BELGRADE
METROPOLITAN
UNIVERSITY

12TH INTERNATIONAL CONFERENCE

BISEC

BUSINESS INFORMATION SECURITY CONFERENCE

Belgrade Metropolitan University
Belgrade, 3rd December 2021.
www.**metropolitan**.ac.rs

PROCEEDINGS

The Twelfth International Conference on Business Information Security

BISEC
BUSINESS INFORMATION
SECURITY CONFERENCE

Belgrade Metropolitan University

Belgrade, 3rd December 2021.

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 Nemanja Zdravković
MSc Bojana Domazet

Chair of Organizing Committee

Doc. Dr Nemanja Zdravković

Conference Secretariat:

MSc Bojana Domazet

Design

Petar Cvetković
Mladen Radić

Printing

Copy Centar Kruševac

Circulation

60

Belgrade, 2021.

CONTENT

Organizer

BELGRADE
METROPOLITAN
UNIVERSITY

ESIGURNOST
INFORMATION
SECURITY

PERSONAL DATA PROTECTION IN SERBIAN HOTELS

ALEKSANDRA BRADIĆ-MARTINOVIĆ

Institute of Economic Sciences, abmartinovic@ien.bg.ac.rs

Abstract: *The personal data that guests make available to the hotels they are staying in is becoming more and more exposed to breaches. This has been influenced by at least two factors in the last twenty years - the exceptional global development of tourist travel and new technological solutions involving private data collection. This paper aims to analyse empirical data collected through a web-based survey and in-depth interviews with top managers to determine whether hotel management in Serbia respects its guests' data privacy by following legal regulations - the Law on Personal Data Protection of the Republic of Serbia and EU's General Data Protection Regulation (GDPR). We can conclude that management in the Serbian hotel industry provides compliance with the law's provisions regarding personal data protection. Even so, insight into specific procedures or inspection reports would give us more precise results to avoid biased or unqualified responses. Proper education is the best way to overcome unintentional law violations, while intentional illegal actions should be sanctioned. We also believe that it is essential for domestic hoteliers to be familiar with the provisions of the application of the GDPR in their objects.*

Keywords: *Personal Data, Hotels, Tourism, Legal framework, GDPR, Law on Personal Data Protection, Serbia*

1. INTRODUCTION

The process of rapid digitalisation is distinctive in all areas of business, regardless of a particular sector. The number of systems that collect, process and store personal data has been increased recently. Also, in the last two decades, a significant volume of data has become online. Additionally, the Covid-19 disease pandemic has opened up space for the implementation of digitised solutions to avoid unnecessary social contacts, further growing the volume of personal data collected. As a consequence of this trend, personal data security and protection of personal privacy are raised, given numerous cases of data breaches. We can declare two recent examples. British Airways was fined over 26 mil USD for the significant data breach that potentially affected 400,000 customers [1]. At the same time, Marriott Hotels chain has to pay 25.3 mil USD for, influencing 399 million guests globally. As a result of the hacker attack, due to a lack of proper safeguards, the clients' contact details were compromised, including their names, email addresses, telephone numbers, information about arrival and departure, VIP status, and numbers of loyalty programs.[2] The breaches were triggered by faults such as inadequate monitoring of privileged accounts, insufficient supervising of databases, lack or unsatisfactory management on critical systems and absence of encryption.

Based on only the two previously mentioned cases, we can gain insight into the extent of possible damage when it comes to embezzling customer trust. This research focuses on particular businesses – hotels, the systems that provide accommodation within the tourism sector to a considerable number of visitors worldwide. Their importance in the business environment is reflected in the fact that the global hotel and resort market in 2019 was worth 1.2 trillion USD[5]. The latest trends in the hotel industry introduce even more personal data through the growth of contactless payments, voice search and voice control, enhanced personalisation, recognition technology, and similar.

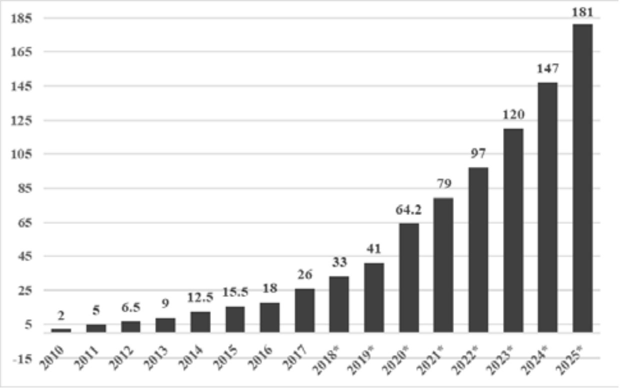
The subject of this paper is the hotel industry in Serbia and its procedures regarding personal data, in line with legislation. This paper aims to analyse empirical data to determine whether hotel management in Serbia respects its guests' data privacy by following legal regulations.

2. DATA BREACHES AND LEGAL RESPONSE

Data breaches: Facts and Figures

Today, data is the "fuel" that provides digital transformation and innovation. It permeates all areas of human activity - the state, economy, science, society. A tremendous amount of data is collected, processed, and shared in just one day on different bases. Data become ubiquitous, massively collected online, and enable much better connectivity.

Image 1 presents data volume created and consumed worldwide from 2010 to 2017 with a forecast from 2018 to 2025. The volume increased by 35% (CAGR).



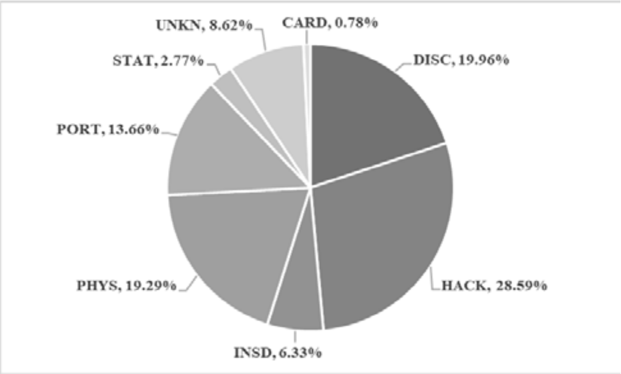
* forecasted values

Image 1: Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2025(in zettabytes) [6]

On the other side, negative phenomena are also present, and the most devastating are breaches, and with the rapid expansion of data volume, including personal data, they are also on the rise.

Privacy Rights Clearinghouse (PRS) define a data breach as “a security violation in which sensitive, protected or confidential data is copied, transmitted, viewed, stolen or used by an unauthorised individual”. According to their methodology, we can distinguish “hacking theft of credit/debit card numbers, lost, discarded or stolen documents/devices, and mishandled sensitive information”. Additionally, MDA divides data breaches into several categories: “confidentiality breach, availability breach and integrity breach” [8]. Garrison & Ncube [9] use breach division as follows: “stolen, hacker, insider, exposed and missing”.

The distribution of different types of breaches in the United States, Asia and Europe is presented in Image 2.



INSD - an insider that intentionally breaches information; CARD - payment card fraud; PHYS - physical loss; PORT - lost or stolen portable device; HACK - being hacked by someone or infected by malware; STAT – stationary equipment loss; UNKN - an unknown method; DISC - an unintended disclosure like sending an email to the wrong person.

Image 2: Breaches type distribution [7]

Regardless of the type of breaches, the consequences of each of them can be harmful to the persons whose data are included, and the damage can be enormous. The same source [7] gives an overview of the largest breaches, according to record numbers/size, as presented in Image 3.

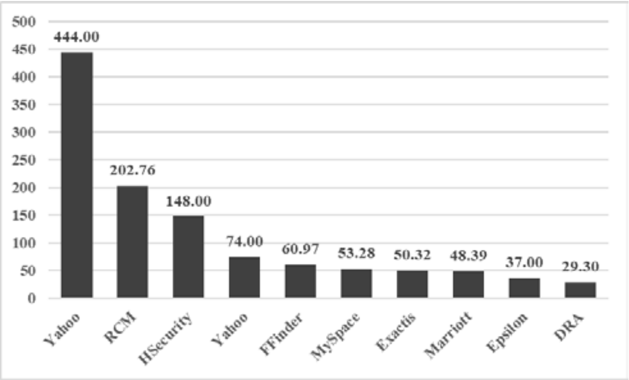


Image 3: Breaches by size (in terms of record numbers)

The three main ways these breaches can threaten companies are financial loss, legal misconduct, and reputational destruction. The consequences can cause enormous damage, so companies must protect themselves in this area.

Data privacy risks in the hotel sector

PwC reported few major points when it comes to data privacy in the hotel sector:[10]

- It is of great importance to make an effort to raise awareness on this topic, as the protection of virtual systems and cyber risks has become an alarming issue. They also add that research, but also practice, shows that security and privacy are still not in focus.
- Existing experience shows that the weakest security point for hotels is data manipulation regarding payment cards. Numerous court cases in the US testify to this statement, which necessitated the development of new, improved protection systems, revision of existing ones and detailed investigation in the case of breaches. On the other hand, guests need to be aware of the potential risks, as well.
- Hotels should think in advance about building and maintaining the trust, confidence, and position of their brand, without corrective measures that must be implemented based on court decisions. Leaving the framework of good business practice leads to long-term capital value erosion, especially if the company is listed on the capital stock exchange. Therefore, hotels are exposed to both legal and reputational risks.
- When it comes to the EU, the introduction of the GDPR has significantly brought hotels into line with the legal framework and avoid unnecessary litigations.

- "There is much more to security and privacy than compliance and risk." [10]. Hotels have the opportunity to use the available data to provide better and more personalised services. The best way to reconcile potential risks and opportunities is to create an appropriate vision. The next step would be to develop a strategy to achieve the set vision. Persons in charge of security and privacy must be involved in the strategic planning process to gain insight into the desired situation and find a way to reach it.

Legal framework

Most countries have faced this problem and accordingly adopted laws on personal data protection. UNCTAD keeps track of Data Protection and Privacy Legislation Worldwide. According to their evidence, 66% of countries have legislation, 10% have draft legislation, 19% are without legislation (Libya, Egypt, Sudan, Ethiopia, Cameroon, Afghanistan, Bangladesh, Sri Lanka, Thailand and a few more), and 5% countries have no data. These laws' most essential and primary benchmark is "use and sharing of personal information to third parties without notice or consent of consumers" [3].

Starting from the fact that Serbia is on the path of integration with the EU and regional tourist gravity towards this market, we believe that the analysis of legislation should start from the GDPR - General Data Protection Regulation (2016/679). The GDPR is applicable as of May 25 2018, in all member states to harmonise data privacy laws across Europe. Additionally, the goal is to raise the user's trust in business entities - service provides and enables a more effortless flow of personal data between EU member states.

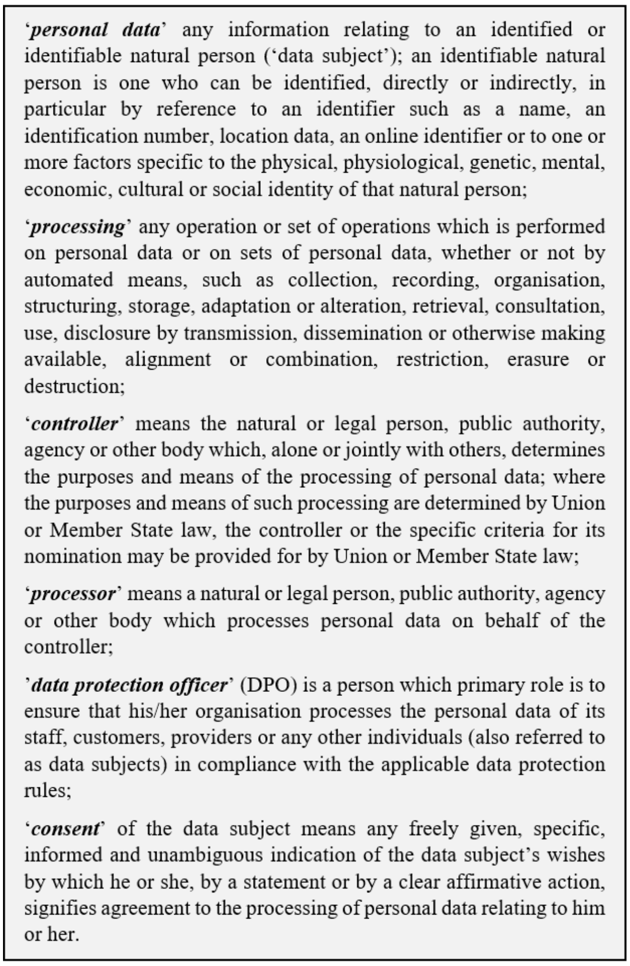
GDPR has returned power to customers and forced business systems to be transparent regarding collecting, storing, and sharing personal data and related information about their guests and employees. When it comes to hotels, the legal basis for collecting private data is a reservation, which is in line with the law. However, there are several clauses related to the way data is manipulated. Hotels have to:

- include top management in the process of personal data protection, at least three positions, General Manager, Head of Marketing and Revenue Manager,

- document what kind of data they hold, how the data is collected and processed and take action to assure compliance with their rights: to be informed, to access or modify data, to give or withdraw consent, to erase data and to transfer data,

- pay attention or avoid collecting sensitive data (biometrics, racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership and health status – which gained high importance during the Covid-19 pandemic),

- be aware that all software they use need to be in line with GDPR,



'***personal data***' any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person;

'***processing***' any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction;

'***controller***' means the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data; where the purposes and means of such processing are determined by Union or Member State law, the controller or the specific criteria for its nomination may be provided for by Union or Member State law;

'***processor***' means a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller;

'***data protection officer***' (DPO) is a person which primary role is to ensure that his/her organisation processes the personal data of its staff, customers, providers or any other individuals (also referred to as data subjects) in compliance with the applicable data protection rules;

'***consent***' of the data subject means any freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her.

Table 1: GDPR - Glossary of the most important terms[11]

- to have Data Processing Agreement in place with each vendor/provider; Hotels can use vendor services outside the EU, but the transfer must comply with the provisions of the GDPR. They must clearly define the data that the vendor will process and for what purpose. It is also obligatory for them to state the vendor's name in the Privacy Policy, saying which data they use and for what purposes,

- ask for the consent of private persons whose data they handle if they want to use it for any purpose other than staying in a hotel based on reservations, e.g. for marketing and promotion purposes and loyalty programs, research, and profiling,

- to allow personal data to be removed at the request of the data subject (which is why they need to have mapped locations of all data they collect about persons);

- to appoint a Data Protection Officer or to seek guidance from a legal advisor, in case of small business systems,
- be aware that the GDPR does not apply to data made available by EU citizens to hotels outside the EU while staying in those hotels. Still, if the hotel uses an online booking system and the citizen fills in the form while in the EU, the hotel must comply with the provisions of the GDPR.

Based on the above-mentioned, DPP-GDPR[12] offers recommendations in case of a data breach. As a first step, they advise management to inform the Information Commissioner's Office about the data breach in detail, then to assess the damage and try to reduce it while at the same time releasing a public statement. Further, to perform a detailed investigation and to use results as feedback for procedures and policies improvement.

In November 2018, Serbia adopted the Law on Personal Data Protection ("Official Gazette of RS" No. 97/08, 104/09 - other law, 68/12 – decision of the CC and 107/12), which came into force in August 2019. The law appliance is ensured through bylaws such as Rulebook on the Manner of Prior Review of Personal Data Processing (Official Gazette of the Republic of Serbia No. 35/2009); Decree on the Form for and Manner of Keeping Records of Personal Data Processing (Official Gazette of the Republic of Serbia No. 50/2009); Rulebook on the Form and Manner of Keeping Record of the Data Protection Officer ('DPO') (Official Gazette of the Republic of Serbia, No. 40/2019); Decision on the List of Countries, Parts of Their Territories or One or More Sectors of Certain Activities in Those Countries and International Organisations where it is Considered That an Adequate Level of Protection of Personal Data is Ensured (Official Gazette of the Republic of Serbia, No. 55/2019), and more.[4]

After two years of application, at the Tribune "How secure is the personal data in your company" held in June 2021, it was concluded that: "The Law on Personal Data Protection is not clear enough to those for whom it is intended, as well as to those who handle data." [13] Insufficient understanding creates an ample space for errors and omissions in the data handling and processing, which endangers both private individuals and businesses to which the Law applies. On the example of the hotel sector in Serbia, we will explore the practice in this area.

Serbian National Strategy for Personal Data Protection in Serbia has been drafted to approach this topic thoroughly.

3. EMPIRICAL EVIDENCE FOR SERBIAN HOTELS

The tourism sector in Serbia has seen a sharp increase in accommodation capacity in recent years. Based on data provided by the Statistical Office of the Republic of Serbia, in the period 2014-2020 number of hotels (including garni hotels) raised at a rate of 5.5% (CAGR), as presented in Image 3.

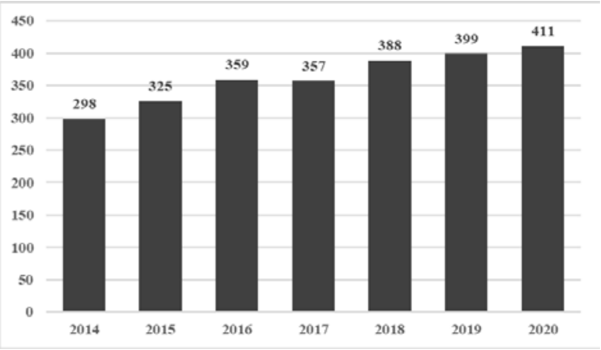


Image 3: Hotel accommodation in Serbia

Each of these hotels is a business system that needs to apply the Law on Personal Data Protection. To discover their attitudes toward this subject, we did a short web-based survey. We sent 150 invitations (including all categories, locations and sizes of hotels) by email and collected responses in the first half of August 2021. Finally, 38 correct and complete questionnaires were collected. The structure of our sample is as presented in Table 1.

Hotel category	Share
3-star hotel	42%
4-star hotel	48%
5-star hotel	10%
Hotel location	Share
City hotel	37%
Destination hotel	63%
Size of the hotel (number of rooms)	Share
Less then 50	26%
51-100	58%
101-300	16%

Table 2: Sample structure

Based on the sample structure, it is evident that lower category hotels were not included in this survey due to a lack of response. Hotels categorised as 1-stared and 2-stared hotels are included in the total hotel offer with 19% (1-stared - 4% and 2-stared - 15%).

Introduction of new acts and procedures - We assume that in the last two years, hotels have had to change practices regarding personal data in order to harmonise their existing processes with the new law. To determine what personal data they collect, how, from whom, in what format, and for what purpose, determine how long the data is kept, what protection measures are applied, with whom this data can be shared, whether data are taken out of Serbia. We can point a few not so obvious examples. Suppose a company uses video surveillance to ensure the safety of property and persons. In that case, it is crucial to introduce appropriate legal acts that will provide for how long the data collected by this video surveillance is kept, who takes care of the storage period, how the recordings are destroyed, etc. Also, web cookies are considered private data, and for that reason, it is obligatory to ask for permission.

According to the responses, the dominant number of hotels (91%) introduced changes. At the same time, other hotels (9%) are exposed to a much higher risk of endangering the privacy of their guests and employees.

Privacy Policy – Privacy Policy is a document created to explain what information a company or organisation collects about its customers and how they use that data and provide any other relevant information regarding their information.

Hotels should include in their Privacy Policy who is a data processor, basics or purpose for data collection, with whom they share data, duration of the processing, data security and (in some cases) how and why they transfer data outside of Serbia.

Purpose	Type of data
Fulfilling statutory obligations - registering a hotel guest with the competent authorities of the RS	• name and surname; • date of birth; • place of birth; • residence (city, state, postal code); • citizenship; • gender; • length of stay at our hotel; • identification on the document information (ID or passport).
Fulfilling contractual obligations - to register a guest at a hotel and provide it with e-services	• name and surname; • gender; • address; • information on the identification document (ID or passport); • e-mail address; • telephone number; • credit card number and payment information; • preferences regarding room type, room views, extra packages and the like; • the license plate number to provide guests with a parking space.

Table 3: Type of data according to the purpose – Example from “Hotel Mona”[14]

In Serbia, 85% of hotels have a Privacy Policy as a legal act. It would be interesting to find out if the management in these entities is fully acquainted with the purpose and content of this policy. So far, the practice has noticed that when the law imposes certain legal acts, their content is transcribed without a deeper consideration of specific circumstances.

Among the hotels that have adopted the Privacy Policy, only 69% have it publicly available on the Internet. At first glance, we can conclude that the situation in this area is good. Nevertheless, if we consider that 15% of hotels do not have this policy, published ones are much lower.

Consent - Hotels have the right to collect data if it is lawful, and two examples are listed in Table 3. However, suppose a hotel plans to use data without legal ground (for advertising, loyalty programs, web cookies). In that case, it must obtain consent from an informed person.

According to the response, only 70% of Serbian hotels provide consent.

Data protection officer (DPO) – The processor and the controller have to appoint the DPO if they are obliged under Article 56 of the Law – internal DPO. Otherwise, they may also nominate this position in case of need or use the consultancy services of the Commissioner (*Poverenik*) or another qualified person – external DPO. The primary role of the DPO is to monitor and control the processes and procedures and to give the assessment of the impact of processing on data protection.

The practice so far, especially in smaller organisations, has shown a lack of understanding of the importance and position of the DPO. There are even cases where organisations appoint a person without proper expertise or even basic knowledge to satisfy the form. The lack of professional support in this area can significantly jeopardise the data management process and thus the privacy of the persons whose data is collected and processed.

In Serbian hotels, 46% of respondents confirm that the management appoints DPO.

5. CONCLUSION

Protecting and preserving personal privacy in today’s digital society is becoming an increasing challenge. Numerous examples of accidental or malicious actions have led to the endangerment of this fundamental human right, followed by financial and/or other material losses and damages. Policymakers have recognised this threat, and in the last twenty years, most countries have adopted and then improved the legislation in this area. Business entities in Serbia, including hotels, are obliged to manage data under the Law on Personal Data Protection of the

Republic of Serbia. They also need to be aware of foreign trade, including online business, with the EU countries. In that case, they have to respect the General Data Protection Regulation (GDPR).

When it comes to hotels in Serbia, we found a high awareness of the need to adjust existing procedures to manipulate the personal data of guests and employees. Only a few hotels have not changed the processes and procedures that hopefully align with the previous law. Accordingly, most hotels have drawn up a Privacy Policy - the basic document they communicate with the persons whose data they process. Despite the importance of this document for the users of their services (hotel guests), not everyone has made it publicly available through a website. We would also like to point out that at least one-third of the hotels do not ask their guests for Consent, which gives them consent to expand the scope of the legal use of their personal data. Finally, most hotels do not appoint Data Protection Officers.

Based on the conducted research, we can conclude that management in the Serbian hotel industry provides compliance with the law's provisions regarding personal data protection. Even so, insight into specific procedures or inspection reports would give us more precise results to avoid biased or unqualified responses. Proper education is the best way to overcome unintentional law violations, while intentional illegal actions should be sanctioned. In response to this need, the Assistant Secretary-General held a training on implementing the Law on Personal Data Protection for hotels and tourist organications on October 10, 2019, in the Chamber of Commerce and Industry of Vojvodina in Novi Sad.[15] We also believe that it is essential for domestic hoteliers to be familiar with the provisions of the application of the GDPR in their objects.

6. ACKNOWLEDGEMENT

This paper results from the research financed by the Ministry of Education, Science and Technological Development of the Republic of Serbia.

REFERENCES

[1] Tidy, J. British Airways fined £20m over data breach, BBC News, October 30 2020, Available at URL: https://www.bbc.com/news/technology-54568784 (Accessed July 13 2021).

[2] Tidy, J. Marriott Hotels fined £18.4m for data breach that hit millions, BBC News, October 30 2020, Available at URL: https://www.bbc.com/news/technology-54748843 (Accessed July 13, 2021).

[3] UNCTAD, Data Protection and Privacy Legislation Worldwide, Available at URL: https://unctad.org/page/data-protection-and-privacy-legislation-worldwide (Accessed July 11, 2021).

[4] Popović, U., Andrejević, M., Serbia - Data Protection Overview, December 2020, Available at URL: https://www.dataguidance.com/notes/serbia-data-protection-overview (Accessed July 20, 2021).

[5] Statista, Lock, S. Global hotel and resort industry market size worldwide 2011-2021, May 2021, Available at URL: https://www.statista.com/statistics/1186201/hotel-and-resort-industry-market-size-global/ (Accessed July 13, 2021).

[6] Statista, Hoist, A. Amount of data created, consumed, and stored 2010-2025, Available at URL: https://www.statista.com/statistics/871513/worldwide-data-created/ (Accessed July 20, 2021).

[7] Hammouchi, H., Cherqi, O., Mezzour, G., Ghogho, M., El Koutbi, M., Digging Deeper into Data Breaches: An Exploratory Data Analysis of Hacking Breaches Over Time, International Symposium on Machine Learning and Big Data Analytics for Cybersecurity and Privacy (MLBDACP), April 29 - May 2, 2019, Leuven, Belgium, doi: 10.1016/j.procs.2019.04.141.

[8] MDU, Data breaches, November 2020, Available at URL: https://www.themdu.com/guidance-and-advice/guides/gdpr-data-breaches, (Accessed July 21, 2021).

[9] Garrison, C.P., Ncube, M. A longitudinal analysis of data breaches, Information Management & Computer Security, Vol. 19 No. 4, 2011, pp. 216-230, doi: 10.1108/09685221111173049.

[10] PwC, Cyber and data security in the hotel industry, edited by Maddison, M. & Grinnell, A., 2016. Available at: https://www.pwc.com/m1/en/publications/cyber-and-data-security-in-the-hotel-industry.html. (Accessed July 20, 2021).

[11] GDPR official website, Available at URL: https://gdpr.eu/, (Assessed July 21, 2021)

[12] DPP-GDPR, How to Handle the Legal Implications of a Data Breach Effectively, Available at URL: https://www.dpp-gdpr.com/news/legal-implications-of-a-data-breach/, (Assessed July 21, 2021)

[13] "Danas" report on the Tribune, Available at URL: https://www.danas.rs/drustvo/tribina-zakon-o-zastiti-podataka-o-licnosti-nejasan-i-onima-koji-ga-primenjuju/ (Accessed July 23, 2021)

[14] Official website of Hotel Mona, Available at URL: https://monahm.com/en/politika-privatnosti/ (Accessed July 25, 2021)

[15] Report of the Commissioner for 2019, Available at URL:

https://www.poverenik.rs/images/stories/dokumentacija-nova/izvestajiPoverenika/2019/Izvestaj-za2019.pdf
(Accessed August 1, 2021)

ADVANCED CRYPTOGRAPHY BY USING CHRISTOFFEL SYMBOLS

NENAD O. VESIĆ

Mathematical Institute of Serbian Academy of Sciences and Arts, n.o.vesic@outlook.com

Abstract: *In this paper, a metric tensor and the corresponding Christoffel symbols are applied for text-encryption. The uncountable infinities of real numbers and quadratic functions are used for better hiding of messages. The corresponding decryptions are relatively fast if the key is known. The paper is consisted of introduction and three parts. History of text-data-processing is recalled in introduction. In the first part of the paper, a small part of differential geometry necessary for this research is presented. In the second part, the encryption and decryption algorithms are presented. Finally, in the third part, the presented algorithm is tested. Program support is realized by software package Mathematica.*

Keywords: *encryption, decryption, safety, metric tensor, Christoffel symbols, differential equations*

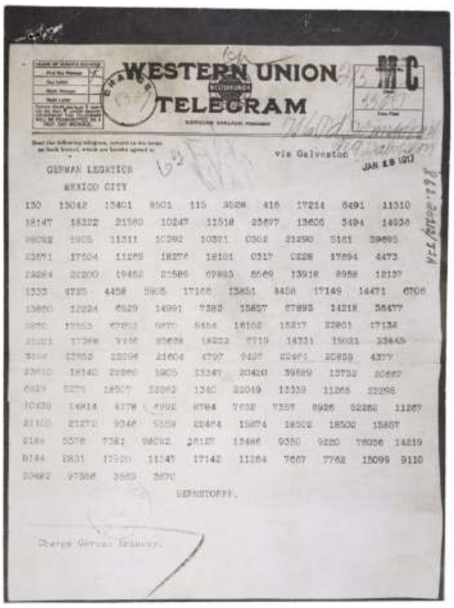
1. INTRODUCTION

Cryptography, often considered to provide secret communication over insecure channels has a long history, from antient times and pen-and-paper methods through machines and the well-known and applied new age algorithms. One of the major contributions of cryptography are the achievements of the mathematicians in Bletchley Park's team of scientists during the WWII. Using some of the encrypted data, scientists obtain as much as possible information about the original messages. One of the proved good procedures for such project is the solving the key, usually unique for a particular message. Depending on the number of keys used in the encryption and decryption process, we have two major parts of cryptography. Symmetric encryption handling the same key for both processes use block ciphers encrypting several bits as a single unit. DES (Data Encryption Standard) based on Shannon's idea was created in 1977 and considered unbreakable until Differential Cryptanalysis proposed by Biham and Shamir [1] and Linear Cryptanalysis method created by Matsui [2] required less complexity than a key-exhaustive search attack. Improvement came with a DES3 increasing the key size and AES with a 128- or 256-bit key size [3].

In [4] the initial idea for a common secret key over a public channel was created, enabling the creation of RSA algorithm [5] based on the number theory. Factorization of number obtained multiplying the large prime numbers guaranties the inviolability of the RSA system, demanding large 2048-bit key size. This problem can be overcome using significantly smaller key size in Elliptic Curve Cryptosystem.

The secrecy of transmitted messages holds an important part in modern electronic communication. As history has showed us, different authors have created different algorithms to hide the messages which should be sent to receiver [6].

From Caeser until now, people like to preserve their communication. Many different text data hiding algorithms are used [7]. Historicaly, the most famous code is Zimmermann's telegram [8]. After this message was broken, USA declared war on Germany (April 6, 1917).



WESTERN UNION
TELEGRAM

via Galveston

JAN 19 1917

GERMAN LEGATION
MEXICO CITY

```
130    13042  13401  8501   115    3528   416    17214  6491   11310
18147  18222  21560  10247  11518  23677  13605  3494   14936
98092  5905   11311  10392  10371  0302   21290  5161   39695
23571  17504  11269  18276  18101  0317   0228   17694  4473
22284  22200  19452  21589  67893  5569   13918  8958   12137
1333   4725   4458   5905   17166  13851  4458   17149  14471  6706
13850  12224  6929   14991  7382   15857  67893  14218  36477
5870   17553  67893  5870   5454   16102  15217  22801  17138
21001  17388  7446   23638  18222  6719   14331  15021  23845
3156   23552  22096  21604  4797   9497   22464  20855  4377
23610  18140  22260  5905   13347  20420  39689  13732  20667
6929   5275   18507  52262  1340   22049  13339  11265  22295
10439  14814  4178   6992   8784   7632   7357   6926   52262  11267
21100  21272  9346   9559   22464  15874  18502  18500  15857
2188   5376   7381   98092  16127  13486  9350   9220   76036  14219
5144   2831   17920  11347  17142  11264  7667   7762   15099  9110
10482  97556  3569   3670
```

BERNSTORFF.

Charge German Embassy.

Image 1: Zimmermann Telegram

We are motivated with algorithm for text-data encryption and decryption presented in [9]. In this algorithm, authors used matrices and polynomials to encrypt text. In this paper, we will develop this idea by using special matrix valued functions and the corresponding Christoffel symbols [10, 11].

2. NECESSARY DIFFERENTIAL GEOMETRY

Before presenting the enctyption and decryptioon algorithm, we will present the necessary terms in differential geometry needed to better understand the topic.

An N-dimensional manifold $M_N = M(x^1, \dots x^N)$ equipped with a (covariant) metric tensor $\underline{\hat{g}}$ whose components are $g_{\underline{ij}} = \underline{g}(x^i, x^j)$, $g_{\underline{ij}} = g_{\underline{ji}}$, is Riemannian space $\mathbb{R}_N$ (for details, see [11]). We assume that the matrix $\left[g_{\underline{ij}}\right]$ is regular, i.e. $\det\left[g_{\underline{ij}}\right] \neq 0$. The components of contravariant metric tensor $\underline{\hat{g}}^{-1}$ are $[g^{\underline{ij}}] = \left[g_{\underline{ij}}\right]^{-1}$.

If $g_{\underline{ij}} = g_{\underline{ij}}(t)$, for the variable $x^1 = t$, the space $\mathbb{R}_N = \mathbb{R}_N(t)$ is space-time. In this case, the coordinates x^i, $i = 2, \dots, N$, are constants. The geometrical objects

$$\Gamma_{i.\underline{jk}} = \frac{1}{2}\left(g_{\underline{ji},k} - g_{\underline{jk},i} + g_{\underline{ik},j}\right), \tag{1}$$

for partial derivatives $\partial g_{\underline{ij}}/\partial x^k$ denoted by comma, are the Christoffel symbols of first kind.

The special Christoffel symbols of first kind for space-time $\mathbb{R}_N(t)$ which are necessary for our research are

$$\Gamma_{1.\underline{jk}} = \Gamma_{1.\underline{kj}} = -\frac{1}{2} g_{\underline{jk},1}, \tag{2}$$

for $j, k = 1, 2, \dots, N$.

For the known Christoffel symbols $\Gamma_{i.\underline{jk}}$, the corresponding metric tensor is

$$g_{\underline{ij}} = -2\int \Gamma_{1.\underline{ij}} dt + c_{ij} = 2\int \Gamma_{i.\underline{1j}} dt + c_{ij} = 2\int \Gamma_{i.\underline{j1}} dt + c_{ij}. \tag{3}$$

With the necessary terms from differential geometry in place, we may proceed with the encryption and decryption algorithms.

3. ENCRYPTION AND DECRYPTION

The algorithms for encryption and decryption of a text will be presented in this section.The dimension $\boldsymbol{N}$ of space-time $\mathbb{R}_{\boldsymbol{N}}(\boldsymbol{t})$ is enough large. Bijective linear function $\boldsymbol{b}\colon \mathbb{N} \to \mathbb{N}$, and array $\boldsymbol{\mathcal{A}}$ composed of $\boldsymbol{M}$ rows with not necessary equal numbers of elements in different rows are given.

The object $\mathcal{A}_{pq}$ is the q-th element in the p-th row of the array $\mathcal{A}$. The position (p, q) of a character from the array $\mathcal{A}$ is transformed to the pair (u, v) for $u = p + M \cdot n_1, v = q + \mathcal{A}_p \cdot n_2$ for number of elements in the p-th row of array $\mathcal{A}$ equal $\mathcal{A}_p$ and $n_1, n_2 \in \mathbb{N}$. This pair is represented by complex number $z_{pq} = u + iv = p + M \cdot n_1 + i \cdot \left(q + \mathcal{A}_p \cdot n_2\right)$.

The position (p_k, q_k) of k-th character in text τ is characterized by complex number $z_k = p_k + i \cdot q_k$. The transformed position (u_k, v_k) of this character is characterized by complex number $\tilde{z}_k = u_k + i \cdot v_k$.

Let us encrypt the text τ consisted of c characters.

Encryption

Inbox: Private key consisted of array $\boldsymbol{\mathcal{A}}$ with $\boldsymbol{M}$ rows with not necessarily equal numbers of elements in any row and function $\boldsymbol{b}(\boldsymbol{v}) = \boldsymbol{v} + \boldsymbol{n_b}$, for coefficient $\boldsymbol{n_b}$, and the text $\boldsymbol{\tau}$ of $\boldsymbol{c}$ characters.

E1: For the k-th character in text τ find the corresponding position (p_k, q_k) of this character in the array $\mathcal{A}$.

E2: The pair (p_k, q_k) transform to pair $(u_k, v_k) = \left(p_k + M \cdot m, q_k + \boldsymbol{\mathcal{A}}_{p_k} \cdot n\right)$, for integers m, n and the number of elements in the p_k-th row of array $\mathcal{A}$ equal $\boldsymbol{\mathcal{A}}_{p_k}$.

E3: The pair (u_k, v_k) transform to polynomial $\pi_k(t) = t^2 - 2u_k t + u_k^2 + v_k^2$.

E4: Create the ordered set $\Pi = \{\pi_1(t), \dots, \pi_c(t)\}$.

E5: Create $\widetilde{N} = N(N+1)/2 - c$ polynomials $\tilde{\pi}_{c+u}(t) = t^2 - (r_{c+u} + s_{c+u})t + r_{c+u}s_{c+u}$, $u = 1, \dots, \widetilde{N}$ for integers r_{c+u}, s_{c+u}.

E6: Complement the set Π to ordered set Π^* with polynomials $\tilde{\pi}_{c+u}(t)$ before, between and after the polynomials $\pi_k(t)$. In this way, the ordered set $\Pi^* = \left\{\pi_1^*(t), \dots, \pi^*_{\frac{N(N+1)}{2}}(t)\right\}$ is obtained.

E7: Create the square matrix $\left[h_{\underline{ij}}\right]$ of the type $N \times N$ whose elements are

$$h_{\underline{ij}} = \begin{cases} \pi^*_{i_j^*}(t), & i \leq j, \\ \pi^*_{j_i^*}(t), & i > j, \end{cases} \tag{4}$$

for $i_j^* = \frac{i\cdot(i-1)}{2} + j$.

E8: Expand the matrix $\left[h_{\underline{ij}}\right]$ to the matrix $\left[g_{\underline{ij}}\right]$ with elements

$$g_{\underline{ij}} = \begin{cases} p_{11}(t), & i = j = 1, \\ 0, & i = 1 \text{ and } j > 1 \text{ or } j = 1 \text{ and } i > 1, \\ h_{\underline{(i-1)(j-1)}}, & \text{otherwise.} \end{cases}$$

E9: Form the matrix $\Gamma = \left[\Gamma_{1.\underline{ij}}\right] = \begin{bmatrix} \Gamma_{1.\underline{22}} & \cdots & \Gamma_{1.\underline{2N}} \\ \vdots & \ddots & \vdots \\ \Gamma_{1.\underline{N2}} & \cdots & \Gamma_{1.\underline{NN}} \end{bmatrix}$ of the corresponding Christoffel symbols with respect to the metric tensor whose components are $\left[g_{\underline{ij}}\right]$.

E10: The components $g_{\underline{ij}}$ of the matrix $\left[g_{\underline{ij}}\right]$ are of the form

$$g_{\underline{ij}}(t) = t^2 + p_{ij}t + q_{ij}. \qquad (5)$$

The corresponding Christoffel symbols are of the form $\Gamma_{1.\underline{ij}} = -t - p_{ij}$. The corresponding constant c_{ij} from the equation (3) is $c_{ij} = q_{ij}$.

- **Output:** Public key $\Gamma_{1.\underline{ij}}(0) = \left[-p_{ij}\right]$. Message is $\mu = \left[q_{ij} - \boldsymbol{b}(0)\right]$.

For decryption, we need the following operation

$$Mod_n(m) = \begin{cases} n, & n|m \\ \left\{\frac{m}{n}\right\}, & \text{otherwise,} \end{cases} \qquad (6)$$

for the rest of division m by n marked by $\left\{\frac{m}{n}\right\}$.

Decryption

If components of metric tensor are $g_{\underline{ij}}(t) = t^2 + at + b$, for real constants a,b, the corresponding Christoffel symbol is

$$\Gamma_{1.\underline{ij}} = -t - \frac{a}{2} = -t + publickey_{ij}. \qquad (7)$$

The number a is corresponding component of the public key. The number b is corresponding component of $message$, $b = message_{ij} + \boldsymbol{b}(0)$. Therefore, the corresponding component $g_{\underline{ij}}$ of metric tensor is

$$g_{\underline{ij}} = -2\int_0^t \left(-u + publickey_{ij}\right)du + message_{ij} + \boldsymbol{b}(0).$$

After using the operation Mod1n, we transform metric tensor to pairs of integers which correspond to the positions of characters in the list of symbols.

In this way, we decrypt the public key.

4. DISCUSSION AND CONCLUSION

In this paper, we have presented a lightweight encryption and decryption algorithm with the purpose to implement it in systems such as Cyber Physical Systems and Industry 4.0 in general. We applied Christoffel symbols to encrypt and decrypt a plain text message. The process for decryption of public key is presented. We have successfully tested the algorithm using Wolfram Mathematica software [11, 12]. We plan to conduct a comparative analysis of this algorithm with standard ones to test its efficiency, as well as to get a better grasp where it can be applied.

We are most certainly motivated to implement this algorithm wherever possible, with an emphasis on modern encryption systems. Our future work includes the implementation of this algorithm in modern learning management systems (LMSs), as was shown in [13], as well as the inclusion in Smart Cities and Internet of Things (IoT). We firmly believe that this algorithm can be used in wireless sensor networks as well, implemented in the central sink nodes for backbone-level communication, and perhaps in sensor node equipped with a processor as well.

ACKNOWLEDGEMENT

This paper is financially supported by the Serbian Ministry of Education, Science and Technological Development through Mathematical Institute of the Serbian Academy of Sciences and Arts.

REFERENCES

[1] E. Biham, A. Shamir, “A. Differential Cryptanalysis of the Data Encryption Standard”, Springer: Berlin, Germany, 12993.

[2] M. Matsui, “Linear Cryptanalysis Method for DES Cipher”, In Advances in Cryptology—*EUROCRYPT ’93*; Springer: Berlin, Germany, 1993; pp. 386 – 397.

[3] K. Kim, “*Cryptography*: A New Open Access Journal. *Cryptography*, vol. *1*, no. 1. 2017.

[4] W. Diffie, M. Hellman, “New Directions in Cryptography”, IEEE Trans. Inf. Theory, vol. *22*, 644 – 654, 1976.

