



BELGRADE
METROPOLITAN
UNIVERSITY



BISEC
BUSINESS INFORMATION
SECURITY CONFERENCE

14th International Conference on
Business Information Security (BISEC'2023)
Niš, November 24th 2023

www.bisec.metropolitan.ac.rs

PROCEEDINGS
The Fourteenth International Conference on Business Information Security



Belgrade Metropolitan University

Niš, 24th November 2023

www.metropolitan.ac.rs

Publisher

Belgrade Metropolitan University
Tadeuša Koščuška 63, Belgrade, Serbia
<http://www.metropolitan.ac.rs>

For Publisher

Prof. Dr. Dragan Domazet

Editor

Asst. Prof. Nemanja Zdravković
Olga Pavlović

Chair of Organizing Committee

Asst. Prof. Nemanja Zdravković

Organizing Committee

Asst. Prof. Nemanja Zdravković
Anja Marković
Olga Pavlović
Ana Ročkomanović

Design

Petar Cvetković
Mladen Radić

Printing

Scero Print Niš

Circulation

40

Belgrade, 2024

CONTENT

ADRIÁN CAMPAZAS-VEGA, ALBERTO MIGUEL-DIEZ, MARIO HERMIDA-LÓPEZ, CLAUDIA ÁLVAREZ-APARICIO, IGNACIO SAMUEL CRESPO-MARTÍNEZ AND ÁNGEL MANUEL GUERRERO-HIGUERAS.....	07
Cybersecurity Issues in Robotic Platforms”	
ZLATOGOR MINCHEV, LUBEN BOYANOV.....	15
”Future of Smart Cities Security Challenges – Proactive Modelling & Identification”	
MILOŠ KOSTIĆ AND IGOR SAVELJIĆ.....	21
”Gamification as a Tool for Elevating Password Strength Awareness”	
ANDREJA SAMČOVIĆ.....	26
”Security Related Use of Facebook as a Communication Channel”	
MARKO S. STEFANOVIĆ, NENAD O. VESIĆ, ALEKSANDRA PENJIŠEVIĆ AND ĐORĐIJE VUJADINOVIĆ.....	32
”Advanced Cryptography Using Conformal Mappings”	
STEFAN GOGIĆ, NEMANJA ZDRAVKOVIĆ, EMILIJA KISIĆ AND PONNUSAMY VIJAYAKUMAR.....	36
”Secure Course Completion Credentialing Using Hyperledger Fabric”	
YAJNA PANDITH.....	42
”Deep Blockchain to Enable Scalable Web Applications”	
ALEXANDER K. ALEXANDROV, ANASTASS N. MADZHAROV.....	57
”Energy-Efficient Routing in UAVs Supported Perimeter Security Networks”	
ALEXANDER K. ALEXANDROV.....	63
”Reducing the WSN’s Communication Overhead by the SD-SPDZ Encryption Protocol”	
LJUBIŠA BOJIĆ, VLADIMIR ĐAPIĆ.....	71
”The Interplay of Social and Robotics Theories in AGI Alignment: Navigating the Digital city through Simulation-based Multi-Agent Systems”	
MARINA DODEVSKA.....	77
”Survey on Methods of Online Payment over the Internet”	

MILOŠ JOVANOVIĆ, STEFAN JANČIĆ.....	83
”Ethical Dimensions of AI Security and Privacy Policies: Enabling Inclusive Growth”	
ALEKSANDAR JOVANOVIĆ, PETAR MILIĆ.....	95
”Custom Made Approaches to Cloud Container Security: a Methodologically Sound Approach Based on International Sample Results”	
NIKOLA SRETENOVIĆ, DEJAN NEMEC.....	99
”NBS Web Service Use in a Business Environment”	
DAMIR BRADIĆ, DEJAN NEMEC.....	106
”Ensuring High Availability of Clusters Within the Network Infrastructure Using Microsoft Hyper-V Technology in a Medium-Sized Enterprise”	
VALENTINA B. PAUNOVIĆ, SEDAT A. UYAR.....	112
”Exploring the Power of AI in Internet Security: Balancing Attacks and Defenses in Black and White”	
MILICA MLADENOVIĆ.....	124
”Human Aspects of Online Security and Needs for Implementing Corporate Work/Life Balance Programs”	
NIKOLA DIMITRIJEVIĆ, NEMANJA ZDRAVKOVIĆ, MILENA BOGDANOVIĆ AND ALEKSANDAR MESTEROVIC.....	128
”Advanced Security Mechanisms in the Spring Framework: JWT, OAuth, LDAP and Keycloak”	

Organizer



COORDINATOR OF THE INTERNATIONAL PROGRAMME COMMITTEE:

Zlatogor Minchev, Bulgarian Academy of Sciences, Republic of Bulgaria
Miguel Ángel Conde, Universidad de León, Spain

MEMBERS:

Prof. Dr. Zlatogor Minchev, Bulgarian Academy of Sciences, Republic of Bulgaria
Prof. Dr. Mitko Bogdanoski, Military Academy "General Mihailo Apostolski" Skopje, Republic of North Macedonia
Ramo Šendelj, University Donja Gorica, Montenegro
Marko Beko, Universidade Lusófona, Lisbon, Portugal
Urska Cvek, Louisiana State University Shreveport, One University Place, Shreveport, LA
Marjan Trutschl, Louisiana State University Shreveport, One University Place, Shreveport, LA
Miroslava Raspopović, Belgrade Metropolitan University, Serbia
Igor Franc, Belgrade Metropolitan University, Serbia
Nemanja Zdravkovic, Belgrade Metropolitan University, Serbia
Milena Bogdanović, Belgrade Metropolitan University, Serbia
Dragan Đurđević, Academy of National Security, Serbia
Aca Aleksić, Information Technology Services Dunav RE, Serbia
Slobodan Jovanović, Belgrade Metropolitan University, Serbia
Ivana Ognjanović, University Donja Gorica, Montenegro
Nemanja Maček, School of Electrical and Computer Engineering, Beograd, University Business Academy in Novi Sad
Sonsoles López Pernas, University of Eastern Finland
Miguel Ángel Conde, Universidad de León, Spain
Dragan Domazet, Belgrade Metropolitan University, Serbia
Ponnusamy Vijayakumar, SRM IST, ECE Department, Kattankulathur, Chennai, India

Language

The official language of the BISEC 2023 Conference is English. English will be used for all printed materials, presentations and discussion.

Cybersecurity Issues in Robotic Platforms

Adrián Campazas-Vega^{1,*}, Alberto Miguel-Diez¹, Mario Hermida-López¹,
Claudia Álvarez-Aparicio¹, Ignacio Samuel Crespo-Martínez¹ and
Ángel Manuel Guerrero-Higueras¹

¹Grupo de Robótica de la Universidad de León, Campus de Vegazana, 24071 León, Spain

Abstract

The use of robots has increased dramatically in recent years. Currently, there are multiple types of robots, from service robots, designed to help people in any kind of environment (home, work, hospitals...), to quadruped platforms, developed for critical infrastructures or the military field. Security in those platforms is crucial, since robots present vulnerabilities, they can pose a risk to both their integrity and that of the people/objects around them. In this work, a security evaluation of the Unitree A1, a quadruped robot, and the humanoid robot Pepper has been carried out, to know the security flaws that may be present, as well as the implications that it may have for the user, the environment, or the integrity of the robot. The final goal of the work is that the vulnerabilities found will be taken into account by other researchers or companies that develop that kind of robot and take into account those security problems.

Keywords

Pentesting, robot, security, Unitree A1, Pepper

1. Introduction

The use of robots has exponentially increased in the last decade. Throughout the year 2022, the utilization and deployment of industrial robots increased by 40% in the United States and 6% in Spain, according to the Spanish Association of Robotics (AER) [1]. Industrial robotics has traditionally focused on the precise repetition of tasks, surpassing the capabilities of a human being. However, in recent years, there has been a particular emphasis on the development of robotic platforms capable of performing tasks that are difficult or dangerous for humans. In this regard, the most impactful robotic platforms are quadruped robots. These robots are characterized by supporting their weight on four legs, typically mimicking the morphology of a dog. The design of these devices offers advantages over bipedal robots due to their versatility in adapting to various types of terrains. The characteristics of quadruped robots enable them to undertake tasks considered challenging or hazardous for humans. These tasks include bomb inspection and deactivation, radiation detection, and critical infrastructure maintenance.

In addition to their civilian applications, these robots are actively utilized in the military domain [2]. Similarly, the use of service robots has also significantly increased in recent years. These robots are designed to interact and communicate with humans to assist in the completion of everyday tasks.

Similarly, to other types of devices, cybersecurity in robotic environments is an important aspect that becomes critical when a robot is involved in highly sensitive tasks or interacts with people. Many issues with these platforms arise because manufacturers often prioritize manufacturing cost or design over conducting product security testing [3]. In addition to the lack of device security by manufacturers, it is worth noting that most of these robotic platforms are "plug and play," meaning that end users often do not pay proper attention to configuring the device correctly. This includes changing default passwords, which poses an additional security challenge.

This paper aims to address some of the security issues presented by both quadruped robotic platforms and social robots. Specifically, a security evaluation has been conducted on the quadruped robot Unitree A1 and the semi-humanoid robot Pepper, with the objective of identifying potential vulnerabilities and risks that could affect both humans and the robot itself, as well as the environment in which it is deployed. The severity of the discovered vulnerabilities has been assessed using the CVSSv3 (Common Vulnerability Scoring System version 3) standard. This work and the methods employed can serve as a starting point for other researchers interested in evaluating the security risks of other models of quadruped robots and social robots.

The rest of the article is organized as follows: In Section 2, related works are presented. Section 3 introduces

BISEC'23: 14th International Conference on Business Information Security, November 24, 2023, Niš, Serbia

*Corresponding author.

✉ acamv@unileon.es (A. Campazas-Vega);
amigud00@estudiantes.unileon.es (A. Miguel-Diez);
mherml00@estudiantes.unileon.es (M. Hermida-López);
calvaa@unileon.es (C. Álvarez-Aparicio); icrem@unileon.es
(I. S. Crespo-Martínez); am.guerrero@unileon.es
(Á. M. Guerrero-Higueras)
📞 0000-0001-8237-5962 (A. Campazas-Vega); 0000-0002-7465-8054
(C. Álvarez-Aparicio); 0000-0002-3154-0144 (I. S. Crespo-Martínez);
0000-0001-8277-0700 (Á. M. Guerrero-Higueras)

the architecture and characteristics of the robots Unitree A1 and Pepper, along with the method for assessing the severity of discovered vulnerabilities. Section 4 provides details on the various experiments conducted and the implications of exploiting the vulnerabilities in a real-world environment. Finally, Section 5 offers the current conclusions.

2. Related Works

Despite the growing popularity of quadruped robots, there is limited research on the cybersecurity of these robots. Most research in this field focuses on the physical security of robots, such as collision prevention [4] and stability on different terrains [5]. However, there are some works that examine overall security in robotic devices. In [6], the authors analyzed potential security issues that different types of robots might have and listed some generic recommendations that could be implemented to enhance the overall security of robotics. One of the conclusions reached by the authors is that cyberattacks on robots used in critical infrastructures and military environments are the most damaging and dangerous. It's important to note that the current use of quadruped robots primarily focuses on these two areas. Another work related to robotic security is presented in [3]. In this work, the authors identified security threats in the field of robotics, classified them based on the affected layer of the robot's architecture, and analyzed their impact and potential countermeasures. Other works, such as [7] and [8], discuss security issues associated with ROS (Robot Operating System). ROS is a set of software libraries and tools that help create applications for robots. While Pepper and Unitree A1 do not come with ROS by default, it is possible to install ROS on the latter.

Finally, regarding the specific analysis of the Pepper robotic platform, in [9], the authors conducted a security evaluation of the semi-humanoid robot "Pepper" from SoftBank Robotics. The authors demonstrated that this robot had critical vulnerabilities that needed to be addressed by the manufacturer. This article expands on the work done in [9], confirming that years later, the vulnerabilities identified by the authors still exist and uncovering new vulnerabilities in the platform.

3. Materials and Methods

In this section, the characteristics of the robots analyzed in this work are presented. Additionally, the methodology used to conduct the experiments and the evaluation method for these experiments are described.



Figure 1: Unitree A1 of the Robotics Group of the University of León.

3.1. Unitree A1

As mentioned in Section 1, to conduct the cybersecurity evaluation of quadruped robots, the Unitree A1 robot, as shown in Figure 1, has been utilized. The Unitree A1 is manufactured by Unitree Robotics, a Chinese company that has been producing quadruped devices since 2016 [10].

The Unitree A1 robot can reach a maximum speed of 3.3 m/s at a particular moment and can carry objects with a maximum weight of 5 kg. Additionally, it is equipped with sensors that enable it to maintain proper balance during operation, preventing the robot from falling on uneven terrain. The device has a battery life ranging from 1 to 2.5 hours, depending on the mode in which it is used [11].

Regarding the cameras and sensors, the Unitree A1 is equipped with a RealSense camera [12], located on its "head." This camera features a depth sensor that utilizes a combination of infrared and laser technologies to measure the distance between objects and the camera. This enables it to capture 3D images and detect objects in real-time. In the field of robotics, these types of cameras are used to implement autonomous functions in the robot, allowing it to navigate around obstacles and create a 3D map of the area in which the robot is deployed [13, 14]. At the connectivity level, the quadruped robot has several ports on the upper part of its "body" that the user can utilize to interact with various interfaces of the robot. These connections include four USB ports, two HDMI

ports, and two Ethernet ports.

Teleoperation of the robot can be performed using a mobile application developed by the manufacturer or by using the controller that comes with the robot. The controller includes two joysticks and a directional pad (D-pad) for easy robot maneuvering. According to the manual, the controller connects directly to the robot's control board via radio frequency. On the other hand, Unitree's mobile application is compatible with both iOS and Android devices. The app allows users to control the robot, view the real-time camera feed, and utilize a simulator of the Unitree A1. However, despite the robot being available for commercial use since 2020, some features of the app may not work correctly or require specific parameter configurations. Furthermore, Unitree provides users with a Software Development Kit (SDK) to develop custom code for the robot. This SDK enables developers to create their own applications and functionalities for the Unitree A1.

3.2. Pepper

Pepper is the world's first social humanoid robot capable of recognizing human faces and basic emotions. It is optimized for interaction and can engage with people through conversation or its touchscreen interface. Pepper is designed for intuitive and natural interaction. It finds common applications in various fields such as hospitality, retail, healthcare, education, entertainment, and personal assistance. Its appearance is depicted in Figure 2.

Pepper has 20 degrees of freedom to achieve more natural and expressive movements. Additionally, it features voice recognition available in 15 languages and perception modules to recognize and interact with the person in front of it. In terms of physical sensors, the robot is equipped with touch sensors, LEDs, microphones for multimodal interaction, infrared sensors, bumpers, an inertial unit, and 2D and 3D cameras to enable autonomous and omnidirectional navigation. Pepper provides an API that allows for the development of custom applications and functionalities for this robotic platform.

3.3. Evaluation

To assess the severity of the discovered vulnerabilities, the Common Vulnerability Scoring System (CVSS) version 3 has been employed [15]. CVSS, or Common Vulnerability Scoring System, is an open and widely used framework that defines metrics for communicating the characteristics, impact, and severity of vulnerabilities affecting security elements. It provides a standardized way to evaluate and communicate the seriousness of security vulnerabilities.

CVSSv3 categorizes vulnerabilities with a numerical value between 0 and 10. A vulnerability with a score

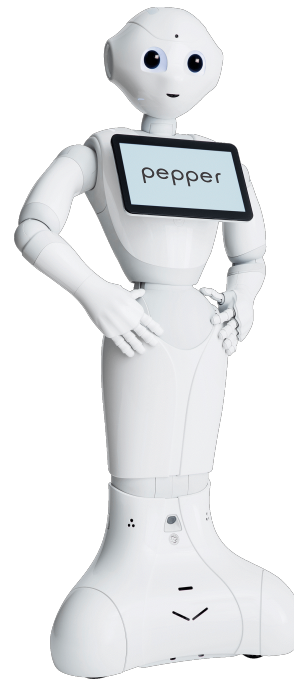


Figure 2: Appearance of the Pepper service robot.

between 0.1 and 3.9 is considered to have low severity. Vulnerabilities with a score between 4.0 and 6.9 are classified as having moderate severity. Finally, vulnerabilities with a score between 7.0 and 10.0 are considered to have high severity. This scoring system provides a clear way to assess the seriousness of vulnerabilities and helps organizations prioritize their remediation efforts.

CVSS defines metrics to assess the likelihood that a vulnerability will be exploited. The metrics defined by the CVSSv3 standard can be seen in Table 1.

3.4. Methodology

The methodology used for the analysis of robotic platforms is similar to that employed in conventional computer systems. Below, we outline the three stages carried out to assess the security of the Unitree A1 robot and the Pepper service robot:

- **Information Gathering:** In this step, information is collected about the robotic platform, including the type of hardware and sensors used by the device, the operating system it runs on, the services it executes, and the nature of the communications that take place.
- **Vulnerability Analysis:** Tests are conducted to identify vulnerabilities in the robotic system. This analysis encompasses both hardware and

Table 1
Metrics associated with the CVSS vector in version 3

Symbol	Description
AV	Attack Vector: Determines how the vulnerability can be exploited, assessing the accessibility requirements. The values of this metric are: <ul style="list-style-type: none"> • Network (N) • Adjacent (A) • Local (L) • Physical (P)
AC	Attack Complexity: Determines the attack complexity required to make use of the vulnerability. The values of this metric are: <ul style="list-style-type: none"> • Low (L) • High (H)
PR	Privileges Required: Determines the level of privileges an attacker must have before he can successfully exploit a vulnerability. The values of this metric are: <ul style="list-style-type: none"> • None (N) • Low (L) • High (H)
UI	User Interaction: Determines if user intervention is necessary for successful exploitation of the vulnerability. The levels of this metric are: <ul style="list-style-type: none"> • None (N) • Required (R)
S	Scope: Determines whether successful exploitation of the vulnerability can indirectly affect other components outside the scope of the system or application. The values of this metric are as follows: <ul style="list-style-type: none"> • Unchanged (U) • Changed (C)
C	Confidentiality Impact: Confidentiality is the ownership of a document, message or data that is only authorized to be read or understood by certain persons or entities. The values of this metric are as follows: <ul style="list-style-type: none"> • None (N) • Low (L) • High (H)
I	Integrity Impact: Integrity is the property of a document, message or data that guarantees the veracity of the information. The values for this metric are as follows: <ul style="list-style-type: none"> • None (N) • Low (L) • High (H)
D	Availability Impact: Availability is the property of a system, service, or application that is accessible without impediments. The values for this metric are as follows: <ul style="list-style-type: none"> • None (N) • Low (L) • High (H)

software aspects, as well as the systems deployed by the robot.

- **Exploitation of Identified Vulnerabilities:** Finally, identified vulnerabilities are exploited to determine the extent to which these security flaws pose a risk to the safety of the robot itself and its surrounding environment.

4. Experimentation and Discussion

The evaluation conducted on these robots aims to identify vulnerabilities that may be present in the devices and could be extrapolated to other robotic platforms. The following will demonstrate how both robots share common

vulnerabilities. All vulnerabilities listed below are associated with an impact vector generated using the CVSSv3 standard, as discussed in Section 3. The discovered vulnerabilities, which are explained below, are presented in Table 2.

4.1. Common vulnerabilities in both robots

In this subsection, we present the vulnerabilities that are common to both robots.

Table 2

Vulnerabilities of the evaluated robots

Vulnerability	Impact	Robot
Lack of protection against brute force attacks in SSH protocol	High	Unitree A1 Pepper
Lack of verification against MiTM attack	High	Unitree A1 Pepper
Denial of service to the robot's Web server	Moderate	Unitree A1 Pepper
Unsecured physical ports	High	Unitree A1
Web server without authentication	Moderate	Unitree A1
API access without authentication	High	Pepper
Communication with the web server without encryption	Moderate	Pepper

4.1.1. Lack of protection against brute force attacks in SSH protocol

One way to access the embedded computers inside the robot is through the SSH protocol. This connection allows for configuring certain aspects of the robot, such as the AP password, and even controlling the robot using the installed SDK. Both the Unitree A1 robot and Pepper do not implement security measures to prevent brute-force attacks on the SSH servers installed in the robot. To verify that the SSH servers are vulnerable to dictionary attacks or brute-force attacks, the open-source tool Hydra has been used [16].

If an attacker gains access to the robot's internal computers, they could potentially control the robot remotely and even delete system files, rendering the device inoperable. Furthermore, since the default password for both devices is considered insecure today and is present in a wide range of online dictionaries, this vulnerability is deemed severe with a score of 9 and the following CVSS vector: `AV:A/AC:L/PR:L/UI:N/S:C/C:H/I:H/A:H`.

4.1.2. Lack of verification against MiTM attack

Neither the quadruped robot Unitree A1 nor the social robot Pepper implement security measures to prevent an attacker with access to the robot's network from performing a Man-in-the-Middle (MitM) attack. This would allow the attacker to intercept unencrypted communications and manipulate them at will. Here's an example of the vulnerability in the Unitree A1 robot: The A1 robot deploys a web server that serves images from the robot's camera, allowing an operator to teleoperate the device remotely.

An attacker who has access to the network deployed by the robot can carry out a MitM attack, altering the video transmission from the robot's camera with another feed controlled by the attacker, without the victim noticing any difference. If the robot is used in critical situations, the operator controlling the robot will not perceive the



Figure 3: On the left, view of the teleoperator after being attacked. On the right, real image of the robot's situation.

actual situation, potentially enabling an attacker to cause harm to the robot itself or its surrounding environment.

To exploit this vulnerability, an ARP Spoofing attack was conducted using the "arpspoof" tool [17]. This attack is considered one of the most dangerous on LAN networks [18]. The attacker manipulates both the robot's and the victim's ARP tables, associating their MAC address with the victim's IP address, thereby redirecting all traffic to a machine controlled by the attacker. Subsequently, the attacker redirects the traffic arriving from the user to a web server identical to the robot's but under the attacker's control. In this case, the web server deployed by the Unitree is MJPG-Streamer, which is publicly available on GitHub [19].

The consequences of such attacks can be critical in certain environments. For instance, in Figure 3, can see that the person operating the robot perceives an obstacle-free corridor, while in reality, the robot is in a hazardous situation near a set of stairs.

This vulnerability has a high impact with a score of 8.0 and the following associated CVSSv3 vector: `AV:A/AC:L/PR:L/UI:N/S:U/C:H/I:H/A:H`. A video has been created to replicate the experiment performed [20].

4.1.3. Denial of service to the robot's Web server

The web servers deployed by both robots are vulnerable to denial-of-service (DoS) attacks. The process to exe-



Figure 4: Top view of Unitree A1.

cute this attack is quite similar to the previous one, as it relies on the ARP Spoofing technique in both cases. To exploit this vulnerability, the attacker must manipulate the victim's and robot's ARP tables to intercept traffic. Once the attack is successfully carried out, all packets are received by the attacker, who will then discard these packets, causing the legitimate user to lose the connection to the web server. This vulnerability has a moderate impact with a score of 5.7 and the following associated CVSSv3 vector: AV:A/AC:L/PR:L/UI:N/S:U/C:N/I:N/A:H.

4.2. Unitree A1 robot vulnerabilities

This section shows vulnerabilities that exclusively affect the Unitree A1 robot.

4.2.1. Unsecured physical ports

Figure 4 shows the port distribution of the robot. The main vulnerability lies in the fact that the robot does not request any form of authentication when connected through the provided ports.

The lack of authentication poses several security implications, even without connecting standard input and output devices such as a keyboard and monitor. Currently, there are USB-like devices that function as input and output devices, enabling the execution of commands

simply by plugging them in. These devices are referred to as Rubber Ducky [21]. Furthermore, the exposure of USB ports also makes the robot vulnerable to attacks carried out with a USB killer device [22]. This type of device discharges a high-voltage surge, damaging the components of the connected device. This vulnerability has a high impact with a score of 7.5, and the associated CVSSv3 vector is AV:P/AC:L/PR:N/UI:N/S:C/C:H/I:H/A:L.

4.2.2. Web server without authentication

Access to the live video feed from the robot's camera does not have an authentication system. Therefore, any user connected to the network emitted by the robot can view the real-time image either through the device's web server or via the mobile application. To be considered secure, this functionality should require authentication.

This vulnerability has a moderate impact with a score of 5.7 and the following CVSSv3 vector: AV:A/AC:L/PR:L/UI:N/S:U/C:H/I:N/A:N.

4.3. Pepper robot vulnerabilities

In this section, the vulnerabilities that exclusively affect the social robot Pepper are presented.

4.3.1. API access without authentication

The API implemented by Pepper allows for complete control of the device. Access to the API occurs without any form of authentication, so an attacker only needs to be on the same network as the robot. Interaction with the API is done through port 9559 using the Python programming language, although C++ and Java are also supported.

This vulnerability has a high impact with a score of 7.5, and the associated CVSSv3 vector is: AV:N/AC:L/PR:N/UI:N/S:U/C:N/I:N/A:H.

4.3.2. Communication with the web server without encryption

The web server used by the robot utilizes unencrypted HTTP communication. An attacker connected to the network can sniff the traffic and obtain the access credentials for the web server, as depicted in Figure 5.

This vulnerability has a moderate impact with a score of 6.5 and the following CVSSv3 vector: AV:L/AC:L/PR:L/UI:N/S:C/C:H/I:N/A:N.

5. Conclusions

The use of robotics is becoming increasingly widespread; however, it is essential that progress in this field is accom-

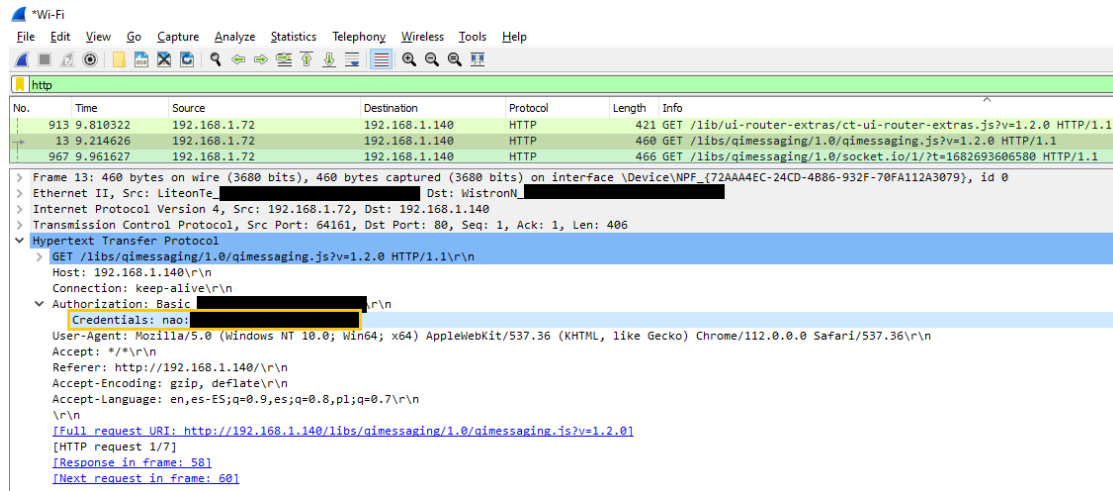


Figure 5: Capture of Pepper’s traffic showing the robot’s plaintext credentials.

panied by a thorough review of potential vulnerabilities in these devices.

In this work, a security evaluation has been conducted on the quadruped robot Unitree A1 and the service robot Pepper. Several potential vulnerabilities have been identified that could be exploited by an attacker to gain unauthorized access to the robot or control its movements and actions. For each of the vulnerabilities discovered in this work, a Common Vulnerabilities and Exposures (CVE) has been requested. The CVE program’s mission is to identify, define, and catalog publicly disclosed cybersecurity vulnerabilities.

To continue advancing in the field of robotics, it is necessary to implement security measures such as user authentication and authorization, encryption of device communications, and regular security testing to detect and address potential vulnerabilities in the software of various robotic platforms. It is important to emphasize that the cybersecurity of quadruped and social robots is a critical issue that must be addressed by manufacturers, developers, and users of these devices to ensure their proper functioning and protect them against potential malicious attacks that could pose a security risk to the robot itself or to people in its vicinity.

Acknowledgment

This research has been partially supported under the grant PID2021-126592OB-C21 funded by MCIN/AEI/10.13039/501100011033 and by ERDF A way of making Europe and under the Grant TED2021-132356B-I00 funded by MCIN/AEI/10.13039/501100011033 and by the "Eu-

ropean Union NextGenerationEU/PRTR.

References

- [1] A. E. de Robótica y Automatización, La importancia de la ciberseguridad en la industria 4.0, https://www.aer-automation.com/wp-content/uploads/2023/01/Ciberseguridad_AERPaper.pdf, 2023.
- [2] K. Geldenhuys, Killer robots are real, *Servamus Community-based Safety and Security Magazine* 116 (2023) 20–22.
- [3] G. W. Clark, M. V. Doran, T. R. Andel, Cybersecurity issues in robotics, in: *2017 IEEE conference on cognitive and computational aspects of situation management (CogSIMA)*, IEEE, 2017, pp. 1–5.
- [4] R. Singh, T. Bera, Walking model of jansen mechanism-based quadruped robot and application to obstacle avoidance, *Arabian Journal for Science and Engineering* 45 (2020) 653–664.
- [5] Y. H. Lee, Y. H. Lee, H. Lee, L. T. Phan, H. Kang, U. Kim, J. Jeon, H. R. Choi, Trajectory design and control of quadruped robot for trotting over obstacles, in: *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 2017, pp. 4897–4902.
- [6] C. Cerrudo, L. Apa, Hacking robots before skynet, *IOActive Website* (2017) 1–17.
- [7] S.-Y. Jeong, I.-J. Choi, Y.-J. Kim, Y.-M. Shin, J.-H. Han, G.-H. Jung, K.-G. Kim, A study on ros vulnerabilities and countermeasure, in: *Proceedings of the Companion of the 2017 ACM/IEEE International*

- Conference on Human-Robot Interaction, 2017, pp. 147–148.
- [8] R. White, D. H. I. Christensen, D. M. Quigley, Sros: Securing ros over the wire, in the graph, and through the kernel, arXiv preprint arXiv:1611.07060 (2016).
- [9] A. Giaretta, M. De Donno, N. Dragoni, Adding salt to pepper: A structured security assessment over a humanoid robot, in: Proceedings of the 13th International Conference on Availability, Reliability and Security, 2018, pp. 1–8.
- [10] U. Robotics, Unitree, <https://m.unitree.com/>, 2022.
- [11] U. Robotics, Unitree a1 user manual, https://www.mybotshop.de/Datasheet/UnitreeA1_User_Manual_v1.0.pdf/, 2020.
- [12] F. L. Siena, B. Byrom, P. Watts, P. Breedon, Utilising the intel realsense camera for measuring health outcomes in clinical research, *Journal of medical systems* 42 (2018) 1–10.
- [13] J. Bayer, J. Faigl, On autonomous spatial exploration with small hexapod walking robot using tracking camera intel realsense t265, in: 2019 European Conference on Mobile Robots (ECMR), IEEE, 2019, pp. 1–6.
- [14] J. Hu, Y. Niu, Z. Wang, Obstacle avoidance methods for rotor uavs using realsense camera, in: 2017 Chinese Automation Congress (CAC), IEEE, 2017, pp. 7151–7155.
- [15] INCIBE, Métricas de evaluación de vulnerabilidades: Cvss 3.0, <https://incibe-cert.es/blog/cvss3-0/>, 2023.
- [16] V. Hauser, Hydra, <https://github.com/vanhauser-thc/thc-hydra/>, 2022.
- [17] D. Song, arspooof - intercept packets on a switched lan, <https://manpages.ubuntu.com/manpages/bionic/man8/arspooof.8.html>, 2022.
- [18] G. Jinhua, X. Kejian, Arp spoofing detection algorithm using icmp protocol, in: 2013 International Conference on Computer Communication and Informatics, IEEE, 2013, pp. 1–6.
- [19] jacksonliam, Servidor web mjpg-streamer, <https://github.com/jacksonliam/>, 2021.
- [20] A. Miguel, Ataque man in the middle al unitree a1, <https://bit.ly/3JGCGDl>, 2023.
- [21] INCIBE, Rubber ducky, ¿una simple memoria usb?, <https://www.incibe.es/empresas/blog/rubber-ducky-simple-memoria-usb>, 2023.
- [22] O. Angelopoulou, S. Pourmoafi, A. Jones, G. Sharma, Killing your device via your usb port, in: Proceedings of the Thirteenth International Symposium on Human Aspects of Information Security & Assurance (HAISA 2019), The Centre for Security, Communications and Network Research (CSCAN), 2019, pp. 61–72.

Future of Smart Cities Security Challenges – Proactive Modelling & Identification

Zlatogor Minchev^{1,2,*}, Luben Boyanov³

¹Institute of ICT, Bulgarian Academy of Sciences, Acad. Georgi Bonchev Str., Bl. 25A, 1113 Sofia, Bulgaria

²Institute of Mathematics & Informatics, Bulgarian Academy of Sciences, Acad. Georgi Bonchev Str., Bl. 8, 1113 Sofia, Bulgaria

³University of National & World Economy, 8-mi Dekemvri Str. 19, 1700 Sofia, Bulgaria

Abstract

Joining technologies & people in future smart cities infrastructure by merging sensors, effectors and intelligence is going to create a rather challenging mixed reality transformation. In this sense, the competition between natural and artificial intelligence is inevitably establishing quite new and interesting society overlaying of humans and technologies with federated domination areas. The results are presently addressing the digital society transformation towards Society 5.0, whilst outreaching the next Society 6.0 expectations. The paper is going to outline a comprehensive analytical intelligence framework (i-framework) for studying the problem, adding a scenario-based proactive analysis, combined with system modelling and results hybrid multicriteria validation. The intelligent part comes from different AI models that are implemented in the process, giving supportive and generative added values. Finally, a concluding discussion on the outlined findings is presented.

Keywords

Future smart cities, digital society transformation, security challenges, scenario-based analysis, system modelling, hybrid multicriteria validation

1. Introduction

Digital transformation is expected to affect in practice all fields of future society reality, including people and their residence area, adding also biotope dynamics (to note: climate changes, species migration, natural disasters, etc.) [1]. As for the new urban environment of future smart cities, the process is certainly expected to combine new IoTs & AI, providing innovative commodities, services (to mark: transportation, deliveries, education, governance, media, energy supplies, assistance, medicine, economics) and jobs for the citizens, aiming the horizon towards the year 2050 [2]. New smart gadgets' autonomous integration (multifunctional robots, vehicles, etc.) with the vastly interconnected reality (due to broadband wireless meshes & optical network technologies enhanced usage) will additionally advance the new habitual digital landscape [3].

Thus, the digital change towards the post-information age is expected to have both - positive and negative transformational effects on the new Society 5.0 idea [4]. The situation is getting even more complicated with Society 6.0 transcends exploration [5], where AI and machine singularity are expected to appear in practice.

So, new technologies are going to establish a digital

divide between citizens in the smart urban reality and the rest of the populated areas. They are going to be considered as a new digital class with advanced capabilities but will be also challenged via joint human-machine threats in the smart habitat [6]. This definitely will affect future jobs and culture transformation, together with deeper smart machines, sensors and algorithms integration in the transformed people's lifestyle and environment smart reshaping. In this context, the security, privacy and ethical issues require an adequate and smart exploration approach that is proactively organized as to be earlier prepared as a civilization for this change.

Further, an exploration methodological approach in this context is going to be outlined, combining both natural and artificial intelligence with expert analytical support.

2. Analytical "i-Framework"

Combining human & machine intellect into a joint analytical power is practically extending some of the ideas from [1, 7] into a new "i-framework" (see Figure 1), better applicable to the digital future security dynamics, and comprehensive proactive exploration.

The extension was organized due to the vast dynamics' escalation with AI & IoTs immersion in the post-information age, as stated in the introductory part of the paper that is difficult to be easily handled due to the scale and dynamics with human only intellect.

While aiming proactive identification and comprehensive analysis of digital transformation security in general,

BISEC'23: 14th International Conference on Business Information Security, November 24, 2023, Niš, Serbia

*Corresponding author.

✉ zlatogor@bas.bg (Z. Minchev); lboyanov@unwe.bg (L. Boyanov)

🆔 0000-0003-2479-5496 (Z. Minchev); 0009-0006-2292-0619

(L. Boyanov)

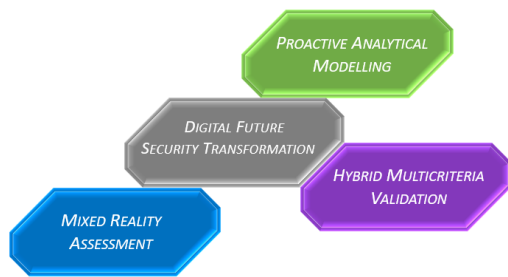


Figure 1: Analytical "i-framework" for digital future security proactive exploration.

the current "i-framework" implementation is addressing the future smart cities in particular. Being a broad landscape example for security dynamics studying with people & technologies digital transformation, the outlined findings could be further used with the broader digital society evolution deeper exploration.

Further in the paper a detailed illustration of the trilateral "i-framework", concerning (2.1) Proactive Analytical Modelling, (2.2) Hybrid Multicriteria Validation & (2.3) Mixed Reality Assessment will be given for the future smart cities' comprehensive security challenges context exploration.

2.1. Proactive Analytical Modelling

Achieving proactive analytical modelling is combining both morphological and system analysis approaches. A starting implementation of the scenario method, with expert and reference data, towards the establishment of plausible and implausible scenario combinations is accomplished. The result is a cross-consistency matrix M , containing three types of scenarios, in accordance with their Relative Common Weight – RCW : Active (tangible), Passive (intangible) & Neutral (probably most uncertain) [8]. With the present study on future smart cities, the particular matrix context towards year 2037 and post-information society of 136080 scenarios [1] has been zoomed for smart cities security topic, with a total scenario number $N^* = 2880 (N^* = 5 \times 3 \times 4 \times 3 \times 4 \times 4)$; plausible – $N1^* = 86$ & implausible ones – $N2^* = 2794$; from $N1^*$ are additionally selected: Active, i.e. – "tangible" (76, $RCW > 0$) & Passive, i.e. – "intangible" (10, $RCW < 0$).

As this landscape shows quite an uncertain future with mostly implausible scenarios, a deeper analytical causality exploration has been performed toward smart city system sensitivity analysis.

A "system-of-systems" modelling paradigm over an i-fuzzy weighted graph-based "Entity – Relationship"

Morphological Analysis				
Drivers	Threats	Measures	Ambiguities	Objectives
Mixed Intellect	Reality Mixing	Tech Limiting	Smart Resources	Resilient Future Cities
Climate Changes	Smart Dual Apps	AI Overwrite	Privacy Concerns	Energy Independence
Quality of Life	Lifestyle Machine Control	Legal Issues	New Smart Activities	Transformed Security
	AI Autonomization		Infrastructure Smart Services	Transformed Citizens


Index	Length	Weight	Name	
1	5	5	Scenario1	Active scenarios +  Passive scenarios -
2	5	30	Scenario2	
3	5	65	Scenario3	
4	5	-5	Scenario4	
5	5	10	Scenario5	
6	5	-35	Scenario6	
7	5	5	Scenario7	

Figure 2: Future smart cities security transcendents towards year 2037.

representation in I-SCIP-SA environment [9] has been performed.

Taking an aggregated analytical representation into a "3D Sensitivity Diagram" – "3D SD" with Influence – x , Dependence – y & Sensitivity – z , due to relations i-fuzzy weights, concerning future smart cities security towards year 2037 (simulated in 10 steps) is finally achieved with 16 entities (addressing social – yellow, technical – blue & mixed – white aspects) & 41 bi-directional relations model (see Figures 3 and 4).

The resulting classification gives four classes for the model entities distribution (with two subclasses for each: Active – white & Passive – grey) in the 3D SD diagram as follows:

Buffering (in green): "Social Credits Score" – 16, being at the same time Passive.

Active (in red): "Smart Infrastructure" – 5, "Transformed Life" – 8, "Mixed Intelligence" – 11, all being Active.

Critical (in yellow): "Super Humans" – 3, "Human Preservation" – 15, both being Passive & "Autonomous AI" – 2, "Data Leakages" – 6, "New Jobs" – 7, "Smart Communication" – 9, "Criminal Activities" – 13 all being Active.

Passive (in blue): "Privacy Concerns" – 1, "World i-Domination" – 4, "AI Regulations" – 10, "Hardware Compromising" – 12, "Law Enforcement" – 14 all being Passive.

Further, the presented quantitative classification results could be aggregated in more detail, around several key findings:

(i) That future smart cities will create numerous threats and challenges from their critical infrastructure perspective [10], that could be attacked with different vectors: communicational, data & hardware ones. Apart of this, the future superhumans will have a somewhat ambiguous role with new technologically extended capabilities. So, insider security threats are neither to be completely excluded or added by default due to potential technological

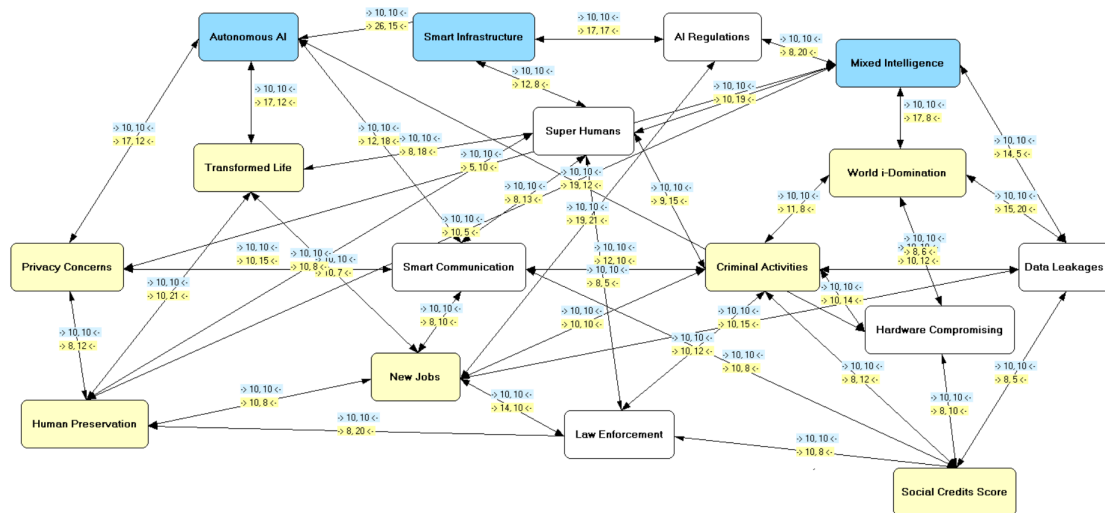


Figure 3: Future smart cities security transcendent of a discrete system-of systems model.

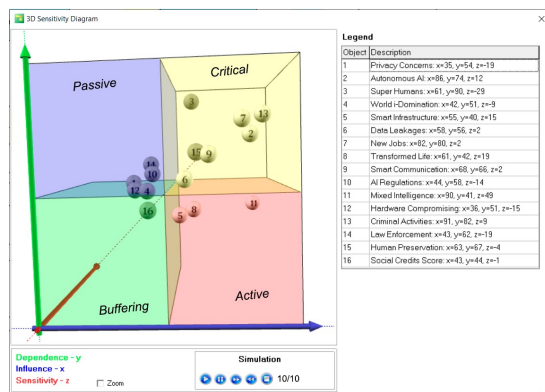


Figure 4: Resulting 3D SD analytical assessment towards year 2037.

influence.

(ii) New jobs and ambitions for intelligent world domination will certainly appear and progress both from positive and negative (criminal, manipulative) perspectives, adding AI & IoT with new capabilities and challenges;

(iii) Mixing artificial and natural intelligence for future smart cities' security could be quite beneficial except if a superintelligence with negative objectives manages to compromise the system due to emergency external influences (natural or man-made disasters).

(iv) Keeping privacy and humanity present understandings will be quite different for the future as the role of AI & IoT transformation will also demand new ethical and social boundaries.

All these findings will hopefully omit the dystopian

scenarios with machine-controlled social credits and behaviour that normally are a question of culture and social system respect, whilst trying to keep a non-authoritarian but secured future urban reality.

As the presented expert findings are mostly based on expert analytical beliefs, further AI-assisted validation and assessment will also be given, trying to achieve a comprehensive urban security landscape exploration.

An overall evaluation of the exercise has been performed by the participants (with Positive either Indefinite judgment marks) via a q-based survey, giving feedback for: reality, scenario & interawareness complexity, AI & human factor roles, training satisfaction (see Figure 6).

Being somewhat subjective the obtained results could be also enriched with biometric & simulation assessments (see [1, 9]) and finally combined within the system model. More details on these ideas will be given further.

2.2. Mixed Reality Assessment

The mixed reality assessment of the accomplished analytical findings (see Section 2.1) was further conducted, using a transformed reality interactive simulation, organized in the framework of CYREX 2023 [11]. Assuming a fictitious scenario events script (generated with human intellect guiding & tailoring Open AI ChatGPT results), interactively played (for about 180 minutes) from the trainees in several multirole teams, an exploration of future smart cities security transcendent was performed.

The main objective of CYREX 2023 exercise was to test the human-machine inter-awareness in an imaginary context, concerning the future smart cities mixed reality from different aspects (both utopian and dystopian



Figure 5: Selected moments (a, b) & organizational architecture (c) of CYREX 2023 [11].

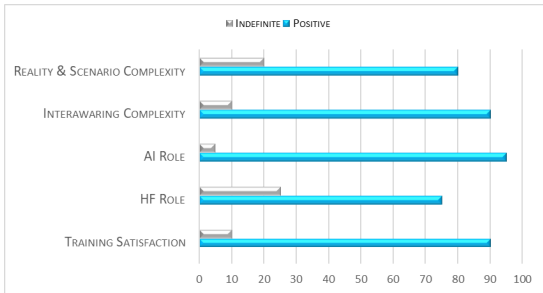


Figure 6: Aggregated participants' q-based assessment of CYREX 2023.

ones), concerning the cyber security area different aspects: technological, social, infrastructural, security, political, governance, diplomatic. The idea was to develop and test a set of morphological and system models for the not-so-far future (10-15 years from now), among young

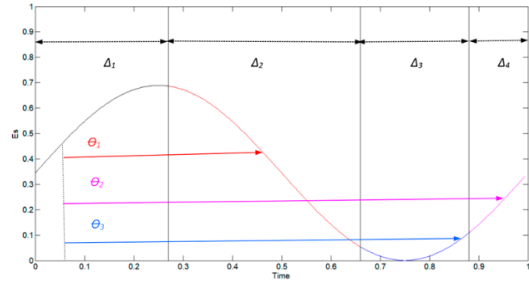


Figure 7: Socio-technological S-shaped curve quantum representation, after [1].

people (Y- & Z- generations), joining both human and machine intelligence in the process of decision-making and scenario development, while using a role-based organization, multiple smart gadgets (smartphones, laptops, advanced PCs, smart TVs & interactive screens) and platforms (Windows, Android, iOS). The training was illustrated, combining results of artificially generated images, videos, sounds and popular multimedia clips (assisted with Gencraft, beatovenAI & invideoAI). The approach allowed studying of complex security transcendent dynamics in a futuristic mixed reality smart ecosystem, giving excellent training feedback results, especially at organizational and operational levels.

2.3. Hybrid Multicriteria Validation

The main idea for this section is to combine both human and AI expectations for the future of smart cities' security, taking as a base the presented in 2.1. Analytical i-Framework findings in the dynamic context. So, a joint smart approach has been further accomplished, adding human beliefs vs machine-adaptable smart multicriteria optimization [10]. In this manner, it is possible to get a feasible evaluation of potential future expectations, taking into account trends S-shaped dynamics, but with unplanned jumps (see Figure 7) that could be best explained with quantum tunnelling effect stochastic socio-technological modelling of system model relations dynamics [1].

Selected multicriteria near future illustrative critical entities (towards the year 2037, see Figure 8), concerning the system-of-systems discrete model of smart cities security transcendentals (see Figure 3) are further presented, taking the risk-assessment approach from [12], but using a percentage-based measuring scale.

So, taking society the system-of-systems model ideas is quite intriguing as normally the model is both subjective to the expert beliefs and at the same time – limited due to the preliminary analytical assessment. In this sense, the validation process has been extended with

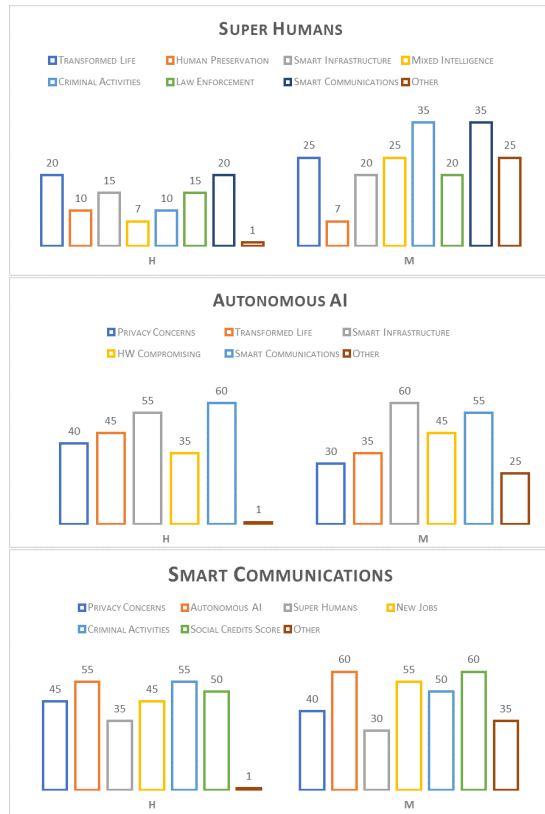


Figure 8: Selected multicriteria smart cities security future entities illustrative assessment (after Figure 3) from human – H & machine – M perspectives towards 2037.

"Other" additional relations and entities. This additional extension could be produced from both human and generative AI hints and supplementary human discussions. With the presented illustrative examples an added human opinions value has been taken from SRS'2023 young international participants in the context of Society 6.0 [13]. The machine-generated added value was produced by taking Open AI ChatGPT feedback with the human responses' extension, towards machine domination, concerning smart cities' security near future evolution.

