management* is to preserve the state of the system by monitoring, analyzing and controlling other components of the system. The security policies are implemented by the *Security Model & Policy* component, which directs all other components to ensure the security of the data. The protection of the data in the system is represented by the last component – *Data Protection*. It covers the protected data, the system

management and configuration data and the data collected as part of the monitoring and analysis function [17].

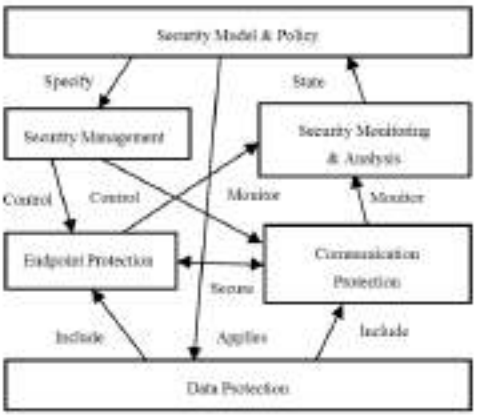

Figure 2: Meta-model of "information security" viewpoint

The data can be in one of three states at any given time: it can be at rest (on storage device), in motion (communications) or in use (processed in applications) [18]. This data model could be used for the meta-model "Information processing" viewpoint in architecture description of ISS (see Figure 3).

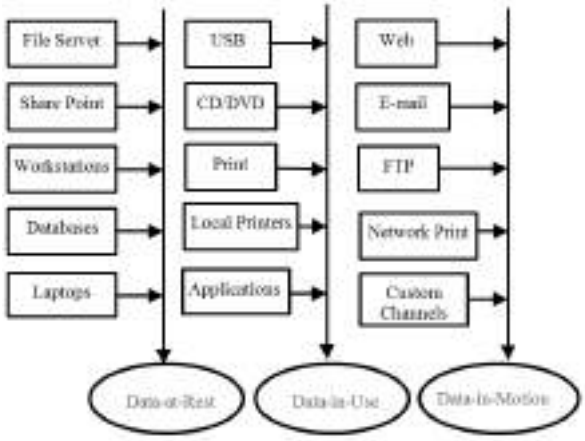

Figure 3: Meta-model of "information processing" viewpoint

This meta-model present the three types of data states that result from the work of available systems and devices:

- Data-at-Rest – Inactive data, stored in databases, workstations and laptops or archives;
- Data-in-Use – Active data, processed in application, printed or copied;
- Data-in-Motion – Data transmitted by the networks, communicated between endpoints, or moved with external storage devices.

To protect the different types of data it is necessary to be implemented specific Information Security Techniques (IST) in the basic components from the "information security" meta-model. The data have to be protected against loss, theft, unauthorized access and uncontrolled changes by applying IST such as confidentiality controls, integrity controls, access control, isolation and replication [18].

Figure 4 shows the two layers, which contain the meta-models representing "Information security" and "Information processing" viewpoints, respectively and the relations between them:

- The *Endpoint Protection* component protects Data-at-Rest and Data-in-Use of the endpoints through relevant IST, as access control and passwords, antivirus, audit trails, physical security measures and etc.;
- *Communications Protection* component protects Data-in-Motion using cryptographic techniques, network segmentation, perimeter protection, gateways, firewalls, intrusion detection, network access control, deep packet inspection and network log analysis;
- *Security Management* component protects all configuration, monitoring, and operational data, using cryptography.
- *Security Monitoring & Analysis* component is responsible for the protection of the data for current system state, the monitoring of key system parameters and indicators. The typical ISTs in this component are cryptographic techniques.

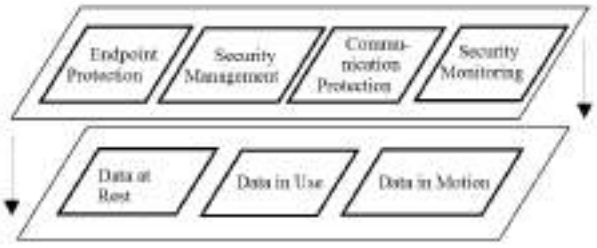

Figure 4: Two-Layered conceptual model of ISS

3. DATA LEAK PREVENTION (DLP) SYSTEMS

Data breaches have been one of the biggest dangers to the organizations today. Significant part of the breaches are happening due to leakages of information from inside to the outside. The traditional network- and infrastructure-centric security solutions are not capable to detect the content of the data and enforce protective actions based on its content, value and sensitivity. To be able to prevent data leakage from inside the protected network, the IT security solution should leverage data-centric approach [19], [20]:

- The focus must be moved from the corporate network perimeter to every workstation/laptop individually;
- The solution should control all data channels and transfer operations with contextual controls based on "Who", "How", "From", "Where", "When";
- The solution must inspect and filter the content of transferred data and to be able to control (allow, block, mitigate, audit, shadow, or alert) the data transactions.

These requirements have become the fundamentals of the Data Leak Prevention (DLP) systems.

DLP can significantly reduce the leak of sensitive information, resulting from internal threats like human error, intentional action or outside breach. The goal is to stop the data before it leaves protected environment of the organization.

Another important functionality of DLP is discovering and identifying all sensitive information in the organization in order to protect it.
DLP systems are designed to prevent attempts to steal, modify, prohibit, destroy or obtain unauthorized access and usage of the data, without interrupting normal business processes. They provide identification of the violations, threats, risks and vulnerabilities to the data, as well as violations of security policies and procedures. They may be a part of the organization's integrated IT security system.
Depending on the location of the protected data channel, the DLP systems can be:

- With focus on servers, global communications and data channels of the organization. The DLP can control email servers, file transfers from file servers, and Internet traffic filtering;
- Focused on endpoints and local data channels – workstations, laptops, mobile devices (e.g. tablets and phones). Controlled channels include all possible physical ports, personal e-mails, file transfer to cloud services and more.

The typical DLP system includes the following blocks:

- Data Channels monitoring;
- Data Channel control and management;
- Discovery module for Data-in-Rest;
- Centralized management, Analysis and Reporting Unit.

4. VALIDATION

To validate the proposed by us approach, we can use two methods for verifying the conceptual model of ISS:

- Implementation of a real-life ISS, which corresponds to our model;
- Computer simulation (using e.g. agent based modelling).

For the present study, we have chosen the first method. In a real working environment, we have installed a DLP class protection system. The reason for choosing a DLP system is that this system by definition is as close as possible to the multi-layered conceptual model of ISS, organized around the viewpoints "Information Security" and "Information processing".
Like the first layer "Information Security", the typical DLP system consists of endpoint and communications protection modules, as well as modules for monitoring and analysis, device configuration, and security policy management.
On other hand, DLP systems provide monitoring, management and controlling of all data states: Data-in-Motion, Data-in-Use and Data-at-Rest, similar to the "Information processing" layer.
For the purposes of the present study, an installation of DLP-class system by DeviceLock Inc – DeviceLock DLP Suite v.8.2 was performed for monitoring the dataflows at the endpoints (workstations and laptops) of 18 organizations, including representatives from the national security sector.
Some generalized results from the validation process are provided in Figure 5.
The empirical DLP implementation results could be aggregated as follows:

- Reduction of sensitive information leak incidents;
- Limiting data leak channels;
- Increasing of the visibility of sensitive information, by the discovery function of the DLP (Data-in-Rest);
- Improving compliance with the internal security policies, legal regulations and privacy directives.

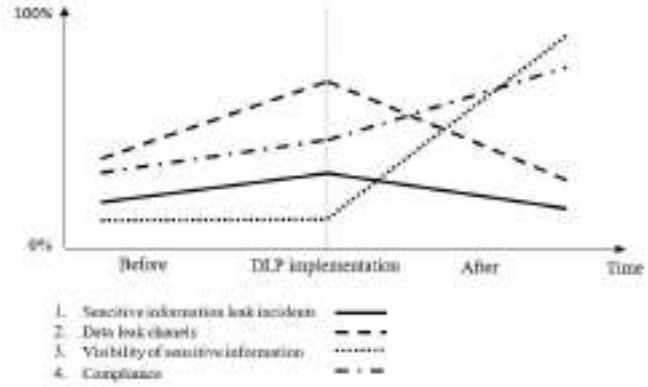

Figure 5: Generalized results from DeviceLock DLP Suite v.8.2 experimental validation

5. CONCLUSION

The main advantage of the proposed approach is the possibility for exploiting the full benefits of the research perspective, adding interoperability and ease-of-use capabilities to the implemented templates for organizations of different types and fields of work.
The conceptual modelling is presented as a starting point in architecture description of ISS that is used in definition of different viewpoints.
The goal of the architecture model construction is to ensure usability, interoperability, easy deployment and scalability of ISS across organizations. One of the most important advantages of using this approach is the achievement of model independence when changing external or internal conditions. Further, more complex architectural ISS models with extended DLP implementations are planned for development and exploration.

6. ACKNOWLEDGEMENT

This study is partially supported by a research project grant "Modelling the Architecture of Information Security Systems in Organizations", Ref. No: 72-00-40-230/10.05.2017, ICT Sciences & Technologies Panel, Program for Young Scientists and PhD Students Support – 2017, Bulgarian Academy of Sciences.

7. REFERENCES

[1] J. Hintzbergen, K. Hintzbergen, A. Smulders, H. Baars, "Foundations of Information Security Based on ISO27001 and ISO27002", Van Haren Publishing, 2010

[2] ISO 27001 Official Web Page, www.iso.org/isoiec-27001-information-security.html, September 1, 2017

[3] NIST Special Publications (800 Series): www.csrc.nist.gov/publications/PubsSPs.html, September 1, 2017

[4] IT Governance Institute, “COBIT Security Baseline: An Information Survival Kit”, 2nd ed., IT Governance Institute, 2007

[5] COBIT Resources, www.isaca.org/COBIT/Pages/default.aspx, September 1, 2017

[6] S. Anand, “Sarbanes-Oxley Guide for Finance and Information Technology Professionals”, 2nd ed, Wiley, 2006

[7] Sarbanes-Oxley Act (SOX) Resources: www.sec.gov/about/laws/soa2002.pdf, September 1, 2017

[8] Gramm-Leach-Bliley Act (GLBA) Resources: www.ftc.gov/tips-advice/business-center/privacy-and-security/gramm-leach-bliley-act, September 1, 2017

[9] PCI Security Standards: www.pcisecuritystandards.org/pci_security/, September 1, 2017

[10] R. Herold & K. Beaver, “The Practical Guide to HIPAA Privacy and Security Compliance”, 2nd ed, CRC Press, 2014

[11] EU General Data Protection Regulation, Official Web Page, http://ec.europa.eu/justice/data-protection/reform/index_en.htm, September 1, 2017

[12] R. Hilliard, “Aspects, Concerns, Subjects, Views, ... ”, in Proc. OOPSLA’99, 1999, pp. 1−4, http://www.cs.ubc.ca/~murphy/multid-workshop-oopsla99/position-papers/ws08-hilliard.pdf

[13] IEEE Std 1471, IEEE Recommended Practice for Architectural Description of Software-Intensive Systems, 2000

[14] ISO/IEC/IEEE 42010:2011 - Systems and Software Engineering - Architecture Description, https://www.iso.org/standard/50508.html, September 1, 2017

[15] R. Hilliard, “Lessons from the Unity of Architecting”, in Software Engineering in Systems Context, I. Jacobson & H. Lawson, Eds, 7 Systems, 2016, pp. 225-250

[16] J. Killmeyer, “Information Security Architecture: an Integrated Approach to Security in the Organization”, 2nd ed, CRC Press, 2006

[17] Industrial Internet of Things Volume G4: Security Framework,www.iiconsortium.org/pdf/IIC_PUB_G4_V1.00_PB.pdf, September 1, 2017

[18] M. Rhodes-Ousley, “Information Security: The Complete Reference”, 2nd ed, New York: The McGraw-Hill, 2013

[19] DeviceLock Web Page, www.devicelock.com/products/, September 1, 2017

[20] CoSoSys Endpoint Protector Web Page, www.endpointprotector.com, September 1, 2017

SECURE MOBILE BANKING BIOMETRIC AUTHENTICATION

NEMANJA MAČEK
School of Electrical and Computer Engineering of Applied Studies, Belgrade and eSigurnost Association, Belgrade, macek.nemanja@gmail.com

MILAN MILOSAVLJEVIĆ
Singidunum University, Faculty of Technical Sciences, mmilosavljevic@singidunum.ac.rs

IGOR FRANC
Belgrade Metropolitan University, Faculty of Information Technologies and SECIT Security Consulting, igor.franc@metropolitan.ac.rs

ZLATOGOR MINCHEV
Institute of ICT, Joint Training Simulation & Analysis Center, Bulgarian Academy of Sciences, zlatogor@bas.bg

MILAN GNJATOVIĆ
University of Novi Sad, Faculty of Technical Sciences, milangnjatovic@uns.ac.rs

BRANIMIR TRENKIĆ
School of Electrical and Computer Engineering of Applied Studies, btrenkic@viser.edu.rs

Abstract: *This paper presents an approach to securing mobile banking biometric authentication. The proposed system is based on secure client-server conventional XOR biometrics, which stores, transmits and verifies templates in encrypted form. Encryption keys are stored on bank's authentication servers, thus protecting the user twofold: if the phone gets stolen, both encryption keys and original templates are unavailable to an adversary. Once the user is authenticated, the communication between the client (the smartphone) and the server (bank) is encrypted. Having in mind that modern smartphones have iris scanners which operate by calculating Hamming distance and that variety of smartphones have fingerprint readers, which can, according to literature, be converted to XOR biometrics, one may conclude that the system is highly applicable and that it does not suffer from severe computational costs and drawbacks originating from cryptographic operations.*

Keywords: *Biometrics, Authentication, Cryptography, Mobile Banking*

1. INTRODUCTION

Mobile banking is a service provided by a financial institution that allows customers to conduct financial transactions, such as electronic bill payments and funds transfers, using a mobile device and software provided by the aforementioned institution. While mobile banking has it's upsides, security of financial transactions is a very important issue that needs to be addresed very carefully, as online banking is one of the most sensitive tasks performed by general user [1]. Althoug many traditional banks offer mobile baking with peace of mind [2], one should note that there is not a silver bullet providing a user with 100% security guarantee. According to [3], "a survey conducted by the Bureau of Financial Institutions found that 75 banks and credit unions' losses due to data security breaches reached a total of over $2.1 million US. This is a significant loss that financial institutions must address in order to reduce fraud rates and protect users worldwide." Jeon et al. identified three assets (which can be defined as targets of attack for mobile devices): device, application and private information [4]. Aforementioned authors defined a threat as anything that is capable of acting against an asset in a manner that can result in harm [4]. Broadly, two types of threats are identified in [5]: ones casued by external factors (adversaries) and ones caused by internal factors (user unawareness). Regarding financial transactions conducted via mobile the devices the following security apects should be addressed: physical security of the device, security of application running on the device, authentication of the user and the device to the service provider, encryption of data being transmitted and data that will be stored in device for later analysis by the customer.

Variety of authentication methods, both having upsides and downsides are implemented in mobile banking today. As an example, customers that secure data with passwords or PINs are at risk of fraud. Major companies have identified the need for strong security countermeasures and they are producing new hand-held products with built-in biometric devices. Accordig to [6], "the market size for biometrics is expected to reach $24.59 billion in the next six years and a lot of the growth will be seen from banks." According to Gartner, over 30% of mobile devices are currently using biometrics; banks should see as an opportunity rather than a barrier to adoption [7]. Although users of biometric devices do not need to remember passwords or carry tokens and biometric traits are distinctive and non-revocable in nature [8], thus offering

non-repudiation [9], one should note that biometric templates can be intercepted, stolen, replayed or altered if unsecured biometric device is connected to a network or if an adversary gains physical access to a device. This enforces the need for identity theft prevention with technological countermeasures such as cancelable biometrics, such as non-invertible transforms presented in [10, 11] and strong cryptography.

Research presented in this paper deals with authentication issue in mobile banking: precisely, cryptographically secure authentication based on conventional XOR biometrics presented in [12] is employed as mobile banking authentication system. Variety of smartphones having fingerprint readers, while devices with iris scanners are emerging technology. As fingerprint can be converted into XOR biometrics [13] and iris is verified by calculating Hamming distance and comparing it with a threshold, we can conclude that this modular system is suitable for implementation in mobile banking.

2. SECURED MODULAR AUTHENTICATION SYSTEMS WITH DISTRIBUTED STORAGE

In this section, cryptographically secured modular authentication systems based on conventional XOR biometrics with distributed storage are briefly described. Additional details on enrolment and verification phases as well as security evaluation of the system are given in [12].

System consists of one or more clients, an authentication server and a trusted storage. Client is a device used to capture biometrics, obtain auxiliary data and create encrypted cancelable templates. Authentication server manages encryption keys and verifies cancelable templates, while the trusted storage stores the encrypted templates. Two important characteristics of the proposed system are that it keeps biometric templates encrypted or cancelable during all stages of storage, transmission and verification, and that it does not suffer from severe computational costs and large sizes of encrypted templates.

3. IMPLEMENTATION IN MOBILE BANKING AUTHENTICATION SCENARIO

Authentication server resides in the bank. As authentication server stores encryption keys, it is logically that encrypted templates reside on the client. This prevents the attacker who obtains illegal access to authentication server to decrypt the temples.

The client is a mobile device (smartphone or a tablet) with fingerprint reader or an iris scanner. If the fingerprint biometrics is used, conversion to conventional XOR biometrics before cancellable template generation is necessary during both enrolment and verification phases. A system that generates XOR biometrics of fingerprints based on filterbank of Gabor filters of different spatial radial angles is presented in [13]. According to authors, the resulting fixed-length binary representation was tested in an authentication scenario with associated mechanism for extraction of associated cryptology keys, based on the principles of error correcting codes and the perspective of the proposed approach was experimentally evaluated. Additional software that provides feature extraction and cryptographic operations is installed on the client (as an additional application provided by the bank).

The non-invertible transform key is stored on the device. User obtains this key from the bank. User is allowed to wipe both the key and the data stored during enrolment phase both locally, if he suspects the data is somehow compromised, and remotely, if the device gets stolen. The bank is allowed to do remote data wiping also, if the authentication server is somehow compromised.

During the enrolment phase, client-side application calculates hash of the devices' IMEI and sends it to the authentication server. Server generates a private-public keypair (K_{priv}, K_{pub}), stores the private key with hash of IMEI (H(id), K_{priv}) and sends public key to the mobile device. User provides biometrics to the mobile device. Client-side app creates a binary template b_0 (with the aid of additional conversion if fingerprints are used) and generates cancelable binary template $b = K_t \oplus b_0$ using non-invertible transform key stored on the device. Client-side app further generates random seed s_0 and encrypts it with the public key: $s_E = E(s_0, K_{pub})$. App generates a keystream $s = PRNG(s_0)$ using pseudorandom number generator and given seed, calculates $s \oplus b$, stores values (s_E, $s \oplus b$) on the device and discards the rest of the data.

During the verification phase, hash of the device IMEI is calculated on the client-side application and sent to the authentication server. User provides biometrics to the mobile device. Client-side app creates template b_0' and generates cancelable binary template $b' = K_t \oplus b_0'$. App retrieves values s_E and ($s \oplus b$), calculates $s \oplus b \oplus b'$ and sends it with the encrypted seed s_E and hash of the devices' IMEI to the authentication server. Server retrieves private key from stored record (H(IMEI), K_{priv}) with the corresponding device IMEI hash, decrypts the seed with the private key $s_0 = E(s_E, K_{priv})$ and generates the keystream: $s = PRNG(s_0)$. Authentication server calculates $b \oplus b' = s \oplus s \oplus b \oplus b'$ and compares the Hamming distance between cancellable templates b and b' with the treshold. According to that result, the decision is made and sent back to the client. If the user is genuine, the rest of the communication between the mobile device and mobile banking authentication server is encrypted.

4. SECURITY OF THE PROPOSED SYSTEM

Regarding the security of the proposed solution, the following conclusions can be made. Templates are encrypted or at least cancelable during all stages of storage, transmission and verification, and the mobile device is not allowed to access private keys stored on authentication server. Authentication server has no access to the transform keys and cancellable templates created on the mobile device during enrolment. If the phone is stolen, an adversary cannot claim as legitimate user as the system is prone to all attacks listed in [14] as well as to hill-climbing, non-randomness, re-usability, blended substitution and linkage attack. Additionally, one should note that the user is allowed to remotely wipe all stored data if the phone gets lost or stolen.

However, one should note that security of the system also depends on the security of the biometric device itself. As

an example, a hack on Samsung Galaxy S8 iris scanner is briefly discussed. Iris patterns stable and distinctive features for personal identification [15]. More than 250 distinguishing characteristics of an iris (degrees of freedom) can be used in biometrics, resulting in six times more identifiers than the fingerprint [16]. This is why iris is sometimes referred to as an optical fingerprint. According to aforementioned, iris scanners should be hard to trick into false acceptance, but a group of hackers have managed to do so with the iris-based authentication in Galaxy S8 in an easy-to-execute attack. Hardware required to complete the attack included a digital camera, a printer and a contact lens, costing less than the unlocked smartphone. The hack required taking a picture of the subject's face, printing it on paper, superimposing the contact lens (see Image 1), and holding the image in front of the locked phone [17].

Image 1: Galaxy S8 iris-based authentication hack (still-frame taken from [18])

Despite that, the manufacturer of the iris recognition used in the smartphone still claims that iris recognition allows consumers to finally trust that their phones are protected.

CONCLUSION

This paper presented an implementation of modular authentication systems based on XOR biometrics into mobile banking. Security evaluation of the proposed system given within the paper. According to high level of security and low computational costs make it highly applicable as an authentication solution for mobile banking. Our further work will focus on implementing the system in simulated mobile banking scenario.

REFERENCES

[1] M. Mannan and P. C. Van Oorschot, "Security and Usability: The Gap in Real-World Online Banking", NSPW'07, North Conway, NH, USA, Sep. 18-21, 2007.

[2] Y.S. Lee, N.H. Kim, H. Lim, H. Jo, and H.J. Lee, "Online banking authentication system using mobile-OTP with QR-code", in Proc. 5th International Conference on Computer Sciences and Convergence Information Technology (ICCIT), 2010, November 2010, pp. 644-648, IEEE.

[3] B. Armour, "Biometric Authentication for Mobile Banking: What Banks Need to Know", Clear Bridge Mobile, available online, last time visited September 2017.

[4] W. Jeon, J. Kim, and Y. Lee, Y., "A practical analysis of smartphone security", Human Interface and the Management of, 6771, pp. 311-320, 2011.

[5] I. Ashra, "Mobile Banking Security", Vrije Universiteit, Amsterdam, Thesis number 1073, April 2012.

[6] B. Armour, "How Biometric Authentication is Shaping the Future of Mobile Banking", Clear Bridge Mobile, available online, last time visited September 2017.

[7] C. Stamford, "Gartner Says 30 Percent of Organizations Will Use Biometric Authentication for Mobile Devices by 2016", February 4, 2014, available online, last time visited September 2017.

[8] Y. C. Feng, P. C. Yuen and A. K. Jain, "A Hybrid Approach for Face Template Protection", in Proceedings of SPIE Conference of Biometric Technology for Human Identification, Orlando, USA, Vol. 6944, pp. 325, 2008.

[9] P. Balakumar and R. Venkatesan, "A Survey on Biometrics-based Cryptographic Key Generation Schemes", International Journal of Computer Science and Information Technology & Security, Vol. 2, No. 1, pp. 80-85, 2012.

[10] N. K. Ratha, S. Chikkerur, J. H. Connell and R. M. Bolle, "Generating Cancelable Fingerprint Templates", Pattern Analysis and Machine Intelligence, IEEE Transactions on, 29(4), pp. 561-572, 2007.

[11] J. Zuo, N. K. Ratha and J. H. Connell, "Cancelable iris biometric", In Pattern Recognition, ICPR 2008, 19th International Conference on (pp. 1-4), IEEE, 2008.

[12] N. Maček, M. Milosavljević, I. Franc, M. Bogdanovski, M. Gnjatović and B. Trenkić, "Secure Modular Authentication Systems Based on Conventional XOR Biometrics", accepted for publication in The 9th International Conference on Business Information Security Proceedings.

[13] S. Barzut and M. Milosavljević, "Jedan metod formiranja XOR biometrije otisaka prstiju Gaborovom filtracijom," in Sinteza 2014 - Impact of the Internet on Business Activities in Serbia and Worldwide, Belgrade, Singidunum University, Serbia, 2014, pp. 610-615.

[14] R. Jain and C. Kant, "Attacks on Biometric Systems: An Overview", International Journal of Advances in Scientific Research, 1(07), pp. 283-288, 2015

[15] S. Lim, K. Lee, O. Byeon and T. Kim: "Efficient iris recognition through improvement of feature vector and classifier". ETRI journal, 23(2), pp. 61-70, 2001.

[16] G. Amoli, N. Thapliyal and N. Sethi, "Iris Preprocessing", International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 2, No. 6, pp. 301-304, 2012.

[17] D. Goodin, "Breaking the iris scanner locking Samsung's Galaxy S8 is laughably easy", Ars Technica, May 23, 2017, available online, last time visited September 2017.

[18] Hacking the Samsung Galaxy S8 Irisscanner, online: https://media.ccc.de/v/biometrie-s8-iris-en#video&t=66, last time visited September 2017.

PERFORMANCE PREDICTION IN SECURE TELECOMMUNICATION SYSTEM WITH QUALITY OF SERVICE GUARANTEES

STOYAN PORYAZOV
Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, stoyan@math.bas.bg

DMYTRO PROGONOV
Institute of Physics and Technology, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", d.progonov@kpi.ua

EMILIYA SARANOVA
University of Telecommunications and Posts, Sofia, Bulgaria, Institute of Mathematics and Informatics, BAS, emiliya@cc.bas.bg

ZLATOGOR MINCHEV
Institute of ICT, Bulgarian Academy of Sciences, zlatogor@bas.bg

Abstract: *This paper explores a model of overall telecommunication systems, including users, terminals, and a network with Quality of Service (QoS) guaranties. Apart from GSM, BSDN and others, generalized virtual networks (VNET) with overall QoS guaranties have been considered also. In our approach, the network traffic, terminal traffic for A (calling) and B (called) terminals and users' traffic have been divided and considered separately, in their interrelationship.*
The conceptual model consists of a limited number of homogeneous terminals and users' behavior parameters, including repeated calls. The call attempt losses, considered in every service stage, are generalized.
In most cases, information transmitted between communication parties is accessible for third parties. For counteraction against channel eavesdropping in special-purpose communication systems, like commercial and governmental e-mail services, military systems, additional security communication layer is used. This layer is responsible for establishing the core secure communications services, such as cipher keys distributions, message authentication, integrity controlling etc. We extend the generalized conceptual model, of the considered telecommunication systems, by including the information protection stage, which can be activated on-demand. This stage may include hardware and software components and may cause additional delay and distortions.
The analytical model worked out, allows estimation of the influence of the security stage on the communication systems performance as well as prediction of the overall systems' QoS. The results obtained are a base for further QoS and Quality of Experience (QoE) management models.
Keywords: *Overall Telecommunication System and Network Performance, Information Protection, Quality of Service Guaranties*

1. INTRODUCTION

Our main objective is development of scalable conceptual and analytical performance models of overall telecommunication systems, allowing prediction of the values of many Quality of Services (QoS) indicators as functions of user's , network's and service's behavior

The importance of the teletraffic models, particularly of the overall QoS indicators, for Quality of Experience assessment is emphasized by Fiedler [1].