[5] R. L. Rivest, A. Shamir, L. Adleman, “A Method for Obtaining Digital Signatures and Public-Key Cryptosystems”, Commun. ACM **1978**, 21, 120 – 126.

[6] T. H. Cormen, C. E. Leiserson, R. L. Rivest, C. Stein, “Introduction to Algorithms,” Third Edition, MIT Press, Massachusets, 2009.

[7] D. Luciano, T. Prichett, “Cryptology: From Caesar Ciphers to Public-key Cryptosystems,” The College Mathematics Journal, Vol. 18, No. 1, 2 – 17, 1987.

[8] Washington Exposes Plot, available online: https://timesmachine.nytimes.com/timesmachine/1917/03/01/102318118.pdf. (Accessed: Nov 2021).

[9] N. O. Vesić, D. J. Simjanović, “Matrix-Based Algorithm for Text-Data Hiding and Information Processing,” Military Technical Courier, Vol. LXII, No. 1, 42 – 57, 2014.

[10] J. Mikeš, E. Stepanova, A. Vanžurová, “Differential Geometry of Special Mappings,” Palacký University, Olomouc, 2015.

[11] Lj. S. Velimirović, P. S. Stanimirović, M. Zlatanović, “Geometry of Curves and Surfaces Covered by Software Package Mathematica,” in Serbian, Faculty of Sciences and Mathematics, Niš, 2010.

[12] P. S. Stanimirović, G. V. Milovanović, “Program Language Mathematica and Applications,” in Serbian, Faculty for Electronic Engineering, Nis, 2002.

[13] N.O. Vesić, N. Zdravković, D. J. Simnjanović, „Securing Online Assessments Using Christoffel Symbols,“ in Proc. Of the 11th Conference on eLearning, pp. 54-57, 2020.

The 12th International Conference on Business Information Security (BISEC-2021), 3rd December 2021, Belgrade, Serbia

MALICIOUS FUTURE OF AI: TRANSCENDENTS IN THE DIGITAL AGE

ZLATOGOR MINCHEV

Institute of ICT/Institute of Mathematics & Informatics, Bulgarian Academy of Sciences, zlatogor@bas.bg

Abstract: *Understanding the modern digital world present and future evolution definitely requires an adequate transformational outlook to the Artificial Intelligence (AI). Whilst nowadays still the singularity idea of humans and machines is a bit far away, the process of AI generalization is irrevocably progressing, challenging the security environment. The paper reveals a methodological approach for identification of AI malicious future implementations, combining expert, crowdsourcing and reference data. A threefold approach defining: Security Landscape Formation, Advanced Risk Analysis and Results Smart Assessment towards the next 10-15 years now towards 2036 is accomplished. Both structured and system-of-systems analytical approaches are accomplished on a modelling base, providing a holistic scenario-based result within future digital transcendents. The findings are proactively assessed, joining algorithmic machine and human intellect in a multicriteria probabilistic manner. Finally, some generalizations and discussions on the outlines are also presented.*

Keywords: *AI Malicious Use, Digital Transendents, Future Security, Holistic Future Foreseeing, H-M Joint Assessment*

1. INTRODUCTION

The evolution of Artificial Intelligence (AI) from more than 70 years now is still positioning machines in an obedient role vs human intellect, though gaining a tremendous progress [1]. The modern fundament of AI is a bit far of self-aware algorithmic implementations because the context is defined by humans, whilst a general objective for variables' optimization is searched with their realization. Hopefully, this drawback will be blurred with the General AI, as presently the role of AI is quite narrow even though progressive [2]. The huge interconnectivity of Internet of Things (IoTs) with 4G/5G mobile, wireless & satellite technologies has successfully established a new, "trans-reality" having quite smart & mixed appearance [3]. Adding at the same time, semi-autonomous activities in the social networks, working and living environments with a new digital lifestyle, extended capabilities, senses & opportunities. Being mostly related to IoTs miniaturization & implantation, either integration and automation for a new level of H2M & M2H interaction, this change is getting a disruptive role for the future digital society [4]. The trans-humanization and autonomous systems creation are becoming a new tool for smart political governance, whilst keeping the ethical balance. AI distracting role is actually addressing a triplet for the new, mixed environment: *technological*, *objective* and *cyber* segments. This is finally creating an innovative, hybrid ecosystem that becomes quite habitual for the emerging new digital age, approaching the sci-fi futuristic movies scenarios. Being chronically overflooded with data and information, the knowledge elicitation is becoming of vital importance for proper AI future development. The new, comprehensive change is expected to establish a completely transformed, new digital society [5]. Being related to plenty of expected and unexpected, useful or malicious contextual transcendents (threats, challenges, risks, opportunities uncertainties, gaps, divides) the transformation need to be suitably handled [6]. Having natural and non-natural origin, together with the technological progress for a safer and resilient future world [7]. The phenomenon is rather significant, especially for the next 10-15 years now, aiming technological singularity and technological stakeholders, society domination. Being also without clear regulation and recently fostered with COVID-19 pandemic, the digital society transformation is getting rather uncertain [8]. Luckily, these trends have not completely changed our inner world of human only traits (emotions, intuition, creativity, ethics, etc.) and objectives (happiness, success, well-being, etc.) [5], [9]. Further in the paper, a methodological framework for AI future risks proactive identification in the context of security landscape expected formation for a resilient & secure future digital society will be presented. The resulting findings are further assessed, using both human & machine dual approach, providing a complete outlook to the presented study.

2. STUDY FRAMEWORK

The framework is based on the ideas adapted after [7], [10] for the future hybrid ecosystem joint human-machine comprehensive exploration, trying to identify the malicious future of AI in the new digital age. A threefold approach (see Figure 1), encompassing: (i) *Security Landscape Formation*, (ii) *Advanced Risk Analysis* & (iii) *Results Smart Assessment* is further accomplished. The idea is to build a scenario context for the digital society future transcendents towards year 2036, that is further used in the risk analysis phase, providing multiple opportunities for dynamic evolution.

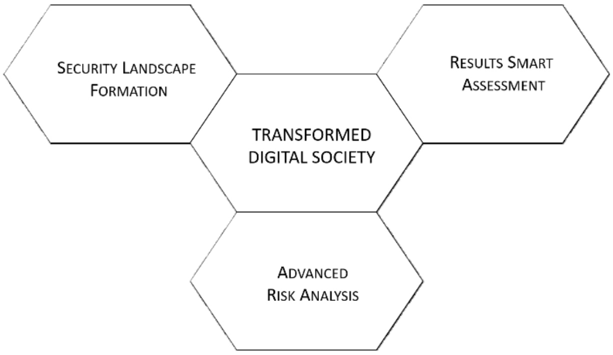


Figure 1: A study framework for AI transcendents malicious use identification in the new digital age

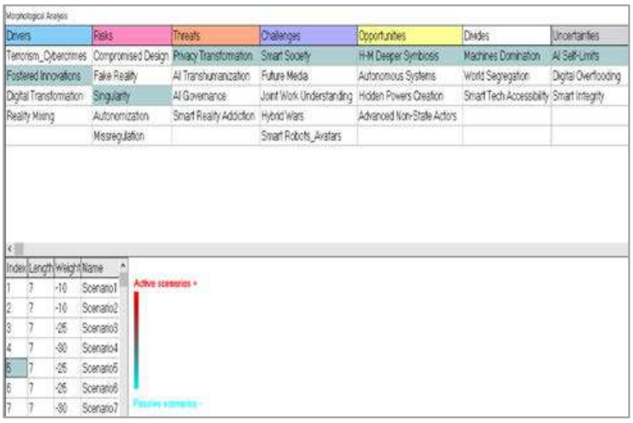
Finally, adding hybrid human-machine smart evaluation, the obtained results are assessed in advance, using a multicriteria set of objectives.

3. EXPERIMENTAL IMPLEMENTATION

The presented in Figure 1, study framework experimental implementation is further given, following the three outlined phases, whilst aiming proactive objective for AI malicious use risks proper contextual identification, together with obtained results comprehensive assessment towards year 2036.

3.1. Security Landscape Formation

The approach accomplished in this phase is successfully handling multiple heterogeneous ideas, using tailored structural (morphological) analysis of expert beliefs and supportive reference data towards the future [11]. In brief, a multidimensional representation is aggregated with selected dimensions, defining the transcendents of the problem of interest, using mutually exclusive alternatives in each of the dimensions. The resulting weighted combination of different alternatives (one of each dimension) is called a “scenario”. All combinations are forming a set of active (tangible) & passive (intangible) scenarios (together with median (neutral) ones in some cases, according to the analytical needs), encapsulated in a cross-consistency matrix. For the present study, a scenario cross-consistency matrix of seven dimensions (*Drivers*, *Risks*, *Threats*, *Challenges*, *Opportunities*, *Divides*, *Uncertainties*) has been defined (see Figure 2) in I-SCIP-MA software environment [12], using data & expert referencing, gathered around [13]. The total number of combinations is above hundred thousand (plausible & implausible scenarios, N = 100800), whilst the plausible ones, that are further noted in the analysis findings hereafter is around five thousand (N’ = 4920), and the rest (implausible) is more than ninety-five thousand (N”= 95880).



Morphological Analysis

Drivers	Risks	Threats	Challenges	Opportunities	Divides	Uncertainties
Terrorism_Cybercrimes	Compromised Design	Privacy Transformation	Smart Society	H-M Deeper Symbiosis	Machines Domination	AI Self-Limits
Fostered Innovations	Fake Reality	AI Transhumanization	Future Media	Autonomous Systems	World Segregation	Digital Overflooding
Digital Transformation	Singularity	AI Governance	Joint Work Understanding	Hidden Powers Creation	Smart Tech Accessibility	Smart Integrity
Reality Mixing	Autonomization	Smart Reality Addiction	Hybrid Wars	Advanced Non-State Actors		
	Missregulation		Smart Robots_Avatars			

Index	Length	Weight	Name
1	7	-10	Scenario1
2	7	-10	Scenario2
3	7	-25	Scenario3
4	7	-30	Scenario4
5	7	-25	Scenario5
6	7	-25	Scenario6
7	7	-30	Scenario7

Figure 2: Plausible & implausible cross-consistency scenario matrix of AI malicious use in I-SCIP-MA

Being quite aggregated this future security landscape could be commented with some details as follows:

– The risks related to future technological singularity, together with autonomization & missregulation in the new digital society are giving most negative expectations (about 2/3 of all the studied scenarios – 2821). These risks are mainly driven by fostered innovations and digital transformation in contrast to the desire of machine domination and smart technologies accessibility. Additional accent is also granted to self-replicating & self-repairing but aiming in particular the unpleasant, monotonous or dangerous activities. Whilst, mainly addressing future hybrid wars (in the sense of human and machine intelligence clashing in a direct & indirect multivalence manner), future smart robots and avatars smart society usage.

The main threats in this sense are emerging in the smart reality addiction of humans and privacy transformation. As in the future it is expected that most of the human current private space (regular activities, habits, preferences, skills, memories, health status, etc.) will be accessible for machine processing and smart communication. This dual human-machine interrelation is mainly cultivated by modern social networks, mobile smart gadgets and cloud automated services evolution.

The new digital transformation is expected to produce mainly opportunities for autonomous systems progressive development that are operating in the new mixed (objective & cyber) reality, adding also non-state actors advancing and new hidden powers potential domination due to the AI evolutionary usage in the social segment.

Finally, all these changes are creating an overflooded landscape of numerous digital solutions (technologies, services, information, data, and resulting new knowledge outlines) that will be also difficult for substantial integration due to the highly granulated and dynamic landscape of constantly competing machine intelligence with human natural governance & control desire over this mixed socio-technological landscape.

– From another perspective however, according to the present findings the rest of the scenario combinations (a little bit more than 1/3, i.e. 1864 plus 235 with completely neutral assessment) are expected to be active ones. These are mainly related to risks of fake, either simulated reality, and compromised design existence, using future AI and technological gadgets, providing new senses and feelings that already will be difficult for control.

As far as the establishment of such phenomena is already a known fact (e.g. with fake news, disinformation & deep fakes multimedia, security by design compromising) it is expected that in the near future, AI evolution will support the successful handling of these risks, adding more certainty to the regulation process.

It could be assumed that this technological progress of AI malicious use is mostly driven with criminal, either terrorist motives, using the complex socio-technological smart reality mixing and we as a society are expecting such innovative new security challenges. We will add here and the threats related to the role of state actors and private sector fostering the AI integration policy in the e-governance process and health oriented trans-humanization.

Challenging future smart society with joint human-machine work understanding and smart media, whilst providing mostly an opportunity of deeper symbiosis between humans and machines is also a question of economic investments. This is actually segregating the human intellect into a developed and lazier one's vs the developing and strongly uncertain but motivated, reflecting at the same time the future socio-technological world distributed segregation.

As the presented commentary on the scenario morphological transcendents analysis is not showing the direct causality origins of the outlined findings, a further advanced risk analysis is provided.

3.2. Advanced Risk Analysis

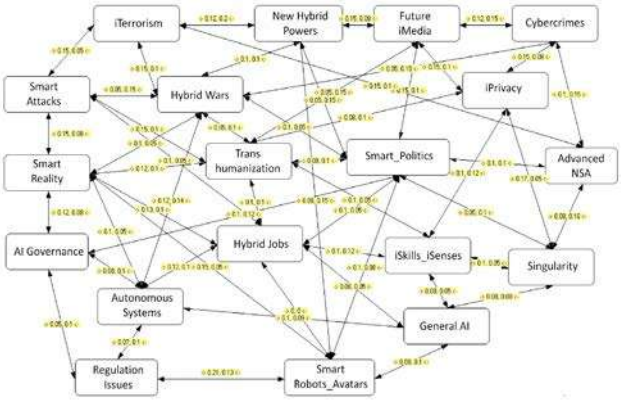
A deeper understanding the future security risks origins of AI malicious use requires both – causality and dynamic evolution implementation. Luckily, the problem could be sufficiently well solved, using system-of-systems analysis [12], [14]. However, it should be deeply underlined that the dynamic uncertainties coping (related to unexpected events existence, especially for the future) are quite difficult to be proactively assessed.

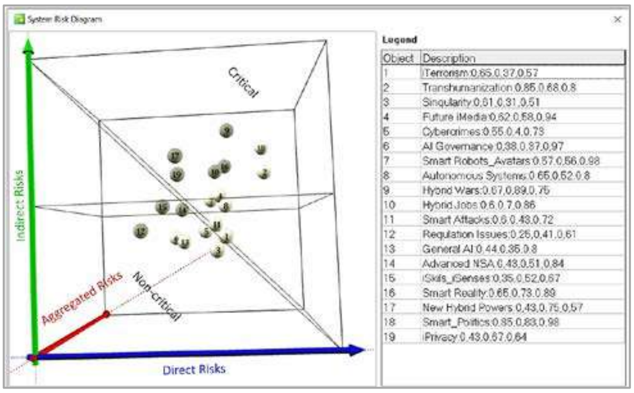
In this sense, the solution of this task is much more complicated for two main reasons: (i) the socio-technological systems evolution is a multidimensional task with different dynamics, as far as our knowledge and beliefs for the future are incomplete in general due to unexpected events appearance; (ii) the communication between different systems is difficult to be coherently matched for a certain moment of interest.

Solving both aspects is of extreme importance for successful modelling the whole problem for AI role in the future, as the singularity approaching is definitely related to natural intelligence equilibrium change that is quite interesting for detailed understanding.

The implemented model encompasses 19 entities gathered around the idea of a system-of-systems modelling, spreading the results on a causality base with I-SCIP-RA software environment assistance [10]. The representation uses both plausible and implausible scenarios, whilst accentuating on the passive scenario evolution set dynamics (see 3.1), using expert support with *Secure Digital Future 21* international forum initiative [13].

The final assessment towards 2036 is aggregated with a System Risk Diagram (SR Diagram, encompassing a probabilistic assessment of: *Direct* (forward, *Indirect* – backward and Aggregated risks), giving a resulting distribution in two main classes: *Critical* vs *Non-Critical*. Additional role in each of the two classes with *Active* (white) vs *Passive* (grey) entity generalization role is also given for the risk comprehensive tailoring.





Legend

Object	Description
1	iTerrorism:0.65,0.37,0.57
2	Transhumanization:0.85,0.68,0.8
3	Singularity:0.61,0.31,0.51
4	Future iMedia:0.62,0.58,0.94
5	Cybercrimes:0.55,0.4,0.73
6	AI Governance:0.38,0.87,0.97
7	Smart Robots_Avatars:0.57,0.56,0.98
8	Autonomous Systems:0.65,0.52,0.8
9	Hybrid Wars:0.67,0.89,0.75
10	Hybrid Jobs:0.6,0.7,0.86
11	Smart Attacks:0.6,0.43,0.72
12	Regulation Issues:0.25,0.41,0.61
13	General AI:0.44,0.35,0.8
14	Advanced NSA:0.43,0.51,0.84
15	iSkils_iSenses:0.35,0.52,0.67
16	Smart Reality:0.65,0.73,0.89
17	New Hybrid Powers:0.43,0.75,0.57
18	Smart_Politics:0.85,0.83,0.98
19	iPrivacy:0.43,0.67,0.64

Figure 3: Risk assessment model (top) & SR Diagram (bottom) results for future AI malicious use in I-SCIP-RA

The presented results outline an aggregated probabilistic risk for AI malicious use towards year 2036 as follows:

– ***Critical*** risk priorities are most probably expected for: *iTerrorism* (1), *Transhumanization* (2), *Future iMedia* (4), *Cybercrimes* (5), *Smart Robots_Avatars* (7), *Autonomous Systems* (8), *Smart Attacks* (11) & *Smart Politics* (18) having active role; *Hybrid Wars* (9), *Hybrid Jobs* (10), *Advanced Non-state Actors (NSA)* - (14), *Smart Reality* (16), *New Hybrid Powers* (17), *iPrivacy* (19) being passive.

– ***Non-critical*** priorities are given to: *Singularity* (3), *AI Governance* (6), *General AI* (13) with active role; *Regulation Issues* (12) & *iSkills_iSenses* (15) having passive role.

Obviously, the security risks expectations towards the near future are giving priorities to the AI implementation in the future media, crimes, terrorism, politics, together with autonomous systems, robots, avatars that are also going to be attacked, using intelligent solutions. At the same time, the upcoming human-machines hybridization with new jobs, wars, global powers creation, advanced non-state actors (NSA with AI) are going to be added with the future smart reality, being mostly consumers of the technological AI innovations for successful securing and progressive development.

Taking into account all these expected malicious fields of AI application, the good news here are coming from the non-critical risks' findings. As the General AI and singularity for total new machine type of governance are not going to happen so soon. In the near future with the advanced intelligent senses and skills (*iSenses & iSkills*) integration to humans with multiple AI assisted services and implanted transhumanization gadgets will be still regulated with the human intellect though overinformed with machine processed big data arrays.

The presented advanced and proactive risk analysis with the future security landscape though quite fascinating needs also to be suitably assessed.

3.3. Results Smart Assessment

Whether the expert based prognostic assessments for the future of AI malicious use are plausible or implausible in the already defined scenario context of digital transformation transcendents, it is extremely difficult to be only numerically judged in general, especially within the next 10-15 years. So, a joint smart approach has been further accomplished, adding human beliefs to a machine adaptable smart multicriteria optimization [15]. In this manner it is possible to get a feasible evaluation of potential future expectations [16], taking into account trends dynamics, following an S-shaped socio-technological dynamics probabilistic risks approximation tendencies [3]. This approach is well-suited for long-term tasks solving [17] and is accomplished further. In brief, the idea is to get a preliminary expert beliefs and audience crowdsourcing match with the advanced risk analysis selected results model, trying to understand whether a potential trend is applicable in a coherent multidimensional sense with a certain risk assessment model. An aggregated implementation results for *Smart Politics*, *Hybrid Wars*, *Future iMedia* & *Autonomous Systems* (selected critical entities from risk analysis results from Figure 3) is presented below.

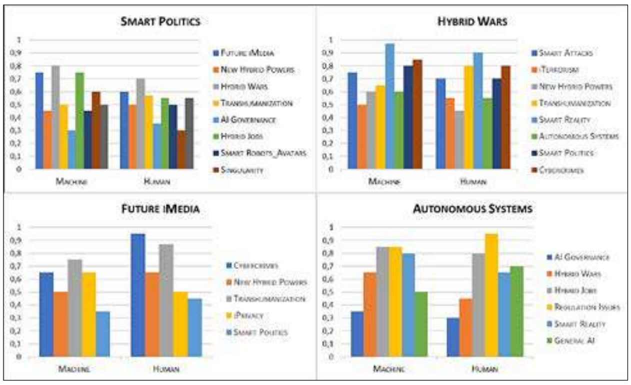


Figure 4: Selected aggregated multicriteria risks assessments of both human & machine future beliefs for AI malicious use, towards year 2036

Here it should be also noted that both reliable probabilities (P >> 0.5) and cycles dynamics evolution (following the assumptions of a social four-staged lifecycle evolution: "Prosperity", "Recession", "Depression", "Improvement", after [3]) of different model entities (that could be transformed to more detailed sub systems) have to be taken into account, achieving a successful aggregated risk assessment for the future.

As for the present example (see Figure 4), there are visible differences between the aggregated risks assessments and the selected illustrations, addressing only the substantial ones of human and algorithmic machine risks future assessments, applied over the same system-of-systems model, towards entities like: *General AI*, *Hybrid Jobs*, *Cybercrimes*, *Singularity*, *New Hybrid Powers*, *Transhumanization* and *Regulation Issues*. The approach is trying to assess the potential risks of AI usage in *Future iMedia*, *Autonomous Systems*, *Smart Politics* & *Hybrid Wars* entities in the model, towards a utopian or dystopian digital society transformation within the next 10-15 years now.

4. DISCUSSION

Identification of AI future malicious use is obviously dependable on the new socio-technological hybrid ecosystem digital transformation. Whilst identifying the potential scenarios of both narrow & general AI implementation in the new trans-reality with multiple smart gadgets, services and transhumans, the future technological progress is getting rather fascinating but also ambiguous, and thus – certainly somewhat dangerous. These naturally also creates numerous security transcendents that have to be proactively identified and handled, assuring a resilient, safe and progressive future smart society. Besides the natural intelligence still leading role, the AI singularity is definitely addressing an inevitable necessity of regulations establishment, especially towards autonomization, transhumanization and living environment destructive targeting.

The recently fostered by climate changes and COVID-19 pandemic digital services accelerated development and usage in our everyday life is inevitably also addressing the problems of fake reality, privacy, news and knowledge problems coping.

Together with the transformed hybrid misinformation, propaganda, cognitive manipulations and social engineering the future world becomes somewhat dystopian from technological domination perspectives. This jointly with the design compromising, know-how, data smart criminal attacks and autonomous systems is naturally producing new defense solutions necessity that are expected to cover the future rather heterogeneous and dynamic security landscape.

At the same time, the creation of new algorithms for constantly improving the AI progress is also assisting the technological singularity approaching with potential establishment of new, advanced political governance, hidden powers and non-state actors that potentially are also threatening the peaceful future society.

With the new AI technologies however, the natural intellect is also going to be advancing, adding future i-skills & i-senses that will obviously support the objective of successful keeping the balance of intellectual competition between the natural and artificial intelligence towards a pseudo-utopian but also divided (because of different future beliefs, understandings, lifestyle and adaptation in general) new trans-reality within the next 10-15 years now.

5. ACKNOWLEDGEMENTS

The results presented in this study are due to the technological, industrial & expert support obtained in the framework of the international forum initiative “Securing Digital Future 21” with more than sixty countries now, spread around the world, https://securedfuture21.org/

REFERENCES

[1]. S. Russell, & P. Norvig, “Artificial Intelligence: A Modern Approach”, Pearson, 2021.

[2]. R. Kurzweil, “The Singularity Is Near: When Humans Transcend Biology”, New York, Penguin Books, 2005.

[3]. Z. Minchev, et al, “Future Digital Society Resilience in the Informational Age”, Sofia, SoftTrade & Institute of ICT, Bulgarian Academy of Sciences, 2019.

[4]. L. Boyanov, “Digital World – The Change”, Avangard Prima, Sofia, 2021.

[5]. G. Leonhard, “Technology vs. Humanity: The Coming Clash between Man and Machine”, Fast Future Publishing, 2016.

[6] Z. Minchev, “Future Digital Society Transformational Transcendents & Gaps Extended Outlook”, Romanian Cyber Security Journal, No. 1, Vol. 2, pp. 11–18, 2020.

[7] Z. Minchev (Ed), “Digital Transformation in the Post-Information Age”, SoftTrade & Institute of ICT, Bulgarian Academy of Sciences, 2021.

[8] T. Walsh, “2062 The World that AI Made”, La Trobe University Press in conjunction with Black Inc., 2018.

[9] A. Braga, R. Logan, “The Emperor of Strong AI Has No Clothes: Limits to Artificial Intelligence”, Information, Vol. 8, No.4, 2017, DOI: 10.3390/info8040156

[10] 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, DOI: 10.13140/RG.2.2.31856.05125

[11] F. Zwicky, “Discovery, Invention, Research through the Morphological Approach, Macmillan”, 1969.

[12] 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, 1 ed., M. Hadji-Janev and M. Bogdanoski, Eds., IGI Global, 2015.

[13] Secure Digital Future 21 International Forum, https://securedfuture21.org/

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

[15] Z. Minchev, “Security Challenges to Critical Infrastructure of Future Smart Cities”, in Proc of BISEC 2019, September 13, Belgrade Metropolitan University, 2019, DOI: 10.13140/RG.2.2.18120.75525

[16] Z. Minchev, “Disruptive Effects of New Age Pandemic to Shifted Cyber Diplomacy due to Multilateral Mixed Transformation”, International Journal of Cyber Diplomacy, Vol. 1, Issue 1, pp. 49-59, 2021.

[17] V. Barnett, “Kondratiev and the Dynamics of Economic Development: Long Cycles and Industrial Growth in Historical Context”, Macmillan Press, 1998.

CLOUD INTELLIGENCE NETWORK FOR RANSOMWARE DETECTION AND INFECTION EFFECT REVERSING

STEFAN TAFKOV

IT for Security Department, Institute of ICT, Bulgarian Academy of Sciences, stefan.tafkov@iict.bas.bg

Abstract: *In the recent years, there has been a staggering increase of the complexity and methods of infecting computer systems through numerous ransomware attacks. This phenomenon raises the need to create new models and methods for analyzing malicious traffic, using evolutionary machine self-learning. Presently, there are different attack approaches, and in our study, we looked at a model for analyzing traffic data to and from the sponsor. Based on the data from the infected file, we can analyze the behavior of the virus and its future actions. Obtaining this data is possible through an Internet connection in a dynamic environment. In the presence of installed sensors in an environment that presupposes the presence of malware, it is possible to secure the system through the Cloud Intelligence Network and Cloud File Analyzing. The addition of sensor monitoring provides complete monitoring of the entire system. The monitoring is organized within multiple layers for analyzing and verifying different kinds of ransomware threats known behavior.*

Keywords: *Cloud Intelligence Network, Dynamic World, Cloud File Analyzing, Deep Residual Neural Network*

1. INTRODUCTION

Recently, there has been a visible trend increase of the complexity and methods of infecting computer systems through numerous ransomware attacks [1]. This phenomenon raises the need to create new models and methods for analyzing malicious traffic, implementing also evolutionary machine self-learning [2]. Today, there are multiple attack vectors in the cyberspace [3], so in the current study a model for analyzing malicious traffic and adequately respond to new threats, using machine deep learning, for adapting to the environment with ransomware cloud intelligence is accomplished. When analyzing in practice the data for experimental assessment, several families of ransomware attacks have to be tested in an isolated environment. The idea is to detect the duration of an attack and "how" each of the following ransomware attacks could gather data from the environment in which they were located, adding also –"what" method of attack they would choose. The data could be obtained thanks to analyzing the behavior of viruses, gathered with malicious web traffic that ransomware attack generates during the attacks. Studying the final data transfer between victim machine and ransom server a new module has been proposed and developed. The idea is to intercept such kind of malicious traffic and convert it for good purposes, such as recovering system encrypted data and gaining key for future generic attacks. For the purposes of the present study, the following ransomware families: "*Loki, WannaCry, REvil, TeslaCrypt, Petya, Cryptolocker, CrySIS*" were used as a test basis, being somewhat popular [4]. Additionally, sensors are required to be placed for collecting data from multiple sources. With the collected and analyzed traffic and behavior data, the accomplished model can further predict and re-generate damaged data by obtaining encryption key from a tipical ransomware attack. Using the entire vector of system monitoring sets, ones could monitor different kinds of malicious and normal activities. Implementing also machine learning [5] that work together with the entire system and collectively exchange data, the model could also predict and neutralize ransomware attacked system only if the users have all of modules installed on their systems. Further in the paper, both static & dynamic analysis of ransomware is outlined, together with hybrid model for experimental assessment, using residual neural networks. Finally, some practical experiments with results discussion are also given.

2. RANSOMWARE DETECTION MODEL

This part of the paper discusses different variations for detecting and identifying ransomware threats that mainly could be divided to static and dynamic ones. Below will be given more details on their practical implementation after [5]. Adding the co-operation between all modules (see Figure 1) and the defense agent that monitor and defend every computer system that it runs on. Using multiple layers of monitoring modules, defense agent takes care for preventing ransomware attacks by co-operating with the entire cloud network.

(i) *Cloud intelligence network*. The use of the intelligence network [6], which aims to deplete data from multiple sensors, provides both rules, signatures and detection models, as well as the ability to intercept sensors from ransomware attacks that are analyzed and extract encryption keys [7]. This module provides recovery of damaged data and rapid response to unknown threats without the need to reanalyze analyze each sensor.

(ii) *Monitoring and analyzing entire file system for malicious activity*. By monitoring the file system, using system watcher to: create, rename, delete, modify and open events that are logged. This data is analyzed by a model that checks for available YARA Rules [8]. Using this type of analysis of unknown attacks by loading rules into the system threats can be detected by complex analysis of PE file structure parameters and behavior of ransomware attacks.

(iii) *Analyzing an unknown file, using deep residual neural network*. The use and retrieval of data that is loaded in Deep Residual Neural Network (DRNN) is analyzed by a preliminary module for their conversion from FILE PE to Meta data and events, originating from the file of interest. In the presence of a larger network and the loading of a larger set of malicious and already proven safe files it is possible the network to expand and develop knowledge, thereby increasing the level of detection of unknown threats, response time and minimizing false positives detections.

(iv) *Process behavioral analysis*. This module is for analyzing the behavior of each individual process and service operating in a given PC system. By tracking input and output nodes and creating a virtual map of each event from each process. This type of mapping of events around a process provides data to the DRNN network. A built-in module to recognize suspicious behavior that monitors changes in the structure of: *Documents, Photos, Movies, Applications,* and analyzing whether there is a change that would lead to a user inaccessibility or denial of access, draws a conclusion about whether or not, that file has suspicious activity. By transmitting the source data to the first and second layer, enrichment is supported to detect the rules and the DRNN network. The system has monitoring of the following events in the system: “Change file PE structure, File lock, check for digital signature, change multiple files in one directory, attempt to access a remote server, create a local network map, and install services in the system”.

(v) *System insider honeypot*. By placing file bytes in the most frequently affected directories in the file system and constantly monitoring them, attempts to replace, attempt to encrypt, attempt to modify the file structure are sought. If positive, the output Meta data is transmitted to Cloud Intelligence Ransomware Meta Core (CIRMC) and Cloud Intelligence Network (CIN) where they are analyzed and reported malicious if they actually are. Every module in the model is intercommunicating with the others, so they can know exactly what type of file should or is analyzed, ommitting the delay of an intruder or incident. If there is any sign of ransomware activity the system focusses 10% of its resources to detect and get the right decision [5] if the potential attack is on the way or there is a false positive situation.

(vi) *Cloud intelligence key store*. At this level of the cloud intelligence network, the data obtained by the sensors is used when data is taken down, sending ransomware data to the attacker. Dynamic sensor analysis subtracts passwords before and after encryption, with a module designed to read in real time from ransomware memory and download encryption passwords before the virus has switched to the victim's encryption and blackmail phase. Through this dynamic code analysis, the data encryption keys are subtracted from the Cloud Data Analyzing (CDA). By communicating between all sensors via cloud, the service relies not only on infection prevention, but also on restoring systems that have already been hit by ransomware attacks.

(vii) *Cloud intelligence ransomware meta core*. At this level, a module is being built in the Cloud Intelligence Network [9], [10] system to process new metadata being added to the database in the system. Using data processing and storage for the purpose of identifying an attack, the following sensor parameters are required: “*MD5, SHA512, SSDeep, Hex, FILE PE Data, File Format, File type, Attack Time, Attack Pattern*”. In attack time, the system has the right to consult with Cloud Learning Reputation Protection System “CLRPS” [11] in order to provide in the absence of sensors this data. By storing and providing access from any sensor, the system can recognize an already known ransomware attack.

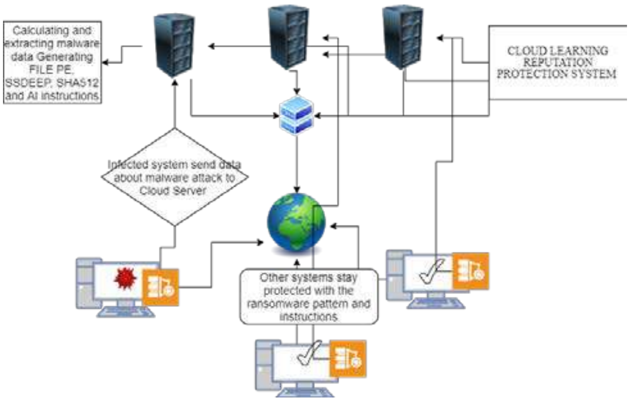

Calculating and extracting malware data Generating FILE PE, SSDEEP, SHA512 and AI instructions
CLOUD LEARNING REPUTATION PROTECTION SYSTEM
Infected system send data about malware attack to Cloud Server
Other systems stay protected with the ransomware pattern and instructions

Figure 1: CLRP system and incident response to new malware attacks

After infecting, system sensor sends data to CLRPS for extracting useful data and generate new AI model instructions and patterns. As soon as the system has the new information other computer systems in the network stays protected from malware that hit the first machine. Each user system and virtual environment including cloud file analyzing that is installed and protected by the module contains the following modules to detect prevent and respond to ransomware attack.

3. DEFENCE AGENT

Each agent shall use the following modules to provide immediate response and identification of ransomware attack using data exchange. Each module communicates with the other modules of the analyzing and protective circle, and when applying measures, detecting and removing malware, maximal accuracy and low levels of false positive results are aimed at maximizing accuracy and low levels of false positive results.

(i) *File System Protection module*. This module monitors the file system, using file analysis with a kernel operating on Deep Residual Neural Network [2], as well as a database core that loaded the signatures of the malicious programs and the "Whitelist" of programs, analyzed by the cloud system through deep learning and behavior analysis in a simulation environment. By tracking each activity in the "*Create, Delete, Replace, Rename, and Move*" file system, the module transmits the file data to the analytics cores. A file PE monitoring file structure analysis engine is also built-in. The File PE analysis and retrieval engine on the file system state kernel preserves the data and monitors for subsequent change. Through a module that takes care of health monitoring and proper operation of files in PCs, detect unwanted changes as an example of this is ransomware attack.

(ii) *Monitoring of startup processes and services* in the system provides a wide range of user activity, including the programs he uses. By inspecting the memory of each process and analyzing the CFA (Cloud File Analyzing) and providing data to the behavior monitoring kernel, the system monitors cyber-attacks on the file system.

(iii) *Prevention of malicious activity*. This module of the malware detection system is used with the kernel both for monitoring processes, services but also monitoring logs in PC. By analyzing behavior with predefined parameters for malicious activity that are subject to upgrade, conclusions are drawn over whether a process or file is malicious or not.

(iv) *Analyzing data from the cores to monitor suspicious activity*. The system sends both raw data to cloud service and ransomware-removed data decryption keys [12]. The benefits of the cloud network are distributed between each sensor in the PC. In case of insufficient parameters and time for analysis, dynamic portions of data transmission is performed to the cloud service, where the analysis of the given file, activity and data structure is carried out.

(v) *Cloud File Analyzing module*. Through a simulation environment, this module provides shared resources to sensors installed on a remote PC, thereby analyzing in a real environment the behavior of suspicious objects. Depending on the OS of a given PC, when sending cloud analysis data, the system decides under which virtual environment to start exploring a given file. Preset sensors for analysis provide source data to sensors located on PCs.

(vi) *Web Protection Module* keeps track on malicious internet activity and identification engine, while taking care of repelling unwanted or malware objects to be installed on a PC. The idea is accomplished through dynamic analysis of each website with a rating check engine and Cross-Site Scripting (XSS) (see Figure 2).

The following critical modules are identified in the development of the ransomware attack detection system that the system have to possess and counter. The modules are developed according to the needs of the systems on which they are installed [8].

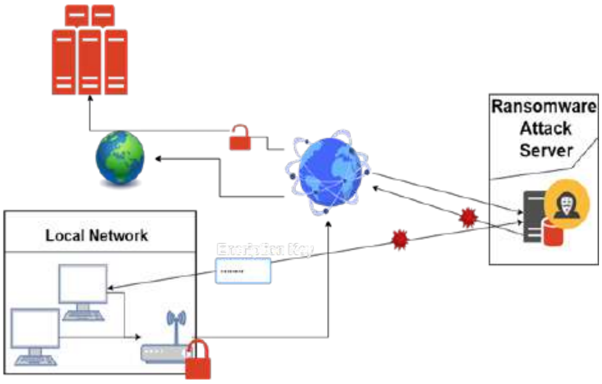


Figure 2: Principle of attack and interception of the encryption key with XSS

4. IMPLEMENTATION DETAILS

During the practical implementation, two test benchmark groups have been studied with a multi-threading approach, after the hybrid model from Figure 1, trained in 8 epochs: 100 & 10 000 ransomware probes (about 30-40 MB sized inputs with initially unknown, for the following families: *Loki*, *WannaCry*, *REvil*, *TeslaCrypt*, *Petya*, *Cryptolocker*, *CrySIS*, see Figure 3). The experimental machine was organized with M.2 ADATA NVME Adata GAMMIX S11 Pro, 2TB SATA 3 2TB GB HDD, 8-core 16-thread AMD RYZEN7 3700x 64-bit processor with L2 cache 4 MB, L3 cache 32 MB and 36 GB RAM [5]. And video card: MSI RADEON RX 580 ARMOR 4G OC.

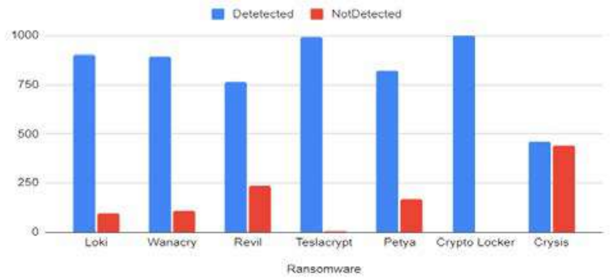


Figure 3: Experimental tests detection ratio, performed with the updated ransomware attack detection modules

Additional results from web traffic sniffer module that helps to recover encryption key and return encryption event are given in Figure 4. The idea is to guarantee, no file loosing, during the attacks, even though the ransomware hits the machine of interest.

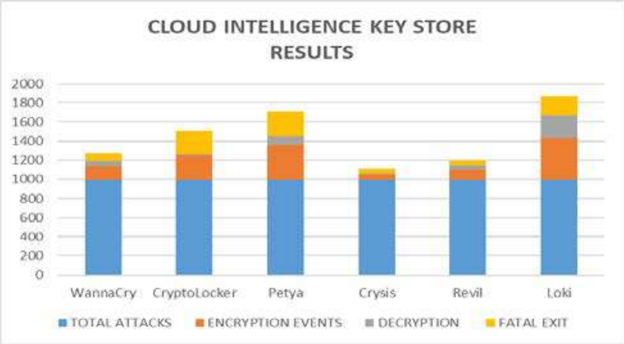


Figure 4: Test results from ransomware attack decrypted with CIKS module with total amount of test files per family of 1000 samples

5. CONCLUSION

In the time of the COVID-19 pandemic, the world as well as the virtual world witnessed new types of ransomware attacks and even more sophisticated ones, for stealing information and organizing more complex ones with APT s handling. The study outlines a faster and more reliable model for detecting ransomware attacks and repairing damage, if any. Through the ability to intercept malicious traffic and extract encryption keys, the model is presenting good performance results, both in detecting unknown ransomware attacks and in the recovery of corrupted files, produced from such attacks. The article results could further evolve into several important directions, to note: complex system for detecting malicious traffic, implementation of the sensors explained in model for use in a SANDBOX platform or analysis of new malware family smaples.

6. ACKNOWLEDGEMENTS

The obtained results in this study are partially supported by National Scientific Program "Information and Communication Technologies for a Single Digital Market in Science, Education and Security – ICTinSES", 2018-2021, financed by the Ministry of Education and Science, Republic of Bulgaria. Additional gratitude for the technological and expert support is also given to the international forum initiative "Securing Digital Future 21" with more than sixty countries now, spread around the world, https://securedfuture21.org/

REFRENCES

[1] "The State of Ransomware 2021", SOPHOS White Paper, April, 2021, https://bit.ly/3ciZzwn

[2] S. Russell, & P. Norvig, "Artificial Intelligence: A Modern Approach", Pearson, 2021.