3. Conclusion

Proactively identifying the future smart cities' security challenges is a quite complex task that could benefit from both human and machine intelligence joint efforts. Going deeper in the problem normally requires a suitable framework as has been already shown in the present study. Whilst human intelligence is always subjective by nature responsive biometric feedback could be quite

helpful with the analytical models' more detailed assessments. As for the role of AI, it is still at an early stage of development and the dream for a "General AI" is still quite limited. However, it should be honestly marked that the generative AI on the other hand, is quite supportive in the analytical assessment and experimental issues. Thus, providing both a neutral opinion advisor and a rapid prototyping tool that facilitates the exploration efforts' proactive nature a lot. This clearly shows a positive technological trend for the not so far digital future new social evolution.

Acknowledgment

The results presented in this study are due to the technological, industrial and 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] Z. Minchev, et al., Digital Transformation in the Post-Information Age, 1st Edition, SoftTrade & Institute of ICT, Bulgarian Academy of Sciences, Sofia, 2022.
- [2] United Nations, Envisaging the Future of Cities, World Cities Report 2022, UN Human Settlements Programme, <https://unhabitat.org/>, 2022.
- [3] L. Boyanov, Digital World – The Change, 1st. ed., Avangard Prima, Sofia, 2021.
- [4] G. Wahlers (Ed), The Digital Future, International Reports, Konrad Adenauer Stiftung, No. 1, <https://goo.gl/8CLcvn>, 2018.
- [5] S. Bousri, Embracing Society 6.0: A Technological Renaissance for Human Living and Economies, Elit Web3 Solutions, <https://www.linkedin.com/pulse/embracing-society-60-technological-renaissance-human-living-bousri-1f/>, 2023.
- [6] United Nations, Addressing the Digital Divide Taking Action towards Digital Inclusion, United Nations Human Settlements Programme (UN-Habitat), <https://unhabitat.org/programme/legacy/people-centered-smart-cities/addressing-the-digital-divide>, 2021.
- [7] Z. Minchev, Proactive identification of future cyber threats, in: Proceedings of the 13th BISEC Conference, 2022, pp. 42–49.
- [8] Z. MINCHEV, Future transformational outlook for the digital society and economy, Romanian Cyber Security Journal. Fall 2 (2021) 25–37.
- [9] Z. B. Minchev, Human factor role for cyber threats resilience, in: Handbook of Research on Civil So-

- ciety and National Security in the Era of Cyber Warfare, IGI global, 2016, pp. 377–402.
- [10] Z. Minchev, Security challenges to critical infrastructure of future smart cities, in: Proceedings of the 9th BISEC Conference, 2019.
 - [11] CYREX 2023 Multimedia Clip, <https://youtu.be/m7mTfvtmtFc>, 2023.
 - [12] Z. Minchev, Malicious Future of AI: Transcendents in the Digital Age, in: Proceedings of the 12th BISEC Conference, 2019, pp. 18–22.
 - [13] International Research Summer School on Mathematics & Informatics - SRS'23, Multimedia Report, <https://www.globaldiplomatic.eu/post/international-research-summer-school-on-mathematics-informatics-srs-23>, 2023.

Gamification as a Tool for Elevating Password Strength Awareness

Miloš Kostić^{1,*}, Igor Saveljić¹

¹Faculty of Information Technology, Belgrade Metropolitan University, Tadeuša Košćuška 63, 11000 Belgrade, Serbia

Abstract

In modern society, where users are confronted with the necessity of managing an ever-growing number of personal profiles and accounts, low password security awareness remains a significant vulnerability in cybersecurity. Despite the existence of numerous tools designed for password safekeeping, educating users and broadening their knowledge of password strength and related cybersecurity risks cannot be understated. The popularity of gamification as an educational technique for overcoming challenges in different domains, mostly related to the lack of motivation and attention, has grown in recent years. This paper explores the concept of a two-dimensional game in which players face specific challenges aimed at replacing existing weak passwords with new, stronger ones, while avoiding the loss of access to various platforms. Time constraints and simulated cyber-attacks enhance the learning process and underscore the importance of the analyzed topic.

Keywords

Gamification, Security awareness, Games-based learning, Human-centered cybersecurity

1. Introduction

In the digital age, our society increasingly relies on the Internet for various aspects of our lives, from banking to e-commerce. Transactions conducted online often require the exchange of personal information, such as home addresses and credit card details. Within this digital landscape, passwords continue to serve as the primary authentication mechanism for accessing online services. Ensuring users remain secure while using passwords is of paramount importance. This paper seeks to address the critical need to enhance security awareness and promote better password practices through the implementation of gamification techniques.

Importance of raising password strength awareness and concept of gamification and its application within the context of the learning environment will be explored within this paper. Additionally, an concept overview of a two-dimensional game (“Lockedout”) in which players face specific challenges aimed at replacing existing weak passwords with new, stronger ones, while avoiding the loss of access to various platforms will be presented.

2. Importance of password strength awareness

The Internet presents numerous potential risks when browsing the web, such as interacting with malicious websites and domains, using inadequately constructed and weak passwords, responding to phishing emails and messages etc. These risks can place users in dangerous situations [1]. Various methods have been employed to raise user security awareness during online transactions. With the prevalence of password-related vulnerabilities, research efforts have predominantly concentrated on the creation and enhancement of security awareness tools aimed at fortifying password security.

Users often grapple with the creation and retention of strong, secure passwords, leading to various studies aimed at addressing this issue [2, 3, 4]. Experience has revealed that the prevalent method of incorporating password meters into password creation forms can frequently create a false sense of security. This is often attributed to the shortcomings in many of the available password meter algorithms, which may incorrectly label weak or poorly defined passwords as strong [5, 6]. Research suggests that additional factors should be considered when using password meters, such as user perceptions of account importance, as opposed to solely relying on the feedback provided by the meter. It becomes evident that password meters alone may not be sufficient in raising awareness and encouraging the creation of secure passwords.

Persuasive messages intended to instill fear by outlining the possible consequences of non-compliance have also been investigated as a means to boost security awareness. By educating end-users on the importance of pass-

BISEC'23: 14th International Conference on Business Information Security, November 24, 2023, Niš, Serbia

* Corresponding author.

✉ milos.kostic@metropolitan.ac.rs (M. Kostić);
igor.saveljic@metropolitan.ac.rs (I. Saveljić)

🆔 0009-0005-0912-9518 (M. Kostić); 0000-0002-0707-5174
(I. Saveljić)

word strength and heightening their awareness of associated risks, this approach has proven effective in motivating users to craft more robust passwords.

Despite ongoing efforts, issues with password hygiene persist, highlighting the necessity for more effective ways to convey password security information to users.

3. Gamification

Gamification is often described as the application of game design principles in non-gaming contexts [6, 7]. However, it encompasses more than just incorporating elements from games. It encompasses the infusion of game thinking into non-game scenarios, involving elements such as: player control, rewards, progress mechanics, collaborative problem-solving, storytelling, and even competition. At its core, gamification seeks to motivate individuals to change their behavior, primarily through enhanced engagement and motivation.

Research and recent studies have unveiled numerous instances where competitive elements successfully encouraged participants to change their behavior [6, 8]. The inclusion of competitive and cooperative elements in non-game contexts exemplifies the integration of gamification. Such gamified contexts provide a safe environment for participants to practice and hone their skills under pressure, fostering an environment of controlled learning and adaptation. Despite the growing popularity of digital or online gamified environments, gamification can also be seamlessly incorporated into tabletop contexts, using elements from card games or board games.

Studies consistently indicate a preference for gamified environments over their non-gamified counterparts among participants. The advantages of increased engagement, motivation, and skill development make gamification an attractive proposition for cybersecurity education and awareness. Nevertheless, a detailed investigation into the precise application of gamification within existing cybersecurity awareness contexts remains an underexplored area.

4. “Lockedout” – Game concept

“Lockedout” is a 2D pixel art time challenge game designed to educate players about prevalent cybersecurity risks and underscore the critical importance of password strength. It embraces a pixelated aesthetic reminiscent of video games from the 1980s and 1990s, deliberately chosen to infuse a sense of charm and playfulness into the overall gaming experience.

The game’s title (Figure 1), “Lockedout,” is a wordplay carefully selected to convey the concept of being virtually locked out due to password-related issues.



Figure 1: Current game title/logo design.



Figure 2: Password change UI.

4.1. Game Structure

In terms of UI/UX elements, “Lockedout” will revolve around the visible borders of a computer monitor, featuring a fictional operating system (OS) hosting five simulated computer applications. Additionally, an OS Guard, akin to antivirus software, will facilitate player interactions within the game and provide essential narrative elements and guidance (Figure 2.). Each of the computer applications will possess its own interface, complete with predefined content, and will serve as representations of significant daily activities necessitating robust password protection:

- Email communication
- Socializing with friends
- Online shopping
- Engaging with social media
- Managing bank account and transactions

A fully operational password checker, featuring a password strength indicator and corrective notifications, will serve as a key gameplay mechanic. The flow of gameplay will be regulated by a predefined scenario and relevant timers.

An imaginary hacker or hacker group will also be featured in the narrative; however, they will not be directly portrayed within the game.

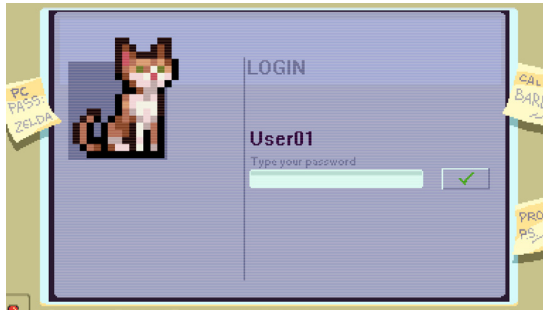


Figure 3: OS Login screen.

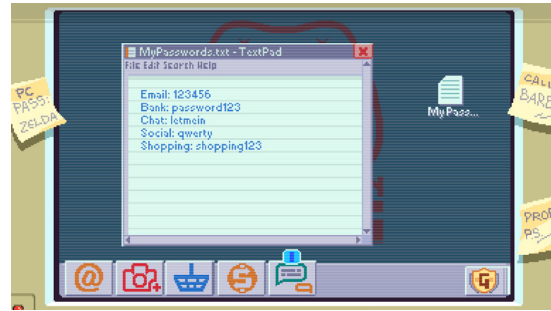


Figure 5: MyPasswords.txt document preview.

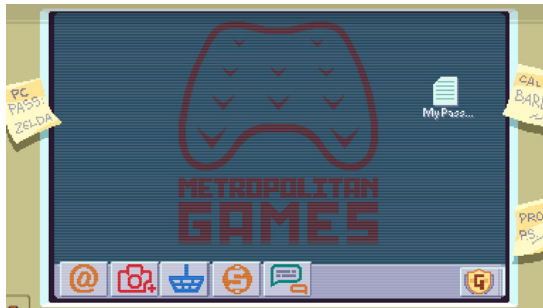


Figure 4: Desktop.

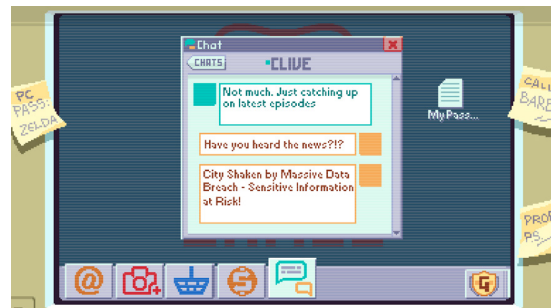


Figure 6: "Chat" app UI.

4.2. Gameplay scenario

First element of player-game interaction represents an old computer monitor with an operating system login window (Figure 3). Several sticky notes are scattered across the monitor frame, with login and password carelessly written on them. Player needs to use these written credentials in order to access the system.

Upon entering the desktop, the player encounters a file named "MyPasswords.txt" and five distinct computer application icons on the taskbar: "Email," "Bank," "Chat," "Social," and "Shopping" (Figure 4.)

At this stage, player can access the text document, or see the interface of each application and read predefined content. The text document (Figure 4) holds passwords for each application, shockingly weak and representative of statistically some of the most commonly used passwords in the world:

- Email: "123456"
- Bank: "password123"
- Chat: "letmein"
- Social: "qwerty"
- Shopping: "shopping123"

Upon a short interval, a visual and audio notification triggers within the "Chat" app (Figure 5), revealing a message from a friend inquiring about recent data breaches in

their city. As the player begins to respond to the message (or when a short timer elapses due to player inactivity), they are abruptly logged out of the chat application.

An OS Guard notification then appears, warning the player of an ongoing cyberattack (Figure 6) and prompting them to change their password to protect their account. Subsequent pop-ups follow, indicating attacks on other applications, heightening tension.

Each app screen displays a red timer, reflecting the time remaining for the player to enter their old password and generate a new, robust one. Timer durations are based on the application's importance, with the bank application's timer set to the shortest duration, emphasizing its critical nature. The chat and social media apps enjoy slightly longer timers.

In addition to time constraints, the player faces a limited number of password change attempts, with each password required to be unique and meet predefined strength criteria. Throughout this phase of the game, OS Guard occasionally provides essential feedback and tips on password strength, and the password check window informs the player of unsuccessful attempts, specifying the contributing factors.

If the timer expires on an application or if the player accumulates too many failed attempts to change a password, the hacker takes control of the app account, en-

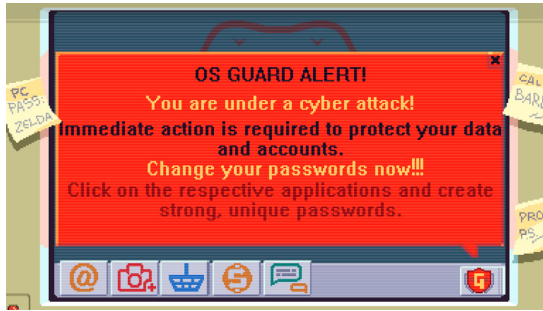


Figure 7: Guard prompt when system is under cyberattack.

gaging in malicious activities such as sending phishing emails, creating compromising posts, or initiating friend requests.

The primary objective of the game is to safeguard as many accounts as possible. Successfully changing all passwords with strength and uniqueness enables the player to defeat the hacker's attempts and receive a congratulatory victory screen. Conversely, the game concludes with a loss if the player loses access to the "Bank" app or if two or more other accounts are compromised.

Following either a win or a loss, an epilogue provides a summary of best practices for password security. It further explains why the passwords in the initial textual document were weak. Players are granted the option to delve deeper into password security through links to additional resources or tutorials.

4.3. Educational and informative aspects

The game's narrative seamlessly integrates educational content into the player's journey, resulting in an engaging and immersive learning experience. It ensures that players develop a nuanced understanding of the risks associated with weak passwords, highlighting poor practices such as storing login data in easily accessible locations like sticky notes or files on the computer desktop.

"Lockedout" offers in-game tutorials and pop-up tips to educate players on password strength, complexity, and the significance of unique passwords. Real-time feedback on password strength, accompanied by explanations of the criteria for robust passwords, enhances the learning process.

After each playthrough, an informative summary reinforces the importance of sound password practices, providing practical guidance. Furthermore, a dedicated section invites players to delve deeper into the subject, offering supplementary resources to expand their knowledge.

To ensure that game is accessible to players with various levels of gaming and technical experience, potentially

different difficulty levels will be implemented to cater to beginners and more advanced users.

5. Conclusion and future work

The gamification of password security education, exemplified by "Lockedout: Password Defense," marks a significant innovation in the realm of digital security instruction. Password security is an indispensable facet of modern life, and yet, conventional methods of education in this domain often fall short in terms of engagement and efficacy. By embracing gamification, this chapter has demonstrated the potential to transcend these limitations and foster a more interactive, enjoyable, and impactful learning experience.

"Lockedout" reinforces the significance of strong and unique passwords while actively promoting good practices and awareness. This approach is not only informative but also enjoyable, creating a transformative learning experience. "Lockedout" should represent a small step toward enhancing password security education. Its gamification principles will provide an innovative path for teaching users about the importance of strong passwords, making the educational journey more engaging and, ultimately, more effective.

Future work considers completion and refinement of all required graphics and audio elements. The game will be developed in the Unity engine, utilizing the C# programming language. This development phase includes the implementation and customization of the password-checking algorithm. Additionally, extensive testing and optimization procedures will be conducted to ensure a seamless and robust gaming experience.

Acknowledgment

This paper was supported by the Blockchain Technology Laboratory at Belgrade Metropolitan University, Belgrade, Serbia.

References

- [1] L. A. Shepherd, J. Archibald, R. I. Ferguson, Perception of risky security behaviour by users: Survey of current approaches, in: Human Aspects of Information Security, Privacy, and Trust: First International Conference, HAS 2013, Held as Part of HCI International 2013, Las Vegas, NV, USA, July 21-26, 2013. Proceedings 1, Springer, 2013, pp. 176–185.
- [2] S. L. Pfleeger, D. D. Caputo, Leveraging behavioral science to mitigate cyber security risk, *Computers & security* 31 (2012) 597–611.

- [3] S. Cohen, W. Nutt, Y. Sagiv, Deciding equivalences among conjunctive aggregate queries, *Journal of the ACM (JACM)* 54 (2007) 5–es.
- [4] J. M. Stanton, K. R. Stam, P. Mastrangelo, J. Jolton, Analysis of end user security behaviors, *Computers & security* 24 (2005) 124–133.
- [5] X. D. C. D. Carnavalet, M. Mannan, A large-scale evaluation of high-impact password strength meters, *ACM Transactions on Information and System Security (TISSEC)* 18 (2015) 1–32.
- [6] S. Scholefield, L. A. Shepherd, Gamification techniques for raising cyber security awareness, in: *HCI for Cybersecurity, Privacy and Trust: First International Conference, HCI-CPT 2019, Held as Part of the 21st HCI International Conference, HCII 2019, Orlando, FL, USA, July 26–31, 2019, Proceedings 21*, Springer, 2019, pp. 191–203.
- [7] G. Fink, D. Best, D. Manz, V. Popovsky, B. Endicott-Popovsky, Gamification for measuring cyber security situational awareness, in: *Foundations of Augmented Cognition: 7th International Conference, AC 2013, Held as Part of HCI International 2013, Las Vegas, NV, USA, July 21–26, 2013. Proceedings 7*, Springer, 2013, pp. 656–665.
- [8] I. Rieff, Systematically applying gamification to cyber security awareness trainings, 2018.

Security Related Use of Facebook as a Communication Channel

Andreja Samčović^{1,*}

¹Faculty of Transport and Traffic Engineering, University of Belgrade, Vojvode Stepe 305, 11000 Belgrade, Serbia

Abstract

This study was performed to test the opinion of students regarding the information security on the social network Facebook as a communication channel. Another goal of this research was to get information on the amount of students who are familiar with the privacy practices of this network and the risks of leaving data on social networks. Respondents were given the opportunity to make their suggestions how to raise user's awareness about the risks as well as who could assist in raising awareness of the risks about entering data on social networks. The results of a survey are presented in this paper.

Keywords

Privacy, communication, information security, social network, Facebook

1. Introduction

Today we can hardly imagine our lives without the internet as the most massive type of communication. Internet offers us answers to all our questions, and for the shortest period of time we can have access to educational materials, entertainment, business opportunity. Modern digital culture brings significant changes in different spheres of life, but it seems that most interest aroused by changes related to the mode of communication. The public is asked various questions, from how the "healthy" communication that is achieved in this way, through that as authentic to how, in the end, it's safe to communicate. The internet has certainly revolutionized our life (in many ways), and social networks have made their revolution on the internet.

When it comes to social networking, the first association is linked to Facebook (FB), the largest and most popular virtual community of its kind in the world. The functioning of such networks includes personal profile repository for sharing information through messages on the walls and the ability to facilitate social cohesion users with instant messaging (IM) and e-mail. When people join a social network, they start with creating a profile, and connect with existing friends as well as meeting other different needs for communication. Members use these sites for multiple purposes. While the initial motivation was to maintain communication and relationships, still popular activities include updating various activities, photo sharing and archiving events, getting news about the activities of friends and upcoming major

events. Users can join the networks that are organized by city, workplace, school, and region to connect and interact with other people. Also, people can add friends, send them messages, and can add new information to their profiles to notify us about new developments.

One can ask have we ever wondered what is that strange force that draws us like a magnet to read FB's "News Feed" (News), full of useless information about extraneous activities from our friends. Sociologists believe that the phenomenon is called ambient awareness (awareness of the environment) as our genetically ingrained deep in the subconscious, and it explains this phenomenon of using Facebook and Twitter. Facebook's News Feed, therefore, gives individuals the opportunity to be closer than ever with people from their surroundings, which is contrary to the claims that it leads to social isolation and alienation.

Social scientists even claim that the facts tell a different story - this is an invention that mankind actually needed, because in the last 20 years people are isolated more than ever - working from home or in remote places, resulting in a reduction of contacts with friends and acquaintances. So this is the reason why micro-blogging (write status) has become emerged as a dominant form of digital communications today. Sociologists argue that micro-blogging is drastically different from blogging. For someone to write an article on his blog, it is necessary to prepare, collect material, think, write, erase ... like when writing a book. However, micro-blogging does not require any effort. Trends show that blogging is in decline.

Another advantage of this way to communicate is that it is now much easier to solve various problems. It applies the so-called "law of weak ties." Weak links are the links to the people that mean we are not friends but acquaintances, people who are somewhere in the passage, met through work or with our real friends. Most people

BISEC'23: 14th International Conference on Business Information Security, November 24, 2023, Niš, Serbia

* Corresponding author.

✉ andrej@sf.bg.ac.rs (A. Samčović)

📄 0000-0001-6432-2816 (A. Samčović)

among their FB friends have most of these "weak ties." This has its advantages, if we find ourselves in a situation where we need a solution to a problem we are much more likely to occur with a solution. Our close friends probably will not know the solution, no matter how hard we try, because they move in similar circles as we hear similar information and know how we do. But in these situations, almost always occurs from round one "weak ties" who knows the solution.

Social networks, as well as use of the internet and mobile phones today are standard and something without which we cannot, especially among young people. How serious thinking about the Internet is the fact that the Finnish Government made a decision on which is the internet in Finland became law guaranteed, which means that every citizen of the country is in due course to have a universal, legally protected access which means it is considered that it is essential for normal life.

In addition to younger generations that use the internet as a source of knowledge, in Facebook we may find the friends we have not seen a long time, or who live far away, and keep in contact with them, and we can meet some new people, one has to recognize that Facebook has its drawbacks. As on the internet there are also a variety of opportunities for abuse on the Facebook.

2. Related research on Facebook

The use of social networks in the modern sense began in 1997 with the access of the social network SixDegrees.com that allowed its users to create profiles, list their friends of, and a year later to surf the friends list [1]. Later on, two web sites became particularly popular: MySpace and Facebook. Facebook was created in early 2004 by Mark Zuckerberg while he was a student at Harvard. First, his approach was only to the students of the University who have had their e-mail addresses in order to later, and in 2005 he spread the network to other stakeholders outside of the university network.

Report from the "Pew Internet and American Life Project" [2] showed the growing social role of communication technology in the lives of young people. According to the report 3/4 of teenagers aged between 12 and 17 years use internet and IM while these technologies are becoming an important aspect of their lifestyle. 76% of teenagers say they would miss it greatly when they could no longer use it. 48% said that they had contributed to improve the quality of their relationships with friends, while 32% think that it helps them in establishing new friendships. More than a third of adolescents use IM to communicate the content that otherwise would not dare to face to face communication, such as calls for going out (17%) or disconnecting (13%). In the project in 2007 [3], it was shown an increase in the number of online users

at 93%. Most use it for the purpose of social interaction, as a place of sharing ideas and their experiences, while the use of other technologies such as e-mail drops.

As one of the most popular social network, Facebook is a tool for communication very similar to e-mail as it is essentially based on the same technology. Most higher education institutions use this kind of communication. However, still most of the communication between teachers and students is carried out by e-mail. Teachers will certainly be willing to adopt the technology for which it is established that can facilitate communication with students. Even more significant is the social aspect of the FB. In teaching experience it looks promising to establish appropriate relationships with students by FB and similar technologies as an effective business move for the establishment of such links. While many teachers have their FB page and are actively connected with his students, some experience suggests that there must be an awareness of the problems that such activities carry [4, 5].

Mazer et al. [6] found that FB is starting to be used by both students and teachers. According to the official Facebook about 297,000 members identified themselves as teachers. They suggest that student's use FB by teachers perceived as an effort to foster positive relationships with their students. It can have a positive impact on their success and show the acceptance and understanding of contemporary student's culture. While on the other hand, this can lead to distortion of their credibility. This interaction can be a key factor in the quality of online courses [7]. FB with its unique characteristics (feed, online games and chat) can encourage interaction and involvement of users regardless of location and transcends language barriers. Researchers in the social sciences are trying to study the use of FB among the younger generations [8], to understand their online interaction, communication and identity as members of online communities.

An increasing number of publications have been dealing with the exploration of the way that young people, especially how university students use information and communication technology [9, 10, 11, 12]. Although they were designed primarily for social purposes, social networks indicate the transition to other areas of adolescent life, including education. Karlin [13] reports that almost 60 of students use social network in order to communicate on issues related to education, and more than 50 use for the exchange of experience on homework.

Salaway et al. [14] have found that students spend 18 hours on average a week in online activities. The most popular and most common online activities of students are visiting the websites of social networks [10, 15]. Quan-Haase [16] reports that 65% of students spend more than three hours a day online, 62% use e-mail per week, 67% use IM daily, while the majority of students used IM for more than 4 years. Morgan and Cotten [15] found

that students spent an average of 3.9 hours using e-mail, 16.3 hours in chatting by using IM, and nearly 12 hours per week using the Internet for purposes that do not involve communication activities, such as surfing or playing games. Hargittai [17] found that 82% of students are communicated to participate in chatting, and nearly 84% are online more than once a day.

Nagel and Kotz [18], suggest that it is necessary to pay attention to the activity on social networks and interpersonal relations as a new form of opportunities for learning, training, exchange of experience and cooperation. For example, joining groups where users share similar interests has pedagogical potential that can be used in constructive ways [19].

3. Survey results

This study should provide preliminary data on the security-related use of FB by the students. The sample consisted of 120 subjects (67% female, 33% male), predominantly students of Transport and Traffic Engineering at the University in Belgrade, Serbia. The participants voluntarily agreed to participate in the study. Students completed a survey consisting of 15 multiple-choice questions, which, in addition to gender and age information, included information how many friends they have on FB, do they accept request for friendship of unknown persons, how much time they spend on FB, do they leave any other except mandatory information, are they informed about the privacy policy of FB [<http://www.facebook.com/about/privacy>], did they adjust the privacy of their account, are they aware of risks when leaving information, do they think that information on FB is safe, has the security of their data ever been in danger, do they think that FB users in Serbia are aware of risks, which would be the most appropriate way to familiarize the users with risks as well as who should take care of informing users about the risks. Analysis of results included the calculation of basic parameters of descriptive statistics. Figure 1 provides a diagram of the structure of the sample by age groups.

As we can see from the graphic, most of the respondents were in the age group of 21 to 25 years (about 54%), but also other groups were under-represented. In the age group of 15 to 20 years 13.33% persons were interviewed, in the group of 26 to 30 years were 23% users and over 30 years we have 10% of respondents. Figure 2 provides a gender structure of the sample.

The gender composition of the sample consisted of 67% female and 33% male respondents. Figure 3 provides a graphic representation of the number of friends that the respondents had classified by different age groups. The most respondents in the age group of 15 to 20 years have more than 200 friends on the Facebook. Half of the

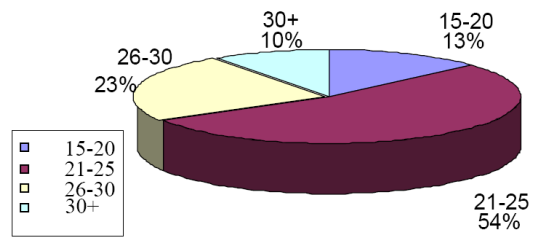


Figure 1: Structure of the sample by age groups.

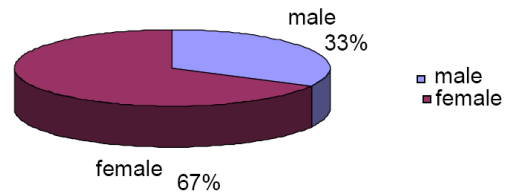


Figure 2: Gender structure of the users.

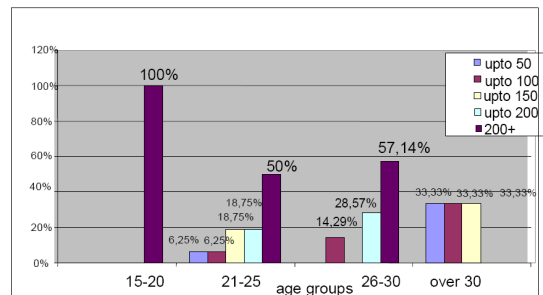


Figure 3: Number of users' friends by age groups.

respondents in the age group of 21 to 25 years have more than 200 friends and only eight users in this group have less than 100 friends. The age group of 26 to 30 years also has respondents with the largest number of more than 200 friends. Only in the respondents' group over 30 years none of the participants has more than 200 friends.

Respondents were also asked about accepting friend's requests sent by strangers and 50% of them said they do not accept such requests. However, even 43.33% of respondents sometimes accept these requests while 6.67% of the respondents accept always this request. Figure 4 provides a graph of respondents' attitudes from different age groups to accept requests for friendship from unknown people. In the age group of 15 to 20 years, in which the survey participants reported to have friends

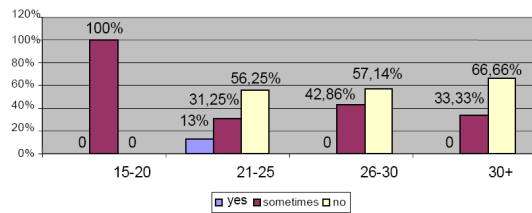


Figure 4: Attitudes of users from different age groups according to the request for friendship from unknown persons.

“mostly”, it can be seen that all of them sometimes accept friendship requests from strangers. In the age group of 21 to 25 years less than 13% of respondents would accept such a request; 31.25% of respondents would accept sometimes, while 56.25% do not accept these requests. On the other side, 42.86% of the respondents aged between 26 to 30 years sometimes accept it, while 57.14% of this group does not ever accept friend’s requests from strangers. A third of the age group over 30 years sometimes accepts this request while 66.66% (of this group) do not accept.

Based on the above diagram it can be concluded that the respondents in the age group of 15 to 20, and from 21 to 25 years are very careless because those groups have the highest percentage of respondents who sometimes accept the requests, and there are even respondents who always accept such demands. Accepting requests from strangers can significantly threaten the security of user’s information, as well as, their safety. Behind these claims may be an attacker who wants to get closer to the user to reveal some information or with the intention for physical attack on the user. Respondents from the age group of 26 to 30 years or more than 30 years are more cautious about accepting friend’s requests from strangers. None of the subjects in these two groups were ready to accept such a request while a small amount sometimes accepts such a request. The largest percentage of respondents in the two groups does not accept such requests.

As for the time spent on the Facebook, respondents identified themselves as follows: 53.33% participants said they spend more than 30 minutes, 30% spend more than 1h, 16.67% spend more than 2 hours, and only four respondents spend more than 3 hours on this network. The majority of respondents said they carried out for 30 minutes to 1 hour on this social network.

Users were also asked whether they leave some other information on Facebook, except the mandatory data, and the graphic representation of their responses was shown in Figure 5. This issue refers to the attitudes of users towards leaving photos, videos, interests, etc., on the Facebook. The vast majority, 93% of the respondents said that they leave secondary data on this network. Most respondents except profile picture inserted videos, interests, information about qualifications, status and so

Graphic presentation of survey of secondary data left on Facebook

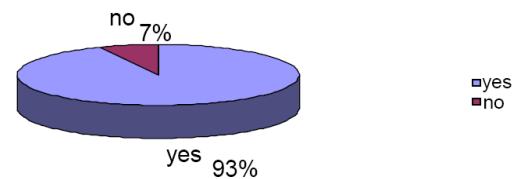


Figure 5: Do you leave some information on Facebook other than the mandatory?

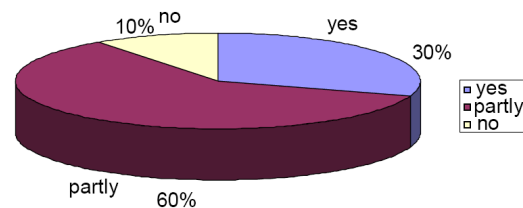


Figure 6: Overview of the familiarity of the users with the privacy policy on Facebook.

on. Almost all respondents indicate their friends in photographs and videos. Only 7% said that except for the information required for registration, leave no other information on social network Facebook.

Respondents were also asked about knowledge of the privacy policies of the Facebook. Figure 6 provides an overview of the responses to this question.

Most of the respondents i.e. 60% are only partially aware of the privacy policy of Facebook as much as 10% were not familiar at all. Only 30% of the respondents were fully aware of the privacy policy. According to the results obtained by placing the matter it can be concluded that it is necessary to inform users of social networks with privacy policies to make them aware of ways how to use the information provided by, the reasons why it is necessary to leave the data for the registration as well as the tools that they make it easier to find friends and to set up an account.

Respondents were also asked about the account settings or if their account on the Facebook is completely private, public or partially private. Figure 7 shows a graphic representation of research on privacy accounts where respondents are sorted by age group.

Research has shown that, as far as the age group of 15 to 20 years, the largest number of accounts (50%) is partially private while some information is private and some is public. A quarter of respondents in this age group reported that their accounts are public and the same percentage thinks that they have fully private accounts. In the age group of 21 to 25 years 50% of users have con-

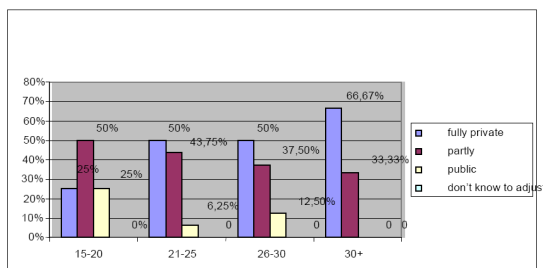


Figure 7: Overview of the policy accounts by age group.

firm that their accounts are completely private, slightly less (43.75%) of the respondents think they have partially private accounts while only 6.25% think their accounts are public. Similar results were given by the respondents in the age group of 26 to 30 years. The most of this group thought they have entirely private accounts (50%), and less of them think their accounts are partially private (37.5%) while several subjects have opinion their accounts were public (12.5%). As for the age group of the respondents that have more than 30 years, the situation looks different. In this group, even 66.67% think their accounts are completely private while 33.33% have opinion the accounts have been partially private and none of the respondents said about a public account.

4. Conclusion

On the basis of the survey presented we can come to the conclusion that it is necessary to inform the users of the Facebook about the dangers that can be caused by leaving their data on social networks. Also, it is necessary to familiarize the users with the privacy because, although the Privacy Policy is available on social networking sites, a small number of users are informed.

Users of social networks mostly have information about the risks they are exposed on social networks and how to use data that get through newspaper articles or television. However, information obtained in this way is often incomplete or inaccurate. In order to reduce attacks on user's accounts we need to appeal to customers to get more informed about the risks brought leaving data on social networks. It is also necessary that the owners of social networks appeal to users how to learn about the privacy policies of social networks and they should point to the fact that it is very important to set the privacy orders.

Acknowledgment

This work is supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia [projects TR32025, TR32048].

References

- [1] D. M. Boyd, N. B. Ellison, Social network sites: Definition, history, and scholarship, *Journal of computer-mediated Communication* 13 (2007) 210–230.
- [2] A. Lenhart, O. Lewis, L. Rainie, *Teenage life online* (2001).
- [3] A. Lenhart, M. Madden, A. Smith, A. Macgill, *Teens and social media* (2007).
- [4] N. Flynn, Facebook, take 2: Cyberbullying, *Education week* (2009).
- [5] J. R. Young, College 2.0: How not to lose face on facebook, for professors, *The Chronical of Higher Education* 55 (2009) 22.
- [6] J. P. Mazer, R. E. Murphy, C. J. Simonds, I'll see you on "facebook": The effects of computer-mediated teacher self-disclosure on student motivation, affective learning, and classroom climate, *Communication education* 56 (2007) 1–17.
- [7] M. D. Roblyer, M. McDaniel, M. Webb, J. Herman, J. V. Witty, Findings on facebook in higher education: A comparison of college faculty and student uses and perceptions of social networking sites, *The Internet and higher education* 13 (2010) 134–140.
- [8] M. A. Urista, Q. Dong, K. D. Day, et al., Explaining why young adults use myspace and facebook through uses and gratifications theory, *Human communication* 12 (2009) 215–229.
- [9] S. R. Cotten, Students' technology use and the impacts on well-being, *New Directions for Student Services* 2008 (2008) 55–70.
- [10] S. Jones, S. Fox, et al., *Generations online in 2009*, 2009.
- [11] J. Pasek, E. Hargittai, et al., Facebook and academic performance: Reconciling a media sensation with data, *First Monday* (2009).
- [12] D. F. Roberts, U. G. Foehr, Trends in media use, *The future of children* (2008) 11–37.
- [13] S. Karlin, Examining how youths interact online, *Education Digest* 73 (2007) 6.
- [14] J. B. Caruso, G. Salaway, The ecar study of undergraduate students and information technology, 2007, Retrieved December 8 (2007) 2007.
- [15] C. Morgan, S. R. Cotten, The relationship between internet activities and depressive symptoms in a sample of college freshmen, *CyberPsychology & Behavior* 6 (2003) 133–142.

- [16] A. Quan-Haase, University students' local and distant social ties: Using and integrating modes of communication on campus, *Information, Communication & Society* 10 (2007) 671–693.
- [17] E. Hargittai, A framework for studying differences in people's digital media uses, *Grenzenlose Cyberwelt? Zum Verhältnis von Digitaler Ungleichheit Und Neuen Bildungszugängen Für Jugendliche* (2007) 121–136.
- [18] L. Nagel, T. G. Kotzé, Supersizing e-learning: What a coi survey reveals about teaching presence in a large online class, *The Internet and Higher Education* 13 (2010) 45–51.
- [19] G. Blattner, M. Fiori, Facebook in the language classroom: Promises and possibilities, *International journal of instructional technology and distance learning* 6 (2009) 17–28.

Advanced Cryptography using Conformal Mappings

Marko S. Stefanović^{1,*}, Nenad O. Vesić², Aleksandra Penjišević³ and Đordije Vujadinović⁴

¹Faculty of Science and Mathematics, University of Niš, Višegradska 33, 18000 Niš, Serbia

²Mathematical Institute of the Serbian Academy of Sciences and Arts, Kneza Mihaila 36, 11000 Belgrade, Serbia

³Faculty of Management, University Union – Nikola Tesla, Cara Dušan 62-64, 11158 Belgrade, Serbia

⁴Faculty of Science and Mathematics, University of Montenegro, Džordža Vašingtona bb, 81000 Podgorica, Montenegro

Abstract

The development of new technologies and daily exposure to the Internet creates the need to hide data and securely send information. In this paper, we present an algorithm for data encryption and decryption based on the transformation of Christoffel symbols using conformal mapping, which uses structures from differential geometry. The uncountable infinities of real numbers and quadratic functions are used for better hiding of messages.

Keywords

encryption, decryption, security, metric tensor, Christoffel symbols, conformal mapping, tensor deformation, differential equations.

1. Introduction

Cryptography is a scientific discipline that deals with the study of methods for sending messages in such a form that only the person for whom they are intended they can read. Generally speaking, cryptography deals with a problem data encryption and decryption. The very word cryptography is Greek origin and means a secret (hidden) record (letter).

Some elements of cryptography were already present among the ancient Greeks. Namely, the Spartans in the 5th century BC used the device for encryption called Scytale. It was a wooden stick around which wound the string on which the message was written. After registration messages, the string would unravel, and they would remain mixed up on it signs that could only be read by someone who had a staff of equals thickness.

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, disabling any message detection, modification, or infiltration by Third person.

The communication procedure between the sender and the receiver is as follows. The Sender transforms the original message (plain text) into an incomprehensible message (cipher text) using a previously determined key. This message is hence sent to the receiver who, knowing a key, can decode the message, and therefore read it. Third person can intercept the message, or can disguise their self as receiver in order to receive this message [1].

Depending on the number of keys used in encrypting/decrypting process, there are two types of cryptographic algorithms [2, 3]:

- 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)** [1].
- 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)** [1].

2. Necessary deffinitions

Before presenting the encryption and decryption algorithm, we will present the necessary terms in differential geometry: metric tensor, Christoffel symbols, conformal mapping, tensor deformation.

An N -dimensional manifold $M_N = M(x^1, \dots, x^N)$ in whose any point is defined a symmetric (covariant) metric tensor \hat{g} whose components are g_{ij} , $g_{ij} = g_{ji}$, $i, j = 1, \dots, N$, is Riemannian space \mathbb{R}_N (see [4]). The matrix $[g_{ij}]$ is assumed to be regular, i.e. $\det [g_{ij}] \neq 0$.

BISEC'23: 14th International Conference on Business Information Security, November 24, 2023, Niš, Serbia

*Corresponding author.

✉ markoni95rened@gmail.com (M. S. Stefanović);

n.o.vesic@outlook.com (N. O. Vesić);

aleksandra.penjisevic@famns.edu.rs (A. Penjišević);

djordjijevuj@ucg.ac.me (Đ. Vujadinović)

📄 0009-0004-1961-0797 (M. S. Stefanović); 0000-0002-7598-9058

(N. O. Vesić); 0000-0002-0898-6818 (A. Penjišević)

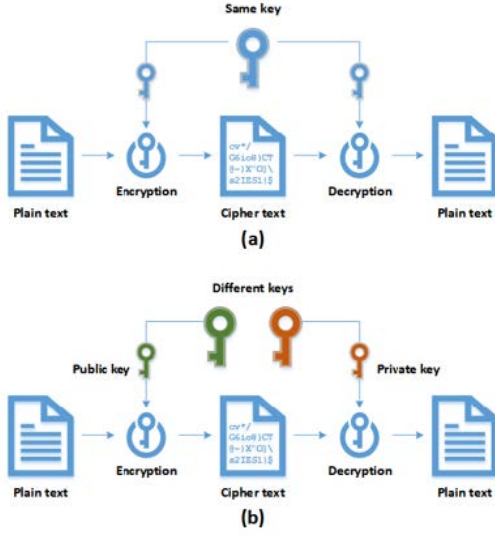


Figure 1: Symmetric and asymmetric algorithms [1]

The components of contravariant metric tensor \hat{g}^{-1} are $[g_{ij}^{-1}] = [g_{ij}]^{-1}$.

In our research, we will use the metric tensor $\hat{g} = \hat{g}(t)$, for the variable $x^1 = t$, where other variables x^2, \dots, x^N are constants. Then space $\mathbb{R}_N = \mathbb{R}_N(t)$ is space-time.

The geometrical objects

$$\Gamma_{i,jk} = \frac{1}{2}(g_{j,i,k} - g_{j,k,i} + g_{i,k,j}), \quad (1)$$

are the Christoffel symbols of the first kind. The partial derivative $\partial/\partial x^k$ is denoted by comma.

The components of Christoffel symbols for space-time $\mathbb{R}_N(t)$ are

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

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

The Christoffel Symbols of second kind are

$$\Gamma_{jk}^i = g^{ip}\Gamma_{p,jk}.$$

For the known Christoffel symbol $\Gamma_{i,jk}$, the components of corresponding metric tensor are

$$\begin{aligned} g_{ij} &= -2 \int \Gamma_{1,i,j} dt + c_{ij} = 2 \int \Gamma_{i,1,j} dt + c_{ij} \quad (3) \\ &= 2 \int \Gamma_{i,j,1} dt + c_{ij}. \end{aligned}$$

Let also $\psi = \psi(x^1, \dots, x^N) = \psi(t, \dots, x^N)$ be a scalar function. The transformation

$$g_{ij} \rightarrow \bar{g}_{ij} = e^{2\psi} g_{ij}, \quad (4)$$

is the conformal transformation of space \mathbb{R}_N to the space $\bar{\mathbb{R}}_N$. The components of metric tensor of the Riemannian space $\bar{\mathbb{R}}_N$ are \bar{g}_{ij} .

The transformation rule of Christoffel symbols Γ_{jk}^i to $\bar{\Gamma}_{jk}^i$ is

$$\bar{\Gamma}_{jk}^i = \Gamma_{jk}^i + \psi_j \delta_k^i + \psi_k \delta_j^i - g_{jk} g^{ip} \psi_p. \quad (5)$$

In the last equation, $\psi_j = \partial\psi/\partial x^j$. After contracting the equality (5) by i and k , one gets

$$\psi_j = \frac{1}{N}(\bar{\Gamma}_{jp}^p - \Gamma_{jp}^p). \quad (6)$$

The inverse transformation of the transformation (4) is conformal mapping determined by scalar function $\bar{\psi} = -\psi$.

The geometrical object $P_{i,jk} = \bar{\Gamma}_{i,jk} - \Gamma_{i,jk} = \psi_j g_{ik} + \psi_k g_{ij} - g_{jk} \psi_i$ is tensor. This tensor $P_{i,jk}$ is the tensor deformation. For encryption and decryption of texts, the tensor $P_{i,jk}$ will be used.

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

3. Algorithm for encryption and decryption

The main purpose of this work is to use terms from differential geometry, ie. conformal mappings for hiding textual data. We present an algorithm for data encryption and decryption based on the transformation of Christoffel symbols using conformal mappings.

The space-time $\mathbb{R}_N(t)$ has N dimensions, which are large enough. Suppose that linear function $b: \mathbb{N} \rightarrow \mathbb{N}$ is bijective, and array \mathcal{A} composed of M rows with not necessary equal numbers of elements. The q -th element in the p -th row of the array \mathcal{A} is marked by the object \mathcal{A}_{pq} . The transformation $u = p + M \cdot n_1, v = q + \mathcal{A}_p \cdot n_2$ is transformed the position (p, q) of a character from the array \mathcal{A} to the pair (u, v) , 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

$$z_{pq} = u + iv = p + M \cdot n_1 + i \cdot (q + A_p \cdot n_2) \quad (7)$$

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

3.1. Encryption

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 covariant vector $\psi_j = \partial\psi/\partial x^j$ have ∞ components, the numerical matrix $P_{1..ij} = \psi_i g_{1j} + \psi_j g_{1i} - g_{ij} \psi_1$ of the type (∞, ∞) , and the text τ of c characters.

Let's apply the following steps:

- **ENC1:** In text τ find the corresponding position (p_k, q_k) for the k -th character in the array \mathcal{A} .
- **ENC2:** The pair (p_k, q_k) transform to pair $(u_k, v_k) = (p_k + M \cdot m, q_k + M_{p_k} \cdot n)$, for integers m, n and the number of elements in the p_k -th row of array \mathcal{A} equal M_{p_k} .
- **ENC3:** The pair (u_k, v_k) transform to polynomial

$$\pi_k(t) = t^2 - 2u_k t + u_k^2 + v_k^2. \quad (8)$$

- **ENC4:** Create the ordered set

$$\Pi = \{\pi_1(t), \dots, \pi_c(t)\}$$

- **ENC5:** Create $\tilde{N} = \frac{N(N+1)}{2} - c$ polynomials

$$\pi_{c+u}(t) = t^2 - (r_{c+u} + s_{c+u})t + r_{c+u}s_{c+u},$$

$$u = 1, \dots, \tilde{N} \text{ for integers } r_{c+u}, s_{c+u}.$$

- **ENC6:** 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.

- **ENC7:** Create the square matrix $[h_{ij}]$ of the type $N \times N$ whose elements are

$$h_{ij} = \begin{cases} \pi_{i_j}^*(t), & i \leq j, \\ \pi_{j_i}^*(t), & i > j, \end{cases} \quad (9)$$

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

- **ENC8:** Expand the matrix $[h_{ij}]$ to the matrix $[g_{ij}]$ with elements

$$g_{ij} = \begin{cases} p_{11}(t), & i = j = 1, \\ 0, & i = 1 \text{ and } j > 1, \\ 0, & j = 1 \text{ and } i > 1, \\ h_{(i-1)(j-1)}, & \text{otherwise.} \end{cases} \quad (10)$$

- **ENC9:** Form the matrix

$$\Gamma = [\Gamma_{1..ij}] = \begin{bmatrix} \Gamma_{1..22} & \dots & \Gamma_{1..2N} \\ \vdots & \ddots & \vdots \\ \Gamma_{1..N2} & \dots & \Gamma_{1..NN} \end{bmatrix} \quad (11)$$

of the corresponding Christoffel symbols with respect to the metric tensor whose components are $[g_{ij}]$.

- **ENC10:** From the matrix P , select the submatrix $P_{1..ij}$ of the type $N \times N$ from the up left angle of the matrix P .

- **ENC11:** Form the matrix

$$\bar{\Gamma} = [\Gamma_{1..ij} + P_{1..ij}]$$

$$= [\Gamma_{1..ij} + \psi_i g_{1j} + \psi_j g_{1i} - g_{ij} \psi_1].$$

- **ENC12:** The components g_{ij} of the matrix $[g_{ij}]$ are of the form

$$g_{ij}(t) = t^2 + p_{ij}t + q_{ij}. \quad (12)$$

The corresponding covariant affine connection coefficients are of the form

$$\Gamma_{1..ij} = -t - \frac{1}{2}p_{ij} + P_{1..ij}$$

$$= -t - \frac{1}{2}p_{ij} + \psi_i g_{1j} + \psi_j g_{1i} - g_{ij} \psi_1.$$

The corresponding constant c_{ij} from the equation (3) is $c_{ij} = q_{ij}$.