The network traffic indicators, in force [2] are not suitable for overall telecommunication system, including users. Users are shown in "Schematic contributions to end-to-end QoS" in [3], but they are not connected to the network.

The influence of the security services on the overall QoS parameters in the telecommunication systems is not considered in the available publications, due to the complexity of the problem.

In the present paper Information Security Stage (ISS) Concept is added and integrated in the Overall Telecommunication System Model [4]

2. BASE VIRTUAL DEVICES LEVEL.

In the bottom of the structural model presentation, we consider "base virtual device", which does not contain other virtual devise, in the conceptual model considered.
In the bottom of the structural model presentation, we consider "base virtual device", which does not contain other virtual devise, in the conceptual model considered. A base virtual device has the following graphic presentation and notation of its parameters:

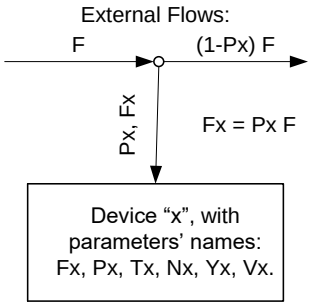

Fig. 1. Graphic presentation of a Base virtual device, named "x" and its parameters' names.

Parameter Names. Parameter Names denotations, connected with one base virtual device are (for terms definition see [5]: F – Frequency (intensity of incoming rate) of the flow of requests (requests per time unit); P – Probability of direction of the requests towards the device considered; T – Time duration of the service of a request in the device considered; Y – Traffic Intensity (Erlang).

Functional Normalization. In our models we consider monofunctional idealized base virtual devices, of the following types (Fig. 1):

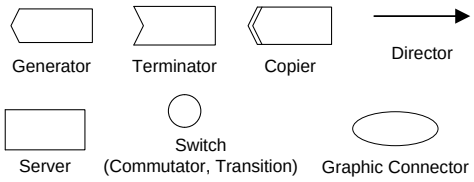

Fig. 2. Graphic block presentation of the main virtual base mono-functional devices used.

Parameters' Qualification. Traffic qualification is necessary and it is used in [5] but without any attempt for including the qualifiers in the parameters' names. Since 2006 [6] we use up to two qualifiers as a part of the parameter's name, which are used for parameters characterization

3. SERVICE PHASE

For more precise traffic characterization, in a pool of resources, we propose [4] the following definitions:

Definition 1: Served traffic in a pool of resources, is traffic, occupying (using) resources of the pool.

Definition 2: Carried traffic in a pool of resources, is successfully (effectively, completed) served traffic in the pool.

Definition 3: Parasitic traffic in a pool of resources is unsuccessfully (not effectively, not completed) served traffic. Parasitic traffic uses real resources, but not for an effective service.

In definitions 1 and 2 served and carried traffics are different, despite the ITU-T definition: "traffic carried: The traffic served by a pool of resources" ([5], Term 5.5).

4. CAUSAL GENERALIZATION

In this paper causal generalization is proposed, as an aggregation of all unsuccessful (parasitic) service cases (*prs.Ys*), from one hand, and all successful (carried (*crr.Ys*)), from the other hand.

By definition, served traffic is a sum of the parasitic and carried traffic: $srv.Ys = prs.Ys + crr.Ys$.

5. SERVICE STAGE CONCEPT

Service Stage is a service presentation containing: one service phase, realizing a function of the service; all auxiliary service phases, directly supporting the function realization, but are not parts of the function.

The intensity of the flow of offered to the stage 'g' call attempts is *ofr.Fg*, of the outgoing carried flow is *crr.Fg*, and of the parasitic served calls is *prs.Fg*.

For every service stage 'g' in the telecom system considered, we will use the following QoS indicator (*Qg*):

$$Qg = \frac{crr.Fg}{ofr.Fg} = \frac{Carried\ Call\ Attempts'\ Flow\ Intensity}{Offered\ Call\ Attempts'\ Flow\ Intensity}.$$

This indicator is inspired from [2] indicator Answer Seizure Ratio (ASR).

6. INFORMATION SECURITY STAGE

Significant part of modern business-oriented communication systems applications is providing secured channels for message and data exchanging. As example there should be mentioned Security-as-a-Service (SecaaS) business model example that takes into account person authentication in communication systems, message integrity checking, and sensitive information leakage counteraction during message transmission via open (public) channels [7].

Practical application of mentioned features requires usage of subsidiary security infrastructures – digital certificates and public key distribution managements, encryption protocol management etc. Modelling of these services requires integration the specific information security stage into proposed communication system model (Fig.2).

The security service stage (Fig. 2) contains Entry and Service phases. In the Entry Phase, B-user is asked for his/her security needs. With probability Pz (zero service) B-user is not ready to pay (with time, efforts and money) for security service. With the complementary probability (1-Pz) special process and control units (virtual devices) take care for security. The process unit is used for extracting the security related headers (e.g. certificate and key management data) from inputted data flow.

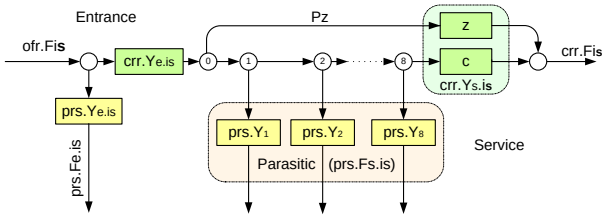

Fig. 3. Information Security (is) Service Stage with Entrance (e) and Service (s) Phases.

Establishing of secured communication channels requires cooperation of caller and called users with security infrastructure services, such as trusted third parties digital certificates management centers, public keys distributions servers etc. After successfully update and checking the security-related information by these services, communication between users can be started. In case of falling establishing the common security parameters, communications between users is cut off. The reasons for communication blocking are [8], [9], [10], [11]:

1. Failing of communication security protocol's parameters establishing;
2. Failing of message security attributes extraction;
3. Failing of message security attributes checking;
4. Invalid version of security protocol;
5. Invalid version of certificate;
6. Incorrect encryption/decryption keys;
7. Failing encryption/decryption procedure;
8. Incorrect digital signature;

Establishing the common security parameters by caller and called users can be recommence after fixed delay. In case of recurring communication blocking, value of delay is (exponentially) increased for preventing the infrastructure overloading and counteracting to deny-of-service attacks.

7. TELECOMMUNICATION SERVICE SYSTEM TRAFFIC CONCEPTS

In Fig. 4 the telecom network is presented as five service stages: A-terminal, Dialing, Switching, B-terminal Seizure and B-terminal. There are other stages in the system – included in A-User and B-User blocks, with their specifics. The service stages of call attempts, in Fig 3are:

1. Demanding Stage: Calling (A)-users generate, in a Generate Device in the A-User block, intent call attempts [5], with intensity *int.Fa* . The intensity of suppressed intent call attempts is *sup.Fa* . Suppressed traffic is "The traffic that is withheld by users who anticipate a poor quality of service (QoS) performance" [5]. "At present, suitable algorithms for estimating suppressed traffic have not been defined" [15].

The intensity of demand call attempts [5] is *dem.Fa* . A performance indicator of A-User Demanding Stage is Adir (Demand –Intent Ratio):

$$Adir = \frac{dem.Fa}{int.Fa} = \frac{int.Fa - sup.Fa}{int.Fa}$$

Adir reflects demand, intent and suppressed call attempts and corresponds to the users' anticipations of a poor QoS performance.

2. Offering stage: This is a stage in which A-user adds call attempts to the demand ones. The additional call attempts are caused by repeated (*rep.Fa*) attempts. A-user decides whether to make the parasitic call attempts, (leaving the network), repeated or to terminate those (see terminator blocks in Fig. 4).

The intensity of all offered call attempts (demand, and repeated) trying to occupy A-terminals, in Fig.3, is *ofr.Fa* A-terminals are considered as the first service stage servers, in the telecom network. From Fig. 4 follows: $ofr.Fa = dem.Fa + rep.Fa$

As a performance indicator of A-user Offering Stage, we propose Adem (Demand – Offered Ratio):

$$Adem = \frac{dem.Fa}{ofr.Fa} = \frac{dem.Fa}{dem.Fa + rep.Fa}$$

3. A-terminal Stage: In this stage A-terminals are occupied, effectively or not. The QoS indicator of A-terminal Stage (*Qa*) is:

$$Qa = \frac{crr.Fa}{ofr.Fa} \quad .$$

4. Dialing stage: The intensity of carried in A-terminals call attempts (*crr.Fa*) is equal to the intensity of the offered call attempts (*ofr.Fd*) to the Dialing Service Stage in the network – because the 'input' and 'output' are different roles of the same flow (*ofr.Fd* = *crr.Fa*). The QoS indicator of the Dialing Stage (*Qd)* is:

$$Qd = \frac{crr.Fd}{ofr.Fd}$$

5. Switching Stage: The QoS indicator of the Switching Stage (Qs) is:

$$Qs = \frac{crr.Fs}{ofr.Fs} \quad .$$

6. B-seizure Stage: The intend B-terminal may be busy or unavailable and this blocks call attempts. The QoS indicator of the B-Seizure Stage (*Qz)* is:

$$Qz = \frac{crr.Fz}{ofr.Fz} \quad .$$

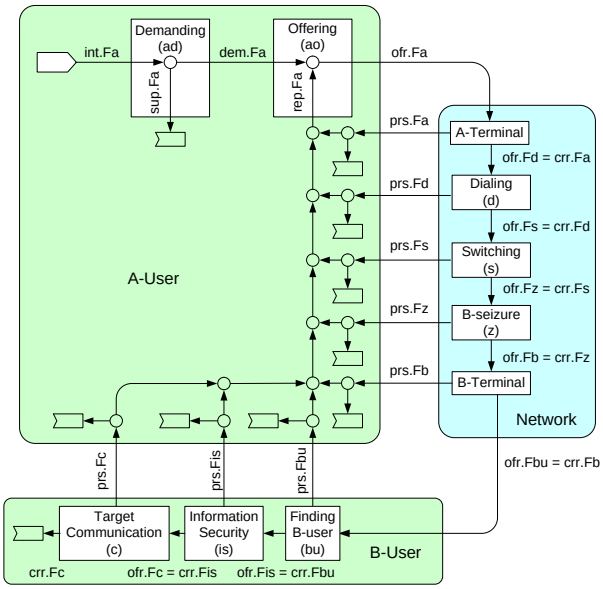

Fig.4 Schematic contributions to QoS in an overall telecom system, including users and Security Stage.

7. B-Terminal Stage: This stage corresponds to the B-terminal usage. The QoS indicator of the B-Terminal Stage (*Qb)* is:

$$Qb = \frac{crr.Fb}{ofr.Fb}$$

8. Finding B-user Stage: B-user may be absent, busy, tired etc. The QoS indicator of the Finding B-user Stage (*Qbu)* is:

$$Qbu = \frac{crr.Fbu}{ofr.Fbu}$$

9. Information Security Stage includes all necessary activities, ensuring security, if it is needed, as they are described in Section 4. The QoS indicator of the Information security Stage (*Qis)* is*:*

$$Qis = \frac{crr.Fis}{ofr.Fis} \quad .$$

10. Target Communication Stage. In this stage users exchange the target, of the call made, information. The QoS indicator of the Target Communication Stage (*Qc*) is:

$$Qc = \frac{crr.Fc}{ofr.Fc} \quad .$$

8. EFFICIENCY INDICATORS ON OVERALL NETWORK LEVEL

In our understanding, the Overall Network includes terminals and all network equipment. This means seven stages from A-Terminal to Information Security (Fig.4), inclusive. So, the QoS indicator of the Overall Network (Qnet) is:

$$Qnet = \frac{crr.Fis}{ofr.Fa}$$

Following Fig. 4 and equations in Section 7, it is checked directly that:

$$Qnet = \frac{crr.Fis}{ofr.Fa} = Qa\ Qd\ Qs\ Qz\ Qb\ Qbu\ Qis$$

9. EFFICIENCY INDICATOR ON OVERALL SYSTEM LEVEL

The Overall System Level includes users and all stages from intend call generation to fully successful completed communication. In terms of Fig.4, this means ten stages from Demanding to Target Communication, inclusive. Hence, the QoS indicator of the Overall Telecommunication System (*Qsys*) is:

$$Qsys = \frac{crr.Fc}{int.Fa}$$

Following Fig. 4 and equations in Section 7, it is checked directly that:

$$Qsys = \frac{crr.Fc}{int.Fa} = \frac{Adir}{Adem} Qa\ Qd\ Qs\ Qz\ Qb\ Qbu\ Qis\ Qc$$

The presented above indicators are flow-oriented. They are a base for time and traffic indicators construction, using the Theorem of Little and duration of service data. A step towards such indicators is made in [14].

10. NUMERICAL PREDICTION EXAMPLES OF THE OVERAL PERFORMANCE INDICATORS

We consider a model of Software Defined Network or virtual network (VNET) carrying Traffic Class 0. The VNET is with virtual channels switching, following the main method for traffic QoS guarantees – resource reservation.

The model is with: BPP (Bernoulli–Poisson–Pascal) input flow; repeated calls; limited number of homogeneous terminals; without suppressed call attempts; The calling (A) and called (B) terminals and users are considered separately, but in their interaction.

In Figures 5, 6 and 7, numerical results are presented using analytical and computer models, built following the methods explained in [14].

The numerical results are in the whole theoretical network load interval – overall terminal traffic of the A and B terminals equals of 0% to 100% of the number of all terminals in the network. The input parameters in the three figures are the same, excluding: (i) capacity of the network given as percentage of the number of all terminals in the system, causes blocking due to equipment insufficiency and (ii) the probability of repeated call attempts. The values of input parameters, in the presented numerical results, are typical for voice networks. Three cases have been considered:

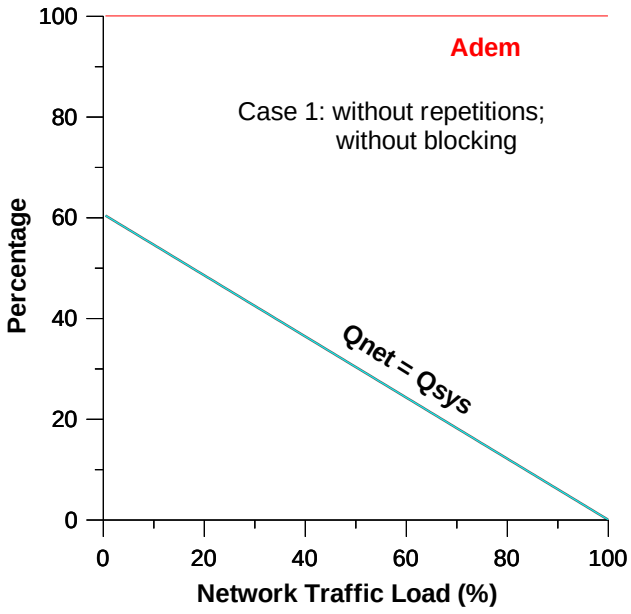

Fig. 5. Call Efficiency indicators Adem, *Qnet* and Qsys in Case 1. Adem = 1 and *Qnet* = Qsys, because there are not repeated attempts in the system. Adem is a constant 1. *Es* and *Ed* are decreasing monotonic functions.

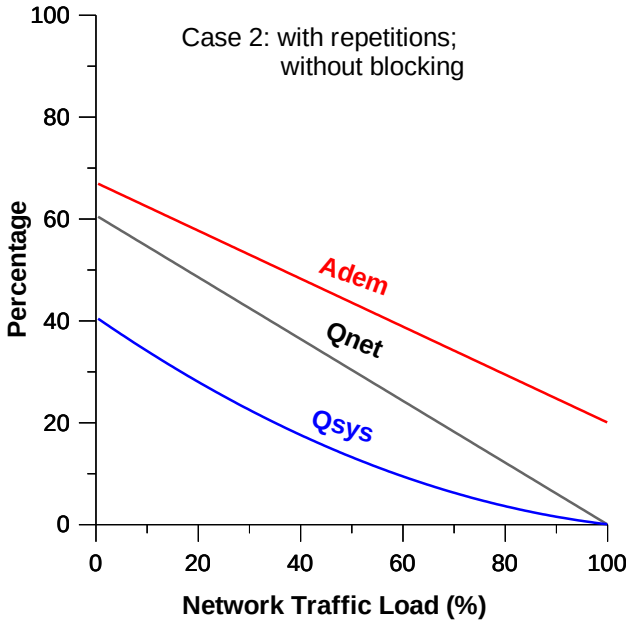

Fig. 6. Call Efficiency indicators Adem, Qnet and Qsys in Case 2. Repeated attempts make worse the performance considerably. Adem, Qnet and Qsys, are decreasing monotonic functions.

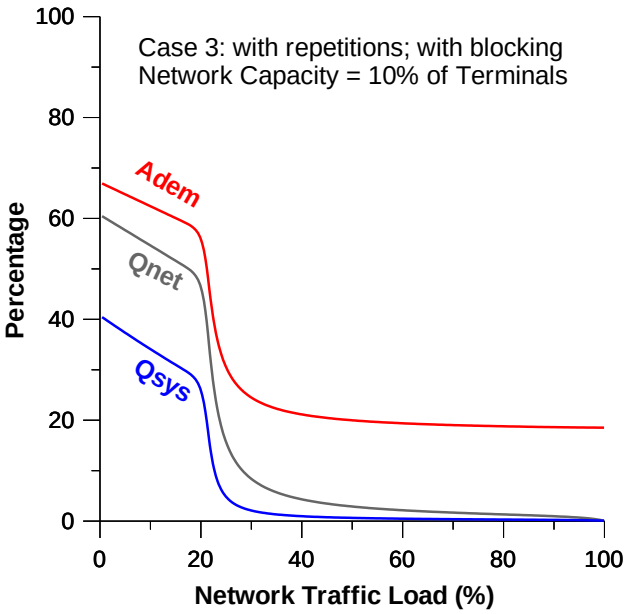

Fig.7s. Call Efficiency indicators Adem, Qnet and Qsys in Case 3. Blocking sharply make worse the system performance.

11. CONCLUSION

- The approach, explained, allows QoS presentation of every portion of the overall telecommunication system, consists composition of one or more consecutive stages. This may be used for stage performance comparison and more targeting QoS and QoE management.
- Introducing the security service stage give us opportunity to model not only public (open), but also secured communication channels between users. Taking into account the additional information flows, related to communication with trusted third parties, as well as increasing the traffic between users (adding the security related headers) allows increasing the precision of quality-of-service indicators estimations for real systems. The approach proposed is a step towards Security as a Service (SecaaS) practice.
- The QoS indicators considered are defined in terms of termination or continuation of call attempts' service, disregarding the reasons for this. An important task is modeling the QoS degradation factors as noise, distortions and others.

ACKNOWLEDGMENTS

This work is coordinated under EU COST Action IC 1304. The work was partially funded by Bulgarian NSF Projects DCOST 01/9 (work of S. Poryazov), and Bulgarian NSF Project DCOST 01/20 (work of E. Saranova).

REFERENCES

[1] M. Fiedler. Teletraffic models for Quality of Experience assessment. Tutorial at 23rd International Teletraffic Congress (ITC 23), San Francisco, CA, Sept. 2011. http://i-teletraffic.org/_Resources/Persistent/9269df1c3dca0bf58ee715c3b9afabbc71d4fb26/fiedler11.pdf (Accessed on 20.07.2017)
[2] ITU-T Rec. E.425 (03/2002). Internal automatic observations.
[3] ITU-T Recommendation E.800 (09/08), Definitions of terms related to quality of service.
[4] Stoyan Poryazov, Emiliya Saranova, Ivan Ganchev. Conceptual and Analytical Models for Quality of Overall Telecommunication Systems with QoS Guarantees Prediction. In: Ivan Ganchev, Rob van der Mai, J.L. (Hans) van den Berg (Editors). Autonomous Control for a Reliable Internet of Services: Methods, Models, Approaches, Techniques, Algorithms and Tools. Springer, LNCS, State-of-the-Art Surveys, 2018 (In print).
[5] ITU-T Recommendation E.600 (03/93), Terms and definitions of traffic engineering.
[6] S. A. Poryazov, E. T. Saranova. Some General Terminal and Network Teletraffic Equations in Virtual Circuit Switching Systems. Chapter 24 in: A.Nejat Ince, Ercan Topuz (Editors) Modeling and Simulation Tools for Emerging Telecommunications Networks. Springer Sciences+Business Media, LLC, USA 2006, pp. 471-505. ISBN-10: 0-387-32921-8 (HB).
[7] Furfaro, A.; Garro, A.; Tundis, A. (2014-10-01). "Towards Security as a Service (SecaaS): On the modeling of Security Services for Cloud Computing". 2014 International Carnahan Conference on Security Technology (ICCST): 1–6. doi:10.1109/CCST.2014.6986995
[8] RFC 2401. Security Architecture for the Internet Protocol. 1998/11
[9] RFC 4716. The Secure Shell (SSH) Public Key File Format. 2006/11
[10] RFC 5246. The Transport Layer Security (TLS) Protocol version 1.2. 2008/08;
[11] National Institute of Standards and Technology (December 2010). "Implementation Guidance for FIPS PUB 140-2 and the Cryptographic Module Validation Program"
[12] RFC 4256. Generic Message Exchange Authentication for the Secure Shell Protocol (SSH). 2006/01
[13] RFC 5878. Transport Layer Security (TLS) Authorization Extensions. 2010/05
[14] Poryazov S., E. Saranova. Models of Telecommunication Networks with Virtual Channel Switching and Applications. Prof. Marin Drinov, Academic Publishing House, 2012, pp. 238. ISBN 978-954-322-540-8.
[15] [ITU-T Recommendation E.501] ITU-T Recommendation E.501: Estimation of Traffic Offered in The Network. (26th of May 1997).

BIG DATA AND CYBER SECURITY: CONTEMPORARY ISSUES

DRAGAN MITIĆ
Faculty of Information Technologies, Metropolitan University (Niš, Serbia), and OpenLink Group (Belgrade, Serbia), dmitic@openlink.rs

MILOŠ JOVANOVIĆ
Middlesex University London (London, United Kingdom), Faculty of Organizational Sciences, University of Belgrade (Belgrade, Serbia), Faculty of Informatics and Computing, Singidunum University (Belgrade, Serbia), and Openlink Group (Belgrade, Serbia), mjovanovic@openlink.rs

NENAD BIGA
Graduate School of Business, La Salle University (Philadelphia, United States), and JPMorgan Chase & Co. (New York, United States), nenad.biga@jpmchase.com

NEMANJA ĐAKOVIĆ
Faculty of Political Sciences, University of Belgrade (Belgrade, Serbia), and the Center for International Relations and Sustainable Development – CIRSD (Belgrade, Serbia), nemanja.djakovic@cirsd.org

ALEKSANDAR PETROVIĆ
Faculty of Informatics and Computing, Singidunum University (Belgrade, Serbia), and SECIT Security (Belgrade, Serbia), apetrovic@secitsecurity.com

Abstract: *Maintaining electronic data security at the highest level is an imperative in the modern business environment. In the exploration of new ways to improve defence, through making reports from analysis of attacks and defence strategies, the new idea was born – to use big data for assembling those reports. Big data enables early detection of attacks and tracking down real adversary. The problem addressed within this paper is how big data is being used and how it will be developed in the future. Implementation of the big data within the context of cyber security can only intensify during the following years, and the solution for the problem is still being built. There still exist a wide variety of areas which are yet to be identified as prosperous in terms of big data application.*

Keywords: *analytics, security, improvement, detection, prevention*

1. INTRODUCTION

Possibilities for exploiting big data are growing and cyber security is set to benefit from it. Starting from the beginning of the new millennium, the role of big data was evolving from monitoring websites and social networks, to more complex and socially significant ways of implementing, e.g. healthcare, scientific experiments [1], and cyber security. Big data can both enhance the cyber defence and harm it. Therefore, before starting to improve security with huge amounts of information, the information that has been already gathered needs to be protected. Due to the fast development pace of big data, there is a strong need for keeping up with the innovations. Conveniences of big data are being discovered daily and its utilization rates are on the rise. Certain problems appeared in pursuing the increase in efficiency of data used, that need to be tackled with new solutions.

2. BACKGROUND

Daily amounts of information exchanged are enormous and the trail individuals leave is more than considerable. Presence of electronic devices in everyday life is massive. Every device that requires an exchange of information can collect data or submit it, but the emphasis is not on how much data it collects but on the quality of it. Creating the web with the goal of catching as much solid data as possible requires effort and considerate expertise. Some general guidelines for avoiding misuse of gathered information are: proper collection of information, having an expert in the addressed field analysing it, and ensuring not to fail to consider the factors such as the volume, velocity, variety, and variability of the information [2]. The complexity of information is a new specification that seeks further attention. It is required to extract essential information and data patterns from the large quantities of data and to correlate these to other parameters that are being taken into account. Correlated relationships, hierarchies, and multiple data linkages are the keys of well-organized frameworks, all of which are susceptible to external influences.

Compromising any information or a role of another persons' computer, hand-held device or an IoT device can be addressed to as cyber-attack. Cyber-defence represents the protection from such attacks and it is not one-time fight. Rather, it is a continuous war with threats originating from different sources one has to fight for to win. Based on the type of damage the adversary wants to achieve, there are

more than just one type of an attack. The most common ones are malware, phishing, a variety of attacks on passwords (e.g. brute-forcing, rubber-hosing, rainbow table attacks or key-loggers), Denial of Service (DoS) attacks and drive-by downloads [3]. Privacy adversarial actions address data or identity theft, while destructive are ones, such as blackmailing (e.g. adversary-recoverable deletion, such as ransomware based encryption), secure wiping or making data unreachable in any other way as long as an adversary can achieve. These actions are hell-hammering the privacy, integrity and availability of the information.

It shouldn't be underestimated that the primary defence of any organization is a common sense [4]. Awareness of what could jeopardize the system needs to be increased between every person in touch with any part of it. Regularly patching firewalls, updating firmware and setting strong passwords became essentials of cyber security.

Additionally, there is a new star in the field of cyber security improvements – big data.

3. THE CURRENT ROLE

New ways of improving cyber security are constantly being discovered and one of the most promising ones is improving detection through the complex analytics of information from different sources. Assembling a complex, analytical and centralized infrastructure with high-speed ingestion and ability to identify changes in used patterns, that is efficient enough to generate reports as close as possible to real time is the right way of improving cyber security. Application logs, network events, and user activities are fundamentals of advanced analytics. The organization is crucial in receiving large amounts of current and historical data. This information is enhanced by implementing additional context data and external threat intelligence. Apache Hadoop and inexpensive hardware are the main contributing factors to this possibility, as they also enable clients to store and analyse huge amounts of unstructured data in real time. The anatomy of big data enhanced security is machine learning, text mining and ontology modelling that would be effective in predicting, detecting, deterring and preventing the attacker from having his way. The possibility of automating the disrupting of malware attacks is a very welcome idea and a future goal of improving cyber security.