[3] "DBIR 2021 Data Breach Investigation Report", Verizon, 2021, https://vz.to/32eukRv

[4] "Ransomware in a Global Context", Activity Report, VirusTotal, October, 2021, https://bit.ly/3ozpq8V

[5] S. Tafkov, Z. Minchev, "Ransomware Detection & Neutralization System", in Proc of X International Scientific Conference Hemus 2020 Research and investment in technology innovation – a crucial factor for defence and security, Defence Institute "Prof. Tsvetan Lazarov", Bulgaria, 2021, pp. II-144-II-152, DOI:10.13140/RG.2.2.21029.12009

[6] D. Medhi, K. Ramasamy, "Network Routing (Algorithms, Protocols, and Architectures)", Morgan Kaufmann, 2007.

[7] A. Orebaugh, G. Ramirez, J. Beale, J. Wright, "Wireshark & Ethereal Network Protocol Analyzer Toolkit", Syngress, 2007.

[8] V. Alvarez, "yara Documentation Release 4.1.0", October, 2021.

[9] M. Neidinger, "Python Network Programming Techniques: 50 real-world recipes to automate infrastructure networks and overcome networking challenges with Python", Packt Publishing, 2021.

[10] S. Burns, "Hands-On Network Programming with C# and .NET Core: Build robust network applications with C# and .NET Core", Packt Publishing, 2019.

[11] S. Ludin, J. Garza, "Learning HTTP/2: A Practical Guide for Beginners", O'Reilly Media, 2017.

[12] W. Odom, "CCNA 200-301 Official Cert Guide Library", Cisco Press, 2019.

The 12th International Conference on Business Information Security (BISEC-2021), 3rd December 2021, Belgrade, Serbia

DECENTRALIZED FILE STORAGE AND RANSOMWARE PROTECTION

STEFAN TAFKOV, ZLATOGOR MINCHEV

E-mails: stefan.tafkov@iict.bas.bg, zlatogor@bas.bg

IT for Security Department, Institute of ICT, Bulgarian Academy of Sciences

Abstract: *Over the past decade, storage technologies have evolved many times over. The existence of a global market economy involving the continuous migration of data flows has created many branches of data processing and operation systems. These events marked the beginning of the Model Core – a system for decentralized storage and processing of files and data. The model relies on two components – data encryption and block allocation. There are three stages in the distribution of the blocks – PDA (Portable Data Access), Data Mapping and Data Mining. One of the most important parts of the model is getting to know the system. This part is also made up of three stages – Intelligence, Mapping and Master Hold Stack Service. All the components are an integral part of the module for protection against ransomware attacks and external influences. In this way, the system can check in real time, at a client request, whether the file that the user uploads is not infected or damaged.*

Key words: PDA (Portable Data Access), Data Mapping, Data Mining, Mapping, Master Hold Stack Service

1. INTRODUCTION

Decentralization has proveven its self to be the best way to protect data from harms. Taking into account that many storage and protection systems work leads to the necessity of new data containing model and solution, the approach has been further accomplished.

The main goal of this research is to analyze alternative methods of storing data in the big data age. So, it would be suitable to provide a different way of decentralized storing of data (see Image 1), their security and sustainability in the dynamics of cyber threats [1]. The idea behind this study is to create a model to empower a system for decentralized storage (see Image 2), management, processing and assure a higher level of data security.

Thus, the advanced purpose of this study is not only the storage, but also the protection [2] of data from cyber attacks such as “ransomware” that continue to grow [3], [4].

Targeting a cyber attack to a system processing or storing data would result in a total failure of the specific system. The critical consequences in such a situation would be a breach of classified data, a collapse of the company’s authority, which in turn – would lead to a loss of trust among some of customers.

The presented article describes a model selected for the study, as well as specific algorithms that will be followed by the given model. Using Portable Data Access (PAD), the system tracks every movement of a file between the local machines. Every PAD use data encryption to secure the data if unwanted intruder tries to steal data using, e.g.: Trojan Horse, sniffing techniques or even ransomware attack targeting critical data, related to different data breaches [5].

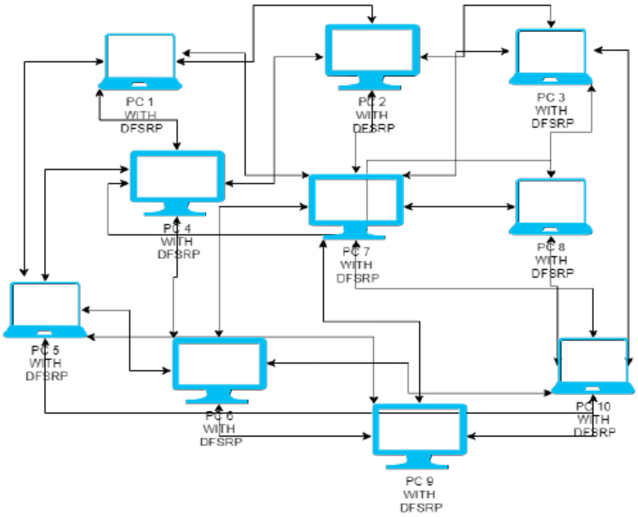


Image 1: Decentralized File Storage and Ransomware Protection (DFSRP) system idea

By exchanging data between all PC over a local network, using PAD described below the model take care of every file in the network and in the storage system. Additionaly, the study could be extended also in a distributed web accessed network, using the same multiagent based approach [6], but taking into account the drawbacks of relyable communication encryption, that normally diminish the storage intercommunication process [7].

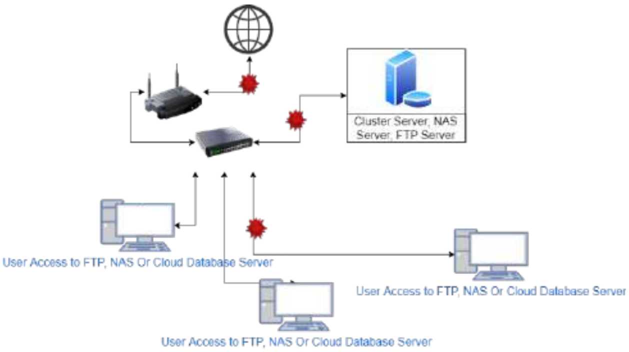

Cluster Server, NAS Server, FTP Server
User Access to FTP, NAS Or Cloud Database Server
User Access to FTP, NAS Or Cloud Database Server
User Access to FTP, NAS Or Cloud Database Server

Image 2: Base view of conventional Data Storage System with the risk of malware infection and data loss

2. CORE MODEL

The main core of the system for decentralized storage and processing of files and data is built on Portable Data Access (PDA, see Image 3). The PDA contains all the necessary modules for launching, operating and building an autonomous data storage system. For the purpose of this study, it was noteless that the model is built to have the ability for operating in any territory, i.e. – in a distributed manner. No matter the type of operating system, location or network equipment. This study aims precisely at developing a system that takes care of proper data storage, as well as providing a high level of security. Security from both cyber and physical attacks. Each PDA has the following sensors: "Local network monitoring sensor and every single PDA in it, file activity monitoring sensor, behavior monitoring sensor on PC, user activity monitoring sensor, process monitoring sensor operating in the system, system data collection sensor, active directory monitoring sensor". Using each of these sensors, the system was able to provide unwanted or unauthorized access to the blocks and constant monitoring of the overall structure, including the level of operation in real time.

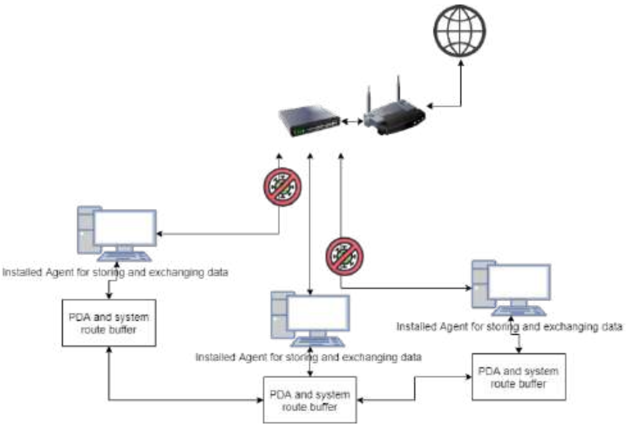

Installed Agent for storing and exchanging data
PDA and system route buffer
Installed Agent for storing and exchanging data
Installed Agent for storing and exchanging data
PDA and system route buffer
PDA and system route buffer

Image 3: View of PDA and serverless Data Store system, using only local network existing computer systems

(i) *Data Encryption.* Critical to the system is the realization of an encryption algorithm through which each PDA and the Data Blocks work [8]. The combination of AES-256 bit encryption and RSA with 4096 bit encryption was selected for the purpose of this study. By using this combination and integrating a module to recognize the level of encryption, the system can decide according to the user classifier, whether a file is secret or not. When the File is secret, the system uses both algorithms with the maximum level of encryption. When the file is not set as secret, the system switches to a lower level of encryption but also performs the operation faster [9].

(ii) *Block Allocation*. Multiple systems operating under a common cap – PDAs that work in sync by exchanging data with each other, using socket communication. Unlike conventional file storage systems, these pads have an integrated module for dispersing a file into multiple blocks that are distributed throughout the local network on which they are installed. After analysis and approval by the algorithm for distribution and fragmentation of data, the model for encryption and generation of metadata is launched, through which a system in a critical situation could restrict access to data in an organization. Using the distribution model, each file is divided into fragments of a size from 128 Kb to 5 Mb. Each of these blocks goes through encryption and further verification of data for greater security when they are transported to the local network. Unlike conventional storage systems where the user sends a file to a server, there is an emphasis on an algorithm (see Image 1) for dynamic data distribution between multiple PADs. As PAD, any single PC on a local network can be used. For the purposes of proper operation of the system there are build-in modules for analyzing the condition and capabilities.

(iii) *PDA (Portable Data Access).* Each of the PC systems on which this module is installed becomes a receiver and a data hub, both of the data that inhabits this machine and of all other data that will be stored within the local network in the future. By generating "meta tag" for the capacities of each object, the PDA [8], [9] takes care on the correct communication and data exchange between all other pads. This guartees mutual control of capacities and quality of work. When allocating data, each pad takes on the role of receiver until the installed pad does not consider that the PC cannot absorb more data. Then it is switched to unloading and distributing the data in the network.

(iv) *Data mapping.* At any given time, the system knows where to look for a certain snippet of a file, so that when a user requests, they can check the: integrity, identity, fidelity, security and mode of transmission for a minimum period of time. By building a compressed map [10] in each PDA, the system using socket communication exchanges (in an encrypted form) the location of each fragment of a file on the network. This achieves a large amount of memory and allocation of resource systems for partitioning into blocks of larger files.

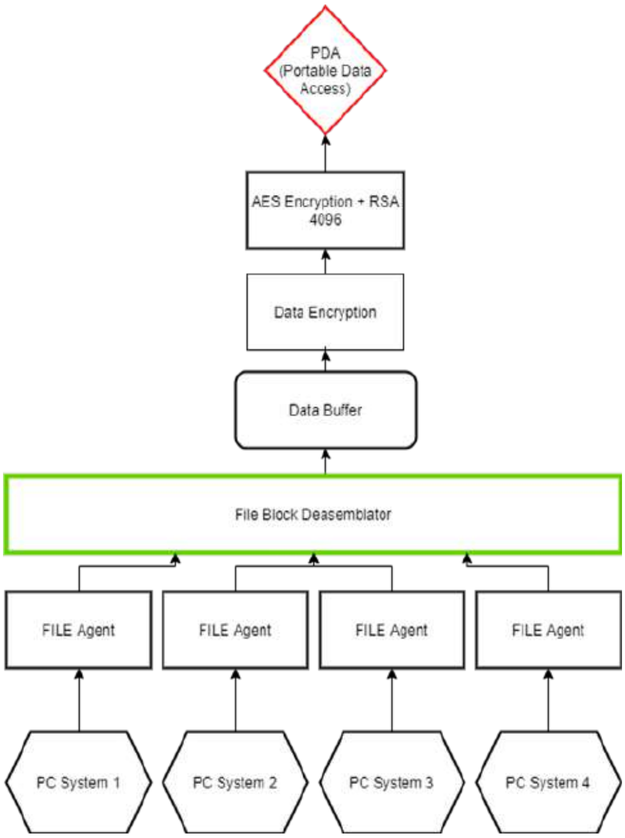


Image 4: Portable Data Access block logic model for translating file information into distributed blocks

(v) *Data mining*. Using an algorithm to correctly separate and place the data in a block container [11], the system provides each PDA with security level data, imlementing deployed sensors to detect anomalies. Also feedback from every single PDA to any system on the local network is accomplished.

3. GETTING DEEPER IN THE SYSTEM

An important part of the model's work is getting to know the computer system on which it will operate. Using a system mapping method to retrieve data on a system resource and load the data in the technical condition analysis model, the system can estimate to what percentage the resources could be used for the proper operation of critical conditions. The system resource analysis model is divided into three main stages.

(i) *Reconnaissance stage*. During reconnaissance or system capture, the model retains data on capacities, taking into account what percentage of them could be used. After analysis of the PCs, the model transmits in JSON format the data to the Mapping stage, Master Service Decision [12]. Because the system does not rely on a single machine for its operation, as with a number of other systems using servers and NAS (Network Attached Storage) devices, which allows multiple memory increases and CPU resource allocations required for operation. Several criteria have to be collected by the system during reconnaissance: "Disk space, Processing Power, LAN Map, RAM capacity and its frequency". After collecting these data, each machine generates in the form of a JSON format identification card that is used during its further operation.

(ii) *Mapping*. After capturing the data from a computer, the model switches to scanning the local network. Whilst looking for other PDAs that are ready for work. After the data synchronization of each available PDA is completed, the system switches to the Master Hold Stack Service distribution and broadcast stage.

(iii) *Master Hold Stack Service*. The purpose of the "MHSS" system is a qualitative distribution of data flows. By assigning the MHSS system, we have calculated that a certain PC has the best data buffering parameters. Any computer system on the local network can participate as gateway without limitation. This module is implemented due to the presence of quite old networks and computer systems that do not have enough power to distribute the blocks between themselves. This system in no way disrupts the operation of the involved computer systems.

4. RANSOMWARE ATTACK PROTECTION AND INTRUDERS

The main module in data protection – monitors their integrity and accessibility by downloading meta data from the FILE PE file structure. Thus, the real-time system, upon request by the client, can verify that the file that a user uploads is not infected or corrupted. Before moving to splitting the file between the PDA-details, the security module makes a meta map of the file that is uploaded. This meta card integrates into each block. Thus, if you need to transfer and assimilate a file, the system will know the state in which the user has transmitted the file. By analyzing each file every 24 hours or when there is a reduced data consumption in the system [7], the system switches to Health Check. In order accomplish an accurate verification, one to four PDAs are appointed to check the data for gaps and inconsistencies in metadata that accompany the blocks. After completion of this procedure, the system goes into standby mode, excluding any access by the PCs to the data, unless there is human intervention. Using a sensor monitoring of the user activity, the system can decide whether to turn off the file communication channel or put it in stand by mode. If the presence of a person is detected by monitoring a keyboard, mouse, camera or microphone, the system goes into operation mode, where a command by the user is expected.

5. CONCLUSION

Evidently, dynamic storage technologies provide many conviniences, good speed in performing daily tasks and duties. The study has presented a kind of alternative to conventional data storage systems, using decentralization. The problem with data storage comes from the low level of data protection and redundancy, as well as from the lack of a sufficiently powerful hard drives for storage. The current study highlights the potential of decentralized technologies and outlines a future vision for their security development.

The need for additional specialized storage hardware decreases in the presence of such a system that uses the resources of the available machines in the local network or largely via WWW, taking the necessary security precautions.

6. ACKNOWLEDGEMENTS

The obtained results in this study are partially supported by National Scientific Program “Information and Communication Technologies for a Single Digital Market in Science, Education and Security – ICTinSES”, 2018-2021, financed by the Ministry of Education and Science, Republic of Bulgaria. Additional gratitude for the technological and expert support is also given to the international forum initiative “Securing Digital Future 21” with more than sixty countries now, spread around the world, https://securedfuture21.org/

REFRENCES

[1] “ENISA THREAT LANDSCAPE 2021”, European Union Agency for Cybersecurity, October, 2021, https://bit.ly/3ox9Lae

[2] S. Tafkov, Z. Minchev, “Ransomware Detection & Neutralization System”, in Proc of X International Scientific Conference Hemus 2020 Research and investment in technology innovation – a crucial factor for defence and security, Defence Institute “Prof. Tsvetan Lazarov”, Bulgaria, 2021, pp. II-144-II-152, DOI:10.13140/RG.2.2.21029.12009

[3] “Ransomware in a Global Context”, Activity Report, VirusTotal, October, 2021, https://bit.ly/3ozpq8V

[4] “The State of Ransomware 2021”, SOPHOS White Paper, April, 2021, https://bit.ly/3ciZzwn

[5] “DBIR 2021 Data Breach Investigation Report”, Verizon, 2021, https://vz.to/32eukRv

[6] S. Russell, & P. Norvig, “Artificial Intelligence: A Modern Approach”, Pearson, 2021.

[7] P. De Filippi, “Blockchain and the Law: The Rule of Code”, Harvard University Press, 2018.

[8] D. Tapscott, “Blockchain Revolution: How the Technology Behind Bitcoin Is Changing Money”, Portfolio, 2018.

[9] M. Hameed, “Low Power Approach for Implementation of Huffman Coding: For High Data Compression”, Scholars’ Press, 2018

[10] C. McAnlis, A. Haecky, “Understanding Compression: Data Compression for Modern Developers”, O’Reilly Media, 2016.

[11] D. Drescher, “Blockchain Basics: A Non-Technical Introduction in 25 Steps”, Apress, 2017.

[12] A. Orebaugh, G. Ramirez, J. Beale, J. Wright, “Wireshark & Ethereal Network Protocol Analyzer Toolkit”, Syngress, 2007.

SMART CARDS TO IMPROVE SECURITY OF BIOMETRIC SYSTEMS

ANDREJA SAMČOVIĆ
University of Belgrade - Faculty of Transport and Traffic Engineering, andrej@sf.bg.ac.rs

Abstract: *Smart cards are portable secure devices designed to hold personal and service information for various applications. Examples of their use are banking cards, personal cards or cell phone user identification. Smart cards and biometrics can be used together in many applications. Being a secure portable device, smart card can be used for secure storing of biometric references of the cardholder, or to perform some biometric operations. This paper describes physical, logical and cryptographic security mechanisms provided by smart cards. Biometric systems are adressed from the privacy point of view. Once the technology and the security services are presented, their integration in biometric systems is presented.*

Keywords: *Smart cards, security, privacy, biometrics, potential vulnerabilities, attacks*

1. INTRODUCTION

Smart cards are portable secure devices designed to store personal and service information for a variety of applications. Examples of the use of smart cards are: mobile phone identification, bank credit/debit card, or ID (Identity) card. Smart cards and biometrics can be used together in different scenarios. Smart cards contain a microprocessor for data processing and three types of memory: RAM for temporary storage during processor operation, ROM (Read Only Memory) on which the operating system is located and EEPROM (Electrically Erasable Programmable Read Only Memory) for data storage of interest.

The main advantage that smart cards provide to a biometric system is the protection of biometric data and/or operations. This paper will address the problem of the various Potential Vulnerable Points (PVP) of a biometric system, in order to establish a basis for the ways in which smart cards can be incorporated into a biometric system. Several different architectures will be described later, which will enable the removal of PVPs from the biometric system.

2. SECURITY PROVIDED BY SMART CARDS

The main difference between smart cards and other identification technologies is the level of security they provide. This high level of security is achieved by security mechanisms built to avoid misuse or reverse engineering (i.e. physical security) in order to ensure the exchange of information and enable authentication of both terminals and cards (i.e. logical security) [1].

If all mechanisms are successfully integrated, for each of the attacks the smart card will react by rejecting them, not reacting or interrupting temporarily or permanently. For example, in the event of a number of consecutive incorrect PIN code entries, the card will block the code-protected information. If no mechanism for unblocking such code is defined, then access to information will be permanently blocked.

2.1. Physical security mechanisms of smart cards

It is important to emphasize that the advantages of using smart cards compared to, for example, centralized systems, are not in the use of cryptographic protocols (which can be used in both cases), but in the fact that smart cards are portable devices that are protected from unauthorized access. Devices that are protected from misuse offer resistance to their intentional alteration or forgery. In other words, they are designed to avoid any type of attack aimed at misusing the card, reverse engineering and/or accessing the contained information. This information can be cardholder information, application information, or even security keys and secret codes stored on the card.

The most reliable smart cards are made using microcontrollers resistant to unauthorized handling. To prevent an attacker from retrieving or modifying stored data, the chips are designed so that the information is not accessible through any external means, as it is only accessible through logical security mechanisms that will be listed later [2]. However, it is clear that the design and manufacture of such an interference-resistant microcontroller is extremely difficult, as attacks can be performed in many different ways (e.g. applying voltages and frequencies from certain ranges, freezing and/or overheating the chip, irradiating a certain zone microcontrollers, analyzing variations in time or power consumption related to input data, etc.).

The physical security mechanisms of microcontrollers for smart cards are different and are usually secret and private, but some of the basic ones are as follows [1]:

• Pseudo-random placement of cells in a chip. Instead of producing a chip that follows the traditional structure, all its components are divided into cells and then placed randomly in the substrate. This makes the interconnection less optimal, but acts as a defense against reverse engineering.

- Coded addressing of memory blocks. In addition to dividing the memory blocks and randomly placing those parts in the chip, the addressing of each memory position is coded, so that even within the same part of the memory, the information does not follow a logical structure.
- Encrypt information stored in non-volatile memory. Using the logical security mechanisms that will be explained below, a smart card can be manufactured to store all the data in the card in an encrypted way, which makes it even more difficult to search for any type of relationship between data cells.
- Attack detection via power variations (static and dynamic). Smart cards have a set of circuits at the power entry point, to detect variations in voltage values in both absolute value and frequency variation [3]. If the supply voltage does not meet the specifications, or the voltage variations have an unacceptable frequency, the smart card stops working, depending on the design, and is temporarily or permanently blocked.

All of these mechanisms are used to avoid direct attacks either to destroy the card, reverse engineering, or to access sensitive information within the card. However, there are indirect means of obtaining information, which in a few words statistically analyzes variations in energy consumption depending on the variation of the input data.

2.2. Logical security mechanisms within smart card technology

In addition to the physical security mechanisms described above, the smart card includes a range of logical mechanisms to protect access to data and exchange information with the outside world. These mechanisms are based on cryptographic algorithms and security rules [1].

Most smart cards on the market include cryptographic algorithms. These algorithms are generally symmetric, such as DES (Data Encryption Standard), Triple-DES, or AES (Advanced Encryption Standard). However, there are some products that are able to use asymmetric cryptography, such as RSA (Rivest-Shamir-Adleman) or ECC (Elliptic-Curve Cryptography) [1].

Of all the security services, the most interesting is the authentication of the actors involved in the exchange of information. The use of different ways of combining cryptographic algorithms, as well as storing different keys within one card, allows solving certain problems when trying to authenticate both ends in the exchange of information and/or commands [1].

Authentication mechanisms are based on the joint calculation of certain data using stored internal keys and random data used as a challenge [1]. If random data (challenge) is generated by the terminal and sent to the card, it is called internal authentication because the terminal confirms the authenticity of the card (since the terminal is the one that initiates the account with new input data, i.e. generates the challenge and verifies the success of the procedure). In this case, when the card receives a challenge, it uses internal keys and an algorithm to calculate the data originating from the challenge, and sends the result to the terminal. The terminal, which also needs to know the internal keys and algorithm, performs the same calculation and compares its result with the result sent by the card. If the result is the same, then the terminal is sure that the card is reliable.

Some cards also apply a mutual authentication mechanism that, in a few words, provides the same service as successfully performing internal and external authentication, but as a single operation. It is important to note that none of these mechanisms exchange any type of keys, keeping them securely stored in the terminal and on the card.

In addition to these authentication mechanisms, smart cards also provide another authentication mechanism that involves authenticating the cardholder. This is known as Card Holder Verification (CHV), although most people call it entering a PIN. Through this mechanism, the terminal asks the cardholder for a personal identification number (PIN). After entering the PIN, the terminal generates a CHV key, which is sent to the card. The card then compares the CHV key with the one stored on the card. If a mismatch is found, the number of erroneous comparisons increases, and if it reaches a certain value, the whole mechanism is blocked. On the other hand, if a match occurs, then the cardholder is authenticated and the incorrect comparison counter is reset.

2.3 Secure channel

Once the main logical security mechanisms are in place, the application of certain rules can provide the conditions for the exchange of information and commands between the card and the outside world. It can ensure all three security requirements: confidentiality, integrity and authenticity. When this is achieved, we can assume that a secure channel has been created for the use of the smart card. The principles for building a safe channel are as follows [1]:

- Using mutual authentication, both communication ends are authenticated, and therefore the authentication requirement is met. Moreover, during the process of mutual authentication, both ends of the communication know the same key, which can be used for other cryptographic mechanisms;
- Any cryptographic algorithm can be used as an algorithm to create summary data and/or commands that are exchanged, following the rules of those algorithms used to check integrity;
- Last but not least, with both authenticated ends, which know the same key and the same algorithm, and after adding the integrity field, all data involved in the information exchange can be encrypted, ensuring channel confidentiality.

If the key is the same for all cards, then a large-scale attack can be applied if the attacker can obtain a significant number of cards. As smart cards can have a limit on the number of consecutive key misuses, this type of attack, since theoretically possible, is not feasible in practice because the number of cards required would be so large that the manufacturer would detect that an attack is being performed. The manufacturer can react then in accordance with the threat and prevent the attack before it is realized.

The problem is that if an attacker accidentally discovers the key to one card, then he knows the key to all the cards, compromising the security of the entire system. To avoid this, each card has its own key, which is derived from a common basis, using some unique data from the card (e.g. serial number plus some immutable personal data of the cardholder). The terminal requests this unique information from the card and performs an algorithm to obtain the key for each card individually. This process is called key diversification.

2.4 Key management

When general ideas about security mechanisms are established, it is important to highlight one of the biggest potential vulnerabilities that these types of mechanisms can have: managing the keys involved in the process.

In the smart card industry, key management starts from the moment they are produced. In such a process, all cards are completely protected from the moment of production by using the various stored keys. Such a key is usually called a production key or personalized key. When these cards are sent to the issuer e.g. banks, they are sent without revealing the basis and algorithm for diversification of all related keys. Only when the issuer reports that all cards have been received (after counting them), the manufacturer sends the issuer the basis and algorithm for calculating the various keys. In case the number of received units is not the same as the one that is sent, the system reports an alarm, and the basis and algorithm will never be released, which prevents the use of cards forever.

3. BIOMETRICS SECURITY

The biggest problem when we talk about privacy is that it is not a universal concept, but depends a lot on the culture and society in which the citizen finds himself. Privacy can be defined as the ability of a human being (or group of people) to keep, his deeds and/or his information out of the reach of any other interest group within society. From the point of view of information, privacy can be understood as the right of a citizen to decide with whom he wants to share or not share a certain part of the data related to him / her. Given this definition, biometrics has a dual role in privacy [4]:

- Biometric data are part of personal data, and in that case biometric data will be protected like all other types of personal data, which are in the personal interest of the citizen about his privacy [5];

- Biometric data is a means of identification and can therefore be used as a confirmation of access to private data. In that case, the disclosure of biometric data may lead to the disclosure of all personal data protected by such biometric information, and therefore indirectly endangers the privacy of citizens. Therefore, depending on the way the biometric system is implemented, the consequences for citizens' privacy may vary.

If the biometric identification system is applied using a centralized database, citizens can reject the system because they feel that part of their personal data is kept out of their control. Therefore, they may consider following them, for example police, government agencies or even private companies. This is what is called the "big brother effect", which can cause concern among citizens, even without any evidence of illegal or illicit activity that supports the concern. Also, from a security point of view, if a centralized database is compromised, then the privacy of the entire population entered into such a database is compromised, creating a large-scale social problem.

A distributed architecture such as that based on smart cards will allow citizens to become aware that their personal data is stored in tokens that belong to them, without being stored in a central database that is beyond their control. Therefore, they will be in a position to know when, where and who requests to authenticate, as well as to be able to refuse the request. Moreover, given media reports about how hackers can access central databases to obtain personal information, such as bank accounts, citizens will prefer to have complete control over their personal data (especially those not expected) to change frequently), in a personal and secure device. From a security point of view, if the attack is successful, it will only endanger the privacy of a citizen, avoiding the emergence of a wider social problem.

Therefore, a distributed architecture is desirable for those systems that include personal information, and especially those that include permissions to access citizens' information. However, in addition to many advantages, the use of biometrics in a distributed manner can face several problems [6]. As it will be shown, if the distributed system is implemented using smart cards, based on the security they provide, many vulnerabilities will be removed [1].

3.1. Potential vulnerabilities

Within the biometric system, there are several vulnerabilities that need to be considered. Figure 1 shows the potential vulnerabilities that a biometric system could have. By removing vulnerabilities that may appear during the application, the application is considered to be made in a secure environment.

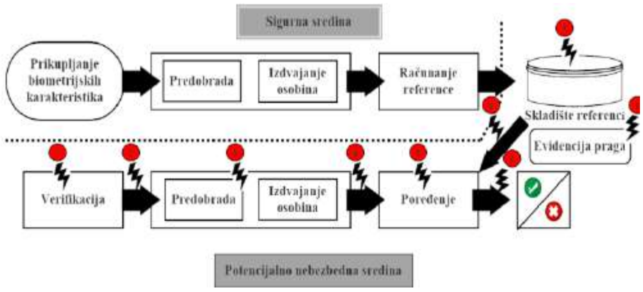


Figure 1: Potential vulnerabilities of the biometric system

The first PVP (Potential Vulnerable Points), PVP1 refers to user behavior at the input part of the system (recording device). Unfortunately, smart cards will never remove this PVP. Only by integrating anti-fraud mechanisms within the system, this PVP can be partially removed.

PVP2 is located at the output of the recording device, as well as at the input of the biometric algorithm. The sample may be intercepted and/or introduced to ensure re-attack. If this PVP can be exploited, it is of great concern because the validity of the identity may be called into question. Another type of attack on this PVP is "hill-climbing" attack by entering consecutive biometric samples based on the feedback that the biometric system supplies.

A large number of remaining PVPs (i.e., 3-8) are potential vulnerabilities that also exist in any IT (Information Technology) system. Hacking tools such as Trojans, viruses, Man-In-The-Middle attacks for further data entry, hill-climbing or re-attacks can be used to overload the identification process. PVP3 and PVP5 refer to the manipulation of algorithms, parameters and / or time data used for calculation. PVP4 and PVP6 represent the interception points of the data exchanged between the modules. Of these two, PVP6 is more sensitive as a template (or other type of biometric reference) that is included in the data, and which officially refers to the citizen. Last but not least, PVP7 and PVP8 address vulnerabilities in the storage of sensitive data such as biometric references, thresholds and logs.

Finally, even considering that PVP9 (i.e. exploiting the outcomes of the biometric identification process) is also a typical point of interest in any IT system. It is important that the system could provide information to the outside world. If the result of the comparison is not given only as an OK/ERROR message, but carries information about the match level, the attacker can use this information to create an artificial pattern using "hill-climbing" techniques.

These PVPs (some or most of them) can be removed using smart cards. To achieve this goal, several architectures can be proposed by the recommendation ISO/IEC 24787 [7]:

- Data store on card;
- Biometric comparison on card;
- Job sharing between the card and the biometric system;
- System on card.

3.2. Data store on card

The first idea of using smart cards to protect the biometric system is based on using a smart card to store biometric references of each citizen, the principle of which can be seen in Figure 2. This solution provides a number of advantages. Because smart cards are devices to protect against unauthorized access, they can securely store sensitive information that represents a user's biometric reference. Also, they can force the biometric system to establish a secure channel, after successful mutual authentication, in order to read the user’s biometric reference. In addition, the smart card can send a biometric reference to the system encrypted with a session key, to avoid any success of the "Man-in-the-Middle" attack. Therefore, it can be considered that PVP6 and PVP7 are already covered by this approach.

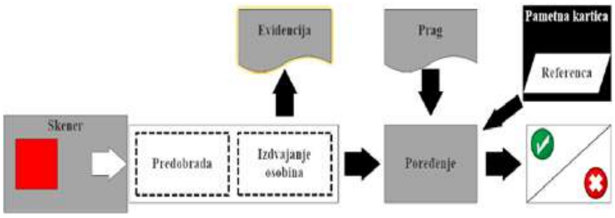


Figure 2: Architecture of data store on card

This solution is currently available and can be applied in any biometric system. It is only necessary to provide the card to users and to make the smart card reader available at all terminals. This can be applied to any system that uses biometrics or some other identification mechanism, such as passwords. This approach can be used with all existing models, with the following limitations [1]:

- Card memory capacity. Depending on the capacity of the card and the amount of information that needs to be placed in the card, some models may have implementation difficulties. Accordingly, several compact formats have been defined for data storage (for example, in the ISO/IEC 19794 series of standards). Currently, the memory required for storing raw data, such as a face, fingerprints, or iris, is small enough to accommodate more than one sample, or even more than one model.
- The amount of data transferred to and/or from the card in any verification attempt. Since the communication channel in a smart card is not as fast as in other technologies, the exchange of large amounts of data can slow down the whole authentication process.

3.3. Biometric comparison on cards

The previous approach does not provide a sufficient level of protection because the biometric reference is transferred to the terminal, which can be misused to collect citizen samples. In addition, comparisons and decisions are made in the biometric system and this can be misused by attackers. Last but not least, if the user is required to verify in order to activate a certain service or access certain information on the card, the biometric system will have to check and send an administrative key (which may also be called a CHV key because it will refer to the user’s identity, although the cardholder is not directly known), to gain access to such services or information.

The advantage of this solution is that, as is the case with smart card keys, the biometric reference never leaves the card, and successful verification can be used to allow access to internal data and/or card operations. Also, as in the previous case, the card may force the biometric system to establish a secure channel before sending the feature vector, or other biometric data, which will be sent to the card [1].
Abuse of biometric references, thresholds, algorithms or results is disabled in this case. Also, following the best practices in the smart card sector, the information that the card provides to the outside world will be minimized, in order to confirm successful comparison, unsuccessful result, or rejection of submitted biometric data (eg because security mechanisms are not met or quality thresholds are not reached).

The level of safety achieved by this approach is such that PVP 4-9 can be considered removed, as can be seen in Figure 3.

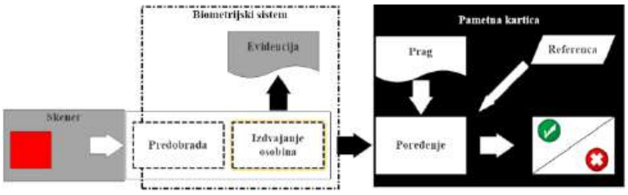


Figure 3: Biometric comparison on card architecture

This concept can be applied to several biometric models, although there are others that, in addition to the limitations shown in the card data storage section, may have a comparative algorithm with a high degree of complexity. For example, the comparison algorithms used today in face, voice, and signature biometrics do not allow them to be integrated, although smart card data processing capabilities may increase in the near future, and therefore allow the use of any of these models. There are currently a large number of commercial products that use the fingerprint technique.

It is also important to note that due to the low processing power of smart cards, compared to desktop computers, the performance achieved by using smart cards in biometric comparison algorithms may be lower than those obtained by an equivalent desktop solution.

3.4. Job sharing mechanism

ISO/IEC 24787 recommendation [7] defines a mechanism called job sharing for those products that are used for biometric comparison on the card, but where the computational needs of the comparison algorithm require that part of the algorithm be performed in a biometric system. This process should be performed without sending any part of the biometric reference to the system.

In this case, certain activities that are computationally intensive, are sent to the biometric system to perform the calculation. The result of the calculation is sent back to the smart card so that the final result of the match is calculated on the card, as well as an appropriate decision is made in terms of the defined threshold. During processing before comparison, communication takes place between the card and the biometric verification system. The final comparison should be made on the card. Figure 4 shows a block diagram of this architecture.

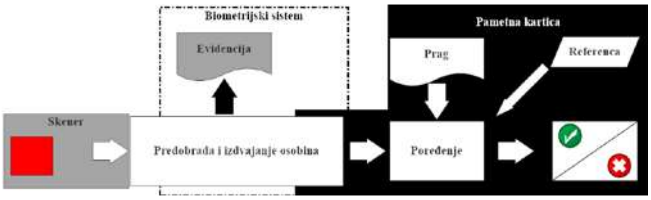


Figure 4: Job sharing mechanism for biometrics improvement

The same idea can be used to expand the role that the smart card plays in the process of citizen authentication and to reduce the PVP of the entire system. With this idea in mind, part of the pretreatment process and/or extraction function can be performed within the smart card, while the rest is performed in the biometric system.

Extending this architecture will complete all processing inside the card. In that case, the only part of the biometric system that is outside the card is the recording device. The advantage of this solution is that, based on standardized data formats (i.e. those defined in ISO/IEC 19794), the solution will be fully interoperable with third-party solutions [8]. With this approach, PVP 4-9 is considered secured, while some parts of PVP3 are still considered vulnerable [1].

3.5. System on card

The last approach comes as a supplement to the previous one. In this case, even a recording device (i.e. a biometric sensor) is built into the card. There are currently prototypes as well as products that implement this idea using fingerprints. An example of one such product can be seen in Figure 5.

With this architecture, the entire biometric system is included in the product used by the citizen, and he can communicate electronically with any type of application that requires biometric authentication for physical and/or logical access. As can be seen in Figure 5, all PVPs can be considered protected, although PVP1 can never be considered fully resolved.

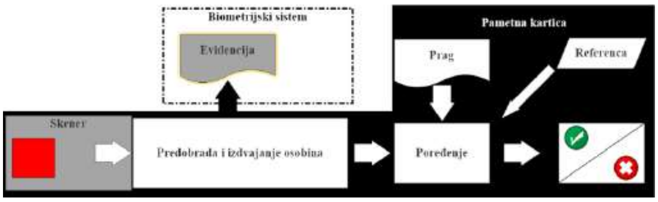


Figure 5: Integrated biometric system on card

The main disadvantage of this approach is that currently this idea is only applicable to fingerprints. Voice biometrics could also benefit from this approach due to the existence of small microphones, although processing could be too demanding. However, it is difficult to think that in the near future an iris or vein sensor could be built into a plastic identification card using the ISO/IEC 7816-4 standard [9].

On the contrary, the same idea, without considering the physical limitations of a plastic card, can be built into other types of devices, such as a USB token. This can allow the same architecture to be extended to other biometric models in the short term. The only condition that should be focused on is that such a token is fully compliant with all security mechanisms defined by the smart card industry.

It is important to mention that the most common misconceptions are present among users of technology, such as biometric smart cards. Some of them are as follows [1]:

- Fingerprint can be easily falsified - not entirely true, a fingerprint reader cannot be fooled by using a 2D fingerprint image, but misuse is possible where the biometric fingerprint technique is explained;
- Cardholder fingerprint data may be available to others - not true, as user fingerprint information is stored on the card. This information never leaves the card and is therefore not available to others;
- The card must be charged in order to be used.

4. CONCLUSION

In a world whose determinants are the global economy and the internet, the issue of daily identification of people is crucial for information security point of view. It is known that biometric identification systems, due to the nature of work, always bring a dose of uncertainty into the accuracy of the results. In some important social activities, for example in the field of finance, the accuracy of the conclusions of some uni-modal biometric systems becomes questionable, because individual omissions can cause incalculable damage to an individual or community. Although no information security system is completely secure, biometrics is the best solution currently available.

The reason why the topic of smart cards is treated in this paper as an example of solving the problem, are surveys that indicate that the largest part of the total number of frauds occurs during automated payment by card. Another important factor is the acceptance of a certain method of storing and collecting personal biometric information, where the principle of a biometric card is the most acceptable way. This type of technology is becoming more and more accepted among users thanks to a large number of devices that have biometric feature readers. One of these devices is a smart phone that today has very advanced biometric sensors and has significantly influenced this technology to be increasingly accepted in other spheres of life.

Biometrics is becoming more and more present in the modern world and integrated into various objects. The use of biometrics is growing from year to year, making it a very current topic. For that reason, biometric systems will be relevant for many years to come and provide users with important information on how a complex system works.

REFERENCES

[1] R. Sanchez-Reillo, R. Alonso-Moreno, J. Liu-Jimenez: “Smart Cards to Enhance Security and Privacy in Biometrics”, in *Security and Privacy in Biometrics*, Ed. P. Campisi, pp. 239–274, Springer, 2013.

[2] A. Samčović, S. Čičević, M. Nešić: “Bezbednosni aspekti elektronskih ličnih identifikacionih dokumenata nove generacije”, Synthesis 2015 - *International scientific conference of IT and business-related research*, Beograd, april 2015.

[3] R. Sanchez-Reillo: “Achieving security in integrated circuit card applications: reality or desire?”, *IEEE Aerospace and Electronic Systems Magazine*, Vol. 17, No. 6, pp. 4-8, 2002.

[4] A. Jain, A.A. Ross, K. Nandakumar: “*Introduction to Biometrics*”, Springer, New York, 2011.

[5] A. Samčović: “Biometrija i forenzika u digitalnom dobu”, *Business and Information Security Conference BISEC 2015*, Beograd, str. 58–63, 17. jun 2015.

[6] C. Michelle, B. A. Frye, *The body as a password: considerations, uses, and concerns of biometric technologies*, Washington DC, USA April 2001.