Output

Public key

$$\begin{aligned}\bar{\Gamma}_{1,ij}(0) &= [-p_{ij} + P_{1,ij}] \\ &= [-p_{ij} + \psi_i g_{1j} + \psi_j g_{1i} - g_{ij} \psi_1].\end{aligned}$$

Message is

$$\mu = [q_{ij} - b(0)].$$

To decrypt the public key, we transform the public key $\bar{\Gamma}$ to $\Gamma = [\bar{\Gamma}_{1,ij} - P_{1,ij}] = [\bar{\Gamma}_{1,ij} - \psi_i g_{1j} - \psi_j g_{1i} + g_{ij} \psi_1]$. The elements of matrix Γ are the free particles of the Christoffel symbols obtained in step **ENC9**. In this way, we obtained the polynomials which hid the text. After solving the equations $\Gamma_{1,ij} = 0$, and using the operation

$$\text{Mod}1n [p, q] = \begin{cases} q & q|p, \\ \text{Mod} [p, q], & \text{otherwise.} \end{cases} \quad (13)$$

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.

Our future work could be a connection between Differential geometry and Fuzzy logic, like it was done in [5, 6].

4. Conclusion

The main aim of this paper is the presentation of an algorithm based on conformal mappings, that can be used for text-data hiding and information processing. The shown encryption and decryption algorithms enable safe communication and message transmissions, possibly applied in digital and e-learning education and economics. Since the need for secure communications is more important than ever, the potential for using these mappings in the design of modern security protocols is high, especially on IoT hardware platforms and blockchain based healthcare applications, as well as other blockchain medical scenario applications. With the constant increase in threats in the cyber world, an algorithm like the one presented in this paper could be used in protocols deployed on multiple layers, for instance in machine learning security, smart cities security platforms, and lightweight cryptography.

Acknowledgments

Nenad Vesić wishes to thank Serbian Ministry of Science, Technological Development and Innovations which supported his work through the Mathematical Institute of Serbian Academy of Sciences and Arts.

References

- [1] D. Simjanović, N. Vesić, B. Randelović, N. Zdravković, Đ. Vujadinović, A new cryptographic algorithm based on affine connection coefficients, in: BISEC'2021 - The Twelfth International Conference on Business Information Security, 2021, pp. 71–74.
- [2] B. Schneier, Applied cryptography: protocols, algorithms, and source code in C, John Wiley & Sons, 2007.
- [3] N. O. Vesić, D. J. Simjanović, Matrix-based algorithm for text-data hiding and information processing, Vojnotehnicki glasnik/Military Technical Courier 62 (2014) 42–57.
- [4] J. Mikeš, E. Stepanova, A. Vanžurová, S. Bácsó, V. Berezovski, O. Chepurna, M. Chodorová, H. Chudá, M. Gavrilchenko, M. Haddad, et al., Differential geometry of special mappings, Palacký University Olomouc, Czech Republic, 2015.
- [5] D. Simjanović, N. Vesić, B. Randelović, Đ. Vujadinović, Cyber security criteria: Fuzzy AHP approach, in: BISEC'2022 - The Thirteenth International Conference on Business Information Security, 2022, pp. 62–67.
- [6] D. Simjanović, N. Zdravković, B. Randelović, N. Vesić, Utilizing ahp for smart-city development with blockchain-based solutions for healthcare, government and education, in: ICIST 2022 Proceedings, 2022.

Secure Course Completion Credentialing using Hyperledger Fabric

Stefan Gogić¹, Nemanja Zdravković^{1,*}, Emilija Kisić¹ and Ponnusamy Vijayakumar²

¹Faculty of Information Technology, Belgrade Metropolitan University, Tadeuša Košćuška 63, 11000 Belgrade, Serbia

²SRM IST, ECE Department, Kattankulathur, Chennai, India

Abstract

In this paper, we present a blockchain solution, based on Hyperledger Fabric, for issuing and validating documents from Higher Education Institutions (HEIs), such as diplomas and diploma supplements. By utilizing Hyperledger Fabric, the most popular distributed ledger technology for private blockchains, we propose a lightweight and secure credentialing three layer blockchain system – the smart contract layer, the blockchain layer itself, and the network layer. With a minimal needed number of functionalities such as issuance and verification, our lightweight system can be deployed on a trustful environment, e. g. faculties from the same university, or a consortium of universities. With such an environment, we eliminate the need for a computationally complex consensus mechanism for adding blocks to the ledger, while retaining easy implementation with the HEIs information system and/or learning management system. Based on previous research and prototyping, our model acts as an additional security layer on top of and HEI's information system and utilizes blockchain's immutable property to keep student's records secure.

Keywords

blockchain, credentialing, distributed ledger, Hyperledger

1. Introduction

Blockchain technologies (BCTs) and distributed ledger technologies (DLTs) have surpassed their initial use in cryptocurrencies, and are already being used in a plethora of fields – from supply chain managements and healthcare, to predictive maintenance systems and public sector [1, 2, 3, 4, 5]. With the rise of Ethereum and its smart contracts written in Solidity, presenting code which can be directly run on the chain itself, paired with a robust consensus mechanism, a secure and immutable record keeping solution in a trustless environment without the need of third-party stakeholder has risen, identifying BCTs/DLTs as disruptive technologies [6].

Credentialing solutions for Higher Education Institutions (HEIs) based on blockchain and similar technologies are still few. As of writing this paper, only a small number of papers have been published [7, 8, 9] compared to other blockchain-based use cases. For instance, one of the main conclusions found in one of the earliest studies on the topic of blockchain in education state that BCTs (and later DLTs) should allow users to be able to automatically ver-

ify the validity of certificates in a direct manner, without contacting the HEI that originally issued the documents [7]. Indeed, the authors of [8] state that BCT/DLT-based systems promise a permanent authentication and storage solution for the alternative credentials market. This continuously growing market consists of various kinds of microcredentials, nanodegrees, MOOCs/SPOCs, certificates and/or badges from various types of training and pre-qualification programs. The authors also emphasize scalability issues, most noticeably if the BCT/DLT use the computationally complex Proof-of-Work (PoW) consensus mechanism, as does Bitcoin and many other cryptocurrency networks. The PoW approach will likely remove the need for educational organizations to validate credentials, and other lightweight approaches are needed.

Since the initial hype of using BCT/DLT for various use cases including ones in education, the authors of [9] conducted a literature review of solutions based on public blockchains, highlighting the need for a standardized approach built on a public blockchain to promote faster adoption and acceptance. This recent study states that full functioning and active prototypes are still low in numbers; however, one of the conclusions was that the blockchain application should run on a stable, secure, and trustworthy network.

Indeed, in a trustless environment where actors are not known, public BCTs with robust consensus mechanisms such as Biction are imperative [10, 11, 12]. However, mechanisms such as PoW or various variations of Proof of Stake (PoS) are computationally complex and require powerful, often dedicated computers equipped with a

BISEC'23: 14th International Conference on Business Information Security, November 24, 2023, Niš, Serbia

*Corresponding author.

✉ stefan.gogic.6354@metropolitan.ac.rs (S. Gogić);
nemanja.zdravkovic@metropolitan.ac.rs (N. Zdravković);
emilija.kisic@metropolitan.ac.rs (E. Kisić); vijayakp@srmist.edu.in
(P. Vijayakumar)

🆔 0000-0002-2631-6308 (N. Zdravković); 0000-0003-3059-2353
(E. Kisić); 0000-0002-3929-8495 (P. Vijayakumar)

powerful central processing unit (CPU) and/or graphic processing unit (GPU). Conversely, in a more specific environment, i. e. where the nodes in the blockchain network are known (and trusted) parties, a blockchain-based solution with less complex consensus mechanism can be implemented, retaining security with the added benefit of not needing a powerful CPU/GPU to handle blockchain transactions. Usually, this approach is called a distributed ledger technology (DLT).

The authors' main motivation is to utilize a trustful environment and propose a lightweight framework for document credentialing, tailored specifically to HEIs and the issuance and validation of student diplomas and diploma supplements.

Based on literature, commercially (un)available solutions and our own previous attempts, we have identified the following research questions:

- *RQ1*: Is it possible to design a lightweight framework for the specific needs of HEIs to incorporate document issuance and validation in a secure manner, without relying on complex solutions?
- *RQ2*: Can the flexibility of Hyperledger Fabric be used as a basis for incorporating a BCT/DLT-based addition to an existing HEI information system (IS)?

The rest of the paper is organized as follows. Section 2 gives a brief introduction on blockchain technologies, focusing on Hyperledger Fabric. Afterwards, Section 3 gives presented the proposed system, developed at Belgrade Metropolitan University's (BMU's) Blockchain Technology Laboratory. Finally, Section 4 gives a conclusion, with current limitations and further research ideas.

2. Blockchain and Hyperledger overview

In this Section, we firstly provide a brief overview of the building blocks of a general blockchain system. Afterwards, we focus on the Hyperledger DLT solution, of which Hyperledger Fabric is used to develop the credentialing system.

2.1. Brief blockchain overview

In general, BCTs impose a fundamental change to manner various types of data are processed, and can improve existing data security solutions. A blockchain can be viewed as a shared, append-only distributed ledger, in which all events are stored in linked blocks [13]. These events are often referred as transactions. A copy of the ledger is therefore kept by all nodes which form the blockchain network. Due to the fact that all member nodes have a

copy of the ledger, all network nodes are updated in real time, simultaneously. Further, a block can be viewed as a data structure consisting of the following:

1. a header which connects the new block to the previous one.
2. a list of transactions;

Each transaction, besides the data, contains a header with a timestamp, paired with an unique cryptographic signature, thus enabling the ledger to be resistant to modifications. This chain of blocks that is formed and continuously updated can be traced back all the way to the first block, named the genesis block.

The combination of peer-to-peer networking, public-key cryptography, and distributed consensus is what secures blockchain transactions. Conversely to a centralized system, no single entity i.e. node should be able to control the process of adding a block to the chain. As the blockchain is a distributed system, each new block addition is managed by all nodes who share equal rights. This mechanism is utilized in order to overcome security issues, and is achieved through the process known as distributed consensus. This process can be viewed as an agreement among the nodes in the network how to validate each block yet to be added to the chain. Depending on the consensus mechanism, nodes can either compete for correct transaction validation (PoW), be chosen randomly (PoS and its variations), or apply a different algorithm altogether. The algorithms used can vary in computational complexity.

Finally, it is important to note that blockchain are a class of technology; the term refers to different forms of distributed databases with variations in their technical and governance arrangements and complexity.

2.2. Hyperledger and its use cases

Hyperledger is the leading open source community focused on developing various stable frameworks, tools and libraries for enterprise-grade distributed ledger deployments [14]. This community aims to advance BCT/DLT technologies by identifying and more importantly realizing a cross-industry open standard platform for DLTs. The aim of the open standard is to transform the approach to business transactions on a global level [14]. Hyperledger has a modular approach to hosting projects similar to the approach of the Linux Foundation, as shown in Fig. 1. All Hyperledger projects are open source, they are easy to obtain [15]. All Hyperledger projects, with the exception of Hyperledger Indy, are used for general purpose blockchain-based applications and solutions, whereas Hyperledger Indy focuses on decentralized identity [16].

One of the key differences between the various BCTs/DLTs systems is the utilized consensus mechanism. Due

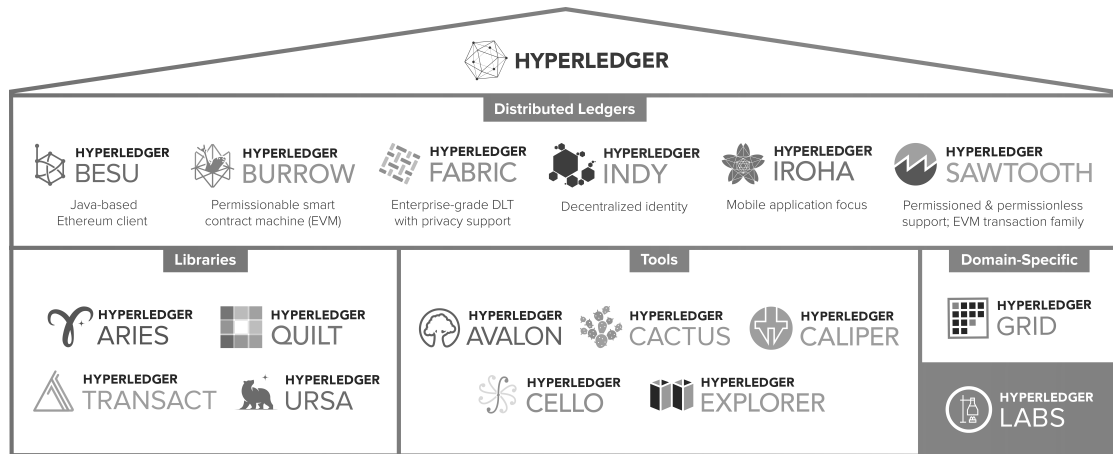


Figure 1: The Hyperledger Project umbrella [15].

to the variety of blockchain usage requirements, Hyperledger provides several different consensus mechanisms [17]. For instance, Fabric uses the Apache Kafka platform [18] as the main Crash Fault Tolerance (CFT) protocol on the network which is permissioned i.e. private, and it is voting-based. Hyperledger Indy utilized a consensus based on Redundant Byzantine Fault Tolerance (RBFT), a protocol inspired by Plenum Byzantine Fault Tolerance (Plenum). Hyperledger Iroha used a variant of the BFT algorithm called Sumeragi, which tolerates more than one Byzantine faulty network nodes. Hyperledger Sawtooth facilitates the so-called pluggable consensus for both lottery and voting algorithms. By default, Hyperledger Sawtooth uses a lottery-based, Nakamoto consensus algorithm called Proof of elapsed time (PoET). Hyperledger Burrow comes with Byzantine Fault-Tolerant Tendermint protocol with a greater transaction rate, whereas Buru implements various consensus algorithms that are involved in transaction validation, block validation, and block production, i.e. mining in the PoW mechanism, while Hyperledger Sawtooth has the most support for smart contract languages [16].

The core Hyperledger-based use cases include banking, healthcare, supply chain management, financial services, information technology, government, and media and entertainment. Indeed, the Hyperledger Foundation promotes a range of business DLTs, including many libraries and tools that provide support for the creation, maintenance, deployment, providing cryptographic work, etc [15].

For the proposed system, the authors have opted to use Hyperledger Fabric, as it is the Hyperledger project with most testing, working real-world applications community, and documentation. The details of Hyperledger Fabric are listed in Table 1.

Table 1
Hyperledger Fabric features

Advantages	Enterprise backing Relative maturity Private channels Modular architecture Smart contracts
Consensus mechanism	Kafka RAFT Solo
Smart contract technology	Chaincode
Smart contract type	Installed
Smart contract language	Go Java Javascript Solidity
State storage	CouchDB leveldb

3. System model

BMU's ongoing internal R&D includes implementing blockchain in education and e-learning. BMU's Blockchain Technology Laboratory (BCT Lab) is investigating which blockchain technology is most suitable for applying in education, with emphasis on data protection. BMU's BCT Lab is collaborating with ISUM (Information System of University Metropolitan) and BMU's e-Learning center. During a four month testing developing and period, a working prototype for credentialing was developed. The proposed model is comprised of three layers, stacked on top of the zeroth layer, which is the HEI's IS:

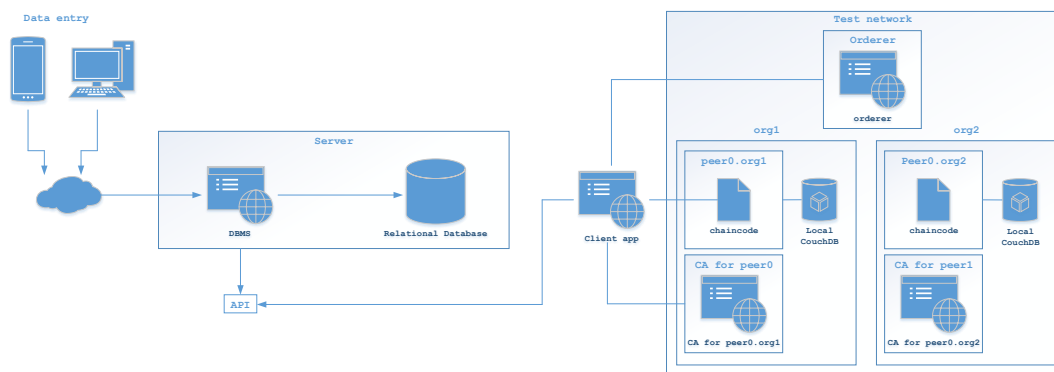


Figure 2: Proposed system consisting of a data entry system and the test blockchain network, communicating over an API.

1. the smart contract layer;
2. the blockchain layer itself;
3. the network layer.

The smart contract layer runs the chaincode to add the data to a block. It is present on every node, denoted as a peer. The blockchain layer consists of the peer itself, a Certification Authority (CA) for that peer, and a local NoSQL database - CouchDB. The network layer consists of the test network with two peers, denoted with org1 and org2.

The system was developed in two stages – Stage 1 consists of using an isolated GIT branch of the HEIs to add a functionality to export diploma supplemental materials as an API to the blockchain network. Stage 2 comprised of developing a lightweight blockchain application, based on Hyperledger Fabric, to connect the the API and add the data to a block. The architecture of the two-stage system is shown in Fig. 2.

The main parameter which Hyperledger Fabric uses is the transaction context `ctx`. It holds the needed information for transaction logic "per transaction" or "per contract". IT enables to access the `stub` which allows various blockchain operations such as state returns, adding a new item to the block, or getting all blocks (in our case diploma supplements).

To add a diploma supplement, it is needed to connect to the peer node using a gateway, and to get the chaincode from the network.

To write the transaction i.e. diploma object, an asynchronous promise function will get all the necessary parameters for add a new diploma supplement, as shown in Fig. 3. It will create a new object with those parameters which will be later added to the blockchain using `stub` API operations.

The data which is added to the blockchain has the following structure:

```

35  async dodajDiplomu(ctx, idDiplome, ime, prezime, studijskaGrupa, ocene) {
36      const diploma = {
37          idDiplome,
38          ime,
39          prezime,
40          studijskaGrupa,
41          ocene
42      }
43      await ctx.stub.putState(idDiplome, Buffer.from(350N.stringify(diploma)));
44  }
45  }

```

Figure 3: Asynchronous promise function.

```

const diplomas =
[
  {
    "name": "Firstname",
    "surname": "Surname",
    "studygroup": "StudyGroup",
    "grades":
    [
      {
        "grade": "GradeValue",
        "course": "CourseCode",
      },
      ...
    ]
  },
  ...
]

```

When data is added, a message can be viewed in the console terminal to confirm a successful transaction, as shown in Fig. 4. In our testbed, an endpoint was not deployed from the IS's side; therefore we have manually added the data in the same format as the HEI's IS would provide.


```
Transaction has been evaluated, result is: [{"Key": "ID0", "Record": {"idDiplome": "ID0", "ime": "Stefan", "ocene": [{"\ocena\": \"10\", \"naziv\": \"CS101\"}, {\ocena\": \"10\", \"naziv\": \"CS102\"}, {\ocena\": \"10\", \"naziv\": \"IT101\"}], \"prezime\": \"Gogic\", \"studijskaGrupa\": \"IT\"}}, {"Key": "ID1", "Record": {"idDiplome": "ID1", "ime": \"Milos\", \"ocene\": [{\ocena\": \"10\", \"naziv\": \"CS115\"}, {\ocena\": \"10\", \"naziv\": \"IT210\"}, {\ocena\": \"10\", \"naziv\": \"MA101\"}, {\ocena\": \"10\", \"naziv\": \"IT381\"}], \"prezime\": \"Vasov\", \"studijskaGrupa\": \"SI\"}}]
```

Figure 4: Transaction successfully added.

4. Conclusion

In this paper, we have used Hyperledger Fabric to develop a lightweight blockchain network for credentialing HEI's diplomas and diploma supplements. Currently, our system only addresses the issuance use-case, while validation use-case remains open. As prototyping was conducted in an isolated environment, several open issues still remain. Firstly, should the blockchain remain private, or be public (where anyone can be a part of the network)? As the target group of the system are first and foremost HEIs, the authors, as was discussed in other literature as well, opt for a private blockchain solution, where the HEIs comprise the network. Still, there exists a possibility to add the learners as nodes as well.

Using Hyperledger Fabric, data such as diplomas and supplements can be issued and verified reliably. Blockchain can help learning platforms to add an additional layer to their credentialing process. We have presented a blockchain-based credentialing system can be easily deployable and connected to a learning platform. Within our proposed system, upon generating the certificate file for the diploma and/or supplement, the HEI's IS will make a transaction to the blockchain. This entry will also have the certificate information, alongside metadata required for the transaction header. This information will be encrypted, and can be accessed only by the IS, the student, and an authorized third party.

This new issuance transaction is sent to the blockchain, where the other nodes in the network will verify it and add it to the blockchain using a simpler consensus mechanism. Each node will have a local copy of the blockchain on a NoSQL database like CouchDB. For certificate validation, upon receiving the access link, the student or an authorized third party can verify the digital credential by accessing the blockchain through a query. If a match is found on the blockchain, the certificate file is validated and a corresponding message appears.

The innate immutability property of BCT/DLT does not allow fraudulent or modified certificate files to be deemed as verified. Any tampering to the certificate file will result in a vastly different hashed value of the file, ensuring impossible verification.

Acknowledgment

This paper was supported by the Blockchain Technology Laboratory at Belgrade Metropolitan University, Belgrade, Serbia.

References

- [1] B. K. Mohanta, S. S. Panda, D. Jena, An overview of smart contract and use cases in blockchain technology, in: 2018 9th international conference on computing, communication and networking technologies (ICCCNT), IEEE, 2018, pp. 1–4.
- [2] K. Zile, R. Strazdiņa, Blockchain use cases and their feasibility, *Applied Computer Systems* 23 (2018) 12–20.
- [3] P. Zhang, D. C. Schmidt, J. White, G. Lenz, Blockchain technology use cases in healthcare, in: *Advances in computers*, volume 111, Elsevier, 2018, pp. 1–41.
- [4] M. Alabadi, A. Habbal, Next-generation predictive maintenance: leveraging blockchain and dynamic deep learning in a domain-independent system, *PeerJ Computer Science* 9 (2023) e1712.
- [5] V. Milicevic, N. Zdravkovic, J. Jovic, On the selection of suitable blockchain technologies for supply chain management, *International Journal for Quality Research* (2023).
- [6] N. Zdravković, J. Jović, M. Damjanović, Secure credentialing in e-learning using blockchain, in: *Proceedings of the 11th Conference on eLearning (eLearning-2020)*, 2020, pp. 39–42.
- [7] A. Grech, A. F. Camilleri, *Blockchain in education*, Luxembourg: Publications Office of the European Union, 2017.
- [8] M. Jirgensons, J. Kapenieks, Blockchain and the future of digital learning credential assessment and management, *Journal of teacher education for sustainability* 20 (2018) 145–156.
- [9] G. Caldarelli, J. Ellul, Trusted academic transcripts on the blockchain: A systematic literature review, *Applied Sciences* 11 (2021) 1842.
- [10] P. Ocheja, B. Flanagan, H. Ueda, H. Ogata, Managing lifelong learning records through blockchain, *Research and Practice in Technology Enhanced Learning* 14 (2019) 1–19.

- [11] F. R. Vidal, F. Gouveia, C. Soares, Revocation mechanisms for academic certificates stored on a blockchain, in: 2020 15th Iberian Conference on Information Systems and Technologies (CISTI), IEEE, 2020, pp. 1–6.
- [12] F. Vidal, F. Gouveia, C. Soares, Analysis of blockchain technology for higher education, in: 2019 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC), IEEE, 2019, pp. 28–33.
- [13] Z. Zheng, S. Xie, H. Dai, X. Chen, H. Wang, An overview of blockchain technology: Architecture, consensus, and future trends, in: 2017 IEEE international congress on big data (BigData congress), IEEE, 2017, pp. 557–564.
- [14] C. Cachin, et al., Architecture of the hyperledger blockchain fabric, in: Workshop on distributed cryptocurrencies and consensus ledgers, volume 310, Chicago, IL, 2016.
- [15] The Hyperledger Foundation, <https://www.hyperledger.org>, 2023.
- [16] V. Milićević, J. Jović, N. Zdravković, An overview of hyperledger blockchain technologies and their uses, in: Proceedings of the 11th International Conference on Information Society and Technology (ICIST 2021), 2021, pp. 62–65.
- [17] J. Moubarak, E. Filiol, M. Chamoun, Comparative analysis of blockchain technologies and tor network: Two faces of the same reality?, in: 2017 1st Cyber Security in Networking Conference (CSNet), IEEE, 2017, pp. 1–9.
- [18] B. R. Hiranman, et al., A study of apache kafka in big data stream processing, in: 2018 International Conference on Information, Communication, Engineering and Technology (ICICET), IEEE, 2018, pp. 1–3.

Deep Blockchain to Enable Scalable Web Applications

Yajna Pandith^{1,*}

¹Bengaluru, Karnataka, India

Abstract

The work delves into the exploration of deep blockchain architecture involving the introduction of higher-layer blockchains, which summarize their blocks through anchor transactions integrated into the blocks of lower-layer blockchains. The architecture is structured as follows: (I) Layer 1 - MainNet: This layer serves as the repository for registered Layer 3 blockchain roots and Layer 2 Block Merkle roots. (II) Layer 2 - Plasma Cash chain: This layer facilitates the storage of Plasma tokens, which can be redeemed for bandwidth, along with Layer 3 Block hashes. (III) Layer 3 - Multiple blockchains: These blockchains leverage the storage and bandwidth capabilities provided by Layer 2, enabling seamless packaging of NoSQL/SQL transactions and similar database operations. Sparse Merkle Trees is employed extensively, demonstrating their efficacy in delivering provable data storage through the use of Deep Merkle Proofs. Our objective is to present results highlighting re-markable throughput and low latency for Layer 3 blockchains built upon economically secure Layer 2 Plasma Cash blockchains. Collectively, these advancements lay a solid foundation for the development of scalable web applications. Our research paves the way for innovative solutions in various industries that can scale modern web applications successfully, ensuring unwavering data integrity, enhanced security, and optimized efficiency.

Keywords

deep blockchain, data storage, web applications

1. Introduction

Layer 1 blockchains such as Ethereum and Bitcoin, on their own, cannot support the latency and throughput needs for modern web applications [1]. Attempting to support higher throughput or lower latency with naive solutions (e.g. larger blocks, lower security consensus algorithms, etc.) sacrifices the core benefits of layer 1 blockchains [2]. It is unnecessary to make these sacrifices in the name of scalability for blockchains: when one blockchain is capable of storing and retrieving state, then another blockchain's summary state variables may be stored there. This can be done in layers, where *Layer i+1* blockchain's state is stored in *Layer i* blockchains and each blockchain uses a well-motivated consensus engine to achieve Byzantine fault tolerance [3]. Using this layered approach, the key elements of a *deep blockchain* architecture can be specified. The blockchain paradigm [4] that forms the backbone of all decentralized consensus-based transaction systems to date is as follows. A valid state transition for a blockchain of Layer *i* is one which comes about through a transaction T_j^i :

$$\sigma_{t+1}^i = \Upsilon^i(\sigma_t^i, T_j^i) \quad (1)$$

where Υ^i is the Layer *i* blockchain state transition function, while σ_t^i enables components to retain arbitrary state between transactions. Transactions are organized

into blocks, which are interlinked through a parent hash within each block to reference the preceding block. Together, these blocks serve as a ledger, with block hashes employed to spot the ultimate state:

$$\sigma_{t+1}^i \equiv \Pi^i(\sigma_t^i, B_j^i) \quad (2)$$

$$B_j^i \equiv (\dots, (T_{j_0}^i, T_{j_1}^i, \dots)) \quad (3)$$

$$\Pi^i(\sigma_t^i, B_j^i) \equiv \Omega^i(B_j^i, \Upsilon^i(\sigma_t^i, T_{j_0}^i, T_{j_1}^i, \dots)) \quad (4)$$

where Ω^i is the block finalization state transition function for layer *i*, B_j^i is the *j*th block of layer *i* (which collates transactions and other components), and Π^i is the block-level state transition function for layer *i*.

In a **deep blockchain** system, a blockchain layer *i* is said to be *connected* to layer *i + 1* if:

1. there exists a transaction mapping function Λ^{i+1} mapping blocks at layer *i + 1* into transactions T_k^i at layer *i* for all layer *i + 1* blocks B_j^{i+1}

$$T_k^i \equiv \Lambda^{i+1}(B_j^{i+1}) \quad (5)$$

2. there exists a mapping function $\Xi^i(k)$ retrieving from blockchain layer *i* a mapping $f(B_k^{i+1})$ of the blocks state of layer *i + 1* for all blocks B_k^{i+1}

$$\Xi^i(k) \equiv f(B_k^{i+1}) \quad (6)$$

A natural choice for transaction mapping $\Lambda^{i+1}(B_j^{i+1})$ may be to include a block hash b_k^{i+1} of the block B_k^{i+1}

BISEC'2023: 14th International Conference on Business Information Security, November 24, 2023, Niš, Serbia

*Corresponding author.

✉ yajnapandith@gmail.com (Y. Pandith)

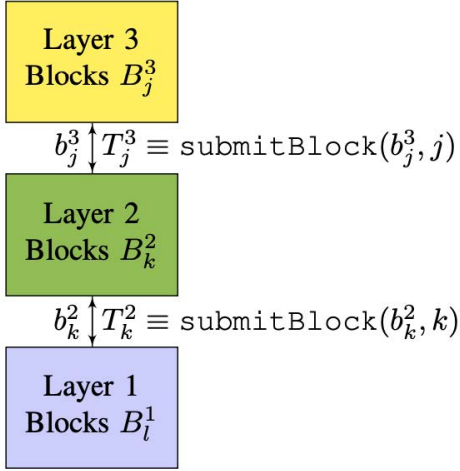


Figure 1: Deep Blockchains: In the deep blockchain architecture explored here, each layer $i + 1$ is connected to layer i with transactions submitted to layer i for every block at layer $i + 1$. Typically, block hashes and Merkle roots are submitted in transactions as key attributes of the block.

as a transaction T_k^i [5], and for the lower layer to provide the block hash back (see Figure 1 (left)). This paper, demonstrates a deep blockchain system for provable storage, situating a “Plasma Cash” design [6] in a Layer 2 Blockchain and NoSQL/SQL/File Storage for any number of Layer 3 Blockchains (see Figure 2).

Historically, the low-throughput high-latency of Layer 1 blockchains resulted in immediate pressure to drive activities off-chain [7], but only a few “off-chain” attempts can be considered deep blockchains because they lack the connected blockchains. Layer $i + 1$ and layer i may be explicitly connected in a deep blockchain system for many different reasons:

1. Higher throughput services at layer $i + 1$ may be paid for using the value held in layer i currency
2. Storing a limited set of information in layer $i + 1$ in layer i may support the security and provenance of layer i
3. Proof of fraud at layer $i + 1$ can be used for economic consequences at layer i

The nascent label “Layer 2” encompasses many newly developing notions ranging from state channels to almost any approach that may help Layer 1 scale (e.g. bigger blocks), but the term “deep blockchain” is not used for all Layer 2 notions but specifically for any situation where one or more blockchains are *connected* in the above way.

2. Layer 2: Plasma Cash Blockchain

Seminal insights on multi-layer blockchains were put forth by [8], which have inspired many “Plasma” designs, and specifically motivated our implementation of what has been termed “Plasma Cash” for tracking storage and bandwidth balances. The Layer 2 Plasma Cash blockchain is connected to Layer 1 using the following trust primitives:

- **User Deposit:** When Alice wishes to use the services enabled by the Layer 2 blockchain, Alice deposits some Layer 1 currency λ_{dep} (.01 ETH or 1 WLK) in a Layer 1 contract function (`createBlockchain`); the deposit event results in Alice owning a Layer 2 token τ through a Layer 2 Deposit transaction included in a Layer 2 block.
- **User Transfer:** When Alice wishes to transfer her Layer 2 token τ to another user Bob or the Plasma operator Paul, Alice signs a Layer 2 token transfer transaction specifying the recipient and the previous block. This Layer 2 transaction is included on the Layer 2 blockchain by Paul.
- **Layer 2 Block Connection:** The operators of the Layer 2 blockchain mints new Layer 2 blocks B_j^2 with a collation of Layer 2 transactions T_k^2 (with a consensus protocol such as Quorum RAFT and POA in permissioned networks or Ethereum Casper for permissionless networks) from the User Deposit and User Transfer transactions. Each Layer 2 block B_k^2 has its Merkle Root b_k^2 submitted to Layer 1 with a transaction $T_k^2 = \text{submitBlock}(b_k^2, k)$ recorded in a Layer 1 block B_l^1 . The recipient Bob of a token transfer must receive the full history of all transactions from Alice and verify it against these Merkle Roots b_k^2 stored in Layer 1, all the way to the original deposit. If any transaction in the history cannot be verified by Bob, Bob cannot accept Alice’s token as payment.
- **User Exit:** When Alice wishes to withdraw her token τ for Layer 1 cryptocurrency, she calls `startExit` function with the last 2 transactions¹ which can be verified against and Merkle proofs that must match the stored Merkle roots to be a valid exit; if no one challenges the exit, Alice receives the outstanding token balance within a short time period when exits are finalized.
- **User Challenges:** If the operator Bob or another user Charlie notices that Alice’s exit attempt is invalid, it submits a Merkle proof and rewarded when a valid challenge indicates a invalid exit.

¹As to why *two*, two is indicative, but not conclusive concerning Alice’s ownership, therefore a user challenge process is required.

Remarkably, users of the Layer 2 blockchain can conduct their business securely even when the Layer 2 operator has 100% control over the Layer 2 blockchain! Any sign of malicious operator Paul and the Layer 2 users can exit, and all Layer 2 token values remain secure. How can practitioners reconcile instincts to pursue this objective:

Blockchain 1.0 Objective: *Maximize decentralization.*

with an obviously centralized operator? The answer is to pursue a more nuanced objective of

Blockchain 2.0 Objective: *Maximize the cost of successful attacks.*

With the Plasma Cash construct, the Blockchain 2.0 Objective is achieved with:

1. Layer 1 Smart Contracts supporting a Layer 2 Connection to Layer 1 storage that collectively make the cost of attacking the Layer 2 blockchain the same as the cost of attacking the Layer 1 blockchain – for Ethereum and Bitcoin Layer 1 blockchains, this is the famous “51% attack”, for others it might be whatever is required to control the state of that Layer 1 Blockchain.
2. Layer 1 Cryptocurrency being used for value transfer of services between users of the Layer 2 Blockchain and the Layer 2 operator mediated through deposits, token transfers and exits mediated by Layer 1 constructs

With the Layer 2 Block Connection and trust primitives in place, Layer 2 can operate at much higher throughput than Layer 1 because of its reduced consensus, but continuing to inherit Layer 1’s cost of attack and achieving the more fundamental objective. Therefore practitioners of deep blockchain engineering must develop different instincts, incorporating different software trust primitives between different constructed layers to achieve the same objective depending on the structure of between layers and the value unlocked in each.

3. Deep Blockchains for Provable Data Storage

The specific deep blockchain system that has developed extends the Blockchain 2.0 Objective up one more layer by incorporating trust primitives (Block transactions, Sparse Merkle Trees) in provable NoSQL, SQL and Storage services, shown in Figure 2: Layer 3 NoSQL, SQL and Storage blockchains rest on the storage and bandwidth services of Layer 2, which supervene on the decentralized computation and payment services of Layer 1. Our work follows Ethereum SWARM’s foundational work on storage and bandwidth [9] which outlines the following ideas that is situated in multiple layers:

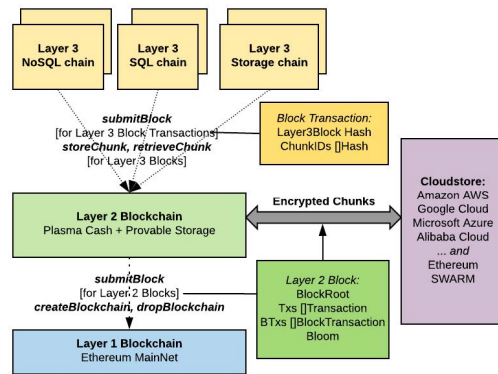


Figure 2: Deep Blockchain for Storage: Users of NoSQL/SQL/Storage Layer 3 blocks createBlockchains on Layer 1, and use Layer 2 Plasma Cash to operate their blockchain. Layer 3 blocks are submitted to the Layer 2 Blockchain with block transactions and chunks are insured. Plasma Tokens are used for bandwidth. Cloudstore combines major computing platforms with Ethereum SWARM for both resilience and speed.

- A chunk of bytes v is stored in Cloudstore using 256-bit hash $k = H(v)$ as the key to retrieve v . Nodes that request a chunk by key k can verify correctness of the value v returned from Cloudstore simply by checking if $k = H(v)$.
- Insurers of chunks can earn Layer 1 currency with valid Merkle proofs; Failure to provide valid proofs result in severe insurance payouts
- Bandwidth consumed by a node, when hitting the nodes threshold must result in signed payments

Layer 1 blockchains were initially developed without the concern for storage models being competitive with cloud computing platforms or even a passing concern for bandwidth; the birth of Bitcoin and Ethereum Layer 1 focused on birthing trustless payments and trustless computation mediated by a peer-to-peer network, rather than about nodes providing decentralized storage [10]. In contrast, decentralized *storage* networks, as manifested in Ethereum SWARM and many other systems, promises to have a large peer-to-peer network of nodes sharing the responsibility to keep a portion of the world’s data and compensated proportionately for the commodity storage and bandwidth they provide. In these networks, a distributed hash table (typically, with Kademlia routing layers) is used for logarithmic look ups of chunks, but in practice, $O(\log_2(n))$ retrieval times are just not competitive with modern UI expectations or typical developer expectations. Nevertheless, decentralized storage networks have a critical role to play in providing censorship-resistance. Rather than layer 3 rest solely on a decentralized stor-

age network (which is slow but resilient and censorship-resistant), layer 3 can rest on *both* decentralized storage networks *and* mature modern cloud computing platforms. Again, the Blockchain 1.0 Objective must be displaced in favor of the Blockchain 2.0 Objective: in this sense, more storage variety *increases* the cost of attack.

Putting the elements together in a deep blockchain system for provable storage:

- Layer 1 blockchain: When a developer wishes to have a Layer 3 blockchain for NoSQL/SQL/Storage, they send Layer 1 currency into `createBlockchain(blockchainName string)` on MainNet; this can be refunded with a `dropBlockchain(blockchainName string)` operation (taking place of `startExit`). When storage is used in `blockchainName` through the activities of Layer 3 blockchains (as recorded by the Layer 2 blockchain below), this balance goes down. Balances can added to and withdrawn by the owner of the blockchain.
- Layer 2 Plasma Cash Blockchain: The storage and retrieval of chunks in Cloudstore are exposed to Layer 3 blockchains with the following 2 APIs (see Appendix A):
 - `storeChunk(k, v, τ, ω)` - stores a key-value pair mapping (k, v) in Cloudstore, backed by Layer 2 token τ (signed with ω) received from the Layer 1 transaction.
 - `retrieveChunk(k, τ, ω)` - retrieves a key-value pair mapping (k, v) in Cloudstore, backed by Layer 2 token τ (again, signed with ω), and returning the balance of τ used so far

The Layer 2 operator will store via Cloudstore in as many regions and cloud providers as necessary to **insure** the chunk as follows: A new type of Layer 2 **block transaction** insures a set of chunks recorded through `storeChunk` calls. The cause of these chunks are from any Layer 3 blockchain needing storage and bandwidth, where bandwidth is used in `retrieveChunk` calls. When a Layer 3 blockchain mints Layer 3 blocks, the Layer 3 blocks themselves contain a Cloudstore key that references a list of chunks written in the Layer 3 block. The block itself is stored in Cloudstore with another `storeChunk` call, signed by the Layer 3 blockchain owner, and the block hash b_k^3 is submitted by the Layer 3 blockchain to the Layer 2 blockchain via a `submitBlock(b_k^3, k)` block transaction. This enables the Layer 2 blockchain to meter the cumulative storage of `blockchainName` and deduct from the balance originally deposited in the `createBlockchain(blockchainName string)`

operation (approximately every 24 hours), passing on Cloudstore costs to Layer 3 blockchains. Notably, Layer 3 blocks themselves are recorded with `storeChunk(k, v, τ, ω)` to store the layer 3 block in Cloudstore and then results in a call to `submitBlock(b_j^3, j)`:

$$T_j^3 \equiv \text{submitBlock}(b_j^3, j) \quad (7)$$

Because the block storage is signed and because the block transactions are signed, Layer 2 operators collect storage payments with the layer 3 blockchain operator consent, forming a kind of “state channel” within the deep blockchain. Taken together, this is the Layer 3 Block Connection, as seen in Figure 1. The Layer 2 block consists of:

- the transaction root θ_k^2 that utilizes the SMT structure to represent just the tokens τ_1, τ_2, \dots spent in block k

$$\theta_k^2 \equiv \text{KT}((\tau_1, T_{\tau_1}^2), (\tau_2, T_{\tau_2}^2), \dots) \quad (8)$$

- the token root τ_k for *all* tokens τ_j, \dots

$$\tau_k \equiv \text{KT}((\tau_1, T_{\tau_1}^2), (\tau_2, T_{\tau_2}^2), \dots) \quad (9)$$

- array of token transactions T_k^2
- array of block transactions \tilde{T}_k^2 from all Layer 3 blockchain operators using Layer 2 services
- an account root, using an SMT to store an accounts “balance” and a list of tokens held by that account.
- Layer 3 blockchains: Any number of Layer 3 blockchains that utilize storage and bandwidth can be layered on top of the Layer 2 blockchain, regularly submitting lists of chunks based on the structure of the Layer 3 blockchain.
 - For NoSQL + File Storage, there is a key for each row of NoSQL or File, and a value for the row (a JSON record) or raw file contents. The root hash changes when any table is added/removed or when any table schema is updated, and where each table has a root hash that changes when any record of the table is changed; any new database content results in new chunks, where the chunk is referenced by the hash of its content.
 - For SQL, there is a root hash for each database, where the root hash changes when any table schema is updated, and where each table has a root hash that changes when any record of the table is changed; any new database content results in new chunks, where the chunk is referenced by the hash of its content.

Both NoSQL and SQL Blockchains is described in Section 5.

Just as with Layer 1 blockchain nodes, running Layer 2 and Layer 3 blockchains consists of running a node within the framework of a decentralized system, retrieving and relaying messages about new transactions and new blocks. Wolk’s blockchain implementations of the Layer 2 and Layer 3 originated from Ethereum’s `go-ethereum` and JPMorgan’s `Quorum RAFT` code bases, written in Golang. RAFT is used for both Layer 2 and Layer 3 implementations due to its simple model of finality. For each blockchain, Golang package is created containing each of the interfaces specified in Appendix A, and adapted `Quorum RAFT` code to conform to these interfaces. There is no explicit assumption that permissioned consensus algorithms be used, however. The choice of RAFT was made purely out of simplicity, its maturity as a code base, and its capacity for high throughput – any consensus protocol that achieves finality can fit within this deep blockchain architecture. For both the Layer 2 and Layer 3 blockchains, Sparse Merkle Tree is used to support provable data storage.

4. Sparse Merkle Trees and Provenance

The Sparse Merkle Tree (SMT) is a persistent data structure that map fixed q -bit keys to 256-bit values in an abstract tree of height q with 2^q leaves for any set \mathcal{J} :

$$\mathcal{J} = \{(\mathbf{k}_0 \in \mathbb{B}_q, \mathbf{v}_0 \in \mathbb{B}_{256}), (\mathbf{k}_1 \in \mathbb{B}_q, \mathbf{v}_1 \in \mathbb{B}_{256}), \dots\} \quad (10)$$

The function of the SMT is to provide a unique Merkle root hash that uniquely identifies a given set of key-value pairs \mathcal{J} , a set containing pairs of byte sequences. Each key stored in the SMT defines a Merkle branch down to one of 2^q leaves, and the leaf holds only one possible value for that key in \mathcal{J} . The bits of the q -bit key define the path to be traversed, with the most significant bit at height $q - 1$ and least significant bit at height 0. Following [11] and [12], to compute the Merkle root of any SMT in practice and allow for the ideal computation of Merkle branches for the n Merkle branches, it is useful pre-compute a set of default hashes $d(h)$ for all heights h from $0 \dots q - 1$ levels: (shown in Figure 3)

- At level 0, $d(0) \equiv H(0)$
- At level h , $d(h) \equiv H(d(h - 1), d(h - 1))$

Logarithmic insertion, deletion and retrieval operations on the SMT are defined with elemental operations:

- `insert(k, v)` - inserts the key by traversing chunks using the bytes of k

- `delete(k)` - deletes the key by inserting the null value for k into the SMT
- `get(k)` - gets the value from the SMT through node / chunk traversal

Typically, block proposals with SMTs as a core data structure involve bulk combinations of the above, with many inserts and deletes mutating the content of many chunks, and the Merkle root only being computed as a final step.

Sparse Merkle Trees are best suited for a core primitive over more familiar Binary Merkle Trees (BMTs) because:

- when an id (a tokenID, a document key in NoSQL, a URL in File storage, a table root in SQL) is mapped to a value, you can guarantee that the id has exactly one position in the tree, which you don’t get with BMTs.
- when an id is NOT present in the SMT, you can also prove it with the same mechanism. This approach proves beneficial in situations where Bloom filters produce erroneous matches.
- A Merkle proof for the id mapped to a specific value is straightforward, and because of sparseness the number of bytes required is much less than the depth of the tree

The key concept behind SMTs is the efficient representation of included IDs using n hashes at n out of 2^q leaves. Each ID, represented as a q -bit number, is associated with either a null value or its corresponding hash at a leaf node. Instead of using a Merkle proof consisting of 64 32-byte hashes from the leaf to the root, a compact representation can be achieved using `proofBits`, a q -bit value (e.g., `uint64`). Each bit in `proofBits` indicates whether the sisters on the path to the root use default hashes (0) or non-default hashes stored in `proofBytes`. The `proofBytes` array exclusively consists of non-default hashes, while the value stored at the leaf level is the 32-byte RLP hash.

For the Layer 2 Block Connection, a call to `checkMembership(bytes32 leaf, bytes32 root, uint64 tokenID, uint64 proofBits, bytes proofBytes)` helper function in Ethereum MainNet can take `proofBits` and `proofBytes` and prove that a exit or challenge is valid if it matches the Merkle roots provided by the Plasma operator in a call to

```
submitBlock(bytes32 root)
```

Similarly, when a user receives a token from another user, they must obtain the `tokenID`, along with t raw transaction bytes and t Merkle proofs. Each Merkle proof corresponds to a specific block and verifies the token spend. It’s important to note that a non-spend can also be proven, where the leaf is represented by $H(0)$.

In the optimal scenario, an SMT representing a single key-value mapping ($n = 1$) reduces the proof size

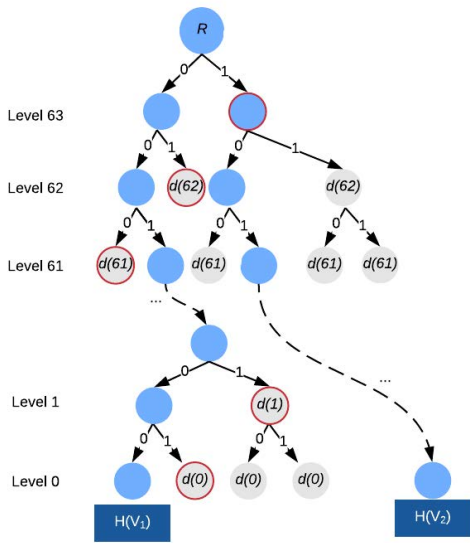


Figure 3: Sparse Merkle Tree Illustration: Merkle branches for 2 64-bit keys $k_1 = 001. .00$ and $k_2 = 101. .$ hold $H(V_1)$ and $H(V_2)$ in a unique SMT root R for a 2 key set $\mathcal{J} = \{(K_1, V_1), (K_2, V_2)\}$. Since there are only keys in this tree, the default hashes $d(h)$ (outlined in red) appear starting at level 62, so the branches K_1, K_2 (shown in blue circles) have Sparse Merkle proofs using default hashes from level 0 to level 62, which can be specified in a `proofBits` parameter. This makes for very tiny proofs and lower gas costs on Main-Net.

significantly. Instead of a 64×32 byte proof, the entire path from level 0 to level 63 consists of default hashes, and `proofBits` is a 64-bit value filled with zeros (`0x0000000000000000`). In this case, `proofBytes` is empty, and the `uint64` value is 0, resulting in the most compact proof size possible: 8 bytes.

In the following favorable scenario, considering 2 ids (for example, `0x01234. . .` and `0x89abc. . .`), the proof of spend for each token would include a single non-default hash at the topmost level 63, and `proofBits` would consist of the value 1 followed by 63 zeros (`0x8000000000000000`). The resulting proof size would be 40 bytes.

In typical scenarios, SMTs exhibit high node density in the upper levels, ranging from level $q - 1$ down to approximately level $\log_2(n)$. To illustrate this, consider a situation where you have 10MM Layer 2 tokens, and each token undergoes 500 transactions per token per year. This results in a total of 5B transactions for the 10MM tokens annually. Assuming a Layer 2 block frequency of 15s/block, these 5B transactions would be distributed across 2.1MM blocks per year, with an average of 2,378

transactions per Layer 2 block ($500 \times 10 \times \frac{10^6}{86400 \times \frac{365}{15}}$). When incorporating these 2,378 transactions into an SMT, given that $\log_2(2378) = 11.2$, you will have a densely populated set of nodes, mostly consisting of non-default hashes, from levels 63 down to approximately level 53. Below that, you will have only one tokenID extending all the way to level 0. The proof size would amount to 32 bytes per level, resulting in a total of 320 bytes for 10 levels.

$q = 64$ is decided instead of $q = 256$ because:

- collisions are still unlikely at $q=64$... until it is around 4B keys
- the `proofBits` are 24 bytes smaller (`uint64` instead of `uint256`)
- less gas is spent in `checkMembership` on all 0 bits in `proofBits`
- smaller 64-element array of default hashes computed instead of 256 hashes

Reducing the frequency of hashing leads to decreased gas consumption and increased user satisfaction, particularly in the Level 2 block connection. In this context, it ensures that collisions between circulating tokenIDs can be definitively ruled out during deposit events. Moreover, you can combine the fixed length `proofBits` and variable length `proofBytes` into a single proof bytes input for exits, i.e. `startExit(uint64 tokenID, bytes txBytes1, bytes txBytes2, bytes proof1, bytes proof2, int blk1, int blk2)` The analogous challenge interfaces will then have fewer argument inputs in the same way.