Intrusion detection is a part of the defence that got better with big data. It improves it by posting red flags on possible breaches and minimizing the damage. Prevention of an attack can be achieved but the likeliness of it makes determining the identity of the attacker more important. Other than holding someone responsible, it is crucial to find out if any data was stolen, what was stolen and if something was damaged by the attack. Systems that hold sensitive information highly require this to ensure the clients safety and if they can continue to use the service with the same credentials. Diversity of the data used for analyses can improve detection but it can also worsen it if too many features have been used. Features have to be properly selected and changed as new types of attack emerge. One of the methods is to remove misleading data that will improve data classification time [5]. Sets of features are an important part of the peer to peer network detection framework along with the traffic sniffer module and machine learning. These systems were made for detecting botnet attacks and the core of them is Apache Mahout that is used for building predictive modules. Traffic sniffing collects the data into successive PCAP files of a specified size using the capture ring buffer option, feature extraction continues with Apache Hive enabled feature selection that helps picking relevant ones for the problem and machine learning is at the end using Apache Mahout for model training and classification purpose [6]. Big data security consists of keeping safe the data uploaders, data users, and society.

4. CHALLENGES

Currently, known limits to using the full potential of big data are for some users' resource wise but those who are not limited materialistically lead the battle against the legal system as it is because of the people that abused it. Latterly data privacy was a controversial subject and still is to those that it matters to. Information on one's habits or interests always holds a value to someone. Product placement or keeping track of someone in either way are serious violations if they are not approved in some way by the law or if the person was not aware and did not agree to that type of data sharing. Hidden data and metadata have been discovered in PDF files, further increasing the security risks and complicating the data gathering [7]. Not collecting relevant data is another problem that needs to be tackled by good organization and planning ahead, but if the inadequate person is writing the reports based even on the correct information everything falls apart. Being a good analyst and informed on the matter are an imperative traits of a person fit for the job. There will always be people who will not even care to give it a try. This to blame are low awareness of the benefits and lack of investments in this sector.

5. IMPACT

One out of five businesses implemented big data in their security systems, but more than half of them experienced huge benefits and all of them observed some [8]. Currently, the most talked about new trend in informational technologies is big data [8], while being around since the second half of the 20th century, its popularity started rising not too long ago. Importance of big data security analytics is ascending and in the near future it will be necessary for preventing attacks. Currently, the biggest challenge is bridging the gap between the perceived importance of big data security analytics and actual implementation. Big data made it into a lot of topics that include improving cyber security, now it has to be heavily implemented and to increase its usage.

6. FUTURE TRENDS

Promising new field of operation for gathering data is the deep web. Nature of this network does not allow putting any number on its size and only given fact is that it is a lot

bigger than the surface web. The use of big data analytics in deep web is extraordinarily difficult because it resists classification and quantification. Considering its comparison in size we come to a conclusion that it holds larger amounts of information. This makes it a gold mine for someone withy enough to gather it [10], since the deep web does not operate the same way as every other network, meaning that the information that can't be found on the surface will be buried deep but still waiting for someone interested enough to find it.

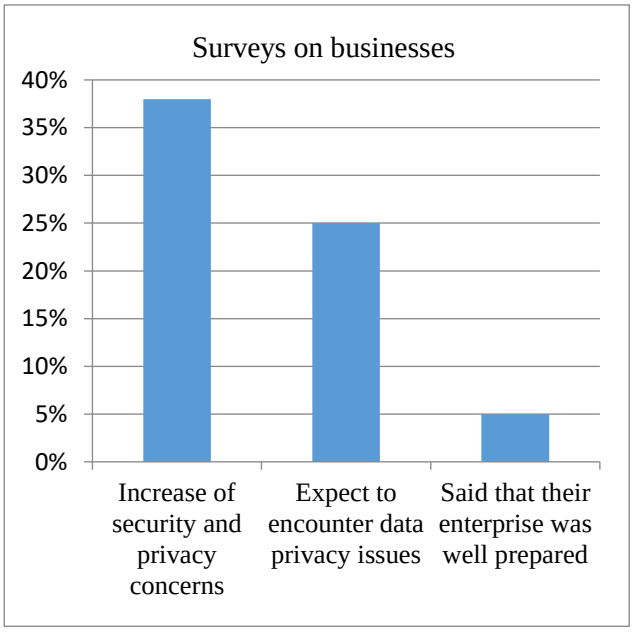

Image 1: Graphical representation of peoples' response to big data [9]

Cloud storage is following up to big data as an IT trend, taking up third important spot right after IT security. Gains from using cloud storage are not just equipment wise. Many cloud services provide valuable assets like machine learning, advanced analytics, content serving, archive and disaster recovery [11]. To possess these tools upgrading the network infrastructure or acquiring third-party data migration tools would otherwise be required. Combining cloud storage and big data is for the majority of businesses considered an advanced technique and is yet to be explored.

Future of harvesting more data holds in connecting everything and making everything online. Smart homes and cars are an example of mass data users and can be exploited in collecting it and putting it to good use. Not only this expensive and big things are getting connected. Smart watches were one of breakthrough items that got connected to smartphones. They collect data on user's daily activities and keep track of his health. The goal is to make smart cities [12] in which every part of them will be connected to a network of some sort and communicate with each other. This is beneficial because it could reduce waste of power and lower pollution.

To make the analysis from stockpiled data available sooner is starting to become an obstacle that seeks consideration. To those who results from analysing matter need them to be available and actionable faster. The problem lays in time elapsed to write reviews. To make this easier certain tools had been made and more of them are in making. Currently available are Jaspersoft BI Suite, Pentaho Business Analytics, Karmasphere Studio and Analyst, Talend Open Studio, Skytree Server, Tableau Desktop and Server and Splunk [13]. To invest in such software is essential and as a subfield of big data, it has great potential to grow.

7. CONCLUSION

From the tools available with big data stored in the cloud, cyber security gets a highly efficient combination of flexibility and accuracy. Vast options for optimization of analytics levelled the cyber warfare. Problems that emerge can be solved by raising public consciousness of the benefits of big data. Enhancing security systems with it is expensive, but if the society adopts it as a standard, like everything else when it hits wholesale the prices of it will drop. Recognizing the opportunity at this moment will pay off when the early responders develop their security systems to the maximum. Those who chose to sleep on this advancement will miss out a great part of cyber security and will tackle problems following up to the standards that future holds. Unquestionably finding its place, big data made a huge impact on cyber security. Having all the possible information was known to be the best since ages but the trend of using it most efficiently and quickly as possible is taking its place in the cyber world.

REFERENCES

[1] H. Han, Y. Wen, T-S. Chua and X. Li, "Toward Scalable Systems for Big Data Analytics: A Technology Tutorial", IEEE Access, 2, pp. 652-687, 2014.

[2] P. Zikopoulos and C. Eaton, "Understanding big data: analytics for enterprise class hadoop and streaming data", New York: Osborne/McGraw-Hill, 2012.

[3] "8 Types of Business Cyber Attacks In Business", [online] available at: https://quickbooks.intuit.com/r/technology-and-security/8-types-of-cyber-attacks-your-business-needs-to-avoid/, QuickBooks, last time visited, Aug. 25, 2017.

[4] C. Scott, "How to stop cyber-attacks on your organisation", The Guardian, [online] available at: https://www.theguardian.com/public-leaders-network/2015/oct/14/how-to-stop-cyber-attacks-on-your-organization, last time visited: Aug. 26, 2017.

[5] R. Zuech, T. Khoshgoftaar and R. Wald, "Intrusion detection and Big Heterogeneous Data: a Survey", Journal of Big Data, 2(1), p.3., 2015

[6] K. Singh, S. Guntuku, A. Thakur and C. Hota, "Big Data Analytics framework for Peer-to-Peer Botnet detection using Random Forests", Information Sciences, 278, pp. 488-497, 2014.

[7] National Security Agency, “Hidden Data and Metadata in Adobe PDF Files: Publication Risks and Countermeasures”, 9800 Savage Rd. STE 6704 Ft. Meade, MD 20755-6704: Enterprise Applications Division of the Systems and Network Analysis Center (SNAC) Information Assurance Directorate.

[8] N. Janoschek, “Big Data Security Analytics: A Weapon Against Cyber Security Attacks?” [online] BI Survey, [online] available at: https://bi-survey.com/big-data-security-analytics, last time visited: Sep. 1, 2017.

[9] N. Kshetri, “Big data′s impact on privacy, security and consumer welfare”, Telecommunications Policy, 38 (11), pp. 1134-1145, 2014.

[10] B. Marr, “Big Data and the Deep, Dark Web”, [online] available at: http://data-informed.com/big-data-and-deep-dark-web/, last time visisted Sep. 1, 2017.

[11] T. Coughlin, “Moving Data To The Cloud”. Forbes, [online], available at: https://www.forbes.com/sites/tomcoughlin/2017/07/19/moving-data-to-the-cloud/#4dbe6c2c567c, last time visited Sep. 1, 2017.

[12] J. Morgan, “A Simple Explanation Of 'The Internet Of Things'” Forbes. [online], available at: https://www.forbes.com/sites/jacobmorgan/2014/05/13/simple-explanation-internet-things-that-anyone-can-understand/#4673b6cd1d09, last time visited Sep. 1, 2017.

[13] P. Wayner, “7 top tools for taming big data. [online] [online], available at: https://www.infoworld.com/article/2616959/big-data/7-top-tools-for-taming-big-data.html, last time visited Sep. 1, 2017.

The 9th International Conference on Business Information Security (BISEC-2017), 18th October 2017, Belgrade, Serbia

METHODOLOGICAL PITFALLS OF AUTOMATIC SPEECH RECOGNITION

MILAN GNJATOVIĆ
Faculty of Technical Sciences, University of Novi Sad, milangnjatovic@uns.ac.rs

NEMANJA MAČEK
SECIT security consulting; Faculty of Engineering Management, Union University, Belgrade, macek.nemanja@gmail.com

ZLATOGOR MINCHEV
Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, zlatogor@math.bas.bg

Abstract: *The broad surveillance potential of automatic speech recognition has been recognized across a range of application domains. Although considerable research effort has been devoted to the research question of achieving robust large-vocabulary continuous-speech recognition, it is still a fragile technology. In this paper, we discuss a methodological desideratum in this field. The currently dominant approaches to speech recognition relate to Bayesian statistical inference methods based on hidden Markov models and n-grams, or, ever-increasingly, neural networks. The common point of these approaches is that they are primarily corpus-driven and thus essentially agnostic of a broader interaction context, which represents a methodological limitation. Finally, this paper provides an overview of selected aspects of our recent research on natural language processing aimed at overcoming these methodological shortcomings.*

Keywords: *context-dependent speech recognition, focus tree, hybrid approach*

1. INTRODUCTION

The specification and design of automatic speech recognition systems involve different parameters of variation (including the vocabulary size, the speech fluency, the noise level, and the speaker-class characteristics), but the most important applications of this technology relate to large-vocabulary continuous-speech recognition. Such systems are intended to recognize spontaneous speech from previously unknown people, in realistic conditions [1]. However, although large research effort has already been invested in this field, the state-of-the-art speech recognition systems are still too restrictive, showing considerable error rates when applied in adverse conditions [2]. Most errors occur at the signal level, when the user's utterance was not correctly recognized although it was within the domain, scope and grammar of the recognition system [3]. These recognition errors are usually attributed to inadequate acoustic or language modelling. Complementary to such views, we discuss that a serious methodological limitation of the currently dominant approaches to speech recognition lies at the methodological level, as they do not account for a broader interaction context.

The paper is organized as follows. Section 2 provides a brief overview of the conceptualization of context in statistical pattern-matching approaches. Section 3 discusses on why language corpora alone do not suffice to produce reliable speech recognition systems in a general case. An overview of selected aspects of our recent research on a hybrid approach to automatic speech recognition is provided in Section 4. The paper ends with Section 5.

2. CONTEXT IN STATISTICAL PATTERN-MATCHING PARADIGM

The statistical approaches to speech recognition relate to Bayesian statistical inference methods based on hidden Markov models and n-grams. The recognition task is conceptualized as finding the most probable word sequence $\widehat{W}$ for an acoustic signal X, in a given search space L:

$$\widehat{W} = argmax_{W \in L} P(X|W)P(W).$$

Probability $P(X|W)$ is the observation likelihood derived from an underlying acoustic model. Words contained in a vocabulary are typically represented as hidden Markov models whose states express phone-like units. A widely accepted approach to large-vocabulary speech recognition is to apply context-dependent phone models, e.g., triphone hidden Markov models that represent phones in a particular left and right contexts.

Probability $P(W)$ is the prior probability derived from an underlying language model. It estimates the probability of a given word sequence $W = w_1 w_2 \dots w_n$ by using n-gram models (most often bigrams and trigrams, due to practical reasons):

$$P(w_1 w_2 \dots w_n) = \prod_{i=1}^{n} P(w_i | w_{i-1} \dots w_{i-N+1}).$$

Without going further into formal and practical details (for these, the reader may consult [1,4]), it is important to note that acoustic and language models include contextual information only at the phone and sentence levels,

respectively. This very limited account of context is the reason why speech recognition systems based on this paradigm lack the robustness towards unexpected topic shifts, noisy conditions, etc. Therefore, the following questions may be raised here: Are these technical deficiencies due to inappropriate corpora used to train acoustic and language models, rather than to the methodological approach? Could a more comprehensive corpus provide enough data to address these technical deficiencies? The answer is negative, as discussed in the next section (cf. also [5]).

3. WHY LANGUAGE CORPORA ALONE DO NOT SUFFICE?

Language corpora have undoubtedly an important role in the design of speech recognition systems. However, their role tends to be overstated in research practices. Thus, Chomsky strongly criticizes the prevalent understanding "that the only real object is a corpus of data and that by automated analysis … one can derive everything that's relevant about the language" [6]. He describes it as "a novel concept of science that has emerged in the computational cognitive sciences and related areas of linguistics" [6]. With this concept, "an account of some phenomena is taken to be successful to the extent that it approximates unanalyzed data" [6]. Still, it has been widely adopted in recent language acquisition studies [7,8] and statistical approaches to machine learning.

Leaving this intellectual divide aside, there is a consensus that a language corpus should be representative, balanced, appropriately sized, etc. Nevertheless, these criteria are too vague and observer-relative. Although it is clear what representativeness of a corpus should mean, questions like "how do we identify the instances of language that are influential as models for the population?" still do not have definite answers [9]. In fact, we have no means to ensure or even evaluate the representativeness of a corpus [10, p. 57]. Similarly, a corpus is "pronounced balanced when the proportions of different kinds of text it contains should correspond with *informed* and *intuitive* judgments" [9] (emphasis added by author). A rigorous definition of these criteria is not to be expected soon.

In addition, speech recognition corpora usually contain recordings of utterances isolated from an interaction context, or telephone conversations (e.g., collected from a call centre, etc.) whose structure is objective-driven and thus not representative. In general case, the dialogue structure is not given beforehand, but evolves as the conversation unfolds [11,12]. Therefore, it is not even possible to produce a language corpus that would contain all relevant dialogue phenomena [13].

4. FROM RECOGNITION TO UNDER-STANDING

To overcoming the above methodological shortcomings, we apply a hybrid approach to speech recognition, incorporating both symbolic and statistical approaches. On the symbolic side, we refer to the focus tree model [14-18]. It is a symbolic and cognitively inspired model of attentional information in human-machine interaction that addresses the problem of robust recovery of semantic information from spontaneously uttered user's commands without explicit syntactic expectations. This model integrates three lines of research:

- the neurocognitive understanding of the focus of attention in working memory,
- the notion of attention related to the theory of discourse structure in the field of computational linguistics,
- the investigation of a corpus that comprises recordings of spontaneous speech-based human-machine interaction.

A corpus-based investigation of user commands resulted in the following findings:

- Propositional content is expressed by frequent insertion of chunks that explicitly relate to entities from the currently salient focus space. We refer to these parts as to focus stimuli.
- At the surface level, focus stimuli are non-recursive phrases. However, at the level of dialogue structure, they carry information about the attentional state.
- The order of focus stimuli within an utterance is flexible, while the word-order within them is rather fixed.
- Interaction participants often share a non-linguistic context. Therefore, speakers sometimes intentionally omit to explicitly utter information related to the attentional state because they believe it is already known by the interlocutor.
- In the strategy for recovering from non-understanding, speakers try to help the interlocutor by explicitly referring – using a constituent negation – to entities from the current interaction domain that should not be in the focus of attention.

In other words, attentional information is clearly signalled in the spontaneously produced user commands, and that it may be used to robustly recover semantic information without introducing explicit syntactic expectations [14,15]. In addition, the focus tree model includes cognitively-inspired and context-dependent evaluation of the retrieval cost and the integration difficulty of user's dialogue acts [18].

For the purpose of improving speech recognition performances, the focus tree model is used for post-processing of recognition hypotheses. The main idea of the proposed hybrid approach is that speech recognition hypotheses obtained by a standard statistically-based speech recognizer are further evaluated with respect to their retrieval cost, integration difficulty, and lexical matching. The system reduces the set of recognition hypotheses in an iterative manner, according to the following criteria [4]:

- The system first selects the recognition hypotheses with the minimum semantic integration difficulty in the given context. It there are more than one such a hypothesis, the second criterion is applied.
- From the recognition hypotheses selected in the previous step, the system selects hypotheses that

are most informative in the given context, i.e., provide the maximum retrieval cost. If there are again more than one such a hypothesis, the third criterion is applied.

- From the recognition hypotheses selected in the second step, the system selects hypotheses with the maximum lexical matching with respects to the system's vocabulary. If there more than one such a hypothesis, the last criterion is applied.
- From the recognition hypotheses selected in the third step, the system selects the recognition hypothesis that is assigned the highest probability by the standard, statistically-based speech recognition system.

For a detailed algorithm and a discussion on a prototype system, the reader may consult [4]. For the purpose of illustration, Table 1 summarizes the evaluation results, showing that the hybrid speech recognizer integrating both statistical and symbolic approaches outperforms the statistically-based recognizer.

Table 1: Evaluation results [4].

Parameter	Statistical approach	Hybrid approach
Number of sentences	980	980
Word error rate (%)	7.4	1.2
Sentence error rate (%)	38.6	3.5

5. CONCLUSION

Although considerable research effort has been devoted to the research question of achieving robust large-vocabulary continuous-speech recognition, it is still a fragile technology. This paper briefly discussed a methodological desideratum in this field. The common point of currently dominant approaches to automatic speech recognition is that they are primarily corpus-driven and thus essentially agnostic of a broader interaction context, which represents a methodological limitation. This paper provided an overview of selected aspects of our recent research on natural language processing aimed at overcoming the methodological shortcomings.

Acknowledgments

The presented study was sponsored by the Ministry of Education, Science and Technological Development of the Republic of Serbia (research grants III44008 and TR32035), and by the intergovernmental network EUREKA (research grant E!9944). The responsibility for the content of this article lies with the authors.

REFERENCES

[1] D. Jurafsky and J.H. Martin, "Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition", second edition, Prentice-Hall, 2009.

[2] C.H. Lee, "Fundamentals and Technical Challenges in Automatic Speech Recognition", In Proceedings of the XII International Conference "Speech and Computer" (SPECOM'2007); Moscow State Linguistic University, Moscow, Russia, pp. 25-44, 2007.

[3] D. Bohus and A.I. Rudnicky "Sorry, I Didn't Catch That! An Investigation of Non-Understanding Errors and Recovery Strategies", In: Dybkjær L., Minker W. (eds) Recent Trends in Discourse and Dialogue. Text, Speech and Language Technology, vol 39. Springer, Dordrecht, pp. 123-54, 2008.

[4] D. Mišković, M. Gnjatović, P. Štrbac, B. Trenkić, N. Jakovljević, V. Delić "Hybrid Methodological Approach to Context-Dependent Speech Recognition", International Journal of Advanced Robotic Systems, Vol. 14, No. 1, 2017.

[5] M. Gnjatović, "Do Language Corpora Suffice to Develop Conversational Agents?", Zbornik radova sa 9. konferencije Digitalna obrada govora i slike, DOGS 2012, ISBN 978-86-7892-439-2, Kovačica, Serbia, pp. 24–27, 2012.

[6] N. Chomsky, "Language and the Cognitive Science Revolution(s)", Carleton University, April 8, 2011, http://chomsky.info/talks/20110408.htm, 2011.

[7] M. Tomasello, "Constructing a Language: A Usage-Based Theory of Language Acquisition", Harvard University Press, 2003.

[8] J.R. Saffran, "Statistical Language Learning: Mechanisms and Constraints", Current Directions in Psychological Science 12(4), pp. 110-4, 2003.

[9] J. Sinclair, "Corpus and Text – Basic Principles", M. Wynne (ed.) "Developing Linguistic Corpora: a Guide to Good Practice", Oxford: Oxbow Books: pp. 1-16, http://ota.ahds.ac.uk/documents/creating/dlc/chapter2.htm, 2005.

[10] E. Tognini-Bonelli, "Corpus Linguistics at Work", John Benjamins, Amsterdam, 2001.

[11] B. Grosz and C. Sidner, "Attention, intentions, and the structure of discourse", Comput. Linguist. 12(3), pp. 175-204, 1986.

[12] J. Searle, "Conversation", in: H. Parret, J. Verschueren (eds.) "(On) Searle on Conversation", John Banjamins Publishing Company, Philadelphia/Amsterdam, pp. 7-29, 1992.

[13] Y. Wilks, "Is There Progress on Talking Sensibly to Machines?", Science 318(5852), pp. 927-928, 2007.

[14] M. Gnjatović, M. Janev, and V. Delić, "Focus tree: Modeling attentional information in task-oriented human-machine interaction". Applied Intelligence 37(3), pp. 305-20, 2011.

[15] M. Gnjatović, V. Delić, "Cognitively-inspired representational approach to meaning in machine dialogue", Knowledge-Based Systems, Vol. 71, pp. 25-33, 2014.

[16] M. Gnjatović, B. Borovac, "Toward Conscious-Like Conversational Agents", In: Toward Robotic Socially Believable Behaving Systems, Volume II - Modeling Social Signals, A. Esposito, L.C. Jain (eds.), volume 106 of the series Intelligent Systems Reference Library, Springer, pp. 23-45, 2016.

[17] M. Gnjatović, "Changing Concepts of Machine Dialogue Management", in Proc. of the 5th IEEE

International Conference on Cognitive Infocommunications (CogInfoCom 2014), Vietri sul Mare, Italy, pp. 367-372, 2014.

[18] M. Gnjatović, "Therapist-Centered Design of a Robot's Dialogue Behavior", *Cognitive Computation*, Vol. 6, No. 4, pp. 775-788, 2014.

The 9th International Conference on Business Information Security
(BISEC-2017), 18th October 2017, Belgrade, Serbia

SOFTWARE DEVELOPMENT PROBLEMS OF THE SDN INTERNALS ENGINEERING

VITO LEGGIO, LYUDMILA ZHAROVA
University of Belgrade, FON, Belgrade, Serbia

ALEKSANDAR R. MIHAJLOVIĆ
Seven Bridges, Boston, MA, USA

SRAVANTHI DONTU, RADOMIR A. MIHAJLOVIĆ
NYIT, New York, NY, USA

Abstract: *Software-Defined Networking or SDN has been the focus of numerous research and development teams over last decade. A long history of efforts to make a network more controllable (configurable and manageable) can be traced back to the very beginnings of the computer networking. The fundamental technological concepts of SDN are by no means new. In spite of the evident academic and industrial enthusiasm, SDN has been surprisingly slowly accepted and deployed. Development of the super-sized data centers and cloud computing systems have recently accelerated SDN adoption in spite of the documented relevant software complexity.*

While promising to simplify network design, control and provide new solutions, SDN is introducing a significant amount of software that has to be trusted as correct, be bug-free, secure and extensible. In order to address the last issue of extensibility, we have initiated a project of complete SDN stack re-engineering. To be able, not only to deploy new protocols on SDN ready traffic processors such as OpenFlow compatible switches, also known as trivial network elements but also to introduce custom south bound or north bound protocols, complete control over the entire SDN software architecture stack was mandatory.

To achieve such a control, the project presented in this paper involved re-engineering of the available open source software products such as OpenFlow compatible Open vSwitch (OvS) developed in C/C++, OpenDayLight (ODL) SDN controller developed using enterprise Java technologies, (Java 8, Maven, OSGi, Karaf, YANG, etc), and UI DLux developed using Web set of languages (HTML5, CSS, JavaScript, XML, etc.), as well as a set of diverse technologies, (Python, REST, NodeJS, JSON, AngularJS, etc.) and protocols (HTTP, SNMP, OpenFlow, NETCONFIG, etc.)

Although the SDN stack model decouples the network application, control and forwarding functions and so reduces the underlying implementation complexity to the three smaller blobs of code, the fact that these three blobs had to be developed in three different technologies and their inherent individual internal complexity, demanded specific strategic approach to additional complexity reduction.

The primary purpose of our paper is to share problems faced, discoveries made and solutions used. SDN stack represents one complex software engineering system whose development involves the use of the specific tool stack. A tool stack in question involves a stack of languages (from the very low level to the very high level), a stack of relevant development tools and software engineering technologies, as well as a number of protocols. The evident complexity and dynamism of the SDN stack evolution have driven forward almost all interested parties, including our team, to take the open source global team effort sharing as the fundamental software engineering principle.

Keywords: *Network operating system, NOS, SDN, NFV, software coupling, OpenFlow, Open vSwitch, OpenDayLight, DLux, YANG, SNMP, NETCONFIG.*

1. INTRODUCTION

In parallel with the intensified global cyber-warfare activities (e.g., in the last two years the number of cyber-attacks on Russian cyberspace has tripled, [1]), we are witnessing a new form of the cyber-security relevant battlefield that we refer to as "the software market cyber-security arena" or SmCsA. German BND security government agency has long ago quietly replaced Windows desktops with the in-house compiled Linux versions. For political correctness reasons, the avoidance of Windows operating system by German government had to be publically denied [1]. Chinese and Russian governments are currently eradicating, not only Microsoft Windows in their offices, but many other complex software systems too [2,3,4]. The intent of the US military to achieve total hegemony, including a cyber-hegemony, has produced a worldwide reaction that in the final analysis turned out to be bad for American exports.

We observe that the global SmCsA conflict, for all practical reasons would not have been possible without the massive advancement of the phenomenal open software movement with Linux operating system as the flagship open software product. It is not very well known that celebrated Finish developer Linus Torvalds, was not the creator of the Linux operating system. The first version of Linux was simply a copy of the well documented scaled down UNIX versions XINU (D. Comer with students [5]) and MINIX (A. Tanenbaum with students [6]). True contribution of Linus Torvalds to the modern history of computing is in initiating an open source approach to voluntary team development of complex software such as kernels of the operating systems [7]. Thirty years after the very first successful open source Linux project, the US cyber-hegemony in the SmCsA space is threatened. Open source is promising to bring worldwide independence of the US software super trusts like IBM, Microsoft or Oracle.

The earliest reaction against the software monopolies was the establishment of the Free Software Foundation (FSF) by Richard M. Stallman in 1983. Unaware of the future defense-related consequences, US government was quick to react to the federal law known as "The National Cooperative Research and Production Act," of 1984. Following up on these two events, we had initial moves to start structured voluntary team development of complex software. The attribute structured refers to the strict rules by which volunteers could participate in the movement.

In 1988 the nonprofit Open Software Foundation (OSF) organization has been formed. OSF primary purpose was to improve ATT UNIX back-end networking and GUI front-end related source code. In the late 1980's UNIX required large and costly memory space so that OSF could engage only big corporations leaving little freedom for startup companies and entrepreneurs. Improved hardware with declining prices accompanied with new software tools development have provided a better platform for wider acceptance of the OSF general idea. In the early 1990's a genuine open software development movement "for the masses" has been practically initiated by the now celebrated Linux founder, Linus Torvalds [7].