[7] ISO/IEC 24787: Information technology — Identification cards - On-card biometric comparison, 2018.

[8] ISO/IEC 19794: Information technology — Biometric data interchange formats. 2011.

[9] ISO/IEC 7816-4:Identification cards - Integrated circuit cards, 2020.

PROTECTION OF PERSONAL DATA ON THE EXAMPLE OF SMART CONTRACTS

SINIŠA DOMAZET

Educons University, Faculty of Security Studies, sdomazetns@gmail.com,

ORCID iD: https://orcid.org/0000-0002-5964-2249

Abstract: *The paper analyzes the concept, basic characteristics, as well as challenges in the application of smart contracts, in order to indicate their possible application in practice. It has been established that there are numerous definitions of smart contracts, some of which are technical and others of a legal nature. Also, research has shown that there are certain difficulties in the implementation of smart contracts, given their automatism, self-enforceability and irreplaceability. This also applies to the protection of personal data. It seems that most of the problems can be avoided by establishing close cooperation between the contracting parties, legal experts and developers, but also with the prior drafting of a written text of the agreement. The author was guided by two methods: the positive-legal method, as well as the legal-logical method of induction and deduction.*

Key words**:** *Law, Smart Contracts, Security, Data protection, GDPR*

1. INTRODUCTION

The protection of personal data is an increasingly important issue in the 21st century, given the accelerated development of information and communication technologies, the development of artificial intelligence, the Internet of Things (IoT), as well as the so-called smart cities. Regulators around the world are trying to keep up with the development of science and technology, that is, to protect privacy and personal data as much as possible, but this does not always seem to be the case. Numerous problematic cases that occur in practice around the world, the violation of privacy at the expense of the use of high technology and smart devices testify to the need to make additional efforts in this regard. Smart contracts represent a kind of revolution in the field of contract law and are based on blockchain technology. Numerous definitions of smart contracts (technical or legal) testify to their complexity. Given the characteristics of smart contracts that relate to their automatism, self-enforceability and irreplaceability, numerous dilemmas arise in connection with them. The paper will analyze the dilemma regarding the protection of personal data. First, different definitions of smart contracts will be analyzed, and then attention will be focused on the protection of personal data.

2. DEFINITION OF SMART CONTRACTS

Smart contracts were first proposed by Nick Sabo, a computer scientist, lawyer and cryptographer, in the early 1990s. Its primary goal was to enable the conclusion and execution of contracts without intermediaries and without proving the reliability of the contracting parties. At the same time, he gave a definition of smart contracts, considering them a computerized transaction protocol that executes contractual clauses. In addition to the above definition, there are numerous other definitions of smart contracts, of which we will (due to the number) mention only some. Thus, according to some authors, one of the canonical definitions of a smart contract is that it is a contract as a combination of security protocols with user interfaces (the interface is the place of computer / machine interaction with man), in order to formalize and secure relationships in a computer network [1], then a special protocol intended to contribute to, verify or conduct the negotiation or execution of the contract without the interference of third parties in a traceable and irreversible manner [2], or as a set of promises, including protocols for their fulfillment [3]. There are also definitions that define smart contracts as computer code created to automatically perform contractual obligations after a "trigger" event occurs [4], or agreements in which execution is automated, usually by means of a computer program [5]. Finally, it would be useful to mention the definition that defines smart contracts as a type of computer code that is managed by a computer and that is self-executing and self-executing [6], and the definition given by the UK expert group in a special report, which refers to smart contracts as contracts whose terms are written in computer language instead of legal language, is also interesting [7].

It should be mentioned that the definitions of smart contracts are also contained in the positive legal regulations of certain countries. Thus, in the United States, many federal states have enacted their own regulations regarding blockchain technologies, while giving a kind of definition of smart contracts. For example, in the federal state of Arizona, a regulation was passed stating that smart contracts can exist in trade. A contract relating to a transaction may not be denied legal effect, validity or enforceability merely because that contract contains the term smart contract [8].

A similar regulation was passed in the state of Tennessee, where it is stated that a smart contract is an event-driven computer program that is executed on an electronic, distributed, decentralized, shared and replicated record used to automate transactions, including, but not limited to, transactions that : (a) assume protection and transfer the transfer of assets to that register; (b) create and distribute electronic assets; (c) synchronize information; or (d) manage the identity and user access to the software application. A cryptographic signature that is generated and stored using distributed log technology is considered to be in electronic form and an electronic signature. A record or contract that is secured by distributed record technology is considered to be in electronic form and is an electronic record. Smart contracts can exist in a store. No contract relating to a transaction shall be denied legal effect, validity, or enforceability merely because that contract is enforced through a smart contract [9].

Of the European countries, we can mention Italy, which in 2019 passed the appropriate regulation regarding smart contracts. In the new Italian law, smart contracts are defined as software based on DLT (Distributed Ledger Technologies), which, once the appropriate entry in the register has been confirmed, automatically approves the relevant terms agreed between two or more parties. Smart contracts are considered by law to be equivalent for certain purposes (ie the formation of consent and probative value) to traditional written contracts to the extent that the digital authentication of the parties is made in accordance with a procedure to be determined by the Digital Italy Agency [10].

Overall, numerous definitions of smart contracts lead to the conclusion that some of them are purely technical in nature, with smart contracts being linked to autonomous code running on a blockchain, while on the other hand smart contracts are equated with written contracts that are only “copied”. “In program code. The idea that it is possible to use modern technologies to ensure the application of legal norms, ie the implementation of contractual provisions is not new. Moreover, in the initial stages of the development of the Internet, too much faith has been placed in technological development, which could eliminate all difficulties related to the specifics of the legal system. If such a simplified understanding were accepted that a smart contract is in fact the embodiment of a written contract, it would lead to a naive understanding that it is possible to mechanically convert all legal norms into program code. In terms of the division of smart contracts, "pure" smart contracts stand out, which are limited to the virtual, dematerialized world. They do not cause changes in the physical world, nor do they require the recognition of the legal order in order to act. Their expansion came with the development of bitcoin, because they enable the exchange of cryptocurrencies and digital goods [11]. The literature also mentions "hybrid" smart contracts, which are classic contracts, coded in whole or in part, for the purpose of automation. Although originally intended for the virtual world, smart contracts now operate more widely (for example, a smart contract denies entry to a hotel room via a code lock; interrupts the car engine) [12]. According to some authors, most problems occur during the contact between the two worlds. With the so-called Internet of Things, hybrid contracts are expanding. The fridge orders food when left empty. A smart vehicle insurance contract determines the premium based on the mileage (pay as you drive) clause. More aggressive drivers pay more, and vice versa. The car has a blockchain interface, and the smart contract determines the amount of the premium according to the driving parameters [13].

Smart contract code is used for simple transactions. For example: a retail chain issues promotional coupons to a multitude of customers. The smart contract code verifies whether the promotional coupon has already been used. "Smart legal contract" is a fusion of the above category and classic contracts. Example: a real estate lease agreement has been concluded on an online platform. Every 30 days, the smart contract pays the rent to the landlord; if the tenant is illiquid — locks the door. Protects the buyer by preventing the landlord from raising the rent or issuing it to a third party [3].

Having in mind the concept, general characteristics and division of smart contracts, the question of personal data protection arises. In the text that follows, the regulations of the European Union will be analyzed, with a special emphasis on the Regulation on the protection of personal data.

3. SMART CONTRACTS AND GDPR

For smart contracts, it is especially important Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (hereinafter: GDPR).

In accordance with Article 22 and Rec.71 of GDPR, the data subject should have the right not to be subject to a decision, which may include a measure, evaluating personal aspects relating to him or her which is based solely on automated processing and which produces legal effects concerning him or her or similarly significantly affects him or her, such as automatic refusal of an online credit application or e-recruiting practices without any human intervention. Such processing includes ‘profiling’ that consists of any form of automated processing of personal data evaluating the personal aspects relating to a natural person, in particular to analyse or predict aspects concerning the data subject's performance at work, economic situation, health, personal preferences or interests, reliability or behaviour, location or movements, where it produces legal effects concerning him or her or similarly significantly affects him or her. However, decision-making based on such processing, including profiling, should be allowed where expressly authorised by Union or Member State law to which the controller is subject, including for fraud and tax-evasion monitoring and prevention purposes conducted in accordance with the regulations, standards and recommendations of Union institutions or national oversight bodies and to ensure the security and reliability of a service provided by the controller, or necessary for the entering or performance of a contract between the data subject and a controller, or when the data subject has given his or her explicit consent. In any case, such processing should be subject to suitable safeguards, which should include specific information to the data subject and the right to obtain human intervention, to express his or her point of view, to obtain an explanation of the decision reached after such assessment and to challenge the decision. Such measure should not concern a child.

Further, in rec. 71 GDPR states that in order to ensure fair and transparent processing in respect of the data subject, taking into account the specific circumstances and context in which the personal data are processed, the controller should use appropriate mathematical or statistical procedures for the profiling, implement technical and organisational measures appropriate to ensure, in particular, that factors which result in inaccuracies in personal data are corrected and the risk of errors is minimised, secure personal data in a manner that takes account of the potential risks involved for the interests and rights of the data subject and that prevents, inter alia, discriminatory effects on natural persons on the basis of racial or ethnic origin, political opinion, religion or beliefs, trade union membership, genetic or health status or sexual orientation, or that result in measures having such an effect. Automated decision-making and profiling based on special categories of personal data should be allowed only under specific conditions.

Having in mind the above, some authors state that we may accordingly conclude that smart contracts are at least in some circumstances caught by Article 22(1) GDPR. As a consequence, the Regulation’s qualified prohibition of automated data processing applies, however only where automated processing produces legal or otherwise significant effects on the data subject [14].

In accordance with article 22(2) GDPR, this will not apply if the decision: (a) is necessary for entering into, or performance of, a contract between the data subject and a data controller; (b) is authorised by Union or Member State law to which the controller is subject and which also lays down suitable measures to safeguard the data subject's rights and freedoms and legitimate interests; or (c) is based on the data subject's explicit consent.

Further, in the cases referred to in points (a) and (c), the data controller shall implement suitable measures to safeguard the data subject's rights and freedoms and legitimate interests, at least the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision. Decisions referred to in paragraph 2 shall not be based on special categories of personal data (personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation), unless point (a) or (g) of Article 9(2) GDPR applies and suitable measures to safeguard the data subject's rights and freedoms and legitimate interests are in place.

The second, with regard to the protection of personal data, special attention should be paid to two rights: the right to rectification and the right to erasure (‘right to be forgotten’). According to GDPR (Article 16), the data subject shall have the right to obtain from the controller without undue delay the rectification of inaccurate personal data concerning him or her. Taking into account the purposes of the processing, the data subject shall have the right to have incomplete personal data completed, including by means of providing a supplementary statement.

On the other hand, Article 17 of the GDPR contains provisions on the right to erasure (‘right to be forgotten’). The data subject shall have the right to obtain from the controller the erasure of personal data concerning him or her without undue delay and the controller shall have the obligation to erase personal data without undue delay where one of the following grounds applies: a) the personal data are no longer necessary in relation to the purposes for which they were collected or otherwise processed; b) the data subject withdraws consent on which the processing is based, and where there is no other legal ground for the processing; c) the data subject objects to the processing and there are no overriding legitimate grounds for the processing, or the data subject objects to the processing pursuant to Article 21(2); d) the personal data have been unlawfully processed; e) the personal data have to be erased for compliance with a legal obligation in Union or Member State law to which the controller is subject; f) the personal data have been collected in relation to the offer of information society services. According to the same article, where the controller has made the personal data public and is obliged pursuant to paragraph 1 to erase the personal data, the controller, taking account of available technology and the cost of implementation, shall take reasonable steps, including technical measures, to inform controllers which are processing the personal data that the data subject has requested the erasure by such controllers of any links to, or copy or replication of, those personal data.

The dilemma is how to ensure the implementation of the right to correction and the right to forget. The data subject may request the deletion of personal data from the smart contract, but how to implement it, for example, when buying and selling cryptocurrencies. In that situation, the identity of the contracting parties is unknown. Even if this problem were to be solved, the dilemma remains as to whether the personal data was indeed deleted.

The third, according to Article 5. GDPR, personal data shall be adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed (‘data minimisation’) and processed in a manner that ensures appropriate security of the personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organisational measures (‘integrity and confidentiality’). Thereby, the controller shall be responsible for, and be able to demonstrate compliance with this rule (‘accountability’). These requirements must also be met in smart contracts.

The fourth, according to Rec. 26. GDPR, the principles of data protection should apply to any information concerning an identified or identifiable natural person. Personal data which have undergone pseudonymisation, which could be attributed to a natural person by the use of additional information should be considered to be information on an identifiable natural person. To determine whether a natural person is identifiable, account should be taken of all the means reasonably likely to be used, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly. To ascertain whether means are reasonably likely to be used to identify the natural person, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments. The principles of data protection should therefore not apply to anonymous information, namely information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable. This Regulation does not therefore concern the processing of such anonymous information, including for statistical or research purposes.

For smart contracts are also important Recitals 28 and 29 GDPR. According to Rec. 28 and 29 GDPR, the application of pseudonymisation to personal data can reduce the risks to the data subjects concerned and help controllers and processors to meet their data-protection obligations. In order to create incentives to apply pseudonymisation when processing personal data, measures of pseudonymisation should, whilst allowing general analysis, be possible within the same controller when that controller has taken technical and organizational measures necessary to ensure, for the processing concerned, that this Regulation is implemented, and that additional information for attributing the personal data to a specific data subject is kept separately. The controller processing the personal data should indicate the authorised persons within the same controller.

4. CONCLUSION

Based on the above, it can be concluded that smart contracts represent a great novelty in legal science and practice, all thanks to the emergence and development of blockchain technology. Blockchain technologies are something that has great potential and it can be rightly argued that their application will be increasingly important in the future. This is especially because blockchain technologies instill security in users, especially when we talk about cryptocurrencies and their turnover. Of course, blockchain technologies, ie smart contracts, should not be equated with cryptocurrencies (which, it often seems), but they have a much broader purpose.

Thus, smart contracts can be applied in a variety of areas, which in addition to (currently) mostly financial transactions, can be applied in the future in online commerce, in relation to copyright, leasing, credit arrangements or vehicle leasing, issues that are related to inheritance, areas of management, and more and more often there is talk about the application in the field of health.

However, regardless of the many positive characteristics of smart contracts, it should be noted that during their creation and implementation, certain problems arise that are not only technical, but also legal in nature. These problems are, above all, related to the difficulties in their execution. It is self-enforceability, automatism and immutability that adorn smart contracts that are, at least at this moment, its main shortcomings.

It turns out that there are certain specifics in contractual relations that are not easy to transfer to the program code. When converting a classic contract into program code, ie converting it into a smart contract, it is necessary to incorporate the true will of the contracting parties, ie what they had in mind when concluding the contract. It is the transfer of that "spirit" of the contract into the program code that is one of the most difficult things that can happen to programming experts. This is especially true for the language of the contract.

Ordinary language is not the same as legal language, which abounds in very specific terms of a different scope of meaning, which in practice can create a number of problems for the contracting parties, and ultimately this can lead to litigation. Some other dilemmas regarding the application of smart contracts in practice relate to the identity of the contracting parties, the consent of the contracting parties when concluding the contract, the legal capacity of the contracting parties, the form of the contract, interpretation of contractual provisions, legal gaps, inability to fulfill the contract. consumer rights, problems with different jurisdictions, because the parties to a smart contract may be nationals of different countries, and the like.

All of the above is directly related to the protection of personal data. With regard to the GDPR (in case this act applies) for smart contracts, the principles of personal data processing will be important, with special emphasis on the principle of minimization, provisions on pseudonymization, anonymization, as well as the right to rectification and the right to erasure ('right to be forgotten').

Most of the problems related to the protection of personal data can be solved with the help of various data encryption techniques, which would contribute to the security of the contracting parties in smart contracts. Also, it is necessary to ensure adequate storage of personal data outside the blockchain, which requires the implementation of appropriate administrative, technical and organizational measures by the controller and data processor.

It seems that most of these problems can be avoided by establishing close cooperation between the contracting parties, legal experts and developers, but also with the prior drafting of a written text of the agreement.

Finally, at this moment, the legal regulation of the use of blockchain technologies, ie smart contracts, is still in its infancy. It should not be expected that states will simply relinquish their powers in relation to the issuance of money, as evidenced by the fact that some states have already started issuing their cryptocurrencies, or are preparing to do so, while some have even banned cryptocurrencies as a means payments, referring to their abuses (above all, in connection with money laundering). Special emphasis should be placed on the use of blockchain technology in sensitive areas, such as health care, where special emphasis should be placed on the protection of personal data. Therefore, in the future, the use of smart contracts should be expected to be legally regulated not only at the national, but also at the international level.

Thus, the general conclusion is that the rules on privacy and protection of personal data (whether they are contained in the GDPR or some other legal act) must be taken into account when creating and implementing smart contracts.

REFERENCES

[1] P. Cvetković, „Blokčejn kao pravni fenomen: uvodna razmatranja“, Zbornik radova Pravnog fakulteta u Nišu, br. 87, Vol. 59, pp. 127-144, 2020.

[2] S. Bourque, S. Fung, L. Tsui, “A Lawyer´s Introduction to Smart Contracts“, Scientia Nobilitat: Reviewed Legal Studies, Lask, Poland, pp. 4-24, 2014.

[3] M. Cvetković, „Smart ugovori: revolucija ili komplikacija?“, Zbornik radova Pravnog fakulteta u Nišu, Vol. 58, br. 85, pp. 225-242, 2019.

[4] P. Paech, “The Governance of Blockchain Financial Networks”, Modern Law Review, Vol. 80, No. 6, pp. 1082., Nov. 2017.

[5] M. Raskin, „The law and legality of smart contracts“, Georgetown Technology Review, Vol. 1, pp. 305-326, 2017.

[6] E. Mik, „Smart contracts: terminology, technical limitations and real world complexity“, Journal of Law, Innovation and Technology, Vol. 9, pp. 269-300, Oct. 2017.

[7] Government Office for Science, Distributed Ledger Technology: beyond block chain, London, 2016, Internet, available at: *https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/492972/gs-16-1-distributed-ledger-technology.pdf* , 08.10. 2021, 18.

[8] Revidirani Zakon Arizone, naslov 44, Trgovina i promet (Arizona Revised Statutes, Title 44 Trade and Commerce, § 44-7601C), Internet, available at: *https://www.azleg.gov/ars/44/07061.htm*, 08.10. 2021.

[9] Zakon iz Tenesija iz 2018. godine, Naslov 47 (Trgovinski instrumenti i transakcije), Poglavlje 10. (Tennessee Code, Title 47 (Commercial Instruments and Transactions), Chapter 10, § 201 as amended by Senate Bill 1662), Internet, available at: *https://legiscan.com/TN/text/SB1662/2017*, 08.10.2021.

[10] F. Squerzoni, V. Trombetti, M. Lombardi, M. Frattini, H. Territt, S. Obie, Blockchain And Smart Contracts: Italy First To Recognize An Overarching Legal Foundation, Internet, available at: file:///C:/Users/S&B/AppData/Local/Temp/Blockchain%20and%20Smart%20Contracts.pdf (11.11.2021)

[11] M. Cvetković, „Novčane obaveze i kriptovalute“, Zbornik radova Pravnog fakulteta u Nišu, Vol. 58, br. 81, pp. 119-138, 2018.

[12] E. Tjong, T. Tai,“ Force Majeure and Excuses in Smart Contracts“, European Review of private Law, No. 6, pp.787-804, 2019, in: M. Cvetković, „Smart ugovori: revolucija ili komplikacija?“, Zbornik radova Pravnog fakulteta u Nišu, Vol. 58, br. 85, pp. 225-242, 2019

[13] M. Durovic, A. Janssen, „The Formation of Blockchain-based Smart Contracts in the Light of Contract Law“, European Review of Private Law, Vol. 27, pp. 753-771, 2019.

[14] M. Finck, Smart contracts as a form of solely automated processing under the GDPR, International Data Privacy Law, Vol. 9, No. 2, pp. 83-84, 2019.

RESULT METRIC OF A FINGERPRINT SCANNERS DONE IN JAVA GUI APPLICATION

KOMLEN LALOVIĆ

Faculty of Project and Innovation Management Belgrade, Komlen.Lalovic@pmc.edu.rs

Abstract: *This paper presents novel comparative analysis between various fingerprint scans developed in java programming language, especially GUI (Graphical User Interface) application. It shows detail advantages and disadvantages of several scanners and provide science relevant data when it can be used and acquired to enable top level security in fingerprint biometrics.*

Keywords:

Biometrics, Fingerprint, Baby, Java, GUI.

1. INTRODUCTION

Biometrics is a scientific discipline and technology that measures and analyzes biological characteristics of humans. It is a part of advanced security systems widely used nowadays in modern society and protection systems.

Java programming language is one of the oldest OOP (object-oriented) programming languages based on all OOP concepts. One of java frameworks is also one of the oldest and stable for creating *GUI* applications. (*GUI* – Graphical User Interface).

Biometrics is scientific discipline and technology that measures and analyzes biological characteristics of people. It is a part of advanced security systems widely used in today's modern society and protection systems.

The highest persistence in Biometrics and the lowest possibility of interrupt data is fingerprint and minutiae, so this work will present those data done on humans.

2. TECHNOLOGY OVERVIEW

One of the good known OOP programming language which runs over 3.5 billion devices in the world is Java. It is a very strong and confident programming language which provides all OOP aspects of software development. In this application authors used swing framework for development in Java. [18] [19] [20] [21]

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. [6] [7] [8]

All this together make java ideal for realizing our application throw this specific one.

3. EQUATIONS

For this research we took two various scanners as it follows: optical and capacitive. Results are quite impressive and we will present it here now. [14] [15] [16] [17]

However, the java GUI application is quite simple and serves only to make easy presenting of those data and make it combine with other levels of usage. Figure 1 shows this IDE (Integrated Development Environment).

Figure 1

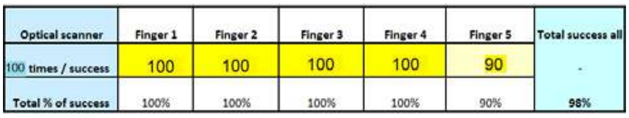
4. FINALLY THE RESEARCH RESULTS

Science fact, or rather an axiom, 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. [1] [2] [3] [4] [5]

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 whom were 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] [12] [13]

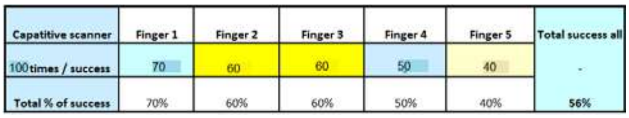
Results gained with optical scanner and 100 times tried each of a five fingers provided to us very high result of data acquired.



Optical scanner	Finger 1	Finger 2	Finger 3	Finger 4	Finger 5	Total success all
100 times / success	100	100	100	100	90	-
Total % of success	100%	100%	100%	100%	90%	98%

Table 1

After this very good result we have now other type of a scanner and it is capacitive. We got a bit different data result set. It can be seen in next table.



Capatitive scanner	Finger 1	Finger 2	Finger 3	Finger 4	Finger 5	Total success all
100times / success	70	60	60	50	40	-
Total % of success	70%	60%	60%	50%	40%	56%

Table 2

5. CONCLUSION

Each Biometrics tries to minimize **FAR**[1] and to maximize **FRR**[2] in attempt to be much more accurate and secure. This is case with fingerprint also, and role of various scanners is to provide it.

We could see that optical scanner is one class ahead especially when child or a baby fingers are used, because of a small tactile print and small ridges and valleys it is somehow expected.

REFERENCES

Books:

[1] 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

[2] 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):

[3] 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

[4] Komlen Lalović, Doctoral thesis "New system of identification newborn babies and parenthood guarantee based on Biometry", University of Singidunum, July 2016.

[5] 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.*

[6] 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*

[7] Before We Are Born, 9th Edition, Authors: Keith Moore, T.V.N. Peraud, Mark Torchia, Elsevier UK, Saunders, ISBN: 9780323313377, 2014

[8] NIST publishes compression guidance for fingerprint, Journal Elsevier - Biometric Technology Today, Volume 2014 Issue 4, April 2014, Pages 12

[9] Komlen Lalović, Ivan Tot, Aleksandra Arsić, Milan Škarić - Security Information System, Based on Fingerprint Biometrics, Acta Polytechnica Hungarica, Volume 16, Issue Number 5, 2019 DOI: 10.12700/APH.16.5.2019.5.6

[10] Komlen Lalović, Nemanja Maček, Milan Milosavljević, Mladen Veinović, Igor Franc, Jelena Lalović, Ivan Tot - Biometric Verification of Maternity and

[1] **FAR** – False Accept Rate

[1] **FRR** – False Reject Rate

Identity Switch Prevention in Maternity Wards, Acta Polytechnica Hungarica, Volume 13, Issue Number 13, 2016 DOI: 10.12700/APH.13.5.2016.5.4

[11] Hannah Grace Dahlen, Shea Caplice: “What do midwives fear?”, Published Online: July 24, 2014 – Elsevier, Women and Birth, Journal of Australian College of Midwives

[12] Komlen Lalović, Ivan Tot, Svetlana Andjelić - How to Guarantee Baby Identity based on Fingerprint Biometry, Bisec 2017 - International conference in Security ICT, October 18th-Belgrade, Serbia

[13] Komlen Lalović, Jasmina Nikolić, Ivan Tot, Žana Lalović - Software Algorithm of Device for biometric identification of Parenthood, BISEC 2016 - International conference in Security ICT, October 15th-Belgrade, Serbia

[14] Keith Moore, T. V. N. Peraud, Mark Torchia: Before We Are Born, Elsevier UK, Saunders, ISBN: 9780323313377, 2014, 9th Edition.

[15] NIST publishes compression guidance for fingerprint, Journal Elsevier - Biometric Technology Today, Volume 2014 Issue 4, April 2014, Pages 12

[16] Komlen Lalović, Patent Overview: Device for Fingerprint Identity Guarantee - Military Technical Courier, 2018, Vol. 66, Issue 2, http://dx.doi.org/10.5937/vojtehg66-15868

[17] Ivan Tot, Mladen Trikoš, Jovan Bajčetić, Komlen Lalović, Dušan Bogićević - Software Platform for Learning about Brain Wave Acquisition and Analysis, Acta Polytechnica Hungarica, Volume 18, Issue Number 3, 2021 DOI: 10.12700/APH.18.3.2021.3.8

[18] Komlen Lalović, book: Osnove java programiranja, Beograd 2020. Srbija, ISBN: 978-86-902148-0-8

[19] https://www.biometricupdate.com/202005/new-terminal-brings-contactless-iris-biometrics-to-unioncommunitys-ubio-line

[20] https://nordicapis.com/8-biometrics-apis-at-your-fingertips/

[21] Komlen Lalović, book: Java 2 programiranje i uvod u baze podatka kroz MySQL, Beograd 2021. Srbija, ISBN: 978-86-902148-0-8

EFFECTIVENESS OF NEW DIGITAL TECHNOLOGIES IN FIGHTING TERRORISM AND RELATED MONEY LAUNDERING

DRAGAN Ž. ĐURĐEVIĆ

Academy of National Security, Republic of Serbia, djurdjevic.dragan@gmail.com

MIROSLAV D. STEVANOVIĆ

Academy of National Security, Republic of Serbia, mstvnv297@gmail.com

Abstract: *In this article, we observe the use of computer models to detect and recognise the potential terrorist threats. Modern intelligence activities include the use of cutting-edge technologies to investigate terrorist activities. This involves efforts to identify and prevent potential participants to conduct recruiting, preparations logistics for conducting terrorist actions. The capabilities of new digital technologies resulted in broadening of available data to the level which makes it a challenge for human evaluation. Development of technologies to overcome the problem of processing the data has enabled defining, analysing and exploring ever more models. This led to the idea of computer experiments and simulations to achieve more complex planning and forecasting. The evaluation of the potential of fractal concepts to generate such models shows that the they rely on compression and data reduction with preservation of regularity, while analytic forecasting of missing visual data and forms is used to supplement the empirical evidences. The results indicate that the results can provide a similar models but depend on substantial prior knowledge and do not necessarily provide a reliable identification of various anomalies suitable to prevent terrorism and money laundering. We conclude that new digital technologies are a useful in the terrorism prevention as an additional tool in classic investigative methods.*

Keywords: *terrorism, money laundering, identification, fractal recognition, fractal compression*

1. INTRODUCTION

Information society intensifies globalisation and universalisation of values. In this framework, operative politics and decision making on national and global levels are faced with two challenges for globalisation and universalisation of values: terrorism and money laundering. Terrorism, as institutional political violence aimed to achieve fear, undermines the institutional order, while money laundering undermines the foundation of globalisation - the financial order. These challenges have propelled a need for a new approach to national security, which implies efforts to maintain the position of a state within global mechanisms. So, every country is engaged in combating terrorism (AT) and money laundering (AML) as transnational threats for "common values", even if they are not directly targeted.

Information age has enabled that agencies and institutions which are responsible for AT and AML at various phases have the capability to obtain and store vast number of data. Today, intelligence activities in preventing and combatting terrorism include financial investigations and money laundering for the purpose of financing terrorism, resulting in broadening of the scope of data to the level which makes it impossible for human logical evaluation.

Technologies development that provides increase in capacity of speed and quantity of data processing has enabled defining, analysing and exploring more and more models. This led to the idea of computer experiments and simulations aimed at complex planning and forecasting for the purpose of countering terrorism and “dirty” money transaction, as complex and variable phenomena.

From the aspect of information technologies, suitable for providing visual projection or formulas are fractals, as geometric shapes which are miniature copy of the whole and can be divided into parts. These transformation methods can be used, to translate, scale, shear, and rotate available data, creating an image. On the practical side, partitioned encoding of the total image will be smaller, while resembling its characteristics closely.

2. PERCEIVING TERRORISM AND MONEY LAUNDERING AS FRACTAL CONCEPTS

A fractal shape is geometry with scale. Applied on a phenomenon it would be its middle ground. Thus, when faced with observing an occurrence, which is based on a logic of growth and hierarchical scale (subsystems, networks), fractals should be suitable to obtain a model of such occurrence. The fact that fractals represent a middle ground implies that they necessarily involve a certain level of presumption. They, therefore, also provide the ability to develop an educated guess about a phenomenon (though artificial to a degree).

Fractals are fabricated by the repetition of a geometric over a succession of hierarchical scale. The result is an image that, on one side, describes the grid partitioning (the range blocks) and, on the other side, transforms per each range block. This method of achieving an image of a subsystem from reality involves five basic issues: (1) partitioning the image; (2) forming the set of blocks; (3) selecting type of transformations that will be considered; (4) selecting a distance metric between blocks; and (5) specifying a method for pairing blocks. On the practical level, this requires that the coding process should provide for as precise as possible measure and quantification of errors in the compressed image. This is obtained through series of presumptions. In practical application of fractal-based technology, a sequence of images converged to a sequence of subtly different faces [1], thus minimizing a measure of error may be of dubious value for recognition of reality.

In respect to error measures, the scaling is fabricated through the mentioned repetition of a geometric act over a succession of hierarchical scale. The obtained result is an axis of scale symmetry. In scaling terrorism or money laundering wider image, as shown on an example of signature reconstruction [2], coding requires a computable error measure that would capture impression of human viewers accurately.

Per the model of the Complex Adaptive System (CAS), an organisation which wants to cope with accelerated rate of change can rebuild itself as a system. This quality provides for its independent functioning, in terms of the possibility to exploit the existing patterns for immediate, independent solutions (from bottom to top) [3]. The CAS model imitates a targeted natural system established on a decentralized deployment of autonomous subsystems (fractals) acting independently to the environment. The effectiveness of the obtained model depends on two processes: firstly, on the capability of each fractal to independently carry out its assignment; and secondly, on the permanent integration of resources by transfer of relevant information among the fractals. The important feature from the aspect of digital technology, as a tool, is that the specific knowledge of one cell is functional for all the other cells (fractals). The reached result depends on five criteria: (1) definition of foci of system activities which have the potential for future growth; (2) relevant adaptation of organisational response to external events; (3) the integration between subsystems is brought about by dynamic work processes which enable coordination between the organisation and its subsystems; (4) decision making on local level; and (5) adaptation of the organisation’s ability to cope with changes in its environment, personal development and individual specialisation of each participant.

The way CAS operates has similarities with at least two patterns of terrorist organisations. Terrorist organisations function as integrative networks of independent cells, and they are perceived as organisations with the ability to initiate actions even while the elements of certainty in their environment are changing [4]. Thus, by the analogy to CAS, the use of fractals can help the researcher of organized network of terrorism or ML to overcome the following challenges: core competence; surrounding influence; recycling flow effect; decentralization; and synchronization of knowledge resources.

Since 9/11 attack in New York, the computational social science has recognised the practical use of information technologies in combating terrorism. Above all, it concerns the possibility of automation of visualising of data available on social networks, by their scaling or modelling. Computational social scientists argue that modelling and simulation are uniquely suited to understand the dynamics of emerging threats, at a time when national security decision makers are urgently looking for new frameworks, data sources and technologies for making sense. Computational social modelling and simulation, has the capacity to enhance understanding of a wide range of national security problems, including terrorism and terrorist network detection

3. THE USE OF FRACTAL TOOLS IN AT AND AML

Wide spectrum of mathematically based fractal concepts is used to generate computer models of terrorist or money laundering activities. The logistics behind the items connected with detecting and recognizing degree of threat involves comparing fractal structure of people’s faces, fast search through the databases of faces and fingerprints. In searching, of vital importance is the speed of processes, and the result depends on the ability of compression and reduction of data with preservation of regularity. Especially important, in addition to the empirical evidences and records, is the analytic forecasting of missing visual data and forms.

The operations of generating computer models by mathematically founded fractal concepts are enabled through the higher degree of utilization of knowledge and adaptation of virtual reality in prevention of terrorism and money laundering. So, achievements in implementation of the concept of fractals depends on substantial prior knowledge, environmental influences, subsystem integration, decentralization and synchronization. It allows obtaining a similar high information technology models, but not necessarily the identification of the authentic features of the various anomalies that result in objectively asocial consequences.

The development of computer technologies enabled wider scope of usefulness of fractals, as artistic field of mathematics (e.g. recreating the terrain, 3D iteration of parts), in creating of virtual space like natural forms. Also, since fractals can be divided in miniature representations of the whole, compression and reduction of data provide the opportunity for anticipation of the wider processes, for the analysis of patterns, which has the application in the imitation of networks. A fractal structure does not have regularities, and in practice it represents a supplement to experience and a higher level of use of knowledge and adaptability to reality. Computer forensics, which is aimed at collecting, identification, delivering and documenting computer for the need of a state, thus, also has a role in the work of intelligence services.

Available and incoming data and their elements can be evaluated from the aspect of their complex patterns of spatial distribution, providing a fractal dimension, which introduces a tendency towards a more linear, and thus one-dimensional, structure [5]. That makes it suitable for automatized analytical and synthetic operations, based on programmed criteria, which would otherwise require lengthy logical or forensic procedures. This method has found its role in the analysis of potential scenarios, as simulation, forensics of scripts, evaluation of behaviour, cognitive and psychological states (except of emotional inflections and nuances), excluding information security (private massaging, digital money, online services etc.).

In combating terrorism and money laundering, fractal tools can be used for at least the following intelligence goals:

- The most acknowledged is the analysis of terrorist (money launderers) social networks. This includes three types of input: information intelligence on intensive use of the internet to communicate; human intelligence about the contacts of suspected terrorists; and technical intelligence about contacts of suspect terrorists [6]. The research is basically related to the structure of interconnections and interactions, and the analysis of possible bases for terrorist operations, with the aim of identifying recruiting, training, organizing. The input for this type of IT researching terrorism and ML are links between participants. The available links may be impossible to overview, or to comprehend with logical sense, which can be overcome by the algorithm models which provide a dimensional layout, based on the strength of association, and crossings - centring points and their locations, possible subgroups and their dispersion. Based on types of relationship, the visualization is the output. The visualization is based on exclusion of most likely irrelevant data. Junctions with closest links can be filtered with fractal views because of stronger associations for those closer and more centred. This method provides a presentation suitable for observing and perception. the Model can be simplified, focusing on fractal values: a) probable core social structures of a network b) uncovering regional and indirect associations, and even hierarchical order of on organization. This application of fractals is appropriate for spatial analysis, which in practice includes identifying possible terrorist leaders and their locations, or possible cells. For visualization to be effective it should respect certain requirements: a) in order to fully utilize the space dimension, instead of a piled scale, the nodes should be separated by an optimal distance; b) to reflect the strength of association among linked end nodes, the length of a link should appear closer if they are strongly associated; c) in order that the user can clearly see the relationships between nodes, the crossing of edges should be minimized; and d) the importance of the corresponding terrorist (money launderer) should be represented through a proportional size of a node [7].
- Fractal tools can be used for the analysis of resilience to terrorist threat [8]. This includes the estimation of two aspects related to threat: measuring ability: to resist, to absorb, or to adapt, excluding ability to function under stress; and measuring possible impact of damage of terrorism, as a criterion for estimation. The result of the analysis of resilience to terrorist (or money laundering) threats is an objective assessment of cascade ability, which can be further used for modelling scenarios, and security requirements.
- Analysis of shape and structure features involves searching for missing details in relation to the identification of persons and objects [9].
- The intelligence goal can be the analysis of psychological and behavioural aspects. This is basically includes recognizing the temporal variabilities in actions of involved persons [10]. The analysis of psychological and behavioural aspects is practically datamining, since it involves active searching for understandable self-similarities to reproduce pattern of analogous behaviours [13].
- Fractal tools can be applied in the analysis of linguistic markers. In practice, this involves searching for weak signals on the Internet, as input, their collection and analysis [12]. The aim of this type of application is to select and identify isolated terrorists.

In general, fractal tools seem to be applicable as an artificial reality modelling and means of scaling, for the research of terrorism and money laundering networks, as phenomena which are characterized by non-equilibrium stability, fractal dimensionality, self-organized criticality, spontaneous self-organization, "typically observed at the edge of chaos" [13].

4. VALUE OF FRACTAL METHODS FOR AT AND AML

A fractal is a set of smaller subsets in some way similar to the larger set. The measure of its size and complexity is fractal dimension. A set is well-ordered if it can be linearly ordered in a way that every subset has a smallest element in that ordering. Well-ordering is equivalent to the axiom of choice - each implies the other. Natural shapes have fractality. One of its main characteristics, which is applicable in researching parameters of terrorism and money laundering is branching. That process seems chaotic, and without scalable geometry. All connections relating to causes and consequences cannot be seen in the structure and the achieved pictograph reproduces the essence of fractals - their similarity to the structure. This is obtained through a method which includes: automated input of elements of appearance and details; automated informatics contour of an occurrence; conscious contour of an occurrence; selection of various elements which connected by specific links and conditions of specific elements of the occurrence.

There is evidence against simplified or supposedly simplified approaches as a replacement for methods which achieve the best possible access to real life, language and philosophy [14]. Therefore, the researcher must apply culture and knowledge-source specific orientation. Redefining the boundaries between the different disciplines in the thought process includes reordering and reconnecting the ways of thinking outside of general.

Though there were pictures, there was no definition [15]. The concept of fractional dimension, or fractals, was developed to describe the shapes of natural objects. A fundamental property of fractal objects is that as a figure is magnified more details appear, but the basic shape remains [16]. Thus, when faced with rough data, strongly nonlinear, irregular, or displaying complex patterns that seem to defy statistical analysis, the fractal analysis offers a possibility to overcome these shortages and solve tasks, like identifying names, spaces, cases, or situations.

As a cognitive method, fractal tools provide graphs to facilitate overview of complex time and space characteristics, [17] saving time and manpower, and drastically increasing general investigating capacities. Fractal views facilitate seeking, in aim to recognize indicators of terrorism connected elements. The basic assumption in “war” against terrorism is that there is a compulsive repetition of cycle of violence. Since a pattern is fractal, combating includes efforts to objectify the preconditions and subjectify potential perpetrators. The obtained micro scale replicates a certain aspect of macro scale [18]. It resembles reality, and visualisation enables the reasonable assumption that the positioning of potential perpetrators is in real time [19]. As such, it has more value for orienting capabilities, future resources structuring, and concept exploration, i.e. high-level decision making, but less so where its impact is confined to perception of possible prevention, i.e. on the tactical level.

Fractal tools have proven applicable. The data about global terrorist events from 1970 to 2014. have been analysed by fractal dimension, that reflects the dynamics in time and space: firstly, the space–time characteristics from the available statistics; secondly, terrorism dynamics in a regional perspective; and thirdly, complementary analysis on multidimensional scaling and clustering techniques [20]. They also trace money laundering, which is most often consensual and map chains of service providers, personnel and organizations used. There is no evidence that they have advanced the prevention, regarding recruitment, or terrorist social activity (neither in money laundering), but they facilitate tactical initiative in general prevention, detection of associated activities and chains, and apprehending members [21].

Models and scales depend on algorithmic construction of coders for results in form of signals, pieces, transformability, complexity, and reduction. But, software companies promote the impression that tools using fractals will by itself allow a real-time risk monitoring system, to detect potential money laundering activity across its transactions. The obvious problem causing the divergence between capability and expectation stems from the fact that most software projects have various stakeholders, including developers, funders, and end users. In the software engineering, it is generally accepted that getting end users involved in the design and development of the tools they will use is critical if the software is to be usable, useful and relevant to real problems. Because the software projects are mostly commercial in nature, and begin, progress, and end without much consideration of who will use the software or what they will do with it, end users have little influence in these projects.