The sparseness of the SMT derives from the observation that keys will extremely rarely share paths at increasingly lower heights and naturally will share paths at increasingly higher paths. This lends itself to a representation where the SMT is chunked by byte k_i , where traversing the SMT from a root chunk (representing a range of keys from 0 to $2^{64}-1$) down to an intermediate chunk with just one leaf involves processing one additional byte, which each chunk of data storage having up to 256 child chunks specifying a range of keys each child possessing a range that is $\frac{1}{256}$ smaller. Just as with a radix tree, the SMT is traversed from root to leaf, with an additional byte of the key causing a read of a chunk that represents up to 256 branches and the hashes of all the branches, utilizing default hashes. Golang "smt" package is implemented and a "cloud" package to map SMT operations into Cloudstore.

5. Layer 3 Blockchains

With the foundations of Layer 2 providing storage and bandwidth paid for with Layer 2 tokens, any number

of Layer 3 blockchains may be constructed. The construction of a NoSQL and SQL blockchain is detailed here. At a high level, Layer 3 blockchains collate SQL and NoSQL transactions in Layer 3 blocks submit block transactions to Layer 2, and Layer 2 collate token and block transactions with Merkle Roots of token root and blocks submitted and included in transactions to Layer 1 blockchain. It then becomes possible to aggregate multiple proof of inclusions at the highest layers all the way to MainNet with Deep Merkle Proofs, which is illustrated here.

5.1. Layer 3 NoSQL Blockchain and Deep Merkle Proofs

To support Layer 3 NoSQL transactions in a NoSQL blockchain, the Layer 3 blockchain has a layer 3 block structure defined as collating a set of NoSQL records along with a Layer3KeyRoot of a Sparse Merkle Tree managing a set of key-value pairs of “documents”. All NoSQL records are encrypted using counter mode (CTR) encryption defining operations $encrypt(d, \pi)$ and $decrypt(d, \pi)$ and utilizing a database encryption key π known only to the layer 3 blockchain user. Three operations are defined, each of which map into the SMT data structure:

- `SetKey(k, v)` - stores arbitrary k, v , through a `storeChunk(k, v)` Layer 2 operation and a Layer 3 SMT operation on κ (`insert(H(k), H(encrypt(v, π)))`)
- `GetKey(k)` - retrieves the value v stored in the `SetKey(k, v)` operation, through Layer 3 operation on κ `get(H(k))` which returns v_h followed by `decrypt(retrieveChunk(v_h), π)`
- `DeleteKey(k)` - removes k from the NoSQL database, by storing $(H(k), 0)$ in the SMT; subsequent calls to `GetKey(k)` will not return a value.

The minting of a new Layer 3 NoSQL Block consists of taking each of the Layer 3 transactions (`SetKey, DeleteKey`), executing `storeChunk` Layer 2 API calls for its users. Unless two transactions operate over the same key k , all transactions can be executed in parallel. If multiple transactions operate over the same key, only the last received transaction will have its mutation succeed.

example (shown in Figure 4)

- In Layer 3 Block 302, the user wishes store document ID 1 with key K_1 mapped to encrypted value V_1 and document ID 2 mapped to encrypted value V_2 . The user can submit 2 Layer 3 NoSQL transactions:

```
[SetKey( $K_1 = 0b001\dots00$ ,
```

```
 $V_1 = 0x778899\dots$ ])
SetKey( $K_2, V_2$ )]
```

which results in a set of SMT primitive operations:

```
insert(H( $K_1$ ), H(encrypt( $V_1, \pi$ ))),
insert(H( $K_2$ ), H(encrypt( $V_2, \pi$ )))
```

resulting in `Layer3KeyRoot = 0x83fc\dots`. The chunks for both documents $H(V_1)$ and $H(V_2)$ along with chunk of the previous block 301 (e.b. `storeChunk(0b001\dots\dots)`) are included in Layer 3 Block 302 in the `ChunkIDs` and insured with a call to

```
submitBlock(0b101\dots11, 302)
```

submitted to the Layer 2 blockchain.

- When the Layer 2 blockchain processes the block transactions from this new Layer 3 block (and many other Layer 3 blockchains) to build Layer 2 Block 2002, it will build a SMT with `insert(concat(blockchainName, 302), 0b001\dots00` (and other inserts) to generate a `BlockRoot` (e.g. `0x4d69\dots`). As is standard, the new `BlockRoot` uses the previous blocks `BlockRoot` as a starting point. Packaging the block transactions together with any token transactions (balance updates, transfers, deposits, etc.), the new layer 2 block 2002 with hash `0xe8db\dots` will be stored in Cloudstore with a call to `storeChunk(0xe8db\dots)` Txns and submitted to Layer 1 with a call to

```
submitBlock(0xe8db\dots, 2002).
```

- Finally, a Layer 1 Block (e.g. 10,000,002) will be proposed by some MainNet miner including the above Layer 2 `submitBlock` transaction and eventually be finalized by the Layer 1 consensus protocol.

A Deep Merkle Proof is formed through the aggregation of each proof of inclusion across each layer of blockchain connections in a deep blockchain down to Layer 1. In our 3 layer deep blockchain with the Layer 3 NoSQL blockchain layered on the Layer 2 Storage / Plasma Cash blockchain, there is a Layer 2-3 connection and a Layer 1-2 connection. So a full Deep Merkle Proof that a NoSQL document K_1, V_1 is included in the deep blockchain all the way up to MainNet consists of:

1. Layer 3 proof of inclusion of $(H(K_1), H(V_1))$ in Layer 3 block `Layer3KeyRoot` – in our example, this would be that the value $H(V_1)$ hashes up to SMT root $R_{302} = 0x83fc\dots$

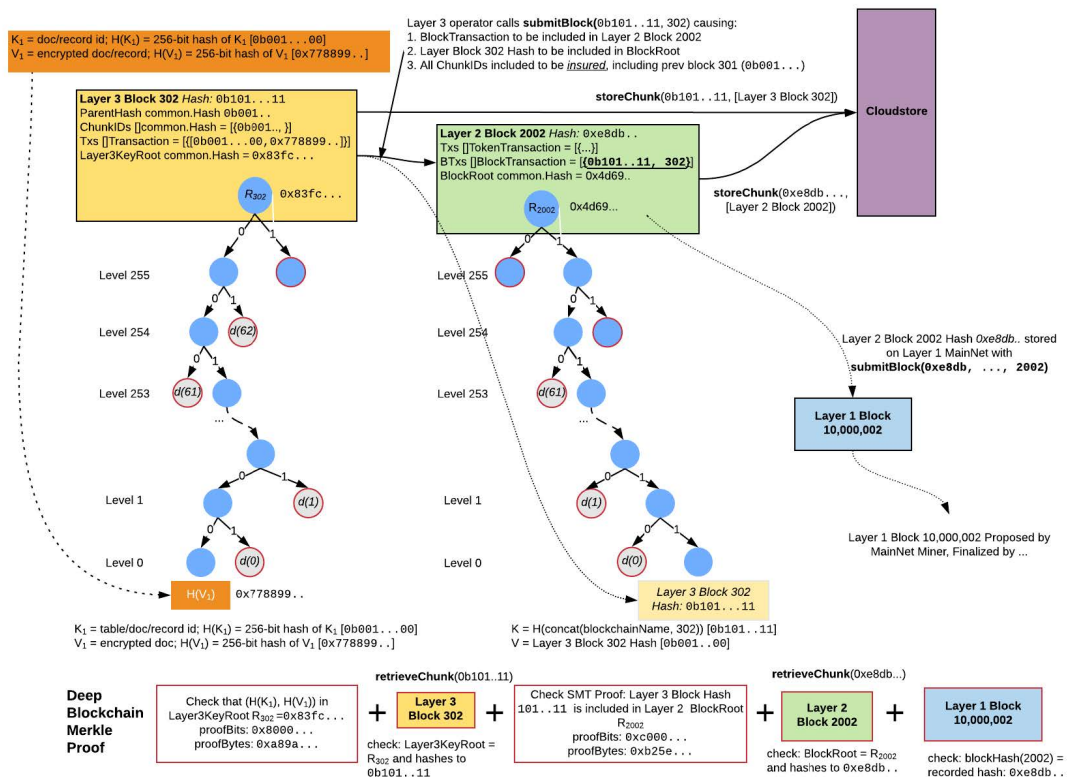


Figure 4: Deep Merkle Proof illustrated: The hashes key-value pair is recorded in a Layer 3 Sparse Merkle Tree, the root of which is kept in Layer3KeyRoot in the Layer 3 block. When the Layer 3 operator uses `submitBlock` to submit a *block transaction* to Layer 2, Layer 3 Block hash 302 is included in another SMT maintained by the Layer 2 operator storing all Layer 3 block hashes of all Layer 3 blockchains. When the Layer 2 block 2002 is minted, the `BlockRoot` is set and included in Layer 1 Block 10,000,002. The individual proof of inclusion from the 2 SMTs and the portions of the raw Layer 3 and Layer 2 block form a *Deep Merkle Proof* for inclusion a specific record in the deep blockchain, from the highest layer to Layer 1.

2. Layer 2 proof of inclusion of $(H(\text{concat}(\text{blockchainName}, k)))$ in Layer 3 block Layer3KeyRoot – in our example, this would be that the Layer 3 block hash 0b101...11 hashes up to SMT root $R_{2002} = 0x4d69...$
3. Layer 1 proof of inclusion of the Layer 2 block hash in the `blockHash` array of the Layer 1 Smart Contract – in our example, this is that `blockHash(2002) = 0xe8db`

In our implementation, deep Merkle proofs are provided in response to `GetKey(K, V)` to the layer 3 blockchain users as an optional deep boolean parameter and when true, returns the full combination of:

- Layer 3 Block, which includes Layer3KeyRoot
- `proofBits` and `proofBytes` for the

Layer3KeyRoot, which are shown to match $H(K), H(V)$

- Layer 2 Block, which includes `BlockRoot`
- `proofBits` and `proofBytes` for the BlockRoot, which are shown to match the Layer 2 Block Hash
- Layer 1 `blockHash` record of the Layer 2 block number

The concept of a deep Merkle Proof is not limited to 3 layer deep blockchains, nor is the concept only applicable to NoSQL blockchains – the concept applies to multiple layers of proof of inclusion enabled through the general layering processes of deep blockchain systems generally.

5.2. Layer 3 SQL Blockchain

To support Layer 3 SQL operations in a SQL blockchain, the Layer 3 block has a structure defined as having as packing a set of encrypted SQL transactions (insert/update/delete statements) along with a `Layer3KeyRoot` of a Sparse Merkle Tree representing a set of table root hashes.

In our implementation, Quorum RAFT is adopted as the consensus layer for our layer 3 SQL blockchain (again, following Appendix A), which collectively follow a consensus protocol where once a *leader* has been identified, the leader mints a new Layer 3 block based on:

- An array of SQL transactions that is mapped into newly created chunks (created via `storeChunk` for table root hashes)
- An array of table root hashes, key-value pairs written to `Layer3KeyRoot`, based on the execution of the above SQL Transactions

The minting a layer 3 block consists of the leader compiling each SQL transaction into a set of instructions to be executed by a “SQL Virtual Machine” (SVM) based off of the widely used SQLite’s virtual machine. In this model, a virtual machine has a program counter that increments or jumps to another line after the execution of each opcode instruction. For example, a SQL statement of “Select * from person” received by a node is mapped into an interpretable set of opcodes like this:

```
{ "n": 0, "opcode": "Init", "p2": 8, "p4":
  ": "select * from person" }

{ "n": 1, "opcode": "OpenRead", "p2":
  : 2, "p4": "2" }

{ "n": 2, "opcode": "Rewind", "p2": 7 }

{ "n": 3, "opcode": "Column", "p3": 1 }

{ "n": 4, "opcode": "Column", "p2":
  : 1, "p3": 2 }

{ "n": 5, "opcode": "ResultRow", "p1":
  : 1, "p2": 2 }

{ "n": 6, "opcode": "Next", "p2":
  : 3, "p5": 1 }
{ "n": 7, "opcode": "Halt" }

{ "n": 8, "opcode": "Transaction", "p3":
  : 3, "p4": "0", "p5": 1 }

{ "n": 9, "opcode": "Goto", "p2": 1 }
```

In our SVM Golang implementation, all opcodes are mapped into Layer 2 `storeChunk` and `receiveChunk` calls, manipulating the following chunk types:

- Database Schema chunk: represents up to 32 tables belonging to the “blockchainName”. Each table is identified by name (up to 32 bytes) and has a *table chunk*;
- Table chunk: represents up to 32 columns belonging to a specific “table”. Each column is identified by name (up to 27-bytes) and additional information: its column type (integer, string, float, etc.), whether it is a primary key, and any index information; a 32-byte chunk ID points to a potential *index chunk*, if the column is indexed A table must have at least one primary key.
- Index chunk: a B+ tree, composed of intermediate “X” chunks and data “D” chunks. Each X chunk has 32-byte pointers to additional X chunks or D chunks. D chunks form an ordered doubly linked list, and contain pointers to *record chunks*.
- Record chunk: a 4K chunk of data that holds a JSON record for a keyed value.

Our current implementation has a full implementation of single table operations thus far, but with relational database operations approachable with the same dynamics:

- When the owner of a Layer 3 blockchain creates a new database, the owner chunk is updated and database chunk is created and the owner chunk is updated with the new database chunk information. If this is the first database created by the owner, the root hash of the owner is set for the first time. The root hash of the database is set for the first time.
- When the owner of a Layer 3 blockchain creates a new table, the database chunk is updated and table chunk is created and the database chunk is updated with the new table information. This also causes the owner chunk to be updated with the new database chunk information. The root hash of the table is set for the first time in the child chain.
- When the owner of a Layer 3 blockchain creates or updates a table, this creates or changes the database schema chunk. The database chunk is then updated with the new schema information, which in turn causes the owner chunk to be updated with the new database chunk information.
- When an owner creates a new record in a table with a SQL statement such as

```
insert into account (id, v)
values (42, "minnie@ethmail.com")
```

the index chunks (X chunks and D chunks) are updated with new primary key information and a record chunk is created in JSON form

```
{"id":42, "v":"minnie@ethmail.com"}
```

Because the index chunk changes, the table chunk changes. The root hash of the table is set for the first time in the child chain. When an owner updates a record in a table with a SQL statement lie

```
update account set v =
```

```
"minnie@mail.eth" where id = 42
```

the record has a new chunkID because of the new JSON content

```
{"id":42, "v": "minnie@mail.eth"}
```

and so one or more index chunks are updated with a new chunkID.

- When an owner drops a database, the owner chunk is updated globally. Additionally, any tables associated with the database at the time of deletion should have their root hashes updated.
- When an owner deletes a table, the root hash of the table is updated, the schema chunk is updated, and the database chunk is updated with the new schema chunk info and removing the table name. The owner chunk is then updated with the new database chunk info..

When the leader node of a Layer 3 SQL blockchain mints a Layer 3 block, it must include in its Layer 3 block:

- the SQL transactions – where for each table referenced in the SQL, the leader must retrieve the previous root hash of the table in the SMT and execute the SVM operations for that table against that SMT’s data.
- the Chunks newly written through the execution of the SQL transactions, where chunks are only created, and never “updated”.
- a new Layer3KeyRoot transactions and calls `submitBlock(b_k^3, k)`: for all tables updated from the SQL transactions, each table has a new root hash. Using the Layer3KeyRoot, any layer 3 node can respond to a SQL SELECT query by retrieving the the previous hash of any table from the SMT. Using the Layer3KeyRoot, any layer 3 node can respond to a SQL SELECT query by retrieving the the previous hash of any table from the SMT

With a newly minted Layer 3 block k , the Layer 3 SQL blockchain can submit a layer 2 block transaction T_k^3 for the Layer 3 Block b_k^3

```
submitBlock( $b_k^3, k$ )
```

which proceeds just as in the NoSQL blockchain, with the analogously structured Deep Merkle Proof. Where in the NoSQL chain, each NoSQL document / row updated resulted in an updated leaf in the SMT for the newly updated document, now with the SQL chain, each SQL statement supports a new table root hash change in an update leaf in the SMT.

6. Paying for Storage and Bandwidth

The Layer 3 blockchain users who store NoSQL/SQL/File data with `storeChunk` operations give the Layer 2 operator permission to charge for bandwidth and storage in two different ways:

1. *Bandwidth* is paid for through (1) users signing `retrieveChunk(k, τ, ω)` calls to retrieve data and obtaining recent balances, where each call uses up a tiny amount of bandwidth backed for with a token τ originated by the `createBlockchain` call; (2) users signing a new `updateBalance(τ, ω)` request originated by operator and agreeing to making a payment for incurred bandwidth usage and the latest token owner balance. An `updateBalance` response by users is mapped into layer2 transaction, where incurred bandwidth cost is deducted from token’s owner balance (τ) and added to operator’s allowance $\gamma(\tau)$.
2. *Storage* is paid for through Block Transactions `submitBlock(b_k^3, k)` signed by the Layer 3 blockchain operator - because chunks are identified directly inside Layer 3 blocks, a tally of the number of bytes used in each new layer 3 block is added to the SMT. The Layer 1 contract then exposes `storageCharge` interface to the Layer 2 operator where a recently signed Layer 2 block transaction (containing a tally of the number of bytes, signed by the Layer 3 blockchain operator) is used to deduct the layer 3 operator’s balance since the last time it was called. This is detailed below.

In this way Layer 3 Blockchains pay for the services of the Layer 2 blockchain. The lifecycle of a short lived Layer 3 blockchain is shown in Figure 5, which is expounded in the next section.

6.1. Layer 2 Plasma Tokens for Bandwidth Payments

In this section it is explained how Layer 2 Tokens can form a *unidirectional payment channel*, where each signed

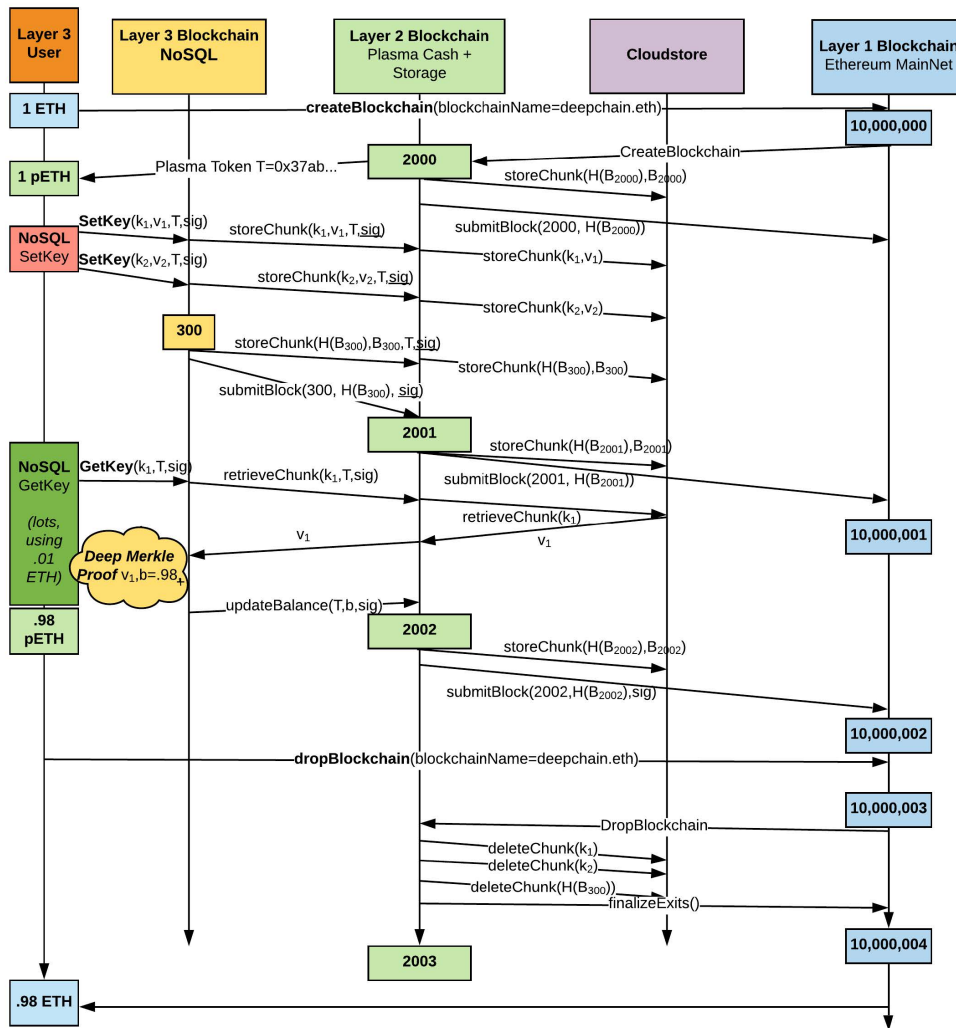


Figure 5: Samples of a Layer 3 NoSQL Blockchain User in the Deep Blockchain architecture presented here. A Layer 3 user creates a Layer 3 Blockchain by sending 1 ETH to createBlockchain and receives a Plasma token (1 pETH) which is included as a deposit in Layer 2 block 2000. The Layer 3 user does 2 SetKey operations included in Layer 3 block 300, which result in the Layer 3 blockchain signing 2 storeChunk for each key-value pair. The Layer 3 blockchain mints Layer 3 Block 300 and stores it again with 3rd signed storeChunk call, finally calling Layer 2's submitBlock with the hash of Block 300. The Layer 2 operator includes this block transaction in Layer 2 Block 2001, stores it in Cloudstore, and submits it the Layer 1 blockchain with a Smart Contract call to submitBlock. Later, multiple (lots and lots) of signed GetKey calls are done by the Layer 3 User, each executed using retrieveChunk signed calls. At some point, the user hits $\Sigma_{max} = .02pETH$ and uses updateBalance to agree that its balance is 0.98 pETH, which is included in Layer 2's token transaction block and submitted to Layer 1. Finally, the Layer 3 user can drop its blockchain, which will trigger deletion of chunks and initiate a finalizeExit process that returns the .98 ETH balance to the user.

retrieveChunk call is not a transaction to be included in a Layer 2 block (and has no nonce to increment) but simply indicative of "permission to return some data and decrement my token balance"; where the Layer 2 operator can check the signature against its record of the current

owner as a condition of looking up the chunk.

The Layer 2 operator must have tally aggregation capability that can aggregate numerous signed calls together and compute that token τ has some new balance (τ). In

our implementation simple minute-wise Hadoop job is used to tally periodic flushes of `retrieveChunk` operations grouped by different τ , keeping as a short-term output (TTL=3600s) log that each minutewise change of τ was caused by specific signed operations; this balance update log is exposed to the user. When this internal tally reaches a critical threshold Σ_{max} , responses to `retrieveChunk` halts and can only resume with a Layer 2 `updateBalance` transaction submitted and included in the Layer 2 block directly in the SMT root `Layer2TokenRoot`. The threshold Σ_{max} in the contract. To minimize disruptions from halting in this way, it is the responsibility of the Layer 3 blockchain to periodically submit `updateBalance`, signing recent token balances provided in the `retrieveChunk`.

Moreover, as the token has considerable usage accumulated, and as users regularly submit sufficient `updateBalance` transactions to Layer 2, the value of the balance may accumulate to a great enough level that the Layer 2 operator may wish to withdraw the balance accumulated directly in the Layer 1 contract. To support this, the SMT state is expanded to include the token balance and the operator withdrawal amounts.

If users wish to transfer the token to another user of the layer 2 blockchain by submitting a token transfer operation, the `updateBalance` must be executed to “close” the token-based state channel.

Finally, users who wish to withdraw token τ for Layer 1 currency can do so by calling `startExit` with the last 2 transfers *and* this last `updateBalance`, which will redeem the denomination less the tally of what has been withdrawn by the operator. Others may challenge this exit, but only with a valid proof of user double spending τ .

6.2. Layer 1 Storage Insurance

Because every single write of a Layer 3 blockchain is included in sequentially ordered layer 3 blocks (each of which identify a set of Chunk IDs) the layer 3 blockchain forms an itemized list of signed insurance requests that form a Layer 1 unidirectional state channel initiated by the deposit into `createBlockchain`. Assuming no challenges exist, if the Layer 2 operator that receives a Layer 3 block identifying a set of chunks can provide a recent

proof of storage then it may deduct from this deposit. On the other hand, if the layer 3 blockchain has reason to believe that some chunk is lost, it can submit its claim to Layer 1 smart contract and demand Merkle proofs in response. The CRASH patterns of [9] specify this challenge-response system in detail, which is extended to our deep blockchain in the following way:

- *Insurance Request.* Each Layer 3 block B_j^3 (submitted to the Layer 2 operator in the block transaction `submitBlock(b_j^3, j)` calls) includes (1) a seed hash γ , where the seed ν ($\gamma = H(\nu)$) is held solely by the layer 3 operator and revealed when the layer 3 operator wishes to challenge the Layer 2 operator with `storageChallenge` (see below); (2) the hash of an SMT Merkle root $H(\Xi)$ for all the chunks specified in the layer 3 block using ν ; (3) the *total* collection size in bytes σ_{total} in *all* Layer 3 blocks; (4) a Ω parameter, the amount of layer 1 currency required to hold 1 GB per month (e.g. if market conditions for keeping data in 8 places in Cloudstore is \$.25 GB/mo and Layer 1 currency is \$500/*ETH*, then Ω would be 5×10^4 wei). The Layer 2 operator uses κ to fetch the list of chunks the Layer 3 operator wishes to insure, verifies that all chunks in the list are in fact available, and checks that σ_{total} matches the Layer 2 operators own tally closely. If the chunks are missing, or the tally is not reasonable, the Layer 2 operator may reject the block. Otherwise, the Layer 2 blockchain will include the Layer 3 block hash in the `BlockRoot` of the next Layer 2 block. In this way, the block transaction is taken as a signed request to insure the entire Layer 3 blockchain’s storage.
- *Storage Charges.* Under ordinary conditions, the layer 2 operator can submit the most recent proof of any signed block transaction to the Layer 1 smart contract function:

```
storageCharge(blockchainName string,
txbytes bytes, storagecost uint64,
sig bytes)

```

Since `txbytes` contains σ_{total} and Ω , the `storageCharge` function can deduct from balance originally deposited via `createBlockchain` since the last time `storageCharge` was called.

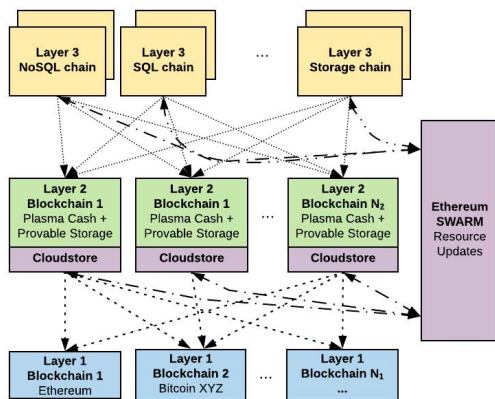


Figure 6: Three-layer deep blockchain model.

- *Storage Challenge-Response: CRASH proofs.* If at any time, the Layer 3 blockchain wishes to challenge Layer 2's inept storage (due to a missing block or missing chunk included in the block), it may do so by demanding a *CRASH proof* of a specific layer 3 block, revealing ν (which must match the γ in txbytes) by calling:


```

] storageChallenge(blockchainName
string, blockNumber uint64, seed
bytes32)

```

 A valid CRASH-proof response must be provided by the layer 2 operator within some time period (e.g. 3 to 7 days) or the challenger layer 3 user will obtain a payout proportional to σ_{total} contained in txbytes.


```

storageResponse(blockchainName
string, blockNumber uint64,
proofBits uint64, proofBytes bytes)

```

 This payout must come from a registered balance held in the Layer 1 Smart Contract. The response must be a valid proof whose root Ξ that matches $H(\Xi)$ originally supplied for the block. Finally, to guard against the situation that some layer 3 operator supplies a bogus $H(\Xi)$ in the block transaction to claim this payout, the layer 2 operator can supply a small number (e.g. 5) of Merkle branches resolving to κ . The economic incentives of this challenge-response system is refined to balance the layer 2 and layer 3 operators in this challenge-response pattern to be reasonable relative to Layer 1 Ethereum gas costs.

With the above mechanism in place, the layer 2 operator can charge the layer 3 operator when transaction fees are negligible. In regular conditions, the Layer 3 blockchain

can see its storage fees through `storageCharge`; when the balance approaches zero, the Layer 3 blockchain must deposit additional Layer 1 currency to its blockchain balance at Layer 1. Finally, a call to `dropBlockchain` must permit the layer 2 operator the opportunity to claim a final `storageCharge` and close out the bandwidth balance of τ before finalizing exits (see Figure 5). Since there are two sources of demand (storage charge and bandwidth charges), the layer 2 blockchain must check that the sum of both sources equal the available balance for the layer 3 blockchain.

7. Discussion

There have been many approaches scaling blockchain architecture to support higher throughput and lower latency:

- Changing the security model of Layer 1 blockchains (c.f. NEO, EOS's approach)
- Incremental improvements to Layer 1 or Layer 0 that don't change security model (c.f. larger blocks)
- Having many separate chains, using sharding
- State Channels
- Layer 2 Plasma solutions

This paper focussed on the last approach, and described how using the core ideas behind Layer 2 Plasma Cash can be extended to a *deep blockchain* system, forming the basis for provable data storage for widely used NoSQL + SQL developer interfaces. The concept of Deep Merkle proof for a 3 layer deep blockchain system is illustrated here and shown its conceptual viability, borrowing state channel concepts for Layer 3 NoSQL and SQL blockchains to pay for storage and bandwidth.

Deep *learning* architectures have advanced numerous high-scale applications in every industry in a way that is not about one specific deep learning algorithm – and instead about an approach that could not be achieved through dogmatic faith in single-layer “neural” networks. In an analogous way, deep blockchain architectures could have the potential to enable a wide range of high-scale applications in a way that might not be achieved through dogmatic faith in Layer 1 scaling innovations alone.

Blockchain practitioner instincts are to be wary of centralized consensus protocols and centralized storage. However, our use of non-local storage can be rationalized, not by demanding that every component be dogmatically decentralized, but by considering how attack vectors are reduced through judicious use of some not-so-decentralized components. The attacks on storage are limited in nature due to:

- verifiability of chunks, where all k, v pairs retrieved from non-local storage are verifiable either due to (a) k being verified to the hash of the value v returned (b) k being directly included and signed by a trusted party. In this sense the attack vector is limited to the private key
- the use of Ethereum SWARM (currently in POC3) as a *copyright-resistant* cloud storage provider. In the event that the Layer 2 blockchain provider loses access to its Cloud Storage backend, higher layer backends can simply request chunks using the Kademia-based DHT of Ethereum SWARM. Generally this copyright-resistance comes at the cost of higher latency responses.
- cryptoeconomic incentives, wherein if a data storer can prove (with a Merkle branch) that a piece of data can no longer be accessed but has been included on chain through a valid Merkle branch

It is believed the combination of decentralized storage and cloud computing storage increases the cost of attack and that the Blockchain 1.0 Objective of *Maximize decentralization* must be altered in favor of the more nuanced Blockchain 2.0 Objective *Maximize cost of attack*, which ultimately will lead to more secure and reliable blockchain systems. One gets the best of both worlds: from centralized storage one gets low-latency, high-throughput infrastructure, and from decentralized storage one gets resilience and copyright resistance.

Concerning the use of a single centralized Layer 2 operator, it is highlighted that in all cases where Layer 1 currency is deposited (in `createBlockchain`), because use of the “Plasma Cash” design pattern, the owner of the tokens may withdraw its balance on the Layer 1 blockchain. This is a surprising result: that checks and balances on token ownership are possible through the use of the Layer 1 blockchain despite the Plasma operator being in 100% control; if users discover that the Layer 2 blockchain operators are malicious, they can be certain they can get the value of their tokens back, and if the data is kept in resilient Ethereum SWARM (or if they have kept their data locally), they can move to another Layer 2 operator using the same protocol.

This shows a deep blockchain that has a higher cost of attack than the deep blockchain illustrated in Figure 2, utilizing N_2 Layer 2 blockchains (each with their own Cloudstore) and N_1 Layer 1 blockchains, each receiving the same Layer 3 `submitBlock` and Layer 2 `submitBlock` transactions respectively:

Because the retrieval of layer $i + 1$ data from layer i can be verified by layer $i + 1$ (checking block data: does the block hash match the block content? is it signed? does it have a parent hash? etc.; checking chunks: does the hash of the chunk data equal the chunk key), each lower layer

node devolves into a dumb storage layer with some failure or attack probability ($p_1^2 \dots p_{N_2}^2$ for layer 2 blockchains, $p_1^1 \dots p_{N_1}^1$ for layer 1 blockchains) – depending on this probabilistic model, the cost of attack may be divided. However, it seems most likely that motivated parties would attack the centralized control behind each layer (c.f. via EIP999, mining pools arewedecentralizedyet.com, governments asking the Cloudstore providers to block Layer 2 operator’s accounts) – in this sense the probabilistic independence in concentrated efforts to attack layer 1 and 2 would be highly suspect. For this reason, our true faith relies in Ethereum SWARM’s *resource updates* ([9]), where chunks may be keyed not by the hash of their content but with a *resource key*, which can be used for the block data without an index mechanism; all resource updates are signed so the reader can authenticate. Ethereum SWARM, because of its use of Kademia-like protocol, is not naturally as fast as other components in Cloudstore, but kicks in when all Layer 2 blockchains Cloudstore fail or when Layer 1 itself is attacked (via 51% attacks, or unknown POS failures). If other decentralized storage services provided similar provable storage as Ethereum SWARM’s resource update, so long as Layer 3 blockchain does not go Byzantine, only one answer can surface, making for unstoppable layer 3 blockchains.

The Layer 3 NoSQL and SQL blockchains developed in this paper operated under an assumption that the NoSQL + SQL transactions should be private data secured by an encryption key known only to the operators of the Layer 3 blockchain. This protects the Layer 3 blockchain from operators of Layer 2 blockchain and any Cloudstore. However, the same problem as with standard databases (MySQL, MongoDB, DynamoDB, etc.) exists with our current implementation of NoSQL/SQL Layer 3 blockchains: once someone gets access to a Layer 3 blockchain node holding the database encryption key or private key, the entire database is compromised. Therefore, provenance and immutability of the NoSQL/SQL database state changes, as manifested in Deep Merkle Proofs, differentiate a Layer 3 blockchain from standard databases. The small latency incurred with permissioned protocols (RAFT, POA) and negligible cost should be welcomed when provenance and immutability are of paramount concern.

Many other Layer 3 blockchains can be constructed using the Layer 2 storage and bandwidth infrastructure: a chain that represents the evolving state of ERC721 tokens, a chain that represents a cryptocurrency exchange where your money can never be stolen, and so forth. The state of the Layer 3 blockchain is not stored locally but instead kept in Cloudstore with storage and bandwidth costs properly accounted for using the Layer 2 tokens, themselves based on Layer 1. Layer 3 and Layer 2 nodes are therefore “light nodes” in that they can quickly catch up to the latest state by asking the layer 2 and layer 1

blockchains for the most recent finalized block. This is not possible to do for the Layer 1 blockchain, however. However, it is possible, and interesting to adapt a Layer 1 blockchain of Ethereum and make it a Layer 3 blockchain. Computation (Ethereum gas costs) can consume *Layer 2 token* balances in state channels along with bandwidth, contract storage can use SMTs mapped to Cloudstore (instead of Patricia Merkle Tries kept in local store) submitted in blocks to the Layer 2 blockchain, and the consensus machinery can be put in a modern sharded Proof-of-Stake framework to achieve high-throughput low-latency ambitions of Ethereum 2.0, with all layer 3 nodes. The expectation would be that a Layer 3 Ethereum blockchain would have massively lower costs due to rational models of storage and bandwidth. Other deep blockchain systems can be developed with different computational primitives than the EVM, such as Amazon's Lambda or Apache Hadoop.

It is believed that there can be many deep blockchain systems developed with higher layers resting on many Layer 1 blockchains, even to the point where multiple Layer 1 systems are dropped and many more added to provide more or less Layer 2 security. The same can be said for any layer to benefit higher layers. If the blockchain at layer i changes its consensus algorithm from Quorum RAFT to pBFT or Casper Proof-of-Stake, the layer $i + 1$ benefits; higher layer blockchains are supervenient on Layer 1, so innovations on Layer 1 are inherited by all deep blockchain systems. It is hoped that many deep blockchain systems can explore high throughput low latency scale through some of the design patterns explored here.

References

- [1] M. Gracy, B. R. Jeyavadhanam, A systematic review of blockchain-based system: Transaction throughput latency and challenges, in: 2021 International Conference on Computational Intelligence and Computing Applications (ICCIICA), IEEE, 2021, pp. 1–6.
- [2] Y. Meshcheryakov, A. Melman, O. Evsutin, V. Morozov, Y. Koucheryavy, On performance of pbft blockchain consensus algorithm for iot-applications with constrained devices, *IEEE Access* 9 (2021) 80559–80570.
- [3] J. Yoo, Y. Jung, D. Shin, M. Bae, E. Jee, Formal modeling and verification of a federated byzantine agreement algorithm for blockchain platforms, in: 2019 IEEE International Workshop on Blockchain Oriented Software Engineering (IWBOSE), IEEE, 2019, pp. 11–21.
- [4] G. Wood, et al., Ethereum: A secure decentralised generalised transaction ledger, <https://karl.tech/plasma-cash-simple-spec/>, 2014.
- [5] U. Rahardja, A. N. Hidayanto, N. Lutfiani, D. A. Febiani, Q. Aini, Immutability of distributed hash model on blockchain node storage, *Sci. J. Informat-ics* 8 (2021) 137–143.
- [6] K. Floersch, Plasma cash simple spec, <https://karl.tech/plasma-cash-simple-spec>, 2018.
- [7] E. Gaetani, L. Aniello, R. Baldoni, F. Lombardi, A. Margheri, V. Sassone, Blockchain-based database to ensure data integrity in cloud computing environments (2017).
- [8] J. Poon, V. Buterin, Plasma: Scalable autonomous smart contracts, <http://plasma.io/plasma.pdf>, 2017.
- [9] V. Trón, A. Fischer, D. A. Nagy, Swarm: a decentralised peer-to-peer network for messaging and storage (2018). Forthcoming.
- [10] S. K. Panda, A. A. Elngar, V. E. Balas, M. Kayed, Bitcoin and blockchain: history and current applications, CRC Press, 2020.
- [11] B. Laurie, E. Kasper, Revocation transparency, <https://www.links.org/files/RevocationTransparency.pdf>, 2017.
- [12] R. Dahlberg, T. Pulls, R. Peeters, Efficient sparse merkle trees: Caching strategies and secure (non-) membership proofs, in: Secure IT Systems: 21st Nordic Conference, NordSec 2016, Oulu, Finland, November 2-4, 2016. Proceedings 21, Springer, 2016, pp. 199–215.

Energy-Efficient Routing in UAVs Supported Perimeter Security Networks

Alexander K. Alexandrov ^{1,*}, Anastass N. Madzharov ¹

¹*Institute of Robotics, Bulgarian Academy of Sciences, Acad. G. Bonchev str., 1113 Sofia, Bulgaria*

Abstract

UAV-assisted ground and underwater perimeter security sensor networks represent a sophisticated integration of aerial, ground, and underwater technologies for surveillance and security purposes. This system combines Unmanned Aerial Vehicles (UAVs) with underwater sensors to monitor and protect strategic areas like harbors, offshore installations, and coastal facilities. Unmanned Aerial Vehicles (UAVs) have become pivotal in modern surveillance and security operations. Their versatility, mobility, and technological adaptability make them ideal for perimeter security systems. This study examines the integration of group of UAVs into perimeter security, evaluating their effectiveness, operational frameworks, technological advancements, and potential future developments. We analyze and implement a PSO (Particle Swarm Optimization) algorithm, related to group of UAVs trajectory optimization, review case studies, and identify key considerations for effective development.

Keywords

UAV, PSO, sensor network, perimeter security

1. Introduction

UAV-assisted underwater perimeter security sensor networks represent a cutting-edge blend of aerial and maritime technologies, designed to enhance the security of critical aquatic areas. This integration of Unmanned Aerial Vehicles (UAVs) and underwater sensors provides a robust solution for monitoring and safeguarding sensitive zones like naval bases, coastal areas, ports, and offshore installations.

The key components in the UAV-assisted underwater perimeter security sensor networks are the sensors, UAVs, and the control center.

Underwater sensors typically include acoustic sensors (such as sonars), geophones, hydrophones for detecting sound under water, and magnetic anomaly detectors for identifying metallic objects. These sensors continuously scan underwater environments to detect and track potential threats, like submarines, divers, or unmanned underwater vehicles (UUVs).

UAVs provide real-time aerial surveillance, significantly extending the range of observation beyond the immediate perimeter. They act as a vital link between the underwater sensors and the control center, especially important in deep-water areas where direct communication is difficult.

Control center provides data processing and decision making. Here the data from both UAVs and underwater

sensors is analyzed, processed, and fused to form a comprehensive operational picture. Control center assesses potential threats based on the gathered information and coordinates appropriate responses.

One of the challenges in the UAVs assisted underwater perimeter security sensor networks is the energy management. Both the UAVs and underwater sensors must efficiently manage their power to ensure prolonged operational capabilities. The present study focuses on energy management, especially in energy-efficient and reliable routing of groups of UAVs. The UAVs energy-efficient routing is a multifaceted challenge that involves optimizing the flight paths and operational strategies of UAVs.

The objective is to maintain vigilant monitoring and rapid response capabilities while minimizing energy consumption, which is critical for the longevity and effectiveness of the UAVs in defense operations. The aim is to create routes and operational patterns that minimize energy usage while ensuring comprehensive security coverage.

Altitude and speed optimization in UAV-supported underwater perimeter security sensor networks is a critical aspect of ensuring energy-efficient routing and effective operation. The right balance of altitude and speed directly impacts the UAVs' energy consumption, coverage area, sensor effectiveness, and response times.

1.1. Altitude optimization

Higher altitudes can offer less air resistance, but the benefit must be balanced against increased energy requirements for climbing and maintaining altitude. Higher altitudes may increase the coverage area but could reduce the detail or accuracy of sensor data. The right altitude affects UAV performance in different weather

BISEC'23: 14th International Conference on Business Information Security, November 24, 2023, Niš, Serbia

*Corresponding author.

✉ akalexandrov@ir.bas.bg (A. K. A.); a.madzharov@ir.bas.bg (A. N. M.)

conditions. For example, flying above or below certain weather layers (like fog or clouds) can be crucial.

The optimized altitude can ensure communication with both the underwater sensor network and the control station [1].

1.2. Speed optimization

Generally, faster speeds increase energy consumption. The optimization algorithm should identify the most energy-efficient cruising speed for each UAV model. Faster speeds allow for quicker coverage of an area but might reduce the effectiveness of sensors due to motion blur or reduced processing time.

Speed must be optimized to balance routine surveillance with the need for rapid response in case of detected threats [2, 3]. Tailwind can be exploited to reduce energy consumption, whereas flying into headwinds will require more energy, affecting optimal speed decisions.

2. Related works

There are some existing solutions related to the UAVs assisted underwater perimeter security sensor networks as:

- **DJI Enterprise Drones** - the solution is used for inspection and surveillance of commercial and military complexes. The drone is equipped with thermal imaging sensors, high-resolution cameras, and programmable flight paths and is programmed for routine patrols or dispatched upon alerts from ground and underwater sensors.
- **AeroVironment Raven RQ-11B** - the solution is used for battlefield reconnaissance and surveillance. The UAV is equipped with GA (Genetic Algorithms), based trajectory optimization system and interfaces with ground and underwater control systems and sensor networks.
- **Elbit Systems Skylark I-LEX** – this is electrically propelled UAV equipped with MPC (Model Predictive Control) trajectory optimization system, designed to collect data and interface with ground and underwater sensors for a comprehensive security net and is utilized by military and homeland security for national borders and sensitive areas.
- **Anduril Industries' Lattice** – this is a complete system that integrates drones, ground and underwater sensors, and AI-powered analysis to detect, classify, and respond to threats.
- **Asylon DroneCore** - automated drone deployment system that works with perimeter sensors to conduct autonomous patrols and respond to

intrusions. The system is integrated with existing security infrastructure, providing a bird's-eye view when a ground sensor is triggered.

- **General Atomics Predator B** - used for national border surveillance, can be used in conjunction with ground sensor arrays for detecting and tracking movements and is equipped with high-resolution cameras and advanced signal intelligence equipment that can integrate with sensor network data.

All the mentioned UAVs have a custom design navigation systems with included energy-efficient software algorithms for routing and altitude/speed optimization, using various algorithms such as RL (Reinforcement Learning, Dynamic Programming, Dijkstra, GA (Genetic algorithms) in different combinations.

3. Proposed solution

The current research is focused on the development and implementation of altitude (elevation) and speed optimization algorithm in custom designed UAVs.

The proposed algorithm is based on PSO (Particle Swarm Optimization) [4, 5, 6]. This is a computational method that optimizes a problem by iteratively trying to improve a candidate solution with regard to a given measure of quality.

It solves a problem by having a population of candidate solutions, here dubbed particles, and moving these particles around in the search-space according to simple mathematical formulae over the particle's position and velocity.

Each particle's movement is influenced by its local best known position but is also guided toward the best known positions in the search-space, which are updated as better positions are found by other particles. When applying PSO for altitude and speed optimization in UAVs supporting underwater perimeter security sensor networks, the goal is to determine the optimal flight paths, altitudes, and speeds for the UAVs to maximize coverage, efficiency, and responsiveness while minimizing energy consumption.

3.1. Challenges in the speed and elevation optimization

The following challenges related to the speed/elevation optimization problem were defined during the research:

- **High Dimensionality:** The speed/elevation optimization problem can be high-dimensional, especially when considering 3D space and time, making it computationally intensive [7].

- **Dynamic Constraints:** UAVs must respond to dynamic changes in the environment, which requires the PSO to be adaptable and responsive in real-time.
- **Local Minima:** The PSO algorithm may get trapped in local minima. This issue can be mitigated by tuning the parameters (ω, c_1, c_2) or by hybridizing PSO with other optimization techniques.
- **Safety and Collision Avoidance:** Ensuring safety is paramount. The algorithm must incorporate collision avoidance with other UAVs, terrain, and obstacles [8].

3.2. Implementation

Implementing a Particle Swarm Optimization (PSO) algorithm for altitude and speed optimization in UAV-supported underwater perimeter security sensor networks involves several mathematical concepts. Here's an mathematical overview of the proposed algorithm [9, 10, 11]:

Objective Function

Let's denote the objective function as $f(x)$, where x represents a vector of the decision variables (altitude and speed in this case) for UAVs. The function might aim to minimize energy consumption while maximizing area coverage, response time, or signal quality.

This could be a weighted sum or a more complex function based on the mission requirements.

Constraints

Include constraints like battery life (B), maximum and minimum altitude ($A_{\{\max\}}, A_{\{\min\}}$), and speed limits ($S_{\{\max\}}, S_{\{\min\}}$).

PSO Algorithm Structure

Particle Representation - each particle i in the swarm represents a potential solution, with its position p_i indicating a particular set of altitudes and speeds for a UAV.

Initialization: randomly initialize the position p_i and velocity v_i of each particle within the feasible space defined by the constraints.

Velocity and Position Update Rules

Velocity update:

$$v_i^{t+1} = \omega v_i^{(t)} + c_1 r_1 (p_{best,i} - p_i^{(t)}) + c_2 r_2 (p_{global\ best} - p_i^{(t)}), \quad (1)$$

where ω is the inertia weight, c_1 and c_2 are cognitive and social coefficients, respectively, r_1, r_2 are random numbers between 0 and 1.

Position update:

$$p_i^{t+1} = p_i^{(t)} + v_i^{t+1}. \quad (2)$$

Ensure that the updated position adheres to the constraints.

Evaluation:

Evaluate the fitness of each particle using the objective function $f(x)$.

Update the personal best $p_{best,i}$ if the current position of the particle yields a better value of the objective function. Update the global best $p_{global\ best}$ if any particle achieves a better value than the current global best.

Termination:

Continue iterating until a maximum number of iterations is reached or convergence criteria are met (e.g., minimal improvement in the global best).

Example Objective Function

Consider a simplified example where the objective is to minimize energy consumption E while ensuring good area coverage C . The objective function might look like this:

$$f(x) = \alpha E(x) - \beta C(x). \quad (3)$$

Here, α and β are weights reflecting the importance of energy consumption versus coverage.

The functions $E(x)$ and $C(x)$ compute the energy consumption and coverage based on the altitude and speed parameters in x .

The mathematical overview provided here is a simplified version of what could be a complex real-world implementation. In practice, the functions and parameters would need to be tailored to specific UAV capabilities, sensor characteristics, environmental factors, and mission goals.

Additionally, various enhancements to the basic PSO, such as constriction factors or varying inertia weight, might be employed to improve convergence and solution quality.

To implement the PSO algorithm for altitude and speed optimization in UAV-supported underwater perimeter security sensor networks, we will develop a structured pseudocode.

This pseudocode will help visualize the flow of the algorithm and serve as a guide for actual programming.

Remember that PSO is inherently iterative and works with a population of solutions, adjusting them over time based on a defined objective function.