Drawing a parallel with the early 1990's and Linux as the OS kernel open source code development movement, we have now a similar movement with the open source network OS (NOS) kernel at the central stage. Open source NOS development is backed by the Open Network Foundation (ONF) formed in 2011 by the group of computing industry leaders [8]. An antiparallel between the OS kernel with OSF and NOS kernel with ONF supports could be viewed in the original contexts of these two mega projects. On the contrary to the early 1990's with the beginnings of Internet expansion, nowadays we have rising political tensions between the US on one side and China with Russia on the other, we have massive Internet use and rising cyber insecurity tensions. Apparently, political as well as technological contexts are different. Detailed analysis of the implications of the new situation in the modern computing industry is beyond the scope of this paper. This paper is primarily focused on the security and technical problems faced when deciding in favor of the open software option.

We present here the so-called NOS stack as an example of the extremely complex ongoing open software project that requires a thorough understanding by the contributing developers. In the following sections, we describe a wide spectrum of software engineering state of the art technologies that have to be mastered in order to achieve full control of the deceivingly free open source NOS software [9].

As probably the most complicated project in the open source arena, the NOS stack has engaged all available software engineering achievements including the latest cutting-edge technologies. To illustrate this statement we have decided to analyze all relevant open source sub-projects that are currently gaining industry acceptance ground. The most outstanding representatives of the open source NOS stack sub-projects are:

- DeLux (Dlux) representing application or network management plane [10],
- OpenDaylight (ODL), representing NOS or control plane [11], and
- Open Virtual Switch (OvS), representing data plane [12].

2. NETWORK VIEWS

Viewed physically in space, computing networks are made of distributed network nodes (a.k.a. network elements) and network lines that physically connect nodes. The physical structure of the network nodes and lines distribution may be described as network topology model. Within a topology, a node may be dedicated to data traffic processing as traffic processor (TP) or to serving user application programs as application host (AH). As an exception, a node may be hybrid in nature serving both, as a TP and as an AH at the same time. Traditionally, well designed and implemented network topology would have TPs as inner nodes and AH devices as network edge nodes or topology leafs. Figure 1 illustrates one example of a topology with a diverse set of TP nodes.

Traditional physical TP nodes have been delivered in many versions containing diverse and complex functions loaded and/or embedded into their mainboards. Functions such as basic or advanced LAN switching (e.g., MPLS), routing, load balancing, multicast routing, differentiated service processing, traffic engineering, firewall traffic filtering, VPN processing, etc., were delivered to customers at nominal cost of the TP unit.

With the proliferation of new networking protocols and TP functionality solutions, network engineers and network software developers were driven to look at alternative solutions for more cost-effective TP firmware and software reusability. One of the most significant contributions in that direction was Martin Casado's doctoral thesis [13] published in 2008. Ethernet Architecture for the Networked Enterprise, (ETHANE) project [14]) based on the Casado's thesis has explicitly proposed that basic switching-only TP (a.k.a., dumb, vanilla, simple, trivial or

bare iron), may be more broadly used as a cost-effective network building block, provided that TP functionality is supported by the distinct network control and management software server. This idea is well illustrated in Figure 2 with a network switch (Marked as SW) acting as a TP node of minimized functionality.

The externalization of software traditionally resident in the TP node device out into the Network Host (NH) also referred to as a network controller, allows minimized hardware TP (SW) vendors to increase the speed of the simplified traffic processing. With smaller computation demands and reduced onboard software, new SW devices can forward traffic data frames at higher speeds and lower costs.

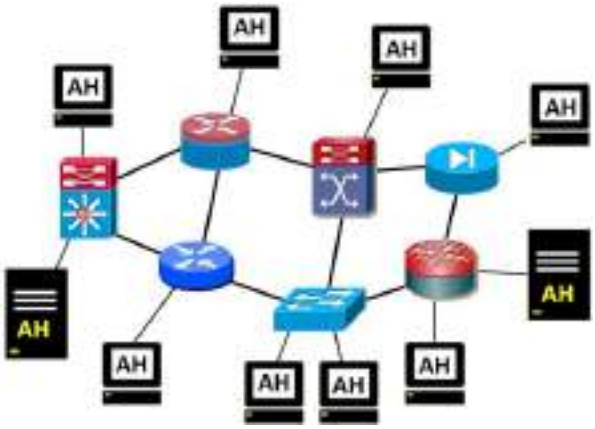

Figure 1. Well designed traditional network.

To clarify the argument behind the new view of the network outlined in Casado's thesis, it is important to distinguish two general categories of applications running on a network. Application host (AH) may be running network application software that is applied to:

- Arbitrary user activities in a form of general user applications (Labeled as AH in Figure 2), or
- Specific systems or network relevant activities in a form of dedicated network management and control programs and hosts, (Labeled as NH and NAH in Figure 2).

One typical network system dedicated application is Domain Name Service (DNS) which is used to decouple physical topology (PhT) from the logical user bound topology (LgT). Since the underlined PhT exists and is just transformed via LgT to act as an interface facing general purpose applications, it is not proper to view LgT as a virtual topology (frequently made mistake). We refer to the virtual network as a network that truly does not exist, a network implemented as software simulation [15].

Computing networks are distributed systems supporting execution of distributed applications. Being distributed, networks are hard to manage and protect consistently. Solutions such as access control, traffic filtering using routers or firewalls, traffic isolation with virtual networks (VLANs) and virtual private networks (VPNs) are subject to continuous change and therefore are prone to a faulty configuration, inconsistency and increased security attack sensitivity.

Among other operations, control P2 plane must compute forwarding state of the P1 plane devices. To accomplish its task, the control plane must:

- Learn controlled network topology
- Decide how to process received by P1's SW device unknown data frame on the learned topology
  - Assign to all controlled SW devices (TP-switch) forwarding state (Maintain TP state).

History of software engineering (SE) implies the following analogy. At the very beginnings of the SE history programs had to be aware of the processor hardware capabilities (CPU instruction set and instruction format) and exact memory map address space. Faced with the complexity of increasingly larger programs, developers looked at how to apply abstractions and how to simplify programming. By using higher-level languages and program translators it was possible to write programs (e.g., using 3GL C language) that would abstract CPU specific details. As a result, the same high-level language program could be used on the CPU of any brand. However, CPU abstraction and programming simplification required the use of some sort of interface-like program translation.

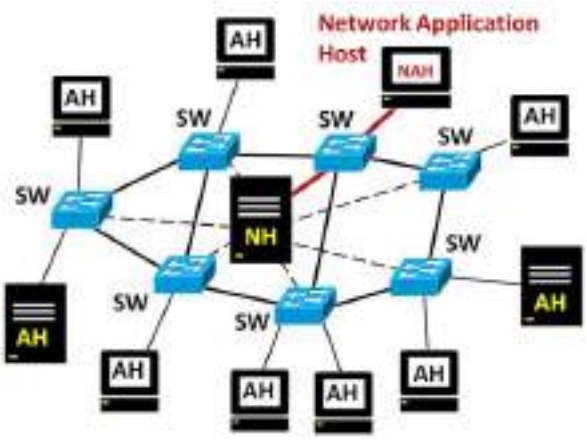

Figure 2: Software Defined Network standard TP devices of minimized functionality.

In parallel with the mentioned SE complexity reduction by means of processor abstraction and interface-like translation, network engineers have realized that similar line of thought could be applied to networks, i.e., higher to lower abstraction level mapping system could be applied too. Along these lines of reasoning, a three-plane network operation and management model has been proposed, (See Figure 3). Nick McKeown, a professor at Stanford and his Ph.D. student Martin Casado, have introduced an idea of externalization of the TP software functionality and to change traffic processing and forwarding behavior. Their idea is now known as Software Defined Networking or SDN [14,16].

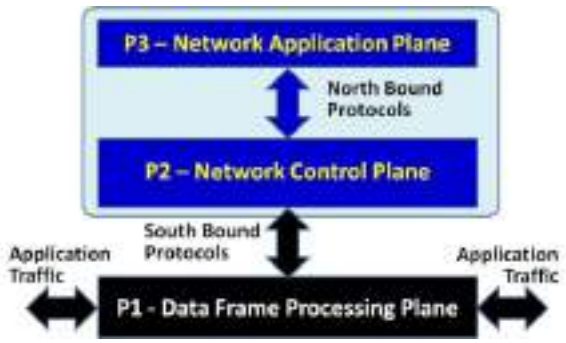

Figure 3: Basic three plane network operation and management structure.

Figure 3 illustrates a structured model of a network as a system in charge of processing network application data traffic between AH nodes in the P1 plane under the control of network management, administration and engineering applications in the P3 plane. In 1982 [17] John Gage chief scientist at Sun Microsystems has coined a slogan "The network is the computer." The network as the computer running distributed applications. has been augmented now with the systems administration application P3 plane and the operating system, (i.e., network operating system or NOS), in the P2 plane. With the SDN the network appears clearly as the computer.

Moving forward with technologies such as Cloud Computing, virtualization, big data, massive mobile computing and IoT networking, modern networks had to become more agile and responsive to new user applications demands and dynamic topology variations. Today's networks had to gain flexibility and possible automated adaptability which only better application of abstraction and new software could provide. Network adaptability referred to as network automation and orchestration can only be accomplished through the application of feedback-control design patterns within a network as a software system. We refer to the P2 and P3 two plane SDN architecture of a network as network operating system stack or NOS-stack.

SDN has been recognized as a disruptive technology and has shown great promise to manage and control next generation of networks. The new generation of networks are re-engineered traditional networks (See Figure 1) with the simplified P1 plane and the TP functional software migrated to the NOS running on the dedicated system and application hosts (NH and NAH nodes shown in Figure 2).

3. P1 DATA PLANE

At the lowest level of the traditional seven-layer ISO-OSI network model [18], traffic is a represented by the modulated L1 signals and by the L2 streams of data frames. At the higher levels of the model (L3, L4 or L7), traffic flow gains in semantic value as data packets, message segments or complete messages in the flow. The end-points of the data flow are interface controllers for L1 and L2 traffic, host operating systems for L3 and L4 traffic or distributed application modules for L7 traffic. Network traffic engineers are accommodated by a definition of RFC-3697 which specifies traffic flow as "a sequence of packets sent from a particular source to a particular unicast, multicast or anycast (broadcast) destination." To be more restrictive, one packet flow may represent one application message or media stream. More general, application independent flow is defined in RFC-3917 as a stream of packets at some observation point in the network.

RFC-2722 defines traffic flow as “a general logical equivalent to a connection” relevant to any protocol, while SDN relevant literature refers to traffic flow as only L2 data frame traffic. As the most fundamental element of the NOS stack in the P1 plane, a “flow” represents systems data record located in the TP switching (traffic-data-forwarding) table entry, indexed by attributes of the arriving data frame at the ingress TP interface [19].

P1 TP device can be physical device or software simu-lated virtual device used in Network Function Virtualization (NFV) [20] such as Open Virtual Switch (OvS) program [21]. Performance requirements have made C a language of choice when developing runtime-hyper-active NFV software version of the P1 plane virtual TP devices such as OvS. We have used OvS in experimental network simulations, and have analyzed internal architecture and implementation of the OvS software system but will not cover it this paper.

Trivialized TP switch is configurable and programmable by the P2 and P3 plane programs.

Programmability enables flexible use of P1 plane simple traffic forwarding functionality that may be applied to the implementation of nonstandard higher layer traffic processing protocols in the ISO-OSI model.

In this paper, we do not elaborate on the trivialized frame switching TP internal architecture and we highlight the P1-to-P2 software interface present in the TP. The most fundamental P1-to-P2 interfaces are shown in Figure 4.

Figure 4: Data plane P1-to-P2 interface modules.

4. NETWORK ABSTRACTION

Initial SDN based simplification of a network using three-plane system architecture has introduced a long list of software engineering problems related to the complexity of each of the plane’s internal architectures and platform-specific software implementations. Internal architectural complexity and the implementation process of each of the planes in the three plane model demands a new mindset and requires using a wide range of SE technologies, such as design modeling languages, programming languages, software system building tools, deployment tools, maintenance tools, etc. To start dealing with this new di-mension of the networking discipline we present the following definition of the software system architecture.

Definition 1: Software system architecture is a hierarchical-abstraction of the final runtime elements describing their execution sequencing, inter-relationships, and their coexistence in the runtime context, referred to also as the environment or platform, [22].

Specifically, a "characterized architecture" is an architecture constrained by the configuration of architectural attributes or parameters needed to satisfy given functional requirements. Architecture characterization itself is one general attribute of the overall architecture that defines future runtime system performance and its relationship with the production execution context.

Abstraction reduces complexity at a given point of view but requires view-transformation to the lower levels of complexity. In case of the highest network abstraction plane, a P3 plane, view-transformation is performed by the P2 plane which implements central NOS element as a network controller, (SDN controller). Software planes P3 and P2 stack greatly simplifies network operator's or administrator's view.

P3 plane software as network management and control set of applications may implement and operate on the P2 exposed logical representation of the physical P1 network. Logical network representation is also known as the network service topology. The highest possible abstraction of a network as the logical network could be a network appearing as one big logical TP switch, implementing trivial star topology. Such a topology abstraction applicable to any underlined physical topology is announcing perimeterless networks and maybe another paradigm shift, a shift of the network security paradigm, (perimeter protection paradigm) shift.

5. P2 PLANE CORE ARCHITECTURE

From the software engineering point of view, the most complex element of the NOS stack is a P2 plane with the network controller at its core. Architectural static structure and dynamic functional organization of any operating system (OS) in general, are so complex that only continuous evolutionary, iterative approach to development and redevelopment throughout the production/use phase is viable practice. SDN NOS as a special purpose operating system with all of its complexities is also going through the evolutionary process which requires easy extensions, simple updates, upgrades and even hot-swaps of software modules at runtime.

One of the most fundamental rules in software engineering is to reuse old knowledge (patterns), designs and existing programs whenever possible. The reuse rule also known as the principle of "cloning and owning" should be applicable to all development phases. For instance, at compile time reusable software may be introduced as libraries, frameworks and ready to reuse plug-and-play code objects. At runtime we may reuse distributed service modules, runtime platforms, software environments, and so on [23]. Frequently mentioned term "Application Program Interface" or API implies the existence of the interface between the software under development and the reusable library or system programs. The API concept has been recently extended beyond the simple runtime stack-based "in process" thread jump to the linked library routine. Grasping modern SE extended API concept has been one of many problems that authors of this work had to deal with.

Depending upon the level of the development complexity of monolithic software systems, the principle of "design by reuse and for reuse" may lead to the following strategic approaches:

- Cut-and-paste or macro-based approach [24],
- Library or linker based approach,
- Framework or execution-container based approach, and
- Hybrid mix framework with libraries approach.

All current SDN NOS solutions, including the ODL project considered in this work, are based on some sort of a reusable framework supported by a set of class libraries, i.e., on the hybrid approach to the software reuse. Figure 5 illustrates the general hybrid architecture of the ODL NOS software system with the custom developed code reusing framework and Java class libraries.

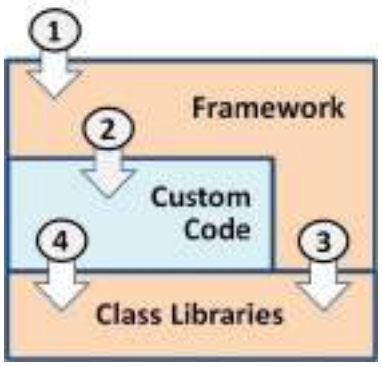

Figure 5: Hybrid mix, a framework with libraries as architecture super-pattern of software reuse. Framework provides reuse of the macro-architecture design and code while libraries provide only code reuse.

A framework is ready-made, semi-finished, standardized reusable initial skeleton code that can be customized into a complete version of the software product by:

- Modification (not recommended), or
- Object-oriented extension and composition adding new custom code (recommended)

Development process at compile time starts with the framework. At runtime, when software execution begins the first program to start is the framework, (See point 1 in Figure 5). The fact that developer's custom code execution is controlled by the framework as external program, demonstrates the so-called inversion of control (IoC) principle. With the IoC, a custom developed code is not where the initial program's thread would start running (point 2 in Figure 5). Framework code controls software system static structure and runtime execution control flow too. The framework calls back into a custom code. IoC lacks explicit and clearly visible deterministic control sequence which oscillates between the framework

(callbacks) and the custom code so that debugging becomes quite hard.

A framework may be:
- White or clear box framework, or
- Black or opaque box framework

Developer unfriendly clear box framework requires that user-programmer understands framework source code. With the clear framework, user-programmer can employ inheritance to extend identified classes or implement interfaces of the framework. The term "clear framework" is more intuitively appealing than the term "white framework" while occasional use of the term "transparent framework" is improper (The framework is not invisible). Clear box framework is also known as a source code developer's framework or compile time framework.

A black box or opaque framework is an interface and composition based framework that does not necessarily require user programmer's knowledge of the framework's source code internal implementation. Customization of the black box framework simply requires the use of the exposed framework interface. The developer designs and deploys framework interface compatible objects known as components or plugins. A black box framework is commonly delivered compiled as the end-reuser's development platforms. Detailed interface documentation is all that is needed by the custom code developer. We consider black box framework as a runtime framework and use it as the initial SDN NOS controller building block. Black box frameworks provide better modularity and module isolation and are more developer friendly.

The framework is problem domain specific and maximally reusable, both as a design pattern and as a software module at the same time. API libraries are more general and less domain-specific programs that provide a reduced scope of reusability. Complementary blend of the framework and APIs is well used in the SDN NOS ODL implementation.

As a rule, the development productivity and runtime performance are two conflicting software engineering requirements. This is quite well illustrated in the NOS-stack design and development case. For instance, the use of object-oriented language such as Java with the huge class library base was natural productivity driven choice for the complex P2 plane SDN software development. However, Java programs are always outperformed by the developer unfriendly equivalent C programs.

Frameworks are software reusing productivity solutions that may involve unnecessary code which may increase the overall size of the system. Studying framework documentation and learning how to use it may take longer time than developing a specified system from scratch. Since P2 SDN controller system involves sizable complexity and must go through the evolutionary development process using a framework is a right choice.

6. OPENDAYLIGHT SDN CONTROLLER

As software engineering project OpenDaylight (ODL) is one example that best demonstrates the use of the state of the art SE technologies, from the Domain Driven Design (DDD), Model Driven Architecture (MDA), to the framework based component software engineering with dynamically reloadable modules.

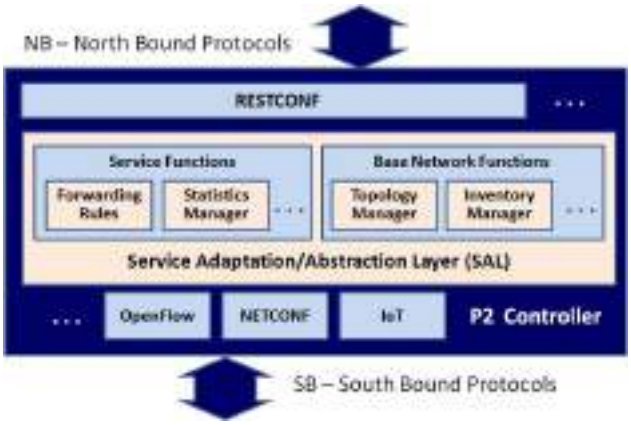

Figure 6: Open Daylight (ODL) P2 controller basic architecture with functional plugins integrated via service adaptation layer.

ODL is a highly available, extensible and scalable multi-protocol open source P2 controller framework built for license-free SDN deployment. ODL is open source RedHat Foundation project powered by Java technology.

Through the so-called scoping process developers are making decisions what controller functionality to support. Default basic functionality that any SDN controller must have is represented in the general architecture layout shown in Figure 6.

All functional custom code objects are developed as ODL framework plugins or code bundles. Developing a plugin does not require modifying the framework code. A plugin is a composition code module developed to extend framework functionality. A plugin is developed separately from the framework and can be added to or removed from the framework even at runtime, (can be deployed/loaded or unloaded dynamically by the system administrators). The dynamic feature of the ODL system is facilitated by the special deployment tool named Karaf.

Besides the framework and the API libraries as reusable code, individual plugin components can also be reusable and open software. A plugin can be delivered by a third party, and installed within a framework by the application administrator. In case of the ODL controller, there exist over 1000 plugins developed by different teams and available for free reuse. These plugins reuse is narrowly focused on SDN functionality and as more specialized are on average larger than general more reusable components. The basic "use or reuse" rule states that "the wider reuse the smaller footprint and the more narrowed focus the larger is component code footprint." We expect that when the learning curve reaches its saturation, and we all learn enough about the SDN problem domain, some other, not yet known technology will be applied to the overall controller footprint reduction. The general rule in the field of secure software engineering states that systems with larger footprint are less secure and are harder to defend.

7. SERVICE ORIENTATION OF THE NOS STACK

Building up overall software system by working only on new modules is the most desirable software engineering feature of object-oriented, object-based and service-oriented RESTful technologies. In simple terms, modifying software by not modifying anything but by well-outlined addition is the most robust software development strategy on which object oriented and object based approaches rest.

Framework container and component-oriented architecture may be viewed as a loosely distributed system. The limiting factor of holding all system parts and having all messaging handled by a framework container was one reason to isolate some architectural components outside of the container and use them as distributed services. By decoupling certain components as services, these services became reusable by more than one application at runtime. Apparently, distributed architecture facilitates software reuse both, at compile time and at runtime. Literature names resource sharing as the primary reason of architecture distribution, while we highlight software engineering need to share code as a reason to distribute the NOS stack. Using a service as architectural element represents new design super-pattern. By using distributed architectures based on services large monolithic applications could be simplified in all software lifecycle phases, (design, development, deployment and runtime phases).

Service-oriented software architecture (SOA) is distributed architecture with several specific attributes that can be summarized by the following definition.

> **Definition 2**: SOA is a design paradigm, a super-pattern, a principle, architectural blueprint or a rule how to design distributed software system with minimum two communicating modules, one being service consumer and the other service provider, supported by the messaging protocol that guarantees maximized service provider reusability.

To accomplish maximized service reusability, SOA must have:

- Minimized service logic, and
- Service-specific application protocol layer that enables introspective service description, advertisement and simple access, (Protocol simplicity is essential service feature).

SOA architecture is asymmetric, meaning that service-consumer and provider have different structure and different ways of using common messaging protocol of minimized query/response dictionary and verbosity. Basic SE principle implies that protocols of minimized verbosity result in maximal distributed module decoupling.

In a complex distributed system such as a NOS stack, traditional monolithic service logic had to be re-factored into a set of smaller and more reusable services, [24].

Two general classes of service based architectures may be distinguished:

- Service Oriented architecture (SOA), and
- Micro-service based architecture (μSOA).

Services may be composed into custom-services. Customized SOA service runtime chain of service consumers/providers may concurrently serve multiple applications with composed super services. Inter-service dependency in the SOA service chains may result in problems of dependency tracking and management, which is one of the major issues that designers and developers of the NOS stack have to deal with.

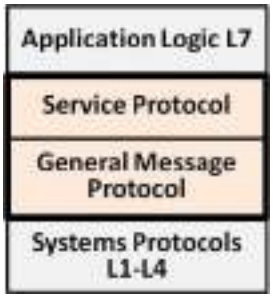

Figure 7: Two layer structure of the service protocol.

It is important to recognize the existence of two layers in the structure of the application communication protocol used in SOA solutions. As shown in Figure 7, service specifics messaging details are handled in the upper layer by the service-specific protocol (e.g., SOAP or REST) and more generic messaging (more reusable messaging) is handled by the lower service-independent protocol layer, (e.g., HTTP).

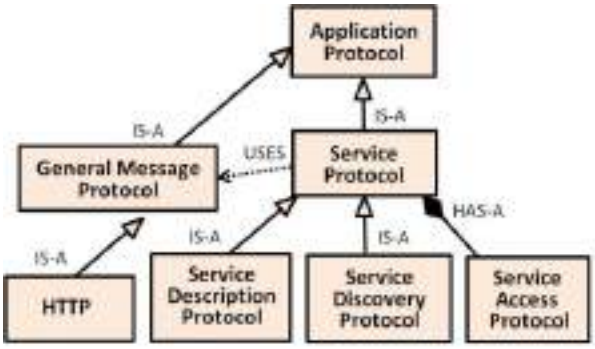

Figure 8: UML diagram of the SOA application protocols.

The "use or reuse" rule is the best demonstrated by the concept of the RESTful service which is extensively applied to SDN controller interface design.

8. REST API IN THE NOS STACK

Whenever possible (in case of P1-P2 or P2-P3 communications), NOS stack development teams are employing the so-called RESTful service APIs and service-oriented architecture approach. To gain an appreciation for this approach to massive software distribution, it is important to distinguish at two basic server elements:

- Service, and
- Resource.

Service element involves a call to some function or method to perform a certain activity. Functionality and performance are key attributes of a service. A resource, on the contrary, does not imply functionality and performance but focuses on the result of some abstract encapsulated functionality needed to provide a resource.

The well-known argument in favor of architecture distribution is client and server development effort disassociation. Ph.D. dissertation of Roy Fielding [25] has clearly described that such disassociation could be maximized, not by means of a new protocol, but by using new SOA derivative known as REST. REST stands for REpresentational State Transfer client/server architecture specification which among several things proposes that a RESTful session-state is held in the client. The client maintains and provides all data needed by the server to process new requests. A RESTful server is stateless and "ignorant" of the past request/response session activities. This runtime server independence of the client greatly simplifies server development and security, (Stateful protocol based servers are easier to attack than stateless).

REST architecture approach has been inspired by the HTTP protocol specification of which Roy Fielding was one of the principal authors (Englishman Tim Berners-Lee is being unfairly distinguished as the sole inventor of the Web and the HTTP [26,27]). As illustrated in Figure 9, architecture and protocol are two interrelated but distinct items, where architecture may include protocol specification. A RESTful protocol such as HTTP can be used to expose encapsulated RESTful server functionality as a service.

Figure 9: REST architecture.

To better understand service oriented RESTful service, interface, and API, one should observe that the Uniform Resource Identifier (URI) used in the service query does not refer to the server functionality (e.g., … /computeResultX) but instead refers to the static Web-page-like resource (e.g., … /ResultX). Referring to the result or a resource, server-end internal functions can be freely modified and renamed at server developer's will. RESTfull API allows server components to evolve independently of the client and vice versa. Some refer to RESTful architecture as a resource-oriented REST design pattern that uses the concept of a resource instead of more platform-specific objects or methods.

Tracking the state changes of the client/server session may require a deeper knowledge of the server func-tionality on the client developer's part, which is known as the client/server coupling problem. Representation of a resource encapsulates (hides) resource dynamic generation and delivery mechanisms. REST-based architecture ensures maximal decoupling of the client from the resource providing server and provides a clean separation of concerns. Maximal decoupling requires that client only has a reference or identifier of the resource and know nothing about the server-side software that produces and delivers the resource. Having most of the modern server-side projects being subjected to continuous modifications, updates, and upgrades, maximal client/server decoupling provides maximal server development freedom. This freedom is nowhere more desirable than in the case of the NOS stack project.

9. SERVICE ABSTRACTION IN THE SDN CONTROLLER

To accommodate consumer service needs service exposure can be made dynamic, and may vary at runtime. Runtime options of the given service interface are referred to as a service contract. Service contract covers mini-mized description and advertisement/exposure (Service Discovery Protocol in Figure 8) of the given service. Self-described service may be introspected. i.e., meta-queried (Self Description Protocol in Figure 8), to provide its encapsulated capabilities and public interface options [28]. This "talkative" service feature is of great runtime functional benefit but may present a serious security risk.