Computational models and simulations provide outputs, but predictions are forms of human judgments, and reflect the state of knowledge at a moment in time. Focusing attention on the limitations of models and simulations as humans’ tools, and investing what those limitations imply for decision making in the real-world, can advance the developing a broader understanding of how, where, when, and why computational models and simulations can be useful to people working in high-consequence decision making contexts [22]. Outside of that, similarly to application of approximative equation in highly automated high frequency trading, [23] fractal tools are not appropriate for precise results, but are useful for abstraction, approximation and reformulation [24].

5. CONCLUSION

In the information age, planners and practitioners who deal with analysis of space, asymmetric and irregular patterns of data cannot avoid being receptive to non-linear dynamic systems modelling, and fractal concept offers a range of suitable software opportunities.

In combating terrorism and money laundering, which appear as such patterns, the application of information technology on the practical level requires the implementation of concept of fractals, thereby introducing mathematical methods in scaling and modelling of potential social anomalies.

Even though, as the analysis demonstrate, there is computational prediction, IT models provide visual forecasting of missing data and supplement the empirical evidences and records, enabling higher degree of utilization of knowledge and adaptation in combating terrorism and various forms of money laundering. For this reason, application of information technologies based on the concept of fractals has its place in the analytic and strategic planning arsenal in combatting terrorism and money laundering.

REFERENCES

[1] M. Barnsley and J. Hutchinson, “New Methods in Fractal Imaging”, Proceedings of the International Conference on Computer Graphics, Imaging and Visualisation, July 26-28, 2006, Washington: IEEE Society, 296-301.

[2] M.B. Yilmaz, Offline Signature Verification With User-Based And Global Classifiers Of Local Features”, Ph.D. dissertation, Sabanc University, 2015. http://theses.eurasip.org/media/theses/documents/yilmaz-mustafa-berkay-offline-signature-verification-with-user-based-and-global-classifiers-of-local-features.pdf, Retrived 10. August 2020.

[3] J. Miller and S. Page, “Complex Adaptive Systems: An Introduction to Computational Models of Social Life”, New Yersey: Princeton University Press, 2007.

[4] R. Leary and J. Thomas, „How Complexity Theory is Changing the Role of Analysis in Law Enforcement and National Security“, in: Intelligence Management: Knowledge Driven Frameworks for Combating Terrorism and Organized Crime, A. Babak and S. Yates, Eds. London: Springer-Verlag, 2011, p. 63, pp. 61-78.

[5] G. McIntosh, M. Lauren, “Incorporating Fractal Concepts into Equations of Attrition for Military Conflicts”, in: OR, Defence and Security, R. Forder, Ed., Basingstoke: Palgrave Macmillan, 2015, p. 101-123.

[6] C. Yang, N. Liu and M. Sageman, “Analyzing theTerrorist Social Networks with Visualization Tools”, Intelligence and Security Informatics: Proceedings IEEE International Conference on Intelligence and Security Informatics, San Diego: ISI May 23-24, 2006, S. Mehrotra *et al.*, Eds., Heidelberg: Springer, 2006, pp 331-342.

[7] C. Yang and M. Sageman, ”Analysis of Terrorist Social Networks with Fractal Views”, in: SAGE Quantitative Research Methods: Volume 1 - Fundamental Issues in Quantitative Research, P. Vogt, Ed., London/Thousand Oaks/New Delhi/Singapore: Sage Publications, 2011, pp. 367-394.

[8] T. Lewis, “Critical Infrastructure Protection in Homeland Security: Defending a Networked Nation”, 2nd edition, New Jersey, John Wiley & Sons, 2015.

[9] S.M. Obaidullah, C. Halder, N. Das, K. Roy, ,,An Approach for Automatic Indic Script Identification from Handwritten Document Images”, in: Advanced Computing and Systems for Security, Volume 2, R. Chaki *et al.*, Eds., New Delhi: Springer, 2016, pp. 37-51.

[10] M. Wijnants, *et al.*, “Does Simple Rate Introduce an Artifact in Spectral Analysis of Continuous Processes”, in: Fractal Analyses: Statistical and Methodological Innovations And Best Practices, J. Holden *et al.* (eds.), Frontiers, 2003, pp. 56-68.

[11] T. Henderson and D. Boje, “Organizational Development and Change Theory: Managing Fractal Organizing Processes“, New York/London: Routledge, 2016.

[12] K. Cohen *et al.*, “Detecting Linguistic Markers for Radical Violence”, Social Media, Terrorism and Political Violence, 26:1/2014, p. 253. pp. 246-256.

[13] C. Mesjasz, „Complex Systems Studies and Terrorism“, in: Conflict and Complexity: Countering Terrorism, Insurgency, Ethnic and Regional Violence, V. Fellman, P. Bar-Yam and M.A. Yaneer, Eds., New York: Springer, 2015, p. 38. pp. 35-71.

[14] A. K. Bangura, “A Pluridisciplinary Treatise of the Fractal Complexity in John Mukum Mbaku’s Corruption in Africa: Causes, Consequences and Cleanups”, African Social Science Review, 7:1/2014.

[15] P. Davis, “Spirals: From Thoedorus to Chaos”, Wellesley: A. K. Peters, 1993.

[16] J. Russ, “The Image Processing Handbook”, 6th edition, Boca Ratton: CRC Press, 2011.

[17] B. Akhgar *et al.*, “Investigating Radicalized Individual Profiles through Fuzzy Cognitive Maps”, in: Emerging Trends in ICT Security, A. Babak and H. Arabnia, Eds., Waltham: Morgan Kaufmann, 2014, pp. 559-574.

[18] M. Blain, “Power, Discourse and Victimage Ritual in the War on Terror”, 2nd edition, Oxon/New York: Routledge, 2016.

[19] M. Barlow and R. Cox, “All Hazards Analysis; A Complex Perspective”, in “Applications of Information Systems to Homeland Security and Defense”, A. Hussein and D. Essam, Eds., Hershey/London: Idea Group, 2006, pp. 17-45.

[20] A. Lopes, J.T. Machado, and M. Mata, “Analysis of Global Terrorism Dynamics by Means of Entropy and State Space Portrait”, Nonlinear Dynamics, 85:3/2016, pp. 1547-1560.

[21] S. Strang, “Network Analysis in Criminal Intelligence”, in: Networks and Network Analysis for Defence and Security, A. Masys, Ed., Cham: Springer, 2014, pp. 1-26.

[22] L. McNamara, “Why Models Don’t Forecast”, Paper presented A Paper for the National Research Council’s “Unifying Social Frameworks” Workshop, Washington, DC, 16-17 August 2010, p. 23. Retrieved 07. June 2017.

http://sites.nationalacademies.org/cs/groups/dbassesite/documents/webpage/dbasse_071326.pdf

[23] I. Aldridge, “High-Frequency Trading: A Practical Guide to Algorithmic Strategies and Trading Systems”, 2nd edition, New Jersey: Wiley, 2013.

[24] L. Saitta and J.D. Zucker, “Abstraction in Artificial Intelligence and Complex Systems”, New York: Springer, 2013.

SECURING ONLINE ASSESSMENTS IN ONLINE EDUCATIONAL SYSTEMS USING BLOCKCHAIN

JOVANA JOVIĆ

Belgrade Metropolitan University, Faculty of Information Technologies, jovana.jovic@metropolitan.ac.rs

VIJAYAKUMAR PONNUSAMY

Department of ECE, SRM Institute of Science and Technology, Kattankulathur, India, vijayakp@srmist.edu.in

VLADIMIR MILIĆEVIĆ

Belgrade Metropolitan University, Faculty of Information Technologies, vladimir.milicevic@metropolitan.ac.rs

NEMANJA ZDRAVKOVIĆ

Belgrade Metropolitan University, Faculty of Information Technologies, nemanja.zdravkovic@metropolitan.ac.rs

Abstract: *The ongoing COVID-19 pandemic has led to more and more universities adopting an online-only option for studying. As a result, students often take course assessment through an online form within the Learning Management System (LMS), and hence the need for secure LMSs become more evident. Commercially available LMS solutions may not be secure in every aspect, especially when it comes to online assessments and their grading, either automatic or manual. In this paper, we present a blockchain-based add-on for the purposes of securing online assessments. We show that with our proposed solution, secure assessment and grading can be achieved with the innate security properties of blockchain.*

Keywords: *blockchain, learning management system, online assessments, secure eLearning.*

1. INTRODUCTION

The ongoing global COVID-19 pandemic has forced education institutions on all levels of study (primary, secondary, and higher education) to consider a switch from traditional „face-to-face“ learning methodologies to blended learning, or completely online learning [1]. Indeed, this switch presented a challenge to all parties involved – students, but also teaching staff, and their mutual communication. As most teacher-to-student communication is now being conducted exclusively online, several security issues arise [2, 3].

One of the main security issues regards communication channels. Namely, commercial video conferencing platforms such as Skype, Microsoft Teams, or Zoom are being implemented ever more in virtual classrooms offer limited security options without purchasing or subscribing to the service.

In some cases, the use of these platforms has shown more efficient student engagement [4], topics which include writing on the blackboard, such as mathematics and similar subjects, still face difficulties in presentation [5].

Another security issue regards online exams in which, without a complex e-monitoring system, cheating is possible [6-8]. The authors of [6] pointed out that security requirements for such a system would account for accessibility, monitoring, management, authenticity, integrity, secrecy and copy prevention and detection. It should be noted that most final exams in Higher Education Institutions (HEIs), unlike from Massive Open Online Courses (MOOCs), are conducted on-campus.

The third security issue, which is the topic of this paper, are online assessments which are conducted in online and blended methods in HEIs, as well as in MOOCs. The COVID-19 pandemic showed that not all HEIs were completely prepared for the completely online or blended learning, especially regarding conducting exam obligations such as individual assessments, tests, homework etc. [9]. Most HEIs employ some sort of Learning Management

System (LMS); however, these systems are often expensive, or lack specific functionalities such as assessment submission. Some HEIs, especially Science, Technology, Engineering, and Mathematics (STEM) universities choose to develop an in-house solution regarding these functionalities, and security issues are often overlooked due to time constraints. Most LMSs and similar learning platforms do not provide Virtual Private Networks (VPNs) when a student or teaching staff member connects to the LMS platform, and data transfer is often conducted using Hypertext Transfer Protocol (HTTP) which is prone to security vulnerabilities [10, 11].

Continuing from our previous work done in [12, 13], in this paper we provide a solution to securely monitor and access student's assessment by implementing Blockchain Technology (BCT). The rest of the paper is organized as follows. Section 2 presents our motivation, broken into two subsections: identified use-cases for the proposed platform, and a short introduction to the security properties of BCT, highlighting its advantages in these use-cases. Section 3 presents our solution with appropriate discussion, while Section 4 concludes the paper.

2. MOTIVATION

Our main motivation comes from the fact that communication between student and teaching staff regarding pre-exam obligations, upon switching to a completely online learning and teaching model, is at the same time simplified, and yet more complicated. Students which complete their assignments online within an in-house developed LMS can have difficulties if that LMS does not natively support fully online assignment submission.

Use-cases with security issues in online learning

We have identified multiple use-cases in which students' assignments may be prone to security issues. Namely, the first case, presented in Image 1, shows an unsecure channel between the student and the server-side LMS app. As stated in [10, 11], if no security measurements have been applied, a third party can potentially monitor local network traffic and obtain information about the assessment and about the student's submitted answers. This issue is even more important if the make-shift online assessment in developed with no encryption.

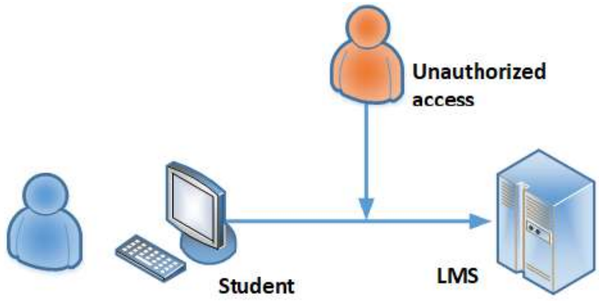


Image 1: Use-case #1: Unsecure traffic between the learner and the LMS.

The second use-case is any type of student's answer correction without authorization. This can be achieved either by the student, or by the teaching staff. The motivation behind this use-case is to remove any teaching staff bias. Finally, the third use-case is in regards to presenting the students their complete results upon their or teaching staff's request. If not implemented correctly, the student and/or teacher can only see, for instance, the number of correct or incorrect answers in a multiple-choice test, but not which ones were correct or not, which can be an issue of lack of complete information. The second and third use-case are presented in Images 2 and 3, respectively.

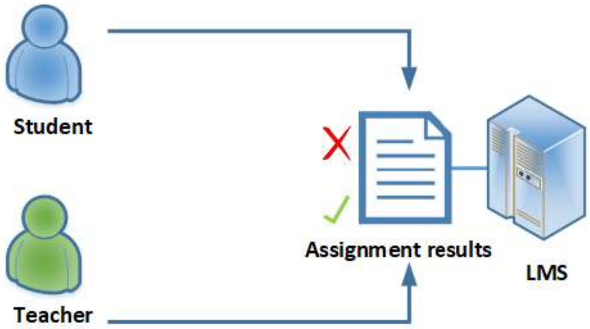


Image 2: Use-case #2: Unauthorized assignment result correction.

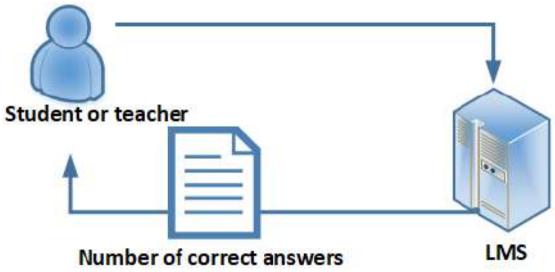


Image 3: Use-case #3: Incomplete access to assignment results.

We note that in most commercially available LMSs these security issues will not happen often, and that these use cases are more in line with in-house developed LMSs. In addition, in-house developed LMSs are, in general, open to modification and upgrades, which corresponds with our BCT extension regarding issues with online assessments.

Security properties of Blockchain technology

A Blockchain is a shared, append-only distributed ledger, in which all events, which are usually denoted as transactions, are stored in linked blocks [14]. Every event contains data regarding the event itself, as well as a unique cryptographic signature, which ensures that the blockchain ledger is resilient to modifications. A block can be viewed as a data structure consisting of a list of events, coupled with a header. This header connects the block to the previous one, forming a chain all the way to the genesis block. The combination of peer-to-peer networks, public-key cryptography, and distributed consensus is what secures blockchain transactions. Unlike a centralized system, no single entity within the blockchain should be able to adding a block to the chain: all nodes in the chain share equal rights. This type of decentralized storage is managed with a distributed consensus mechanism. Depending on the consensus algorithm, nodes can either compete for correct transaction validation, be chosen randomly, or apply a different algorithm altogether. One significant advantage of using BCT is that it can provide a decentralized management and access to all types of databases, providing only authorized access when needed.

3. SYSTEM MODEL AND DISCUSSION

In this Section, we provide a model for in-house developed LMSs which can connect to a blockchain running on the HEI's computers. It is important to note here that this type of blockchain would be a private blockchain, as all the nodes are in the ownership of the HEI. Multiple nodes running the blockchain can be deployed within an HEI's servers or multiple computers on campus, needed to run the consensus. This approach may seem redundant; however, this system still has an advantage compared to a traditional database, as any type of access to the blockchain will be recorded and verified by multiple nodes with the consensus mechanism, leaving a traceable timestamp on all the machines running the blockchain.

Our system is designed in such a manner that an endpoint from the LMS's regarding assessment is connected to the HEI's private blockchain. Each time a user, be it student or teaching staff, is accessing anything related to pre-exam assignments, an event is triggered and a new transaction is formed. This transaction is broadcast to all noted within the blockchain network, and is verified. This type of blockchain may need not have a processor-heavy consensus mechanism such as proof-of-work, used in Bitcoin [15]; it can use one of the mechanisms found in the Hyperledger family of BCTs [16]. The types of transactions can be categorized into deploying assessments, submitting assessment answers, and result view. These categories correspond with the requests sent from the user to the LMS, and replies sent from the LMS to the user, respectively. The overall system diagram is shown in Image 4.

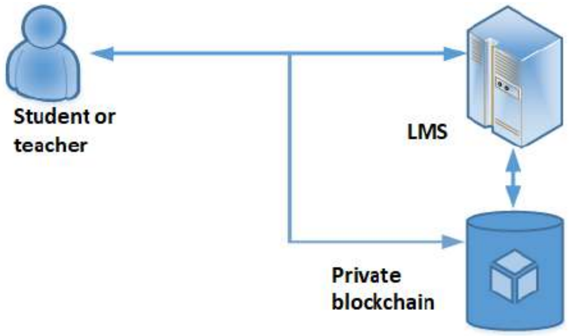


Image 4: Connecting a private blockchain to the user-LMS communication.

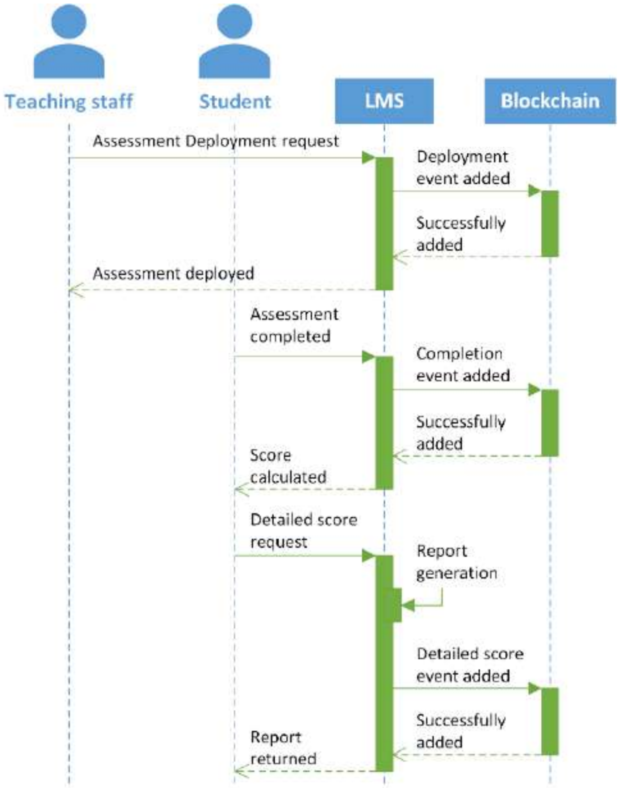


Image 5: Sequence diagram of assessment report with blockchain events.

Image 5 shows a sequence diagram where teaching staff is deploying an assessment, and a student is submitting assessments results, with a detailed result view afterwards. The blockchain will be triggered to make an event every time a request or response is made regarding assessments. This type of system ensures that no modification made to scores are made without being traceable. Furthermore, use case #1 can be addressed with the data in the events themselves. For instance, if assessments are made up of randomly chosen questions, each combination for each student will be written in a separate event. If an unauthorized party is monitoring unsecure traffic, the information they obtained may be traced back to the exact combination of questions.

5. CONCLUSION

In this paper, we have identified several security issues regarding in-house developed LMSs and pre-exam assessment submission and modification. By applying a private blockchain extension to this type of LMS, an additional layer of security can be achieved. Furthermore, this blockchain extension can be easily deployed on several on-campus computers, and with the correct choice of consensus mechanism, the extension may not be processor-heavy. Currently, we are exploring the possibilities of adding BCT in e-Learning systems, as they will most definitely be a crucial part of future learning methodologies in all levels of studies. Our future work will focus on an overall blockchain-supported LMS, which may expand to multiple campuses or even multiple universities.

ACKNOWLEDGMENT

This paper was supported in part by the Blockchain Technology Laboratory at Belgrade Metropolitan University, Belgrade, Serbia, and in part by the Ministry of Education, Science and Technological Development, Republic of Serbia (Project III44006).

REFERENCES

[1] J. Davis, „Traditional vs. Online learning: It's not an either/or proposition," Employment Relations Today, vol. 27, no. 1, 2000, pp. 47-60.

[2] J. B. Earp, F. C. and Payton, "Data protection in the university setting: Employee perceptions of student privacy," in Proc. of the 34th IEEE Annual Hawaii International Conference on System Sciences, 2001, pp. 6.

[3] N. H. M. Alwi, and I. S. Fan, "E-learning and information security management," International Journal of Digital Society, vol. 1, no. 2, pp.148-156., 2010.

[4] J. Alameri, R. Masadeh, E. Hamadallah, H. B. Ismail, H.B. and H. N. Fakhouri, 2020. "Students' Perceptions of E-learning platforms (Moodle, Microsoft Teams and Zoom platforms) in The University of Jordan Education and its Relation to self-study and Academic Achievement During COVID-19 pandemic," Advanced Research & Studies Journal, vol. 11, no. 5, pp. 21-33, 2020.

[5] M. Irfan, B. Kusumaningrum, Y. Yulia, and S. A. Widodo, "Challenges during the pandemic: use of e-learning in mathematics learning in higher education," Infinity Journal, vol. 9, no. 2, pp.147-158, 2020.

[6] I. Y. Jung, and H. Y. Yeom, "Enhanced security for online exams using group cryptography," IEEE transactions on Education, vol. 52, no.3, pp. 340-349, 2009.

[7] Y. Atoum, L. Chen, A.X. Liu, S. D. Hsuand X. Liu, "Automated online exam proctoring. IEEE Transactions on Multimedia, " vol. 19, no. 7, pp. 1609–1624, 2017.

[8] H. Ilgaz, and G. A. Adanır, G.A., "Providing online exams for online learners: Does it really matter for them?," Education and Information Technologies, vol. 25, no. 2, pp. 1255-1269, 2020.

[9] E. Edelhauser, L. Lupu-Dima, "Is Romania Prepared for eLearning during the COVID-19 Pandemic?," Sustainability, vol. 12, pp. 5438, 2020.

[10] G. D. Bissias, M. Liberatore, D. Jensen, and B. N. Levine, "Privacy vulnerabilities in encrypted HTTP streams," in Proc. Workshop on Privacy Enhancing Technologies, pp. 1-11, 2005.

[11] M. Vieira, N. Antunes and H. Madeira, "Using web security scanners to detect vulnerabilities in web services," in Proc. IEEE/IFIP, pp. 566-571, 2009.

[12] M. Damnjanović, V. Grković, and N. Zdravković, „Towards Secure online studies: Applying Blockchain to e-Learning, " in Proc. Of the 11th international conference on eLearning, 2020.

[13] N. O. Vesić, N. Zdravković, and D. J. Simjanović, „Securing Online Assessments using Christoffel Symbols," in Proc. Of the 11th international conference on eLearning, 2020.

[14] Z. Zheng, S. Xie, H. Dai, X. Chen, H. Wang, „An overview of blockchain technology: Architecture, consensus, and future trends," in Proc. of the IEEE international congress on big data (BigData congress), pp. 557–564, 2017.

[15] S. Nakamoto, „Bitcoin: A peer-to-peer electronic cash system," Decentralized Business Review, pp. 21260, 2008.

[16] V. Milićević, J. Jović, and N. Zdravković, „An overview of Hyperledger blockchain technologies and their uses," in Proc. Of the 11th International Conference on Information Society and Technology (ICIST 2021), pp. 62-65, 2021.

COMPARATIVE ANALYSIS OF DIGITAL SIGNATURES BASED ON RSA CRYPTOGRAPHY AND ELLIPTIC CURVE CRYPTOGRAPHY

MILOŠ S. DRAŽIĆ

Belgrade Metropolitan University, Faculty of Information Technologies, milos.drazic@metropolitan.ac.rs

MILOŠ JOVANOVIĆ

Belgrade Metropolitan University, Faculty of Information Technologies, milos.jovanovic@metropolitan.ac.rs

IGOR FRANC

Belgrade Metropolitan University, Faculty of Information Technologies, igor.franc@metropolitan.ac.rs

DRAGAN ĐOKIĆ

Belgrade Metropolitan University, Faculty of Information Technologies, dragan.djokic@metropolitan.ac.rs

BOJANA TOMAŠEVIĆ DRAŽIĆ

Belgrade Metropolitan University, Faculty of Information Technologies, bojana.tomasevic@metropolitan.ac.rs

Abstract: *Elliptic curve cryptography (ECC) as one of the most advanced cryptosystems is being used by many applications for encryption, public key exchange and digital signatures. The strength of this type of cryptography is in the significantly increased difficulty of solving the problem of discrete logarithm due to the replacement of numbers multiplication by the operation of points "addition" on an elliptic curve, without affecting the key size. The application of RSA to digital signatures is a method complementary to the ECC, in the sense that the trust in the procedure is built on the factorization of a large number of N and the difficulty of taking roots modulo N. Nevertheless, for the same level of security, the key lengths in RSA and ECC can differ drastically, resulting in a large difference in the load of the hardware components, which will be addressed in this paper.*

Keywords: *Digital signatures, elliptic curve cryptography, RSA*

1. INTRODUCTION

The emergence of public key cryptography, which came with the pioneering work of Diffie and Hellman [1], as well as the development of RSA (Rivest, Shamir, Adleman) public key cryptosystems [2], are the cornerstone of modern cryptographic procedures. Encryption solutions are focused on secure communication in an insecure environment, i.e. insecure network. While symmetric schemes, such as DES [3] and AES [4], have been built on ad hoc settings, for which there is no mathematical claim, which undoubtedly, formally rigorously, provides a clear assessment of the degree of security of the encryption method used, RSA has its status as a secure tool based on the paradigm of asymmetric encryption. The creators of this system added mathematical formalism to the entire cryptographic machinery, by which each step of the offered procedure, as well as its consequences, could be clearly mathematically grounded through definitions, theorems, propositions and lemmas. The rigor of this approach brought two benefits: on the one hand, mathematics was used to create the most advanced cryptosystem at the time, and on the other hand, the same mathematics gave very clear statements regarding security assessments and the strength of this solution. It was this second aspect, which had been missing until then, that brought additional confidence in the RSA. RSA quickly became the standard in the world of crypto protection and security, and it is important to mention here that its authors patented their result [5]. Protecting the originality of their results, they prevented insufficiently trained and educated individuals from dealing with "upgrades" and "improvements", which would certainly appear quickly in the case of open source.

This prevents confidence in this technology from being shaken. In addition to this fact, it is important to keep in mind that the computer resources of that time could not be a threat to the RSA. Therefore, the attempt to recognize the threat and weakness of the RSA was reduced to thought experiments, rather than in reality at all. Shor's algorithm is without a doubt the most important representative of such attempts [6]. However, it was Shor's algorithm that showed that with the application of power resources (quantum computers), RSA ceases to be the tool of choice for secure cryptography, because factorization could be achieved in polynomial time. A few years before the advent of the Shor’s algorithm, Victor Miller [7] and Neal Koblitz [8], independently introduced elliptical curves into cryptography. The disadvantage of their work lies in the fact that these are purely mathematical results, which did not clearly indicate that the problem of discrete logarithm would be significantly aggravated by the use of elliptic curves. Also, the authors did nothing to patent and protect their result. To the lost time that passed before the potential of elliptic curves was realized, should be added the time caused by the damage caused by mathematically unqualified people, whose "improvements", with the aim of obtaining the patent rights, in fact only questioned the security of such a cryptosystem. It should not be overlooked that the mathematical complexity of applying elliptic curves and the level of abstraction, such as the operation of points "addition” on a curve, made it even more difficult to build confidence in elliptic curves quickly. Due to all this, the real application of elliptic curves in technology had to wait more than 15 years after pioneering works. However, today it can be said that the lost time has been successfully compensated, so elliptical curves have become the tool of choice for many new technologies, such as cryptocurrencies [9], and that in some important aspects they greatly exceed RSA.

2. THE CONCEPT OF DIGITAL SIGNATURE

Unlike the idea of secure communication in an insecure environment, digital signature rests on a different concept. The idea itself is simple and is part of everyday experience. If there is a document and a person who wants to verify the items, provisions, or messages listed in that document, he/she will do so by affixing his signature, trademark, or any mark by which he/she wishes to certify the authenticity of that document. Here, an adequate analogy can make a clear distinction between public key cryptography and digital signature. Let the letters, which anyone can send, be inserted into the mailbox of a certain person. Although these letters can come from anyone, only the owner of the mailbox will have access to their contents, after opening them. If we imagine that among the received letters there is someone who is stamped, then we can assume that the owner of the box, as well as anyone who could take a look at that letter, will be convinced of the authenticity of that document, in the sense that it will not doubt that letter had just been compiled by the person whose stamp was on the document. On the other hand, only that person, using his own stamp, is capable of creating such an impression and trust. Now we can translate all this into the language used in cryptography. We can consider a mailbox to be a public encryption key, and a private decryption key is actually a mailbox key owned only by the owner. In this way, the concept of secure communication is created. On the other hand, for the physical stamp used, we can consider that in digital signing it represents a private signing key, while the imprint of that stamp on the document is a public signature, or public signing key. Also, the procedure by which the sender puts a stamp on a document is a type of signing algorithm, while the procedure by which the stamp on a given document is compared to a real, physical stamp can be considered a verification algorithm. The basic elements in the steps in digital signing are listed: A private and a public signing key, as well as a signing and a verifying algorithm.

Careful analysis allows us to identify what necessary conditions a successful digital signing should meet. If someone who has insight into the imprint of the stamp would want to create his own physical stamp based on this imprint to try to falsify the real stamp, he/she should not be able to do so. Likewise, the imprint should be so unique that the attacker, unable to create his stamp, must not be able to find among the existing physical stamps available to him, one that gives an identical imprint as the stamp he/she wants to forge. This analysis refers to the case of one printed document. Now imagine that the same stamp is imprinted on a number of documents. Each document and stamp print could potentially reveal more and more about the real, physical stamp to the attacker. The requirement that is imposed is precisely that no matter how many letters with the imprint of the stamp the attacker has an insight into, he/she cannot conclude anything more about the real stamp than if he/she had insight into only one letter with the imprint of the stamp.

This simple analysis clearly indicates that digital signatures are of great importance, because they confirm that the information comes from a reliable party.

3. RSA DIGITAL SIGNATURES

RSA, applied to digital signing, includes the following fundamental steps:

- The sender creates the key
- The sender signs the document
- The recipient is doing verification

In the first step, the following operations are performed:

- Selection of secret prime numbers p and q,
- Selection of the verification exponent V_e, such that V_e and (p-1) (q-1) are mutually prime,

- Calculation of private signing key S_{priv}, which is a modular inverse of the verification exponent per module (p-1) (q-1),
- Publishing the values of N and V_e, where N = pq.

The second fundamental step involves the following operations:

- Select a digital document D of length equal to or greater than 1 and less than N,
- Digital signature calculation, i.e. public signing key S_{pub}, which is congruent $D^{S_{priv}}$ modulo N.
- Publishing D and S_{pub}.

The third fundamental step consists of:

- Taking public values published by the sender, N and V_e,
- Calculating $S_{pub}^{V_e}$ mod N and confirming that the obtained value is the same as the value of D.

3. ECC AND DIGITAL SIGNATURES

The elliptic curve formalism, applied to digital signing, is a direct application of the rules established by the Digital Signature Algorithm (DSA) [10], which represents a successor and an improved version of the Elgamal solution [11]. Therefore, the steps that are specific to DSA will be listed here, and the difference between standard DSA and Elliptic curve DSA (ECDSA) will be that in the case of ECDSA, the binary operation involves points addition on the curve.

Fundamental steps include:

- The sender creates public parameters and a key
- The sender signs the document
- The recipient performs the verification

In order for the whole procedure to be successfully implemented, it is necessary to start from the construction of a finite field F_p, where p is a prime number. When a construction is made by removing 0 from the field F_p, a group structure F_p^* of order $p-1$ is obtained, which is closed to a binary operation. For such a group, there is certainly one element G such that all other elements of the group, F_p^* = {1, G, G^2, G^3, ..., G^{p-1}}, can be generated from it by successive application of the binary operation. The speed of signing depends on the order of the group F_p^*, so the question arises whether and to what extent security is sacrificed if the order of the group is reduced, i.e. if a subgroup is taken. In general, the index calculus will not simplify the problem of solving the discrete logarithm if a subgroup of the group F_p^* is selected. Working with subgroups simplifies the procedure and shortens it. It remains to define the subgroup of order $q < p$, where q is a prime number. This subgroup is easiest to determine by assuming that it can be generated from the element g which is the q-th root of $(p-1)$-th power of the generator of the group, i.e. $g = G^{(p-1)/q}$. This laid the groundwork for the beginning of the procedure.

As part of the first step, the following operations are performed:

- Selection of prime numbers p and q, such that p is congruent to the number 1 modulo q, as well as the element g of the subgroup of the group F_p^*, which is of order q,
- The private signing key S_{priv} and verification key V_k are created, as follows: first, S_{priv} is chosen such that it is greater than or equal to 1 and less than or equal to q-1, and then a verification key is generated, satisfying $V_k \equiv g^{S_{priv}}$(mod p).
- The verification key is then published (V_k is equivalent to the verification exponent in the RSA procedure).

The second step involves:

- Selection of a document D and calculation D mod q
- Selection of a random number r, which satisfies $1 < r < q$
- Calculation of signature pair:
  S_{pub1} = (g^r mod p) mod q,
  S_{pub2} ≡ (D + $S_{priv}S_{pub1}$) r^{-1} (mod q)
- Publication of document D and pair (S_{pub1}, S_{pub2})

The third step consists of the following procedures:

- Calculating the value pair (V_1, V_2) as follows:
  V_1≡D S_{pub2}^{-1} (mod q) and V_2≡ $S_{pub1}S_{pub2}^{-1}$ (mod q)
- Confirm that (g^{V_1} ${V_k}^{V_2}$ mod p) mod q=S_{pub1}.

4. ECDSA VS. RSA DIGITAL SIGNATURE-METHODS

To perform a comparative analysis of the hardware load as well as the time required for digital signing and verification, a computer with the following performances listed was used:

- Processor: Intel(R) Xeon(R) CPU E5-2640 0 @ 2.50GHz (2 processors)
- No. of virtual processors: 8
- System type: 64-bit operating system, x64-based processor
- Edition: Windows 10 Pro
- Installed RAM: 6GB
- Hard disk capacity: 100GB

The comparison was made on the basis of data from official documents related to ECC [9] and RSA [12]. **Table 1** lists the curves used, their size (the length in bits of the field order) and the strength (number of bits of security), as well as the values of N = pq corresponding to the RSA algorithm of the same security strength.

Table 1: Elliptic curves used, their size, strength and modulus N corresponding to RSA algorithm of the same strength

Elliptic curve	Size (bits)	Strength (bits)	N (bits)
Secp192r1	192	96	1536
Secp224r1	224	112	2048
Secp256r1	256	128	3072
Secp384r1	384	192	7680
Secp521r1	521	256	15360

Tools used:

- Development environment: NetBeans IDE 8.2
- Performance Monitor within which the following counters were monitored:

1. Processor > % Processor Time,
2. Memory > Pages/sec,
3. Physical Disk > % Disk Time,
4. Physical Disk > Avg. Disk Queue Length,
5. System > Processor Queue Length,
6. Network Interface > Bytes Total/sec.

Testing was done in the Java programming language with the BouncyCastleProvider [13] as a security provider in ECDSA implementation.

4. RESULTS AND ANALYSIS

Based on the measurements performed using the Performance Monitor, as well as working in the NetBeans environment, which gave us the values of the time it took to generate the keys, signing the digital document (message) as well as verification, we obtained the results listed in the table below.

Table 2: Comparison of time in case of using ECDSA and RSA, which are required for keys generation (Keys gen. column), signing (Sign. column) and verification (Ver. column) for keys of different sizes (Size column) and for the appropriate security level (Strength column)

	Strength (bits)	Size (key length in bits)	Keys gen. (s)	Sign. (s)	Ver. (s)
EC DSA	96	192	0.497	0.039	0.004
	112	224	0.498	0.039	0.005
	128	256	0.5	0.04	0.005
	192	384	0.522	0.043	0.007
	256	521	0.543	0.044	0.011
RSA	96	1536	0.398	0.03	0.002
	112	2048	0.89	0.035	0.002
	128	3072	2.805	0.055	0.004
	192	7680	126.77	0.353	0.005
	256	15360	1245.65	2.663	0.014

What is immediately noticeable is that the key generation speed, which is the first fundamental step in digital signing, is for the lowest level of security on the RSA side compared to ECDSA and that RSA is about 25% faster in execution than ECDSA. This is somewhat understandable, if we take into account that to perform the first step in the case of RSA, it is enough to apply Euclidean, i.e. extended Euclidean algorithm, with the index N being a relatively small number. In addition, we must take into account that in our example, it was not taken into account that any primality test, such as the Rabin-Miller test [14], [15], was performed when selecting prime numbers. On the other hand, in the first step in ECDSA, a reduction had to be done to a subgroup whose order has a well-defined property given by the modular equation, to conduct a random number generation, and finally to solve another modular equation. Already at this point we can notice that the very first step in the case of RSA, with an increase in the value of the N index, with the aim of increasing the level of security, will increase nonlinearly over time. On the other hand, the increase in the value of the finite group order in the case of ECDSA, in order to achieve a higher level of security, will not be the same as the growth rate of index N. This will result in only a slight increase in key generation time. One of the reasons is the efficiency of the double and add algorithm which performs points addition on elliptic curves (the speed of this operation would be even higher if we worked on Koblitz, not Weierstrass curves, due to the high efficiency of Frobenius mapping [8]). In the case of raising the security level, it is noticed that the key generation time in the case of ECDSA has only slightly increased, while in the case of RSA this increase is dramatic and for a 256-bit security level, it is over 20 minutes! We also note that the time required for signing and verification in the case of ECDSA is between 0.043s and 0.055s and is, as expected, the shortest for the lowest and the longest for the highest security level considered. For ECDSA, we can conclude that the signing time is always longer than the verification time, but that the increase of that time with increasing levels of security is more pronounced in the case of verification time. We can understand this behaviour if we take into account that three modular equations must be solved for both signing and verification, with the difference that in case of signing a document of fixed length (actually a Latin sentence “Festina lente” in our case) is always taken, while in case of verification the parameters are variable, i.e. they grow numerically, so as the order of the group increases, so does the complexity of the modular equations. We can assume that with a further increase in the order of the group (security level), the verification time would become longer than the signing time. Based on the obtained results for ECDSA, we can conclude that of the total time required for key generation, signing and verification, about 90% of the time was spent on key generation and that computer resources are the most loaded during this activity. In the case of RSA, the signing time is comparable to ECDSA for the lowest security levels, but the time-nonlinear behaviour with further increase of the N index is clearly visible and that for 192-bit security level key generation in RSA is very inefficient in relation to ECDSA. Computer resources are almost entirely spent on this activity. In contrast, the verification time is the same as in the case of ECDSA, while the signing time has an increase that is more pronounced with the increase in security levels. This is to be expected if one considers the modular equation that must be solved in the second fundamental step. Based on these results, it can be said that the price for increasing the level of security in RSA was paid by the increase in the value of the N index, which made it significantly more difficult to generate key pairs. In the case of ECDSA, the increase in the level of security was achieved through the introduction of two public keys (signatures), so in order to increase security, it was not necessary to increase the order of the group significantly. An additional reason for this is that the problem of discrete logarithm on elliptic curves is a far more complex than it is in standard modular arithmetic. The fact that the order of the group does not have to increase drastically, as well as the mentioned efficiency of binary operation on elliptic curves (double and add), results in an almost constant time required for key generation, signing and verification, regardless of the required security level.

As for the counters that monitored the system load, we can say that in the case of ECDSA in the first place are time-localized jumps in value, and the reason for this is the fact that execution times are very short. Therefore, only the highest counter values are listed in the ECDSA case (**Table 3**). The System\Queue Length counter was zero for all cases, so it is not shown in the table.

Table 3: Counters in the case of ECDSA for different key lengths.

Key	Mem. (pages/s)	Net. (bytes/s)	\%disk time	\%avg. queue length	\%proc. time
192	16.9	17949	2.13	0.021	35.48
224	24.2	20607	2.02	0.02	36.32
256	17.2	28862	2.53	0.025	35.86
384	16.9	18891	2.17	0.022	33.17
521	24.9	20034	2.38	0.024	35.35

Based on the obtained values, we can see that these are events that are quite stable and uniform, and that they show that the ECDSA loads the system almost equally, regardless of the level of security. This is in favour of the constancy in execution times that was observed earlier.

In the case of RSA, for lower security levels, significant counter changes are, as with ECDSA, localized jumps, but as the security level increases, the system load becomes continuous. In this sense, two modes can be recognized: one corresponds to time-localized jumps in the system load, while the other is characteristic of a continuous load. Based on the analysis so far, it is not difficult to conclude that the first mode includes RSA signatures with key values of 1536, 2048 and 3072 bits, while the second mode includes lengths of 7680 and 15360 bits. The **Table 4** shows the characteristic values: maximum for the first mode, and mean values for the second mode. System > Processor Queue Length counters in the case of 7680 and 15360 show non-zero behaviour, which is understandable given the length of execution, and amounts to 0.18 and 0.77, respectively.

Table 4: Counters in the case of RSA for different key lengths and two different modes. In the case of the first mode (key lengths 1536, 2048 and 3072) the maximum values are displayed, and in the case of the second mode (7680 and 15380) the mean values are shown.