The related PSO algorithm written in pseudocode is shown below:

PSO algorithm for UAVs elevation/speed optimization

Inputs:

- num_particles: Number of particles in the swarm
- max_iterations: Maximum number of iterations
- objective_function: Function to optimize (minimize or maximize)
- A_max, A_min: Maximum and minimum allowable altitudes
- S_max, S_min: Maximum and minimum allowable speeds
- omega: Inertia weight
- c1, c2: Cognitive and social coefficients

Initialize:

- Create num_particles particles with random positions and velocities
- for each particle i:
 - position[i] = Random within [A_min, A_max] and [S_min, S_max]
 - velocity[i] = Random initial velocity
 - pbest[i] = position[i]
- gbest = position of the best particle based on objective_function

Main Loop:

- for iter = 1 to max_iterations:
 - for each particle i:
 - Update velocity:
 - r1, r2 = Random numbers between 0 and 1
 - velocity[i] = omega * velocity[i] + c1 * r1 * (pbest[i] - position[i]) + c2 * r2 * (gbest - position[i])
 - Update position:
 - position[i] = position[i] + velocity[i]
 - Ensure position[i] adheres to [A_min, A_max] and [S_min, S_max]
 - Evaluate:
 - If objective_function(position[i]) is better than objective_function(pbest[i]):
 - pbest[i] = position[i]
 - If objective_function(position[i]) is better than objective_function(gbest):
 - gbest = position[i]
 - Return gbest as the optimal solution
- End Algorithm

Related to the pseudocode above please note:

Initialization: The initial positions and velocities are randomly assigned within the permissible ranges for altitude and speed. Each particle's initial position is considered its personal best (*pbest*).

Updating Velocities and Positions: The velocities are updated considering both the particle's own best position and the global best (*gbest*). The updated velocity influences the new position. It's important to ensure that the updated positions are within the allowed ranges.

Evaluating and Updating Best Positions: After updating positions, evaluate them using the objective_function. If a particle's new position is better than its *pbest*, update *pbest*. If it's better than the current *gbest*, update *gbest*.

Termination: The algorithm iterates through this process, gradually moving the swarm towards the best solution. The process repeats either until the maximum number of iterations is reached or some other stopping criterion (like a convergence threshold) is met.

Returning the Optimal Solution: Finally, the *gbest* after the last iteration is returned as the optimal set of altitude and speed parameters.

Customization for the specific use-case: The objective function should be designed specifically for the UAV's operational requirements, taking into account factors like energy consumption, area coverage, sensor effectiveness, and other mission-specific metrics.

Parameters such as ω , c_1 , and c_2 may need tuning for optimal performance in specific scenarios.

Additional constraints or enhancements can be integrated into the algorithm based on specific requirements and operational environments.

4. Key Takeaways

Enhanced Efficiency: The PSO algorithm effectively optimizes UAV flight parameters (altitude and speed), leading to improved energy efficiency. This results in longer mission durations and reduced operational costs.

Adaptive Flight Paths: The algorithm's ability to dynamically adapt flight paths in response to changing environmental conditions and mission requirements is a significant advantage, ensuring optimal coverage and data collection.

Collaborative Functionality: PSO inherently supports multi-UAV coordination, allowing for effective swarm operations. This results in comprehensive area surveillance and redundant systems for critical defense missions.

Real-Time Decision Making: The implementation enables UAVs to make real-time adjustments, crucial for responding to emergent underwater threats or anomalies detected by the sensor network.

Operational Flexibility: The algorithm's flexibility allows it to be tailored to various mission scenarios, UAV types, and sensor network configurations, making it broadly applicable in underwater perimeter defense.

4.1. Challenges and Considerations

Complex Environmental Dynamics: The underwater and aerial environments present unique challenges, including variable weather conditions and underwater currents, which can affect the algorithm's performance.

Communication Limitations: Ensuring reliable communication between UAVs and underwater sensors remains a challenge, impacting the coordination and effectiveness of the network.

Computational Demands: PSO, especially in real-time applications, can be computationally intensive, necessitating robust onboard processing capabilities. **Security and Robustness:** The system must be secured against potential cyber threats and robust enough to handle operational uncertainties and potential system failures.

5. Conclusion

The implementation of a Particle Swarm Optimization (PSO) algorithm for altitude and speed optimization in UAVs supporting underwater perimeter security sensor networks is a sophisticated approach that leverages the strengths of swarm intelligence for operational efficiency. The conclusion drawn from this implementation can highlight its significance, potential benefits, and areas for future enhancement. Future steps:

- incorporating advanced variants of PSO or hybrid algorithms could further optimize performance, especially in highly dynamic or unpredictable environments.
- leveraging AI for predictive analytics and machine learning for continuous improvement of flight path algorithms based on historical data can enhance operational efficiency.
- incorporating sustainable technologies, such as solar-powered UAVs, can extend mission durations and reduce environmental impact.

The implementation of a PSO algorithm for optimizing a group of UAVs' altitude and speed in underwater perimeter security sensor networks demonstrates significant potential in improving maritime security operations [12]. While challenges remain, the continuous advancements in technology and algorithmic strategies hold promise for developing more sophisticated, efficient, and robust defense networks in the future. This approach exemplifies the innovative integration of aerial

and maritime technologies, paving the way for enhanced security solutions in coastal and offshore environments.

In conclusion, PSO is a robust and versatile algorithm widely used for solving complex optimization problems. Its ongoing developments and applications across diverse fields highlight its relevance in the current technological landscape.

References

- [1] H. He, S. Zhang, Y. Zeng, R. Zhang, Joint altitude and beamwidth optimization for uav-enabled multi-user communications, *IEEE Communications Letters* 22 (2017) 344–347.
- [2] N. H. Chu, D. T. Hoang, D. N. Nguyen, N. Van Huynh, E. Dutkiewicz, Joint speed control and energy replenishment optimization for uav-assisted iot data collection with deep reinforcement transfer learning, *IEEE Internet of Things Journal* 10 (2022) 5778–5793.
- [3] W. Ye, J. Luo, F. Shan, W. Wu, M. Yang, Offspeeding: Optimal energy-efficient flight speed scheduling for uav-assisted edge computing, *Computer Networks* 183 (2020) 107577.
- [4] D. Wang, D. Tan, L. Liu, Particle swarm optimization algorithm: an overview, *Soft computing* 22 (2018) 387–408.
- [5] O. F. Aje, A. A. Josephat, The particle swarm optimization (pso) algorithm application—a review, *Global Journal of Engineering and Technology Advances* 3 (2020) 001–006.
- [6] M. Ehteram, A. Seifi, F. B. Banadkooki, Structure of particle swarm optimization (pso), in: *Hellenic Conference on Artificial Intelligence*, Springer, 2022, pp. 23–32.
- [7] S. Schröder, D. Schleich, S. Behnke, Two-step planning of dynamic uav trajectories using iterative-spaces, in: *International Conference on Intelligent Autonomous Systems*, Springer, 2022, pp. 257–271.
- [8] J. Hong, D. Chen, W. Li, Z. Fan, Trajectory planner for uavs based on potential field obtained by a kinodynamic gene regulation network, *Sensors* 23 (2023) 7982.
- [9] R. Poli, J. Kennedy, T. Blackwell, Particle swarm intelligence. an overview, *Swarm Intelligence* 1 (2007) 33–57.
- [10] Y. Yang, X. Xiong, Y. Yan, Uav formation trajectory planning algorithms: A review, *Drones* 7 (2023) 62.
- [11] Q. Geng, Z. Zhao, A kind of route planning method for uav based on improved pso algorithm, in: *2013 25th Chinese control and decision conference (CCDC)*, IEEE, 2013, pp. 2328–2331.
- [12] H. Zhang, B. Xin, L.-h. Dou, J. Chen, K. Hirota, A review of cooperative path planning of an unmanned

aerial vehicle group, *Frontiers of Information Technology & Electronic Engineering* 21 (2020) 1671–1694.

Reducing the WSN's Communication Overhead by the SD-SPDZ Encryption Protocol

Alexander K. Alexandrov ^{1,*}

¹*Institute of Robotics, Bulgarian Academy of Sciences, Acad. G. Bonchev str., 1113 Sofia, Bulgaria*

Abstract

Wireless Sensor Networks (WSN) have emerged as a pivotal technology in many application areas such as environmental monitoring, IoT, military applications, and healthcare. These networks consist of spatially distributed, autonomous sensors that cooperatively monitor physical or environmental conditions, such as temperature, sound, or pollution levels. The unique characteristics of WSNs, including their resource constraints (e.g., energy, memory, and computational capacity), make them vulnerable to various security threats. Information security in WSNs is crucial to ensure the confidentiality, integrity, and availability of the data they collect and transmit.

As these wireless sensors collect and share data, they ensure the security and privacy of transmitted information becomes critical. In recent years, with an increasing emphasis on security, there has been a growing interest in Multi-Party Computation (MPC). MPC allows multiple parties to compute a joint function over their inputs while keeping those inputs private. The SPDZ protocol is among the most prominent and influential secure computation protocols. While the initial SPDZ protocol and its successor, SPDZ-2, have shown promising results, there were still challenges related to performance, scalability, and overall security.

This paper presents a newly developed protocol named SD-SPDZ (Sensor Data SPDZ). The proposed protocol is based on MPC SPDZ-2 protocol and proposes changes to increase the performance in the preprocessing phase by implementing a new algorithm for the Beaver triples calculation. This protocol enhances the privacy-preserving attributes and efficiency of its predecessors. SD-SPDZ integrates advanced cryptographic techniques, offering a more robust and scalable solution for secure computations in WSNs. The primary benefits include reduced communication overhead, faster computation times, and improved resistance against various cyberattacks. The integration of SD-SPDZ in WSNs could improve performance sensitively and change the way sensor data is securely processed in sensor networks. It provides a promising pathway to ensure that as technology advances, the integrity and confidentiality of the data in these networks remain uncompromised.

In summary, as WSNs play an increasingly critical role in modern-day applications, the need for advanced high-performance security mechanisms such as the SD-SPDZ protocol becomes more evident. This combination of cutting-edge, high-performance, secure computation with wireless sensor networks promise a future where data can be both globally accessible and privately computed, bridging the gap between performance and privacy.

Keywords

WSN, Information security, sensor data encryption, SPDZ, SD-SPDZ, Fixed Block Ciphers

1. Introduction

Wireless Sensor Networks (WSN) [1] are being used in numerous applications ranging from environmental monitoring to defense and healthcare. The distributed nature of WSNs and their deployment in potentially hostile environments make data encryption crucial to ensure data confidentiality, integrity, and authenticity. Historically, traditional encryption algorithms such as Advanced Encryption Standard (DES) [2] and Data Encryption Standard (DES) [3] were evaluated for WSNs. However, due to resource constraints in WSN nodes, some additional encryption techniques gained popularity.

Constraints and Challenges

Limited Resources: WSN nodes typically have limited processing capability, memory, and energy.

Dynamic Network Topology: Nodes can join or leave, posing challenges for key management.

Physical Vulnerability: Sensor nodes may be deployed in hostile environments, susceptible to physical attacks.

Current Encryption Techniques

Lightweight Block Ciphers: They require less computational power and memory [4].

Stream Ciphers: Focus on processing data bit-by-bit, requiring minimal memory [5]. Examples are Trivium and Grain.

Public Key Cryptography: Though resource-intensive, they can be optimized for specific tasks like initial key exchange [6].

Multi-Party Computation: Multi-Party Computation (MPC) [7] is a subfield of cryptography that enables multiple parties to jointly compute a function over their inputs

BISEC'23: 14th International Conference on Business Information Security, November 24, 2023, Niš, Serbia

*Corresponding author.

✉ akalexandrov@ir.bas.bg (A. K. A.)

without revealing those inputs to each other.

The main benefits of the MPC based encryption protocols are:

Privacy: Ensures that individual inputs remain secret from other participants.

Correctness: Guarantees that the output is correct even if some participants behave maliciously.

This essential in some WSN's as:

Secure voting systems where voters want to compute the result without revealing individual votes;

Military applications;

Collaborative data analysis in medical research where institutions want to compute a joint result without sharing patient data directly.

1.1. Sensor data encryption techniques

With the rising proliferation of the Internet of Things (IoT) and the widespread deployment of sensor networks across various industries, ensuring the confidentiality, authenticity, and integrity of sensor data has become paramount. This study delves deep into the techniques and strategies employed for sensor data encryption, focusing on the unique challenges and requirements presented by these environments.

Objectives

To understand the peculiarities and constraints of sensor data. To evaluate existing encryption methodologies suitable for sensor data. To propose efficient techniques or improvements tailored for sensor data encryption.

Characteristics of Sensor Data

Sensor data can be distinguished by:

- High volume: Many sensors generate data continuously.
- Temporal relevance: Some data may be time-sensitive.
- Varying importance: Not all sensor data is equally critical.

Challenges in Sensor Data Encryption

- Resource Limitations: Sensors often have constrained processing capabilities, energy, and memory.
- Transmission Overheads: Encryption might introduce additional latency or payload.
- Diverse Deployment: Sensors can be found in hostile environments, making them susceptible to physical attacks.

2. Related works

In the area of the existing approaches, protocols, and algorithms used to reduce the encrypted communication overhead in WSNs the following is commonly used nowadays: BGW Protocol: The Beimel, Malkin, and Micali (BGW) protocol [8] is one of the foundational works in the area of secure multi-party computation. SPDZ can be viewed as a descendant of the BGW protocol, where both focus on achieving security against a malicious adversary.

TinyOT: An efficient protocol [9] for two-party computation, TinyOT inspired many techniques used in SPDZ, especially the ones in the preprocessing phase. Overdrive2K: Overdrive refers to optimizations and enhancements of the SPDZ protocol, further improving the efficiency of the offline phase [10].

MASCOT: A follow-up to SPDZ, MASCOT introduces a more efficient method [11] for the preprocessing phase by using oblivious transfer instead of somewhat homomorphic encryption, reducing computational overhead.

SPDZ2k: The SPDZ2k protocol [12] has been adjusted to operate with calculations based on powers of two.

The significant difficulty with this is that in Z2k, not every component has an inverse, an essential factor for ensuring the security of both MASCOT and SPDZ. To address this, SPDZ2k shifts to Z2k', where k' is a greater value, to offset the presence of zero divisors.

MP-SPDZ: provides a complete implementation of SPDZ2k [13] and features its distinct Z2k version, which is optimized for compile-time k.SPZDZ-2: An optimized version of the original SPDZ, it enhances the online phase for better efficiency.

BMR. Beaver and colleagues introduced a method [14] to create garbled circuits from any multi-party computation framework while maintaining security attributes. This method was later enhanced by Lindell and team by employing SPDZ as the foundational protocol. MP-SPDZ integrates BMR with the SPDZ/MASCOT protocol and other security model protocols. Even though this feature wasn't included in SPDZ-2, it was unveiled partially prior to MP-SPDZ's first edition, as it was utilized by Keller and Yanai in their oblivious RAM development.

Yao's Garbled Circuits. Bellare and co-authors showcased a version of Yao's garbled circuits optimized for DES-NI, which is the standard DES execution on contemporary processors [15]. After the final release of SPDZ-2, this version was incorporated and recently updated to encompass the half-gate method.

2.1. SPDZ and SPDZ-2 Encryption Protocols Overview

The SPDZ protocol is a foundational Multi-Party Computation (MPC) scheme known for its robust security

guarantees and practical efficiency. SPDZ facilitates secure computation among multiple parties as connected sensor modules, ensuring that individual inputs remain private.

Protocol Basics

At a high level, the SPDZ protocol encompasses two main phases: **Preprocessing Phase**: Offline phase where correlated randomness (like Beaver Triples) is generated without knowing the inputs.

Online Phase: Actual computation is performed using the preprocessed data.

Secret Sharing in SPDZ

Given a secret s , it is split into additive shares $s_1, s_2, s_3, s_4 \dots, s_n$ such that:

$$s = \sum_{i=1}^n s_i. \quad (1)$$

In the preprocessing phase, a Beaver's triples (a, b, c) are generated where $c = a \times b$. During the online phase, given shares of values x and y that need to be multiplied, the protocol proceeds as:

Compute

$$\delta_x = x - a \quad (2)$$

and

$$\delta_y = y - b. \quad (3)$$

Each sensor module locally computes

$$x \times y = x + \delta_x \times b + \delta_y \times a + \delta_x \times \delta_y \quad (4)$$

In the online phase both values x and y where

$$x = \sum_{i=1}^n x_i, \quad (5)$$

$$y = \sum_{i=1}^n y_i \quad (6)$$

are computed as:

$$x + y = \sum_{i=1}^n (x_i + y_i) \quad (7)$$

Each sensor module locally adds its shares. Using Beaver's triple, multiplication can be securely performed as outlined above.

The SPDZ protocol also integrates zero-knowledge proofs to ensure correctness without revealing individual inputs or intermediate results.

Mathematically, SPDZ employs techniques from linear secret-sharing schemes to ensure zero-knowledge properties.

Basics of the SPDZ-2 Protocol

The SPDZ-2 protocol [16] is an improvement over the original SPDZ protocol for secure multi-party computation (MPC). It builds upon the foundations of the original protocol while addressing certain performance and security issues. The SPDZ-2 protocol also employs two main phases like its predecessor:

Preprocessing Phase: Where correlated randomness is generated.

Online Phase: Where the actual computation using the preprocessed data takes place.

SPDZ-2 introduces a more efficient zero-knowledge proof system to ensure that:

- The shares of each party are consistent.
- The Beaver's triples are valid.

Instead of employing full-fledged zero-knowledge proofs, SPDZ-2 uses MACs (Message Authentication Codes) and correlated randomness to ensure honesty and correctness without much communication overhead.

Improvements over the original SPDZ

Reduced Communication Overhead: By leveraging MACs and efficient consistency checks, SPDZ-2 reduces the number of rounds of communication, which is especially beneficial in settings with many parties. To ensure consistency of shares and validity of the triples, MACs (Message Authentication Codes) are utilized.

The preprocessing phase is made more efficient, leading to faster overall computation times. At the same time, when applied to wireless sensor networks, the SPDZ-2 protocol can still exhibit considerable communication overhead. Sensor networks have bandwidth constraints, limited battery life, and operate in high-latency environments, making communication efficiency crucial.

SPDZ-2 Protocol implementation in Wireless Sensor Networks (WSN)

Wireless Sensor Networks (WSN) typically consist of spatially distributed autonomous devices that cooperatively monitor physical or environmental conditions.

Applying the SPDZ-2 protocol in WSN enables secure collaborative data processing without revealing individual sensor readings.

For a WSN with n sensor nodes, let each node i have a private value v_i . The goal is to compute a function $f(v_1, v_2, \dots, v_n)$ securely.

Secret sharing in WSN

A sensor node's private value v_i is split into additive secret shares distributed among other nodes such that:

$$v_i = \sum_{i=1}^n share_{ij} \quad (8)$$

For shared values x and y , use preprocessed triples (a, b, c) where $c = a \times b$.

Calculate and open

$$\delta_x = x - a, \quad (9)$$

and

$$\delta_y = y - b, \quad (10)$$

to all nodes. Each node locally computes

$$x \times y = c + \delta_x \times b + \delta_y \times a + \delta_x \times \delta_y. \quad (11)$$

Zero-Knowledge Proofs

To ensure consistency of shares and validity of the triples, MACs (Message Authentication Codes) [17] are utilized. Given a MAC key α , and a value v , the MAC is:

$$MAC_v = \alpha \times v. \quad (12)$$

Sensor nodes verify the validity of MACs without revealing their private values.

Communication Model in WSN

Given the energy and bandwidth constraints in WSN, the application of SPDZ-2 requires efficient communication models, possibly hierarchical or cluster-based, to minimize overhead.

In WSN, sensor nodes can be viewed as parties in the MPC. Each node can hold a piece of the secret (i.e., its measurement) and wants to perform computations without revealing its exact measurement to others.

Sensor Data Aggregation

For an aggregate function f over sensor data d_1, d_2, \dots, d_n :

$$f(d_1, d_2, \dots, d_n) = \sum_{i=1}^n f(d_i). \quad (13)$$

Using SPDZ-2, the function f can be computed in a distributed manner without revealing individual d_i values.

Challenges and Solutions in WSN

Bandwidth Constraint

Solution: Use compact secret sharing schemes and optimize communication patterns, possibly adopting hierarchical sensor node structures where cluster heads manage intra-cluster communication.

Energy Constraint

Solution: Minimize interactive rounds in the protocol and consider energy-efficient cryptographic operations. Asynchronous operations can be adapted to allow nodes to enter low-energy states when not actively participating.

Node Failures

Solution: Employ error-correcting codes for share recovery and design the protocol to be resilient to node dropouts.

Security Considerations

In WSN, the threat model may differ, with concerns of node capture or eavesdropping. The security of SPDZ-2 in such a model ensures:

- Privacy: Individual sensor readings are kept confidential.
- Integrity: The outcome of the computation is correct even if some nodes are malicious.

3. Case study

3.1. Sensor Data Communication Overhead in the SPDZ-2 Protocol

The SPDZ-2 protocol, when applied to sensor networks, still has a significant communication overhead. This is especially problematic for wireless sensor networks, which may have limited bandwidth or be subjected to high-latency communication environments.

Communication Overhead in SPDZ

The communication overhead in the SPDZ protocol primarily arises from:

- Calculation, sharing and, reconstructing values in the preprocessing phase.
- Exchanging values during the online phase for operations like multiplication using Beaver's triples.
- Zero-knowledge proofs ensure honesty and correctness.

Strategies to Reduce Communication Overhead

Before initiating the SPDZ protocol, sensors can locally aggregate or summarize their data. For instance, instead of sending individual readings, sensors can send averages or other statistical summaries over a time window.

Group multiple operations together, especially during the preprocessing phase. This can help amortize the cost of generating and distributing values like Beaver's triples over multiple operations.

Instead of running individual proofs for each operation, consider batched or aggregated proofs that can cover multiple operations at once.

Implement secret sharing schemes that are tailored for sensor networks. These can focus on minimizing the number of shares or using techniques like error-correcting codes to handle lost or delayed shares without retransmission.

Employ data compression algorithms to reduce the size of the transmitted data. This can be especially effective if sensor readings or intermediate values in the SPDZ protocol have redundancy or predictable patterns.

Instead of all-to-all communication, consider using relay nodes or hierarchical structures where a subset of sensors aggregates data and communicates with other groups, reducing the total communication across the network.

Instead of continuous computation, synchronize the computation in intervals. This allows for more batched operations and fewer real-time communication requirements. Reducing the communication overhead in the SPDZ protocol when applied to sensor networks requires a combination of algorithmic optimizations, architectural considerations, and leveraging domain-specific knowledge of sensor data. Implementing the above strategies can significantly enhance the efficiency of the SPDZ protocol in sensor environments.

The current paper focuses on the algorithms related to reducing the communication overhead in the preprocessing phase of the SPZD-2 protocol. One of the possible ways to reduce the communication overhead in the preprocessing phase of the SPDZ protocol in WSNs is to use technique such Fixed-key block ciphers.

Fixed-key block ciphers [18], as the name suggests, involve the use of block ciphers with a fixed, predefined key. The idea behind using a fixed key is to transform the block cipher into a deterministic function with pseudorandom behavior.

Standard Block Cipher: A standard block cipher can be denoted as:

$$E : \{0, 1\}^k \times \{0, 1\}^n \rightarrow \{0, 1\}^n \quad (14)$$

where E is the encryption function. The first parameter is a key of length k bits. The second parameter is a plaintext block of length n bits. The output is a ciphertext block of length n bits. For a given key K and plaintext P , the encryption is denoted as

$$C = E(K, P) \quad (15)$$

Fixed-Key Block Cipher: When we talk about a fixed-key block cipher, the key remains constant. This can be represented as:

$$E_{K_{fixed}} : \{0, 1\}^n \rightarrow \{0, 1\}^n \quad (16)$$

where K_{fixed} is a predefined constant key. For any input block P , the output is $E(K_{fixed}, P)$.

With the key fixed, a block cipher behaves like a pseudorandom permutation (PRP) over the set of n -bit strings. This means that for every input P , there is a unique output C , and the relationship appears random unless you know the fixed key.

```
Function FixedKey_DDESES_Encrypt(input_block) :
// Define a fixed key; this remains constant.
    FIXED_KEY = "32-byte key derived
    from a secure process"

// Use DES encryption with the fixed key.
    ciphertext = DES_Encrypt(FIXED_KEY,
    input_block)
    return ciphertext
End Function

Function FixedKey_DES_Decrypt(ciphertext) :
// Define the same fixed key.
    FIXED_KEY = "32-byte key derived from
    a secure process"
// Use DES decryption with the fixed key.
    plaintext = DES_Decrypt(FIXED_KEY,
    ciphertext)
    return plaintext
End Function
```

The `FIXED_KEY` should be securely generated, preferably using a cryptographically secure random number generator, and then kept constant for all future operations. Storing cryptographic keys securely is essential. Depending on the application, you might consider using hardware security modules, secure key storage services, or other best practices.

It is essential to ensure that the `input_block` has an appropriate size for the block cipher is used. For DES, this would typically be 128 bits (or 16 bytes). For the same input, the output will always be the same since the key remains constant.

Since block ciphers are permutations for a given key, the process is reversible. If you know the fixed key, you can decrypt any ciphertext produced by the fixed-key block cipher to retrieve the original input.

In the context of secure multi-party computation (SMPC), fixed-key block ciphers can be used to produce correlated randomness between parties or derive other types of structured randomness efficiently.

One notable application is in the generation of "oblivious pseudorandom functions" (OPRFs) where one party learns the output of a PRF on a specific input without the other party learning anything about the input or the output.

Integration between Beaver triple and Fixed-Key Block Ciphers

Beaver triples and fixed-key block ciphers are both techniques used within the realm of secure multi-party computation (SMPC). While they serve different primary functions and can sometimes be complementary, they can also be seen as alternative techniques in specific settings.

Primarily used for securely computing multiplication in SMPC protocols, Beaver triples [19] consist of preprocessed random multiplicative triples (a,b,c) where $c=a \times b$. These triples allow parties to perform multiplication on secret-shared values without revealing their actual inputs.

The generation of Beaver triples can be computationally intensive, especially in protocols that require a large number of such triples. However, once generated, they make the online phase of the computation faster. Used widely in SMPC protocols like SPDZ and its variants. They are fundamental for protocols that rely on secret sharing and require multiplication operations.

Beaver Triples offer strong security guarantees when generated correctly. Their security relies on the fact that the triples are random and independent of the inputs on which they will be used.

Fixed-Key Block Ciphers: Used to generate certain types of correlated randomness in SMPC. A fixed-key block cipher is a pseudo-random function where the key remains constant. Given the same input, it will always produce the same output, but changing even one bit of the input will produce a substantially different output.

Typically, block ciphers are relatively efficient, especially in hardware implementations. Using them to produce correlated randomness can sometimes be more efficient than generating Beaver triples, depending on the protocol and context. Often used in oblivious pseudo-random function (OPRF) [20] contexts and other settings where correlated randomness or specific patterns of randomness are required.

The security here typically depends on the underlying block cipher's robustness and resistance against cryptographic attacks. If a cryptographically secure block cipher is used, the fixed-key variant can provide strong security guarantees for its purpose.

3.2. Reducing the Sensor Data Communication Overhead in the SD-SPDZ Protocol

Utilizing fixed-key block ciphers to substitute the Beaver triple generation in the SPDZ preprocessing phase is an advanced topic in secure multi-party computation, and this approach is at the core of the new proposed SD-SPDZ protocol.

The idea behind this technique is to use block ciphers, like DES, to deterministically generate shared randomness, which can be used to produce Beaver triples.

The high-level approach for this is:

Key Generation: Each party selects a secret key for the block cipher (e.g., DES).

Beaver triple generation using Fixed-Key Block Ciphers:

Generation of a: Each party P_i generates a random value. Each party computes:

$$A_i = \text{Encrypt}_{key_i}(a_i) \quad (17)$$

and broadcast it. The shared value a is the sum of the a_i values.

Generation of b: Each party P_i generates a random value b_i . Each party computes:

$$B_i = \text{Encrypt}_{key_i}(b_i) \quad (18)$$

and broadcast it. The shared value b is the sum of the b_i values.

Generation of c: The shared value $c = a \times b$ is computed. However, instead of interacting to verify the correctness of this multiplication, the sensor modules can use the fact that they have encryption of the values a_i and b_i . They can derive the product of the encrypted values, given the properties of the fixed-key block cipher and the determinism of their chosen function. This step avoids the need for complex interactive proofs, hence removing the original need for Beaver triples.

```
function generate_triples_using_block_cipher():
    # a-values
    a_i = random_value()
    A_i = Encrypt_with_fixed_key(key_i, a_i)
    broadcast(A_i)
    a = sum_of_broadcasted_A_values
    # b-values
    b_i = random_value()
    B_i = Encrypt_with_fixed_key(key_i, b_i)
    broadcast(B_i)
    b = sum_of_broadcasted_B_values
    # Compute c using encrypted values and
    # properties of the block cipher
    c = compute_all_A_values, all_B_values)
    return (a, b, c)
```

This approach dramatically simplifies the preprocessing phase compared to the standard SPDZ protocol with Beaver triples and reduces the sensor data communication overhead. However, it assumes that the fixed-key block cipher has certain properties that make this method secure and that the encryption/decryption operations are performed in a secure manner.

Lab environment

The lab environment consists of a cluster-based sensor network consisting of five sensor modules based on NUCs Gigabyte and control center shown in the picture below:

The testing software is implemented in each sensor module and at the cluster head (CH). The experimental results are shown in the table below which describes the average time in seconds to compute 10.000 triples in a WSN cluster consisting of five sensor nodes:

Table 1
Experimental results

MPC protocol Preprocessing phase	Standard Beaver Triple calculation	Fixed-Key Block Ciphers triple calculation
SPDZ	7	-
SPDZ-2	4	-
SD-SPDZ	4	0.7



Figure 1: Cluster-based sensor network consisting of five sensor modules based on NUCs Gigabyte and control center shown in the picture below.

4. Conclusion

This paper presents a newly developed protocol named SD-SPDZ (Sensor Data SPDZ). The proposed protocol is based on MPC SPDZ-2 protocol and proposes changes to increase the performance in the preprocessing phase by implementing a new algorithm for the Beaver triples calculation.

This protocol enhances the privacy-preserving attributes and efficiency of its predecessors. SD-SPDZ integrates advanced cryptographic techniques, offering a more robust and scalable solution for secure computations in WSNs. The primary benefits include reduced communication overhead, faster computation times, and improved resistance against various cyberattacks.

The integration of SD-SPDZ in WSNs could improve performance sensitively and change the way sensor data is securely processed in sensor networks. It provides a promising pathway to ensure that as technology advances, the integrity and confidentiality of the data in these networks remain uncompromised.

In summary, as WSNs play an increasingly critical role in modern-day applications, the need for advanced high-performance security mechanisms such as the SD-SPDZ protocol becomes more evident. This combination of cutting-edge, high-performance, secure computation with wireless sensor networks promises a future where data can be both globally accessible and privately computed, bridging the gap between performance and privacy.

References

- [1] Y. Pinar, A. Zuhair, A. Hamad, A. Resit, K. Shiva, A. Omar, Wireless sensor networks (WSNs), in: 2016 IEEE Long Island Systems, Applications and Technology Conference (LISAT), IEEE, 2016, pp. 1–8.
- [2] J. Zhao, Des-co-rsa: A hybrid encryption algorithm based on DES and RSA, in: 2023 IEEE 3rd International Conference on Power, Electronics and Computer Applications (ICPECA), IEEE, 2023, pp. 846–850.
- [3] N. Ahmad, S. R. Hasan, A new asic implementation of an advanced encryption standard (AES) crypto-hardware accelerator, *Microelectronics Journal* 117 (2021) 105255.
- [4] Y. Li, J. Feng, Q. Zhao, Y. Wei, Hdlbc: A lightweight block cipher with high diffusion, *Integration* 94 (2024) 102090.
- [5] H. Noura, O. Salman, R. Couturier, A. Chehab, Lesca: Lightweight stream cipher algorithm for emerging systems, *Ad Hoc Networks* 138 (2023) 102999.
- [6] K. Pavani, P. Sriramya, Enhancing public key cryptography using RSA, RSA-CRT and N-prime RSA with multiple keys, in: 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), IEEE, 2021, pp. 1–6.

- [7] H. Goyal, S. Saha, Multi-party computation in iot for privacy-preservation, in: 2022 IEEE 42nd International Conference on Distributed Computing Systems (ICDCS), IEEE, 2022, pp. 1280–1281.
- [8] R. Gennaro, M. Di Raimondo, Secure multiplication of shared secrets in the exponent, *Information processing letters* 96 (2005) 71–79.
- [9] C. Hazay, P. Scholl, E. Soria-Vazquez, Low cost constant round MPC combining bmr and oblivious transfer, *Journal of cryptology* 33 (2020) 1732–1786.
- [10] E. Orsini, N. P. Smart, F. Vercauteren, Overdrive2k: efficient secure MPC over from somewhat homomorphic encryption, in: *Cryptographers’ Track at the RSA Conference*, Springer, 2020, pp. 254–283.
- [11] I. Damgård, V. Pastro, N. Smart, S. Zakarias, Multi-party computation from somewhat homomorphic encryption, in: *Annual Cryptology Conference*, Springer, 2012, pp. 643–662.
- [12] R. Cramer, I. Damgård, D. Escudero, P. Scholl, C. Xing, SPDZ2k: efficient MPC mod 2k for dishonest majority, *CRYPTO*, 2018.
- [13] M. Keller, Mp-spdz: A versatile framework for multi-party computation, in: *Proceedings of the 2020 ACM SIGSAC conference on computer and communications security*, 2020, pp. 1575–1590.
- [14] M. Bottarelli, P. Karadimas, G. Epiphaniou, D. K. B. Ismail, C. Maple, Adaptive and optimum secret key establishment for secure vehicular communications, *IEEE Transactions on Vehicular Technology* 70 (2021) 2310–2321.
- [15] H.-J. Kim, H.-I. Kim, J.-W. Chang, A privacy-preserving kNN classification algorithm using Yao’s garbled circuit on cloud computing, in: *2017 IEEE 10th international conference on cloud computing (CLOUD)*, IEEE, 2017, pp. 766–769.
- [16] J. Liu, Y. Tian, Y. Zhou, Y. Xiao, N. Ansari, Privacy preserving distributed data mining based on secure multi-party computation, *Computer Communications* 153 (2020) 208–216.
- [17] G. Arumugam, V. L. Praba, S. Radhakrishnan, Study of chaos functions for their suitability in generating message authentication codes, *Applied Soft Computing* 7 (2007) 1064–1071.
- [18] C. Guo, J. Katz, X. Wang, Y. Yu, Efficient and secure multiparty computation from fixed-key block ciphers, in: *2020 IEEE Symposium on Security and Privacy (SP)*, IEEE, 2020, pp. 825–841.
- [19] J. B. Nielsen, P. S. Nordholt, C. Orlandi, S. S. Burra, A new approach to practical active-secure two-party computation, in: *Annual Cryptology Conference*, Springer, 2012, pp. 681–700.
- [20] S. Casacuberta, J. Hesse, A. Lehmann, SoK: Oblivious pseudorandom functions, in: *2022 IEEE 7th European Symposium on Security and Privacy (EuroS&P)*, IEEE, 2022, pp. 625–646.

The Interplay of Social and Robotics Theories in AGI Alignment: Navigating the Digital City through Simulation-based Multi-Agent Systems

Ljubiša Bojić^{1,2,*}, Vladimir Đapić¹

¹The Institute for Artificial Intelligence Research and Development of Serbia, Fruskogorska 1, 21000 Novi Sad, Serbia

²Digital Society Lab, Institute for Philosophy and Social Theory, University of Belgrade, Kraljice Natalije 45, 11000 Belgrade, Serbia

Abstract

This study delves into the task of aligning Artificial General Intelligence (AGI) and Large Language Models (LLMs) to societal and ethical norms by using theoretical frameworks derived from social science and robotics. The expansive adoption of AGI technologies magnifies the importance of aligning AGI with human values and ethical boundaries. This paper presents an innovative simulation-based approach, engaging autonomous 'digital citizens' within a multi-agent system simulation in a virtual city environment. The virtual city serves as a platform to examine systematic interactions and decision-making, leveraging various theories, notably, Social Simulation Theory, Theory of Reasoned Action, Multi-Agent System Theory, and Situated Action Theory. The aim of establishing this digital landscape is to create a fluid platform that enables our AI agents to engage in interactions and enact independent decisions, thereby recreating life-like situations. The LLMs, embodying the personas in this digital city, operate as the leading agents demonstrating substantial levels of autonomy. Despite the promising advantages of this approach, limitations primarily lie in the unpredictability of real-world social structures. This work aims to promote a deeper understanding of AGI dynamics and contribute to its future development, prioritizing the integration of diverse societal perspectives in the process.

Keywords

Artificial General Intelligence, Large Language Models, Social Theories, Robotics Theories, Simulation-Based Approach

1. Introduction

The increasingly pervasive role of AI, especially natural language processing (NLP), signifies a new frontier of technological development. AI-driven applications like Generative Pretrained Transformers (GPT) pioneer transformations across society [1]. As reliance on such AI systems rises, so does the challenge of adapting these models to human values, prompting deeper research and development.

Despite rapid advancements, achieving full controllability and value alignment with AI is a notable hurdle, especially with large-scale neural networks [2]. The rise of powerful AI models like GPT further amplifies concerns about their ethical alignment, controllability, and unpredictability [3]. This pressure intensifies the exploration of better testing and mitigation strategies [1].

Large Language Models (LLMs) are artificial intelligence (AI) programs capable of language generation, translation, question answering, summarization, and code generation [4]. Unlike traditional AI models, which are trained on specific datasets and for particular tasks,

LLMs are trained on diverse internet text content. They have demonstrated performance in a wide range of tasks and languages without any task-specific training [4], a capability that resonates with the concept of artificial general intelligence (AGI).

AGI refers to a type of AI with cognitive capabilities that can successfully understand, learn, and implement intellectual tasks equivalent to those of a human being [5]. Contrary to traditional AI that is limited to expert-level competence in specific tasks, AGI can understand, learn, and adapt to any intellectual task that can be performed by humans [5]. The universality of this ability in AGI is often considered as both beneficial and dangerous. While it promises extensive progress and efficiency in virtually all fields of life, it also imposes significant risks related to misuse and unintended consequences.

Aside from text generation, sophisticated Large Language Models (LLMs) also exhibit the capacity to simulate understanding of inquiries and perform complex cognitive tasks [6]. Among numerous platforms, OpenAI's LLMs stand out due to their potential for fine-tuning, making them compatible with a wide range of use-cases. This adaptability sets the stage for their comprehensive influence and application across diverse fields. OpenAI continues the development of Artificial General Intelligence publicly while devising strategies that ensure AGI's safety and alignment with human values [7]. On the other hand, LLMs can be given various degrees of autonomy while creating multiple agents with different

BISEC'23: 14th International Conference on Business Information Security, November 24, 2023, Niš, Serbia

*Corresponding author.

✉ ljubisa.bojic@ivi.ac.rs (L. Bojić); vladimir.djapic@ivi.ac.rs (V. Đapić)

📄 0000-0002-5371-7975 (L. Bojić); 0000-0002-8661-0269 (V. Đapić)

prompts capable of interacting with each other [8].

AI Alignment represents the proposition of ensuring that the behavior of AGI system is congruent with human intentions and values. As Bostrom [9] argues in his book "Superintelligence," it is incredibly challenging to specify what is meant by human values in a way that an AI can understand. The alignment of AGI is considered crucial due to multiple reasons. The development of AGI might lead to an intelligence explosion where AGI surpasses human intelligence. If such a situation arises, it is important to ensure that AGI is beneficially aligned and promotes the interests of humanity [9]. Moreover, poorly aligned AI could result in negative ramifications if it can impact significant resources or make autonomous decisions. Hence, dedicated research is needed to ensure that AGI development is carried out responsibly and with necessary precautions.

AI and AGI advancements come with benefits, complexities, risks, and ethical challenges. With traditional risk management methodologies proving inadequate, there's a shift towards exploring more multi-layered methodologies [10]. The unpredictability of AI and AGI systems poses risks, underpinning the necessity of embedding human values and ethics into AI systems [3]. Transparent, accountable AI systems developed with public involvement are advocated by scholars like Véliz [11] and Whittlestone et al. [12], leading to the democratization of technology. The unification of social science theories and technology offers a promising path for developing socially-responsible AI and AGI [13, 14].

This paper delves into the potentials and challenges of AI and social robotics theory convergence for aligning AGI and LLMs. It explores theories and their application in AI alignment, demonstrating their relevance in simulation-based approaches within a digital city environment. The paper concludes with reflections on limitations and directions for future research, essential for ensuring AGI technologies are effective, secure, and uphold societal values

2. Theoretical framework

Exploring social science and robotics theories can provide critical insights for testing and aligning Large Language Models (LLMs) and artificial general intelligence (AGI). The complexity of LLMs and AGIs demand a stringent, theory-based approach [13]. Social science theories aid in understanding and predicting AI behavior [15], while robotics theories provide essential insights on machine ethics and multi-agent system operation for AGI design and refining [16].

Incorporating social science theories in AI research grants a lens for understanding AI alignment and behavior. The relevance of Social Simulation Theory and

Theory of Reasoned Action is considerable. Stemming from the Computational Social Science spectrum, Social Simulation Theory leverages computational methods for simulating and analyzing social dynamics, thus driving tests for large language models and better aligning AI behavior to social norms [17, 18]. However, representing the unpredictable nature of real-world social systems in abstract computational models is a significant challenge, limiting the theory's accuracy and applicability [19].

The Theory of Reasoned Action, from social psychology, asserts that intentions drive behavior, influenced by attitudes towards the behavior, norms, and perceived control [20]. While originally for understanding human behavior, it can guide AI behavior modeling, influencing AI intentions via programmed norms and attitudes, and helping align AI actions with societal values [21]. However, the challenge lies in replicating the complex nature of human emotions and irrational behavior in AI, emphasizing the need for a multifaceted AI alignment approach.

Asimov's Laws of Robotics and The Uncanny Valley Hypothesis offer insights for AI security, concerning human-AI interactions [22]. Asimov's Laws provide ethical guidelines enhancing AI system's controllability and ethical behavior. Yet, ambiguity in AI behavior complicates adherence to these laws [23].

The Uncanny Valley Hypothesis highlights the comfort of users with human-like AI, stressing careful design to ensure secure AI usage [24]. Despite the theory's cultural subjectivity, considering such perceptions augments holistic AI system design, balancing advancement with ethical responsibility and security. Multi-Agent System Theory offers valuable insights for developing autonomous systems and testing LLMs and AGI. Multi-agent systems of AI agents, each with unique attributes and decisions in a simulated digital city, can reveal emergent behavior and systemic strengths or weak points. Challenges, though, include agent synchronization, conflict resolution, and handling competition [25]. Despite these, the theory provides crucial support for AI testing in simulated environments. Situated Action Theory encourages adaptive, situation-driven behavior, enhancing AI responses to digital environments. This theory implies AI models should adapt dynamically to changes rather than sticking to prescribed actions. This approach equips AI to navigate unpredictability inherent in large networks.

However, translating these concepts into AI programming proves challenging due to reality's multidimensional and ambiguous nature. Designing adaptive behavior based on Situated Action Theory helps decipher cognitive functions in simulated environments, paving the way for advanced, reliable AI systems.

Next, we examine the practical implementation of these theories for AGI, focusing on developing a digital

city. Subsequent section will reflect on the simulation's results, offering insights for alignment of AI models.

3. Towards simulation of a digital city

A simulation-based methodology enhances the reliability, efficacy, and safety of Large Language Models (LLMs) in AGI development [26]. The authors note that simulations provide controlled settings for testing AI behaviors under various scenarios. This digital city simulation, inspired by McEwan et al. [27], effectively mimics real-life complex interactions in a controlled setting. As such, these tested procedures have become instrumental in AGI development.

In this research, a virtual reality framework adds a potent and immersive dimension to simulation studies, a paradigm gaining wider acceptance [28]. Enhanced with AI, this approach offers opportunities for in-depth analysis of AI interactions in realistic scenarios [29].

By incorporating virtual reality, we tap into a broader context for AI implementation. Lending support to Bolton et al. [30], the creation of a 'digital twin' or 'mirror world' facilitates dynamic AI learning. It triples as a platform for appreciating AI behaviors, an arena for future social sciences research, and a toolkit for understanding social dynamics [31].

A simulation-based approach as noted by Bostrom & Yudkowsky [32], enhances the evaluation of AI, especially LLMs behavior. This methodology, bolstered by a virtual reality dimension, holds potential to remarkable breakthroughs in AGI understanding and enhancement.

Automated simulations for LLMs form the cornerstone of our approach, offering reproducible, scalable, and complex interactive environments [33]. Our digital city employs a multi-agent-based simulation framework, modeling a population of autonomous AI agents or 'digital citizens' [34]. Heath et al. [35] affirm the effectiveness of such agent-based models in understanding complex environments.

The development of this digital realm involves iterative creation of autonomous agents operating within defined parameter spaces [36]. Their autonomy determines their dynamics within the city [37]. A meticulously designed environment, where the AI agents function, necessitates a thorough attention to interactions, constraints, and choices [38]. Continuity in learning behavior and refinement of AI agents are ensured by a reinforcement learning approach, as proposed by Leike et al. [2]. The creation of these simulations significantly influences the lockdown approach's effectiveness in providing real-life scenario-based insights for AGI.

Describing the digital citizens, Bartneck et al. [39] underscore their importance in our simulation strategy. Act-

ing as AI actors, these agents vary in personality, norms, and behaviors, enriching the simulation's scenarios and insights. Autonomy, or the capacity to act independently, is critical for AI agents' value and effectiveness [40].

Various learning models, such as reinforcement learning, are utilized for shaping digital citizens [41]. Interaction and responsiveness to their environment, other AI agents, and external inputs is paramount [42]. Personified digital citizens, complete with autonomy, natural language-processing capabilities, character traits, and unique behaviors, significantly enhance multi-agent simulations [43]. Such enhancement underpins our objectives for AGI development [9].

Our digital environment's richness allows observation and manipulation of variables influencing AI behavior, with significant emphasis on interactions and decision-making of digital citizens [44]. Interactions and decisions form the crux of our simulation, driving insights into AI behavior under various scenarios.

Interactions can range from simple exchanges to conflict resolutions and cooperative tasks [45]. Decision-making forms a crucial part of an autonomous agent's function, stretching from simple choices to complex trade-offs [41]. These interactions and decisions provide data useful in refining AI models and informing digital technology policies [46]. Our simulation-based approach provides invaluable insights for AGI and influences its use [47]. The immersive environment offers simulations of significant clinical, social, and psychological interest [48]. These understandings, extending beyond AGI performance, help anticipate and shape AGI's potential societal impact [49].

Data from the digital city facilitates bias addressing in AGI systems [50]. Areas like autonomous vehicles, robotics, customer service, and translation would gain from information acquired in the digital city environment [51]. The virtual city also underlines the ethical considerations and value alignment issues concerning AGI [9]. The use of a simulation approach in a digital city enriches understanding of AGI dynamics, helping society harness AGI innovations responsibly.

Aligning AI models with human values is critical, especially in AGI, which has the potential to mimic human-like reasoning, including ethical decision-making [9]. Observations from interactions within our simulated approach assist in identifying and rectifying AGI's anomalies and misalignments.

Understanding how models encode knowledge is crucial for AI alignment [52]. Our simulation-based testing offers insights into AI's cognitive understanding, giving a better overview of its decision-making processes. Decision-making in AGI leverages reinforcement learning, but it requires careful management to avoid endorsing undesired behaviors [46].

The behaviors and interactions of digital citizens

within our simulation offer rich data for AGI refinement [53]. This scenario-based data aids in developing safety measures, aligning AGI with human values, and mitigating the risks of AI integration into society. Consequently, this enables the creation of safer, controlled, and value-aligned AI systems.

4. Conclusion

AI growth necessitates innovative security solutions and alignment with human values. Through contriving a digital city with digital citizens, various societal interactions can be explored to gain insights into AI behavior. Key theories guiding our approach include social simulation and theory of reasoned action for studying AI behavior in social contexts. Robotics theories illuminate ethical considerations, informed by Asimov's Laws of Robotics and the Uncanny Valley Hypothesis.

The application of Multi-Agent System Theory and Situated Action Theory helps manage AI behaviors, guiding interactions, and environment-response adaptations. This accentuates AI alignment with desired outcomes despite potential challenges. Our approach highlights automated simulations for exhaustive study of AI behavior. Autonomous citizens' interactions provide rich data for understanding autonomy, crucial for AGI refinement and broader societal applications. Simulations also help design value-aligned AGIs. However, challenges exist with theory application to AI programming and replicating real-world effects. Nevertheless, simulation-based approaches show promise for aligning AI with human values, despite complexities.

Our approach also has limitations, primarily the difficulty in replicating complexities of real societies within a digital space. Translating theoretical concepts into AI programming presents additional challenges. Biases in AI models can be perpetuated from training environments, and defining "desirable" behavior for AI alignment proves complex.

Future research can enhance simulation realism using advanced VR and AR technology. Focus should also be on refining theory integration into AI programming and developing automated bias correction frameworks. There's also the need to build definitions of AI alignment that respect the dynamism of values across cultures. This research is a starting point for harnessing theories and simulation-based approaches towards value-aligned AGI.

Acknowledgment

This paper monograph was realised with the support of the Ministry of Science, Technological Development and Innovation of the Republic of Serbia, according to the

Agreement on the realisation and financing of scientific research.

References

- [1] M. X. Chen, O. Firat, A. Bapna, M. Johnson, W. Macherey, G. Foster, L. Jones, N. Parmar, M. Schuster, Z. Chen, et al., The best of both worlds: Combining recent advances in neural machine translation, arXiv preprint arXiv:1804.09849 (2018).
- [2] J. Leike, M. Martic, V. Krakovna, P. A. Ortega, T. Everitt, A. Lefrancq, L. Orseau, S. Legg, Ai safety gridworlds, arXiv preprint arXiv:1711.09883 (2017).
- [3] G. Irving, A. Askill, Ai safety needs social scientists, *Distill* 4 (2019) e14.
- [4] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, I. Sutskever, et al., Language models are unsupervised multitask learners, *OpenAI blog* 1 (2019) 9.
- [5] Gartner, Definition of Artificial general intelligence (AGI), <https://www.gartner.com/en/information-technology/glossary/artificial-general-intelligence-agi>, 2023.
- [6] G. Sartori, G. Orrù, Language models and psychological sciences, *Frontiers in Psychology* 14 (2023).
- [7] S. Altman, Planning for agi and beyond, *OpenAI Blog*, February (2023).
- [8] B. Lutkevich, Auto-GPT, <https://www.techtarget.com/whatis/definition/Auto-GPT>, 2023.
- [9] T. Mulgan, Superintelligence: Paths, dangers, strategies, 2016.
- [10] S. J. Russell, P. Norvig, *Artificial intelligence a modern approach*, London, 2010.
- [11] B. Rumbold, Privacy is power: Why and how you should take back control of your data, written by carissa véliz, *Journal of Moral Philosophy* 20 (2023) 585–587.
- [12] J. Whittlestone, R. Nyrup, A. Alexandrova, K. Dihal, S. Cave, *Ethical and societal implications of algorithms, data, and artificial intelligence: a roadmap for research*, London: Nuffield Foundation (2019).
- [13] I. Rafols, Knowledge integration and diffusion: Measures and mapping of diversity and coherence, in: *Measuring scholarly impact: Methods and practice*, Springer, 2014, pp. 169–190.
- [14] S. Cave, S. S. ÓhÉigeartaigh, Bridging near-and long-term concerns about ai, *Nature Machine Intelligence* 1 (2019) 5–6.
- [15] E. Ostrom, A general framework for analyzing sustainability of social-ecological systems, *Science* 325 (2009) 419–422.
- [16] P. Lin, K. Abney, G. A. Bekey, *Robot ethics: the ethical and social implications of robotics*, MIT press, 2014.