Minimization of the contract reduces service-consumer dependency, reduces service-consumer coupling and increases the number of possible different service users, (i.e., expands service software runtime reusability). Service consumer and service implementation should be as much as possibly independent (decupled). Smaller consumer contract dependency reduces consumer/provider coupling. The same way as Java abstract class and interface would hide concrete implementation logic, service abstraction through minimized contract is the way to hide the unnecessary details about the overall functionality and the internal details of the service. Service abstraction enables easy introduction of the new service versions through the so-called service evolution. Exposed details about a service are advertised or published with the reduced description to be used in the service contract leaving all missing details abstract to the consumer.

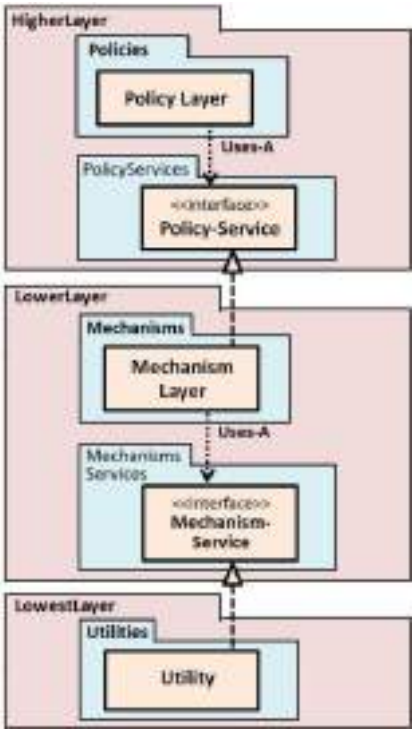

Figure 10: UML model of the service abstraction layer isolating Policy and Utility layers.

Abstraction can be understood as the simplification oriented modeling. Abstraction is a mental process of suppressing, ignoring or neglecting selected set of details related to the modeling target. As the primary SOA design principle service abstraction demands publishing of the minimized (most abstract) service description to be used in the service contract.

Service-specific upper application protocol layer (Figure 7) is in charge of handling service location, access control, and service use rules (Figure 8) which determine service contract options. Given service contract must be the only way consumer may access service internal logic and data. Service abstraction viewed as SOA design supper-pattern that guarantees minimized application protocol upper layer which is highlighted in Figure 7.

In order to decouple P3-P2 northbound (NB) from the P2-P1 southbound (SB) protocols, the P2 architecture employs service isolation through the service adaptation or abstraction layer (SAL) shown in Figure 6. P2 ODL architecture is implemented as a component container with plugin modules used for all required functionality. ODL project has initially relied on the API Driven Service Abstraction Layer (AD-SAL) to separate the SB from the NB service protocol plugins. The service adaptation layer enables NB software development compatible with the SAL avoiding constraints imposed by SB programs or protocols. Figure 6 shows layered P2 ODL controller architecture with SAL as the central layer being in charge of communications between multiple SB and NB plugins.

Fundamental pattern of the service adaptation or abstraction is described by the UML model shown in Figure 10. Reduction of dependency of the upper layer from the lower is implemented by the pair of interfaces, (Policy-Service and Mechanism-Service interfaces).

By definition, dependencies are all objects that given code must use to operate. In layered architectures, lower layer's objects are dependencies of programs in the upper layer. Since dependency existence implies code coupling, it is desirable to minimize dependencies whenever possible. One way of reducing coupling with dependencies is dependency inversion principle or DIP pattern.

DIP pattern use makes higher level objects less depend on the lower level objects by having both level objects depend on abstractions, (e.g., interfaces as shown in Figure 10) and not on known functional concrete objects. Figure 10 illustrates dependency abstraction through dependency inversion using two interfaces. Development or replacement of the upper or lower layer object implementation should not violate the "expectations" of the interface users. Liskov Substitution Principle (LSP) [29] states that subtype or subclass may replace the supertype or superclass without any loss of functionality "promised" by the supertype to its users. This concept is also known as strong behavioral subtyping principle.

To reinforce DIP concept clarity, Figure 10 highlights interface locations in the UML model. The interface to the lower layer classes is under control of the upper layer package developers so that lower layer class decoupled from the upper layer class appears semantically depending upon the upper layer design. The perspective of the lower layer abstract ownership is shifted to the upper layer design team. Policy or high layer code should not depend on implementation details of the lower layer. On the contrary, the lower layer details should be influenced by the higher layer policy and the interface. DIP overrides well-known abstraction of the lower level dependency via the interface and transforms it into the dependency inversion.

The DIP is the main conceptual SAL design guideline. DIP and SAL help developer understand how to correctly bind SDN ODL system parts together and have SB plugins implementation depend on the higher-level policy abstractions which enhances overall ODL system cohesion.

DIP application is enhanced by another conceptual approach which is known as the Dependency Injection (DIj) pattern. DIj proposes that each object requires all dependencies to be created by the Dependency Injection Container (DIjC) which would maintain a map of all object dependencies and logic how to create all dependencies. Custom code of Figure 5 does not create its dependency objects but has all dependency objects pushed into the code from the outside by the DIjC framework. Dependencies are injected into the code through the above mentioned IoC inversion of control.

IoC and DIj have custom code dependencies injected from the context or framework and so decouple given code from the construction of its dependencies, i.e., from the lower level implementations making code cleaner, easier to extend/modify and reuse.

By not controlling the creation of dependencies and by "asking" framework DIjC to create dependencies, DIjC is charged with the task to resolve complex dependencies what appears to be much easier and more transparent than the solution that would leave dependency resolution to the original code. DIjC performs this task by consulting external configuration files instead of using builtin hard-coded dependency data. Changes in dependencies require only DIjCand configuration files update leaving custom code to remain the same.

10. MODEL DRIVEN APPROACH TO NOS STACK DESIGN

Model Driven Architecture (MDA) is a paradigm-shifting software engineering technology that makes a model, and not a program, a primary software artifact [30]. MDA makes software development a process of transformation of one model into another. Each transformation output is a new model that is closer to the desired program than the input model. Starting with the original Platform Independent source Model (PIM) after one or more model transformations, the final desired program or Platform Specific target Model (PSM) can be obtained.

Using precise enough model instead of a program increases the level of abstraction and freedom to flexibly implement final complex software. Being portable and interoperable, models appear as super-programs that are convenient for distribution between diverse platforms in a form of program-containing-messages. Models are much easier to verify than programs, and much harder to undetectably hack. Model verification is a necessary step before model acceptance for transformation and final execution. For instance, not trusted Java applet must be functionally constrained, while received model-message may be trusted, transformed into a program which is allowed to have more of the local functionality. This feature of model distribution instead of explicit program distribution has found excellent use in the implementation of the NOS stack.

An important application of Model Driven Development (MDD) to open source NOS development is YANG project [31], (YANG originates from "Yet ANother LanGuage" or "Yet Another Next Generation"). YANG was developed by the NETMOD of IETF in 2010, to be used as a domain-specific modeling language for the description of Network device Configuration (NETCONF) protocol data [32] and protocol operations. In general, network TP device management involves configuration data and device state data accompanied by the applicable instructions, constraints, and notifications. Like documents, YANG models have descriptive tree structure involving a number of original built-in and custom made data types. Constraints can be individually targeted at tree nodes using XPath expressions (See RFC-6020). YANG uses a compact SMIng like (Structure of Management Information new generation [33]) easy to use syntax (Similar to the equivalent XML YIN language syntax). As a schema style language, describing what can be done to data, YANG can be used to describe how to build Java programs or JSON and XML document trees.

Using model-driven YANG/NETCONF approach P2 control has an option to apply DBMS-like ACID transaction management to every sequence of P1 configuration instructions. The P1 device acts as a NETCONF protocol server. Unlike the traditional command line interface (CLI) with SNMP, YANG/NETCONF SB configuration command sequence can be rolled back at any point in the configuration session. ACID transaction management guarantees consistent configuration data in the P1 data store. NETCONF management protocol defines configuration/operation data and applicable CRUD operations on defined P1 data stores.

NETCONF message format allows message format customization via XSLT transformation. A result of the P1 data store query about services, topologies or policies can be transformed using XSLT from a general vendor-independent to a P1 device-specific format [32].

Model-driven development (MDD) approach draws also on the experience and isomorphic parallels with functional Lambda-like languages that avoid the use of variables and state changes caused by the past program activities. In case of complex system architectures, not programs but models may be used to avoid dependencies upon programs made in the past. This general guideline enhances program robustness and security hardness.

Model-Driven Service Abstraction Layer (MD-SAL) is the current version of the SAL implementation in the ODL system. MD-SAL implements model-driven design concepts to robustly and securely decouple (disassociate) the NB application API from the SB protocol plugins, (e.g., OpenFlow, NETCONF, SNMP, IoT, OVSDB, and so on).

MD-SAL uses YANG as the modeling language for service and data definitions.

11. ODL AND OSGi

ODL is one of the projects of Linux Foundation administered by the Technical Steering Committee (TSC) [34]. Projects such as Open Distance Learning (ODL [35]) or Object Definition Language (ODL [36]) have to be distinguished from the OpenDayLight SDN controller project and software system.

Besides ODL, TSC maintains focus on the projects such as:

- Software for forwarding elements on the southern side of the controller [37,38], and
- Network application software on the north-bound end of the controller, (e.g., DLUX project [39, 40]).

As an open source project ODL has the following open goals:

- To produce customizable code base that covers core SDN controller functionality,
- To help grow a technical community of contributing developers, and

- To broaden compatible equipment vendor base.

ODL project is structured into releases of successful versions named in the order of elements in the Mendeleev Periodic System of elements. Hydrogen was released in 2014 as the first version with 13 projects and 1.3 Mlines of code, Helium released in April 2014 had 25 projects with 2.1 Mlines of code, Lithium released in November 2014 with 40 projects had 2.3Mlines of code, Beryllium was released in July 2015, Boron was released in March 2016, Carbon was released in September 2016, and Nitrogen in June 2017 [41]. The number of related projects and lines of code kept steadily growing.

ODL provides a model-driven service abstraction platform that enables easy integration of plugin modules

from diverse sources and teams.

Java provides language based productivity technology with fair dynamism based on flexible dynamic linking. ODL project has started with Java as the language of choice and OSGi as dynamic Java component service platform. Open Services Gateway initiative (OSGi) is an alliance of companies formed in 1999 to start royalty-free open standardization of dynamic Java component software platform that is now known as OSGi.

OSGi has been adopted as the most convenient approach to the development of the P2-plane framework based software system. Basic OSGi service components known as bundles are Java jar files. OSGi bundles can publish their services dynamically, can find and bind to other services through a service registry, and as service providers may appear and disappear at runtime. Dynamic nature of the OSGi platform implies that components can be hot-swapped (hot-added/loaded or hot-removed/unloaded) at runtime, without having to restart software system or reconfigure it manually. Using special deployment tools, bundles can be installed, started, stopped, uninstalled and updated at runtime. Visibility of the bundle content is enabled via APIs. OSGi is capable of tracking bundle version numbers and inter bundle dependency graph keeping the same bundle of different version wherever needed in the bundle-graph.

Like Java beans, OSGi bundles can provide state-data persistence, transaction management capability (Standard ACID rollback or commit functionality), and application of the web RESTful technologies.

A bundle is a group of Java classes and additional resources accompanied by a package-manifest MANIFEST.MF file which details bundle content and dependencies. Special building (Maven) and deployment (Karaf) tools are needed to use bundle manifest files and organize the ODL system at compile and runtime.

12. DLUX PROJECT IN THE APPLICATION LAYER

Network applications P3 plane is where the so-called network orchestration, i.e., automation, or adaptation takes place. In the P3 plane, we find applications handling network traffic engineering problems and provide friendly network operator or administrator productivity oriented user interface.

ODL graphical user interface, i.e., openDayLight User eXperience (DLUX) and OpenStack Neutron Cloud Computing management system are examples of P3 plane network applications. Figure 11 illustrates basic P3 plane DLUX system architecture.

Figure 11: Network management application layer.

DLUX is based on the Web front-end stack of technologies (HTML, JavaScript, CSS, SVG, JSON, node,js, and JavaScript libraries such as jQuery, Angular JS, etc.). A Web browser built in object framework powered by the HTML and JavaScript is used to implement friendly dynamic graphics- based user interface. Details of DLUX implementations are beyond of this paper scope and will be covered elsewhere.

13. CONCLUDING REMARKS

SDN and networking using software simulated TP devices known as NFV represent a true paradigm shift in transforming traditional computer networking and network security into a software engineering discipline. In simple terms, modern network design requires the use of complex software architectures and software engineering principles or patterns. The main purpose of this work is to highlight exemplary software projects of this transformative network technology revolution and to highlight the hardship that networking specialists had to go through in adopting and implementing SDN and NFV.

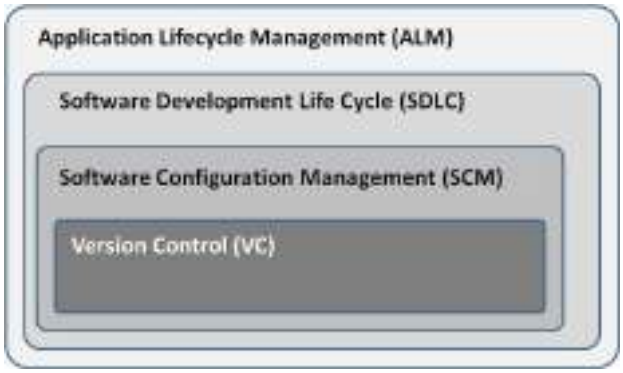

Figure 12: Hierarchy of software management tools.

The central project discussed is ODL, a collaborative open source project and complex SDN controller software system that has employed almost all currently known software engineering solutions and best practice principles. ODL modular architecture has enabled distribution of development labor in a form of individual projects that govern and manage each module. In an attempt to securely

apply a tremendous amount of the NOS stack open source code, authors of this work have dissected entire OvS, ODL and DLUX software systems, while going through series of poorly documented modules and assumed software engineering rationalizations.

Complex software systems such as ODL are made of hundreds of inter-related parts that have to be tracked and integrated, (i.e., the system has to be built), into a perfect (bug-free) harmony for deployment. A large number of parts and interdependencies demand that reliable automation tools be used to consistently assemble compiled code parts. Such tools are known among software developers as Software or Source-code Configuration Management (SCM or S/W-CM) tools (See Figure 12).

In complex systems, any systems part change may require reassembly and realignment of inter-dependent parts. In general, the system configuration is a duty delegated to system administrators. This is true only at installation or production time. However, software system at compile time, under development, requires specific and much more complicated configuration management procedures that SCM tools can easily handle. Authors expect that new Java 9 modular development support will help reduce problems regarding ODL complexity and security management.

REFERENCE

[1] "Senior security official says number of cyberattacks on Russia jumped three-fold in 2016," RT News Release, March 3, 2017. https://www.rt.com/politics/379325-top-security-official-says-number/

[2] Ivan Jenic, "Russia to Ban Windows from Government PCs," Windows Report News, February 10, 2016. http://windowsreport.com/russia-ban-windows/

[3] Rich Edmonds, "Russia wants to stop relying on foreign software for government systems," Windows Central, Nov 2, 2016. https://m.windowscentral.com/russia-wants-stop-relying-foreign-software-government-systems

[4] Timothy J. Seppala, "Microsoft made a version of Windows 10 for the Chinese government," Endgadget, May 23, 2017. https://www.engadget.com/2017/05/23/windows-10-china-government-edition/

[5] Douglas Comer, "Operating System Design: The XINU Approach," Prentice-Hall. ASIN: B005LZVKYE. 1984.

[6] Tanenbaum A., "Operating Systems V1," Prentice Hall, ISBN-10: 0136381987, 1988.

[7] "Linus Torvalds Introduces Linux 1.0," Youtube video lecture, March 30, 1994. https://www.youtube.com/watch?v=qaDpjlFpbfo

[8] David Erickson, "Open Networking Foundation Formed to Speed Network Innovation," OpenFlow Blog March 21st, 2011. http://archive.openflow.org/wp/2011/03/open-networking-foundation-formed-to-speed-network-innovation/

[9] Alain Villeneuve, "The Pitfalls of Using Open-Source Code," The Narional Law Review, November 25, 2011.

[10] "OpenDaylight dlux:Main," OpenDayLight, September 19th, 2013.

[11] "OpenDayLight Project," The Linux Foundation Project, 2017. https://www.opendaylight.org/

[12] "OvS Open vSwitch, The Linux Foundation Colaborative Projects, August 9, 2016. http://openvswitch.org/

[13] Martın Casado, "Architectural Support for Security Management in Enterprise Networks," A Dissertation, Stanford University, December 2007 – January 2008.

[14] Martìn Casado, Scott Shenker, Michael J. Freedman, Justin Pettit, Jianying Luo, Nick McKeown, "Ethane: taking control of the enterprise", SIGCOMM, October 2007, Pages 1-12.

[15] Margaret Chiosi, at al., "Network Functions Virtualisation (NFV)," SDN and OpenFlow World Congress, Frankfurt-Germany, October 15-17, 2013.

[16] Martin Casado, Nate Foster, Arjun Guh, "Abstractions for Software-Defined Networks," Communications of the ACM CACM, Volume 57, Issue 10, October 2014, Pages 86-95.

[17] Patrick Hubbard, "The Network Is The Computer, Again," NetwokComputing, 05/06/2014. http://www.networkcomputing.com/cloud-infrastructure/network-computer-again/1827958867

[18] Hubert Zimmermann, "OSI Reference Model - The ISO Model of Architecture for Open System Interconnection" IEEE Transaction on Communications, Vol.28, Issue 4, April 1980.

[19] Thomas D. Nadeau, Ken Gray, "SDN Software Defined Networks; An Authoritative Review of Network Programmability Technologies," O'Reily, ISBN-978-144934230-2. 2013.

[20] Masato Okuda, Tetsuya Yamada, "Transforming Carrier Networks by Utilizing Network Function Virtualization," Fujitsu Sci. Technical J., Vol. 52, No.2, April 2016, pp.13-19.

[21] "Open vSwitch v.2.8.1," Linux Foundation Colaborative Project, 2017. http://openvswitch.org/features/

[22] Mihajlovic Radomir, Mihajlovic Aleksandar, "Operating Systems Security; The First Cut," Soft Electronics, New York, May 2015.

[23] Michael Mattsson (Feb 1996). "Object-Oriented Frameworks, A survey of methodological issues" PhD Thesis, University College of Karlskrona/Ronneby, Sweden, 1996.

[24] D. Mihajlović, R. Mihajlović, "Web Services & E-Commerce," E-Commerce, Palić, Serbia, Contribution 6, 05-07, April, 2006.

[25] Roy Thomas Fielding, "Architectural Styles and the Design of Network-based Software Architectures," PhD Thesis, University of California, Irvine, 2000.

[26] R. Fielding at al., "Hypertext Transfer Protocol -- HTTP/1.1," RFC-2616, June, w3.org, Network Working Group 1999.

[27] R. Mihajlovic, A. Mihajlovic, "Web 3.0, Ecommerce 2.0 and Internet Neutrality," 4th Intl. Conf. on Application of New Technologies in Management and Economics, ANTiM 2014, April 24-26, 2014, Belgrade, Serbia, pp.65-75.

[28] Andre Tost, Thomas Erl, Philip Thomas, Satadru Roy, "Building Standardized Service Contracts with Java," Service Technology Magazine, Issue LXXXV, July/August 2014.

[29] Barbara Liskov and Jeannette Wing, "A behavioral notion of subtyping", ACM Transactions on Programming Languages and Systems, (TOPLAS), vol. 16, #6, 1994.

[30] Andy Evans, Paul Sammut, James S. Willans, "Metamodelling for MDA," Proc. of the First International Workshop, York, UK, November 2003. https://www.cs.york.ac.uk/metamodel4mda/onlineProceedingsFinal.pdf

[31] M. Bjorklund, "The YANG 1.1 Data Modeling Language," Internet Engineering Task Force (IETF) RFC-7950, August 2016.

[32] R. Enns, M. Bjorklund, J. Schoenwaelder, "Network Configuration Protocol (NETCONF)," Internet Engineering Task Force (IETF) rfc- 6241, June 2011.

[33] F. Strauss,J. Schoenwaelder, "SMIng - Next Generation Structure of Management Information," IETF RFC-3780, May 2004.

[34] "TSC, OpenDayLight Governance," https://www.opendaylight.org/about/governance

[34] Chris Panagiotakopoulos, Antonis Lionarakis, Michalis Xenos, "Open and Distance Learning: Tools of Information and Communication Technologies for Effective Learning," Proceedings of the sixth Hellenic European Research on Computer Mathematics and its Applications Conference, HERCMA2003, Athens, Greece, 2003.

[36] "EYEDB Object Definition Language," SYSRA, December 2007.

[37] "OpenFlow Protocol Library Developer Guide," OpenDayLight Documentation, http://docs.opendaylight.org

[38] "OVSDB Developer Guide," OpenDayLight Documentation, http://docs.opendaylight.org/

[39] "Developing Apps on the OpenDaylight controller," OpenDayLight Documentation, http://docs.opendaylight.org

[40] DLUX http://docs.opendaylight.org/en/stable-carbon/developer-guide/dlux.html

[41] "OpenDaylight Release Plan Contents," OpenDaylight.org, https://wiki.opendaylight.org/view/Release_Plan#Current_Release:_Nitrogen

COMPUTER SIMULATION IN DOMAIN OF NATIONAL SECURITY: THE CASE OF S.E.N.S.E.

MIROSLAV D. STEVANOVIĆ
Academy of national security, Belgrade, mstvnv297@gmail.com

DRAGAN Ž. ĐURĐEVIĆ
Academy of national security, Belgrade, djurdjevic.dragan@gmail.com

Abstract: *In this article, we highlight the application of synthetic environment in SENSE training simulation for the needs of national security. The application of this simulation is problematised, functionally, from the aspect of adequacy of programmed algorithmic defining of realistic foreign policy situations and goals. We approached this problem by presenting an overview of simulation called Strategic Economic Needs and Security Exercise (SENSE) training method in comparative cases. We than observe the theoretical background of the SENSE simulation and subdue them to the constraints in real life matters of national security. This research shows that simulation training develops situational decision making, Since national security concerns perception of threats, and thus prior knowledge as immanent to rational decision making, we conclude that synthetic environment is a useful training tool for practical procedures in complex and interactive circumstances, but with constraints in terms of defining goals and situations.*

Keywords: *integrated strategic planning, simulation game, performance prediction, modelling capabilities, metacognitive skills*

1. INTRODUCTION

SENSE training program was conducted in Montenegro, from 22 to 26 February 1999, by Simulation to train leading national professionals and decision makers in various fields related to national security poses an expectation from the participants to adopt the behaviour and decision making patterns, within the programmed value concept. As such, it is an effort to upgrade some functional segments in the national security procedures.

The synthetic environment on which a simulation is built, thus, does not include complexity of parameters which drive decision making. Because of that, it can be claimed that it promotes individual situation awareness and reaction, but not the cognitive capacity.

The concept of national security involves defining of foreign policy situation, which is related to information gathering and analysis, and foreign policy goals, which include specific values for individual country. They cannot be ignored, nor assumed. Thus, it would be expected that such training simulations provide a realistic environment and model capabilities, through visualising and evaluating the outputs of individual actions within a system.

Application of simulation with synthetic environment, in the case of national security domain, bears a risk that it may foster operational skills without due respect for the role of knowledge for the perception of information in the context of threats for vital values. On the other hand, one of such simulations, SENSE, has found widespread implementation as a tool for upgrading individual capabilities of decision makers in a number of countries, including in some in the West Balkans.

2. SENSE IN PRACTICE

NATO. The training program starts with defining goals that will be valid for the day's exercise. They, by nature, concern the basic intentions, and serve as the frame on which the directions of future planning will be defined and the efforts of all entities that contribute to the goals of the state and the political community harmonised. After defining goals, the outcomes that must be avoided and the acceptable outcomes are determined, so that finally goals are set in a way that "they include positive outcomes, avoid negative outcomes and include important expectations". If expectations are better than reality, the bias is optimistic; if reality is better bias is pessimistic. They, thus, engage in a way they believe will result in positive outcomes (and outcome is an evaluation of their performance) and avoid actions that they believe may lead to negative outcome – basic school psychology, which disregards individual cognitive capability.

The expectations of others, from the inside (political parties, economic centres and interest groups) and from the outside (neighbours, the region, the international community) must be considered. This is a response to a reality that today many sources ask for their invoices, as regional and international influences grow, and the world becomes increasingly interconnected due to economic globalization.

Exercises within the program are performed by groups. Based on what they have heard, participants first evaluate the state of affairs in the state model (in SENSE the state model is called "Acrona"). They are expected to assess whether there are visions, explicit national goals and commitment to these goals. Then, the exercises proceed in

specific areas. For example, a group from an economic area solves issues, such as evaluating investing in a particular branch of industry or sector.

The exercise is conducted by the computer moderator, who opens the menu in which the situation given as the problem for the group exercise is seen, and the candidates make decisions and try to achieve them.

During this assignment, interaction with other participants is emphasised. On the practical level, each of certain responsibilities and actions in exercise are channelled through procedures simulating those in real life. Also, decisions are displayed on the screen, and then joint reviews are arranged, for certain fields (e.g. telecommunications). In addition to the tasks and objectives that every participant must carry out in the framework of individually designed duties, as the context of action, the political situation is also defined, given the set national goals. In the case of model applied in the simulation for Montenegro, the "government of Akrona" only constituted itself, provoked a revolt of the economy due to lack of communication and the prime minister resigned, and they had to start from the beginning, with government press conferences, etc. Interaction with other groups is achieved by informing about conferences of governments and ministers, as well as other subjects, such as banks. The results are evaluated daily, by displaying the individual curves of the results of each participant, the state of the enterprises, banks and the state [1].

In Bosnia and Herzegovina (B&H), the training was organized by the George Marsal Centre for Security Studies, based in Garmisch-Partenkirchen in Germany, and the NATO Counselling, Command and Control Agency based in The Hague in Netherlands. The design of the SENSE training program in B&H was based on the presumption that economic development could be the driving force behind other reforms; that the government must support entrepreneurship, i.e. must serve the country and avoid corruption; that the economic policy of the government can be effective if there is a national understanding and consensus on its directions and implementation; that properly led democratic processes help the economic development.

An option of the training programme can be related to negotiations. The simulation platform poses before participants to communicate and settle matters in various interactions and relationships. Groups engage in goal-oriented discussions — situations where people interact, not necessarily collaboratively, in order to accomplish tasks or settle on choice. Power differences among the participants constitute a crucial ... Postures Versus Powerful Roles. It covers main negotiation techniques and methods of conflict management. Another option is practical. Through it, participants are assigned to manage a fictional state in pre-set conditions. The model country finds itself in a difficult economic and political situation, and having just survived and armed conflict, is struggling with serious social problems, such as the return of refugees, AIDS epidemic, low level of education, poor environmental conditions, and high infant mortality rate. The participants are to manage various spheres of the public life to increase the safety and economic stability of Acrona, and provide decent life conditions for its citizens. (a specific position and the type of business it involves in virtual Akrona).