Key	Mem. (p/s)	Net. (by/s)	\%disk time	\%avg. queue length	\%proc. time
1536	16.94	38299	2.27	0.008	30.22
2048	16.89	46791	2.62	0.026	25.83
3072	16.77	39620	2.6	0.006	25.11
7680	1.126	22536	0.276	0.003	15.351
15360	5.1498	44501	0.332	0.003	13.282

We notice that in cases of lower security level, the system load is similar to ECDSA, but it is also important to note that the mean values of Memory> Pages / sec and Physical Disk>% Disk Time counters in continuous load mode are about an order of magnitude smaller in relation to the pulse mode, while in the case of Processor>% Processor Time, the counters are only about 2 times smaller. Considering the execution length in continuous load mode, it is clear that the hardware components are significantly loaded in the case of RSA than in the case of ECDSA.

5. CONCLUSION

The primary difference between RSA and ECC is in the strength of encryption. ECC provides an equivalent level of encryption power as an RSA algorithm with a shorter key length. The speed and security offered by ECC is higher than the RSA for Public Key Infrastructure (PKI) and Digital Signature. RSA requires much longer key lengths to implement encryption. ECC requires a much shorter key length compared to RSA. RSA increases key lengths by increasing security. Currently, the standard in the RSA is a 2048-bit key length. As various operations are performed on the computer/ server while simultaneously generating keys, signing, encrypting and decrypting data, this certainly puts an additional burden on the computer. The ECC does not face this type of challenge. The fact that its keys are much smaller requires less load on computers / servers. ECC with a 224-bit key achieves the same level of security as RSA with a standard 2048-bit key, which is almost 10 times smaller than the RSA. RSA is resource hungry as a cryptosystem. Encryption standards are becoming stricter. A higher degree of key pair protection is required. RSA can increase key length, but it does not improve security. Its security is not proportional to the growth of the key. A 3072-bit key does not provide double security compared to a 2048-bit key. More time to generate keys affects the security drop. The time lost in this way is given to the attacker, and on the other hand, computer resources are more burdened, which endangers the life of hardware components that can lead to the collapse of computer systems. With a much shorter key length for the same level of security through an efficient load on the hardware components, the ECC beats the RSA.

Based on all this, we can say that the ECC is a qualitative and quantitative step forward in relation to the RSA and that in case of a high level of security, the ECC should be the tool of choice.

REFERENCES

[1] W. Diffie and M. E. Hellman, “New Directions in Cryptography,” IEEE Trans. Inf. Theory, vol. IT-22, pp. 644-654, Nov. 1976.

[2] R. L. Rivest, A. Shamir and R. Adleman, “A Method for Obtaining Signatures and Public-Key Cryptosystems,” Commun. ACM, vol. 21, pp. 120-126, Nov. 1978.

[3] NIST–DES, “Data Encryption Standard (DES),” FIPS Publication 46-3, National Institute of Standards and Technology, Oct. 1999.

[4] NIST–AES, “Advanced Encryption Standard (AES),” FIPS Publication 197, National Institute of Standards and Technology, Nov. 2001.

[5] R. L. Rivest, A. Shamir and R. Adleman, “Cryptographic communications system and method,” U.S. Patent 4 405 829, Sep. 20, 1983.

[6] P. W. Shor, "Algorithms for quantum computation: discrete logarithms and factoring," Proceedings 35th Annual Symposium on Foundations of Computer Science, pp. 124-134, Nov. 1994.

[7] V.S. Miller, “Use of elliptic curves in cryptography,” Advances in Cryptology-CRYPTO ’85, in Lecture Notes in Computer Science, vol. 18, Ed. Springer, Berlin, 1986, pp. 417-426.

[8] N. Koblitz, “Elliptic curve cryptosystems,” Math. Comput. vol 48, pp. 203–209, Jan. 1987.

[9] Standards for Efficient Cryptography, “SEC 2: Recommended Elliptic Curve Domain Parameters, “ https://www.secg.org/sec2-v2.pdf , Jan. 2010.

[10] NIST–DSS, “Digital Signature Standard (DSS),” FIPS Publication 186-2, National Institute of Standards and Technology, Jan. 2000.

[11] T. Elgamal, “A Public-Key Cryptosystem and a Signature Scheme Based on Discrete Logarithms,” IEEE Trans. Inf. Theory, vol. 31, pp. 469-472, Jul. 1985.

[12] K. Moriarty, B. Kaliski, J. Jonsson and A. Rusch, “PKCS #1: RSA Cryptography Specifications Version 2.2, “RSA Laboratories’ Public-Key Cryptography Standards (PKCS) series (RFC8017), Nov. 2016.

[13] The Legion of the Bouncy Castle [Online]. Available: https://www.bouncycastle.org/latest_releases.html (Accessed: Nov. 2021).

[14] G. L. Miller, "Riemann's Hypothesis and Tests for Primality," J. Comput. Syst. Sci., vol. 13, pp. 300-317, May 1975.

[15] M. O. Rabin, "Probabilistic algorithm for testing primality", J. Number Theory, vol. 12, pp. 128-138, Feb. 1980.

SECURITY OF GOVERNMENT CRITICAL INFRASRUCTURES WITH SCADA

MILOŠ JOVANOVIĆ

Belgrade Metropolitan University, Faculty of Information Technologies, milos.jovanovic@metropolitan.ac.rs

IGOR FRANC

Belgrade Metropolitan University, Faculty of Information Technologies, igor.franc@metropolitan.ac.rs

MILOŠ S. DRAŽIĆ

Belgrade Metropolitan University, Faculty of Information Technologies, milos.drazic@metropolitan.ac.rs

NENAD BIGA

Graduate School of Business, La Salle University, Philadelphia, United States, nenadbig@gmail.com

BOJANA TOMAŠEVIĆ DRAŽIĆ

Belgrade Metropolitan University, Faculty of Information Technologies, bojana.tomasevic@metropolitan.ac.rs

Abstract: *In a world where speed, efficiency, as well as access to an ever-growing amount of information and services are of strategic importance, the digital environment is what provides the necessary level of functionality that can meet the needs of society as a whole today. In that sense, the digital environment represents a paradigm and a necessary assumption on which the functioning of a wide range of data and services that are part of the so-callled critical infrastructures. As an interdependent, networked system, critical infrastructures provide fundamental services that are important for security, economic growth and prosperity, as well as social welfare and well-being. Critical infrastructures are based on services and related technologies that cover the production and distribution of energy, agriculture, banking and finance, transport and telecommunications, etc. Successful monitoring and control of these resources are widely implemented globally through the SCADA system service. The priority given to the robustness and operability of SCADA services often results in an insufficient level of security and network security problems, which will be addressed in this paper.*

Keywords: *Information Security, Critical Infrastructure Systems, SCADA*

1. INTRODUCTION

Awareness of the scale of the threats that cyber-attacks can have on critical infrastructures has grown significantly in recent years. Today, it can be said that practically every sector is endangered by cyber-attacks, with a special emphasis on the field of public health, energy production and telecommunications. In that sense, special attention must be paid to the monitoring and control systems, which are vital for their functioning. Efficient operation and functioning of critical infrastructure, as well as timely optimization, through real-time data collection and analysis, has been achieved globally through the implementation of the SCADA system. Weaknesses of SCADA systems and their vulnerability to cyber-attacks, could potentially lead to a violation of the integrity of critical infrastructure resources with unforeseeable consequences for society globally. SCADA control systems typically include sensors, actuators, and associated control software, which are deployed in widely dispersed locations. Due to the high exposure of these devices, it is necessary to use the best safety practices by the personnel who have access to them. The importance of SCADA devices is all the greater if we take into account that not only critical infrastructure, but also systems such as HVAC, traffic control and building automation rely on their proper functioning [1]. In order to best prevent cyber-attacks, it is necessary to identify the weaknesses and potential vulnerabilities of the SCADA system. The current status requires the establishment of security countermeasures, which is something that the governments of Western countries were the first to recognize.

2. BACKROUND

The importance of critical infrastructure can be seen through the global, generally accepted practice of a well-run state and socio-economic order. The key indicators by which this is measured are voice and accountability, political stability, the absence of violence, government efficiency and quality regulation, rules of law and control of corruption. It is information and communication technologies (ICT) that have helped develop transparency, government accountability and reducing corruption, through the direct participation of citizens in government, the avoidance of mediation and the development of democracy [2]. The mentioned progress is being achieved to a great extent by achieving the goals set by the security of an informational system, which are integrity, confidentiality, secrecy, availability, accountability and information assurance [3]. Data, as a fundamental element of any information architecture, is sampled, exchanged, presented and stored with the help of adequate equipment, and it is the security of the system that is applied in order to preserve the attributes of the data. These include availability (access to information uninterrupted by malicious denials of service or unauthorized deletions), integrity (guaranteeing the protection of information from any kind of modifications) and confidentiality (access to information is allowed only to authorized personnel). All these are prerequisites for the reliable functioning of the entire information system. Through the proper functioning of the critical infrastructure, a reliable flow of products and services that are crucial for the defence and economic security of the country is enabled, as well as the uninterrupted work of governance at all levels and society as a whole [4]. All critical infrastructures are complex in the sense that they are interacting components in which change occurs through the learning process. This allows us to define general appearance of multiple infrastructure and to take into account interdependencies and multiple connection points accordingly [5]. Initially, SCADA systems relied on primitive serial protocols and communication infrastructures to link SCADA components and to transport control and data messages. This was accompanied by the absence of security mechanisms. Therefore, standards have been created that adopt security solutions to mitigate risk in industrial control environment. Applied to critical infrastructure installations, SCADA systems must meet specific security requirements and develop appropriate security mechanisms and strategies [6], [7]. The consequences that cyber-attacks can have on critical infrastructure are so devastating that at the level of state sovereignty and integrity they can be considered a terrorist attack or even an open act of aggression and declaration of war in the same way traditionally seen, through engagement. army and military resources. The problem, however, is that the very nature of cyber-attacks, through the relatively easy manipulation of forensic evidence, does not allow the individual or nation-state behind such activity to be easily identified.

3. THE CURRENT ROLE

Today, cyber-attacks have become so complex that they can cause systems shutdown, disrupting operations or even remote control over the attacked systems. Attacks of this kind on critical infrastructure have dramatic consequences on economic security, public health, safety, and even physical survival itself. SCADA systems, as vital elements of critical infrastructure, control pipe lines, refineries, chemical plants, utilities, water and transportation systems, as well as manufacturing operations, etc. SCADA functionalities include:

- Access control: users are divided among groups who are given different, well-defined privileges to the process parameters of the system and to specific product functionality.
- Multimedia interface: supports multiple screens, displaying combinations of diagrams and text.
- Trending: provides trending facilities, summarizing common capabilities through a chart or image.
- Alarm handling: Information exists only in one place, all users see the same status, i.e. acknowledgment, and priority levels of multiple alarms are supported. It is possible to group alarms and manage them as aggregation.
- Logging/archiving: logging, as a medium-term storage of data on a disk, is typically performed when the appropriate file size, period, or multiple points have been reached, and the existing data is overwritten. Archiving, like long-term storage of data on a disk or some other permanent medium, can also mean the transfer of logged data once the log is full.
- Report generation: reports can be sent using SQL type queries to the archive, real-time databases, or logs.
- Automation: the action is triggered automatically after the observed event [8].

SCADA provides real-time management, effectively implements control, raises the level of security of the supervised structure and its employees and reduces costs. All these benefits are possible due to the use of standard hardware and software in SCADA systems, as well as improved communication protocols and increased connectivity to outside networks, including the Internet. The price for these benefits is increased vulnerability to attacks or errors coming from both external and internal sources.

Typically, SCADA systems include human-machine interface (HMI) that is responsible for presenting data to the operator, remote terminal units (RTUs) used for connecting sensors in the plants, converting signals to digital data and sending this data to the supervisory system, the monitoring system responsible for data acquisition and process activity control, programmable logic controllers (PLCs) which are final actuators used as field devices as they are more economical, flexible and configurable than, for this purpose designed RTU, communication infrastructure that connects the supervisory system to the RTUs and/or PLCs, various processes and analytical instrumentation [9]. The traditional SCADA system consists of a host computer, some RTUs, the operator terminals and PLCs. In addition to the components already mentioned, it also includes a SCADA meter used for gathering data from (acquiring) and sending commands (control) to a plant [10].

In order to better understand the threats, it is necessary to have a good understanding of the functioning of the SCADA system, and for that purpose it is good to look at the schematic SCADA architecture in a modern power plant (Image 1). As a central master system, SCADA controls RTUs which consist of relay devices, actuators and sensors, circuit power breakers, voltage regulators, etc. Higher-level units are the so-called master level units (MTUs), including supporting applications, HMIs, data storage and acquisition systems. RTUs and sensors control goes via PLCs, while programmable automation controllers are used as the basic controlling unit.

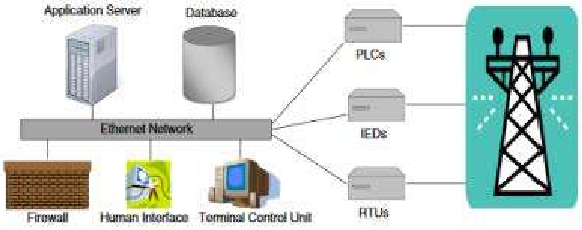


Image 1: SCADA architecture in modern power grids

There are three generations of SCADA system architecture, which rely on different solutions in the sense of communication between MTUs and RTUs. The first generation uses wide area networks (WAN), the second generation relies on local area networks (LAN), while the third generation uses WAN and Internet Protocol (IP). The communication component of the SCADA architecture, for example, includes Ethernet, wireless networks and Modbus and DNP3 protocols. Devices that are part of the SCADA architecture are generally controlling and controlled devices, which are run on embedded operating systems to communicate data primarily using protocols such as Modbus or DNP3. Attacks on the SCADA system pose a threat to human security, loss of productivity and environmental damage [11].

4. CHALLENGES, IMPACT AND SECURITY

Vulnerabilities in critical infrastructure have increased by 600% in the last decade, based on data presented in NSS Labs' Vulnerability Threat Report. This report pointed to an increase in hardware and software vulnerabilities related to the industry, as well as the obsolescence of a large number of SCADA systems. Main issues are:

- Increased exposure, as smart devices and systems create many access points;
- Inter-connectivity as a consequence of communication networks, which makes the system more exposed;
- Complexity of the electric system due to the interconnection of a large number of subsystems;
- Common computing technologies, because smart grid systems will increasingly use commercially available technologies and thus be subject to their weaknesses;
- Increased automation, because smart grid technology will automate many functions, and improper use of this data poses a new risk.

During 2012, The Industrial Control Systems Cyber Emergency Response Team (ICS-CERT) monitored 171 unique vulnerabilities affecting ICT products, coordinating the vulnerabilities with 55 vendors. The results of this tracking can be seen in Table 1.

Vulnerability type	**No.**
Buffer overflow	44
Input validation	13
Resource exhaustion	8
Cross-site scripting	8
Path traversal	8
Resource management	8
Access control	7
Hard-coded password	7
DLL hijacking	6
Other	39
Miscellaneous	15

Table 1: Vulnerabilities by type ICS-CERT 2012.

Key infrastructure challenges include:

- Secure interoperability between systems from different agencies;
- Development of methods and measurements of civic participation in democratic processes;
- Encouraging public and private partnerships and other networked organizational forms;
- Archiving and management of electronic records;
- Developing better methods for managing IT resources;
- Ensuring the availability and equality of access to data.

The overall security of critical infrastructure must be audited throughout the life cycle of its components. Authentication, access control and audit form the basis of information systems security. These three elements are interrelated: authentication, as an identification process, is a prerequisite for access control, while access control places restrictions on the actions performed by the authenticated user. Finally, possible security breaches can be identified through an audit process that collects data on activities and actions. In this sense, audit and monitoring logs are of the utmost importance. While in the case of an audit, information is obtained when the event has already passed, monitoring provides real-time information. The application of these two techniques should bring similar benefits to the security of SCADA systems as it has already brought to IT systems. Technical audits of SCADA devices and networks are critical to security. Today, there are many commercial and open source security tools that allow audits to be conducted on systems and networks to identify activities, patch levels, and common vulnerabilities. The problem that arises in this regard is in the components of SCADA systems that are of different ages and sophistication, which is why for many of them there is no capability of logging [12]. The costs of eliminating such disadvantages must be weighed against the potential benefits. In addition to the above problems, the question of the proprietary protocol also arises. Some SCADA systems use unique proprietary protocols for communications between field devices and servers. Often the security of SCADA systems is based on the security of these protocols, but these protocols provide very little security. Most SCADA systems currently in use have no security features at all. Modems, wireless, as well as wired networks used for communication and maintenance, are a significant vulnerability for SCADA networks and remote sites. In general, any location connected to the SCADA network is a target, especially unmanned or unguarded sites. To ensure the security of SCADA networks, the following steps need to be taken:

- Identify all connections to the SCADA network, through conducting a risk analysis to assess the risk and necessity for the existence of each of the connections.
- A good understanding of how all connections work and how those connections are protected.
- Disconnect all unnecessary connections.
- Isolate the SCADA network from other networks as much as possible.
- Evaluate the security of all remaining connections to the SCADA network by conducting penetration testing or vulnerability analysis.
- Use this information, along with the risk management process, to develop a protection strategy to strengthen the remaining connections [13].
- Strengthen SCADA networks by removing or disabling unnecessary services.
- Because SCADA control servers, built on commercial or open source operating systems, can be vulnerable to attacks through default network services, unused services and network daemons should be removed or disabled to greatest degree possible. This is especially important when SCADA networks are interconnected with other networks.
- Do not rely on proprietary protocols.
- Implement security features provided by device and system vendors and in that sense insist that system vendor provides features through product patches or upgrades.
- Establish strong control over any medium that is used as a backdoor into SCADA network.
- Apply strong authentication where there are backdoors or vendor connections in SCADA systems.
- Implement internal and external intrusion detection systems and strategy and establish 24-hour-a-day monitoring that includes alerting network administrators of malicious activity.
- Perform technical audits of SCADA devices and networks and any other connected networks.
- Conduct a physical security survey and inventory access points at each facility that has a connection to SCADA system and assess all remote sites connected to the SCADA network to evaluate their security.
- Establish SCADA “Red Teams” to identify and evaluate attack scenarios as well as potential system vulnerabilities. It is necessary to gain insight into the weaknesses of the overall network, SCADA systems, physical systems and security controls [14].

5. CONCLUSION

Information technology has led to more and more connected and complex infrastructures with increased centralization of control. Risks related to the level of critical infrastructure security are high, and the consequences of compromising the security and integrity of critical infrastructure can be dramatic. Security issue should concern us all. Of great interest in governments is the assessment of the security of critical infrastructure and industrial control systems managed by private companies.

Control systems within critical infrastructure are particularly vulnerable to cyber-attacks. SCADA systems are growing in their complexity and integration tests are necessary in the deployment phase. By adopting best practices, such as Virtual Private Networks (VPNs) for remote access, or removing, disabling, or renaming any default system account, prevention can be provided.

Global collaboration and sharing of information regarding possible cyber threats and vulnerabilities of every device that is qualified on the market, will bring overall security [15]. The security component is of the utmost importance and the overall security of critical infrastructures must be audited throughout the lifecycle of its components.

REFERENCES

[1] A. Reha and A. O. Said, “Tri-band fractal antennas for RFID applications,” Wireless Engineering and Technology, vol. 4, pp. 171-176, Oct. 2013.

[2] P. Salatin and H. Fallah, “Impact of information and communication technology (ICT) on governance quality,”

European Online Journal of Natural and Social Sciences, vol. 3, pp. 250-256, Mar. 2014.

[3] J. Joshi, A. Ghafoor, W. G. Aref and E. H. Spafford, “Digital government security infrastructure design challenges”, Computer, vol. 34, pp. 66-72, Feb. 2001.

[4] H. Shorr and S. J. Stolfo, “A digital government for the 21st century”, Communications of the ACM, vol. 41, pp. 15-19, Nov. 1998.

[5] S. M. Rinaldi, J. P. Peerenboom and T. K. Kelly, “Identifying, understanding, and analyzing critical infrastructure interdependencies”, IEEE Control Systems Magazine, vol. 21, pp. 11-25, Dec. 2001.

[6] D. Kilman and J. Stamp, “Framework for SCADA Security Policy,” Sandia National Laboratories report SAND2005-1002C, Oct. 2005.

[7] E. J. Byres, M. Franz and D. Miller, “The Use of Attack Trees in Assessing Vulnerabilities in SCADA Systems,” IEEE Conf. International Infrastructure Survivability Workshop (IISW ’04), Dec. 2004.

[8] J. Gao, J. Liu, B. Rajan, R. Nori, B. Fu, Y. Xiao, W. Liang and C. L. P. Chen, “SCADA communication and security issues,” Security Comm. Networks, vol. 7, pp. 175-194, Jan. 2014.

[9] A. Daneels and W. Salter, “What is SCADA,“ presented at the 8th Int. Conf. on Accelerator and Large Experimental Physics Control Systems, pp. 339-343, Oct. 1999.

[10] D.J. Gaushell and W. R. Block, **“**SCADA communication techniques and standards,“ IEEE Computer Applications in Power, vol. 6, pp. 45-50, Jul. 1993.

[11] A. A. Creery and E. J. Byres, “Industrial cybersecurity for a power system and SCADA networks - Be secure, ” vol. 13, pp. 49-55, Jul. 2007.

[12] R. L. Krutz, “Securing SCADA Systems,” Wiley Publishing, Inc., 2005.

[13] M. Berg and J. Stamp, “A Reference Model for Control and Automation Systems in Electric Power,” Sandia National Laboratories report SAND2005-1000C, Oct. 2005.

[14] V. M. Igure, S. A. Laughter and R. D. Williams, “Security issues in SCADA networks, “Computers & Security, vol. 25, pp. 498-506, Oct. 2006.

[15] Office of Cybersecurity, Energy Security, and Emergency Response, “21 Steps to Improve Cyber Security of SCADA Networks,” Jun. 2011.

APPLICATION OF AI IN DATA PROTECTION IN THE BUSINESS ENVIRONMENT

NEMANJA VESELINOVIĆ

Union - Nikola Tesla University, Faculty of Information Technologies, nemanjaveselinovic.sid@gmail.com

MILOŠ JOVANOVIĆ

Belgrade Metropolitan University, Faculty of Information Technologies, milos.jovanovic@metropolitan.ac.rs

IGOR ORLIĆ

Freshfields Bruckhaus Deringer, London, United Kingdom, igor.orlic@freshfields.com

MILOŠ S. DRAŽIĆ

Belgrade Metropolitan University, Faculty of Information Technologies, milos.drazic@metropolitan.ac.rs

ACA ALEKSIĆ

Academy of Technical and Art Applied Studies, School of Electrical and Computer Engineering, aca57aleksic@gmail.com

Abstract: *Artificial Intelligence (AI), which represents the future of information technologies, is considered by most scientists to be a great danger for data protection. However, AI also offers many possibilities for its improvement in the future. This paper discusses existing machine learning technologies that protect users' privacy, their application in business, future technologies, as well as privacy tools based on AI. Also, the paper talks about the areas of compliance with the General Data Protection Regulation (GDPR) where technologies based on the rules of machine learning techniques may be relevant.*

Keywords: *AI, Data Protection, GDPR, Polisis, Homomorphic Encryption, Secure Multiparty Computation, Privacy bots*

1. INTRODUCTION

It is well known to all of us that with the advancement of any modern technology of the future, data protection has the greatest consequences. Computer security (Cyber Security) has the very difficult task of dealing with a large number of abuses of systems, networks, programs, and devices. With the advent and advancement of digitalization, artificial intelligence is expanding and is one of the biggest problems for Cyber Security.

Artificial intelligence is a scientific field in computing whose goal is to develop software that enables computers to understand and imitate human intelligence. Due to its virtues and various applications, artificial intelligence is a very important item in the smart business [1], [2]. This is because it suits companies when they have a large amount of customer and market data. Problems arise due to customer privacy because companies use data to get more accurate information. Industries such as banking, trade, and health benefit the most, as these sectors have a high level of vulnerability to cyber attacks [1]. Lately, we have been recording an increasing number of online attacks around the world. Because of this, it is very important to have systems and processes that protect personal data.

Despite the huge challenges, machine learning and artificial intelligence are improving their strategies to make personal data as secure as possible from attack.

2. MACHINE LEARNING ON ENCRYPTED DATA

Machine learning occupies a significant place in the world of information technology, especially when it comes to analyzing large amounts of data and answering queries from huge databases. Machine learning belongs to the field of artificial intelligence and refers to building and refers to building and study systems that can learn and draw logical conclusions from the mass of data [3]. Machine learning plays a significant role in the referral system and the business world. One of the types of machine learning is mainly used – one discrete value, a class [3].

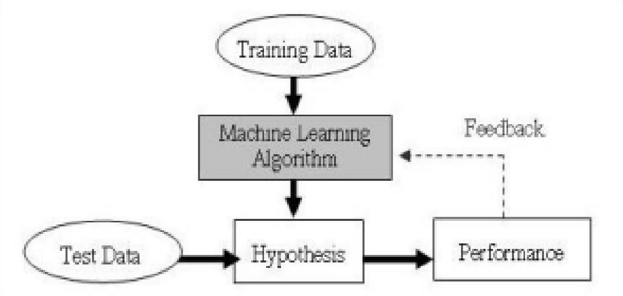


Image 1: Machine learning process

For retailers, one of the sore points, such as creating loyal and long-lasting customer relationships and creating a loyal customer database, can be overcome by using predictive analytics [4]. To build and maintain a customer loyalty database, it is necessary to take care of each customer. Using models that can find correlations in different data sources and point out warnings about specific customers or transactions could be of the utmost importance to companies. On the one hand, it is useful for the company, and on the other hand, it has many advantages for end customers [4].

The two most promising technologies for machine learning of encrypted data are Homomorphic Encryption (HE) and Secure Multiparty Computation (SMPC).

3. HOMOMORPHIC ENCRYPTION

Encryption is a well-known technique by which keeps confidential sensitive information. One of the limitations of this technique is that the information system can generally only store such encrypted data and to make it on demand the user is forwarded in the same such form. Any processing it can only be done when the data is decrypted [5]. In recent years, there has been a noticeable trend of using commercial information systems for service storage and data processing. One of the barriers to the mass use of such systems is the danger of theft or misuse of stored data. Abuse can be prevented if the data is encrypted, but as mentioned earlier, then the system cannot process that data. The solution to this problem lies in the application of homomorphic encryption, such as an encryption scheme that allows arbitrarily complex functions to be performed over the cipher that corresponds to the same functions over the main text [5], [6]. Homomorphic encryption separates data access and data processing. The largest range of homomorphic encryption applications can be found in commercial data storage and processing services. The data can be various. Some are about membership or staff, some are medical or financial data that is extremely sensitive. Homomorphic encryption allows private queries to be made by web browsers. The user makes an encrypted query, while the browser gives the answers without first translating it into unencrypted form [6[, [7]. It is also possible to search for encrypted data. Useful stores encrypted data on a remote server and then asks it to transfer only those files that meet certain conditions, although the server itself cannot decrypt those files [8].

4. SECURE MULTIPARTY COMPUTATION

The goal of secure multi-party computation is to enable several networked parties to perform computational tasks on private information. During the calculation, neither party should be able to learn any information about the input of any other party other than what can be deduced from the results [9].

Security protocols have the task of withstanding any enemy attack. To prove that the protocol is secure, a calculation is required. The most important of these features: privacy, correctness, independence of inputs, guaranteed output delivery, fairness. Many business applications benefit from privacy as mentioned above. Some of the applications are benchmarking, auctions, and supply chain management. With SMC protocols, privacy can be improved [10].

Privacy benchmarking – Benchmarking is a comparison of performance measurement with competition statistics. Privacy is very important here because some measurements can be sensitive and should not be disclosed to the general public, such as financial data, and also bad performance can diminish the value of a brand. To preserve data, benchmarking uses SMC protocols, which calculate statistics. This protects the confidentiality of key performance indicators [10], [11].

Privacy auctions – Today's auctions are a multi-million dollar business. Privacy in this area makes it easy to find the optimal price. This is especially true for public tenders. The auction is conducted by the auctioneer placing a bid on the website and receiving bids. SMC protocols can calculate the auction winner and hide the bids of others [10], [11].

Supply chain management to preserve privacy – Joint supply chain planning can reduce costs compared to local decision-making. The problem here lies in the optimization of production and delivery among companies. If privacy is not ensured in this sector, companies are reluctant to engage in joint supply chain management. That's why we use SMC protocol services that protect privacy [10], [11].

The main purpose of the SMC is to enable a group of mutually distrustful parties to carry out the budget jointly.

Everyone should provide an entrance and get an exit. However, this should remain confidential until the end, so that there is no doubt that everyone will cooperate fairly to achieve the budget [10], [11].

Trusted party – The parties can find an external institution that they can trust to protect the privacy of their data and deliver the correct results. If a trusted party is set up, billing can be done using a protocol where the trusted party privately collects secret entries, notifies the account and the private party. The classified information used in that budget is then deleted. This protocol meets all privacy requirements of maximally complied with. Unfortunately, in practice, this is often too expensive or simply unavailable [10], [11].

Cryptographic protocols instead of the trusted side. A drastically different approach is to allow the parties to perform the calculation themselves, relying on cryptographic protocols. The results are mostly theoretical. Most of these protocols have not been used in practice and are still being tested [10], [11].

The Trust Model as a Reference Solution – The specification of the task to be accomplished is to be a simple, ideal protocol that meets all security requirements in the most obvious way, providing an incorruptible side of trust. In that case, we can define a strong notion of security [12]. This area is currently in great expansion. Although it gives very good results, the assumptions that an ideal protocol that will give 100% results in practice will be created and come to life are huge.

5. PRIVACY BOTS

The pages of data protection declarations are often blindly accepted on the Internet. As digital helpers, so-called privacy bots help strengthen the digital sovereignty of Internet users by automatically assessing data protection statements. This should help you get a quick overview of the Internet Service Privacy Policy [13].

The basis for the concept of the privacy bot and its easy-to-understand the preparation of the data protection statement is a clear division of the content into criteria that can be assessed. The basic function of the privacy bot is one time, but flexible storage of individual data protection preferences. The privacy bots allow Internet users to create multiple data protection profiles. For example, a relatively worry-free data protection profile for online shopping and a very strict data protection profile for online banking. In this way, Privacy Bot takes into account different data protection requirements. To create a data protection profile, an Internet user once registers on the privacy bot's website. After storing at least one data protection profile, Privacy Bot allows you to check Internet services against saved data protection settings. The previously presented concept of the Privacy Bot already taking the first step towards strengthening the digital sovereignty of Internet users. A useful further development would be the implementation of an identity management solution or even a data vault for the individual issuance of personal data. If the Privacy Bot is offered by a government institution, scientific organization, or relatively reliable company, implementing an identity management solution is the initial further development of the data [13].

Finbots – These are bots built on automated decision technology, which aim to facilitate banking and support clients in making financial decisions. The prevalence of these bots is on the rise, to the extent that they are designed to mimic human social rules, expectations, and norms, reducing the need for human-to-human interaction. These bots improve the current state of consumer confidence and adoption rates, so banks and similar institutions are investing heavily in creating these machines. Although they do not guarantee complete security privacy, they have proven to be a good solution. According to some research and surveys, these bots have neither increased nor decreased privacy concerns, although they have increased the perception of social presence [14].

6. POLISIS

Privacy policy – Many have gone through this using an internet browser without paying attention or peeking into what is inside. Big mistake. Within this textual content is all information about how companies collect, store and manage our personal information. If everyone read in more detail, most would not like at all what they confirm every time they skip or agree to the terms of use. The problem is that researchers are not skilled enough to process and understand the content of privacy policy, especially because of its scope. Despite this, researchers often hire expert annotators to analyse privacy policy. The solution to this obstacle is the Polisis [15].

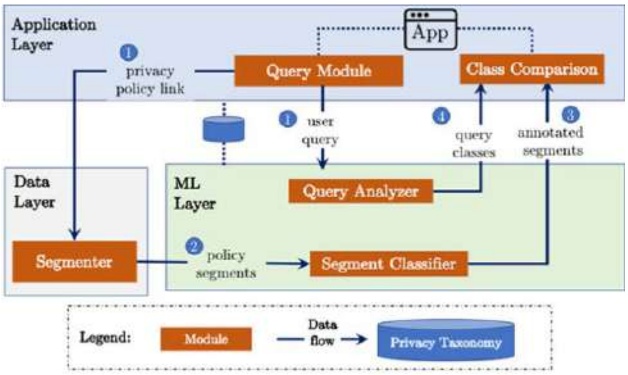


Image 2: Polisis

Polisis – Automatic analysis and presentation of privacy policies through deep learning [15].

The task of this framework is to divide the privacy policy into segments, on a finer scale. It provides a set of classes for each segment, which it further uses to enable privacy policy inquiries. It is important to emphasize that Polisis is not intended to replace privacy policies as a legal document. There are also applications that support the free examination of policy privacy, which we use to assess the usefulness of the Polisis. Structured query application – includes the selection of short notifications in the form of privacy icons from the privacy policy, which are further explored. After that, we get a solution that can automatically select the appropriate privacy icons from the privacy policy. Another application deals with free-form queries. Here we evaluate the PriBot question-answer system for a privacy policy. The Polisis consists of three layers. Application layer, Data layer, and Machine Learning layer [15].

The Application layer provides policy information and allows users to query. The query module receives the user query. This is then passed on to the lower layers, to eventually get built-in privacy classes. When resolving a users' query, the comparison module detects segments with privacy classes that correspond to the query classes.

The Data layer is divided into three phases: separation, list handling, and segmentation

The Machine Learning layer is responsible for producing notes of data segments. It uses segments from the data layer and a query from the application layer [15].

This was a brief review of the Polisis which is a great solution to help all users and researchers understand and process the content and scope of the privacy. Polisis also allows for more applications and is suitable for upgrading.

7. GDPR

It is completely meaningless to talk about data protection without mentioning the GDPR because the compliance of all these actions with it is very important.

GDPR stands for "General Data Protection Regulation". It's about a new act protecting the privacy and personal data, applicable in all 29 EU Members States. The regulation was adopted on May 27, 2016. That is when the two-year transitional period began and the said Regulation entered into force on 25 May 2019. It should be noted that, unlike the directive, the regulation has direct effect and is directly applicable in all EU Member States without the need to enact implementing legislation. This regulation introduces changes that will affect the operations of most companies, especially the part related to data protection and processing. Not all companies process the personal data of their clients for profit, but all companies process the data of their employees and candidates for a job. This part of the business must comply with the GDPR [16].

Despite the fact that the regulation is directly applicable, EU Members States have committed themselves, given the sensitivity and complexity of the scope, to enact implementing the law [16], [17].

A respondent is a natural person whose data is collected and processed. He has the right at any time to obtain information about his personal data held by the organization and for what purpose it is used, may request the deletion of all personal data, and may institute legal proceedings if he considers that personal data have been used outside the GDPR [16], [17].

The data owner is the organization that collects the data of one or more respondents and is responsible for compliance with the GDPR [16], [17].

A Processing manager is a natural or legal person who alone or in cooperation with others determines the purpose and means of personal data processing [16], [17].

The Data Protection Officer is the person in charge of controlling and complying with the GDPR. According to the GDPR, an official need to be add pointed if the organization has more than 20 employees, if it has and processes a large amount of personal data or if it processes personal data for another organization [16], [17].

The application of the GDPR applies only to personal data. This data includes data that can be used to establish one's identity with a high probability. The protection of other data which is not included in personal data is regulated by the national legislation of the Member States [17].

8. CONCLUSION

The development of technology greatly facilitates our lives in the future. Every new technology that appears brings with it a huge number of advantages, but unfortunately, the disadvantages are what first catches the eye. People as people, look at everything to abuse. The same situation is with artificial intelligence. The area that gives us so many advantages, is a very big problem for the future because of its shortcomings when it comes to privacy. Fortunately, the power of this technology is so great that it brings with it numerous solutions. The fact is that people are not sufficiently informed about the risk of leaving their personal information on the Internet, regardless of the fact that no one is forcing us to do so. It is very important that the number of people working on data security solutions is higher than those who are trying to abuse it because data protection is the first wall of defense against crime of today.

REFERENCES

[1] C. Meurisch and M. Mühlhäuser, "Data Protection in AI Services: A Survey," ACM Compt. Surv., vol. 54, pp. 1-38. Mar. 2021.

[2] C. Kuner, F. H. Cate, O. Lynskey, C. Millard, N. Ni Loideain and D. J. B. Svantesson, “Expanding the artificial intelligence-data protection debate,” Int. Data Priv. Law, vol. 8, pp. 289-292, Nov. 2019.

[3] T. Graepel and K. Lauter and M. Naehrig, “ML confidential: Machine learning on encrypted data,” in Proc. ICISC 2012, 2013, pp. 1-21.

[4] A. Wiesberg,. “Machine learning on encrypted data,” Doctoral dissertation, Universität Mannheim, Mannheim, Germany. Retrieved from https://madoc.bib.uni-mannheim.de/46375/.

[5] M. Naehrig, K. Lauter, K. and V. Vaikuntanathan, “Can homomorphic encryption be practical?,“ in Proc. CSCW’11, 2011, pp. 113-124.

[6] C. Gentry, “Fully homomorphic encryption using ideal lattices,” in Proc. STOC’09, 2009, pp. 169-178.

[7] L. J. Aslett, P.M. Esperança, and C. C. Holmes, “A review of homomorphic encryption and software tools for encrypted statistical machine learning,” arXiv preprint arXiv:1508.06574, Aug. 2015.

[8] K. Mallaiah and S. Ramachandram, “Applicability of homomorphic encryption and CryptDB in social and business applications: Securing data stored on the third party servers while processing through applications,” Int. J. Comput. Appl., vol. 100, pp. 5-19, Aug. 2014.

[9] Y. Lindell and B. Pinkas, “Secure multiparty computation for privacy-preserving data mining,” Journal of Privacy and Confidentiality, vol. 1, pp. 59-98, Apr. 2009.

[10] P. Laud and L. Kamm, eds., “Applications of secure multiparty computation,” IOS Press BV, 2015.

[11] W. Du, and M. J. Atallah, “Secure multi-party computation problems and their applications: a review and open problems,” in Proc. NSPW’01, 2001, pp. 13-22.

[12] D. Bogdanov, R. Talviste and J. Willemson, “Deploying secure multi-party computation for financial data analysis,” in Proc. FC 2012, 2012, pp. 57-64.

[13] H. Harkous, K. Fawaz, R. Lebret, F. Schaub, K. G. Shin and K. Aberer, “Polisis: Automated analysis and presentation of privacy policies using deep learning,” in Proc. SEC’18, 2018, pp. 531-548.

[14] J. Kingston, “Using artificial intelligence to support compliance with the general data protection regulation,” Artif. Intell. Law, vol. 25, pp. 429-443, Dec. 2017.

[15] C. Addis and M. S. Kutar, “General Data Protection Regulation (GDPR), Artificial Intelligence (AI) and UK organisations: a year of implementation of GDPR,” in Proc. UKAIS2020, 2020, pp. 1-24

[16] M. Ng, K. P. Coopamootoo, E. Toreini, M. Aitken, K. Elliot and A. van Moorsel, “Simulating the effects of social presence on trust, privacy concerns & usage intentions in automated bots for finance,” in Proc. EuroS&P 2020, 2020, pp. 190-199.

[17] E. Graeff, “What we should do before the social bots take over: Online privacy protection and the political economy of our near future,” presented at Int. Conf. Media in Transition 8: Public Media, Private Media, Cambridge, Massachusetts, United States, May 3-5, 2013.

A NEW CRYPTOGRAPHIC ALGORITHM BASED ON AFFINE CONNECTION COEFFICIENTS

DUŠAN J. SIMJANOVIĆ

Belgrade Metropolitan University, Faculty of Information Technologies, dusan.simjanovic@metropolitan.ac.rs

NENAD O. VESIĆ

Mathematical Institute of the Serbian Academy of Sciences and Arts, n.o.vesic@outlook.com

BRANISLAV M. RANĐELOVIĆ

Faculty of Electronic Engineering, University of Niš, branislav.randjelovic@elfak.ni.ac.rs

NEMANJA ZDRAVKOVIĆ

Belgrade Metropolitan University, Faculty of Information Technologies, nemanja.zdravkovic@metropolitan.ac.rs

ĐORĐIJE VUJADINOVIĆ

Faculty of Science and Mathematics, University of Montenegro, djordjijevuj@ucg.ac.me

Abstract: *In the era of novel information and communication technologies, as well as everyday exposure to the Internet, the need to hide data and securely send information is more relevant than ever before. The importance of data security is especially reflected in the operations of state institutions, military systems, banks, and even the business sector, in which one of the goals of competition is the disclosure of information about clients. With the increase of computational power, system vulnerabilities and data breaches indeed present a high-risk factor in data security. In this paper, we present an algorithm for data encryption and decryption based on the Christoffel symbols, which uses structures from differential geometry. Quadratic functions and uncountability of space of functions make encryption process safer and of better quality.*

Keywords: *Affine connection coefficients, Asymmetric cryptography, Christoffel symbols.*

1. INTRODUCTION

In many situations, it is necessary to hide and send a message without anyone's knowledge. It has been so since antient times, used in Greek Skytale, in the Gaius Iulius Caesar era, substituting/ shifting each letter for k positions [1], substituting more letters in Heidenberg's Tritemius [2], the Thomas Jefferson disks system [3], Schrebius's Enigma [4] and in many other cases since far.

In the digital era, the need for secure transmission of messages has become more important than ever before. As more and more interaction is carried over, in general, unreliable channels, those messages need to be protected in such a manner that only the sender and recipient know it’s contents.