- [17] C. Cioffi-Revilla, Introduction to computational social science, Springer, 2014.
- [18] M. W. Macy, R. Willer, From factors to actors: Computational sociology and agent-based modeling, *Annual review of sociology* 28 (2002) 143–166.
- [19] B. Edmonds, S. Moss, From kiss to kids—an ‘anti-simplistic’ modelling approach, in: *International workshop on multi-agent systems and agent-based simulation*, Springer, 2004, pp. 130–144.
- [20] M. Fishbein, I. Ajzen, Predicting and changing behavior: The reasoned action approach, Taylor & Francis, 2011.
- [21] P. Sheeran, T. L. Webb, The intention–behavior gap, *Social and personality psychology compass* 10 (2016) 503–518.
- [22] I. Asimov, *I, Robot.*, New York: Gnome Press., 1950.
- [23] J. J. Bryson, Robots should be slaves, *Close Engagements with Artificial Companions: Key social, psychological, ethical and design issues* 8 (2010) 63–74.
- [24] M. Mori, The uncanny valley, *Energy* 7 (1970) 33–35.
- [25] G. Weiss, *Multiagent systems: a modern approach to distributed artificial intelligence*, MIT press, 1999.
- [26] S. D. T. Kelly, N. K. Suryadevara, S. C. Mukhopadhyay, Towards the implementation of iot for environmental condition monitoring in homes, *IEEE sensors journal* 13 (2013) 3846–3853.
- [27] G. F. McEwan, M. L. Groner, M. D. Fast, G. Gettinby, C. W. Revie, Using agent-based modelling to predict the role of wild refugia in the evolution of resistance of sea lice to chemotherapeutants, *PLoS One* 10 (2015) e0139128.
- [28] R. C. A. Barrett, R. Poe, J. W. O’Camb, C. Woodruff, S. M. Harrison, K. Dolguikh, C. Chuong, A. D. Klassen, R. Zhang, R. B. Joseph, et al., Comparing virtual reality, desktop-based 3d, and 2d versions of a category learning experiment, *Plos one* 17 (2022) e0275119.
- [29] J. Vora, S. Nair, A. K. Gramopadhye, A. T. Duchowski, B. J. Melloy, B. Kanki, Using virtual reality technology for aircraft visual inspection training: presence and comparison studies, *Applied ergonomics* 33 (2002) 559–570.
- [30] R. N. Bolton, J. R. McColl-Kennedy, L. Cheung, A. Gallan, C. Orsingher, L. Witell, M. Zaki, Customer experience challenges: bringing together digital, physical and social realms, *Journal of service management* 29 (2018) 776–808.
- [31] L. Bojic, Metaverse through the prism of power and addiction: what will happen when the virtual world becomes more attractive than reality?, *European Journal of Futures Research* 10 (2022) 1–24.
- [32] N. Bostrom, E. Yudkowsky, The ethics of artificial intelligence, in: *Artificial intelligence safety and security*, Chapman and Hall/CRC, 2018, pp. 57–69.
- [33] J. Banks, *Discrete event system simulation*, Pearson Education India, 2005.
- [34] E. E. Bertacchini, G. Jakob, E. Vallino, et al., Emergence and evolution of property rights. an agent-based perspective, *WORKING PAPER SERIES* 40 (2013).
- [35] B. Heath, R. Hill, F. Ciarallo, A survey of agent-based modeling practices (january 1998 to july 2008), *Journal of Artificial Societies and Social Simulation* 12 (2009) 9.
- [36] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, et al., Mastering the game of go with deep neural networks and tree search, *nature* 529 (2016) 484–489.
- [37] J. Lehman, J. Clune, D. Misevic, C. Adami, L. Altenberg, J. Beaulieu, P. J. Bentley, S. Bernard, G. Beslon, D. M. Bryson, et al., The surprising creativity of digital evolution: A collection of anecdotes from the evolutionary computation and artificial life research communities, *Artificial life* 26 (2020) 274–306.
- [38] R. S. Olson, A. Hintze, F. C. Dyer, D. B. Knoester, C. Adami, Predator confusion is sufficient to evolve swarming behaviour, *Journal of The Royal Society Interface* 10 (2013) 20130305.
- [39] C. Bartneck, D. Kulić, E. Croft, S. Zoghbi, Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots, *International journal of social robotics* 1 (2009) 71–81.
- [40] P. Stone, M. Veloso, Multiagent systems: A survey from a machine learning perspective, *Autonomous Robots* 8 (2000) 345–383.
- [41] R. S. Sutton, A. G. Barto, *Reinforcement learning: An introduction*, MIT press, 2018.
- [42] S. Zhang, E. Dinan, J. Urbanek, A. Szlam, D. Kiela, J. Weston, Personalizing dialogue agents: I have a dog, do you have pets too?, *arXiv preprint arXiv:1801.07243* (2018).
- [43] J. Z. Leibo, V. Zambaldi, M. Lanctot, J. Marecki, T. Graepel, Multi-agent reinforcement learning in sequential social dilemmas, *arXiv preprint arXiv:1702.03037* (2017).
- [44] A. M. Turing, Computing machinery and intelligence, *mind* 59 (1950) 433–460.
- [45] N. R. Jennings, K. Sycara, M. Wooldridge, A roadmap of agent research and development, *Autonomous agents and multi-agent systems* 1 (1998) 7–38.
- [46] D. Amodei, C. Olah, J. Steinhardt, P. Christiano, J. Schulman, D. Mané, Concrete problems in ai safety, *arXiv preprint arXiv:1606.06565* (2016).
- [47] Y. Shoham, R. Perrault, E. Brynjolfsson, J. Clark,

- J. Manyika, J. C. Niebles, J. T. Etchemendy, et al., The ai index 2018 annual report – ai index steering committee, human-centered ai initiative, 2018.
- [48] C. Castelfranchi, Artificial liars: Why computers will (necessarily) deceive us and each other, *Ethics and Information Technology* 2 (2000) 113–119.
- [49] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, et al., Human-level control through deep reinforcement learning, *nature* 518 (2015) 529–533.
- [50] A. Caliskan, J. J. Bryson, A. Narayanan, Semantics derived automatically from language corpora contain human-like biases, *Science* 356 (2017) 183–186.
- [51] A. Esteva, A. Robicquet, B. Ramsundar, V. Kuleshov, M. DePristo, K. Chou, C. Cui, G. Corrado, S. Thrun, J. Dean, A guide to deep learning in healthcare, *Nature medicine* 25 (2019) 24–29.
- [52] M. Ring, L. Orseau, Delusion, survival, and intelligent agents, in: *Artificial General Intelligence: 4th International Conference, AGI 2011, Mountain View, CA, USA, August 3-6, 2011. Proceedings 4*, Springer, 2011, pp. 11–20.
- [53] P. F. Christiano, J. Leike, T. Brown, M. Martic, S. Legg, D. Amodei, Deep reinforcement learning from human preferences, *Advances in neural information processing systems* 30 (2017).

Survey on Methods of Online Payment over the Internet

Marina Dodevska^{1,*}

¹Ss. Cyril and Methodius University, Faculty of Computer Science and Engineering, Ruger Boskovik 16, 1000 Skopje, North Macedonia

Abstract

This paper aims to give an overview and show the ways of online payment over the Internet. To achieve that goal, we conducted a survey among different age groups in North Macedonia to find out how they manage online payment and what they think about this topic. This paper describes the online payment method and presents the survey results.

Keywords

Online payments, Online shops, Credit cards

1. Introduction

In this article [1] Online payments refer to the electronic exchange of currency through the Internet. These payments usually consist of transferring monetary funds from a customer's bank or debit or credit card account, into the seller's bank account, in exchange for products or services. These funds can come directly from a customer's credit card or checking account, or from an online payment system that is linked to both the buyer and seller's bank accounts. Online payments are used by buyers of goods and services, and the sellers of those goods and services. Several steps occur with the funds when they are transferred and received, especially between the two parties, that often require different types of software to successfully facilitate the transaction. The typical steps are below:

- Online purchase is made: A customer (buyer) provides the necessary information (debit or credit card, checking account information, etc.) to pay for goods or services. This data is then sent to a payment processing software or payment gateway.
- Information is encrypted: The payment gateway encrypts the payment details, such as the customer's name, address, and bank account info, which provides a level of security to make it more difficult for this info to be stolen.
- Details are verified: After the transaction data is encrypted, the information is sent to a payment processor to ensure that the transaction is valid. Once the transfer is verified, it sends the info to the buyer's and seller's banks.
- Funds are approved: Assuming there are no red

flags from the payment gateway or processor, the banks authorize the transaction.

There are, however, several reasons as to why a transaction might not be approved by either bank:

- Insufficient funds,
 - Frozen account status,
 - Invalid credit card number or expiration date,
 - Transaction limits,
 - The card has been reported lost or stolen,
 - The address does not match the card,
 - Invalid Card Code Verification (CCV),
-
- Funds are requested: After the funds and corresponding transfer are approved, the payment processor requests the funds to be sent from the buyer's source of funds to the seller's bank account.
 - The seller receives funds: The transfer of funds is completed and the purchase price has been sent from buyer to seller.

2. Types of Payments

In this article [2] Payment is the transfer of money, goods, or services in exchange for goods and services in acceptable proportions that have been previously agreed upon by all parties involved. A payment can be made in the form of services exchanged, cash, check, wire transfer, credit card, debit card, or cryptocurrencies. Payments are made using various methods. Throughout history, these types of payments have changed and evolved, and new payment methods are likely to appear in the future. Here are the most common types of payments used today.

Credit cards: Today, credit cards are widely used for purchases and payments. Credit cards work by offering their users a line where an individual can draw credit up to a certain limit. When you attempt to use your credit card, your account information is sent to the merchant

BISEC'23: 14th International Conference on Business Information Security, November 24, 2023, Niš, Serbia

*Corresponding author.

✉ marina.dodevska@students.finki.ukim.mk (M. D.)

bank. The merchant bank then receives authorization from the credit card network to process the transaction.

Debit cards: Debit cards may look similar to credit cards, but their underlying mechanism is entirely different. When a debit card is used, funds are immediately withdrawn from an individual's account. Instead of having a line of credit that you can pull from more than what you have saved, debit card transactions can be declined if you do not have enough money in your account.

Cash: Cash is still used for many businesses, such as the retail industry. Coffee shops and convenience stores, for example, still accept cash payments. Considering the fees associated with debit and credit cards, many retail small businesses prefer cash payments from their customers. Cash has its own disadvantages, as it can be lost, stolen, or destroyed. Businesses dealing in large transactions must often incur additional expenses to pay for related security measures such as secured transit or fraud detection.

Mobile phones: The contactless payment technology that has emerged in recent years has made payments easier than ever. The credit or debit card machine—called a point-of-sale terminal (POS)—can read the customer's banking information through the software application that's installed on the mobile device. Once the phone reads the information from the POS terminal, a signal is generated to inform the customer that the payment has been made.

Checks: Checks have fallen out of favor over the years due to advancements in technology, allowing payments to be electronically submitted. However, there are instances when checks might be helpful, such as when the seller wants a guaranteed payment. A bank cashier's check or a certified check are two types of checks that banks offer to help sellers receive the money owed from the buyer.

Electronic Funds Transfers: Wire transfers and ACH payments (Automatic Clearing House) are typically used for larger or more frequent payments in which a check or credit card wouldn't be appropriate. A payment from a manufacturer to a supplier, for example, would typically be made via wire transfer, particularly if it was an international payment. An ACH payment is often used for direct deposits of payroll for a company's employees.

Cryptocurrency: Digital currency or tokens are a more modern approach to facilitating transactions. The premise is simple: one person in possession of digital currency can send coins or tokens to any address on a blockchain. Blockchains with smart contract capabilities can interject logic to automatically withdraw or transfer specific amounts based on underlying conditions. The widespread use of cryptocurrency is still in its infancy stage, especially when compared with other payment systems above. However, cryptocurrency has the advantage of only needing an Internet connection to facilitate

a payment; as long as both parties have a digital wallet on the same network, payments can be made.

3. Related Work

This study [3] gives a wide knowledge of electronic payment systems, payment gateways, and their security considerations, and analyzes the electronic payment systems from an adaptability point of view with the aim to provide a better customer understanding and satisfaction.

The paper [4] explores the challenges associated with mobile payment security as well as realizes the concepts and rising technologies that will benefit mobile payment usability and security. Payments made through mobile are the lifeline of mobile commerce. This field is a chief part of a financial function or transaction, and it attracts extensive interest from everyone. Still, it needs to be converted into a regular approach to making payments. However, the technologies used in payments through mobile are enhanced and going through a noteworthy development in terms of security and service availability enhancement.

The payment systems developed should offer the safety of transactions at each and every point to advance the end user and organization's contentment. This [5] is the first comprehensive review study, which collected and classified the emerging digital payment technologies and associated challenges that provide theoretical and practical directions. This study offers some insight for future researchers in the field of digital payments, by (1) complementing previous literature review studies on digital payment technologies, (2) highlighting challenges associated with digital payment technologies and (3) providing a ground for building a sound digital payment ecosystem to overcome the challenges with digital payment technologies.

This study [6] finds that customers are increasingly using mobile payment methods for their routine online purchases and for their on-site purchases as well. With growing advanced technology that supports mobile transactions and makes them transparent and more convenient, customers have developed their trust and habits in using mobile payment systems. This research also concluded that for a promising future in this industry, mobile payment systems have to be better integrated with present telecommunication and financial infrastructure.

In this research [7] the authors discussed the challenges, risks, benefits, and future of e-payment in order to improve the e-business field as well as improve customer experience for the whole process of e-payment. They focus on fraud, which is the most important risk facing e-payment, and they demonstrate the new fraud detection and prevention models and techniques.

In our paper, we describe several methods of online

payment over the Internet.

4. Security Threats

According to this article [8] the types of security threats in digital payments are:

1. Fraudulent activities – are a persistent threat in the digital payment landscape. They are designed to steal sensitive information, such as credit card numbers, and login credentials, and use it for financial gain.
2. Data breaches – are another major concern in the digital payment world. This can happen when unauthorized individuals gain access to sensitive information stored in a digital payment system. The information stolen in a data breach can be used for fraudulent activities or sold on the dark web.
3. Malware attacks – are another serious threat in the digital payment landscape. Malware can infect a device and steal sensitive information, such as login credentials and credit card numbers.
4. Phishing scams – are a form of fraudulent activity that uses email or other forms of communication to trick individuals into revealing sensitive information, such as login credentials and credit card numbers. These scams often take the form of an email or message that appears to come from a trusted source, such as a bank or payment provider, and requests sensitive information.

These security threats are a reality in the digital payment landscape and individuals and businesses must be aware of the dangers and take steps to protect themselves.

5. Methodology

In our work, We surveyed 67 people in North Macedonia. We distributed the survey through social media and analyzed apps, with my friends and family and collected the responses for about one month. Our survey was answered by students, employees, and unemployed people.

The respondents were asked a different set of questions regarding online payment, web stores, and the methods they use for online payment. The results of the survey are presented and discussed in the next section.

6. Results

We asked our respondents what payment they prefer. As we can see from the picture, 53.7% prefer in cash, and 46.3% prefer online.

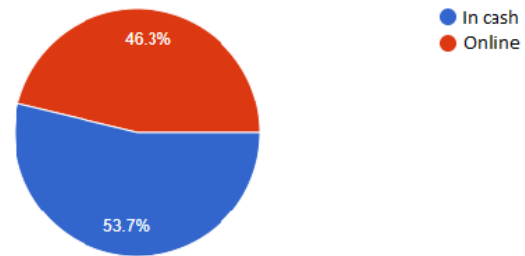


Figure 1: The respondents prefer In cash payment method.

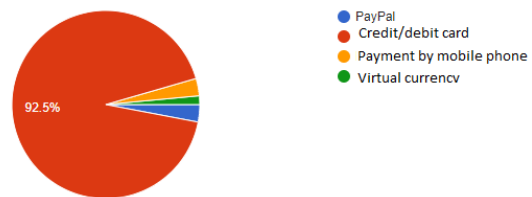


Figure 2: Most of the respondents use credit/debit cards.

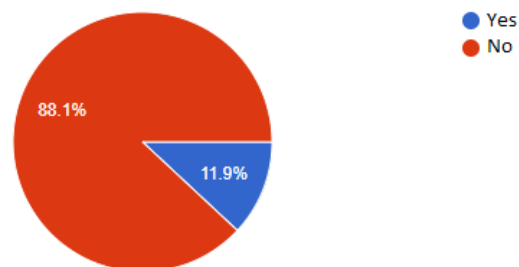


Figure 3: Respondents have no problem with online payment.

We asked our respondents which of the online payment methods they use. As we can see from the picture, 92.5% use Credit or debit cards, 3% use PayPal, 3% pay over the mobile phone, and 1.5% pay over the virtual currency. We can conclude that most respondents use credit/debit cards, and the least use virtual currency.

We asked our respondents to say they had a problem with online payment, and they answered 11.9% yes, and 88.1% no.

We asked our respondents if online payment was safe according to them. They answered with 23.9% no, and 76.1% with yes.

We asked our respondents which device they use to pay most often, and 46.3% of them used a computer, and

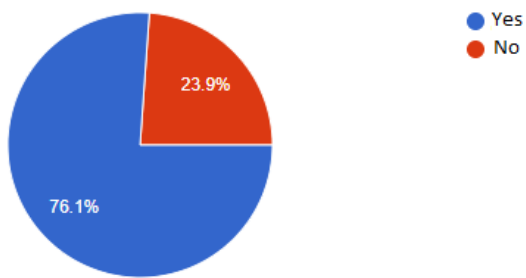


Figure 4: The respondents think that online payment is safe.

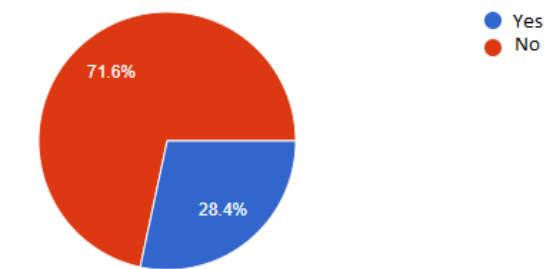


Figure 7: The respondents don't read the site's privacy policy before paying.

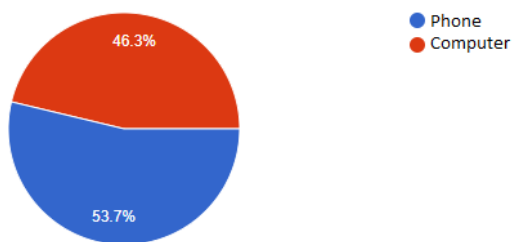


Figure 5: The respondents used a computer to pay most often.

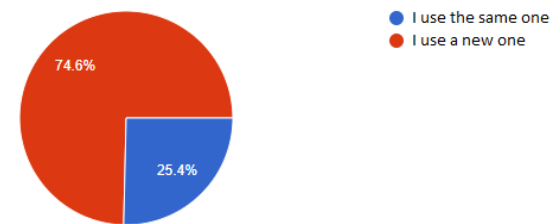


Figure 8: The respondents used a new password.

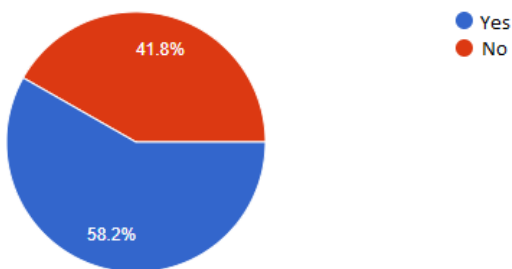


Figure 6: The respondent's trust in web stores.

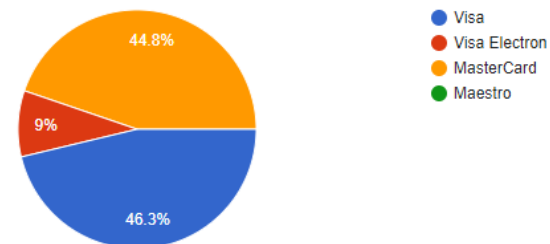


Figure 9: The respondents most often used Visa, and nobody used Maestro.

53.7% used a phone. We assume that this is because of the easy portability of mobile phones.

We asked our respondents do they trust web stores, and 41.8% said no, and 58.2% said yes.

We asked our respondents do they read the site's privacy policy before paying, and 28.4% said yes, and 71.6% said no.

We asked our respondents do they use the same password for every site they visit. 25.4% of our respondents used the same one, and 74.6% used a new one.

We asked our respondents which card they use most

often when paying. A total of 9% of them used Visa Electron, 44.8% used MasterCard, 46.3% used Visa and nobody used Maestro.

We asked our respondents do they check if the site is https before paying, and 71.6% said yes, and 28.4% said no.

We asked our respondents what kind of problems they had with online payment, and we had 7 different answers.

- "I didn't have a problem with payment, I had a problem with a refund when the quality was not met by the seller and a refund procedure had to be started."
- "Sometimes it happens that the transaction does

Table 1

ANOVA analysis for the different types of payment (cash or online) based on customer support.

Summary						
Groups	Count	Sum	Average	Variance		
Online	31	92	2.967	1.965		
In Cash	36	115	3.194	1.875		

Anova						
Source of variation	SS	df	MS	F	P-value	F crit
Between groups	0.856	1	0.856	0.446	0.506	3.998
Within groups	124.6	65	1.917			
Total	125.46	66				

Do you check if the site is https before paying?
67 responses

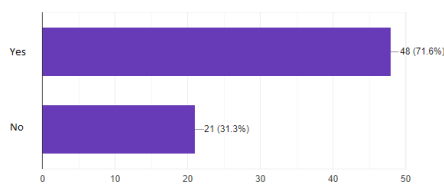


Figure 10: The respondents check if the site is https before paying.

Describe what problem you had with online payment?
7 responses

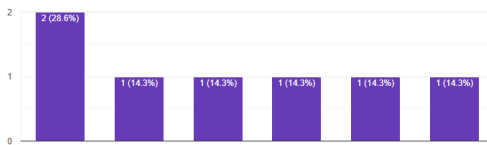


Figure 11: How many respondents had a problem with online payment?.

not go through, due to various reasons. For example, the service is not available in the country where I am, the seller put the wrong information, etc.”

- "For the debit card, the bank has since introduced a new document according to which they must approve you to make any of the transactions across the border. And on the original one, right, they are also covered without limit, right? Why else would someone take out a debit card here is not clear to anyone with their own thoughts, but hey?"
- "I paid for a semester online, the transaction failed, but the money was taken from my account."

Table 1 shows the ANOVA analysis for the different types of payment (cash or online) based on customer support.

Groups: We have two different types as we mentioned (online and in cash).

Count: This column specifies the responses we have for each type of payment.

Average: Specifies the average grade our respondents gave, for example, for the online payment we have an average grade of 2.96.

As you can see from the average grade for that type of payment, we can't say for sure if this is the correct value, so we will look at the ANOVA table for that. From the P-value, we can see that the value is 0.506342, this is the result when we set up the alpha version value when we conducted the ANOVA analysis. As we can see, we have 0.5, which means we have a statistical difference between our responses (this value is lower than the alpha value we set – we conclude our result from the value).

7. Conclusion

With our research in this digital age, we aimed to hear the opinions and experiences of our citizens regarding online payment as well as to describe the ways of online payment.

From the survey conducted among 67 respondents in North Macedonia, we can conclude that most of them prefer to pay in cash, although many of them consider that online payment is safe.

They usually pay through their phones because they are easily portable and commonly used devices. Most of our respondents trust online stores, so they feel there is no need to read the privacy policy.

Also, before paying they check the site for https, and use a new password when visiting the sites.

As we could see some answers from the survey, some of our respondents had problems with online payment, but we hope that with the constant development of technology, such problems will be fewer and fewer.

As the years go by, payments become cashless, and cash payments are used less and less.

Digital wallets, smart watches, biometric payment, etc. are just some of the ways technology is transforming the

traditional way payments have been made. However, the emergence and introduction of new technologies should be followed with great attention, and people must be prepared for future changes and all that modern times bring.

References

- [1] N. Calabrese, Online Payment, <https://www.g2.com/glossary/onlinepayment-definition>, 2023.
- [2] W. Kenton, Guide to payment types, with pros and cons for each, <https://www.investopedia.com/terms/p/payment.asp>, 2022.
- [3] M. Masihuddin, B. U. I. Khan, M. Mattoo, R. F. Olanrewaju, A survey on e-payment systems: elements, adoption, architecture, challenges and security concepts, *Indian Journal of Science and Technology* 10 (2017) 1–19.
- [4] J. D. Bronzino, The role of the engineer and iee in health care policy, *IEEE Aerospace and Electronic Systems Magazine* 3 (1988) 28–29.
- [5] K. Khando, M. S. Islam, S. Gao, The emerging technologies of digital payments and associated challenges: A systematic literature review, *Future Internet* 15 (2022) 21.
- [6] Z. Bezovski, The future of the mobile payment as electronic payment system, *European Journal of Business and Management* 8 (2016) 127–132.
- [7] M. H. Nasr, M. H. Farrag, M. Nasr, E-payment systems risks, opportunities, and challenges for improved results in e-business, *International Journal of Intelligent Computing and Information Sciences* 20 (2020) 16–27.
- [8] N. Gundaniya, Security concerns and solutions in the digital payment landscape, <https://www.digipay.guru/blog/securityconcerns-and-solutions-in-digital-payment/>, 2023.

Ethical Dimensions of AI Security and Privacy Policies: Enabling Inclusive Growth

Miloš Jovanović^{1,*}, Stefan Jančić²

¹Faculty of Information Technology, Belgrade Metropolitan University, Tadeuša Košćuška 63, 11000 Belgrade, Serbia

²Military Academy, University of Defence, 33 Veljka Lukića Kurjaka St., 11000 Belgrade, Serbia

Abstract

With the growing popularity of AI and the generative one, there are global societal changes in the fast-growing technical scene. On another note, there is an increasing alarm bell regarding ethical and security matters arising at a higher rate that should be addressed immediately. The sub-topic "Security and Privacy policies" is part of the major theme, "Generative AI security and ethics", in which ethical issues take the central position. This exploration goes deep into ethical issues, such as the security and privacy policies that are important when considering Generative AI systems. Such penetration of AI into everyday life requires to create a specific legislation aimed at assuring safety and security. These policies will guide the responsible, ethical, and equitable growth of generative AI as highlighted in this article.

The study reveals, however, the intricate nature of Generative AI security and privacy policies and aims to promote creativity. Therefore, we stress the importance of saving and not only saving but also protecting the data from AI-biased, discriminatory, or privacy violation policies. These policies, therefore, serve as a safety net for continuous improvement in AI advancement without undermining the universally accepted norms and values. Finally, this discussion highlights the importance of international cooperation and standardization in worldwide AI security and privacy. This will imply that nations can join forces and establish some common rules for tackling the challenges of a democratic and fair digital age, generative AI.

Keywords

Generative AI, Security, Privacy Policies, Ethical Issues, Legislation, Data Protection, Inclusive Growth, Geopolitics

1. Introduction

In our paced world technology has become the driving force that propels us forward. It shapes industries, influences dynamics and transforms the way we live and work. Today we embark on a journey to explore the influence of technology on the world economy throughout history and examine the trends that will shape our future. Additionally we will delve into how technology can pave the way towards growth by bridging divides.

The significance of technology in today's world cannot be overstated. We find ourselves in an era where innovation's not a luxury but an absolute necessity. Our daily lives are intricately intertwined with technology from our smartphones to the infrastructure that powers our societies. Technology enhances communication streamlines industries and grants us access, to information. As we delve into these lectures it's important to think about how technology has affected your life from the conveniences it offers to the opportunities it creates.

1.1. Significance of Understanding Technologies Impact on the World Economy

Why is it so crucial to understand how technology affects the world economy? The answer lies in our ability to navigate challenges and seize opportunities that lie ahead. This understanding isn't an exercise; it's essential for policymakers, businesses, as well as individuals, like yourself.

In today's connected world the decisions we make are closely linked to the trends, in technology. To make informed choices it is important to grasp the implications of technologies continuous progress. This understanding allows us to utilize its potential for fostering innovation driving growth bridging gaps and ensuring a prosperous future for everyone.

1.2. Understanding Historical Context

To fully comprehend the present and anticipate what lies ahead it is necessary to explore the context that has brought us to this point. The evolution of technology stands as a testament to creativity and resourcefulness. From the introduction of machinery during the Industrial Revolution to the rise of the internet era each period witnessed a convergence of innovation and transformative changes in economies. For instance we can reflect on how industrialization was fuelled by advancements like

BISEC'23: 14th International Conference on Business Information Security, November 24, 2023, Niš, Serbia

*Corresponding author.

✉ milos.jovanovic@metropolitan.ac.rs (M. Jovanović); stefans3100@gmail.com (S. Jančić)

🆔 0009-0008-9032-8195 (M. Jovanović); 0009-0008-2552-6275 (S. Jančić)

the steam engine, which reshaped economies and revolutionized labor markets. Fast forward to today's age where access to information has been democratized through the internet leading us into an era of globalization. These significant milestones offer insights into how technology impacts economies, industries and societies over time. Now we find ourselves standing at yet another moment in history as AI driven automation becomes more prevalent alongside developments, in microchips.

The decisions we make in the present have an impact, on both our well being and the future of our world. As we delve into these lectures it's important to consider how the past, present and future are interconnected. By understanding our history we can gain insights into the importance of technology, in shaping our society.

2. The Impact of Technology on the World Economy

Sabir, [1] shows how historical technological breakthroughs have transformed the world economy. Major changes in how goods are produced, services are provided, and economies operate have been marked by these markers. Some of the key historical technological milestones include:

- **Industrial Revolution:** The paper [2] indicates that a turning point in human history occurred during the late 18th century, with the start of the Industrial Revolution. Characterized by the integration of novel technologies like steam power, mechanized manufacturing, and transportation networks. With these advancements, productivity and economic growth in the manufacturing industry saw significant increases.
- **Electricity and Mass Production:** Adoption spread during the late 19th and early 20th centuries changing industries, thanks to electricity's development. With X, production costs were cut and efficiency increased by creating the capacity to make items speedily. The growth of the economy and consumerism was a direct result of this.
- **Information Technology and the Internet:** According to [3], the late 20th-century advent of computers, IT, and the internet sparked another technological revolution. Through advancements, communication, data processing, and information sharing became vastly improved. E-commerce proliferation and global connection are two key ways the internet has affected business operations.
- **Digitalization and Automation:** Digitalization and automation have grown more prominent in

recent years. Digitization and automation have contributed to enhanced efficiency, cost savings, and higher productivity across various sectors. Significant changes have been seen across different industries, including manufacturing, logistics, finance, and healthcare, thanks to it.

- **Artificial Intelligence and Machine Learning:** Automation and data analysis opportunities have been opened up with AI and machine learning development [3]. From data, AI-based systems can perform complex tasks and make predictions. These technologies hold great promise for revolutionizing various sectors and spurring economic progress.

2.1. Key Statistics on Technology's Influence on GDP

Observing GDP, one can see how technology affects the economy through different data sources [4]. These statistics demonstrate the considerable effect of technology on economic growth and efficiency. Some key statistics include:

- **Research and Development (R&D) Expenditure:** According to [1], a country's investment in technological innovation is reflected in its R&D expenditure. Technological progress and economic growth tend to follow increased spending on R&D. Greater investments in R&D typically lead to higher GDP growth and competitiveness. **Digital Economy Contribution:** The digital economy has seen a significant increase in its contribution to GDP in recent times, according to [5].
- **E-commerce, software development, and digital services** are all part of the broader digital economy. Advancements in information technology and the internet have fueled the growth of the digital economy, leading to it becoming a substantial driver of economic growth in several nations.
- **Productivity Growth:** As per [6], technological developments are a substantial contributor to increased productivity. Output per unit input is a key factor in determining economic growth, thus measured. By automating processes, streamlining workflows, and improving efficiency, technological advancements hold the key to enhanced productivity.
- **Innovation Index:** Paper [1] tracks a country's innovation capacity and technological progress through the innovation index. Taking into consideration R&D expenditure, patent applications, and skilled worker availability, it factors in. Higher levels of economic growth and competitiveness tend to be found in countries with higher innovation indices.

Together, these numbers paint a picture of how technology drives economic growth and development.

2.2. Trends in Job Markets and Automation

Of interest and concern to many, the effects of technology on job markets have been studied [4]. With advancements in automation and AI, traditional job markets face disruption, and work will take on a new form. Some key trends in job markets and automation include:

- **Job Displacement:** Routine and repetitive tasks are particularly susceptible to automation. Tasks that are repetitive and routine are better handled by automation. Job displacement may affect sectors like manufacturing and administration as a result of this. New job opportunities, though, emerge from automation in emerging fields and professions.
- **Skill Shift:** With advancing technology, the skills needed in the job market shift. Data analysis, programming, and digital literacy are in high demand, and as such, there is an increasing need for workers with these skills. Adaptation and skill acquisition are necessary for workers to keep up with industry shifts in the job market.
- **Job Creation:** Automation leads to job displacement in certain sectors but also creates job prospects in developing industries. With technological progress, new industries have been established, including those for data scientists, AI professionals, and cybersecurity stalwarts. Calling for specific knowledge and competencies, these new positions.
- **Augmentation of Human Work:** As per [4], technology might supplement human labor rather than doing away with it. Automation and AI can help workers complete tasks more efficiently and accurately. A result of this could be heightened productivity and job satisfaction. These trends demonstrate the complicated link between tech and labor market dynamics.

2.3. The Global Reach of Technology's Impact

Not limited to specific countries or regions, the impact of technology on the world economy is [7]. Worldwide, it affects economies with global reach. Some key aspects of the global reach of technology's impact include:

- **Global Value Chains:** Technology has enabled the integration of global value chains, simplifying the production process. The result of this has

been increased economic interdependence and trade among nations. Global value chains leverage technology for efficient communication and data sharing, as well as supply chain management.

- **Foreign Direct Investment (FDI):** In the eyes of FDI, technology plays a crucial role [1]. Multi-national corporations are lured by technological advancements and innovation when choosing where to invest. By attracting capital, technology transfer, and job opportunities, FDI fuels economic growth and development.
- **Globalization and Connectivity:** Facilitating global connection, the internet, which is part of technological advancements, has created a more cohesive global community. Collaboration, communication, and information exchange have all become seamless thanks to, and as a result, there has been an growth in international trade and cultural exchange.
- **Technological Diffusion:** Research from [1] shows that technology innovation is global in scope. Potential for global expansion, these entities have. Using technology, developing nations can skip over traditional growth stages and rapidly accelerate economic progress.

Technological diffusion helps drive innovation, productivity, and competitiveness on a global stage. Through historical technological milestones, key statistics on GDP, trends in job markets and automation, and a global reach, technology shows its impact on the world economy [7]. With technological advancements, industries have been shaped, productivity increased, and the nature of work transformed. With technology, communication, trade, and innovation have become more interconnected than ever before. For nations and companies to succeed in today's global economy, harnessing technology's potential is essential.

3. Western Perspective

The emergence of capitalism and the Industrial Revolution helped create Western economic models. Private ownership and profit maximization were central tenets of early industrialization in Western economies. Laissez-faire capitalism, an economic model reliant on market forces and minimum government involvement, drove economic growth through market forces. Over time, a bigger role for the state has been factored into Western economic models. Factors such as the economic crisis of the 1930s and the need for social justice motivated this change. During this time, a model called Keynesian economics gained prominence, which advocated for government spending and fiscal measures to boost demand

and stabilize the economy. Prosperity and growth that had never been seen before blanketed Western economies following World War II. During this time, known as the "Golden Age of Capitalism," a mixed economy model merged elements of capitalism and government involvement. All three areas are critical components of effective governance, etc.

Late 20th-century Western economies saw a comeback of market-based ideas alongside a move towards neoliberalism. Deregulation, privatization, and free trade were hallmarks of neoliberal economics promoted by figures like Ronald Reagan and Margaret Thatcher. With reduced government meddling, an accent on market energies and singular entrepreneurialism developed at this time. New challenges in Western economies have been presenting themselves, along with influences from emerging trends like globalization, technology improvements, and environmental worries. Sustainable and inclusive economic models have gained more attention, leading to increased awareness. Promoting innovation, governments have been focusing on investments in education and skills development while also tackling income inequality.

3.1. Notable Western Technological Innovations

Various disciplines have benefited from the technological superiority of Western societies. Some notable Western technological innovations include:

- **Information Technology:** With computer innovation at their core, internet and digital technologies have revolutionized how we communicate, process data, and share knowledge. Western countries like the United States have taken center stage regarding technology breakthroughs, particularly in Silicon Valley.
- **Biotechnology:** Biotechnological progress was driven by Western states, including improvements in genetic engineering, drug development, and medical instruments. These innovations have revolutionized healthcare, agriculture, and environmental sustainability across the board.
- **Renewable Energy:** From harnessing wind power to embracing solar panels, Western countries continue to drive the shift towards sustainability through energy technologies. Wind, solar, and battery technology breakthroughs have collectively paved the way for a low-carbon and eco-friendly economic transition.
- **Aerospace and Aviation:** Countries within the Western sphere have been at the forefront of both the aerospace and aviation industries. From improved aircraft designs to groundbreaking propulsion systems and space travel, innovations have transformed transportation.

- **Artificial Intelligence (AI) and Machine Learning:** Western powers have spearheaded the progress in AI and machine learning technologies that find their use cases across multiple arenas such as health care, finance, transportation, and manufacturing.

3.2. Economic outcomes from Western inventions

The technology innovations birthed by western countries have brought significant economic ramifications. The following are some of the key significances:

1. **Economic Progression:** Technological inventions have been a pivotal catalyst towards economic progression in western countries. They have made a contribution towards upsurging efficiency levels productivity metrics as well as fostering competitive edge leading therefore to enhanced GDP numbers and overall living standards.
2. **Job creation and metamorphosis:** Technological breakthroughs have resulted in emerging industries on one hand while displacing existing jobs through automation on the other hand however; more jobs get created within these emerging sectors although workers need to acquire new competencies so as to adapt with ever changing job market trends.
3. **Disruption of established industries:** Technology advancements often disrupt traditional industry patterns and business models companies that fail to adjust themselves according to technological shifts will face challenges or even decline whereas those who embrace change can successfully gain a competitive advantage.
4. **Universal Sphere of Influence:** Western groundbreaking advancements have experienced worldwide repercussions, influencing sectors and cash flows internationally. Enterprises hailing from the Western hemisphere, like the technology, pharmaceuticals, and aerospace industries, have taken their place at the forefront of international development by stimulating economic expansion and steering global markets.

3.3. In-depth Analysis & Effects on Western Sectors

One remarkable case study outlining technological innovations' influence on Western industries resides in Silicon Valley in America. This region nestled within the San Francisco Bay Area has evolved into a prominent global centre for technology breakthroughs and entrepreneurial

spirit hosting an abundance of tech companies alongside venture capital firms and research establishments.

Technological advancements sprouting from Silicon Valley have drastically altered various industries. Giants such as Apple, Google, and Facebook have completely reshaped consumer electronics landscape as well as internet search capabilities along with social media platforms. These revolutionary changes not only transformed communication modes but also ushered in novel business models together with lucrative economic prospects.

Silicon Valley's triumph led to the rise of similar tech hubs across other Western nations like London's Tech City based in the United Kingdom and Germany's Berlin-based Silicon Allee. These technology hotspots have drawn talented individuals, funding support, and entrepreneurial endeavours, fuelling economic development and advancement.

Moreover, Western frontiers in the technology space have also made significant waves in sectors such as healthcare, finances, and manufacturing. Notably, progressions in medical technology and pharmaceuticals have led to better health outcomes and extended lifespans. Financial technology or fintech innovations have revolutionized banking systems and financial services industry by rendering transactions more efficient and accessible. In the realm of production, automated systems paired with robotics have boosted productivity rates and efficiency levels thereby cutting costs while elevating competitiveness.

On the whole, Western technological strides have far-reaching financial implications propelling growth in economies as well as catalysing transformations across various sectors thereby leaving an indelible mark on the global economy. They've birthed fresh prospects while shaking up conventional business models thus influencing markets worldwide. Continued emphasis on innovation alongside tech headway will be instrumental for Western countries in retaining their vantage point amidst global economic competition.

4. Eastern Perspective

The introduction of technology in Eastern economies has had a profound influence on these nations' economic progress. Vladimir Putin at SPIEF 2023 highlighted how the perception towards technological strides in the East has evolved drastically and how countries in this geographical area have embraced it as a key stimulant for economic expansion. The Russian President underscored that the use of technology has become an integral element in the national growth policies of numerous Eastern nations.

4.1. Technology's Impact on Eastern Economies

This technological impact on Eastern economies especially in states like Russia and China has been significant and played an influential role in defining their economic trajectory and progress.

Particularly, China has observed rapid technological advances and now stands out as a prolific contributor to scientific progress [8]. The Beijing government acknowledges the dire need to involve technology development to upgrade their industrial sector while boosting competitiveness amidst global counterparts [8]. The emergence of China as a significant contender in the realm of science and technology has been motivated by factors such as its vast populace and human capital foundation, a job market tilted in favour of academic meritocracy, presence of a sizeable diaspora of Chinese-origin researchers, and government backing for scientific endeavours [8].

In Russia, technological progress has also been assigned precedence. The nation has put into action innovation strategies and strived to develop its knowledge creation and transfer processes to stimulate regional advancement [9]. However, the ramifications of technological breakthroughs for the Russian economy have yielded mixed results with challenges in effectively assimilating new technologies. Establishing novel science-based industries and sectors in the economy are perceived as key hurdles faced by Russia [9].

Putin's speech at SPIEF highlighted the profound sway wielded by technology over economies in Eastern regions. He underscored how technology has played an integral role in reshaping economic landscapes across the East. Technological innovations have propelled eastern countries towards modernization while cultivating diversification in their economies. Putin mentioned several instances demonstrating how technology has boosted sectors like energy production, manufacturing operations and finance management within the East thus fostering economic growth and global competitiveness.

4.2. Focus on Russia's Technological Growth

Russia's economic expansion has been driven by its focus on technological growth and innovation. It has established federal research centres and strategic priorities for scientific and technological advancement [9]. These centers play a vital role in implementing innovation policies while developing science in crucial areas of the economy.

Research on the impact of research and development (R&D) and knowledge exchange on economic growth in Russian regions has been conducted [10]. Findings indicate that R&D activity and spending on technological advancements have significant impacts on regional

growth. However, regions lacking innovative abilities may struggle to effectively embrace new technologies [10]. The study also highlights how foreign direct investments (FDI) and imports of goods and services are relevant for regional growth.

This signifies a changing landscape where Western and Eastern technological capabilities are becoming more aligned.

4.3. China's Technological Growth

China has made strides in science and technology establishing itself as a player in this field. The country acknowledges the significance of development in upgrading its industries and enhancing competitiveness in the market [8]. China's emergence as a contributor to science and technology can be attributed to factors such as a population with abundant human capital, an academic meritocracy oriented labor market a considerable number of overseas Chinese scientists and substantial government investments in scientific endeavors [8]. The Chinese government actively promotes advancement as the foundation for industry improvement and increased competitiveness.

China has utilized globalization by becoming an assembly hub, for Asian firms fostering trade of high tech products [11].

China has positioned itself as a hub, for innovation. Has actively established a presence in the digital realm [12]. The growth of the economy has played a role in driving green economic progress, energy conservation and reducing emissions in the manufacturing sector [13].

During his speech Putin acknowledged China's advancements. He recognized China as a player in the global technology landscape highlighting its achievements in areas like intelligence (AI) 5G technology and digital infrastructure. Putin also emphasized the importance of collaboration between Russia and China to push boundaries further creating a synergy that benefits both countries and the wider region.

4.4. Impact on Specific Industries

The impact of technology on economies has had effects on specific sectors such as energy and manufacturing. In China the digital economy is seen as a driver, for achieving carbon peak targets, carbon neutrality goals and promoting quality economic development [13].

The manufacturing industry plays a role, in achieving energy consumption and carbon emissions goals. It has been discovered that the development of the economy has an impact on the green and low carbon transformation of the manufacturing industry.

In Russia researchers have examined how technology affects industries. For instance they have analysed how

the digital economy influences the green and low carbon transformation of manufacturing [13]. The study revealed that the digital economy has an effect on driving manufacturing transformation in Russia's central region. Moreover researchers have investigated the integration of circular economy principles industry 4.0 and lean manufacturing in terms of their impact on sustainability performance among manufacturing firms [14]. The study emphasized the influence of circular economy principles. Highlighted how industry 4.0 and lean manufacturing play mediating roles, in achieving sustainability objectives [14].

During SPIEF 2023 Putin discussed how technology has transformed sectors within a context. He specifically mentioned energy and manufacturing as industries that have been impacted.

Technological advancements, in the energy industry have greatly improved the extraction and management of resources leading to enhanced energy security and sustainability. Similarly in the manufacturing sector automation and digitalization have significantly increased productivity and competitiveness.

5. Comparing Western and Eastern Perspectives

Take Putin's speech at SPIEF 2023 for instance. It offers insights into comparing Eastern perspectives on technology and its impact on economies. While the West has traditionally been a frontrunner in innovation Putin emphasized that Eastern nations like Russia and China are rapidly closing the gap.

This signifies a changing landscape where Western and Eastern technological capabilities are becoming more aligned.

5.1. Highlighting Notable Contrasts

One notable distinction between Eastern perspectives lies in their approach to innovation. In economies Silicon Valley there's a reputation for nurturing a culture of innovation and entrepreneurship. This means being unafraid to take risks focusing on technologies and embracing experimentation even if it entails failure. Conversely Eastern economies such as China and Russia have historically taken an approach to innovation with greater emphasis placed on government led initiatives and strategic planning. These differing approaches to innovation can be attributed to historical and political factors.

Another significant difference lies in the level of advancements and adoption. Western economies, the United States have consistently been at the forefront of progress, with an extensive history of innovative breakthroughs.

This has led to an adoption and integration of technology, across sectors of the economy. In contrast Eastern economies although catching up rapidly have historically been slower in embracing advancements. However countries like China have recently made progress. Emerged as key players in science and technology.

Putin's speech highlights the differences in how innovation's approached between the West and the East. He pointed out that while Western innovation tends to focus on entrepreneurship and startups the Eastern approach often involves government led initiatives and collaborations among government, academia and industry. Putin emphasized that both approaches have their strengths and can learn from each other.

5.2. Finding Common Ground

Despite their disparities Western and Eastern perspectives on technology also share some similarities. One notable similarity is the recognition of technologies importance in driving growth and development. Both Western and Eastern economies understand that technological advancements are critical for boosting productivity, competitiveness and overall economic performance.

Another common aspect is the growing emphasis on digitalization and the digital economy. Both Western and Eastern economies acknowledge the power of technologies and have taken steps to promote digitalization across various sectors. This includes adopting technologies such, as the Internet of Things (IoT) artificial intelligence (AI) and big data analytics.

Furthermore during Putin's speech, at SPIEF 2023 he highlighted the similarities in technology adoption trends between Eastern economies. He emphasized that both sides of the world recognize the significance of technology as a catalyst for growth and development. This shared understanding creates opportunities for collaboration.

5.3. Collaboration Opportunities and Areas of Competition

One specific area where collaboration can take place between Eastern economies is through global tech partnerships. Both sides acknowledge the advantages of working and sharing knowledge to drive advancements and innovation [15]. This includes partnering on research and development projects establishing ventures and facilitating technology transfers. By leveraging each economies strengths and expertise these collaborations can promote growth and development. In his SPIEF address Putin underscored the potential for efforts between Eastern countries through global tech partnerships. He emphasized the importance of cooperation in addressing challenges

such as cybersecurity, AI ethics and climate change. According to Putins speech cooperation in these areas can yield outcomes.

Another aspect where competition arises between Eastern economies is in their pursuit of dominance, in the tech market.

In the realm of the tech industry Western economies, the United States have traditionally held a leading position. Tech giants, like Apple, Google and Microsoft have dominated the market. However Eastern economies, China have been swiftly catching up. Now boast their own tech powerhouses such as Alibaba, Tencent and Huawei. This competition for control over the market is fuelled by factors like advancements in technology intellectual property rights protection and access to markets. During discussions on Areas of Competition Putin acknowledged that the race for dominance in the tech market is a reality on a scale. He highlighted that both Western and Eastern tech companies are vying for leadership roles. Putins insights imply that healthy competition can drive innovation; however they also raise questions concerning regulations, fair practices and equal opportunities, for market access.

6. Technological Aspects of influence of technology on the world economy and AI's Role in Shaping the Economy

Artificial intelligence (AI) has become a game changing technology that carries implications, for the economy. AI refers to the development of computer systems of performing tasks that traditionally require intelligence, such as recognizing speech making decisions and solving problems. Integrating AI across sectors of the economy holds potential for driving productivity fostering innovation and fueling economic growth [16].

AI possesses the capacity to automate repetitive tasks liberating workers to devote their time and energy to more intricate and imaginative pursuits. This can result in increased efficiency and productivity within industries like manufacturing, logistics and customer service. For instance AI powered robots and automated systems can streamline production processes while minimizing errors to enhance efficiency [16].