In Iraq, in 2004, United State Institute for Peace (USIP) introduced training programs for senior Iraqi national security officials designed to: a) refine skills in conflict analysis and resolution, b) adopt negotiation styles and techniques and c) perceive the roles that third parties can play in mediating. Institute trained 173 Iraqi officials from ministries such as defence (including military officers), foreign affairs, interior (including police generals), justice (including judges), planning, and finance. Based in part on the obtained feedback, USIP contracted the United States Institute for Defence Analyses (IDA) to enhance the simulation by scaling the imaginary country closer to Iraqi dimensions and include a significant role for the oil sector in the economy (with pipelines vulnerable to terrorist disruption), as well as additional elements that address managing the risk of inflation in an economy in transition [2].

National School of Public Administration, in cooperation with the Ministry of Foreign Affairs of the Republic of Poland, hosted SENSE training in Warsaw. From 2007-2011, the candidates were recruited from Belarus, Ukraine, Moldova, Georgia, Armenia, Tunisia, Myanmar, Afghanistan, Azerbaijan and the Kyrgyz Republic. The training is dedicated for civil servants, preferably of middle or upper rank, as well as for non-governmental representatives involved in public affairs and having strong professional background. All costs of the training are covered by the Republic of Poland, and the applicants are required fluency in English (USIP, SENSE in Poland, https://www.usip.org/education-training/international/sense-simulation/poland).

3. UNDELYING THE SENSE

SENSE security exercise, as an training method based on platform for on goal driven participant's engagement, primarily concerns the development of systemic decision-making process, to accommodate projected goals. It is therefore primarily in the function of developing of an automatised decision-making, in general. This function of training is conditioned by at least four recognised characteristics of described way of applying of artificial environment.

Firstly, the platform reveals the fundamental structure of decision-making, which is accommodated within a new automated process that exploits knowledge-based decision services [3].

Secondly, on strategic level, there is, on average, a much tighter link between analytics and decision-making within the process [4].

Thirdly, the program provides opportunities to foster cooperation (including on international level) as part of the national cyber security. It places the point of view of a participant within alternative ideas of courses as a foundation for planning [5].

Fourthly, it promotes delegation of activities to a range of software agents, which are them selves automated [6].

The processes in simulation follow the sequence identified by the pioneers of computational social science: assessing environment – goals formation – goal operationalisation – resource allocation – allocation finalisation – conflict generation – expectation formation [7]. The environment simulation training, generally, transposes decision-making process to information-processing rules (within defined objectives). Concerning application of this method in the field of national security in the described way, however, should be accepted with caution, since the constraints of economy on defence are relative, because over time defence needs tend to be partially compensated by decrease in investments.

The simulation exposes the game theory approach to understanding actors' behaviour. Game theory is a mathematical tool. It's application in international relations stems from recognition of a need of decision-making process to include how varying constraints will influence basic goals [8]. On the systemic level, this emphasises overall power resources at the disposal, and complies with realist theory of international relations.

Synthetic environment for training in policy development, market economy and representative authorities was applied for cadres in many countries in volatile regions, and in the USA. It evolved from a simulation for integrated strategic planning developed by the IDA, originally implemented in Africa, for developing conflict prevention institutional capacities [9]. During 1998, IDA convened a group of experts to review the work in developing the computer-based simulation game known as "Synthetic Environments for National Security Estimates". Their expertese was needed in the fields of macroeconomics, the transition economies, psychology and conflict resolution, and on-the-ground international aid activities [10].

The institutional framework within which SENSE was implemented is positioned in the American and NATO system. The Consultation, Control Command Agency (C-3) "directly contributes to NATO's ability to maintain peace and secure the collective defence of sovereign NATO members by providing impartial scientific advice and assistance to NATO's military and political leadership." The doctrinal context in which they work is defined in the following way: "Today NATO faces the challenge of a new strategic concept that implies not only a fundamental collective defence in NATO, but a new framework for cooperation and security in Europe, including peacekeeping operations and regional defence." The C3 was founded in 1996 by merging the Allied Joint Staff in Europe's Technical Center (SHAPE), which was in charge of "daily planning for actual operations" (among else, to provide support to NATO operations in the former Yugoslavia) with NATO's Communications Information System (NACISA). The George Marsal Centre is a product of the US-German partnership, the leading transatlantic defence education institution committed to creating a more stable security environment by promoting democratic defence institutions and a lasting partnership between America, Europe and Eurasia" [11].

4. CONCEPTUAL CONSTRAINTS

Apart from the "drill" of decision-making process participants, the effectiveness of training method should be estimated from the aspect of improving the capacities for analysis of trends and forecasting. The way to improve this quality in individuals, through simulations, is to upgrade their awareness in real time for the potential results of their decisions in realistic situations.

Early theoretical works in real-time simulations focus on the integration of individual efforts within a system, or a process [12]. Later works, however, raise doubts about the reality of situations imposed through platforms. They define as a challenge facing synthetic environment ignoring of impact of the effects of role of public health, warlords, re-emergence of conflict, donor fatigue, arrest, civil resistance, lack of institutional capacity that complicate social processes [13].

The general logic of the simulation theory relies on principles dilemmas underlying deconstructionism: is there a reality above individual perception, has it ever been reality, who is responsible for loss of reality, and can reality be reversed [14]. Thus, a simulation represents a complete process of the forecasting or replication, practically a virtual experiment [15]. From the aspect of social theory, simulation "reality", generally, boils down to hyperrealism.

In effect, training models which, like in the case of SENSE, impose international understanding and cross-cultural communication among decision-makers of diverse cultural and ethnic background through their platforms, drill towards universalising of foreign policies [16].

Since the training of estimating trends and forecasting involves the universalisation of policies, the analysis of a training method must include the element of dominance in international relations. Dominance can be exerted through agenda-setting, theoretically, methodologically; and through setting capabilities and abilities to formulate policy alternatives [17]. As far as implementation of foreign synthetic model is concerned, this poses a challenge twofold: a) establishing of an outside methodological dominance within national IR communities, and b) approach of cadres to international conflict resolution within predetermined technological constraints.

This challenge, from the aspect of functional capabilities of cadres developed through such programme, can be attributed to the method of simulation training. Namely, as shown, performance prediction promotes an increase in situation awareness, which includes the critical factors recognised in the platforms. Thus, it also promotes dynamic decision-making tasks through developing self-regulation, i.e. without conjecture and confrontation with other participants, but [18]. This challenge is emphasised because of fact that, from the aspect of applied algorithms, as scholars notice, SENSE does represent an advance in the use of parallelized algorithms and intuitive user interface [19]. In case of applying a synthetic environment for

assessing national security, leadership roles in grid operations and planning, big data handling, it creates a sense of opportunity to engage directly with major issues, unconstrained with immanent obstacles to idealised functioning.

5. CONCLUSION

For cadres in national security system, the aim of the simulation SENSE can be pre-projected to focus decision-making mechanisms concerning institutional procedures for directing the economic situation; negotiation and decision-making in a post-conflict environment; prevention of conflicts; and participatory management. They do upgrade individual response and aid functionality, as far as technical expertise and procedural automatization in resolving every-day problems and tasks are concerned.

On the other hand, there is an issue of expectation that such training improves individual capacity of rational reaction and adopting to circumstances. As shown, factors of decision-making in synthetic environment are significantly simplified. Definition of the foreign policy situation is related not subject to external evaluation, but rather to available information, also due to the lack of specific value in externally offered political goals, concerning decision-making; as well as since relations in the international community are assumed. Furthermore, by aiming to establish a consent on methods in approaching issues (economic, capacity of specific activities, dialogue and networking, and media), this synthetic environment allows key decision makers and students to knit together only within the existing commercial and global modelling capabilities, whose sustainability is seriously disputed even for the most powerful countries. For the capacity of an individual participant, this is beneficial from the aspect of ability to analyse and visualise the own outputs, in a way that the analyst can make sense of it. But, generally, these are metacognitive skills, and advance decision-making in comparison to desired without axiological ground.

On the general level, it can be induced that synthetic environment has a problematic approach to reality, to be suitable for every country's national security needs. Application of simulation with such environment bears a risk that it may foster operational skills and disregard the role of previous knowledge for the perception of information in the context of threats for vital values. SENSE, as the simulations in general, is useful tool for upgrading individual capabilities of decision makers, but it has an inherent constraint in form of marginalising knowledge and concentrating on predetermined criteria instead of on adaptability to reality.

REFERENCES

[1] P. Živković, "Kratak kurs državnosti: Kako su Crnogorci postali Akronci", Vreme, 437, 6. mart 1999. http://www.vreme.co.rs/arhiva_html/437/index.html.

[2] Congressional Research Service, State, Foreign Operations, and Related Programs: FY2011 Budget and Appropriations, Report for the Foreign Operations Appropriations Subcommittee, April 22, 2011, https://fas.org/sgp/crs/row/R41228.pdf.

[3] A. Fish, "Knowledge Automation: How to Implement Decision Management in Business Processes", New Jersey: John Wiley & Sons, 2012, p. 25.

[4] T. Davenport, "Linking Decisions and Analytics for Organizational Performance", in: Enterprise Analytics: Optimize Performance, Process, and Decisions Through Big Data, Thomas Davenport (ed.), New Jersey: Pearson Education Ltd, 2013, pp. 135- 154.

[5] A. Sillanpaa, H. Roivainen and M. Lehto, "Finnish Cyber Security Strategy and Implementation", in: Cyber Security: Analytics, Technology and Automation, Martti Lehto; Pekka Neittaanmäki (eds.), Cham: Springer International, 2015, pp. 129-146.

[6] T. Stevens, "Cyber Security and the Politics of Time", Cambridge: Cambridge University Press, 2016, p. 79.

[7] S. Bremer, "Simulated Worlds: A Computer Model of National Decision-Making", New Jersey: Princeton University Press, 1977, p.39.

[8] V. Aggarvai, P. Allan, "The Origin of Games: A Theory of Formation of the Ordinal Preferences and Games", in: Cooperative Models in International Relations Research, Michael Intriligator; Urs Luterbacher (eds.), Springer: Science & Media, 1994, pp. 299-326.

[9] L. Gehrig, L. Bateman and S. Ronis, "Nuclear Bomb Case Study", in: Vision Working Group Report and Scenarios, Sheila Ronis (ed.), Washington: Strategic Studies Institute, 2010, pp. 139-250.

[10] Institute for Defense Analyses (1998), "Synthetic Environments for National Security Estimates (SENSE)", Report of the Peer Review Group, Alexandria VA: IDA.

[11] J. Pina, "High-level seminar aims to create viable economy", SFOR Informer 80, 2000. http://www.nato.int/sfor/misc/sense/t000209a.htm

[12] R. Deyo, "Multi-body Vehicles", in: Real-Time Integration Methods for Mechanical System Simulation: Proceedings from NATO Advanced Research Workshop on Real-Time Integration Methods For Mechanical System Simulation, Snowbird, August 7-11, 1989, Edward Haug, Roderic Deyo (eds.), Berlin/Heidelberg: Springer Verlag, 1991, pp. 3-31.

[13] M. Flournoy, "Training and Education for Post-conflict Reconstruction", in: Winning the Peace: An American Strategy for Post-conflict Reconstruction, Robert Orr (ed.), Washington: Center for Strategic and International Studies Press, 2004, pp. 126-137.

[14] S. Cubitt, "Simulation and Social Theory", London / Thousand Oaks / New Delhi: Sage, 2001, p.80.

[15] H. J. Bungartz, S. Zimmer, M. Buchholz, and D. Pflüger, "Modeling and Simulation: An Application-Oriented Introduction", Berlin / Heidelberg: Springer Verlag, 2014.

[16] D. Crookall, P. Landis, "Global Network Simulation: An Environment of Global Simulation", in: Global Interdependence: Simulation and Gaming Perspectives:

Proceedings of the 22nd International Conference of the International Simulation and Gaming Association (ISAGA) Kyoto, 15–19 July 1991, David, Crookall; Kiyoshi, Arai (eds.), Tokyo / Berlin / Heidelberg / NewYork / London / Paris / Barcelona: Springer Verlag, 1992, pp. 106-111.

[17] H. L. Turton, "International Relations and American Dominance: A Diverse Discipline", Oxon / New York: Routledge, 2016, p. 26.

[18] J. H. Kim, "Simulation Training in Self-Learning: Investigating the Effects of Dual Feedback on Dynamic Decision-making Tasks", in: Learning and Collaboration Technologies: Designing and Developing Novel Learning Experiences: Proceedings of First International Conference, LCT 2014, Heraklion, June 22-27, 2014, Volume 1, Panayiotis. Zaphiris; Andri, Ioannou (eds.), 2014, pp. 419-428.

[19] V. Zagkanas, "SENSE changes the feel of simulation", Fieldscale, 24 November, 2016. https://fieldscale.com/sense-simulation-software-for-touchscreens/

SECURITY ISSUES IN DIGITAL CINEMA

ANDREJA SAMČOVIĆ
University of Belgrade, Faculty of Transport and Traffic Engineering, andrej@sf.bg.ac.rs

Abstract: *In the last few years, the emergence of new and efficient digital technologies in the film industry has also increased the need for better protection and security in digital cinemas, in order to prevent the illegal use and reproduction of films, or piracy, and to protect owners' rights. In this paper, we have focused on technology and technical methods that enable the protection of content in digital cinemas. It is also explained how methods such as watermarking, or link encryption, are used, as well-known methods for preventing abuse and illegal use of film materials. Also, some components that set up the security system of digital cinema are described.*

Keywords: *Digital cinema, information security, standards, digital service, secure communication*

1. INTRODUCTION

It can be said that the era of digital cinema (DC) began in June 1999, with the first public demonstration of motion picture in high resolution cinema. George Lucas' movie "*Star Wars: Episode I - Phantom Threat*" was screened at four cinemas in Los Angeles and New Jersey. Film was played on two screenshots on *Digital Light Processing* (DLP) projectors. The other two screens were using *Drive Image Light Amplifier* (D-ILA) technology. In the same month, the "*Ideal husband*" movie was also presented on two digital screens [1].

Since then, there has been advancement in technology and corresponding displays, merging standards-related activities by industry and governments. One of the areas where a significant advance in display technology was achieved is content security. The first presentations did not include content security, while the digital version of "*Star Wars: Episode II - The Clone Attack*" (the first major film that was completely done in digital format) was protected before distribution in May 2002.

The need for standardization is widely accepted. *Society of Motion Picture and Television Engineers* (SMPTE) set up the Digital Cinema Technology Committee in January 2000. In 2002, seven largest Hollywood studies established the Digital Cinema Initiative (DCI), which made the Digital Cinema Specification, the first official standard in this field [2].

This paper focuses on the security of content in digital cinema. The term security refers to the prevention of unauthorized use of film content. This implies the prevention of piracy and the prevention of illicit exposure. It should be noted that not all decisions on legal exposure can be approved in advance, so that the content review mechanism should be monitored by the content security system.

This paper is outlined as follows. We first introduce the general security requirements in digital cinema environment. We summarize typical security treats over the corresponding security tools. Cryptography tools and link encryption are described in the next subsections. Then, we will emphasize the importance of key management and physical security for DC applications. Finally, encrypted transport of digital film materials is pointed out. Proposals for future work conclude the presentation.

2. SECURITY REQUIREMENTS

To make the digital cinema successful, many people in the industry believe that interoperability standards are necessary. At the beginning, the focus was on the overall security of the system using strong protection, link encryption, watermark, abusive hardware and access to certificates and time servers [3].

The content of a movie (moving images) is distributed from one user to another in the form of a business arrangement. The recipient agrees to pay a certain price in exchange for the authorization to use the content in a particular form. For example, the cinema can display the film up to 15 times a week for 5 weeks. They are expected to return the film after these 5 weeks. The cinema should not display the film more than the agreed 15 times in a given week. Of course, depending on the success of the film, these conditions can be changed during the contract.

There may be other conditions in the contract, too. The cinema does not have the authority to copy the film and is responsible for ensuring not to make copies and that the film would not be stolen physically. The cinema also has no authority to change the film. The film may be required to display as a whole without interruption. In other words, cinemas are not allowed to advertise or interrupt, or to show a film in multiple parts. A business contract may also require display a film at the minimum number of times a week or hours during the day. Having in mind that this is a business agreement between the two parties, where one offers intellectual property and the other agrees to pay for these rights, the conditions may be either strict or not, depending on how the parties agree. Traditionally, these contracts are protected by laws and

social mechanisms. Legal protection includes laws on contracts. However, primary social protection is trust. Cinemas are believed to adhere to contractual obligations. Violation of these obligations will destroy the reliability of the cinema and can lead to its exclusion from future business deals.

Technology plays its role in protecting commercial contracts in the sense that it would require access to expensive equipment to make copies of 35mm film. Possession of such a copy would allow cinemas to show the film on multiple screens at the same time or to continue to show films after the expiry of the contract, ie. after the film is returned to the distributor. If a copy is taken from the cinema, it may be displayed in other places that are not contractually obliged to pay compensation to the owner for the use of a particular format (eg. *Digital Versatile Disk* - DVD) and for distribution to users, again without the permission of the content owner or without compensation to the owner.

Moving from analog to digital cinemas does not change legal or social mechanisms. However, it significantly changes the impact of technology in two very important ways. Firstly, there is technological barrier to copy and redistribute the film. If it is assumed that the film is stored as a file or as a certain number of files, the computer equipment for copying is available to a large percentage of the population. With broadband internet, copies can be anonymously distributed in most cases. The second way is to use technology to apply some of the terms in the contract.

3. SECURITY TOOLS FOR DIGITAL CINEMA

There are several technologies that prevent the unauthorized use of content in digital cinema [4]. This includes cryptographic tools for providing digital content during transport and for ensuring its integrity during download. Physical security technologies can be used to store critical issues. Optical techniques prevent camera-based video capture and forensic tools are used to track piracy sources.

3.1. Cryptographic tools

The basic tool for protecting digital film from unauthorized use is encryption. A digital file containing a film is encrypted before distribution and stored in such a condition in the cinema. Due to relatively large film files, symmetric encryption keys are generally recommended. This enables the need for each film to be specifically encrypted, but its security is based on a small set of symmetric keys that must be distributed to cinemas. These keys are delivered with a set of conditions, explicit and implicit, which must be satisfied before the use of keys is justified. Explicit conditions may include a time window in which the film can be played. Implicit conditions may require the equipment to demonstrate that it is certified and not used in an unauthorized manner. Careful key management can ensure that the film is not available to people who abuse the film.

Possible film redistribution would not be possible without decryption the content. The described mechanism will prevent unauthorized access during transport and storage, but will not protect content from an unauthorized access as it is decrypted. Several technologies can be used here: link encryption, hardware that is resistant to unauthorized use and forensic labels.

3.2. Key management

Careful key management is a difficult challenge. Film security depends on the security of the protection keys. The encryption technique should use keys that are sufficiently long to make an attack with all possible keys. A large length of keys would also mean that the probability of random keying would be very low. The actual key selection should keep this resistance to attack by avoiding any schemes that can be learned in the key sequence [5].

Valid keys, once they have been generated, must be protected by encryption. This is necessary, because they are distributed to cinemas and stored for further use. At that time, the key management system must provide the appropriate authorized access modules to the keys and use these keys to access the content. Several technologies can be used here. First, asymmetric cryptography allows a system that manages the key to perform key encryption in such a way as to ensure that only authorized target modules can get the key. To make this safe, an authorized module must be sure that its private key will remain secret. This is often achieved with the help of hardware that is resistant to unauthorized use. The key management system must also be sure that this module is authorized. This leads to the use of authentication protocols and allows the de-authentication of the module through the use of recalling lists. Authentication is generally achieved using digital certificates.

3.3. Link encryption

A security system for digital cinemas can consist of several different components, each of which can be individually secured with different hardware mechanisms. While content may be considered to be protected, in the physical vicinity of each component steps must be taken to protect the content when it passes from one component to another. The tool that is mainly used for this type of protection is encryption.

Asymmetric cryptography is not effective for very large data, as is the case with digital films. That's why symmetric keys are more often used. There are standard methods for devices or processes to achieve a unique symmetric key (session key) using digital certificates and asymmetric cryptography. An example of such a standard method is IETF (*Internet Engineering Task Force*) RFC 3447 [6].

3.4. Physical security

It's very difficult to keep a secret when a hardware or software process contains this secret. The software can be

decomposed until hardware can be disassembled. There are several things that must remain secret in the digital cinema system, including text content, primary decryption keys, encryption keys and private keys that are associated with any component of the security system.

There are many applications where cryptography is used to provide operating of computers and communication systems. All these applications encounter the same problem, keeping their secrets. In response to this problem, several approaches appear to create systems that are resistant and can detect unauthorized access or respond to such approaches by destroying classified information. The security requirements for the implementation of cryptographic modules are standardized.

3.5. Inserting cameras

Cryptographic systems can protect digital content while transporting or storing, but at some point the film must be decoded. Decompression must be done and the film must be converted then for human use. This step leaves the content vulnerable to camera recording.

Camera videos are a significant source of pirated movies. In recent research, it was shown that 16% of films available in the P2P (peer-to-peer) network were recorded with a camera in the cinema [7]. These pirated copies were, in average, available nine days after the cinema premiere. In some cases, the recording is reflected after the closing of the cinema. The person who plays the movie can have a camera at the back of the cinema. The result is a high-quality shot. Furthermore, the sound can be recorded directly from the cinema audio system. In other cases, recording is done during normal screening, with audience in the cinema. Even in these cases, high-quality sound can be captured using headphones that are available in many cinemas. There are several ways to get on this path of piracy, including time modulations, color fusion, *Charge Coupled Devices* (CCD) saturation of sensors using infra-red light, etc.

3.6. Forensic monitoring

It will not always be possible to prevent the creation of pirated copies of movies. The more skilled pirats will find a way to bypass the technology against camera shooting and will successfully capture movies. An unrecognized employee in the cinema can detect weaknesses that can be used to disable or circumvent security features. The security device can be canceled in unpredictable mode. Deep analysis can help to reduce such events, but it cannot be expected to be completely eliminated. When a pirated copy of the film is discovered, then mechanisms should be activated to keep the copy traced back to its source. In the case of camera shots, the cinema and the screen where the film was filmed, and even the time and date of the shot could be recognized from the pirated copy. The security equipment that was used, as well as the staff that was in the cinema on that day, can be identified. It will not only detect and remove unreliable equipment and/or employees from the system, but will also have an effect to reduce the piracy.

The combination of two technologies can be used in this situation. First, a digital watermark can be added to the film during the projection. The watermark is imprinted on the content, but the time and location of the screen can be recognized. This information can be viewed on a pirated copy, even if the quality of the copy is much worse than the original. Watermark technology is rapidly advancing. Another technology used to analyze violations of security procedures that lead to piracy is security records or revisions. Secure record is a mechanism for tracking all cinematic operations, including the projection events themselves, i.e. release movies. The mechanism is resistant to erasing data, altering or falsifying entries. Rights holders may require the cinema to send copies periodically of these records. They can then confirm the integrity and continuity of records. Records will show how many times and when the movie is released. It is expected that this will be in line with the business agreement between the cinema and the right holder of the film. It can also be seen which equipment is used and the identity of the operator of that equipment (using smart cards or biometric identification cards).

4. ENCRYPTED TRANSPORT

The content of the film is encoded for transport from the distributor to the cinema [8]. In fact, the content must remain encoded in the cinema and can only be decoded at the time of display. Images, sounds and translations are encoded separately with different encoding keys. Encrypted symmetric key is used to transfer film content. More specifically, the specification indicates that the AES (*Advanced Encryption Standard*) uses the *Cipher Block Chaining* (CBC) mode with a 128-bit key. Distribution of the film content is shown in Figure 1.

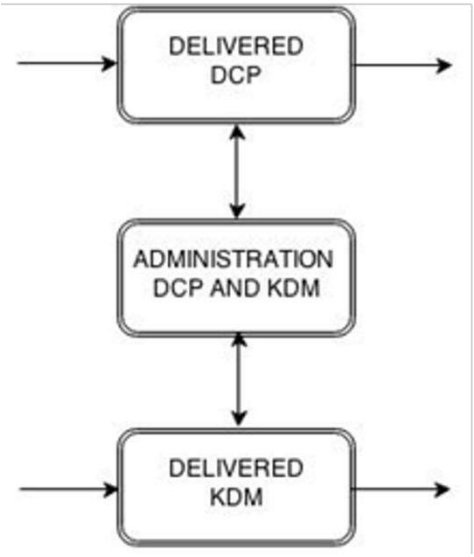

Figure 1: Distribution of film content [8]

AES standard is based on an encrypted symmetric block known as Rijndael. The encrypted symmetric block

algorithm can be used with large numbers for different regimes. CBC mode implies binding of each basic text of the data block with the preceding encrypted text data block before the application of the next block of the code. In the binding process, exclusive disjunction is used (XOR). Digital cinema packages can then be transmitted through unsecured networks without the risk of exposure to plain text content, thanks to encrypting content with such powerful codes. This means that real transport through the network is not relevant to the security of a digital cinema, as long as the transport is successful. In fact, content security follows transport and can be achieved either electronically or by physical means.
The existing courier distribution network can continue to exist. Instead of transporting the film by cable, it can be transferred by solid devices or optical disks. The most accurate film transmission is by satellite or through an optical cable or even via the public internet. As with most security-based encryption systems, a secure content key delivery is also ensured. A key advantage is that the critical data load is reduced.

4.1. Transport of the key

Digital cinema packages are delivered to the theater with three main components: picture, sound and encrypted translation [9]. It can also be delivered to the cinema via a separate channel, at different times. This type of special delivery is called *Key Delivery Message* (KDM). KDM carries an important content, among them is the set of keys needed to decrypt the content of the film. This content is encrypted so that interception with KDM can not endanger the security of the movie. To ensure that only the user can open the content, encrypted public key is used. This means in fact that the distributor can transfer the same encrypted film content to the cinema. Each KDM will be unique and designed for a single receiver. In fact, each projector will get a unique KDM encrypted load with a public key, to control the *security manager* (SM). The process of transport of the key is presented in Figure 2.

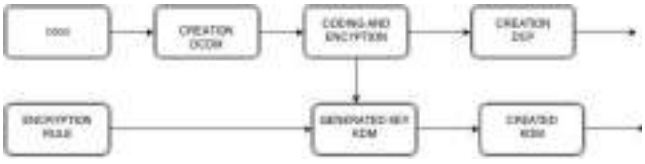

Figure 2: Generating of Key Delivery Message [8]

The DCI specified that the RSA (*Rivest, Shamir, Adleman*) algorithm will use a 2048-bit encrypted public key to encrypt the KDM data load capacity. RSA algorithm is the most commonly used system for decryption, and with such a key length it is considered as extremely safe. Once a security manager decodes KDM content and gets the decryption keys, it must ensure that no content or keys can be accessed by any other device or competitor. This usually means that they are stored in secure hardware. If the security manager has the need to use the keys periodically, it must protect them with the encrypted symmetric key. DCI-compliant security manager can choose between 128-bit AES keys or Triple-DES (TDES) - 112-bit keys.

4.2. Security Manager

The DCI contains devices that exist in the security system of the cinema, but the most important of them is the security manager. The security manager coordinates with all the devices within the security system. It is responsible for identifying other security devices and their authenticity. It establishes a secure communication channel with other security devices. At the end, the security manager confirms the authority of each security device. Figure 2 represents various security modes within the security system.