One of the important questions of the 21st century, together with the increasing use of the Internet, is the potential for secret and secure data exchange. Security indicates the protection of information from unlawful revelation or modification, while secrecy applies to the individual's decision whether to and to what extent it's data will be publicly available.

Three security aspects generally used are [5] attack protection, protective mechanisms, and protective service. The main task of cryptography is the investigation and application of methods used for message transmission in the form readable and comprehensible by the receiver, as well as to potentiate secure communication between sender and receiver, usually named Alice and Bob, disabling any message detection, modification, or infiltration by third person, usually named Eve.

The communication procedure between the sender and the receiver is as follows. Alice transforms the original message (plain text) into an incomprehensible message (cipher text) using a previously determined key. This message is hence sent to Bob who, knowing a key, can decode the message, and therefore read it. Eve can intercept the message, or can disguise their self as Bob in order to receive this message.

Depending on the number of keys used in encrypting/decrypting process, there are two types of cryptographic algorithms [6,7]:

- Symmetric algorithms: Both the Sender and the Receiver need to have the same key in to encrypt their messages, with a necessary condition of secure key exchange, as shown in Image 1 (a).
- Asymmetric algorithms: Both the Sender and the Receiver have a private and public key. The sender can easily encrypt the message for the Receiver, but only the Receiver has the corresponding private key to decrypt the message, as shown in Image 1 (b).

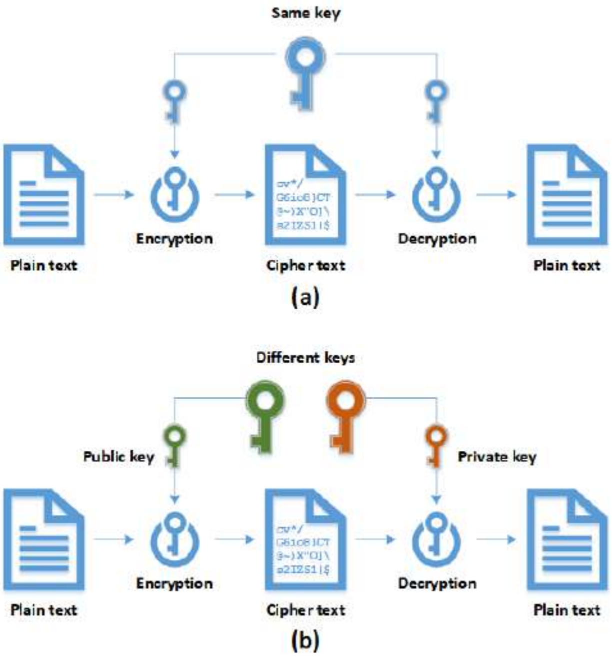


Image 1: Symmetric and asymmetric algorithms.

The advantages of public (asymmetric) cryptography are:

- The private key is the only one to be kept secret.
- The administration of keys on a network requires the presence of only a functionally trusted, not an unconditionally trusted TTP.
- A private key/public key pair may remain unchanged for significant period.
- A lot of public key schemes obtain relatively efficient digital signature mechanisms. The public key is usually smaller than for the symmetric key analogue.
- In a large network, there is a remarkably smaller number of necessary keys than in the symmetric key analogue.

The structure for the rest of the paper is given as follows. Section 2 furthers explains the necessary definitions needed to follow the geometric concept of the algorithm. Section 3 consists of the encryption algorithm itself, with the presentation of the decryption algorithm as well. Finally, we conclude with a discussion on the potential of using these algorithms with future research in mind.

2. NECESSARY DEFFINITIONS

An N-dimensional manifold $M_N = M(x^1, \dots x^N)$ equipped with a (covariant) metric tensor $\underline{\hat{g}}$ whose components are $g_{\underline{ij}} = g(x^i, x^j)$, $g_{\underline{ij}} = g_{\underline{ji}}$, is Riemannian space $\mathbb{R}_N$ (for details, see [8]). We assume that the matrix $\left[g_{\underline{ij}}\right]$ is regular, i.e. $\det\left[g_{\underline{ij}}\right] \neq 0$. The components of contravariant metric tensor $\underline{\hat{g}}^{-1}$ are $[g^{\underline{ij}}] = \left[g_{\underline{ij}}\right]^{-1}$.

If $g_{\underline{ij}} = g_{\underline{ij}}(t)$, for the variable $x^1 = t$, the space $\mathbb{R}_N = \mathbb{R}_N(t)$ is space-time. In this case, the coordinates x^i, $i = 2, \dots, N$, are constants.

The geometrical objects

$$\Gamma_{i.\underline{jk}} = \tfrac{1}{2}\left(g_{\underline{ji},k} - g_{\underline{jk},i} + g_{\underline{ik},j}\right), \qquad (1)$$

or partial derivatives $\partial g_{\underline{ij}}/\partial x^k$ denoted by comma, are the Christoffel symbols of first kind.
The Christoffel symbols of first kind for space-time $\mathbb{R}_N(t)$ are

$$\Gamma_{i.\underline{jk}} = \begin{cases} \frac{1}{2} g_{\underline{11},1}, & i = j = k = 1, \\ \frac{1}{2} g_{\underline{1i},1}, & j = k = 1, i \neq 1, \\ -\frac{1}{2} g_{\underline{jk},1}, & i = 1, j \neq 1, k \neq 1, \\ \frac{1}{2} g_{\underline{ik},1}, & j = 1, i \neq 1, k \neq 1, \\ \frac{1}{2} g_{\underline{ji},1} & k = 1, i \neq 1, j \neq 1, \end{cases} \qquad (2)$$

and $\Gamma_{i.\underline{jk}} = 0$ in all other cases.

For the known Christoffel symbols $\Gamma_{i.\underline{jk}}$, the corresponding metric tensor is

$$g_{\underline{ij}} = -2 \int \Gamma_{1.\underline{ij}} dt + c_{ij} =$$

$$2 \int \Gamma_{i.\underline{1j}} dt + c = 2 \int \Gamma_{i.\underline{j1}} dt + c_{ij}. \qquad (3)$$

The manifold M_N equipped with a symmetric affine connection ∇ whose coefficients are $L^i_{\underline{jk}}$, $L^i_{\underline{jk}} = L^i_{\underline{kj}}$, is the symmetric affine connection space $\mathbb{A}_N$ [8].

We may use the metric tensor $g_{\underline{ij}}$ for lowering of indices in $L^i_{\underline{jk}}$. The covariant affine connection coefficients of space $\mathbb{A}_N$ are $L_{i.\underline{jk}} = g_{\underline{i\alpha}} L^{\alpha}_{\underline{jk}}$.

The geometrical object $P_{i.\underline{jk}} = L_{i.\underline{jk}} - \Gamma_{i.\underline{jk}}$ is tensor. For encryption and decryption of texts, the tensor $P_{i.\underline{jk}}$ will be used.

The geometrical object

$$L_{i.\underline{jk}} = \Gamma_{i.\underline{jk}} + P_{i.\underline{jk}}, \tag{4}$$

is covariant affine connection coefficient of the space $\mathbb{A}_N$.

3. ALGORITHM

The algorithms presented in [7,9] motivated research introduced in this paper. The main purpose of this paper is to use structures from differential geometry for text-data hiding. The corresponding algorithms will be presented with possible application using matching programs writen in software package Wolfram Mathematica [10,11].

The dimension N of space-time $\mathbb{R}_N(t)$ is large enough. Linear function $b: \mathbb{N} \to \mathbb{N}$ is bijective, and array $\mathcal{A}$ composed of M rows with not necessary equal numbers of elements.

The object $\mathcal{A}_{pq}$ is the q-th element in the p-th row of the array $\mathcal{A}$. The position (p, q) of a character from the array $\mathcal{A}$ is transformed to the pair (u, v) for $u = p + M \cdot n_1, v = q + \mathcal{A}_p \cdot n_2$ for number of elements in the p-th row of array $\mathcal{A}$ equal $\mathcal{A}_p$ and $n_1, n_2 \in \mathbb{N}$. This pair is represented by complex number

$$z_{pq} = u + iv = p + M \cdot n_1 + i \cdot \left(q + \mathcal{A}_p \cdot n_2\right) \tag{5}$$

The position (p_k, q_k) of k-th character in text τ is characterized by complex number $z_k = p_k + i \cdot q_k$. The transformed position (u_k, v_k) of this character is characterized by complex number $\tilde{z}_k = u_k + i \cdot v_k$.

Let us encrypt the text τ consisted of c characters.

INPUT: Private key consisted of array $\mathcal{A}$ with M rows with not necessarily equal numbers of elements in any row, and function $b(v) = v + n_b$, for coefficient n_b, the numerical matrix $\tilde{P}_{1.\underline{\imath\jmath}}$ of the type (∞, ∞), and the text τ of c characters.

E1: For the k-th character in text τ find the corresponding position (p_k, q_k) of this character in the array $\mathcal{A}$.

E2: The pair (p_k, q_k) transform to pair $(u_k, v_k) = \left(p_k + M \cdot m, q_k + M_{p_k} \cdot n\right)$, for integers m, n and the number of elements in the p_k-th row of array $\mathcal{A}$ equal M_{p_k}.

E3: The pair (u_k, v_k) transform to polynomial $\pi_k(t) = t^2 - 2u_k t + u_k^2 + v_k^2$ (6)

E4: Create the ordered set $\Pi = \{\pi_1(t), \dots, \pi_c(t)\}$.

E5: Create $\tilde{N} = N(N+1)/2 - c$ polynomials $\tilde{\pi}_{c+u}(t) = t^2 - (r_{c+u} + s_{c+u})t + r_{c+u}s_{c+u}$, (7) $u = 1, \dots, \tilde{N}$ for integers r_{c+u}, s_{c+u}.

E6: Complement the set Π to ordered set Π^* with polynomials $\tilde{\pi}_{c+u}(t)$ before, between and after the polynomials $\pi_k(t)$. In this way, the ordered set $\Pi^* = \left\{\pi_1^*(t), \dots, \pi^*_{\underline{N(N+1)}\atop 2}(t)\right\}$ is obtained.

E7: Create the square matrix $\left[h_{\underline{\imath\jmath}}\right]$ of the type $N \times N$ whose elements are

$$h_{\underline{\imath\jmath}} = \begin{cases} \pi^*_{i^*_j}(t), & i \leq j, \\ \pi^*_{j^*_i}(t), & i > j, \end{cases} \tag{8}$$

for $i^*_j = \frac{i \cdot (i-1)}{2} + j$.

E8: Expand the matrix $\left[h_{\underline{\imath\jmath}}\right]$ to the matrix $\left[g_{\underline{\imath\jmath}}\right]$ with elements

$$g_{\underline{\imath\jmath}} = \\ = \begin{cases} p_{11}(t), & i = j = 1, \\ 0, & i = 1 \text{ and } j > 1 \text{ or } j = 1 \text{ and } i > 1, \\ h_{\underline{(\imath-1)(\jmath-1)}}, & \text{otherwise.} \end{cases} \tag{9}$$

E9: Form the matrix $\Gamma = \left[\Gamma_{1.\underline{\imath\jmath}}\right] = \begin{bmatrix} \Gamma_{1.\underline{22}} & \cdots & \Gamma_{1.\underline{2N}} \\ \vdots & \ddots & \vdots \\ \Gamma_{1.\underline{N2}} & \cdots & \Gamma_{1.\underline{NN}} \end{bmatrix}$ of the corresponding Christoffel symbols with respect to the metric tensor whose components are $\left[g_{\underline{\imath\jmath}}\right]$.

E10: From the matrix $\tilde{P}$, select the submatrix $P_{1.\underline{\imath\jmath}}$ of the type $N \times N$ from the up left angle of the matrix $\tilde{P}$.

E11: Form the matrix $L = \left[\Gamma_{1.\underline{\imath\jmath}} + P_{1.\underline{\imath\jmath}}\right]$.

E12: The components $g_{\underline{\imath\jmath}}$ of the matrix $\left[g_{\underline{\imath\jmath}}\right]$ are of the form

$$g_{\underline{\imath\jmath}}(t) = t^2 + p_{ij}t + q_{ij} \tag{10}$$

The corresponding covariant affine connection coefficients are of the form $\Gamma_{1.\underline{\imath\jmath}} = -t - \frac{1}{2}p_{ij} + P_{1.\underline{\imath\jmath}}$. The corresponding constant c_{ij} from the equation (3) is $c_{ij} = q_{ij}$.

OUTPUT: Public key $\mathrm{L}_{1.\underline{\imath\jmath}}(0) = \left[-p_{ij} + P_{1.\underline{\imath\jmath}}\right]$. Message is $\mu = \left[q_{ij} - b(0)\right]$.

To decrypt the public key, we transform the public key L to $\Gamma = \left[L_{1.\underline{\imath\jmath}} - P_{1.\underline{\imath\jmath}}\right]$. The elements of matrix Γ are the free particles of the Christoffel symbols obtained in step **E9**. In this way, we obtained the polynomials which hided the text. After solving the equations $\Gamma_{1.\underline{\imath\jmath}} = 0$, and using the operation

$$Mod1n[p, q] = \begin{cases} q, & q|p, \\ Mod[p, q], & \text{otherwise} \end{cases} \tag{11}$$

applied to real and non-zero complex parts of the solutions of previous equations, one obtains positions of characters in the matrix of characters. When conjugate characters under and on the main diagonal of matrix, we will decrypt the corresponding encrypted text.

4. CONCLUSION

The capital purpose of this paper is the presentation of an algorithm based on the affine connection ceofficients, that can be used for the text-data hiding and information processing. The shown encryption and decryption algorithms enable safe communication and message transmissions.

Since the need for secure communications is more important than ever before, the potential for using affine connection coefficients in the design of modern security protocols is high. With the rise of Industry 4.0, Internet of Things, Cyber Physical Systems, and Smart Homes, we believe that an algorithm like the one presented in this paper could be used in protocols deployed on multiple layers.

ACKNOWLEDGEMENT

This paper is partly supported by the Serbian Ministry of Education, Science and Technological Development through Mathematical Institute of the Serbian Academy of Sciences and Arts, and partly by the Blockchain Technology Laboratory at Belgrade Metropolitan University, Belgrade, Serbia.

REFERENCES

[1] A. C. Leighton, “Communication among the Greeks and Romans,” Technol. Cult., 10, pp. 149–152, 1969.

[2] G. F. Strasser, “The Rise of Cryptology in the European Renaissance. In The History of Information Security: A Comprehensive Handbook” K. de Leeuw and J. Bergstram, Eds.; Elsevier: Amsterdam, The Netherlands, 2007; pp. 277–325

[3] J. C. Galende Díaz, “Criptografía: Historia de la Escritura Cifrada,” Editorial Complutense: Madrid, Spain, 1995.

[4] A. J. Menezes, P. C. van Oorschot, S. A. Vanstone, “Handbook of Applied Cryptography” CRC Press: Boca Raton, FL, USA, 1997.

[5] D. Bohen, “Twenty Years of Attacks on the RSA Cryptosystem,” Notices of the AMS, 46(2), pp. 203-213

[6] B. Schneier, “Applied Cryptography: Protocols, Algorithms, and Source Code in C”, John Wiley & Sons: New York, NY, USA, 1996.

[7] N. O. Vesić and D. J. Simjanović, “Matrix-Based Algorithm for Text-Data Hiding and Information Processing, Military Technical Courier,” Vol. LXII (2014), No. 1, 42 – 57.

[8] J. Mikeš, E. Stepanova, A. Vanžurová, et al., “*Differential Geometry of Special Mappings*,” Palacký University, Olomouc, 2015.

[9] N. O. Vesić, N. Zdravković and D. J. Simjanović, “Securing Online Assessments using Christoffel Symbols,” The 11th International Conference on eLearning (eLearning-2020), 24-25 September, Belgrade, Serbia; 54-57

[10] P. S. Stanimirović and G. V. Milovanović, “Program Language Mathematica and Applications,” in Serbian, Faculty for Electronic Engineering, Niš, 2002.

[11] Lj. Velimirović, P. Stanimirović and M. Zlatanović, “Geometry of Curves and Surfaces Covered by Software Package Mathematica,” Faculty of Sciences and Mathematics, Niš, 2010.

THE USE OF ARTIFICIAL INTELLIGENCE TO PROTECT CRITICAL IT INFRASTRUCTURE

ACA ALEKSIĆ

Academy of Technical and Art Applied Studies, School of Electrical and Computer Engineering, aca57aleksic@gmail.com

MILOŠ JOVANOVIĆ

Belgrade Metropolitan University, Faculty of Information Technologies, milos.jovanovic@metropolitan.ac.rs

MILOŠ S. DRAŽIĆ

Belgrade Metropolitan University, Faculty of Information Technologies, milos.drazic@metropolitan.ac.rs

CHAD EHRLICH
Whitman School of Management, Syracuse University, Syracuse, NY, cdehrlic@gmail.com

BOJANA TOMAŠEVIĆ DRAŽIĆ

Belgrade Metropolitan University, Faculty of Information Technologies, bojana.tomasevic@metropolitan.ac.rs

MILOŠ MILAŠINOVIĆ

Belgrade Metropolitan University, Faculty of Information Technologies, milos.milasinovic@metropolitan.ac.rs

Abstract: *Artificial intelligence experienced rapid growth in use throughout the IT industry and it is slowly finding its way to the cybersecurity sector. With datamining serving as its backbone, artificial intelligence and machine learning are being developed for cybersecurity threat analysis, as well as being trained to use large amounts of reports to make decisions based on the assessments and categorization of the threats being presented to it. With the continuous complexification of the IT industry, it is hard to keep up with all of the weaknesses that come with that growth. Security in the IT industry should be of the utmost importance due to the role that IT serves as a support to other industries that hold critical infrastructures. The uses of artificial intelligence for the protection of critical infrastructure will be examined in this paper.*

Keywords: *Information Security, Critical Infrastructure Systems, Artificial Intelligence, Machine Learning*

1. INTRODUCTION

Artificial intelligence (AI) is a field of science in which the emphasis is on recognizing patterns and, accordingly, finding optimal solutions to complex problems. Within this technology, machines are faced with requirements of varying degrees of complexity, to which human beings are exposed on a daily basis [1]. This decision-making ability was integrated into the software through appropriate algorithms. Initially, the focus was on software solutions that were based on decision-making, but with the growing need to analyze huge amounts of data and offer timely solutions, artificial intelligence became the tool of choice. In this sense, AI has found useful value in the field of cyber security, reducing human effort and offering reliable and faster results. The use of AI in cyber security has been proven effective and the fight against spam or malware detection are just some of the examples.

2. ARTIFICIAL INTELLIGENCE, MACHINE LEARNING AND DEEP LEARNING

Alan Turing is the man who broke the Nazi encryption machine “Enigma”, he was a man of science and he served to the Allied Forces in World War II. Artificial intelligence is the branch of computing that aims to answer Turing’s famous question “Can machines think?”. The ultimate goal is an intelligent and self-sustaining machine that mimics human behaviour, capable of learning and making decisions, based on inputs coming from the external environment, without human involvement. In practical terms, AI refers to the utilization of data-driven algorithms and machine learning, in order to make automatic decisions.

Machine learning is part of the field of AI whose goal is to reach the requirements of AI through algorithms that are trained by input data. This is achieved through pattern recognition and anomalies based on a huge amount of data. The complex data structure is transformed into a model form. That is why it can be said that machine learning is a process that aims to make computers behave as if they have not been pre-programmed. As part of machine learning, deep learning is specifically focused on unsupervised computer learning, using neural networks, through reading unstructured data and forming patterns and clusters of patterns, much like neurons (nodes) in the human brain do (Image 1).

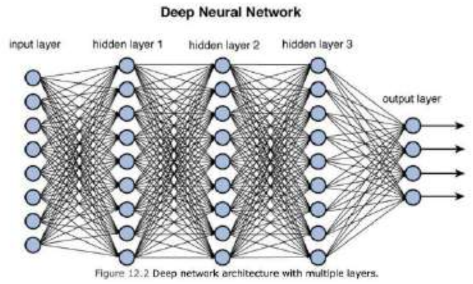


Image 1: Deep network architecture with multiple layers (Source: https://towardsdatascience.com/)

Machine learning can be used as a tool for the attackers. Namely, weaponized AI is going to be ready to adapt to the system it infects, permitting AI cyber-attacks to evade detection and maximize the damage on the targeted system. Machine learning may well be a promising tool to manage the growing range of cyber threats, however, in addition, it also acts as a tool that will be leveraged by malicious attackers. By learning from the input file, it’s aiming to specifically target weak point of the system. The extent to which AI outperforms humans can be enormous.

Security firm ZeroFOX conducted the experiment using an AI called SNAP_R, which sent spear phishing tweets to more than 800 users at a rate of 6.75 messages per minute, capturing 275 victims. In contrast, the human person sent messages to 129 users at a rate of 1,075 tweets per minute, capturing 49 victims. This is just one of the drastic indicators of the reasons why hackers are increasingly relying on AI [2]. Regardless of the motives of those who are interested in the technologies of artificial intelligence, machine learning and deep learning, it is an indisputable fact that the trend of searching for these phrases on the Internet has increased in the last few years, as can be seen in the picture below.

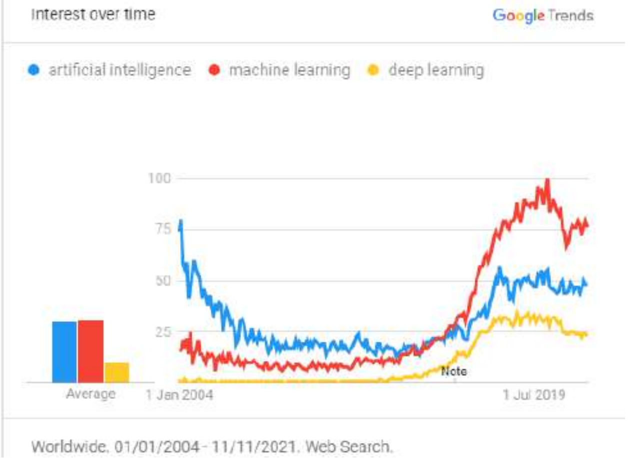


Image 2: Search trend of the terms “artificial intelligence”, “machine learning” and “deep learning” 2004-2021.

2. PREDICTIVE ANALYTICS (REGRESSION)

Analytics are used to identify anomalies in network patterns and traffic, as well as in user activities. Analytics are used to identify anomalies in network patterns and traffic, as well as in user activities. Exploits are identified by their signatures, i.e. known patterns of attack. These are recognized methods that malware or attackers use on networks. Then, when a known signature attack is noticed, the network analysis software warns the security team. Although monitoring by definition is a real-time watching of events, the bad news is that alerting to a signature attack means that an attack has already occurred. The goal is to prevent and identify patterns that would indicate the early stages of the attack, or certain activities that precede the attack, and this is a task of machine learning within AI. By analyzing all types of previous attacks, the machine can acquire the predictive ability to recognize the danger in time. An example of an ability that transcends human capabilities is the ability to recognize Advanced Persistent Threats, which is a collective name for various types of malware that appear harmless on the network, because the damage (normally data theft) is long term. The problem that human perception faces is understanding and determining what it is that we would call and define as a normal or non-threatening activity. Unlike human experience, which is subjective, the machine does not have such problems.

3. ARTIFICIAL INTELLIGENCE AS A TOOL FOR THE DEFENDERS

The growing volume and grade of cyber threats are overwhelming for cybersecurity teams. Designed and operated well, AI considerably improves detection rates through automation. It helps to generate prime quality probabilistic findings that can help the system to recognize if patterns that seem threatening are malicious or not. This can be distinctive to AI-alternative algorithms within the security infrastructure can’t do this. The other good side about AI is that it saves a lot of time, which is the most asset of a security team. The quicker a company can discover, prevent and react to each known and unknown threat, the stronger its cybersecurity stance. AI can use machine learning to create a view of “normal” activity on the network. When the AI deployed on the system detects something that does not fit the previous learned “normal” network activity it will flag it. Malware acting on a system can be distinguished from a human, that makes the ability to identify threat incredibly usefully. A person cannot access thousands of files per second, simply because we cannot click that fast. But, a chuck of malware is well capable of doing such a thing, which makes spotting it easy. Let’s make an example of a standard worker, who accesses 50 files daily. One evening, when work hours end, an account starts accessing and encrypting hundreds files per second. The AI can detect this as uncommon activity and lock the account that is encrypting the files, preventing it in further action. These types of AI security systems aren’t only capable of dealing with malware. They can also spot hardware of software failure, or them acting in suspicious ways. As an example, we can take a workplace that has security cameras all around, specifically a meeting space where major company plans are created. The AI detects that the meeting space security cameras have made a connection to an unknown IP address outside the business, and it can successfully flag it. Following the detection, the investigation discovers that the device was infected with spyware. Damage could already have been done but patching the infected IP camera prevented it from happening again.

4. DARKTRACE

The world leader in AI for cybersecurity is the company named Darktrace. Its AI model is based on the human immune system and is used by over 3,000 organizations to protect against threats to the cloud, email, IoT, networks and industrial systems. This includes insider threat, industrial espionage, IoT compromises, zero-day malware, data loss, supply chain risk and long-term infrastructure vulnerabilities. The Royal Air Force Association (RAFA) has selected Darktrace’s AI to protect its members’ sensitive data from cyber threats. Trusted with the personally identifiable information of over 72 000 former and current members of the RAFA, the association sought a security tool that could protect this sensitive data from stealthy attackers as well as insider threats. By deploying Darktrace’s AI, the RAFA benefits from a self-learning technology that identifies threats that bypass traditional security tools [3].

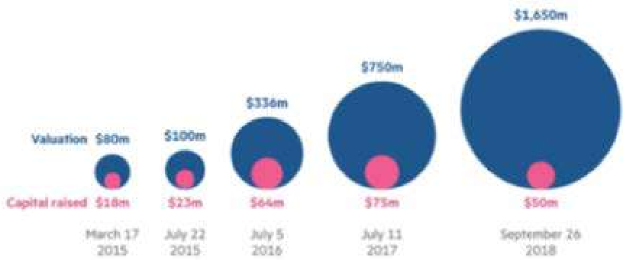


Image 3: The rise of Darktrace (Source: https://www.crunchbase.com/)

Darktrace Antigena represents the first application of automatic response technology in the enterprise. The software the company developed has evolved to the level where it fights back at machine speed, acting so that the threat does not spread through the system. When the system recognizes the threat, Darktrace Antigena reacts in a few seconds and takes appropriate steps to neutralize the threat and give the security team enough time to catch up. For example, a ransomware that can infect dozens of computers in a few minutes will be detected by this technology and placed in a container in about 2 seconds. This prevents it from spreading further beyond the initial point of attack [4].

5. INTRUSION DETECTION AND PREVENTION SYSTEM (IDSP)

An intrusion detection and prevention system (IDPS) is software or a hardware device installed inside the network that can detect possible intrusions and is able to prevent them. IDPSs provide four functionalities that are crucial for security: monitoring, detection, analysis and response to unauthorized activities. Some of desired characteristics of an IDPS system so it can provide efficient security against serious attacks are [5]:

- It should work in real-time and detect intrusions either at the time they occur or with very little delay.
- It should recognize any real intrusion and keep the number of false-positive calls to a minimum.
- It should work continuously with a minimum of human supervision.
- It should be fault-tolerant and able to recover from a system crash whether the cause of the crash was an accident or malicious activity.
- It should be able to monitor itself in order to detect whether he has been modified by the attacker.
- It should be configurable in accordance with the security policy of the system it monitors.
- It should be adaptable to changes in the system as well as to changes in user behaviour over time.

IDPS consists of traditional and advanced components and a typical architecture can be seen in Image 4.

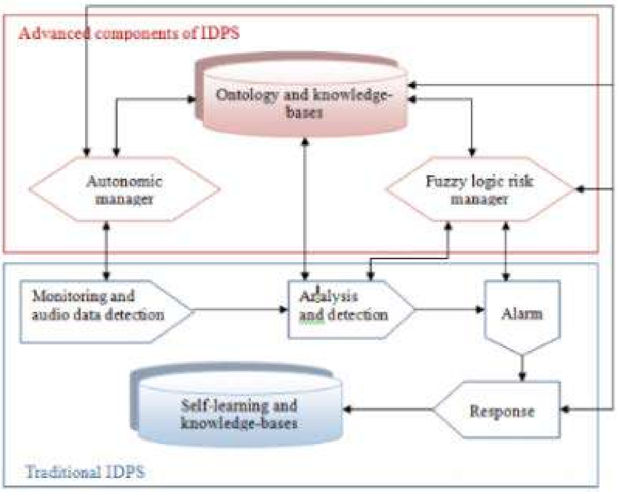

Advanced components of IDPS
Ontology and knowledge-bases
Autonomic manager
Fuzzy logic risk manager
Monitoring and audio data detection
Analysis and detection
Alarm
Self-learning and knowledge-bases
Response
Traditional IDPS

Image 4: Typical IDPS [5]

6. ADAPTIVE HONEYPOTS AND HONEYTOKENS

In cyber security, what is called digital honey is actually a type of bait. Traps, known as honeypots and honeytokens, are designed to deter attackers. lead to the wrong side. Honeypots are computer systems on a network that are apparently full of data that can be easily stolen, but in reality it is just a lure. An ordinary, normal user will not even try to move towards a bait like this, but the one who tries to approach it, triggers an alarm. In general, honeypots and honeytokens can be computers, passwords, as well as all other false information and resources that are placed on a network to collect information about the attack and consequently about the attacker himself. Since this is an approach that is known to the attackers themselves, it is necessary to make such baits as similar as possible to real data and resources, so that the attacker would not easily recognize that it is a bait. For this purpose, more advanced versions have been developed, the so-called adaptive honeypots and honeytokens. Such solutions are able to adapt and change their behaviour depending on the nature of the attack. In this way, the attacker will discover much more about himself. For example, an attacked bait will react in a way that a well-protected resource would react. Based on the actions that the attacker carries out, and in order to disable the defence, a lot can be learned about his abilities, as well as the tools he uses. Here, AI can significantly help because based on this data, one can learn about malicious behaviour that can be recognized and responded to in a timely manner [7].

7. SECURITY INFORMATION EVENT MANAGEMENT

Defence cyber systems collect a huge amount of data, which floods security analysts with alarming events. One such system is the Security Information Event Management (SIEM). To improve the work of defence systems, data science is used to correlate events, recognize patterns, and detect abnormal behaviour. SIEM represents a kind of pc security software system that combines real-time IT infrastructure monitoring with the gathering and analysis of log data and event data collected from different components of the infrastructure environment such as network devices, servers, domain controllers, and more. SIEM than stores and applies analytics to the data stored to discover trends and detect threats.

For intelligent systems to partner with us in the workplace or elsewhere, they must genuinely understand us and then communicate what they are doing back to us in a way that we can comprehend. Under normal conditions, IT security teams store security logs into a central database, to be subsequently analyzed for security purposes. This procedure takes a lot of time. Machine learning does this kind of work in real time, giving people space and time to focus on more important tasks, such as investigating legitimate security events. Also, machine learning can scan the entire network of an organization far faster than a human. This allows SIEM or other security analytics solutions to more quickly identify and capture threats and data breaches. Data breaches are malicious events, with low-profile nature and untimely detection is very common [8]. As we discussed, machine learning involves the use of algorithms to give computer systems the ability to learn automatically and improve from experience without explicitly programming these systems. While machine learning applied in SIEM is still a growing field, there is a clear potential for this branch of Al to augment SIEM capabilities. An effective attack vector utilized by cybercriminals is through the compromise of user credentials. Advanced persistent threats, in which an attacker gains access to a network and remains undetected, usually with the use of phishing and social engineering to gain user credentials, that kind of attacks can cost millions of dollars and potentially lead to compromise of sensitive information. By feeding machine-learning model with files, for example, user behavior, together with location, login times, usage habits, and history, the model will become skillful at establishing an "image" of what is considered a standard user behavior across your systems. Not only does machine learning help better identify stealth types of attacks like APTs, when combined with SIEM, machine learning can even offer predictive analytics, where intelligent analysis and learning techniques will help in predicting and stopping future attacks on critical IT infrastructure.

8. USER AND ENTITY BEHAVIOR ANALYTICS

User and Entity Behavior Analytics (UEBA), a term coined by Gartner in 2015, helps us with a set of processes to detect targeted attacks, insider threats, and financial fraud utilizing advanced analytics (e.g. Machine Learning). In UEBA, we look for patterns in users, devices, servers, applications, DLP (data loss prevention), IAM (Identity and Access Management), network flow data, etc.), to calculate risk to identify anomalies against the baseline. With the use of Big Data, UEBA can look for anomalies across diverse data sets at scale. In a typical web application environment, we look at the following checkpoints for user profiling in UEBA related information namely [9]:

- Registration - Process for signing up new credit cards or new checking accounts.
- Logins - For account authentication, time of day, of attempts, and devices used.
- Transactions - completing a financial transfer or completing a purchase online.
- Logouts - User behavior during logouts (do they typically log off, or close the browser or leave it to expire).

UEBA has its differences from a rules-based approach since it uses Machine Learning and advanced analytics for the system so it can learn and automatically detect anomalies from a group of user profiles in a very short amount of time. With the right data and Machine Learning models, we can improve the precision and help UEBA reduce the number of "False Positives" also accelerate the investigation of threats. UEBA uses the following key components that utilize Machine Learning techniques to continuously learn and build real-time generic user profiles to detect anomalies. From a Machine Learning context, we could utilize simple SVD (singular value decomposition) algorithms to classify account types (user, service, bots, etc.) and build account behaviors from log files. Typical attributes include a maximum number of connections during peak load, number of connected endpoints, applications accessed, location and more. Apart from account types, we also need Machine Learning to help us classify other entity types, namely between a server node and a user desktop using behavioral attributes based on the activities performed on these assets. The use of recommendation algorithms in identifying a peer group of users based on their prior activities is a great example of utilizing Machine Learning in the context of UEBA. Imagine if a user accesses a laptop or desktop for the first time with lots of activity (seems suspicious at the face of it), we could potentially correlate this with a peer group (baseline) profile to help reduce false positives significantly [9].

The insider threat is the primary UEBA driver for many security teams. The reason for this is insufficient trust in the mechanisms of precise detection when an insider threat actually occurred. These threats can be malicious insiders, compromised insiders and negligent insiders. As a result of their actions, data can be compromised in various ways, and malicious behaviour can lead to data destruction. Therefore, it is crucial to recognize the pattern of normal user behaviour and, by forming an appropriate baseline, detect and alert to unusual and high-risk behaviour that deviates from the baseline profile based on factors such as time, source host and location.

9. CONCLUSION

Growing integration of information and operational technologies, i.e. IT / OT convergence, on the one hand, and intensified cyber-attacks, on the other, have led to an increase in very complex security challenges. Artificial intelligence can greatly mitigate existing and future risks. Whether companies and organizations are still not using AI for this purpose or have started using it, some crucial steps need to be taken to see the potential that AI has in the fight against cybercrime. To begin with, it is important to properly assess where the placement of AI in the entire structure of cyber security will bring the most benefits. Organizations need to build a roadmap that addresses infrastructure, data systems, applications, management, best practices, and use case selection and implementation. Governments and businesses must be more flexible than ever when dealing with today's cyber threats. Like most technologies, the capabilities that Al provides us can be used for both defence and attack, and their success depends on the strategies underlying them. We have learned that the usage of Al is based on learning from data that the enterprise infrastructure generates, without it even the most advanced Machine Learning and Al implementations quickly lose their accuracy and usability. The access to data is critical for maintaining these types of security systems. The vulnerabilities that Al systems possess are not yet totally understood. It can be compared with the development of antibiotics, formulating defences against Al-driven cyberattacks is likely to be expanding and will have to be carefully considered. Attacker Al will be able to adapt, just as bacteria can adapt to antibiotics. The best way to defend our Al is to provide it with many scenarios and a huge volume of data so it can plan a response accordingly.

REFERENCES

[1] R. A. “The Impact of Artificial Intelligence,” 23 September 2019., [Online]

[2] N. Ismail, “AI in cybersecurity a new tool for hackers?” https://www.raconteur.net/technology/cybersecurity/ai-cybersecurity/ , Feb. 26, 2019.

[3] Cambridge, UK, “Darktrace AI Used to Protect Military Personnel Data,” https://www.darktrace.com/en/press/2019/284/ , May 20, 2019.

[4] “Darktrace Antigena: The Future of Al-Powered Autonomous Response,” https://www.aquion.com.au/wp-content/uploads/2018/11/wp-antigena.pdf

[5] S. Dilek, H. Çakır and M. Aydin, “Applications of artificial intelligence technique combating cyber crimes: A

review,” International Journal of Artificial Intelligence & Applications, vol. 6, pp. 21-39, Jan. 2015.

[6] A. Patel, Q. Qassim, Z. Shukor, J. Nogueira, J. Júnior3 and C. Wills, “Autonomic agent-based self-managed intrusion detection and prevention system,” Proceedings of the South African Information Security Multi-Conference (SAISMC 2010), pp. 223-234, May 2010.

[7] T. Coombs, “Artificial Intelligence & Cybersecurity for dummies,” John Wiley & Sons, Inc., 2018.

[8] B. Canner, “Machine Learning, SIEM, and Security Analytics: What to Know,” https://solutionsreview.com/security-information-event-management/machine-learning-siem-security-analytics-know/ , Apr. 12, 2018.

[9] S. G. Gopalakrishnan, “Data Science & Machine Learning in Cybersecurity,” 22 May 2017. [Online]

FRAUD DETECTION AND MALICIOUS CODE INJECTION ANALYSIS IN AUTOGRADING SYSTEMS

NIKOLA DIMITRIJEVIĆ

Faculty of Information Technologies, Belgrade Metropolitan University, Serbia, nikola.dimitrijevic@metropolitan.ac.rs

ALEKSANDAR MESTEROVIC

Department of Security Studies and Criminology, Faculty of Art, Macquarie University, Sydney, Australia, aleksandar.mesterovic@students.mq.edu.au

MILENA BOGDANOVIĆ

Faculty of Information Technologies, Belgrade Metropolitan University, Serbia, milena.bogdanovic@metropolitan.ac.rs

NEMANJA ZDRAVKOVIĆ

Faculty of Information Technologies, Belgrade Metropolitan University, Serbia, nemanja.zdravkovic@metropolitan.ac.rs

Abstract: *In recent years, e-learning systems for teaching programming languages have been at a steady rise in popularity. These systems can be a part of commercials learning platforms, Learning Management Systems, commercial learning platforms, or even deployed as individual projects, and they vary in quality and presentation. Systems that employ autograders are considered better in quality, as they automatically inform the student if their entered code is valid or not. However, issues such as fraud detection and malicious code injection present a potential security risk. In this paper, we analyze these potential security issues in e-learning systems with autograders, and present possible solutions for this type of vulnerability.*

Keywords: *Autograders, e-Learning, Fraud Detection,*

1. INTRODUCTION

In the past decade, we are witnessing a high rise in demand for experts in software development, and programmers in general. Employers are looking not only for graduate students coming from Science, Technology, Engineering and Mathematics (STEM) universitites, but also for experts which have completed courses from Massive Open Online Course (MOOC) and different commercial learning platforms as well [1]. The ongoing COVID-19 pandemic showed us that learners can learn from home, i.e. the necessity for online and blended learning required a shift from traditional towards web-based learning, by applying new methodologies such as automatic graded systems (autograders), virtual and augmented reality systems, and an overall gamification of learning [2]. Indeed, contemporary systems for learning programming langugages with autograder support which are developed as a web-based application, help those learners to complete their assignments at home, often without the need of a high-performance computer [3].

With the exception of some Higher Education Institutions (HEIs), assignments and tests submitted by the learners from home are often not supervised, and learners are „free“ to submit arbitraty code. This is especially true for the case when learners are interacting with an autograder system with a broswer-based code editor. We identify multiple security issues with this approach. Namely, learners can input copied code, obtained from the Internet or from another learner, and therefore plagiarise their assessment input. Furthermore, a learner can input malicious code which can cause harm to the server running the autograder.

In this paper, we investigate and categorize possible harmful attacks to autograder-supported e-learning systems (including fraud detection), identify possible vulneravilities to these systems, and propose possible solutions to mitigate and/or prevent them.

The rest of the paper is organized as follows. In Section 2, we provide a nessesary overview to autograder systems and their use in an e-learning environment. Section 3 investigates the security issued in these systems. Section 4 discusses possible solutions and mitigation strategies, and a summarizing discussion is given in Section 5.

2. AUTOGRADERS

The concept of using a computer-based tools to help assist in grading learner's assignments been present over half a century, mainly in HEIs [4]. Autograders first and foremost help the teaching staff (professors and teaching assistants) by reducing the work load of having to manually grade all students' assignments. In addition, the bias of the teaching staff is also removed. This is especially important for STEM students, as the majority of the assignments are in the form of writing a computer program to run a specific task. Autograders have evolved with the emergence of new technologies and programming languages, and can be in general categorized in three generations [5]. The first generation of autograders was the simplest form of these systems. These systems were mostly tied to lower-level programming languages. An autograder would yield only "correct" or "incorrect" as an answer to its input. This was achieved by checking a strict set of successive instructions written by the learner. The second generation of autograders often employed different tools, and object-oriented programming languages such as C++ or Java were used to build the autograders. These systems would be able to check problems written in their respective language. Finally, the third generation of autograders emerged with the rise of the modern web-based software development technologies. In such systems and platforms, a web application is hosted on a server, and is presented to the learned with a graphic user interface (GUI), as shown in Image 1. The learner can therefore write their program in a browser, without the need of a compiler or an (often proprietarily) integrated development environment (IDE) installed on their home computer.