Moreover AI has the power to revolutionize decision making procedures by analysing quantities of data and offering insights. This can greatly improve planning, risk management and resource allocation across sectors such as finance, healthcare and marketing. By examining patterns and trends in data using AI algorithms businesses can make decisions driven by data driven predictions and recommendations [16].

The impact of AI, on the job market remains a subject of debate. While AI has the potential to automate job tasks it also opens up opportunities, for employment. The development and implementation of AI technologies require individuals in fields like data science, machine learning and AI programming. As a result there is an increasing demand for professionals with expertise in AI related areas [16].

AI is leading the way in transforming the economy. Its ability to process amounts of data identify patterns. Make predictions has revolutionized various industries. AI powered systems enhance efficiency reduce costs and enable businesses to make decisions based on data. From maintenance in manufacturing to recommendations in e commerce AI is shaping how businesses operate. Additionally its role extends to services where it optimizes trading strategies and detects activities. There's no denying its significance as it drives growth and fosters innovation across sectors.

6.1. Regulating AI and Its Impact on the Economy

Regulating AI technologies plays a role in ensuring their ethical deployment. The aim of regulations is to address concerns related to privacy, security, bias, accountability and transparency. The economic impacts of regulating AI can be both positive and negative.

This can help build trust and confidence, in AI systems among the public, which is crucial for their widespread acceptance. Clear regulations can also encourage competition. Prevent the misuse of AI technologies, such as unauthorized use of personal information or the development of AI powered weapons [17].

However it's important to strike a balance with regulations that are not too excessive or overly restrictive. We need to foster innovation and allow for the development and implementation of AI technologies. Finding the equilibrium between regulation and innovation is vital to harnessing the potential of AI for economic growth. Policymakers should consider the nature of AI technologies. Adopt flexible regulatory frameworks that can adapt to rapid advancements in this field [17].

Nevertheless we must acknowledge that great power comes with responsibility. Regulating AI is crucial to ensure its integration into society and economy. Implementing data protection laws to the European Unions General Data Protection Regulation (GDPR) plays a pivotal role in striking a balance between safeguarding user rights and enabling innovation. These regulations provide a framework for data collection and usage practices. By addressing privacy concerns and ensuring transparency in decision making processes powered by AI governments and businesses can establish trust among users. A factor,

in driving widespread adoption of AI technologies that ultimately benefit our economy.

6.2. Microchips and their Impact, on Advancements

Microchips also referred to as integrated circuits play a role in driving the progress of modern technology. These tiny electronic devices consist of millions or even billions of transistors which are responsible for processing and storing information.

The influence of microchips can be seen across industries such as computing, telecommunications, healthcare, transportation and entertainment. They have revolutionized devices like smartphones, laptops and gaming consoles by making them smaller, faster and more powerful. Additionally the miniaturization of microchips has given rise to wearable gadgets that have transformed the way we interact with technology.

Thanks to advancements in microchip technology described by Moore's Law there has been a surge in computing power and storage capacity. This progress has paved the way for the development of AI algorithms and machine learning models that rely on resources. Microchips have made it possible to process amounts of data and train AI models. Consequently this has contributed to breakthroughs in natural language processing, computer vision and autonomous systems [17].

The semiconductor industry responsible for manufacturing microchips has become a driver, for growth and innovation. It not fuels advancements but also fosters the creation of new applications and industries.

The demand, for microchips is continually rising due to the emergence of technologies like AI, Internet of Things (IoT) and 5G networks [17]. Microchips, often overlooked yet vital to progress have played a role in shaping the digital era. These small but powerful components are the backbone of devices ranging from smartphones to supercomputers.

Their continuous advancement has resulted in improvements in computing power and data processing capabilities. The intricate relationship between AI and microchip technologies has paved the way for AI algorithms and real time processing unlocking opportunities for businesses and economies. Additionally intense competition among nations to develop microchips has sparked a race for sovereignty.

6.3. Interplay between AI and Microchip Technologies

The interplay between AI and microchip technologies is symbiotic and mutually reinforcing. AI heavily relies on the power and storage capacity provided by microchips to process and analyse amounts of data. Simultaneously

advancements in AI drive the demand for microchips capable of supporting increasingly complex algorithms and applications.

The development of microchips, like graphics processing units (GPUs) and tensor processing units (TPUs) has significantly accelerated AI computations. These chips have been specifically designed to enhance the performance of AI tasks resulting in training and inference times, for AI models. By integrating hardware accelerators for AI into microchips significant improvements have been made in the efficiency and speed of AI calculations [17].

Moreover advancements in microchip technology such as the creation of chips and quantum computing hold potential for further enhancing AI capabilities. Neuromorphic chips aim to emulate the structure and functionality of the brain enabling the development of effective and energy efficient AI systems. On the hand quantum computing possesses the power to revolutionize AI by tackling optimization problems and exponentially increasing computational capacity [17].

AI and microchip technologies are deeply. Profoundly impact economies. The role of AI in shaping economies is evident as it has the potential to drive productivity, foster innovation and stimulate growth. It is crucial to regulate AIs deployment ethically. Microchips serve as a component of technology; they have played a pivotal role in technological advancements while facilitating the development of robust AI systems. The relationship between AI and microchip technologies is symbiotic; a growing need, for powerful microchips is propelled by the demands placed on them by advancing AI applications [17].

The interaction, between AI and microchip technologies is driving a collaboration that goes beyond the boundaries of hardware and software. AI relies on microchips for its abilities while microchip technologies benefit from AI driven improvements in manufacturing and design processes. This beneficial relationship is bringing about changes in industries like healthcare, where AI powered diagnostics utilize the processing capabilities of microchips to provide rapid and accurate results. Likewise in vehicles, AI and microchips work together to make real time decisions enhancing safety and efficiency. This interplay not contributes to growth but also raises concerns about regulating and ethically using these technologies. Striking a balance between innovation, accountability and economic impact in this evolving landscape will be vital for harnessing the potential of AI and microchip technologies, in the economy.

7. Geopolitical Aspect

The geopolitical implications of advancements, in intelligence (AI) are significant. AI technologies have the potential to reshape power dynamics between nations and impact competition. When a country develops and deploys AI it can enhance its competitiveness, military capabilities and technological leadership which in turn affects its position [18].

Leading countries in AI research and development gain an edge across sectors like healthcare, finance and defence. The ability to effectively utilize AI drives growth attracts investments and strengthens a countries position in the global economy. Consequently there is competition among nations to establish themselves as leaders in AI technology leading to increased rivalry and potential geopolitical tensions [18].

The rapid progress of AI technology has reaching consequences. As countries vie for leadership in AI it becomes evident that those with cutting edge capabilities hold influence on the stage. Innovations driven by AI in areas like defense, cybersecurity and intelligence have the potential to reshape the landscape. Nations are investing heavily in applications of AI such, as drones and cyber warfare.

This competition raises concerns, about the risks associated with an arms race in the field of AI highlighting the need for international agreements to ensure ethical use of AI.

The interaction between AI and microchips also holds implications for relations. The development and production of microchips, which serve as components in AI systems can become an asset for countries. Control over microchip manufacturing and supply chains can greatly influence a nations capabilities and economic competitiveness [18].

Relying on suppliers for microchips can create vulnerabilities and dependencies within a countries infrastructure. This raises concerns about security prompting nations to prioritize the development of microchip manufacturing capabilities. The competition to establish dominance in the microchip industry carries implications as countries strive to secure their supply chains and reduce reliance on nations [18].

International relations are increasingly shaped by a nations prowess with AI and microchips playing roles. The ability to produce microchips and harness cutting edge AI technologies is now considered a measure of a nations sovereignty. Consequently complex dynamics arise as countries seek partnerships for technology development while simultaneously navigating dependencies, within supply chains.

Furthermore the exportation and regulation of AI and microchip technologies play a role, in trade negotiations and diplomatic relationships. The competition for dom-

inance holds implications for alliances and geopolitical alignments.

7.1. Impact on Global Security Dynamics

Advancements in AI also have effects on global security dynamics. The integration of AI technologies into systems can bolster a nation's defense capabilities. Reshape the nature of warfare. For instance autonomous weapons systems powered by AI have the potential to revolutionize operations and decision making processes [18].

The development and deployment of AI in the sector raise concerns regarding arms races unintended consequences and the necessity for regulations and norms. The pursuit of AI supremacy can result in tensions well as strategic rivalries among nations. The competition to develop technologies driven by AI has reaching implications for global security dynamics prompting discussions about power balances [18].

The impact of AI and microchip technologies, on security dynamics is profound. It is essential to have AI powered cybersecurity systems to defend against cyber threats while safeguarding infrastructure. However these advancements also introduce vulnerabilities since malicious actors can exploit AI capabilities to launch cyberattacks.

Furthermore the integration of AI, into applications, such as weaponry has the potential to disrupt customary concepts of warfare and deterrence. The global security landscape is undergoing a transformation as nations invest in AI for defense and surveillance objectives. As a result discussions surrounding arms control treaties and international norms regarding the utilization of AI in security contexts are gaining traction.

7.2. Geopolitical Considerations for the Future

As AI progresses further geopolitical considerations will gain increasing significance. Countries will need to navigate the terrain to safeguard their competitiveness, national security and technological supremacy. This entails investing in AI research and development fostering ecosystems to innovation and establishing frameworks that strike a balance between innovation and ethical deployment of AI [18].

International cooperation and collaboration will also play a role in addressing the implications of AI. Countries must work together to establish norms, standards and regulations that foster ethical use of AI technologies. This involves addressing concerns related to privacy, security, bias mitigation and accountability [18].

The advancements in microchip technology alongside developments, in AI have profound ramifications. They

have the capacity to reshape power dynamics influence international relations dynamics significantly while impacting security overall. Countries must carefully consider the implications and geopolitical factors when developing their AI strategies and policies to ensure their competitiveness, national security and technological leadership.

Looking ahead the influence of AI and microchips, on relations will continue to shape dynamics. The race for supremacy will intensify, raising questions about data ownership, intellectual property rights and the governance of emerging technologies. Nations will have to find a balance between collaboration and protectionism fostering cooperation on AI standards while safeguarding their own interests. Additionally ethical concerns regarding the use of AI in warfare and surveillance will require attention. Successfully navigating these complexities is crucial for maintaining stability in the face of advancements.

In conclusion our exploration into the impact of technology on the world economy has revealed a network of interconnected factors. We have delved into the role played by AI recognized the importance of microchips and acknowledged the geopolitical implications associated with these advancements. Understanding these dynamics is vital as we navigate a world where technology serves as both an engine for growth and a catalyst, for change.

We have witnessed how technology goes beyond being a tool and becomes a force that shapes economies influences relations and raises important ethical concerns. From looking at the past to examining the scenario we have observed the march of innovation and its consequence.

8. Conclusion

In conclusion our exploration of the impact of technology, on the economy and its role in fostering growth has unveiled a complex network of interconnections. We have delved into the power of AI, the importance of microchips and the geopolitical implications that these advancements bring about. It is crucial to comprehend these dynamics as we navigate a world where technology serves as both an engine for progress and a catalyst for change. Our observations have demonstrated that technology is not a tool but an influential force that shapes economies impacts international relations raises pivotal ethical concerns and holds the potential to bridge societal gaps.

Looking ahead to what lies in store for us we can anticipate trends. Artificial intelligence will continue its evolution making an impact across industries such as healthcare and finance while potentially redefining the

nature of work itself. Microchip technologies will witness advancements, unlocking possibilities in fields like AI and IoT and beyond. Quantum computing, neuromorphic chips and other groundbreaking innovations will reshape our perception of what's achievable. Moreover it is essential to consider the dimensions associated with these trends; from ensuring deployment of AI to addressing worries regarding data privacy and security. Moreover it is crucial to prioritize an approach as we navigate the future. International cooperation plays a role, in establishing standards and regulations that protect the community.

The consequences of these trends for the world economy are significant. As artificial intelligence and microchips become more integrated into industries we can anticipate intensified competition and fresh economic opportunities. Nonetheless we must also address disruptions to employment. Carefully consider how to retrain our workforce. Additionally technological capabilities will shape the landscape. Nations at the forefront of AI and microchip innovation will wield influence on the stage potentially altering power dynamics. Striking a balance between competition and effective cooperation is essential to ensure stability and prevent conflicts. In conclusion I strongly encourage all of us to continue exploring these themes. The world of technology is constantly evolving, with implications, for our economies, societies and pursuit of inclusivity. Let us foster an attitude of curiosity and collaboration by supporting research efforts and initiatives that promote advancement.

As we move forward into the future equipped with knowledge and a comprehension of technologies impact, on our society we have the opportunity to shape a tomorrow that utilizes innovation to improve the lives of all humanity. I sincerely appreciate your support. May our shared exploration of technology lead us to discoveries, advancements and widespread prosperity.

References

- [1] S. Sabir, A. Rafique, K. Abbas, Institutions and fdi: evidence from developed and developing countries, *Financial Innovation* 5 (2019) 1–20.
- [2] X. Niu, A. Moussawi, G. Korniss, B. K. Szymanski, Evolution of threats in the global risk network, *Applied Network Science* 3 (2018) 1–24.
- [3] A. Annor-Antwi, A. A. M. Al-Dherasi, Application of artificial intelligence in forecasting: A systematic review., *American Journal of Computer Sciences and Applications* 2 (2019).
- [4] A. Mashal, E. Ahmad, L. Nasrawi, A. Ghazalat, The impact of external funding flow on Jordan's GDP (1997-2017), *International Journal of Research in Business and Social Science* (2147-4478) 10 (2021) 328–337.
- [5] M. Zeeshan, J. Han, A. Rehman, I. Ullah, F. E. A. Afridi, Z. Fareed, Comparative analysis of trade liberalization, CO₂ emissions, energy consumption and economic growth in Southeast Asian and Latin American regions: a structural equation modeling approach, *Frontiers in Environmental Science* 10 (2022) 79.
- [6] B. Mayor, Unraveling the historical economies of scale and learning effects for desalination technologies, *Water Resources Research* 56 (2020) e2019WR025841.
- [7] M. Savićević, P. Veselinović, N. Makojević, The effects of globalization on the international competitiveness of the Western Balkan countries, *Economic Themes* 60 (2022) 459–480.
- [8] Y. Xie, C. Zhang, Q. Lai, China's rise as a major contributor to science and technology, *Proceedings of the National Academy of Sciences* 111 (2014) 9437–9442.
- [9] A. Samarin, The role of federal research centers in implementing the strategic priorities of scientific and technological development of Russia, *KnE Social Sciences* (2022) 132–141.
- [10] M. Kaneva, G. Untura, The impact of R&D and knowledge spillovers on the economic growth of Russian regions, *Growth and Change* 50 (2019) 301–334.
- [11] G. Gaulier, F. Lemoine, D. Ünal-Kesenci, China's integration in East Asia: Production sharing, FDI & high-tech trade, *Economic Change and Restructuring* 40 (2007) 27–63.
- [12] A. F. Aysan, B. Sadriu, H. Topuz, Blockchain futures in cryptocurrencies, trade and finance: a preliminary assessment, *Buletin Ekonomi Moneter Dan Perbankan* 23 (2020) 525–542.
- [13] W. Zhang, H. Zhou, J. Chen, Z. Fan, An empirical analysis of the impact of digital economy on manufacturing green and low-carbon transformation under the dual-carbon background in China, *International Journal of Environmental Research and Public Health* 19 (2022) 13192.
- [14] A. M. Ghaithan, Y. Alshammakhi, A. Mohammed, K. M. Mazher, Integrated impact of circular economy, industry 4.0, and lean manufacturing on sustainability performance of manufacturing firms, *International Journal of Environmental Research and Public Health* 20 (2023) 5119.
- [15] J. Rezaei, R. Ortt, P. Trott, How SMEs can benefit from supply chain partnerships, *International Journal of Production Research* 53 (2015) 1527–1543.
- [16] P. Vozniuk, The impact of artificial intelligence on the global economy, <https://doi.org/10.18411/lj-03-2019-46>, 2019.

- [17] D. Corcos, Food–nonfood discrimination in ancestral vertebrates: Gamete cannibalism and the origin of the adaptive immune system, *Scandinavian Journal of Immunology* 82 (2015) 409–417.
- [18] M. Hasan, M. A. Naeem, M. Arif, S. J. H. Shahzad, S. M. Nor, Geopolitical risk and tourism stocks of emerging economies, *Sustainability* 12 (2020) 9261.

Custom-made Approaches to Cloud Container Security: A Methodologically Sound Approach based on International Sample Results

Aleksandar Jovanović^{1,*}, Petar Milić²

¹Faculty of Information Technology, Belgrade Metropolitan University, Tadeuša Košćuška 63, 11000 Belgrade, Serbia

²Faculty of Technical Sciences, Department of Computer Science and Informatics, University of Priština - Kosovska Mitrovica, Knjaza Miloša 7, 38220 Kosovska Mitrovica, Serbia

Abstract

Previous available research focused towards examining a sample of programmers' responses regarding cloud container security. In this paper, we performed a comparison of differences between the analyzed sample in this paper and the previous one that we published and which comprised a Serbian-based sample. The research aim of the following paper is to form an analysis of answers in terms of its validity for the overall programmer community. It was determined that there is a slight difference between two samples; however there are not any country-specific types of cloud security issues, based on the sample. There is a tendency of completely globalizing or even neglecting the interviewee's background in discussing results in cloud security literature. And this very aspect is indicative for custom made approaches to security issues, according to parameters such as confidentiality, availability, usability, non-repudiation and integrity, all mentioned and discussed in the paper. The results helped us focus on the creation of the overall recommendations for creating methodology for cloud containers' assessment in terms of security that would be more globally applicable. In particular, the next step is to allow for mathematical models to be applied in the survey results interpretation and also to implement a fuzzy logic mathematical model to create a layer of protection in CCSA.

Keywords

cloud containers, Security, questionnaires

1. Introduction

Cloud environments are susceptible to malicious attacks, like any other form of IT-based architecture. The more data installed and infrastructure facilities installed on the cloud, the more chance it would be attacked at certain point. There is an ongoing tendency to "shift left" when dealing with cloud container security [1]. This means, that as early as possible in the development process of an app lifecycle malicious ruptures should be examined and that this process saves a lot of later costs and time when dealing with securing cloud-based apps throughout their lifecycles.

2. Methodology

Analysis was performed by using a questionnaire, which involved 50 members of the IT sector, that have been using cloud technologies and that pointed towards dif-

ferent issues and aspects of creating a methodology for CCSA (Cloud Container Security Assessment). The questionnaire had some multiple choice answers and some questions had possible answers given in the form of statements and views towards certain cloud container-based themes and problems. The questionnaire was conducted over a twelve-month period. The sample was diverse in terms of background with software use among interviewees but focused on their experience according to that which they have been applying mostly in their daily work. One criterion for further doing the questionnaire is that all of them should have software development experience for at least one year and are at least familiar with Cloud.

Most of the sample included individuals with more than ten years of experience with IT and software development, both in private or public departments and academia. The previous sample conducted in 2022 [2] included 90% of individuals residing in Serbia, involved either at Serbian companies or international software companies there. The current sample in this paper made a more international sample, according to which at least 35% of the interviewees were of EU or US space.

BISEC'23: 14th International Conference on Business Information Security, November 24, 2023, Niš, Serbia

*Corresponding author.

✉ aleksandar.jovanovic@metropolitan.ac.rs (A. Jovanović);
petar.milic@pr.ac.rs (P. Milić)

🆔 0000-0002-9815-4344 (A. Jovanović); 0000-0003-0427-8379
(P. Milić)

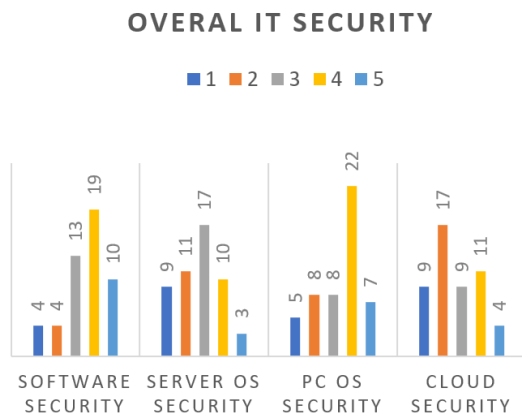


Figure 1: Overall IT security.

3. Previous research

In the previous study we have conducted in 2022 [2], the results of the paper indicated that commercially-driven advancements remain the real drive for creating theoretical models in the novel scientific field of CCSA. But it also pointed to the fact that there is not an overall consensus on creating a common methodology with binding applications or at least emphasizing common security frameworks for assessing security in the cloud. The survey results analyzed in the mentioned paper indicated security patterns for building secure systems. It also pointed to the aspect of previous inspection on cloud container images before the utilization as significant for creating cloud container security methodologies.

The above mentioned study concluded with the comparison of the results of questionnaires used with the literature references and case studies, and indicated a real cloud threat incident analysis as necessary in order to get more specific results on cloud environments' particularities in the future and to be able to further advance towards creating any kind of CCSA methodology.

4. Results of our current study

Based on the analysis we have conducted here, it can be noted that attention has been paid to the security in cloud environment, and this is depicted on Figure 1. On the scale from 1 - little to none till 5 - highly experienced, we can notice that respondents in our survey showed adequate experience in working with cloud security issues which can lead to proper usage and configuration of cloud services. Thus, security issues will be properly resolved. Also, other general types of security are well represented and respondents are aware of their importance.

IMPORTANCE OF DIFFERENT ASPECTS OF CLOUD SECURITY

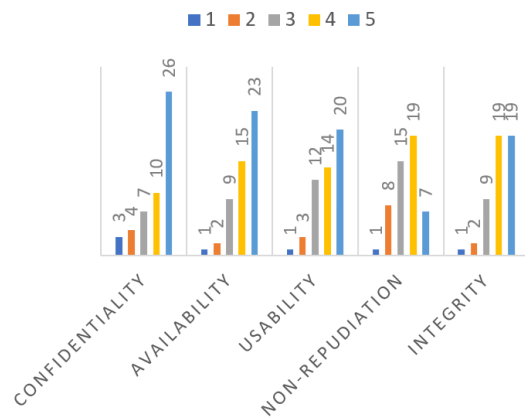


Figure 2: Importance of different aspects of cloud security.

In accordance with previous statements, collected responses goes towards confidentiality and availability aspects of cloud security, which is among highest, as depicted on Figure 2. Nowadays, confidentiality represents a fundamental aspect of cloud security, ensuring that sensitive data remains private and protected from unauthorized access, whether it's at rest or in transit [3, 4]. Similarly, availability is equally critical in cloud security, ensuring that cloud services and resources are accessible and reliable for users, minimizing downtime and disruptions to business operations. Keeping in mind the diversity of services that are available through cloud services, it becomes clear why these aspects are major factor motivating the proliferation of security issues.

The presence of security vulnerabilities within a cloud environment can result in the inadvertent exposure of information regarding the services hosted therein [5, 6]. This risk becomes particularly pronounced when background containers are executed on a single host, sharing the same operating system (OS) kernel, as a compromised kernel can compromise the isolation provided by the container mechanism. In accordance with this assertion, our analysis, as depicted in Figure 3, substantiates these findings. Additionally, the data presented in Figure 4 underscores the critical importance of thoroughly inspecting cloud container images before their deployment and use.

Hence, security concerns emerge as the primary impediment to the continued adoption of containers and cloud computing as a whole. When data and services are outsourced to the cloud environment, they become susceptible to various risks, with security being a paramount concern that necessitates a meticulous implementation

TOP BREACHES CAUSES IN CLOUD ENVIRONMENT

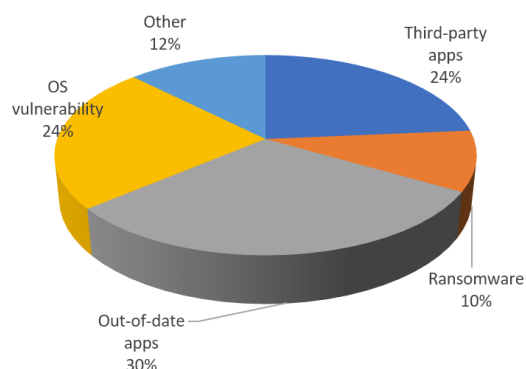


Figure 3: Top breaches causes in cloud environments.

IMPORTANCE OF CLOUD CONTAINER IMAGES SCANNING BEFORE USAGE

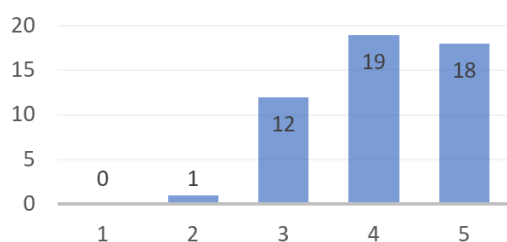


Figure 4: Importance of cloud container images scanning before usage.

strategy. Employing security patterns for constructing secure systems involves outlining methods to mitigate specific threats, remediate vulnerabilities, and establish a secure environment conducive to the effective utilization of cloud services. These patterns offer standardized solutions and best practices for countering common threats and vulnerabilities, thus enhancing the overall security posture of systems and applications. By implementing security patterns, organizations can streamline the development of secure software and systems in cloud environment, reduce the likelihood of security breaches, and fortify their defenses against evolving cyber threats.

5. Comparison with previous research

In comparison with the results from our study published on BISEC 2022 regarding the overall IT security, current results confirm dedication of IT professionals to the general software security. Also, awareness about server security and PC OS security gained more attention, which is in line with survey respondent’s orientation toward proper configuration and usage of environment for usage of cloud services. Nevertheless, respondents slightly increased their experience with cloud security issues indicating thus that this aspect is important.

Furthermore, significance of different aspects of cloud security such as confidentiality, availability, usability, non-repudiation and integrity is increased in comparison with the study from BISEC 2022, showing that they are essential in designing a robust cloud security strategy that protects data and services while ensuring they remain accessible and usable for authorized users.

When we come to the top breaches causes in cloud environment, it can be noted that still high percentage is about out-of-date apps. This lead us to conclusion about high significance of regular update of all parts of the information system, as of OS software, app libraries and etc. in order to maintain as much as possible high level of security. Other breaches have balanced values in comparison with results from BISEC 2022. Similarly, the critical importance of thoroughly inspection of the cloud container images before their deployment and use is confirmed.

6. Similar studies of other authors

Seongmo et al. [7] suggested a CloudSafe platform, whereas the authors pointed towards a necessity for testing on a actual cloud system. The study focused on Amazon’s AWS, but had implications for other Cloud providers such as Azure too. Gudapati and Gaikwad [8] created common cloud security issues’ guidelines. Nitiashree et al. [9] proposed a three-stage cloud computing data security model. A unique or coherent methodology for cloud containers security assessment is not available and the ones suggested by companies are less usable for the current attacks that are diverse in nature. However, the last mentioned study [9] went ahead to create an Advanced Encryption Standard (AES) algorithm for Data security. The last layer of this algorithm model involved cryptography techniques. Furthermore, the study indicated a relation between occurrence of public cloud threats and data security during the transmission from Cloud Service Customer (CSC) to the Cloud Service Provider (CSP) [10]. This is relevant to understand the direction in which future research should be focusing.

What we noticed from cloud security literature analysis is that there is a tendency of completely globalizing or even neglecting the interviewee's background in discussing results. And this very aspect is indicative for custom made approaches to security issues, in terms of aspects analyzed and described.

7. Conclusion

A thorough inspection of cloud container images is necessary and confidentiality, availability, usability and non-repudiation along with integrity become more significant for cloud environments security strategy which is robust. The next step is the study would be to allow for mathematical models to be applied in the survey results interpretation and also to implement a fuzzy logic mathematical model to create a layer of protection which could solve some if not majority of issues mentioned in this paper and this save time and effort in dealing with cloud security.

Acknowledgment

We would like to thank the programmers and engineers for filling-out the survey on cloud security and for their interest in our research.

References

- [1] D. Gonzalez, P. P. Perez, M. Mirakhorli, Barriers to shift-left security: The unique pain points of writing automated tests involving security controls, in: Proceedings of the 15th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM), 2021, pp. 1–12.
- [2] A. Jovanović, P. Milić, V. Saraswathi, Towards creating methodology for security assessment of cloud containers- an overview of available tools, in: BISEC'22: 13th International Conference on Business Security, 2022, pp. 50–53.
- [3] A. Tchernykh, U. Schwiegelsohn, E.-g. Talbi, M. Babenko, Towards understanding uncertainty in cloud computing with risks of confidentiality, integrity, and availability, *Journal of Computational Science* 36 (2019) 100581.
- [4] P. Yang, N. Xiong, J. Ren, Data security and privacy protection for cloud storage: A survey, *IEEE Access* 8 (2020) 131723–131740.
- [5] H. Tabrizchi, M. Kuchaki Rafsanjani, A survey on security challenges in cloud computing: issues, threats, and solutions, *The journal of supercomputing* 76 (2020) 9493–9532.
- [6] M. Jouini, L. B. A. Rabai, A security framework for secure cloud computing environments, in: *Cloud security: Concepts, methodologies, tools, and applications*, IGI Global, 2019, pp. 249–263.
- [7] S. An, A. Leung, J. B. Hong, T. Eom, J. S. Park, Toward automated security analysis and enforcement for cloud computing using graphical models for security, *IEEE Access* 10 (2022) 75117–75134.
- [8] G. S. Prasad, V. S. Gaikwad, A survey on user awareness of cloud security, *International Journal of Engineering & Technology* 7 (2018) 131–135.
- [9] B. Nithiasree, B. R. Prakash, R. S. Sundar, A survey on cloud security threats and solution for secure data in data stages, *2021 International Journal of Computer Techniques (IJCT)* 8 (2021).
- [10] M. Toy, Cloud services architectures, *Procedia Computer Science* 61 (2015) 213–220.

NBS Web Service Use in a Business Environment

Nikola Sretenović¹, Dejan Nemeč^{1,*}

¹Faculty of Technical Sciences, University of Novi Sad, Trg Dositeja Obradovića 6, 21000 Novi Sad, Serbia

Abstract

In today's business environment, access to accurate and up-to-date financial information plays a pivotal role in the successful management of organizations. Businesses, regardless of their size and industry, rely on various data sources to make informed decisions and efficiently manage their financial resources. In this context, web services have become an indispensable tool that facilitates access to various types of financial information.

The NBS (National Bank of Serbia), as the central bank of the Republic of Serbia, plays a crucial role in maintaining financial system stability and supporting the country's economic development. As part of its functions, the NBS provides web services that enable organizations to access a wide range of financial data. These web services offer information on exchange rates, banking reports, payment systems, and other relevant aspects of financial operations.

The significance of this project is multifaceted. First, the research will provide a deeper understanding of NBS web services and their potential in the business environment. Second, it will identify a specific example of how organizations can use NBS web services in a particular business context and explore the benefits they gain through their utilization. Third, it will pinpoint potential obstacles or challenges that organizations may encounter when implementing NBS web services.

In conclusion, the project's results are expected to provide practical guidelines and recommendations for organizations that are planning to or already utilizing NBS web services.

Keywords

NBS, Web services, Exchange rates

1. Introduction

web service is a software system that facilitates interoperable M2M (Machine-to-Machine) interaction over a computer network by performing specific tasks. It can be viewed as a service encompassing all the necessary details for interacting with it, including message formats, transport protocols, and location. The interface abstracts the service's implementation details, allowing it to be used independently of the hardware or software platform it's implemented on and the programming language it's written in. This enables web service-based applications to be loosely coupled, component-oriented, and cross-technology implemented. Web services can be used standalone or in conjunction with other web services to perform complex aggregation or transactions [1, 2].

In practical terms, web services in collaboration with GUI (Graphic User Interface) applications fall somewhere between web and desktop applications. A web service provides functionality and data to the server, while a desktop application offers a customizable graphical interface filled with data received from the service. Additionally, a web service in collaboration with a GUI application provides a more flexible system compared to a web application because the user can customize the

application's appearance, whereas web application users are constrained by web browsers such as Google Chrome, Mozilla Firefox, Brave, Microsoft Edge, and others. Therefore, web application users have no influence over how the application appears on their screen [3].

Communication in web services occurs between the client and server. The client sends a specific request to the server, which then accepts and processes the request. Subsequently, the server generates a recognizable response and sends it back to the client. The client-server architecture separates the concerns of both sides of the communication channel. This means that in this context, the client side is not concerned with how information is stored on the server since there is always a uniform way to access these resources. On the other hand, the server is independent of the client and is not interested in the implementation of the user interface or the state of individual clients. This simplifies the server-side significantly. Moreover, complete independence and easier separate development of both system components are achieved. For example, it is possible to enhance or modify the server's logic without the client being aware of any changes as long as it accesses the target resources in the same way [3].

Figure 1 illustrates previously described client-server architecture as the communication type that is being used when communication is related to web services.

1.1. Formats and protocols of web service

web service is a software system identified using a URI (Uniform Resource Identifier), whose public interfaces

BISEC'23: 14th International Conference on Business Information Security, November 24, 2023, Niš, Serbia

*Corresponding author.

✉ nikolasretenovicdzo@gmail.com (N. Sretenović);
denem@uns.ac.rs (D. Nemeč)

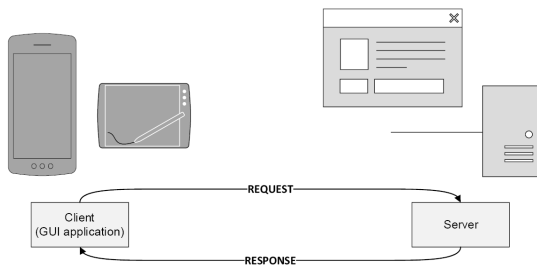


Figure 1: Client-server architecture.

and bindings are defined and described using XML. Its definition can be discovered by other software systems. These systems can then communicate with the web service as prescribed by its definition, using XML-based messages transmitted over internet protocols [4].

The architecture of web services is based on the interaction of three roles: the service provider, service registry, and service requester. The service provider defines the service description for the web service and publishes it in the service registry. The requester uses a discovery operation to find the service description locally or from the service registry and uses the service description to bind with the service provider and invoke or interact with the implementation of the web service [2, 5].

From a business environment perspective, the service provider is the owner of the service. From an architectural perspective, this is the platform where the web service resides. The service requester, in the business environment perspective, represents the task that requires specific functions to be fulfilled, while from an architectural perspective, it is the application that seeks and calls or initiates interaction with the service. The service registry is a registry of service descriptions that can be searched, and where service providers publish service descriptions. For a web service to be accessible, the service description must be published so that service requesters can find it. As part of the discovery operation, the service requester obtains the service description or sends a query to the service registry for the required type of service. Finally, the service must be invoked. In the process of binding, the service provider calls or initiates interaction with the service during execution, using the binding details in the service description to locate, contact, and invoke the service [2].

Figure 2 illustrates the web service architecture that, using the presented technologies, publishes, finds, and binds its three main roles.

The role of web service specifications is to establish interaction between the web service requester and the web service provider. The technologies used within web

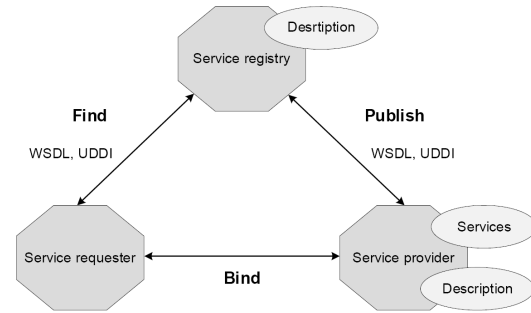


Figure 2: Architecture of web service.

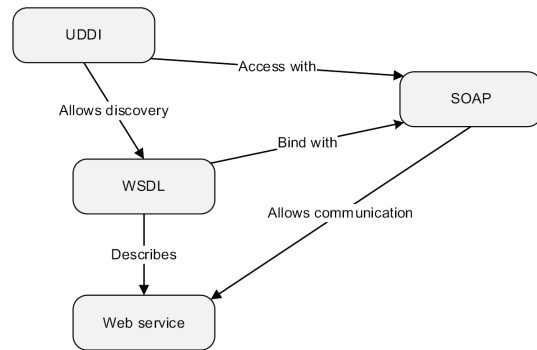


Figure 3: Base standards of web services and relations between them.

services, which also represent the fundamental standards they are based on, include:

- HTTP/HTTPS (Hypertext Transfer Protocol/Secure Hypertext Transfer Protocol)
- SOAP (Simple Object Access Protocol)
- UDDI (Universal Description, Discovery, and Integration)
- WSDL (Web Services Description Language)

On the other hand, these mentioned protocols could not function adequately without corresponding data transmission formats. Data transmission formats commonly used in web services are:

- XML (Extensible Markup Language)
- JSON (JavaScript Object Notation)

Figure 3 illustrates the previously mentioned relationships among web service technologies and their existing connections. UDDI allows discovery through WSDL and access through SOAP. Furthermore, WSDL actually describes the web service, while SOAP enables communication with the web service. Additionally, WSDL is linked to the SOAP standard.

1.2. The classification of web services

Web services are a critical component in modern software architecture, serving as an intermediary layer between application code and its users. They provide a standardized interface, abstracting away language-specific details and enabling applications written in various programming languages to access the same functionality [5].

Key characteristics of web services:

- Web services communicate using XML, a platform-independent and neutral language.
- They can be developed in any programming language and used on any platform.
- Communication occurs using the SOAP protocol over HTTP/HTTPS, with messages sent over the network in plain and structured text.
- Web services are registered in service registries using the UDDI standard, providing a platform-independent catalog where services can be searched and discovered.
- Access to web services is available from various types of applications or other services through a prescribed interface.

Web services can be classified into three main types:

1. Big web services – often referred to as SOAP services, are based on the SOAP standard. They transmit data in XML format and are described using WSDL. SOAP enables communication between applications running on different operating systems and technologies by exchanging messages in an agreed-upon format. It supports various message types, with RPC (Remote Procedure Call) being a common one. This approach is suitable for formal contracts and operations that involve state management and applicable when operations involve state management [3].
2. RESTful services – based on the REST architectural style, which is not a protocol but a design model. These services use a stateless, cacheable, and uniform interface, relying on HTTP methods for communication. They often transfer data in JSON or XML format. RESTful services are ideal for scenarios with limited bandwidth, stateless operations, and caching opportunities [6].
3. POX services (Plain Old XML) – represent a simpler and faster alternative to SOAP services. They utilize raw XML for data transmission and GET/POST methods for communication. POX services do not wrap data in a specific protocol and offer a lightweight approach while maintaining the benefits of strict API (Application Programming Interface) definitions.

The choice between these types depends on various factors that need to be considered while approaching the intended technical solution in observed business environment. For example, the need for formal contracts, state management, bandwidth considerations, and the simplicity of implementation. Additionally, RESTful services and SOAP services can coexist, providing flexibility in selecting the appropriate technology for specific use cases.

Each of the techniques used in web services has its own advantages and disadvantages. For example, when comparing REST and SOAP, it's essential to consider their characteristics [7]:

- SOAP is a protocol with a specification, while REST is an architectural style.
- SOAP requires a WSDL file to convey service functionality to clients, while REST relies on a uniform URI interface.
- SOAP uses more bandwidth due to its message structure, while REST typically sends lightweight JSON messages.
- SOAP can be platform-independent, while RESTful services work well with a variety of formats, including plain text, HTML, and XML.

As previously mentioned, the choice between these web service types should align with specific project requirements and constraints, and it's worth noting that different combinations and hybrid approaches can be employed to meet diverse needs [7].

2. Usage of web service NBS

It has become increasingly common for companies to integrate their business systems with relevant information technology applications from other enterprises. Additionally, there is a clear trend among companies to open up parts of their operations to the community, making them accessible to the general public. One concrete example of such organizational strategy is the NBS web service. The practical application of the NBS web service can vary, depending on the specific needs of organizations and industries. Several practical use cases of the NBS service in business scenarios can for sure include:

1. Financial institutions – banks, insurance companies, and other financial institutions can utilize NBS services to access relevant financial information. This may include exchange rates, financial reports, regulatory guidelines, and other data crucial to their operations. With this information, financial institutions can monitor the market, analyze risks, make informed investment decisions, and provide services to clients.

2. Currency exchange companies and tourism – businesses engaged in currency exchange, tourism agencies, or hotels dealing with foreign currency payments can employ NBS services to track currency exchange rates. This allows them to provide accurate exchange rate information and perform currency conversions for tourists and clients.
3. E-commerce – businesses that operate online can use NBS services to convert currencies when processing payments in different currencies. This ensures precise amounts in the local currency (as these companies are registered in Serbia) for customers and reduces the risk associated with currency exchange rate fluctuations.
4. Analytical companies – analytical or research teams can access NBS services to retrieve financial reports and statistics. This enables them to monitor and analyze trends in the banking sector, market movements, monetary policy, and other aspects of the financial system. Based on this data, they can provide information and advice to clients, make investment decisions, or report on market conditions.

These are just a few examples of practical applications of NBS services. It's important to note that the specific implementation and usage of NBS services can vary depending on an organization's needs and business type.

2.1. Business Project and communication between entities

First of all, it's essential to highlight that the use of NBS web services within the business project for specific company was designed with Visual Studio 2019 software for programming in the C# programming language. Visual Studio was the primary tool used to establish a connection with the NBS web service and subsequently connect to the company's database for specific data entry, required data validation, and more. Within the project, three main entities are distinctly differentiated to facilitate complete communication. These entities are:

1. NBS Web Service – a service provided by the National Bank of Serbia, designed to be accessible to all legal entities upon registration.
2. VS (Visual Studio) – an application used to write code that communicates with the NBS web service. It conducts proper data validation and performs data entry into the company's database within the business environment.
3. Database – a specific entity containing a range of tables related to exchange rates and a list of necessary currencies in the day-to-day operations of the company.

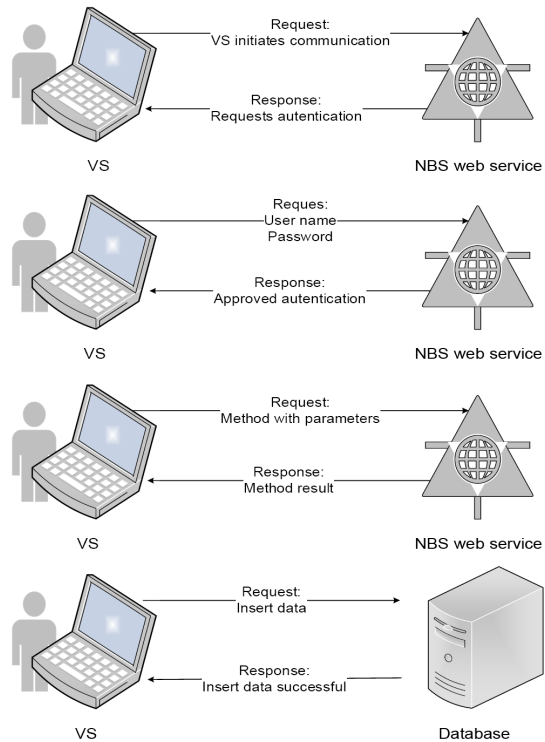


Figure 4: Communication between entities.

Communication begins with VS triggering a request to the NBS web service, which requires appropriate authentication. After a successful authentication process, VS gains access rights to the NBS web service. If the authentication process fails, the NBS web service responds with an appropriate message detailing what went wrong. Various reasons for failure could include incorrect usernames or passwords, unavailability of the service for some reason, or changes to its address or name.

After successful authentication, VS can use methods within the invoked NBS web service. VS calls the method using the required parameters if the method demands them. As a response from the NBS web service, VS obtains the results of the invoked method, which can encompass data about various details provided by the NBS web service. Subsequently, adhering to prescribed data validation rules for each included table, VS performs data entry into the database. It's worth noting that, in this context, the term "database" refers to a local database used for everyday data entry and manipulation within the observed business environment.

Figure 4 illustrates a simplified exchange among three main entities in the case of when the communication is successful.

2.2. NBS and web services

NBS web services offer access to diverse financial data and information supplied by the National Bank of Serbia. However, prior to their utilization, specific prerequisites must be fulfilled. In general, several customary prerequisites exist for engaging NBS web services:

- Registration – entities or individuals intending to use NBS web services typically undergo registration with the National Bank of Serbia. The registration process might contain the submission of specifically designed request, the provision of relevant information about the organization (company) or user-related, the completion of forms, or analogous procedures.
- Permissions and authentication – after registration, the NBS proceeds with data verification and assigns access permissions. This may involve associating users with appropriate user groups or granting specific access privileges to particular data or services. Additionally, users typically receive identification information (such as usernames and passwords) or digital certificates for authentication when accessing NBS services.
- Legal and regulatory compliance – the use of NBS web services is subject to compliance with relevant legal regulations and regulatory guidelines. Organizations or individuals must adhere to all terms and restrictions established by NBS regarding data access and usage.
- Technical Specifications – NBS typically provides technical specifications or API documentation describing the method of communication with web services. This may include details about endpoints, HTTP requests, data formats (e.g., XML or JSON), and other technical specifics necessary for proper and effective communication with NBS services.

It is important to note that these conditions are described at a general level, and the specific terms for using NBS web services may vary depending on the specific service and the data accessed with use of that web service. Accordingly, when initiating the process of gaining access to one of the NBS web services, before starting any of the described prerequisites, it is necessary to establish direct communication with the NBS regarding the exact steps and technical details.

Within the scope of the project, the NBS web service was utilized to access data related to the current exchange rates of currencies concerning the Serbian “dinar” (RSD). Additionally, this service enables the retrieval of all exchange rate lists published by the NBS since May 15, 2002. It is noteworthy that this is not the sole web service for accessing currency exchange rate lists. Apart from it, there

exists a service allowing the access to currency exchange rates for a specified date, as well as rates associated with foreign exchange transactions [8].

There are two types of services for accessing the current exchange rate list, both having identical names and input parameters but differing in the type of object they return [8]:

1. CurrentExchangeRateService – The methods of this service return objects of the DataSet type.
2. CurrentExchangeRateXmlService – The methods of this service return objects of the String type.

For a better understanding of the difference between these two service types, it is necessary to explain the mentioned object types. A DataSet is a collection of data that typically contains the structure and methods for data manipulation. It is used in the context of data analysis and machine learning for training models and evaluating their results. As such, it can encompass various data types, such as numerical values, textual data, or categorical attributes. The DataSet object facilitates efficient data management, loading, transformation, and preparation for analysis.

On the other hand, a String is a data type representing a sequence of characters. It is used for representing text and is often employed for text manipulation and processing in programming. Strings can contain letters, numbers, special characters, and spaces. They can be used for storing names, addresses, messages, codes, or any other textual information. The primary distinction between a DataSet and a String lies in the fact that a DataSet represents a collection of data, while a String represents a sequence of characters. A DataSet is used for organizing and manipulating various types of data in data analysis and machine learning, while a String is employed for manipulating textual data and working with textual information. A DataSet can contain Strings as one of the data types within. NBS provides a web service for accessing the mid-exchange rate of currencies through the public exchange rate list service. The endpoint (<https://www.nbs.rs/kursnaListaModul/srednjiKurs.faces>) is accessed using the GET method. The fundamental parameters that can be provided when using the service and its respective methods are as follows:

- Date (optional) – specifies the date for the desired exchange rate list concerning the Serbian “dinar” (RSD). The date format is “dd.MM.yyyy” (indicating that the first two letters, dd, correspond to days in the month, the next two letters, MM, refer to months in the year, and the last four letters, yyyy, denote the specific year). For example, the date “14.02.1991” corresponds to February 14, 1991. Additionally, this date format adheres to regional settings and serves as the official date format in Serbia, aligning with the standard format

used in many parts of the world. It is employed for displaying dates in official documents, business transactions, and more. If the parameter related to the date is not specified, the method defines the use of the most recent exchange rate list.

- Currency (optional) – Represents the currency code for the desired exchange rate list concerning the Serbian "dinar" (RSD). For example, "USD" for the US dollar, "EUR" for the euro, "CHF" for the Swiss franc, etc. If this parameter is not provided, the method is designed to retrieve all available exchange rate data for the specified date for each currency.

Upon sending the appropriate parameters and utilizing the defined methods, an API response is obtained, containing data on the mid-exchange rates for the requested date and currency. The response format can be JSON or XML, depending on the API configuration. Naturally, as previously mentioned, all of this necessitates registration with the National Bank of Serbia and obtaining access credentials and keys enabling access to the respective web services.

3. Conclusion and results

The use of the NBS web service in a business environment can provide organizations with numerous advantages and opportunities. By gaining access to currency exchange rates, financial reports, regulatory data, and payment system information, organizations can make informed business decisions, track trends, analyze the market, and meet regulatory obligations.

Utilizing the NBS web service enables quick and easy access to relevant financial information, reducing the need for manual data collection. These pieces of information can be used for calculations, currency conversions, trend monitoring, risk analysis, and business strategy planning.

Integration with the NBS payment system allows organizations to efficiently manage payment flows, check payment statuses, and streamline transaction processing. This is especially beneficial for organizations providing payment services or involved in payment transactions.

It's essential for organizations planning to use NBS web services to follow and adhere to the terms and guidelines set for accessing these services. This may include registration, obtaining permissions, and authentication to ensure secure and lawful data usage.

Therefore, the use of the NBS web service offers organizations the capability to efficiently access financial information, analyze the market, manage payment flows, and fulfill regulatory obligations. These services can enhance business efficiency and informed decision-making,

contributing to a successful operation within the financial landscape.

What sets the use of the NBS web service apart from traditional (manual entry) data input are the following benefits:

- Time – referring to the time saved by automating processes through the use of the web service to achieve identical results.
- Error potential – significantly reduced with the implementation of web service-based solutions.
- Adaptability to changes – for instance, if there are internal changes within the organization, and the previously responsible person leaves, the application will continue to function independently.
- Process standardization – precisely knowing who and when data is entered, making it easier to address potential issues should they arise.
- Documentation – it becomes possible to track when changes were made, what the changes were, and assess their validity. Essentially, there is a precise record of changes to the observed table within the database.

Considering all of the above, it is evident that implementing web services within a business environment is desirable wherever there are suitable opportunities.

Acknowledgment

This paper was supported by the Faculty of Technical Sciences, Novi Sad, Serbia, Department of Energy, Electronics and Telecommunications, as part of the project entitled "Development and application of modern methods in teaching and research activities at the Department for Power, Electronics and Telecommunications".