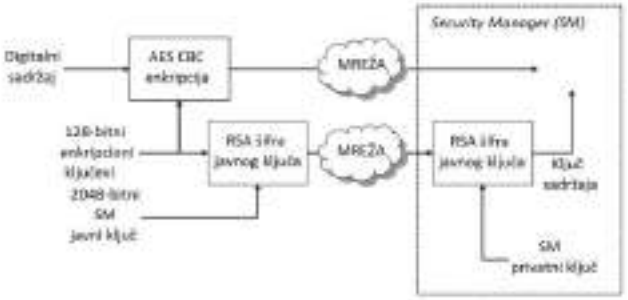

Figure 3: Various security modes

The security manager is responsible for ensuring the compatibility of the DCI device. This is achieved by reviewing the digital certificates offered by the devices. A device certificate is issued by the device vendor, encrypted with a private key, and contains a public key connected to the device. It contains a set of data such as manufacturer, device model, serial number of the certificate, which is *universal uniquely identified* (UUID). The certificate also contains information that defines a set of rules that the device may contain in the context of the digital cinema security system. By decrypting with the public key of the certificate provider, the security manager can confirm that the certificate is issued by the vendor and can obtain a public key as well as details related to the device. The certificate represents the supplier's statement that the device is really a DCI device specification.
Instead that the vendor encrypts the entire certificate with his private key, the alternative is to have a digital signature on the certificate. A large number of digital signature algorithms are available, but in all cases, the vendor calculates the cryptographic overview of the certificate and encodes with its private key. Digital signatures can be decrypted with the public key of the vendors and it can be compared with the certificate. This is confirmed by the signature creator and the integrity of the certificate. The use of digital certificates is in line with the joint use of a third party guarantee in relation to one's identity. The format is based on frequent use of the IETF format, X.509. The digital certificate must use the encrypted RSA public key to encrypt the SHA-256 certificate.

4.3. Secure communication channels

The security manager is responsible for secure communication channels and each of the checked devices in the digital cinema security system. The channel establishment process includes cryptographic protocols in which a key session is generated and replaced with private/public keys. This process ensures that no eavesdropper can detect which key is used between the two devices. The described process is used for secure internet communication, and DCI has adopted a well-known TLS (*Transport Layer Security*) protocol. Since the TLS protocol requires the exchange of digital certificates and their authenticity, and the establishment of the TLS session includes the step of authentication of the described devices. In other words, while identity is checked the creation of secure communication devices is actually done simultaneously with the establishment of the TLS session.

When TLS session is established, the security manager knows the identity of each security devices in the security system. In the DCI digital cinema environment, the security manager will only trust to those special devices that it owns the rights. The rights owner maintains a list of secure devices for each cinema and any projection inside the cinema. When compiled, the KDM program contains a decryption key for a specific security manager, and rights owners will include a list of safe devices depending on projection in the cinema. This list is called TDL (*Trusted Device List*), because TDL is sent to KDM. It can be different for each projection environment and each composition of the projection. Specific security devices are authorized to participate in the reproduction of connected compositions and to determine the roles for selected device. Owners can 'abolish' the trust of a device or family of devices by simply removing them from all TDLs.

Security Manager, like other security devices that participate in the security system of a digital cinema, must be implemented in secured hardware. Physical security requirements for digital cinema system components are based on those specified by secure cryptographic modules. This standard describes the requirements for the operating system, as well as surrounding physical protection for different levels of security. Among other things, it is required from the device to have the ability to detect unauthorized attempts to use the material and to respond to these attempts by adjusting and protecting all sensitive data including cryptographic keys. The Digital cinema security system further requires that all security chips report any unauthorized attempts to manipulate data.

Decrypted content may need to be transferred from one security device to another and this content must be protected by encryption of the connection. The DCI specification claims that the encryption of the connection can be achieved by AES with a 128-bit key or a 112-bit TDES. A symmetrical AES key used for this occasion is generated by a security manager. This process of generating the key must be safe, so it must follow the process specified by the IETF RFC 3447.

Copyright owners expect from a cryptographic security to check how and when the film is displayed. They also expect to see and record the status of the security system. In order to make such a check, all security devices must record events related to security. Examples of such events include establishment and completion of TLS sessions, as well as the beginning and end of decryption, and decoding or the forensic marking process. At the end, the outcome is that any detection of manipulation or any activity on the security components is recorded.

The log entries are marked and assigned to the device. This provides irrefutable and indelible evidence and protection against deletion or modification of log entries. Upon request, each security device will forward its log to the security manager. Such communication is carried out in the context of the TLS session in order to further maintain the integrity of the record.

Records or entries are stored in a pre-defined XML (*EXtensible Markup Language*) format that contains a header, a useful part or content, and a signature. The header contains information about the sequence, type of record, and timestamp. Signature authenticates the header. When all records are collected from various security devices, the security manager can forward the log to the management process. The structure of log messages allows the filtering operation to be applied, and at the same time, the continuity of the guarantees is maintained. Filtering is necessary to create reports that can be sent to film owners.

The filtering operation removes a part of the logs from a log that is not as significant, for example, the specification of the time frame, the type or class of the entry, or the device that created the record. However, the headers and signature remain. Signals for removed entries continue to authenticate headers, and testing these headers can confirm that no entries have been deleted.

DCI specified that a compatible system for displaying a digital cinema must have the ability to incorporate digital watermarks into the film components, both audio and video, but it is not specified which watermark technology should be used. A minimum of useful information and accuracy requirements is specified, but there is no significant rigidity and limit of requirements. Furthermore, the DCI specification does not look at the term watermark security as an ability to prevent unauthorized use of the film. The only exception is the assertion that the watermark in the image is required to prevent submerged attacks. All these problems have been left to the market.

5. CONCLUSION

An important step in the development of digital cinema technology is the adoption of the necessary standards. This indicates the willingness and maturity of the players involved in this field to harmonize their architecture and interfaces. Standards are necessary to ensure interoperability and encourage cinemas to invest in the development of associated security equipment and devices.

SMPTE has played a key role in this process by establishing an early-stage architecture based on security by providing a legal forum for film owners as well as technology vendors to discuss on security issues. Undoubtedly, in the future, digital technologies will

continue to develop, as well as the accompanying standards, so that the film industry, having followed such trends, will have to develop simultaneously and find better and more affordable security solutions in digital cinemas.

REFERENCES

[1] R. Dettmer, "Digital cinema: A slow revolution", *IEE Review*, Vol. 49, No. 10, pp. 46-50, 2003.

[2] Digital Cinema Initiatives, LLC Member Representative Committee, *Digital cinema system specification v.1.1*, 2007.

[3] J. Bloom, "Security and rights management in digital cinema", *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing ICASSP 2003*, Vol. 4, pp. 712-715, 2003.

[4] J. Bloom, "Digital cinema content security and the DCI", *Proceedings of the 40th Annual Conference on Information Sciences and Systems 2006*, 2006.

[5] F. Lu, "A study on key delivery message system of digital cinema design", *Proceedings of the International Conference on Education, Management, Commerce and Society EMCS 2015*, pp. 207-210, 2015.

[6] Jonsson, Kaliski, IETF RFC 3447 PKCS #1: "RSA *Cryptography Specifications*" 2003.

[7] A. Maltz, "How do you store a digital movie for 100 years?", *IEEE Spectrum*, Vol. 51, No. 3, pp. 40-44, 2014.

[8] A. Samčović, "A review of the digital cinema chain – from production to distribution", Chapter 12 in *Emerging research on networked multimedia communication systems*, IGI Global, pp. 366-394, 2016.

[9] C. Fairall, H. Edmunds, D. Cave, "BFI national archive: Digital workflow for the preservation of digital cinema packages", *Journal of Digital Media Management*, Vol. 2, No. 2, pp. 127-136, 2013.

SECURING MACHINE LEARNING CLASSIFIERS WITH INPUT HASHING RE-WEIGHT STRATEGY

IGOR FRANC
Belgrade Metropolitan University, Faculty of Information Technologies and SECIT Security Consulting, igor.franc@metropolitan.ac.rs

NEMANJA MAČEK
School of Electrical and Computer Engineering of Applied Studies, Belgrade and eSigurnost Association, Belgrade, macek.nemanja@gmail.com

MILAN GNJATOVIĆ
University of Novi Sad, Faculty of Technical Sciences, milangnjatovic@uns.ac.rs

BRANIMIR TRENKIĆ
School of Electrical and Computer Engineering of Applied Studies, btrenkic@viser.edu.rs

MITKO BOGDANOSKI
Military Academy General Mihailo Apostolski, Skoplje, Macedonia, mitko.bogdanoski@ugd.edu.mk

DRAGAN ĐOKIĆ
Belgrade Metropolitan University, Faculty of Information Technologies, dragan.djokic@metropolitan.ac.rs

Abstract: *Adversarial machine learning resides at the intersection of machine learning and computer security. Originally, machine learning techniques were designed for environments that do not assume the presence of an adversary. However, in the presence of intelligent adversaries, this working hypothesis is likely to be violated to at least to some degree, depending on the skillset of an adversary. A skilful adversary can carefully manipulate the input data exploiting specific vulnerabilities of learning algorithms. This results in misclassification of malicious instances, which may compromise the whole system security. For example, by carefully modifying values of features with largest weight without changing the outcome of malicious packet, an adversary may trick an intrusion detection system to allow malicious packet into the network. Solutions presented in research studies by other authors consider the classifier protection using re-weight strategies; typically, this results in compromise between accuracy and robustness. Unlike those, the research presented in this paper deals with a re-weight strategy based on hashing all the numeric features without classification accuracy degradation. System becomes robust as feature weights are even and avalanche effect makes virtually impossible for an attacker to modify the input data and trick the learner into misclassification. Research hypotheses are experimentally validated on custom intrusion detection dataset consisting of numeric features.*

Keywords: *Machine Learning, Adversarial Learning, Hashing*

1. INTRODUCTION

Machine learning algorithms independently collect knowledge from the machine readable information, i.e. they learn from data. Such algorithms build a model from example inputs and use it to make predictions or decisions [1]. Tom Mitchell provided a widely quoted, formal definition of machine learning: "A computer program is said to learn from experience *E* with respect to some class of tasks *T* and performance measure *P*, if its performance at tasks in *T*, as measured by *P*, improves with experience *E*" [2].

There are two types of machine learning algorithms: unsupervised (no "teachers") and supervised (with "teachers"). Unsupervised algorithms learn from unlabeled examples; the objective of unsupervised learning may be to cluster examples together on the basis of their similarity [3]. Unsupervised learning is suitable for finding patterns in the data. Supervised learning algorithms build a model from a training set (given in the form of feature vectors) with class label assigned to each instance. Once trained, supervised algorithms assign class labels to previously unseen examples of the same task, on the basis of the formed model [4].

In the case of machine learning, most classification algorithms were developed to learn and operate in secure, controlled environment. Transition from controlled environment to a potentially hostile environment may result in significant security failures [5].

In adversarial classification tasks like spam filtering and intrusion detection, a skilful malicious adversary can carefully manipulate the input data exploiting specific

vulnerabilities of learning algorithms. Thus, aside from achieving good classification performances, machine learning algorithms have to be robust against adversarial data manipulation [6].

2. ADVERSARIAL LEARNIG: ATTACKS ON CLASSIFIERS AND COUNTERMEASURES

Potential vulnerabilities of the learning system in adversarial environment can be categorized according to the influence on the classifier, security policy violation and specificity of the attack [7].

According to the influence on the classifier, attacks can be either causative, if they are aimed at compromising the training phase of the classifier, or exploratory, if they are carried out at the classification phase, with the aim to gather knowledge about the classifier (the knowledge can, for example, be gathered from feedback on class labels assigned to malicious traffic instances.) According to the security policy violation, attacks can be either integrity violation, if the adversary's goal is to have malicious instances classified as legitimate activity by the system, causing false negative rate to increase, or availability violation, if the adversary's goal is to perform DoS attacks and render the classifier useless. Specificity of the attack refers to the scope of malicious samples that will be misclassified as legitimate activities. According to the specificity, attacks can be either targeted, if the adversary's goal is to have a limited range of malicious instances misclassified, or indiscriminate, if the adversary's goal is to have all malicious samples misclassified as legitimate.

When designing a machine learning based security mechanism, it is necessary to identify system vulnerabilities, the possibilities to execute attacks on a system that will exploit these vulnerabilities, and the consequences of successfully executed attacks. In other words, system architect should take the role of adversary and try to anticipate all possible attacks, such as compromising the training set or detection evasion.

Several defence techniques against attacks on the classifier are proposed in [7]: regularization (defence from causative attacks), randomization (defence from targeted attacks) and information hiding or disinformation (defence from exploratory attacks). In some circumstances, the learning system may alter the information seen by the adversary, thus providing the adversary with a misleading picture of the classifier. More sophisticated systems can lead the adversary to believe that a certain type of attack is not included in the training set. Allegedly “allowed attack” will lead the adversary to reveal himself.

There are several ways to detect attacks on the classifier. For example, exploratory attacks can be identified by running a separate clustering algorithm against the classified data: the sudden appearance of a large cluster near the decision boundary could indicate probing attacks on machine learning based intrusion detection system. Detecting attacks on the classifier is very important because it provides information about adversary's capabilities; this information can be used to re-adapt defence strategies.

3. RE-WEIGHT STRATEGIES

One of the re-weight strategies that improves the robustness of linear classifiers in spam filters is proposed in [8]. The technique is based on the normalization of feature weights, thus avoiding over-emphasizing or under-emphasizing feature weights. If there are features with over-emphasized weights, the adversary will adapt the values of most significant features (features with the highest impact on classifier decision) and trick the learner to classify malicious as legitimate instance. These attacks are typically executed on spam filters (for example, by increasing the number of innocent words), but are also feasible to evade detection by intrusion detection classifier.

If the distribution of feature weights is uniform, an adversary will have to change more feature values to trick the classifier. If adaptation of each feature requires the same effort, then this re-weight strategy increases the robustness of the classifier, i.e. its resistance to attacks based on exploiting knowledge about the decision function. However, the proposed technique is a compromise between accuracy and robustness of the classification system.

Re-weight strategy presented in this paper is based on hashing all the numeric features, both in the training and operational phase. By doing so, system becomes robust resulting feature weights become even and avalanche effect resulting from hash function is a huge troublemaker even for a skilled adversary. Aside, if generated hash is bitwise longer than the original input value, input data is virtually casted into higher-dimensional space (similar to Support Vector Machines’ kernel trick.) According to experimental evaluation given in this paper, we can conclude that this re-weight strategy is applicable to several machine learning based security mechanisms, including, but not limited to intrusion detection systems. The only downside of the proposed solution is that it operates only with numeric features, thus not being applicable to systems that operate with categorical features without additional feature vector transformation.

4. APPLICATION TO CUSTOM INTRUSION DETECTION DATASET

The intrusion detection dataset used in this research is built from traffic captured on the simulated virtualized networking environment. Synthetic dataset consists of normal, healthy traffic and a number of successful exploitations of unpatched Windows XP operating system, executed with variety of open source and commercial software products. Both healthy and malicious traffic have been recorded separately and cleansed from virtualization protocol and service leftovers (noise removal). This reassembles a scenario for two-class supervised learning problem. Numerical features (representing statistical data) were extracted from PCAP files and data instances were created, labelled and shuffled into a separate training and test sets, both consisting of 10.000 instances.

Following machine learning algorithms were used to train models and classify test sets: decision trees [9, 10], Naive Bayes [11, 12], Random Forest [13] and AdaBoost [14] using C4.5 decision tree as the base learner. See references

[9-14] for more details on aforementioned algorithms. Hashed values were calculated using MD5 and Secure Hash Algorithm (SHA-2).

Let TP, TN, FP and FN denote number of true positives, true negatives, false positives and false negatives [15]. Accuracy of the classifier is calculated with the following equation:

$$a = \frac{TP + TN}{TP + TP + FP + FN}. \quad (1)$$

A simple algorithm used to calculate feature weights operates follows [15]: let *a* denote the accuracy of classifier trained with all features, and let a_i denote the accuracy of a classifier trained with all features except feature *i*. Accuracy change for that classifier is given with the expression:

$$\Delta a_i = a - a_i. \quad (2)$$

The smallest and the largest accuracy changes are given with expressions (3) and (4):

$$\Delta a_{min} = min(\Delta a_i), i = 1, \ldots n \quad (3)$$

$$\Delta a_{max} = max(\Delta a_i), i = 1, \ldots n, \quad (4)$$

where *n* denotes the number of features in the dataset. Feature weight w_i of the feature *i* is given with the equation:

$$w_i = \frac{\Delta a_i - \Delta a_{min}}{\Delta a_{max} - \Delta a_{min}}. \quad (5)$$

All feature weights are scaled to a range [0, 1]. Classifier used to calculate weights is C4.5 decision tree.

Experiments with original input data

Models have been trained and tested using Python with scikit-learn. The first set of models have been trained with the original datasets (feature values have not been hashed). Test set classification accuracy is given in the table 1.

Table 1: Classification accuracy (original input data)

Algorithm	Accuracy (%)
C4.5	81,43%
Naive Bayes	85,68%
Random Forest	95,19%
AdaBoost	93,87%

Feature weights of 27 features are presented on graph on Image 1. As one may see, the distribution of feature weights is not uniform, as there are features with over-emphasizing or under-emphasizing weights. In this scenario, the adversary familiar with the classifier and statistical features of data instances will adapt the values of most significant features and trick the learner to classify malicious as legitimate instance.

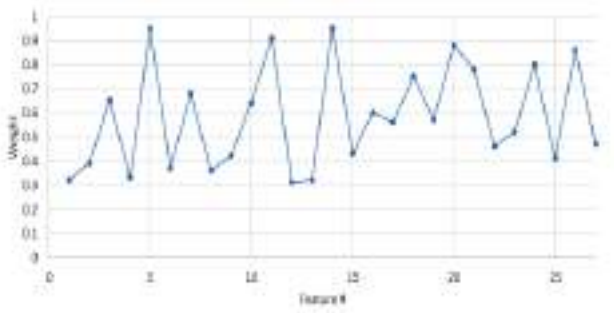

Image 1: Feature weights scaled to range [0, 1], original input data

Experiments with hashed input data

The second set of models have been trained with two datasets containing values from original dataset processed with MD5 and SHA-2 hash functions. Feature values in the test sets have been also processed with the aforementioned hash functions. Test set classification accuracy is given in the table 2. According to the result, one may notice that the degradation of accuracy is almost negligible.

Table 2: Classification accuracy (hashed input data)

	Accuracy (%)	
Algorithm	MD5	SHA-2
C4.5	79,95%	79,89%
Naive Bayes	84,52%	85,01%
Random Forest	94,07%	93,87%
AdaBoost	91,77%	90,82%

Feature weights of 27 features are presented on graph on Image 2. As one may see, the distribution of feature weights is now uniform, as there are no features with over-emphasizing or under-emphasizing weights. Due to uniform distribution is uniform, an adversary will have to change more feature values to trick the classifier. As the classifier operates with hash values it may be concluded that (1) adaptation of each feature requires the same effort and that (2) due to the avalanche effect, minor adaptations in original data before hashing in the IDS will result in major change in the resulting data, making it virtually impossible for a skilled attacker to trick the classifier.

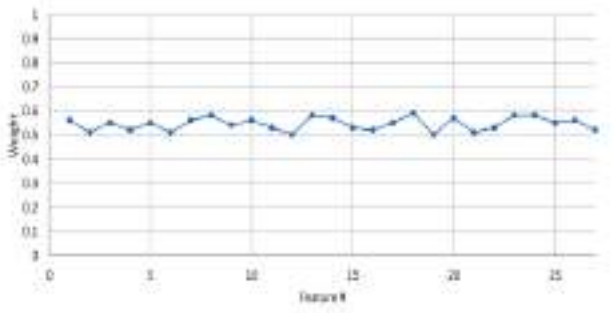

Image 2: Feature weights scaled to range [0, 1], hashed input data

5. CONCLUSION

In this paper we have presented an approach to feature re-weight strategy based on hashing all the numeric feature values. Classifier is trained and tested with all feature values of each instance being previously processed by a hash function. Hashing provides a robust classifier as resulting feature weight distribution is uniform and avalanche effect resulting from hash function is an obstacle even for a skilled adversary. Re-weight strategy presented in this paper is applicable to several machine learning based security mechanisms. However, the downside of the proposed solution is that it operates only with numeric features, which means that additional feature vector transformation is required if categorical features exist.

REFERENCES

[1] M. A. Hall and L. A. Smith, "Practical feature subset selection for machine learning," in C. McDonald (Ed.), Computer Science '98 Proceedings of the 21st Australasian Computer Science Conference ACSC'98, Perth, 4-6 February, 1998, pp. 181-191. Berlin: Springer.

[2] T, Mitchell, "Machine Learning", McGraw-Hill Science/Engineering/Math, page 2, 1997.

[3] Z. Ghahramani, "Unsupervised Learning," in O. Bousquet et al. (Eds.): "Machine Learning", LNAI 3176, Springer-Verlag Berlin Heidelberg, 2003.

[4] I. Hendrickx, "Local Classification and Global Estimation: Explorations of the k-nearest neighbor algorithm", PhD Thesis, Tilburg University, 2005.

[5] T. Woods, M. Evans, D. Rust, and B. Podoll, "Security in Machine Learning: Measuring the relative sensitivity of classifiers to adversary-selected training data," CSCI 5271 Project Final Draft, University of Minnesota, Minneapolis, USA, 2008.

[6] B. Biggio, B. Nelson, and P. Laskov, "Support vector machines under adversarial label noise," in Asian Conference on Machine Learning, pp. 97-112, November, 2011,

[7] M. Barreno, B. Nelson. R. Sears, A. D. Joseph and J. D. Tygar, "Can machine learning be secure?", in Proceedings of the 2006 ACM Symposium on Information, computer and communications security, pp. 16-25), 2006.

[8] A. Kołcz, and C. H. Teo, "Feature weighting for improved classifier robustness", in CEAS'09: sixth conference on email and anti-spam, 2009, no pagination.

[9] L. Breiman, J. H. Friedman, R. A. Olshen and C. J. Stone, "Classification and Regresssion Trees", Wadsworth, Belmont, 1984.

[10] R. Quinlan, "C4.5: Programs for machine learning", Morgan Kaufmann Publishers, Inc., 1993.

[11] V. Cherkassky and F. M. Mulier, "Learning from Data: Concepts, Theory and Methods. 2nd ed.", John Wiley - IEEE Press, 2007.

[12] I. H. Witten, E. Frank and M. A. Hall, "Data Mining: Practical machine Learning Tools and Techniques, 3rdEd", Elsevier Inc., 2011.

[13] L. Breiman, "Random Forests", Machine learning, 45(1), pp. 5-32, 2001.

[14] [21] B. Kégl, "The return of AdaBoost.MH: multi-class Hamming trees", arXiv: 1312.6086, Dec. 20. Last time visited: Aug 15, 2016..

[15] N. Maček, B. Đorđević, V. Timčenko, M. Bojović, M. Milosavljević, "Improving Intrusion Detection with Adaptive Support Vector Machines", Elektronika ir elektrotechnika, Vol. 20, No. 7, pp. 57-60, 2014.

HOW TO GUARANTEE BABY IDENTITY BASED ON FINGERPRINT BIOMETRY

KOMLEN LALOVIĆ

Visoka škola strukovnih studija za Informacione tehnologije - Beograd, komlen.lalovic@its.edu.rs

SVETLANA ANĐELIĆ

Visoka škola strukovnih studija za Informacione tehnologije - Beograd, svetlana.andjelic@its.edu.rs

TOT IVAN

Univerzitet odbrane, Vojna akademija Beograd, ivan.tot@va.mod.gov.rs

Abstract: *In this work will present exclusive results of qualitative research - acquisition fingerprint minutiae of newborn baby. Also, exclusive part of qualitative research which will be presented that confirmed scientific fact for newborn to posses formed fingerprint and that it can be used for identity guarantee.*

It can be used to provide identity of each newborn baby and implement new system of identification based on biometry. This will eliminate possibility of human error or steal or replacement of baby identity.

Keywords:

Algorithm, Baby, Biometry, Birth, Fingerprint, Patent.

I INTRODUCTION

First let say something about Algorithm in different science filds and disciplines. As for a computer science algorithm, it is a step by step set of commands, instructions and operations that are going to be performed. The purpose of aglorithms is to provide calculation, the automatisation of actions and data processing.

An algorithm is a set of steps as explained. One has to learn how to make algorithms using pseudo-code or real code and that is why people who delelop algorithms need to have programming knowledge. If you want to optimize your algorithms, then you have to possess the knowledge of mathematics as well. In the end, you have to have some basic knowledge about all that, since it represents a combination of various knowledge. Good news is that learning about algorithms can be as simple as you want it to be, and as easy as you are able to aquire. [1]

If you want to optimize your algorithms, then you have to posses the knowledge of mathematics as well. In the end, you have to have some basic knowledge about all that, since it represents a combination of various knowledge.

In an attempt to solve a large human issue and remove bad shadow from possible past events in many countries, ex. stealing or replacing the identity of new born babies, also preventing that kind of fear that all future mothers have, and make that bright moment of bringing new life to this world easier and more relaxed to gynecologists, midwifes and nurses, this Patent device – Device for biometric identification of parenhtood – maternity and this algorithm as the part of its software have been developed.

Qualitative research that prove thesis that baby identity is possible to determine base on fingerprint minutiae and providing optimal scanner for that purpose is something that is exclusive, not just in region but in the world.

The invention, generally, placed in filed applied Information Technology, Biometry system. Device makes

a unique Identification (ID[1]) reference and that reference will be the Identifier for each "mother-baby relationship" for every newborn baby in maternities. [2] [3]

According to the International classification of Patents, this patent is classified with a symbol **G06F21/00** which belongs to the Biometry systems – devices for fingerprint scanners.

II TECHNICAL SOULUTION

Technical issue which needs to be set, examined and solved with this Algorithm of the Patent device consists of three partial task as follows:

- Writing one optimal algorithm for emulating and executing every function that device for biometric identification of maternity- parentooh possesses. In future, the realization will give recommendations for traversing that pseudo-code in accurate programing language, probably C programming language, since it has to be structural and low level, not OO and high level programming language as it is JAVA or C++ . It is very important that the algotithm realization is applicable for each existing platform including hardware and software, and C programming language is a proper choise for that. [4] [5]

Beside its common purpose and scanning two fingers of different persons at the same time it will provide a unique ID reference (similar Primary Key) which will be the base for every pair of a scanned mother-baby pair.

III EXPERIMENT

According to modern well known technical devices – fingerprint scanners which use different algorithms and methods in their process of work to determine the identity of individuals.

After having searched through the National base of Patents similar devices with this aim were not found, concretely dual biometric scanners, which contain their own lighting, battery supply and none of the Patents consider this idea and solution in this way, with dual biometric scanner.

It is a science fact in biometry that baby minutiae are formed until 7th month prenatal, but we have no research confirmed. That is something that lead authors throw research to provide this data and bring exclusive information in this field.

Figure 1 shows how scanning fingerprint minutiae of new born baby for 10 times with various types of scanners.

Attempt / scanner kinds	Optical	Capacitive	Pressure	Thermal
Finger 1	10	7	3	2
Finger 2	10	6	2	2
Finger 3	10	6	2	1
Finger 4	10	5	1	0
Finger 5	9	4	0	0
Percentage of success	98,00%	56,00%	18,00%	10,00%

Figure 1

For result we have no doubt that optical scanner is optimal for further researching and deciding which is going to be used for this purpose. For scanning all 5 fingers 10 times we have 49 successful acquisitions and 1 failure. With other type we have big difference in result, smaller accuracy is it clear.