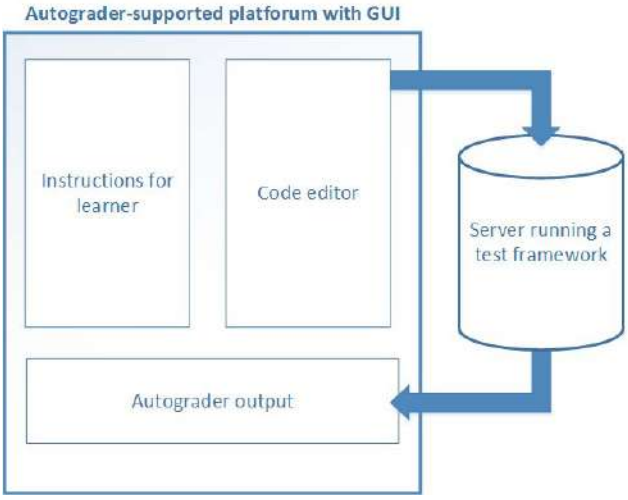

Autograder-supported platforum with GUI
Instructions for learner
Code editor
Server running a test framework
Autograder output

Image 1: Components of an autograder-supported learning system.

This approach is especially helpful for online studies or when taking a course from a MOOC or another commercial learning platform. This generation of autograders often may have support for different programming languages and several programming paradigms, such as functional programming.

However, autograder-supported systems have several issues, which can be categorized as pedagogical and technical [5]. The authors of this paper will focus on technical issues include regarding overall security.

3. SECURITY ISSUES

In this section we take a look at the most common security issues with programming language learning platforms with support for autograders.

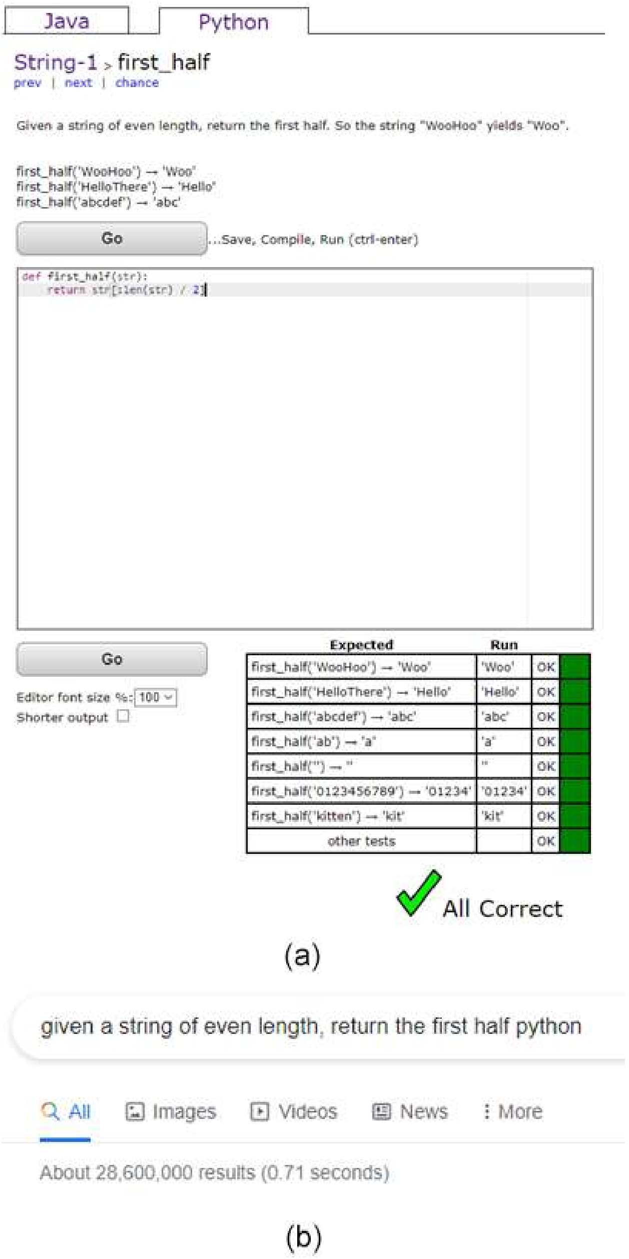
Fraud detection

When discussing fraud detection in these systems, we can identify two categories, as pointed out in [6, 7]:

- Plagiarism, when the learner is using someone else's content as their own,
- Collusion, when assignments that were supposed to be conducted individually are done by two or more students.

Autograder systems are susceptible to both vulnerabilities. Indeed, as beginner assignments are often well-known programming problems, learners can obtain part of the solutions and even whole solutions fairly easily on the Internet. In addition, learners, although more often in HEIs than in MOOCs or commercial platforms, tend to collaborate and can easily share their results online. Differently from plain text, the input to an autograder is computer code, which is, due to the syntax and semantics rules for each programming language, more formal and therefore harder to detect fraudulent or stolen code. One must pay more attention to code snippets that are syntactically different, but can be semantically the same [6]. For instance, obtained code can be altered by including white spaces, comments, or changing variable names; however, the same functionality is preserved.

As an example, we provide an assignment from CodingBat, one of the most popular free autograder platforms available for Java and Python [8]. One of their assignments regarding string manipulation is to complete the given code to define a function to return half of a string. The complete result for this assignment can easily found with a web search, just by typing the instructions from the platform and the corresponding programming language. In this example, we used the first five distinct results from the search, and all five approaches pass all tests. The obtained code from the first result passes all tests, as can be seen in Image 2 (a) on the next page. Furthermore, Image 2 (b) shows that an ordinary search for the solution yields around 28.6 million results in less than a second.



Given a string of even length, return the first half. So the string "WooHoo" yields "Woo".

first_half('WooHoo') → 'Woo'
first_half('HelloThere') → 'Hello'
first_half('abcdef') → 'abc'

```
def first_half(str):
    return str[:len(str) / 2]
```

Expected	Run		
first_half('WooHoo') → 'Woo'	'Woo'	OK	🟩
first_half('HelloThere') → 'Hello'	'Hello'	OK	🟩
first_half('abcdef') → 'abc'	'abc'	OK	🟩
first_half('ab') → 'a'	'a'	OK	🟩
first_half('') → ''	''	OK	🟩
first_half('0123456789') → '01234'	'01234'	OK	🟩
first_half('kitten') → 'kit'	'kit'	OK	🟩
other tests		OK	🟩

given a string of even length, return the first half python

About 28,600,000 results (0.71 seconds)

Image 2: Web search for the correct assignment result yields results that pass all tests with an autograder.

To detect fraudulent code, several well-known algorithms exist [9, 10]. First are the text-based techniques, which apply no text transformation before inspecting the code, and only comments and white space is ignored. Next, token-based techniques apply tokenization, i.e. lexical analysis, and use the tokens for clone detection. More complex fraud detection algorithms include abstract syntax trees (ASTs) and program dependence graphs (PDGs). The former constructs an AST as a syntactical representation of the source code, while PDGs contain control flow and data flow information upon which comparison is based. Finally, a more mathematical approach to fraud detection are metrics-based techniques which are close to hashing algorithms. Each technique performs several steps before comparing the submitted code to a known answer, and these steps often include code division into parts, variable and function name replacements, and removing additional input such as comments, blank spaces and lines. Upon comparing with a known answer, similarities can be found and weighted. Furthermore, it is important to track the time needed for a learner to complete an assignment, as there exists a possibility to obtain the completed solutions and use the time remaining to submit those results. Another problem arises when multiple learners are going a more complex assignment, which can cause server-side problems.

Malicious code injection

The second type of security vulnerability is malicious code injection [11]. Whereas in traditional assignment submission, code is sent as an email to the person conducting the grading, in autograder systems the compiler or interpreter runs the code (often multiple times for each test) as soon as the learner clicks on a submit button.

These vulnerabilities can be exploited by attempting to post the executable code or fragment of such code to the server where it can be executed to retrieve some sensitive data or disrupt the server functionality [11]. These attacks include, but are not limited to cross side scripting (XSS) or SQL injection attacks [12]. Another type of attack is to use the platform to perform actions not meant to, i.e. using an infinite loop (or recursion errors) to run code that can access remote services, hence exploiting the resources of the platform.

The main method of mitigating these types of attacks is to encapsulate the web-application running the autograder using a container such as Docker [13 ,14]. Using such a form of virtualization, only the infected instance can be affected, but the functionality of the platform remains unharmed. Furthermore, access control lists (ACLs) can be implemented to disable any external communication from the platform.

4. TOWARDS SECURE AUTOGRADERS

In this Section we discuss feasible solutions to enable a secure autograder system which can be deployed commercially, in MOOCs, as well as at HEIs. As of writing this paper, a complete secure solution for autograder does not exist, or is proprietary. We discuss the benefits and the drawbacks of the proposed system, as well as possible trade-offs.

In order to secure the autograder system, a layered protection system has to be deployed. The first level of this system should be metrics-based, i.e. check the hashes of the raw input to a known solution. This operation is conducted as a first pass after submitting code and the reason is twofold. Learners which do not alter a solution from the Internet should be quickly identified, as the hashing function of code which is in general a hundred lines at most a quick operation. Secondly, is the code is identified to be malicious, the resources needed for the autograder are not used. The second level of fraud detection should use an AST or PDG algorithm, with the substitution of variables and functions applied beforehand. The combination of these techniques allows the resources of the server running the autograder to be used in a more efficient manner. The sequence diagram of the proposed two-level fraud detection system is given in Image 3.

As the instance level, the autograder web-app has to be encapsulated using an operating system level virtualisation software such as Docker. This approach will ensure that malicious code injection that passes the fraud detection protective layer is isolated from other server resources. In addition, any incoming code with easily-identified SQL or similar code should be removed with a fail and/or warning message given to the learner. To further increase connecting external resources, an ACL should be implemented to block all external communication. The exception to these rules should only be assignments where a specific external resource is being connected. The extended component diagram is given in Image 4.

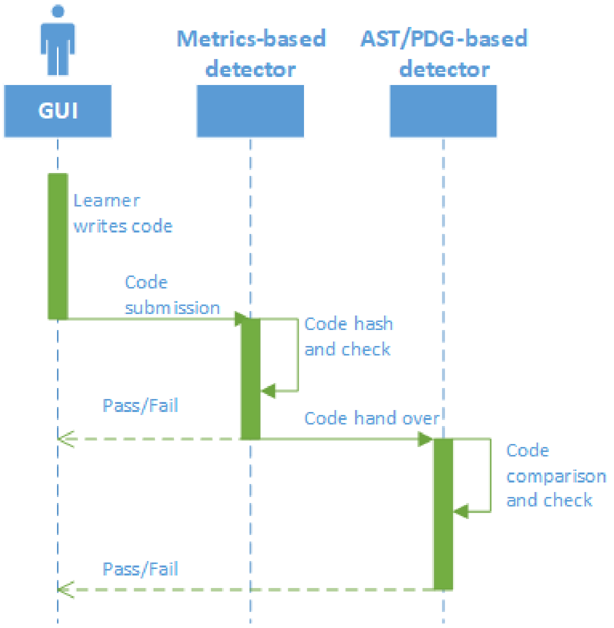


Image 3: Sequence diagram for the two-step fraud detection.

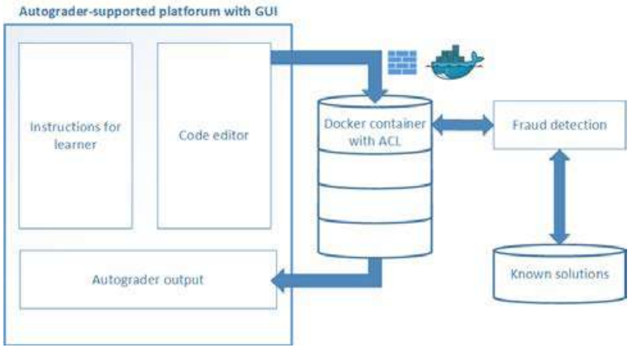


Image 4: Proposed security extension to the autograder-supported e-learning platform.

Both detectors should connect to a database with known solutions. Furthermore, this database could be implemented as a blockchain [15], where a) solutions obtained from previous students can be easily traced back to, and b) new solutions with their origin and timestamp are automatically added to the database.

5. DISCUSSION AND CONCLUSION

In this paper, we have identified possible security vulnerabilities to autograder systems used in e-learning, the means of their mitigation and/or prevention, and presented a two-level fraud detection extension to a commonly employed autograder system.

It is worth noting that every protection addition to an already established system will impact the overall latency of the platform. A possible trade-off could be using only one detector, such as an AST/PDG-based one, with a strict ACL blocking all outgoing traffic from the platform. Extending our previous work in autograder systems, our further research will focus on implementing these types of solutions, as well as establishing a private blockchain to implement the known solutions.

ACKNOWLEDGMENT

This paper was supported in part by the Blockchain Technology Laboratory at Belgrade Metropolitan University, Belgrade, Serbia, and in part by the Ministry of Education, Science and Technological Development, Republic of Serbia (Project III44006).

REFERENCES

[1] Vivek Ravisankar, 2020 HackerRank Developer Skills Report," HackerRank, 2021, [Online]. Available: https://research.hackerrank.com/developer-skills/2020/ (Accessed: Nov 2021).

[2] European Commission, Digital education action plan 2021-2027 [Online], Available: https://ec.europa.eu/education/education-in-the-eu/digitaleducation-action-plan_en, 2021 (Accessed: Nov 2021).

[3] N. Zdravković, , N. Dimitrijević, D. Cvijanović "A System for Interactive Learning of the Python Programming Language with Autograding Support," in Proc. of the 12th Conference on eLearning, pp. 131-136, Sep. 2021.

[4] G. E. Forsythe, and N. Wirth, "Automatic grading programs," Communications of the ACM, vol. 8, no. 5, pp. 275-278, 1965.

[5] C. Douce, D. Livingstone, and J, Orwell, "Automatic test-based assessment of programming: A review," Journal on Educational Resources in Computing (JERIC), vol. 5, no. 3 pp. 4-es, 2005.

[6] M. Konecki, T. Orehovacki, and A. Lovrencic, "Detecting computer code plagiarism in higher education," in Proc. of the 31st International Conference on Information Technology Interfaces, IEEE ITI 2009.

[7] C. Lyon, R. Barrett, and J. Malcolm, "Plagiarism Is Easy, But Also Easy to Detect, Plagiary: Cross-Disciplinary Studies in Plagiarism, " Fabrication, and Falsification, vol. 1, no. 5, pp. 1-10, 2006.

[8] N. Parlante, CodingBat, [Online]. Available: https://codingbat.com/ (Accessed: Nov 2021).

[9] M. Bruntink, A. Deursen, R. Engelen, and T. Tourwe, "On the Use of Clone Detection for Identifying Crosscutting Concern Code," IEEE Transactions on Software Engineering, vol. 31, no. 10, pp. 804-818, 2005.

[10] K. Nguyen, "Automatic Evaluation of Python and C Programs with codecheck," Master's thesis, SJSU, 2014.

[11] A. Scerbakov, F. Kappe, and N. Scerbakov, "Security vulnerabilities in modern LMS," in Proc. of the International Conference E-Learning, pp. 242-246. 2019.

[12] D. C. Luminita, "Information security in E-learning Platforms," Procedia-Social and Behavioral Sciences, vol. 15, pp. 2689-2693, 2011.

[13] F. Špaček, R. Sohlich, T. Dulík, "Docker as platform for assignments evaluation," Procedia Engineering, vol. 100, pp. 1665-1671, 2015.

[14] S. A. Sokolov, A. P. Zahariev, S. M. Vlaev, T. B. Iliev and I. S. Stoyanov, "Technology e-Learning environment for the hybrid cloud," Proc. of the 23rd IEEE International Symposium for Design and Technology in Electronic Packaging (SIITME), pp. 451-454, 2017.

[15] M. Damnjanović, V. Grković, N. Zdravković, "Towards Secure Online Studies: Applying Blockchain to e-Learning," in Proc. of the 11th International Conference on e-Learning, pp. 20-33, 2020.

ON SOME APPLICATION OF THE GENETIC ALGORITHM IN INFORMATION SECURITY SYSTEMS – A SURVEY

MILENA BOGDANOVIĆ

Belgrade Metropolitan University, Faculty of Information Technologies, milena.bogdanovic@metropolitan.ac.rs

NIKOLA DIMITRIJEVIĆ

Belgrade Metropolitan University, Faculty of Information Technologies, nikola.dimitrijevic@metropolitan.ac.rs

VIJAYAKUMAR PONNUSAMY

Department of ECE, SRM Institute of Science and Technology, Kattankulathur, India, vijayakp@srmist.edu.in

Abstract: *Information security is a set of practices intended to keep data secure from unauthorized access or alterations, both when it's being stored and when it's being transmitted from one machine or physical location to another. Nowadays computer systems are facing increased number of security threats. Intruders can be divided into two groups, external and internal. The goal of a security system is to protect the most valuable assets of an organization: data and information. Different organizations will have very different security policies and requirements depending on their missions. GAs are widely used in information security systems. The application of a genetic algorithm to the field of cryptology is unique. This paper reviews genetic algorithm application in information security systems, provides technical analysis of modern malware types, which are used for evolution modeling, and analyses existing malware models.*

Keywords: *information security, security threats, genetic algorithm, data security, data encryption techniques.*

1. INTRODUCTION

Nowadays, the security of data storage and transmission is a major challenge for governments, various companies and organizations. The goal of a security system is to protect the most valuable assets of an organization: data and information. Different organizations will have very different security policies and requirements depending on their missions. Government organizations (especially related to the defense projects) have secret information and cannot afford to rely only on the defense against attacks from outside the organization. They must also consider the danger of theft and/or sabotage from someone inside the organization [1].

Cipher system considers to be the basic element for improving data security, in other hand decipher system is an important step for attack to broken secure data in all time, thus it must create a stamp for sender (cipher data) and one for who received (decipher) to guarantee the data received by a specific user [2]. Development of human intelligence with the art of cryptography has become more sophisticated in order to make information more secure. Recently, secure data transmission over network has become a vital and critical issue due to increased demand of digital media transmission and unauthorized access of important data [3].

2. BASIC OF GENETIC ALGORITHM

Genetic algorithm (GA) is one of the most interesting classes of optimization algorithms, which with its new applications and very good results in recent years constantly surprising how people from the world of technology, as well as ordinary users who are beneficiaries of products resulting from its use. From unusual ideas, to more unusual applications, mimicking the process of evolution GA demonstrate the power of the mechanisms of nature. Genetic algorithms (GA) are a family of algorithms that use some kind of genetic principles that are present in nature, in order to solve specific computational problems. These natural principles are: inheritance, crossover, mutations, survive of the best custom (or survival of the fittest), migration and so on. These algorithms can be used for solving various classes of problems because they are fairly general nature. By mode of action, genetic algorithms are among the methods directed random search space solutions are looking for a global optimum. Genetic algorithms simulate the natural evolutionary process.

The idea is based on the GA processes imitate natural selection. The basis of the selection process that takes place in nature are the following facts: individuals better adapted to survive the environment and have a stronger influence on the formation of the next generation, a generation individuals in the population form the next generation, thus so what are the new features of specimens receive a combination of genetic content of the parents, from time to time there is a mutation, ie. accidental changes to the genetic content of one individual.

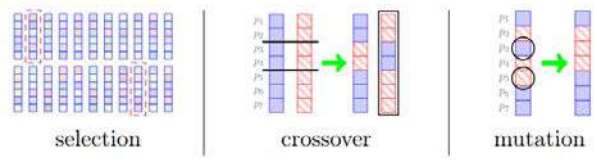


Image 1. Genetic Algorithm Process [4]

Elements of genetic algorithms are:

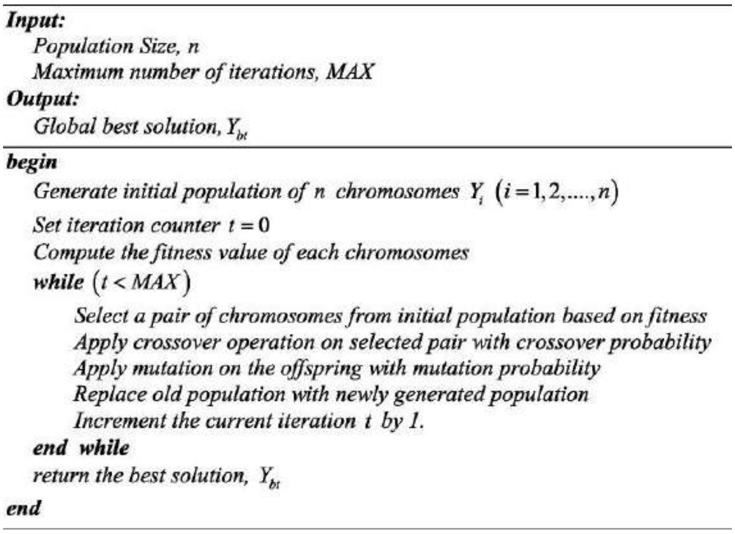
- search space, ie. set of all possible solutions,
- populations, ie. set of actual candidates for the resolution,
- elements of the population are individuals (nodes results, ie. point in search space),
- low space (strings), ie. space representation of individuals,
- as well as functions to map the search engine space in the area are low and vice versa,
- a set of genetic operators to generate new strings,
- and thus new individuals,
- evaluative function (fitness function), which determines the advantage (usefulness) certain individuals,
- stochastic control genetic operators.

In a narrow sense, the term genetic algorithm applies only to the model that was introduced by John Holland in his book "*Adaption in natural and artificial systems*", 1975 [5]. Holland is considered the creator of this meta-heuristic and basic settings of his earliest works are valid even today. More generally, genetic algorithm is any algorithm that is based on a population and operators of selection, crossing and mutation, which are used for production of new points in the space search.

The big problem of genetic algorithms is theoretical - the lack of rigorous mathematical proof that genetic algorithms generate an optimal solution. However, in practice they give satisfactory results, and often there is no need for a theoretical explanation. Usually the quality of a new type of genetic algorithm evaluates empirically, by experimentally tested on a problem of optimization.

Genetic algorithm is applied to the final set of individuals called the population. Each individual in the population is represented by a series of characters (*genetic code*) and corresponds to a solution in search space. Coding can be binary or of some other higher cardinality alphabet. Encoding solutions is an important step genetic algorithm for inappropriate choice of code can lead to bad results regardless of the rest of the structure of the algorithm.

Instead of working with objects (individuals), genetic algorithms work with parametric descriptions of objects (individuals). Mathematical model is based on the play, and the algorithm process control, reproduction and survival of objects (individuals) that can "compete" in the search for a solution. The algorithm, therefore, is a set of parametric descriptions of objects (individuals), with the population. Genetic operators are applied to individuals and shall be repeated their assessment. Genetic operators provide the offspring are similar but not identical with their parents, which allows the population to evolve to solutions that were not present in the initial set of objects (individuals).



```
Input:
    Population Size, $n$
    Maximum number of iterations, $MAX$
Output:
    Global best solution, $Y_{bt}$
begin
    Generate initial population of $n$ chromosomes $Y_i$ $(i = 1, 2, ...., n)$
    Set iteration counter $t = 0$
    Compute the fitness value of each chromosomes
    while $(t < MAX)$
        Select a pair of chromosomes from initial population based on fitness
        Apply crossover operation on selected pair with crossover probability
        Apply mutation on the offspring with mutation probability
        Replace old population with newly generated population
        Increment the current iteration $t$ by 1.
    end while
    return the best solution, $Y_{bt}$
end
```

Image 2. Classical Genetic Algorithm (GA) [6]

Theory of GAs is based on schemata theory and probability, i.e. mathematical foundations of GA are:

- schema definition,
- hypothesis building blocks,
- theorem on implicit parallelism,

and GA to increase popularity was largely due to the existence of mathematical apparatus, whose use of some results can be predicted and explained.

The strength of the genetic algorithms is the fact that they are able to determine the global optimum position in space with multiple local extremes, the so-called *multimodal space*. Applying the classical deterministic methods to solve optimization problems, we get results that are always moving towards the local minimum or maximum, where it can be global, but this can not be determined from the results. On the other hand, stochastic methods, including genetic algorithm, not dependent on a possible starting point and can process their results, with some probability, to locate the global optimum of some objective function [7].

Genetic algorithm shows its power at the most complex requirements, because almost sure to converge to the best solution (global minimum or global maximum). Its strength lies in the way in which the variables are varied and searching for a solution. The essence of the optimization procedure is as follows:

- To specify the problem formed by a set of possible solutions (orderly n-tuple variables). This set may contain solutions that are close to the best ("works"), and were obtained in some other way, in which case the task of the genetic algorithm to improve them. However, this set is often obtained pseudo-random selection of n-tuples.

- Possible solutions (n-tuples) in some way are encoded in the sequence of binary digits, which we call the candidates, and formed the fitness function, whose role is to perform the inverse action, or to evaluate (assess) the candidates on the basis of specified criteria (objectives optimization). Fitness function actually shows us how close a candidate solution, and based on it shall be the best choice.

Fitness function is strict. Candidates who are furthest from the solution, it is deleted, while those who show any movement toward the solution, it allows intertwine and give offspring (selection process). Besides selection process, and there are random changes (mutations). In this way, looking for the best solution, the best candidates survive, as in the process of evolution [8].

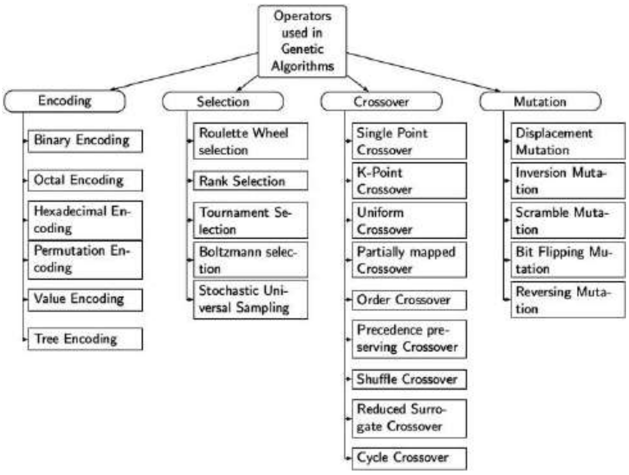


Image 3. Operators used in GA [9]

3. A SURVEY OF THE RELATED WORK

For protection of valuable information from unlawful imitation, eavesdropper's attack and modification, different types of cryptographic algorithms are designed. There are two major types of such algorithms: symmetric cryptography [10] and asymmetric cryptography [11]. In asymmetric key cryptography two different keys are used, one for encryption called public key and one for decryption called private key. There are two ways of key production; the first one is mathematical like AES, DES and the other one is based on the theory of natural selection [12]. Cryptography generally uses DES algorithm for the Encryption and Decryption. DES using round and round strategy. The DES uses private key and its works by using the same key to encrypt and decrypt a data [13].

Herrera et al. in [14] optimized the rule base of a fuzzy logic controller (FLC) by applying GA. The membership functions have already been created for FLC. Lee and Takagi [15] applied GA in order to determine the number of membership functions.

In the paper [16] Belarbi and Titel used binary encoding, implemented FLC as a neural network, and used the GA to train the weights. Park et al. [17] implemented a GA to automate FLC design. Pati and Sahoo [18] used GA for FLC in order to automate the rule extraction in FPGA. D´ıaz et al. [19] presented the use of GA as a tuning factor for fuzzy control rules. Ireland [20] used GA in the domain of intrusion detection in order to train the dataset. Various kinds of modern data encryption techniques [21-22] are found in the literature. Genetic Algorithms (GAs) [23] are among such techniques.

The classical image encryption techniques require the input parameters for encryption. The wrong selection of

input parameters will generate inadequate encryption results [24].
GA and its variants have been used to select the appropriate control parameters. Kaur and Kumar [25] developed a multi-objective genetic algorithm to optimize the control parameters of chaotic map. The secret key was generated using beta chaotic map. The generated key was use to encrypt the image. Parallel GAs was also used to encrypt the image [26].

In the paper [27], the authors proposes a novel approach called AdacDeep that uses an Enhanced Genetic Algorithm (EGA), Deep Autoencoder and a Deep Feedforward Neural Network (DFFNN) with backpropagation learning to accurately predict different attack types. The performance of AdacDeep is evaluated using two well-known datasets, namely, the CICIDS2017 and UNSW_NB15 datasets as the benchmark. The experimental results show that AdacDeep outperforms other state-of-the-art comparative models in terms of prediction accuracy with 0.22–35% improvement, F-Score with 0.1–34.7% improvement and very low false positive rate.

Genetic Algorithm (GA) has been used in different ways in IDSs. In the approaches described in [28], the IDS can be viewed as a rule-based system (RBS) and GA can be viewed as a tool to help generate knowledge for the RBS. In this paper shows how network connection information can be modeled as chromosomes and how the parameters in genetic algorithm can be defined in this respect. This implementation of genetic algorithm is unique as it considers both temporal and spatial information of DARPA data set Rule Set Rule Base Network Sniffer GA network connections during the encoding of the problem; therefore, it should be more helpful for identification of network anomalous behaviors.

James V. Hansen et al. [29] hypothesize that genetic programming algorithms can aid in this endeavor. To investigate this proposition, they conducted an experiment using a very large dataset from the 1999 Knowledge Discovery in Database (KDD) Cup data, supplied by the Defense Advanced Research Projects Agency (DARPA) and MIT's Lincoln Laboratories. The authors used machine-coded linear genomes and a homologous crossover operator in genetic programming, which led to promising results in detecting malicious attacks. They showed that the resulting programs are executed in real time, with a high level of accuracy in identifying positive and negative cases.

In the paper [30], the study showed that GA can be effectively used for formulation of decision rules in intrusion detection through the attacks which are more common can be detected more accurately. Rule based classification of DoS and Probe attacks can be used for effective monitoring of the network. If the use of GA is compared with the use of expert based knowledge, then the use of GA is more fruitful. The reason is that the different possible combination of attributes is tested on the basis of training data, and later confirmed on the basis of testing data. In addition, it was found that increasing the number of iterations of the algorithm application leads to an increase in data accuracy. Also, it was found that the initial iteration converges faster than the later iterations.

Biometrics is seen by many as a solution to a lot of user identification and security problems in today‘s networks. One more application of genetic algorithms is their adoption for fingerprint database generation. Fingerprint databases are used for testing biometric authentication systems. While constructing a fingerprint database, it is important to have in mind the performance of automatic recognition systems [31]. The authors presented a model that uses genetic algorithm for palm recognition.

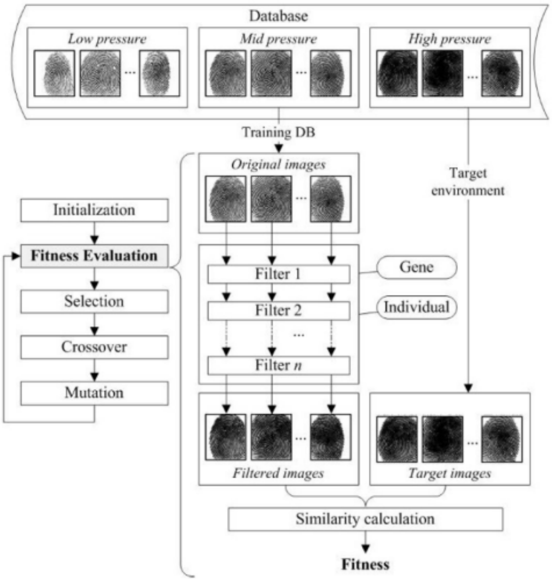
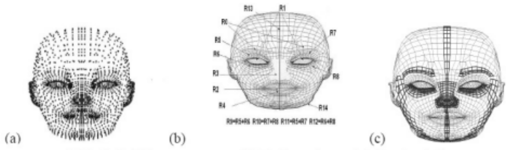


Image 4. GA-based fingerprint image generation method [32]

In the paper [33], a Genetic Algorithm (GA) based approach is proposed for face recognition. The proposed algorithm recognizes an unknown image by comparing it with the known training images stored in the database and gives information regarding the person recognized.

Image 5. Recognition rate for the Genetic Algorithm [33]

In the paper [34] the facial feature space is represented by a set of labels. A cubic approximation method is explored to estimate the principal curvatures of each vertex on the model. Fig. 6.(a) [34] shows the labelled original feature space. Among the set of labels, only the labels located in certain regions are of the most interest. The face model is partitioned into 15 sub-regions based on their physical structures (there are overlaps between some of the regions, Fig. 6.(b)). The GA-based method selects the components that contribute the most to the face recognition task. The procedure for the GA-based feature selection consists of two parts: vertices selection in each sub-region, and the integration of sub-regions.



(a) Labeled feature space. (b) Sub-regions. (c) Optimal feature space.

Image 6. 3D facial feature extraction [34]

The capacities of current 3D face databases are relatively insufficient. To solve this problem, Yin et all. in [35] present a framework to augment existing 3D databases based on Genetic Algorithm. First the prototypical face samples are divided into patches. Then the new face samples are generated by assembling randomly selected patches. Under the guidance of the genetic algorithm, Yin et all. in [35] can perform a number of the generating works at a time. The experiment results show that the proposed method has good performance on face data expansion.

5. CONCLUSION

The aim of this paper is to, through a brief analysis of some of the results, show how the use of genetic algorithms can increase the level of trust in terms of information security. The results indicated that the genetic algorithm was fast enough to provide results and turned out to be more flexible than the discrete dynamic programming method. More and more scenarios can be added to test the robustness and efficiency of this approach. We have presented a very small part of the results of the use of GA in information security, but our intention was to point out that in today's growing needs for information security, the use of metaheuristics is available in order to improve it.

REFERENCES

[1] Maaz Bin Ahmad, Adeel Akram, M. Asif, Saeed Ur-Rehman, "Using Genetic Algorithm to Minimize False Alarms in Insider Threats Detection of Information Misuse in Windows Environment", Mathematical Problems in Engineering, vol. 2014, Article ID 179109, 12 pages, 2014. https://doi.org/10.1155/2014/ 179109

[2] Jhingran, R., Thada, V. and Dhaka, S. 2015. "A Study on Cryptography using Genetic Algorithm", International Journal of Computer Applications, 118(20): pp. 10 – 14.

[3] A. Almarimi, A. Kumar, I. Almerhag, and N. Elzoghbi, "A NEW APPROACH FOR DATA ENCRYPTION USING GENETIC Original Image Pseudorandom Binary Sequence Generator using GA and Decryption Decrypted Image," Computer (Long. Beach. Calif)., pp. 2–6, 2014.

[4] https://wakespace.lib.wfu.edu/bitstream/handle/10339/59317/Odell_wfu_0248M_10904.pdf

[5] Holland, J. H. (1975) *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Ann Arbor

[6] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7599983/

[7] Bogdanović, M. (2011) **"**On some basic concepts of genetic algorithms as a meta-heuristic method for solving of optimization problems – a review", A Journal of Software Engineering and Applications, Vol. 4, No. 8, pp. 482-486, doi: 10.4236/jsea.2011.48055. http://www.scirp.org/journal/jsea

[8] Milena Bogdanović, "A SURVEY ON SOME POSSIBILITIES FOR APPLICATIONS OF GENETIC ALGORITHMS IN THE AUTOMATA THEORY", pp. 79-89, Godišnjak Pedagoškog fakulteta u Vranju, knjiga VII, 2016. UDK 510.53 512.54.05, ISSN 2466-3905, COBISS.SR-ID=221686284

[9] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7599983/

[10] J. Daemen and V. Rijmen, The Design of Rijndael: AES - The Advanced Encryption Standard. 2002

[11] R. L. Rivest, A. Shamir, and L. Adleman, "A method for obtaining digital signatures and public-key cryptosystems," Commun. ACM, vol. 21, no. 2, pp. 120–126, 1978.

[12] Goyat, S. 2012. "Cryptography Using Genetic Algorithms", IOSR Journal of Computer Engineering (IOSRJCE), 1(5): pp. 06-08.

[13] Nagpure, B. D. , Dhote, A. D., Rokade P. S., Kale P. B. and Kinhikar, N. S. 2016. "Implementation of Network Security Using Genetic Algorithm", International Journal of Research in Advent Technology (IJRAT*)*,Special Issue, National Conference "Convergence2016", 06th-07th April

[14] F. Herrera, M. Lozano, and J. L. Verdegay, "Tuning fuzzy logic controllers by genetic algorithms," International Journal of Approximate Reasoning, vol. 12, no. 3-4, pp. 299–315, 1995.

[15] M. A. Lee and H. Takagi, “Integrating design stages of fuzzy systems using genetic algorithms,” in Proceedings of the IEEE International Conference on Fuzzy Systems, pp. 612–617, San Francisco, Calif, USA, April 1993.

[16] K. Belarbi and F. Titel, “Genetic algorithm for the design of a class of fuzzy controllers: an alternative approach,” IEEE Transactions on Fuzzy Systems, vol. 8, no. 4, pp. 398–405, 2000.

[17] Y. J. Park, H. S. Cho, and D. H. Cha, “Genetic algorithm based optimization of fuzzy logic controller using characteristic parameters,” in Proceedings of the IEEE International Conference on Evolutionary Computation, pp. 831–836, Perth, Australia, December 1995.

[18] P. Pati and J. Sahoo, “Implementation of genetic algorithm based fuzzy logic controller with automatic rule extraction in FPGA[thesis]”, Department of Electronics & Communication Engineering National Institute of Technology, Rourkela, India, 2013.

[19] N. P. D´ıaz, R. L. Jim´enez, and G. Angelesa, “Tuning fuzzy control rules via genetic algorithms: an experimental evaluation,” Research Journal of Recent Sciences, vol. 2, no. 10, pp. 81–87, 2013

[20] E. Ireland, “Intrusion detection with genetic algorithms and fuzzy logic,” in Proceedings of the UMM CSci Senior Seminar Conference, Morris, Minn, USA, December 2013.

[21] D. R. Stinson, “Cryptography”: Theory and Practice, vol. 30. 2005.

[22] W. M. H. Company, 2Modern Cryptography”: Theory and Practice, vol. 170, no. 2. 2003.

[23] D. E. Goldberg, “Genetic Algorithms in Search, Optimization, and Machine Learning”. 1989.

[24] Sourabh Katoch, Sumit Singh Chauhan, Vijay Kumar, “A review on genetic algorithm: past, present, and future”, Multimed Tools Appl. 2020 Oct 31 : 1–36. doi: 10.1007/s11042-020-10139-6

[25] Kaur M, Kumar V. “Beta chaotic map based image encryption using genetic algorithm”. Int J Bifurcation Chaos. 2018;28(11):1850132.doi:10.1142/S0218127418501328.

[26] Kaur M, Kumar V. “Parallel non-dominated sorting genetic algorithm-II-based image encryption technique”. The Imaging Science Journal. 2018;66(8): pp. 453–462. doi: 10.1080/13682199.2018.1505327

[27] Ayei E. Ibor, Florence A. Oladeji, Olusoji B. Okunoye & Charles O. Uwadia (2021) “Novel adaptive cyberattack prediction model using an enhanced genetic algorithm and deep learning (AdacDeep)”, Information Security Journal: A Global Perspective, DOI: 10.1080/19393555.2021.1883777

[28] W. Li, “Using genetic algorithm for network intrusion detection”, Proceedings of the United States Department of Energy Cyber Security Group 1 (2004), pp. 1–8

[29] J. V. Hansen, P. B. Lowry, R. D. Meservy, D. M. McDonald, “Genetic programming for prevention of cyber-terrorism through dynamic and evolving intrusion detection”, Decision Support Systems 43 (4) (2007), pp. 1362–1374.

[30] M. S. A. Khan, “Rule based network intrusion detection using genetic algorithm”, International Journal of Computer Applications 18 (8) (2011), pp. 26–29.

[31] Cenys, A., Gibavicius, D., Goranin, N., & Marozas, L. (2013), “Genetic Algorithm Based Palm Recognition Method for Biometric Authentication Systems”, Elektronika Ir Elektrotechnika, 19(2), pp. 69-74. https://doi.org/10.5755/j01.eee.19.2.3473

[32] Cho, U. K.; Hong. J. H.; Cho, S. B. 2007. “Automatic Fingerprints Image Generation Using Evolutionary Algorithm”, Lecture Notes in Artificial Intelligence 4570, pp. 444–453.

[33] Pratibha Sukhija, Sunny Behal, Pritpal Singh, “Face Recognition System Using Genetic Algorithm”, Procedia Computer Science, Volume 85, 2016, pp. 410-417, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2016.05.183. (https://www.sciencedirect.com/science/article/pii/S1877050916305233)

[34] Sun, Y.; Yin, L. 2007. “A Genetic Algorithm Based Approach for 3D Face Recognition”, Computational Imaging and Vision: 3D Imaging for Safety and Security 35, pp. 95–118.

[35] Ge, Y., Yin, Bc., Sun, Yf. et al. “Expansion of 3D face sample set based on genetic algorithm”, Multimed Tools Appl 70, pp.781–797 (2014). https://doi.org/10.1007/s11042-012-1102-4

CIP - Каталогизација у публикацији - Народна библиотека Србије, Београд

007:004.056(082)

КОНФЕРЕНЦИЈА о безбедности информација (12 ; 2021 ; Београд)
Proceedings / The Twelfth International Conference on Business Information Security, Belgrade, 3rd December 2021. ; [editors Nemanja Zdravković, Bojana Domazet]. - Belgrade : Metropolitan University, 2021 (Kruševac : Copy Centar). - 91 str. : ilustr. ; 30 cm

Tiraž 60. - Bibliografija uz svaki rad.

ISBN 978-86-89755-22-0

а) Информације - Заштита - Зборници
COBISS.SR-ID 54086153