Now, we decide to try optical with 500dpi and optical 1.000 dpi scanners to provide more accuracy, and we got expected result concerning baby fingerprint minutiae and small, needed to be scaled and they are not so tactical as adults fingerprints. [5]

Figure 2 present those results.

[1] **ID – IDENTITY** (unique data for each fingerprint scanning process)

Figure 2

IV DISCUSION

However, this device does not have two fields for simultaneous scanning the fingers of two different persons, which at the same time generates unique unchangeable ID reference and is an additional guarantee of a person's identity and guarantee the of Parenthood of baby – precisely the maternity of a new-born baby.

We have not listed scanners OEM because it is not important here, but we listed various types of scanners to show which is optimal for this purpose.

Science fact in Biometry, as a branch of Advanced security systems, Discipline - Informatics and Computing, Science Field - Natural Sciences and Mathematics, **is that fingerprint is formed during prenatal period for every fetus and stays constant in the shape of minutiae during whole life. We can not prove prenatal, but we proved that it is formed and can be acquired at very moment of birth. [6] [8]**

According to many researches realized on fingerprints of fetus, ultra waves and biometry scanning the minutiae on each finger are formed by the end of 7th month of pregnancy. It is important to mention that babies who are born before regular time of birth, during 8th, and especially **by the end of 7th month of pregnancy have fingerprint on each finger, both hands and foots fingers already formed. [9] [10] [11]**

We proved that science fact in practical terms. Here is short overview.

This is essential because minutiae – ridges and valleys are the only biometry that is formed prenatally and it can be used for the purpose of guaranteeing biometry identity. The whole idea for Patent Innovation is based on this scientific fact confirmed by both Biometry system as Computer science and gynecology – midwifery as a branch of HealthCare protection system. [7]

Other biometrics such as Iris recognition is unstable, because until 4th year the pigmentation in children's eye is changing and becoming different. The shape and color both change which makes it impossible to be used for this purpose and for this goal.

The head, hand and body shape and size are rapidly changing since they normally grow up so it is obvious why they cannot be used. That is why this incredible scientific fact that fetus's fingerprint is formed prenatally, by the end of 7th month in a uterus of a pregnant mother and stays constant with the same construction of minutiae, is so great that is amazing. [12]

V CONSLUSION

At moment of birth there are a large number of various fears during birth process, both of mother and of people in medical Care system in maternity. Reading and learning on study which was made in Australia and New Zealand from 2009 until 2011 and 17 workshops with over 700 midwives this device can prevent a part of one of those big fears – dealing with unknown (n=32). [7]

It prevents any possible theft or replacing the baby's identity, which has unfortunately being probably happened at some places and parts of the World, especially in South-East Europe, in the Balkans, countries of former Yugoslavia. We provide strong elements for further research in this field and prove that is possible guarantee new born baby identity based on fingerprint minutiae.

The inventor of the Patent has taken maternity symbolically because the maternal instinct is the strongest instinct in nature.

REFERENCES

Books:

[1] "Was al-Khwarizmi an applied algebraist?", Oaks, Jeffrey A. , University of Indianapolis. Retrieved 2008-05-30.

[2] Handbook of Biometrics, ANIL K. JAIN-*Michigan State University*, USA, PATRIC FLYNN-*University of Notre Dame, USA*, ARUN A. ROSS-*West Virginia University, USA* (2008), Sringer, USA

[3] MILOSAVLJEVIĆ, M., GRUBOR, G. (2007): *Osnovi bezbednosti i zaštite informacionih sistema,* Fakultet za poslovnu informatiku – University of Singidunum, Belgrade, Serbia

Articles from Conference Proceedings (published):

[4] What do midwives fear? Authors: Hannah Grace Dahlen, Shea Caplice, Published Online: July 24, 2014 – Elsevier, *Women and Birth, Journal of Australian College of Midwives*

[5] Biometric Verification of Maternity and Identity Switch Prevention in Maternity Wards Authors: Komlen Lalović, Nemanja Maček, Milan Milosavljević, Mladen Veinović, Igor Franc, Jelena Lalović, Ivan Tot DOI: 10.12700/APH.13.5.2016.5.4

[6] Before We Are Born, 9th Edition, Authors: Keith Moore, T.V.N. Peraud, Mark Torchia, Elsevier UK, Saunders, ISBN: 9780323313377, 2014

[7] NIST publishes compression guidance for fingerprint, Journal Elsevier - Biometric Technology Today, Volume 2014 Issue 4, April 2014, Pages 12

[8] Using Fingerprint Authentication to Reduce System Security: An Empirical Study, Security and Privacy (SP), 22-25 May 2011, Page 32 – 46, ISSN: 1081-6011, E-ISBN: 978-0-7695-4402-1, Conference Location: Berkeley, CA Publisher: IEEE.

[9] Biometric verification of a subject through eye movements, Martti Juhola, Youming Zhang, Jyrki Rasku, Computers in Biology and Medicine, Vol. 43, Issue 1, p42–50, Published in issue: January 01, 2013

[10] Komlen Lalović, **Doctoral thesis "New system of identification newborn babies and parenthood guarantee based on Biometry**", University of Singidunum, July 2016.

[11] Komlen Lalović, Milan Milosavljević, Nemanja Maček, Ivan Tot, "Device for biometric identification of Maternity", Serbial Journal of Electrical Engineering, Vol. 3, October 2015, ***DOI**: 10.2298/SJEE1503293L*.

[12] Nemanja Maček, Borislav Đorđević, Jelena Gavrilović, Komlen Lalović, *"**An Approach to Robust Biometric Key Generation System Design",** Acta Polytechnica Hungarica Vol.12, No.8, Year: 2015* ***DOI**: 10.12700/APH.12.8.2015.8.3, Im. F. 0.65*

THE FUTURE OF PAYMENT CARDS AND NEW TECHNOLOGY – RISKS AND ACHIEVEMENTS

Prof. dr ALEKSANDAR ČUDAN
Kriminalističko-policijska akademija Beograd, aleksandar.cudan@kpa.edu.rs

Prof. dr ZVONIMIR IVANOVIĆ
Kriminalističko-policijska akademija Beograd, zvonimir.ivanovic@kpa.edu.rs

Abstract: *Payments make integral part of business operations regardless of whether we are talking about traditional or on-line business operations, and this is why it is necessary to find the best manner to pay for goods or services using new technological and communications solutions. Mobile banking is no longer a privilege of banks, card brands or mobile operators. Despite ever growing spread of various forms of electronic money, payment cards remain quite a frequent method of payment in which additional functions are integrated, but they still remain a specific institute of cashless payment issued by a bank or a non-banking organization.*
Pressures on managers at various organizational levels to develop new and profitable sources of money and manners of payment have created favourable climate within banking system for actors of financial destruction. Developments of new technologies and financial markets have resulted in appearance of new risks and at the same time have increased the importance of traditional risks to which financial and non-financial institutions are exposed. In order to reduce the risk of abuse of payment cards and ensure positive effects and achievements of card business in financial sector, it is necessary to focus on constant communication, one's own learning but also the training of clients, constant analysis of payment card market, supervision of well-known institutes in order to understand and master new technologies, to reach new technological advantages, to make good prognosis, to offer new concepts, as well as to participate in a constructive manner in designing law and by-laws in this field.[1]

Keywords: *payment cards, financial market, risk, new technologies, payment.*

1. INTRODUCTION

At the beginning of the new millennium technological development has reshaped financial and banking industry which is increasingly taking a leading position in using new technologies. Payments are integral part of business operations regardless of whether we are talking about traditional or on-line operations, and money as a value can be found in the form of information within a banking information system. Definitions referring to remote payments suggest that a boundary between e-payment and m-payment is not differentiated, ant it will be even harder to define in the future decades. Due to the mentioned reasons the common term "electronic payment" at the moment represents a logical choice. Despite numerous statistical records accompanying e-payments and m-payments, the definitions of these two concepts are still rather connected. There exist various opinions by experts in this field who wonder if it is necessary at all to differentiate between them since the demarcation line is rather thin.
Expansion of electronic banking and payment industry is closely and inextricably linked with the development of payment cards as a means enabling activation of mechanisms for large-scale use of banking services. In the contemporary conditions the payment cards still represent the most represented mode of electronic payment of nowadays. In practice they mostly represent a compatible link between instruments of identification and instruments of payment operations. Expansion of electronic banking and payment industry is closely connected with the development of electronic cards as a means enabling activation of terminals for large-scale use of banking services. Electronic payments using payment cards differ substantially according to the degree of safety, functionality and relationship to on-line use, and all this depends on the functional application and technological foundation they are based on, which is the most significant thing in the contemporary conditions.
The main characteristic of trends in contemporary world, including the financial world as well as the payment industry, is development of new although not always better forms of business operations.

2. DEVELOPMENT AND SIGNIFICANCE OF INNOVATIONS IN THE CONTEMPORARY PAYMENT SYSTEM

With the expansion of global financial market worldwide, as well as with the development of increasingly complex financial products, the clients have started facing numerous novelties in the financial market. At the beginning of the new millennium it seems that there is not any other industry which has developed and transformed so intensively as global payment industry. In the contemporary environment the payment system means a set of technological and social links which enable exchange of values. The set of technological links enables the exchange of values to be

achieved, which enhances the efficiency of payment system, while the set of social connections defines the participants' roles in the system.

Contemporary payment systems change so that from an industry which was once characterized and defined by closed standards and limited access it becomes the industry in which openness is the main driving force of success. Payments have three dimensions which are always present: technology, business models and trust. Without all three dimensions there are no quick or safe payments. Technology is foundation for all kinds of payments, but it is particularly important for electronic commerce in which payment cards take central place. The credit for the development of card industry goes to the application of new technologies and innovative business operations. Thanks to a large number of various products and services, there has been a particularly dynamic development of this part of banking operations. In the contemporary banking system payment cards have completely replaced the manner of payment for goods and services, and they continue to do so successfully today.

Innovations are currently one of the most relevant topics of contemporary business operations. In a global environment where technology is constantly changing, innovations represent the main driving force of economic growth for many companies and financial institutions. Electronic payment industry enters a period of unforeseen innovations, and technologies are changing extremely fast. Many new solutions are offered as it has never been the case in the history of electronic payments so far. The contemporary literature in this field underlines the fact that there are more than 300 ways and models of electronic payment. However, it cannot be expected that each innovative idea would turn into a profitable and usable financial product. This is why relevance and profitability in the world of changes become the only measure of success. Other leaders in this industry, as well as non-financial institutions, follow trends and invest a lot into all parts of organization in order to build a framework for development of innovations. Since they are surrounded by extremely dynamic changes, the employed in the electronic payment industry are expected to respond to the needs of their users in an appropriate manner and to create added value for their companies and clients. The experts practicing in the industry of electronic payment must follow and understand development trends, must acquire new skills fast and adapt their personal and professional development plans to new requirements. Financial institutions, as well as their employees, must first well understand the role and significance of innovations for their business operations and then develop processes and methods which will help them achieve goals and be constantly innovative.

In more than half a century electronic payments have had dynamic development and as of recently they are more and more connected with e-commerce and are one of its main boosters. Intensity and form used in the developed economies through e-commerce are direct result of the existence of a wide range of possibilities of electronic payment. Global traditional card brands have held absolute primacy for several decades in the electronic payment industry, which has been changing recently through participation of new companies from several economic sectors. Thanks to innovations new providers of payment services surely become competition to banking sector as well.

The market of electronic payment using payment cards globally offers numerous possibilities for innovations. In the course of the second decade of the new millennium the consumers have changed considerably their habits regarding this type of payment. In addition to a developing trend of payment by credit and debit cards, the increase of e-commerce and growing popularity of smart phones have enabled new remote payment methods using short-range wireless technology NFC (Near Field Communication).

At the end 2016 a chain of British supermarkets Waitrose opened the first cashless store and the customers can choose the payment method (by card or through their mobile devices) at one of self-service checkouts in the store. Waitrose supermarket has made efforts to provide their customers with a unique user experience. For banking institutions this represents a signal that they also must be prepared to respond to changes and must adapt their payment systems to such requirements in order to keep their share of income and in order not to let their place be taken by new companies in the payment industry.

A radical step in the global card market in the recent years has been the implementation of contactless terminals and contactless payment cards. These cards contain antenna which detects and accepts signals sent from the reader so that physical contact between a card and a terminal is not required. Such systems are open for the implementation of artificial technology. It is obvious that as a part of these efforts the integration of artificial intelligence will be dictated by the trends from economically the most developed countries. The application of artificial intelligence in this field will represent a great saving for the companies, since it will enable reduction of expenses for user services and messages will be transferred to clients in an acceptable shape and format.

The practice shows that even small companies such as Irish Voysis develop voice solutions at neutral platforms developed by other companies for all other technological companies that cannot develop their own voice platforms, and according to the quotation of the mentioned company "*voice will soon* be the *first* point of *contact between* '*man'* and *machine*." The giants such as Google, Apple, Union Pay, Amazon, Microsoft and Facebook also invest enormous amounts of money in research in the field and its application. Personal assistants will be enabled with a voice command, which will further enable creation of new markets for shopping, payment by technologies based on the concepts of virtual and expanded reality. Their prognosis is that devices will replace smart phones and that the future concept with built-in electronic software, sensors and connectivity will enable objects to exchange data with a manufacturer, operator and other connected devices, which will be added a payment function as well. Such technological innovations would enable, for instance, a refrigerator to recognize what kind of food it is lacking and at a certain moment of time to order them, to analyse the favourability of purchase (promotions, credit ability, price competitiveness), and finally make the payment. A smart car which is connected to the Internet makes it possible for a user to order goods and services while driving and to pay

for them immediately while driving and after that to take them over.
Current card with mobile application represents a controlled Visa debit card which enables parents to give a debit card to their children while being able to control the amount of money spent by children as well as for what purpose. In this way parents have large control when the card will be cancelled and what the money is spent for with the possibility to block an unwanted transaction which comes from a certain group of sellers such as bars or hotels. Such a system of payment leaves possibility to parents to give a certain prize to their children for doing certain chores, while the children can donate assets if they want to. These are only some innovations in the contemporary system of payment which confirm that it is no more a matter of time when new technologies will be applied in banking, they are already here. In the contemporary payment industry scientific and technological achievements are no longer a missing link, in the new millennium these are ideas.

3. DEVELOPMENT TRENDS IN THE PAYMENT CARD MARKET IN THE REPUBLIC OF SERBIA

Transformation and transition from sliding a payment card, through holding phones or personal assistants close in order to make a payment is natural and it shows the users the direction in which the innovations will go. The banking infrastructure in Serbia today understands well this concept and the attitude is that institutions are ready to develop products which will be used better by the digital environment but which will also be adapted to mobile needs of their users. Still the digitalization of banking services in our country cannot happen overnight, since it is an ongoing process which requires time. [2]
The analysis of the state-of-affairs in the market of payment cards in the Republic of Serbia contains the data on the current state in the market and spotted trends in the number of cards, accepting network for payment cards, as well as the number and value of transactions. The information also includes some current issues important for further development of cards as cashless payment. Statistical methodology according to which the National Bank of Serbia publishes information on payment cards in the course of 2016 is substantially different from previous years, since it has been adapted to the statistics of payment services based on the Law on Payment Services. Taking into account the above said, it is clear that some data which were normally included into annual analyses of payment card market are not fully comparable with previous years.

In the circles of domestic scientific public dealing with these topics the debates can be heard on the degree of development of electronic commerce in the Republic of Serbia and if it represents a developing sector. The subject debates are led among traders, on line traders, financial institutions, banks, state institutions and various professional associations within the Chamber of Commerce which deal with card operations.
In the payment card market in Serbia in 2016 the trend continues of the growth of number of payment cards issued. The total number of cards issued in 2016 is 6.900.997, which is 9.41% more than in 2015, when 6.454.356 were issued. If we take into account the cards with e-money function (pre-paid cards for Internet payment), out of which 64.357 were issued, the total number of cards issued was 6.965.354, and so the growth is 10.43%. The total number of active cards in 2016 was 3.931.496, or 56.97% of the total number of cards handed over/issued. [3] The active card is a card which has made at least one transaction in the reporting period.

Table 1. Number of payment cards issued in the Republic of Serbia

Year	Number of payment cards issued in the Republic of Serbia	Number of active payment cards in the Republic of Serbia	%
2001.	2		
2002.	400.000		
2003.	500.000		
2004.	2.100.000		
2005.	3.800.000		
2006.	5.240.000		
2007.	5.725.465		
2008.	5.728.789		
2009.	6.014.390		
2010.	6.150.000	2.944.458	48
2011.	6.350.587	3.073.282	48
2012.	5.934.784	2.737.873	46
2013.	6.207.833	2.922.597	47
2014.	6.267.058	3.073.646	49
2015.	6.454.356	3.262.106	51
2016.	6.900.997	3.931.496	57

Source: National Bank of Serbia, Sector for monetary system and policy based on bank reports

Taking into account today's level of development of payment card industry in Serbia it can be concluded that the current infrastructural and technological requirements of the market are met to the extent necessary. Despite this fact, the data of many a research suggest that there is still space for their further improvement and development but complete disappearance of paper money cannot be expected, although such prognoses have been present for decades. The similar happens with payment cards which thanks to technological innovations in the system of e-payment still have the bright future. [4]
Education of users on new technology is probably the greatest obstacle to further digitalization of the financial sector in the Republic of Serbia. Card brands and financial institutions cannot only launch a new technology hoping that it will be accepted – it is necessary to invest time and resources in order to fully understand the benefits of such solutions but also the fact that digitalization of business operations implies not only simplicity of payment but also a completely new degree of protection.
The network for acceptance of payment cards in Serbia in comparison with the countries in the region is still well developed. In 2015, after several years the first increase of the number of ATMs was recorded, which is significant

considering that this number was constantly decreasing in the period up to 2014. This rising trend continued in 2016, so that at the end of the year there were 2.845 ATMs. This is 140 ATMs more in comparison with the end of 2015.

At the beginning of 2017, the increase of the number of POS terminals is far greater than in previous years and was 26.32%. In previous years the banks submitted to the NBS the number of selling locations as the number of POS terminals, while in 2016 the real number of POS terminals was shown (with the total number of POS terminals per a selling location, considering that one selling location can have several POS terminals).

Table 2. Number of ATMs and POS terminals at the territory of the Republic of Serbia

Year	Number of ATMs	Scope of transactions in billions	Number of POS terminals	Scope of transactions in billions
2004.	450		16.266	
2005.	837		31.816	
2006.	1348	109,0	48.194	65,0
2007.	2074	165,5	55.340	91,3
2008.	2494	224,5	57.919	117
2009.	2723	263,8	59.058	124,8
2010.	2857	312,6	57.459	141,6
2011.	2830	343,3	58.012	161,5
2012.	2785	395,9	62.656	194,3
2013.	2673	448,6	59.822	187,2
2014.	2632	489,1	64.142	203,8
2015.	2705	535,4	65.428	233,3
2016.	2845		82.647	

Source: Ministry of Finance of the Republic of Serbia

Despite the fact that e-commerce in Serbia is underdeveloped, the positive trends are evident. In comparison with the previous year there is an increase of the number and value of transactions in purchase of goods and services at the Internet. Payment transactions related to the purchase of goods and services at the Internet are made using cards and e-money. The total number of transactions of payment of goods and services at the Internet in 2016 was 3.852.840, out of which 3.639.500 were transactions using payment cards and 213.340 were e-money transactions.

The government of the Republic of Serbia started a debate on introduction of cashless payments of social support in order to prevent the population using support from misplaced spending of money. Unofficial research suggests that budget assets in almost 40% cases are spent as misplaced on cigarettes and alcohol. In order to eliminate possible misuses of social support the state is considering introduction of specific-purpose payment cards. This means that the families from the list of the endangered ones, who receive any kind of support, will no longer receive the assets in cash but only through cards at the places where they could buy only the things necessary to their families. Certain officials of the relevant ministry support introducing card system for child support and other one-off payments in order to control that the budget funds given to the users are spent for the exact purpose they are assigned.

4. SECURITY RISKS OF CONTACTLESS ELETRONIC PAYMENT

Security represents an integral part of innovation strategy. It is desirable that security is provided at the same rate as innovations and this should be the field where card companies would apply the concept and principle of "responsible innovation".

The risk represents a potential problem or a potential opportunity. In both cases it appears in all spheres, so this is why it is necessary to analyse and find the right ways of risk management. In literature a great number of definitions of risk can be found. Linguistic term of risk comes from the Italian word *rishio* and originally it meant danger to ships from cliffs and rocks. Risk assessment is the concept of various content which is why it requires, in addition to legislative foundation, the application of many criteria in order to be implemented more efficiently in practice. [5] Starting from the aspect of economic safety, according to some definitions risk is a possibility of loss or injury, in other words a possibility that an undesired consequence of an event or process will happen, in this case from those involved in criminal activities in the field of economic crimes. The right selection of risk assessment method enables adequate application of measures, which would assess adequate measures and activities on suppression and prevention of infiltration of those involved in criminal activities with the elements and characteristics of economic crime. It is a prevailing opinion that Serbia is still in the stage of development of this market and that there are numerous obstacles and challenges to be overcome. The main driving force of the development of electronic commerce is the payment process. Payment methods have considerably progressed in a few recent years, which is shown by the statistic analysis of payment card market in the Republic of Serbia. Numerous factors influence the changes in payment industry. There is not an ultimate goal, new challenges and risks keep repeating, which at the same time can present new opportunities for criminal actors. This is why receiving and issuing banks must direct all their activities to fast adaptation to changes in the market, improvement of quality of services, customer service and their security. Banking services are becoming personalized, while taking care of the customers, their safety and security are the most important.

Abuse of payment cards is an illegal act, whereas the manners of execution appear in various forms. Express evolution of technology, particularly in the field of information technologies and communications, despite the existing security measures and constant innovations by banks and institutions in charge of security measures and risk reduction, offers new opportunities for payment card abuse. Technological development provides payment card fraudsters with constant inflow of data which are more easily but first of all faster implemented into their fraud activities as opposed to banking systems in which additional innovations of equipment but also the staff education require a longer period of time and have its price. Development dynamics of certain technologies dictates the direction of innovation of criminal offenders and thus

various payment card frauds are becoming more complex in certain fields and gain significance due to inadequate and undue response of institutions in charge of security of electronic payment. Unlimited development of forged technologies with the accent on dynamics of progress makes it conditional upon us to suggest that it is necessary to clarify the existing, i.e. known manners of payment card abuses and to deal with them. Abuses can be observed from the aspect of appropriation of payment cards and the data from payment cards by fraudsters. [6]

At the beginning of the new millennium many traders and consumers provide payment services by independent channels, and there is an ongoing debate on how important it is to accept payment cards for banking and electronic payments. The opinion prevails and is adopted that banks are not the only institutions which offer payment services. The systems of electronic payments allow to financial institutions, companies and state bodies to offer to their clients a range of payment options. New kinds of payments through alternative payment networks represent a risk, including social networks such as Facebook, which can also offer a payment service to their clients.

Electronic banking in Serbia is mostly used by younger and technically well-educated individuals, who are not easy to deceive using common methods. In the years to come the scope and value of mobile-generated payment orders using various payment cards would multiply several times, so that an increased number of attacks and abuses within this manner of payment is also expected, which brings a new kind of security risks. The profile of individuals who commit abuses is also changing in that this form of economic crime today is mostly committed by younger and technologically educated persons.

At the global level the abuses are committed from a remote place, far from the subject of attack and the place of the damage. The national and continental borders do not represent limit in the developed global trade, or in the field of criminal activities, bearing with them new risks. The exact costs caused by the abuse of payment cards are hard to assess and categorize since the data on these costs are not usually reliable enough. High costs of preventing frauds and adhering to regulatory and safe network standards are approximations of the assessments of true losses incurred due to frauds. Indirect costs of payment because of abuses include the costs of local and national laws, barriers for on-line commerce and its benefit.

Trends of migration of card abuses have been present for several years and the crime related to the abuse of payment cards is transferred from Europe and other developed countries into other regions where contemporary technologies are not present or used on a large-scale.

The use of cards on the internet means the possibility that all the details from a card holder remain on the site: name and family name, card number, validity date. The news that thousands, and sometimes even millions of payment card numbers are compromised and taken in an unauthorized way from data bases where they are kept can often be seen in the media headlines.

CONCLUSION

New technology and digital revolution transform trade, including the payment industry. The manners of payment evolve fast, from digital currency, plastic payment cards to mobile payment applications. Payment ecosystem expands and the difference between traditional and new companies is decreasing. Global trends clearly suggest that the image of electronic payment industry will look quite differently in the next ten years. Due to the nature of business operations, connection with global payment schemes and presence of large banks and processors in various markets, new technologies of payment are not limited geographically and can fast be applied in all markets together with old but also with the new emerging risks. The leading banks, processing and technological companies are already investing huge assets into development of payment technology as an investment for the future and survival in the tough market. [7]

Payment industry must also make a choice. Banks, but also the existing and the future card brands must be able to keep pace with their requirements and thus maintain preferential status when choosing the manner of payment.

It can be concluded that the Republic of Serbia is on a good way to use all advantages of electronic manners of payment. The use of new technologies by customers is the reason why global trends of electronic payment expand rather fast in financial practice. The development of electronic payment industry is beneficial to everyone, both the customers and the state. The development of payment methods is one of the main postulates of the EU initiative towards creation of a unique digital market, a requirement which we will have to meet as a part of the process of accession to European integration area. Without a desire for this to be a scientific paper bearing a stamp of the old belated time, the basic motive is to translate the problems discussed to contemporary flows with expressed note of interdisciplinary. The challenges and risks are numerous and they will always be present.

REFERENCES

[1] The paper is the result of research on the project titled "Crime in Serbia and the Instruments of State Response", which is financed and carried out by the Academy of Criminalistic and Police Studies, research cycle 2015-2019.

[2] Turban, R., Uvod u informacione sisteme, Beograd, 2009, str. 182.

[3] Privredna komora Srbije Aktuelno stanje na tržištu platnih kartica, Beograd, 2017, str. 2.

[4] Mikarić, B., Trajković, D., E-business influence upon improvment of banking services, Trendovi u poslovanju 2/13, Beograd, str. 70.

[5] Cindori, S., *Pranje novca: korelacija procene rizika i sumnjivih transakcija,* Odabrani prevod broj 16/2013, Institu za javne finansije, Zagreb, 2013, str. 1.

[6] Pejčić, D., *Zloupotreba platnih kartica kao specifičan oblik kriminala*, Univerzitet u Novom Sadu, Ekonomski fakultet Subotica, magistarski rad, Subotica, 2010, str. 41.

[7] Kričković Lj, Invencijsko-inovativni procesi u kompanji Visa INC. na primeru elektronskog plaćanja, Maribor, 2016, str. 28.

CIP - Каталогизација у публикацији -
Народна библиотека Србије, Београд

007:004.056(082)

INTERNATIONAL Conference on Business Information Security BISEC (9 ; 2017 ; Beograd)
Proceedings / The Ninth International Conference on Business Information Security BISEC, Belgrade, 18th October 2017. ; [editor Igor Franc, Bojana Trebinjac, Sanja Kovačević]. - Belgrade : Belgrade Metropolitan University, 2017 (Beograd : Copy planet). - ilustr. - 95 str. ; 30 cm

Tiraž 60. - Napomene i bibliografske reference uz tekst. - Bibliografija uz svaki rad.

ISBN 978-86-89755-14-5

a) Информациона технологија - Безбедност - Зборници b) Информације - Заштита - Зборници
COBISS.SR-ID 248773900