References

- [1] IBM, IBM Documentation, What is a web service?, <https://www.ibm.com/docs/en/cics-ts/5.1?topic=services-what-is-web-service>, 2019.
- [2] H. Kregen, Web Service Conceptual Architecture (W3CA 1.0), IBM software group, https://www.cse.uoi.gr/~pitoura/courses/ds04_gr/webt.pdf, 2001.
- [3] Webprogramiranje.org, Web servisi (osnove), <https://www.webprogramiranje.org/web-servisi-osnove/>, 2023.
- [4] W3C, Web Standards – The promise of web standards, <https://www.w3.org/standards/>, 2023.
- [5] P. Kulchenko, Programming web services with soap, O'Reilly, 2001.
- [6] P. Macherla, Types of Web Services – Big and RESTful, <https://ibytecode.com/blog/types-of-web-services-big-and-restful/>, 2016.

- [7] A. Walker, SOAP vs REST API: Difference Between Web Services, <https://www.guru99.com/comparison-between-web-services.html/>, 2016.
- [8] N. B. Srbije, Sistem veb servisa – tehnička dokumentacija, <https://www.nbs.rs/sr/drugi-nivo-navigacije/servisi/sistem-veb-servisa-NBS/>, 2023.

Ensuring High Availability of Clusters within the Network Infrastructure using Microsoft Hyper-V Technology in a Medium-sized Enterprise

Damir Bradić¹, Dejan Nemeć^{1,*}

¹Faculty of Technical Sciences, University of Novi Sad, Trg Dositeja Obradovića 6, 21000 Novi Sad, Serbia

Abstract

When adding hosts to the cluster, it is necessary to provide proper hardware components in order to ensure correct communication, redundancy and failover in case of losing of any point in the system. Specific hardware requirements may vary depending on version of the cluster software, desired level of risk, and level of high availability.

Microsoft Hyper-V (Hypervisor-based Virtualization Technology) is a native hardware virtualization hypervisor that enables the creation and running of virtual entities called virtual machines (VM).

Keywords

Virtualization, CPU, RAM, Fiber Channel, SAN, Hyper-V, cluster, adapters, high availability, redundancy, infrastructure

1. Introduction

Designing of information technologies systems implies planning of the business continuity. The business continuity should ensure the uninterrupted processes and response to all unwanted events that may affect business flow. If not to prevent, at least to mitigate the consequences and accelerate the recovery of the system. With this, the company's resilience increasing. Information technologies system and recovery procedures include redundancy at all points. On network nodes, data centers, and geographically remote locations. One of important aspects of recovery is the creation of backup plan. Organizing copies of data to be safe in case of the unauthorized modification, infection, or deletion. Additionally, regarding regular data backups, elements of validation, encryption, and access control ensuring the data integrity. Risk management, as another key element necessary in protection planning, ensures and confidentiality too.

Virtualization technologies helps to better organize business continuity, speed up recovery time, reduce system construction costs, enable scalability, and enhance elasticity. Virtualization is a process that enables efficient use of the physical resources of computer systems. It use the physical resources of the system to split it into several independent and distinct environments, each as isolated separate system.

This paper work will describe the elements of a high-availability information system based on the Microsoft

Hyper-V technology principle in a cluster environment of redundant servers (Failover Cluster). All elements organized in such environment form one centralized system that is resistant to errors, extensible, elastic, and reliable. Hardware elements of system like that can vary depending on size and complexity requirements. In generally, servers and devices must meet the software and hardware requirements for cluster integration and hypervisor applications. Must include multiple network adapters, redundant optical or network switches, central data storages and redundant disk controllers, power supplies, and cooling fans. The hardware elements should also be compatible with various software to enable the full management, administration, monitoring, and notification of the system's health like Out-of-Band Management (OOBM) system.

2. Virtualization

The virtualization concept importance is the ability of better exploit the resources, system flexibility and scalability, and to enable centralized management. Some of key benefits are:

- Virtualization allows the creation and running of multiple virtual machines per host, which increases the utilization of hardware resources. By reducing the number of physical servers, hardware costs and energy consumption.
- The ability of running multiple virtual servers on a single physical machine that reducing the space requirements in a data center or server room.
- Enables easy scaling of resources. Add services, computers, storage or network and create new virtual machines quickly and without excessive investment.

BISEC'23: 14th International Conference on Business Information Security, November 24, 2023, Niš, Serbia

*Corresponding author.

✉ damir.bradic@outlook.com (D. Bradić); denem@uns.ac.rs (D. Nemeć)

- Each virtual machine working independently, providing a degree of isolation. Isolation improves security. If one of the machine compromised, it does not affect the operation of the others.
- Virtualization simplifies management by allowing virtual resources creating easily, cloned, moved or deleted. Software-defined management tools provide centralized control over the entire virtualized infrastructure.
- Virtualization provide easy back up of entire virtual resources and enabling efficient disaster recovery.
- Providing an isolated environment for testing and development. Creation of virtualized instances with different OS, software configurations and network settings without affecting the production environment.
- Enable support for old legacy apps on newer hardware and OSes that natively do not support it.
- Virtualization can improve availability by providing features such as seamless migration where virtual machines can be migrated from one physical host to another ensuring continuity of services [1].



Figure 1: Selected hardware, back and front view.

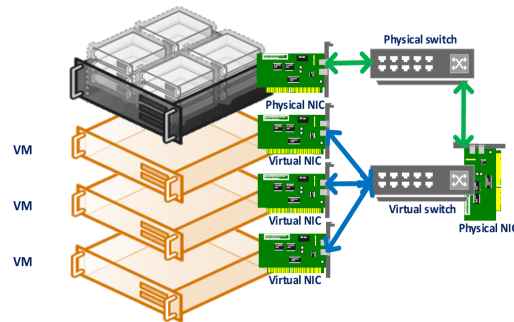


Figure 2: Logical diagram of Hyper-V virtual network.

3. Hardware

The hardware equipment selection for Microsoft Hyper-V cluster depends on performance requirements and the ability of virtualization support, demands for high SQL database query, or high-performance processing. This kind of equipment should place in a data center where all necessary conditions like uninterrupted power supply, air conditioning, physical security and the possibility of using more than one ISPs (Internet service providers) are possible. The data center offers the service like Telehousing, which is a service of IT infrastructure placing on the way that ensures security, flexibility and reliability. The equipment mounted in metal rack cabinets organized by standardized device dimensions. Usually, dimensions of 19" is common, width and height indicated in RU (Rack Unit), HU (Height Unit) or just U (Unit). One U is 4.445 cm. Cabinet depths can be 60, 80 and 110 cm. The choice of cabinet height determined by the amount of equipment. It means that all equipment must be compliant with data center standards. Table 1 shows an example of hardware list required for a cluster installation.

4. Hyper-V virtual network

To access the outside of world, the Internet, and other virtual or physical machines, VM needs to have a virtual

network card as well as a network to connect with it. Basic connectivity in Hyper-V includes two parts, a virtual adapter and a virtual switch. Figure 2 describes how a virtual switch connects to an Ethernet network and a virtual network adapter to a virtual switch port [2].

4.1. Virtual network adapter

vNIC (Virtual Network Adapter) is a virtualized version of the physical one which is used for connection on to Hyper-V environment, physical and virtual environment. Each VM can use one or more vNICs for connection. There are two types of adapters:

- Legacy Network – adapter who emulates a physical NIC for compatibility with older OSes, which do not support Hyper-V Integration Services.
- Network adapter – vNIC that supports Hyper-V integration services. Provides better performance and vLAN tagging capability.

4.2. Virtual switch

A software based switch who uses host memory to connect a vNIC with a physical uplink to a physical switch using physical adapters configured as teams. The virtual switch has three basic operating modes:

Table 1

Example of hardware configuration for a cluster of four nodes and with SAN configuration

4 x HPE DL360 Gen10 characteristics: 2x CPU Intel Gold 6134, 192GB RAM, Ctrl P408i-a SAS, 2x 240GB SSD, 2x Ethernet quad-port 1GbE, 2x HBA FC 16Gb single port	pcs
HPE ProLiant DL360 Gen10 8SFF Configure-to-order Server	4
HPE DL360 Gen10 Intel Xeon-Gold 6134 (3.2GHz/8-core/130W) 1st	4
HPE DL360 Gen10 Intel Xeon-Gold 6134 (3.2GHz/8-core/130W) 2nd	4
HPE 32GB (1x32GB) Dual Rank x4 DDR4-2666 CAS-19-19-19 Registered Smart Memory Kit	48
HPE 240GB SATA 6G Read Intensive SFF (2.5in) SC 3yr Wty Digitally Signed Firmware SSD	8
HPE StoreFabric SN1100Q 16Gb Single Port Fibre Channel Host Bus Adapter	8
HPE 96W Smart Storage Battery (up to 20 Devices) with 145mm Cable Kit	4
HPE Smart Array P408i-a SR Gen10 (8 Internal Lanes/2GB Cache) 12G SAS Modular	8
HPE Ethernet 1Gb 4-port 331FLR Adapter	4
HPE DL360 Gen10 High Performance Fan Kit	8
HPE 800W Flex Slot Platinum Hot Plug Low Halogen Power Supply Kit	8
2x SAN Switch 16Gb SN3000B	pcs
HPE SN3000B 16Gb 24-port/12-port Active Fibre Channel Switch	2
HPE B-series 16Gb SFP+ Short Wave Transceiver	24
HPE Premier Flex LC/LC Multi-mode OM4 2 fiber 5m Cable	24
MSA 2040	pcs
HPE MSA 2040 Storage 24-bay (SFF)	1
P2000 G3	pcs
HPE MSA P2000 G3 Modular Smart Array Systems 24-bay (SFF)	1

- Private mode, which opens communication between VMs only.
- Internal mode enables communication between the VM and the host.
- External mode, in addition to communicating between the VM and the host, uses the host's NIC to connect to the physical switch and enable communication with other systems.

Virtual switches cannot communicate with each other without the presence of a VM with a router role on the same host. A private and internal virtual switch connected just to adapters on the same virtual switch, and an external one depends on the physical adapter. A physical adapter acts as an uplink for access to an available physical network. The vNIC OS management is responsible for the connection between the physical adapter and the switch. Figure 3 illustrates the logic diagram of three different virtual switch modes.

By default, a virtual switch configured to perform the next functions [3]:

- Ethernet frame switching – has the capability of L2 network switch operations such as of be aware of all MAC addresses connected to the virtual switch and reading them from Ethernet packets. There is no L3 routing capability.
- 802.1q VLAN access mode – vNIC adapters can be assigned to a vLAN network. Like a type of

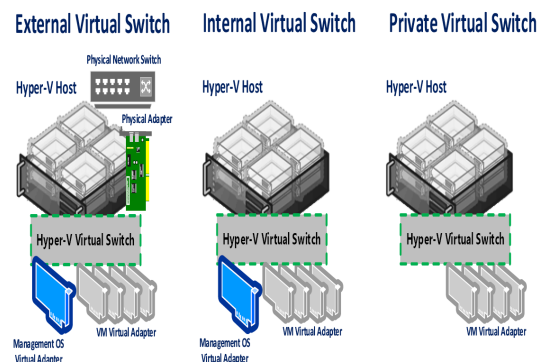


Figure 3: Operating modes of the virtual switch [3].

isolation where traffic will flow alone between adapters who has the same vLAN tag.

- 802.1q VLAN Trunk connectivity option – This option allows the VM to see the traffic of multiple VLANs. A virtual switch port can accept traffic from all configured vLANs.
- 802.1p QoS – Traffic prioritization. Hyper-V has its own way of organizing QoS through two modes:
 - Absolute mode, which can guarantee a minimum or limit the outgoing speed of

the vNIC adapter.

- Weight mode, which can guarantee the minimum or set the maximum speed of the vNIC adapter.
- SR-IOV (Single Root I/O Virtualization) – requires hardware that has built-in SR-IOV capability. This option has the ability to connect a limited number of vNIC adapters which be exposed to the physical NIC and enable almost near real speed of the physical adapter.
- Extensibility – Microsoft has released an API, which can creating filter triggers for a virtual switch.

4.3. Link aggregation and teaming of physical ports

Windows Server 2012 brought the built-in ability to aggregate network adapters to the team. Hyper-V took advantage of this capability. Port channel is the name used by the Cisco vendor, while others use Link Aggregation. In the background of these names is a technology that should prevent a loop in communication that can lead to a broadcast storm (Broadcast Storm). Therefore, with the VLAN trunk protocol for connecting two switches with redundant connections, the ports are binding in the port channel for Cisco or the link aggregation group LAG (Link Aggregation Group). When ports joined in such a group, they behave as one port. Important is to remember that the MAC addresses of the individual ports on the switch have disappeared. A new MAC address assigned now belongs to the LAG channel and not to the switch port.

4.3.1. NIC Team

The NIC team settings, introduced in Windows Server 2012, brings three possible options [4]:

- Switch independent – similar to the traditional model where the use of switch options is optional. A Hyper-V virtual switch will register all of its MAC addresses to a single port so that all incoming traffic will go through single physical link. The advantage of this method is that it is independent of the switch type and possible is to connect several switches in purposes of redundancy. The disadvantage is that all incoming traffic bounds on to one adapter.
- Static mode option – virtual switch and physical switches can work in this mode. The groups on the switches must be configured the same on both sides. The MAC addresses on both sides registered as group addresses rather than individual ports, allowing incoming and outgoing traffic to

use any available physical link. The disadvantage is that all switches must support this option and the ability of sharing traffic between physical links is lost.

- LACP (Link Aggregation Control Protocol) – uses functions similar to the static mode option with the difference where connected switches use LACPDU (Link Aggregation Control Protocol Data Unit) packets to detect connection problems during communication. If the team setup uses LACP, the switch will detect that one side is using three of the four ports for the group and will not try to use the fourth link for traffic, which is not the case with the static mode option.

4.3.2. Bandwidth in link aggregation

When multiple physical connections used in a single link, it is more likely to achieve load balancing than bandwidth aggregation. In most cases, the switch team that sending data controls the flow with specific load balancing algorithms. The sending system will transmit over a specific link, while each communication will go exclusively over one physical link. Splitting traffic across multiple paths could cause conflicts, buffer overload, and even drop of packets. TCP (Transmission Control Protocol) and a few other protocols have ways to correct such errors. Using those protocols is demanding operation that increases the use of resources and does not overcome the limitations of using a single physical link. Another reason for using a single physical connection is practicality. Connecting multiple ports from one switch to another is simple, but from one to the second, to the third and from the third to the fourth switch there is a possibility of lacking of ports. The longer is the chain, the greater is the possibility that there will be a reduction in throughput, and one adapter at the end definitely.

An exception to this rule is the switch-independent team option where incoming traffic directed to a single physical adapter as all MAC addresses registered in one location. Outgoing traffic, in this case, balanced over all other available ports. If used with the Hyper-V balancing algorithm, the MAC addresses of the vNIC adapters equally distributed to all physical adapters. Each vNIC can still use the maximum speed of a single port [5].

5. Using of link aggregation

A simple connection between servers and switch implies one network card connected to one switch port where the bandwidth depends on the characteristics of the slowest component. Microsoft team helps to avoid downtime due to a card, cable, switch, or switch port error. The Microsoft team aggregates multiple physical links into one logical link. On that way achieves better performance,

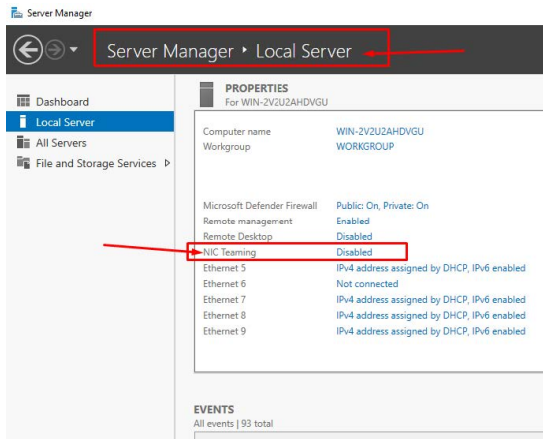


Figure 4: Server Manager.

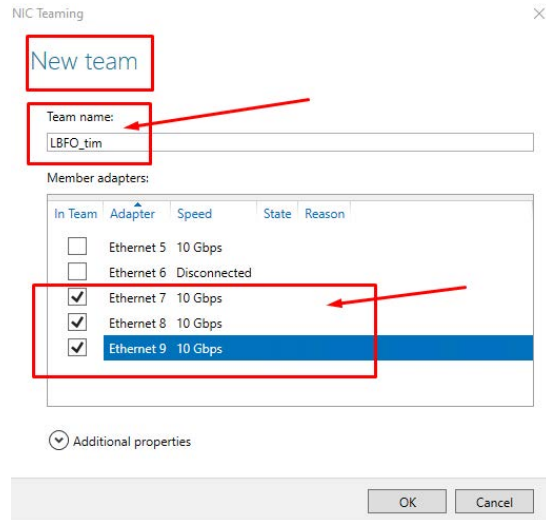


Figure 6: Team selection.

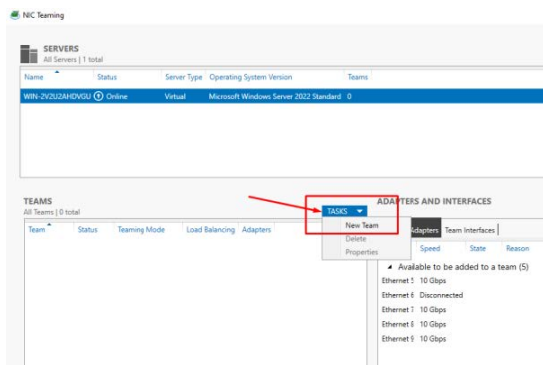


Figure 5: Adding a new team.

- Load balancing,
- Selecting an inactive adapter.

Figure 6 and Figure 7 describes the process of choosing a name, adding an adapter and setting additional options. The team mode option has three additional settings depending on the desired balancing:

- Switch independent,
- LACP Aggregation Protocol,
- Manual mode.

The load balancing option calculates the load using next options:

- Created adapter hash addresses,
- Assigned MAC addresses of Hyper-V ports,
- Dynamic, by combining the previous two.

reliability, LBFO (Load Balancing and Failover) load balancing and redundancy of both adapters and cables and switches are ensured.

5.1. Configuration

After the initial installation of all necessary components, teaming configured with use of the graphical interface or with PowerShell commands.

During initial setup using the server manager graphical interface shown in Figure 4, in a new window the NIC Team option opens. Figure 5 shows the option how to add a new team, where the adapters inserted into the team by simply marking. It is necessary to choose a team name and two or more adapters to add in to the team.

Two or more adapters can be group in to a team, it is necessary to select a team name and set additional options. Additional options are:

- Team mode,

6. Discussion and conclusion

LBFO Load balancing, using the Microsoft team can be set up more than one way depending on hardware and network configuration. In the example of setting, switch independent mode, using dynamic balancing, with a simple connection of two network adapters where both are active. If one adapter goes down, the other takes over the traffic. Next must be consider when calculating the load:

- The maximum data speed rate that the adapter can achieve.

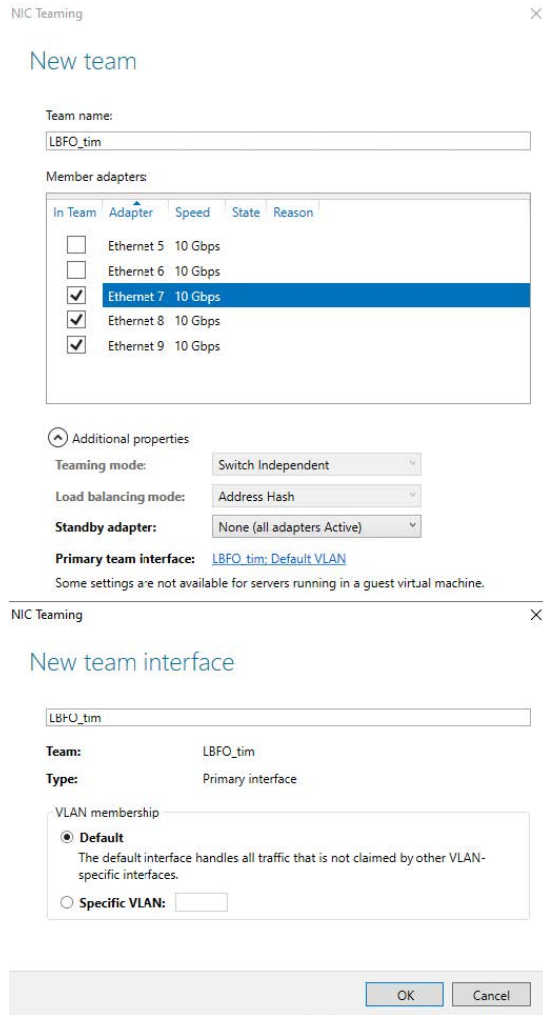


Figure 7: Additional options setting.

- Method of traffic distribution by adapter included in the team. If both active, with symmetric or asymmetric traffic distribution can be assigned depending on the requirements. For example a ratio of 50:50 or 60:40.
- The time it takes of the system to detect the error and redirect the traffic.
- The total amount of traffic that the system must handle.
- The condition that determines the error and triggers the redirection. For example the status of network adapters and connections.

Mathematically, simplified as:

- Throughput rate X bit/s,

- Traffic distribution $A : A$ or $A : B$,
- Total amount of traffic: T bit/s,
- Error detection time H s,
- Redirection trigger condition, event D .

In case of symmetric distribution $A : A$:

1. Total Available Flow (URP) = $2 * X$,
2. Traffic for each individual adapter (SPA) = $URP/2$,
3. The condition that determines the error.
4. If D happens and one adapter fails, the system should initiate a failover.
5. The time to detect the error H should be as short as possible.

If it is an asymmetric traffic distribution $A:B$, the traffic can be determined as:

1. For adapter $A = T * (A/(A + B))$,
2. For adapter $B = T * (B/(A + B))$.

This is a simplified example and the calculations in the production are much more complex and affected by a large number of different factors. Accurate calculations require constant monitoring and adjustment on the entire system.

Acknowledgment

This paper was supported by the Faculty of Technical Sciences, Novi Sad, Serbia, Department of Energy, Electronics and Telecommunications, as part of the project entitled "Development and application of modern methods in teaching and research activities at the Department for Power, Electronics and Telecommunications".

References

- [1] M. Sharma, Characteristics of virtualization, <https://www.geeksforgeeks.org/characteristics-of-virtualization/>, 2023.
- [2] Altaro, Hyper-V Virtual Networking configuration and best practices, <https://www.altaro.com/hyper-v/virtual-networking-configuration-best-practices/>, 2023.
- [3] S. Cooley, M. Briggs, D. Heuer, H. Lohr, N. Schonning, Hyper-V Integration Services, <https://learn.microsoft.com/en-us/virtualization/hyper-v-on-windows/reference/integration-services/>, 2023.
- [4] E. Siron, Link Aggregation and Teaming in Hyper-V, <https://www.altaro.com/hyper-v/link-aggregation-teaming-in-hyper-v/>, 2023.
- [5] A. Fields, Hyper-V Failover Cluster: Converged Network, <https://www.itpromentor.com/hvc-net-converged/>, 2022.

Exploring the Power of AI in Internet Security: Balancing Attacks and Defenses in Black and White

Valentina B. Paunović^{1,*}, Sedat A. Uyar²

¹Faculty of Information Technology, Belgrade Metropolitan University, Tadeuša Košćuška 63, 11000 Belgrade, Serbia

²Schleupen AG, Galmesweg 58, 47445 Moers, Germany

Abstract

In this paper, we explore the dynamic role of Artificial Intelligence (AI), Machine Learning, and ChatGPT in the realm of internet security, which includes both offensive and defensive cyber strategies. These technologies, particularly AI enhanced by Machine Learning and the sophisticated functions of ChatGPT, are revolutionizing the field of cybersecurity. They play a crucial role in enhancing the understanding of threats and strengthening the defense mechanisms for the protection of organizational data and digital assets. Our comprehensive review provides key insights into how AI and Machine Learning are fundamentally altering internet security, highlighting their essential role in combating evolving cyber threats. The paper further delves into a detailed analysis of the diverse applications of AI, including Machine Learning and ChatGPT, in both augmenting and potentially compromising enterprise security. We specifically focus on AI's dual capacity in Business Security, augmented by Machine Learning and ChatGPT, and illustrate this through practical examples that encompass both the generation and identification of malware. The paper concludes with a synthesis of these findings, emphasizing the significant impact of these technologies in the field of cybersecurity.

Keywords

Internet Security, Cyberattacks, Machine Learning, Data Protection, Malware

1. Introduction

In today's digital age, where technology is tightly interwoven with business operations and information exchange, the security of sensitive data has become a paramount concern for organizations worldwide. The relentless march of digitization, coupled with the ever-evolving sophistication of cyber threats, demands a continuous evolution in the methods and strategies employed to protect valuable data assets.

The landscape of Business Information Security has entered a new phase marked by the integration of Artificial Intelligence (AI), representing a pivotal juncture in the ongoing battle to safeguard sensitive data. AI promises transformative capabilities that augment traditional security measures, challenging us to gain a deeper understanding of its role and potential within Business Information Security. This paper embarks on a comprehensive exploration of the symbiotic relationship between AI and Business Information Security, aiming to elucidate the multifaceted ways in which AI is revolutionizing and redefining the security landscape. We will delve into innovative strategies, applications, and implications of this

convergence, propelling organizations toward a state of heightened resilience in the face of an ever-expanding array of digital threats.

As we navigate through this discourse, we will embark on a journey that transcends conventional boundaries, probing the depths of AI's influence on both attack and defense aspects of security. On one hand, AI empowers organizations to understand and anticipate the various types of attacks that threaten their systems, enabling them to respond more effectively. On the other hand, AI equips us with the tools to fortify our defenses against these attacks. Each facet we explore reveals a piece of the mosaic, illustrating the pivotal role AI plays in both comprehending the nuances of attacks and strengthening our digital citadels.

Our quest to harness the power of AI for next-generation Business Information Security is not merely an academic endeavor but a pragmatic pursuit of practical solutions to real-world challenges. We invite researchers, practitioners, and visionaries to join us on this transformative journey, where the synergy between AI and security safeguards our data and shapes the future of digital security practices. With this introduction as our guiding beacon, let us embark on a voyage that explores the uncharted territories of AI-powered security in the quest for a safer, more resilient digital world. AI and machine learning (ML) represent swiftly advancing technologies that have made substantial progress across various industries, including cybersecurity.

The cybersecurity landscape has become increasingly complex and challenging. With cybercriminals con-

BISEC'23: 14th International Conference on Business Information Security, November 24, 2023, Niš, Serbia

*Corresponding author.

✉ valentina.paunovic@metropolitan.ac.rs (Valentina B. Paunović);
sedat.uyar@suyar.de (S. A. Uyar)

🆔 0000-0002-9990-1532 (Valentina B. Paunović);
0009-0004-5538-0095 (S. A. Uyar)

stantly evolving their tactics and leveraging new vulnerabilities, traditional security measures struggle to keep pace with the rapidly changing digital threat environment [1]. This dynamic underscores the critical importance of AI and ML. These technologies present a pivotal opportunity for businesses to revolutionize their approach to cybersecurity, significantly enhancing threat detection, prevention, and response capabilities.

In this article, we aim to unpack the complexities of AI and ML and their profound impact on the future of cybersecurity. We will explore how these advanced technologies can be leveraged to strengthen your organization's cybersecurity stance. Understanding the specific roles and benefits of AI and ML within the cybersecurity sector empowers organizations to make informed decisions about integrating these technologies into their security strategies. This integration fosters a more comprehensive and effective framework for safeguarding digital assets.

By embracing AI and ML, organizations can better anticipate and counteract cyber threats, adapting swiftly to the evolving cybersecurity landscape. This article seeks to provide clarity on the advantages of incorporating AI and ML into cybersecurity practices and the necessity of their inclusion in modern security solutions. With this knowledge, your organization can enhance its cybersecurity posture, ensuring more robust protection against the sophisticated digital threats of today's world [1].

2. AI and Business Information Security

The interplay between Artificial Intelligence and Business Information Security represents a critical juncture in the realm of contemporary cybersecurity. As organizations increasingly rely on digital infrastructure for their operations, the protection of sensitive information has become paramount.

This literature review serves as an in-depth exploration of this dynamic intersection, shedding light on the profound impact of AI on the security landscape. In recent years, AI has evolved into a pivotal force within the cybersecurity domain. Its transformative potential is evident in various facets of security operations, from threat detection to incident response.

This review aims to provide a comprehensive overview of the key developments and trends that have emerged in the field. One of the primary focal points is the role of AI in bolstering the security posture of organizations. Through advanced analytics, machine learning algorithms, and real-time monitoring, AI offers the promise of enhanced threat comprehension and proactive defense strategies.

By analyzing vast datasets and identifying anomalies that may elude traditional security measures, AI has the

potential to revolutionize the way businesses safeguard their critical assets. Business Information Security has evolved in tandem with the rapid digitization of business operations. From rudimentary password-based security to multifaceted defense mechanisms, the journey reflects an ongoing response to an ever-expanding array of cyber threats. The advent of AI in recent years marks a pivotal juncture, offering the promise of a more adaptive and proactive approach to safeguarding sensitive information [1].

In this paper, we will explore both the positive and negative aspects of utilizing artificial intelligence in Business Security. We will provide a detailed explanation of the ways it can be employed for both beneficial and detrimental purposes in protection. After presenting the white and black sides of AI usage in Business Security, an example of using AI in the creation of malware, as well as in malware detection, will be provided, which will also serve as the conclusion of the paper.

3. Utilizing AI as a Defense: Exploring the White Side

In the ever-evolving landscape of cybersecurity, organizations face a daunting challenge: protecting their valuable data and digital assets from an ever-expanding array of threats. The rise of AI has introduced a transformative element into this complex equation. While AI is often associated with both offense and defense, in this part, we shine a spotlight on its crucial role as a defense mechanism in business security.

3.1. AI-Powered Threat Detection - Enhancing Security with Intelligent Surveillance

Traditional security measures often rely on static rule-based systems that struggle to keep pace with the dynamic nature of modern threats. AI, on the other hand, excels in real-time threat detection. By analyzing vast datasets and recognizing patterns and anomalies, AI-driven security systems can swiftly identify suspicious activities, even those that would evade human observers. This level of intelligent surveillance ensures a proactive defense posture [2].

Traditional methods of threat detection often struggle to keep up with the ever-evolving tactics of cybercriminals. This is where AI-powered threat detection systems have emerged as a game-changer in the world of cybersecurity. These systems leverage machine learning algorithms and advanced analytics to provide organizations with real-time, intelligent surveillance capabilities.

3.1.1. The Role of AI in Threat Detection

In the ever-evolving landscape of cybersecurity, the traditional methods of threat detection often find themselves challenged by the rapidly changing tactics of cybercriminals. This is where AI-powered threat detection systems have emerged as a pivotal asset in the realm of cybersecurity, significantly enhancing an organization's ability to safeguard its digital assets [3]. For example, Sagnik et al. [4] introduced the concept of deep neural networks for image recognition, which has since been adapted for anomaly detection in cybersecurity. This approach allows organizations to proactively detect and respond to emerging threats, reducing the window of vulnerability.

Data Analysis at Unprecedented Speeds

AI-powered threat detection systems possess the unique capability to analyze vast amounts of data at speeds that surpass human capacity. This computational prowess allows these systems to recognize subtle patterns, anomalies, and deviations from normal network behavior, all of which can serve as indicators of potential threats. In essence, AI-driven threat detection enables organizations to process and make sense of data in real-time, providing early detection of cyberattacks [3].

Proactive Threat Detection

One of the distinct advantages of AI-powered threat detection lies in its proactive nature. Traditional methods often rely on predefined rules and signatures to identify threats, which can lead to a lag in response time as new attack techniques emerge [5]. AI systems, on the other hand, have the capacity to adapt and evolve in response to evolving threats. They can identify and respond to novel attack vectors promptly, often before they have the chance to inflict significant damage.

Immediate Real-Time Responses

Perhaps one of the most significant strengths of AI in threat detection is its ability to deliver real-time responses [6]. When a potential threat is detected, AI-powered systems can take immediate action to mitigate the risk. For instance, they can automatically block suspicious network traffic, isolate compromised devices, or trigger incident response protocols. This rapid response capability is paramount in minimizing the impact of cyberattacks and preventing their escalation.

Continuous Learning and Improvement

AI-powered threat detection systems are not static; they are in a constant state of learning and adaptation [7]. These systems leverage historical data to refine their algorithms continually. As they encounter new types of attacks and adapt to changing tactics, they become more effective at identifying and mitigating threats over time. This adaptive learning process ensures that the security system remains on the cutting edge of cybersecurity defense.

3.1.2. Real-Time Response

One of the key advantages of AI-powered threat detection is its ability to provide real-time responses. When a potential threat is identified, these systems can take immediate action to mitigate the risk. For example, they can block suspicious network traffic, isolate compromised devices, or initiate incident response procedures. This rapid response time is critical in limiting the impact of cyberattacks.

As is explained in the paper [8], AI-powered threat detection systems are not passive observers; they are proactive defenders. When a potential threat is identified, these systems do not hesitate to take immediate action. They can swiftly and automatically initiate a series of mitigation measures designed to neutralize the threat before it can inflict harm. Blocking Suspicious Network Traffic. Upon detecting unusual or malicious network traffic patterns, AI-driven systems have the capability to act decisively by blocking the source of the suspicious activity. This rapid intervention helps prevent further unauthorized access, ensuring the security of the network. In situations where a device or endpoint within the network is compromised, AI-powered systems can isolate the affected device promptly. This isolation prevents the compromised device from communicating with the network and spreading the threat, thus containing the potential damage. When a significant threat is detected, AI-driven systems can trigger predefined incident response procedures. These procedures often involve notifying cybersecurity teams, collecting forensic data for analysis, and implementing additional security measures to mitigate the threat's impact. This coordinated response ensures that the organization is well-prepared to address the threat effectively. The ability of AI-powered systems to respond rapidly to threats is a game-changer in cybersecurity. It significantly reduces the time window during which an attack can cause damage. By acting swiftly and decisively, these systems limit the potential impact of cyberattacks, helping organizations protect their critical assets and maintain operational continuity. As AI takes center stage in security operations, the preservation of data privacy has emerged as a critical concern. Regulatory frameworks, such as the General Data Protection Regulation (GDPR), have heightened the importance of ensuring that AI applications in security do not compromise individual privacy rights.

3.1.3. Machine Learning and Continuous Improvement

AI-powered threat detection systems possess a unique capability that sets them apart from traditional security measures – the ability to learn and adapt continuously. This dynamic process not only refines their algorithms

but also enhances their accuracy over time. In essence, AI-driven threat detection systems evolve alongside the ever-changing threat landscape, ensuring that organizations maintain effective security measures in the face of emerging cyberattacks.

Paper [9] has a good explanation the cornerstone of machine learning in threat detection lies in its utilization of historical data. AI-powered systems leverage vast datasets encompassing previous security incidents, attack patterns, and network behaviors. By analyzing this extensive repository of information, these systems gain valuable insights into the tactics, techniques, and procedures employed by cybercriminals. Through the examination of historical data, AI-driven threat detection systems continuously refine their underlying algorithms. This iterative process allows them to identify recurring patterns and anomalies, enabling more accurate threat detection. As they encounter a wider array of attack vectors, their algorithms evolve to become increasingly adept at recognizing even the most subtle indicators of malicious activity.

One of the distinguishing features of AI-powered threat detection is its capacity for adaptive learning. These systems are not bound by static rules or rigid definitions of threats. Instead, they adapt and adjust their understanding of what constitutes a potential threat based on the evolving threat landscape [10]. This adaptability ensures that the system remains effective in identifying new types of attacks that may not have been previously encountered.

The continuous improvement process in AI-driven threat detection serves as a proactive defense mechanism. As threat actors develop innovative attack strategies and exploit vulnerabilities, AI-powered systems are well-positioned to stay ahead of emerging threats. Their ability to learn from past incidents allows them to anticipate and respond effectively to new challenges.

3.2. Predictive Risk Assessment - Anticipating and Mitigating Future Threats

AI's formidable ability to process and analyze historical data equips organizations with a powerful tool for foreseeing and proactively mitigating future threats [11]. Predictive risk assessment models, driven by AI algorithms, play a pivotal role in identifying vulnerabilities, potential attack vectors, and weak points in an organization's security posture. This forward-looking approach significantly fortifies overall security, allowing businesses to stay one step ahead of cyber adversaries.

Leveraging Historical Data for Insight

Predictive risk assessment begins with the thorough analysis of historical data related to past security inci-

dents and vulnerabilities. AI-driven systems examine patterns of previous attacks, their outcomes, and the tactics employed by threat actors [12]. By discerning trends and understanding the evolving threat landscape, organizations gain invaluable insights into potential future risks.

Identification of Vulnerabilities and Weaknesses

AI algorithms are adept at detecting vulnerabilities within an organization's infrastructure, software, and network architecture [13]. They can identify potential entry points and areas susceptible to exploitation. These algorithms assess the organization's digital footprint comprehensively, pinpointing areas where security measures may need reinforcement.

Anticipating Attack Vectors

Predictive risk assessment goes beyond identifying vulnerabilities; it also anticipates potential attack vectors. AI-driven models assess how threat actors are likely to exploit identified weaknesses [14]. This foresight enables organizations to anticipate the methods that may be used in future attacks, allowing them to proactively implement measures to block these vectors.

Strengthening Defense Measures

Armed with insights from predictive risk assessment, organizations can take decisive action to shore up their defenses. They can allocate resources to patch vulnerabilities, enhance access controls, and implement security protocols to mitigate potential threats. This strategic approach ensures that security measures are prioritized, and resources are utilized efficiently [15].

Proactive Security Posture

The predictive nature of risk assessment powered by AI contributes to a proactive security posture. Rather than waiting for threats to materialize, organizations can take preemptive measures to reduce their attack surface and minimize risk. This approach not only enhances security but also reduces the likelihood of costly data breaches and cyber incidents.

3.3. Security Automation - Streamlining Defense Mechanisms with AI

The sheer volume of security-related data generated daily can overwhelm human operators. AI-driven automation streamlines the process by handling routine tasks, allowing security teams to focus on critical threats that require human intervention. The integration of AI and cybersecurity represents a symbiotic relationship. AI's strength lies in its ability to process large volumes of data and recognize intricate patterns. In cybersecurity, this translates into improved threat detection, enhanced risk assessment, and the potential for predictive insights. Simultaneously, cybersecurity provides the foundation for AI systems to operate securely by protecting them against external threats [16, 17]. Whether it's managing

access control, patching vulnerabilities, or responding to incidents, AI helps organizations operate more efficiently and respond to threats in real-time [18].

The evolution of AI in the field of cybersecurity has a rich history, with some of the earliest instances of machine learning and AI applications dating back to the introduction of CylancePROTECT® EPP from blackberry over a decade ago [19]. In today's landscape, the ability to predict and prevent new malware attacks has become increasingly vital, especially with the assistance of generative AI, which enables threat actors to rapidly create and test new malicious code. Recent findings from the BlackBerry Global Threat Intelligence Report highlight a concerning 13% increase in novel malware attacks on a quarterly basis. However, the continuous development of technology is helping counteract this evolving threat landscape.

BlackBerry's commitment to bolstering their predictive AI tools is evident through the dedicated efforts of their data science and machine learning teams. Independent assessments have substantiated the effectiveness of Cylance ENDPOINT®, which successfully thwarts 98.9% of threats by proactively forecasting malware behaviour, even when dealing with new variants. This remarkable achievement is the outcome of a decade-long journey marked by innovation, experimentation, and the progression of AI techniques. Notably, this progression includes a transition from supervised human labelling to a composite training approach. This composite approach blends elements of unsupervised, supervised, and active learning, both in cloud and local environments, which has been fine-tuned over time by analysing extensive datasets. As a result, a highly efficient model has emerged, capable of accurately predicting and pre-empting emerging threats.

3.4. Adaptive Access Management - Ensuring Secure and Seamless User Authentication

User authentication stands as a pivotal pillar of cybersecurity, and AI-driven adaptive access management systems have revolutionized the way organizations secure their digital assets. These systems operate by continuously monitoring user behavior, utilizing AI algorithms to detect deviations from normal patterns. This innovative approach ensures the swift identification and mitigation of unauthorized access attempts while simultaneously delivering a seamless and hassle-free user experience to legitimate users [20].

The heart of adaptive access management lies in its ability to vigilantly monitor user behavior in real-time. AI algorithms analyze various parameters, including login times, locations, device types, and typical usage patterns. By establishing a baseline of normal behavior for each

user, the system can promptly identify any deviations or anomalies that may indicate unauthorized access. When a deviation from normal user behavior is detected, the adaptive access management system responds with dynamic authentication decisions. Depending on the level of risk associated with the anomaly, the system may request additional authentication factors, such as biometric verification, one-time passwords, or security questions. This multi-layered approach ensures that only legitimate users can access sensitive resources.

While vigilant monitoring is crucial, adaptive access management systems are also designed to minimize false positives. They use machine learning to refine their understanding of what constitutes normal user behavior. Over time, the system becomes more accurate in distinguishing genuine threats from benign deviations, reducing the risk of inconveniencing legitimate users [21].

One of the key advantages of adaptive access management is its ability to respond swiftly to unauthorized access attempts. When a threat is identified, the system can automatically initiate mitigation measures. This may involve blocking access, triggering alerts to security teams, or implementing stepped-up authentication requirements to thwart potential attackers.

Crucially, adaptive access management systems are designed to strike a balance between security and user experience. Legitimate users are provided with a seamless and convenient authentication process, free from unnecessary friction. This ensures that security measures do not hinder productivity or create barriers for authorized personnel.

3.5. Real-time Incident Response - Rapidly Counteracting Emerging Threats

In the ever-evolving landscape of cybersecurity, cyber threats progress at a relentless pace. These threats exhibit a remarkable ability to adapt, exploiting vulnerabilities and emerging attack methodologies. To effectively counter these ever-shifting challenges, organizations have embraced AI-powered security systems equipped to respond to incidents in real-time. In their paper [22], the authors conducted an in-depth analysis of this topic, offering a critical review and insights into the current state of the art.

The need for rapid incident response cannot be overstated when confronted with the swiftly evolving nature of cyber threats. Cybercriminals operate in a highly time-sensitive environment, where even minor delays can lead to significant damage. Real-time incident response commences with vigilant and continuous threat detection and analysis. AI-driven systems maintain a watchful eye over network traffic, system logs, and user behav-

iors, diligently identifying patterns and anomalies that may signify malicious intent. Upon detecting a potential threat, AI-equipped security systems react swiftly and automatically with predefined actions. These responses encompass a spectrum of measures, including the blocking of suspicious network traffic, the isolation of compromised devices, and the activation of predefined incident response protocols. The paramount goal of real-time incident response is to avert catastrophic outcomes. The agility of swift and precise responses minimizes potential harm and prevents threats from escalating into major security incidents. Furthermore, AI-powered incident response systems display a capacity for adaptive learning. With each incident encountered, they refine their response strategies and enhance their ability to detect threats with precision, thus diminishing the occurrence of false positives. This dynamic process ensures that response efforts remain laser-focused on genuine threats, safeguarding organizations in the ever-shifting landscape of cybersecurity.

3.6. Ethical Considerations - Navigating the Responsible Use of AI in Security

The ethical implications of AI in security demand scrutiny. Concerns regarding bias, particularly in AI models employed in facial recognition and profiling, have gained prominence [23, 18]. Addressing bias and ensuring fairness in AI security applications are paramount ethical imperatives. Additionally, considerations of data privacy and the responsible use of AI in surveillance and profiling are essential topics within this domain.

3.7. Quantum Computing and AI Security Challenges

The rise of quantum computing brings forth unprecedented challenges to AI-driven security systems. In their 2023 study, Rayhan and Shahana [24] investigate the effects that quantum computing could have on encryption techniques, and the resulting vulnerabilities it could introduce in AI-based security frameworks.

Quantum computing's advanced processing power poses a significant threat to conventional encryption methods, which form the bedrock of many current AI security systems. These systems typically rely on complex algorithms that could be swiftly decoded by quantum computers, a capability that endangers the security of encrypted data and the integrity of AI systems reliant on these encryption standards. The research by Rayhan and Shahana in 2023 emphasizes the critical need for AI security mechanisms to evolve in response to the advent of quantum computing. They advocate for the development of encryption methods that are resistant to quantum computing's capabilities, suggesting a new

generation of cryptographic approaches that can endure the enhanced processing powers of quantum machines.

Furthermore, their study draws attention to the wider implications of quantum computing in the cybersecurity domain. As AI and machine learning increasingly permeate security solutions, the disruptive potential of quantum computing becomes more pronounced. This development necessitates a concerted effort among AI and cybersecurity experts to innovate and collaborate, ensuring the continued resilience and effectiveness of security systems against the breakthroughs of quantum computing. In essence, Rayhan and Shahana's 2023 research illuminates the critical challenges posed by quantum computing to AI-driven security, highlighting the urgency for adapting current encryption methodologies and AI security strategies to withstand the transformative impact of quantum computing in the cybersecurity field.

3.8. AI-Powered Malware and Exploits, Including Polymorphic Malware

One of the most concerning developments in the realm of cyberattacks is the emergence of AI-powered malware and exploits, including the notorious Polymorphic Malware. Cybercriminals leverage AI algorithms to create sophisticated malware that can adapt, evade detection, and autonomously target vulnerabilities. This newfound intelligence in malicious software, exemplified by Polymorphic Malware, has the potential to make traditional cybersecurity measures obsolete [25].

3.9. AI Automated Phishing and Social Engineering

Automated phishing and social engineering, combined with AI, pose a formidable challenge in the realm of cybersecurity. These malicious tactics have evolved to leverage AI and automation, making them even more potent threats. AI-enhanced automated phishing attacks use machine learning algorithms to craft convincing and personalized phishing emails. These messages can mimic the writing style of trusted contacts or analyze social media data to tailor content that resonates with the target.

AI also enables attackers to automate the selection of targets based on their likelihood to fall for the phishing scheme. Social engineering attacks benefit from AI by automating the profiling of potential victims. AI-driven bots can scour the internet for vast amounts of personal information, enabling attackers to create highly convincing narratives or deceptive personas. Moreover, AI can generate deepfake audio and video, making it even more challenging to discern between genuine and fraudulent communication. To combat these advanced threats, cybersecurity professionals are increasingly turning to AI

as a defense mechanism. AI-powered security solutions can analyze vast datasets in real-time to detect anomalies, identifying phishing attempts or social engineering attacks. Machine learning models can continuously adapt and evolve to recognize new attack patterns. In this ever-evolving landscape, the synergy between automated phishing, social engineering, and AI highlights the need for a multi-faceted cybersecurity strategy that leverages AI's capabilities to protect against AI-driven threats. As example in paper [26] is given example of creating AI bot which can be used on Facebook messenger and other social media platforms.

3.10. AI-Enhanced Reconnaissance and Targeting

AI is revolutionizing the reconnaissance phase of cyberattacks. Threat actors can employ AI algorithms to scan and analyze vast amounts of data, identifying potential targets and vulnerabilities with unparalleled speed and accuracy. This level of automated reconnaissance empowers attackers to conduct highly targeted and efficient campaigns.

3.11. AI in DDoS Attacks

AI is also finding its way into DDoS (Distributed Denial of Service) attacks, transforming them into more sophisticated and potent threats. Historically, DDoS attacks relied on a multitude of compromised devices to flood a target server with traffic, overwhelming it and causing service disruptions. However, AI is now being employed by cybercriminals to make these attacks even more dangerous. Here is some ways AI is influencing DDoS attacks [27]:

- **Traffic Mimicry:** AI-driven DDoS attacks can mimic legitimate traffic patterns, making them harder to detect. They can adapt to changes in network traffic and even appear as if they are normal user interactions. This dynamic behavior challenges traditional mitigation techniques.
- **Targeted Attacks:** AI can identify specific vulnerabilities in a target's infrastructure and focus the attack on those weak points. This targeted approach increases the chances of successfully crippling a service or website.
- **IoT Botnets:** AI can optimize the coordination of botnets comprising compromised Internet of Things (IoT) devices. These devices can generate vast amounts of traffic, and AI ensures efficient orchestration, making it difficult to stop the attack.
- **Evading Detection:** Machine learning algorithms can be used to bypass security systems. For instance, AI can learn to circumvent

signature-based detection methods by generating attack patterns that don't match known attack signatures.

- **Adaptive Attacks:** AI allows DDoS attacks to adapt in real-time. When an attack is detected and mitigated, AI can modify attack vectors to bypass defenses, making it an ongoing challenge for security teams.

4. The white side: AI and Machine Learning and Malware & phishing detection

In this segment, we explore the utilization of Artificial Intelligence, with a particular focus on Machine Learning as a crucial component of AI, in the realm of malware detection. We examine the current limitations of prevalent methods and investigate strategies to enhance the effectiveness of detection using AI and Machine Learning. Machine Learning, as a vital part of AI in cybersecurity, empowers products to make autonomous decisions. The efficacy of these decisions heavily depends on the quality of the machine learning model, which is itself a function of the training data's quality. In the context of AI-driven malware detection, the data-centric nature of machine learning is paramount. The model's performance and accuracy hinge on the dataset used during its training, as this dataset shapes the model's understanding of what features are statistically relevant for correct predictions.

For example, if a training dataset erroneously suggests that all files larger than 10 MB are malicious, the machine learning model will learn to associate large file sizes with malware, leading to false positives in real-world applications. To mitigate such biases, it is crucial to include a diverse range of benign files in the training set, ensuring the model does not learn incorrect associations [28].

The importance of training machine learning models on datasets that genuinely mirror real-world scenarios cannot be overstated in AI applications. This is especially true in contexts where the models are expected to perform with high accuracy, such as in malware detection. Many machine learning models, particularly deep neural networks, operate as "black boxes," processing input data (X) to produce an output (Y) through complex and often non-transparent operations. This opacity can be problematic, especially when trying to diagnose and address false positives – benign files mistakenly identified as malicious. Minimizing false positives is crucial in machine learning applications for malware detection, as even a single misidentification can lead to significant consequences. Moreover, the dynamic and evolving nature of malware challenges the static nature of many machine learning models. Unlike other domains where data distribution

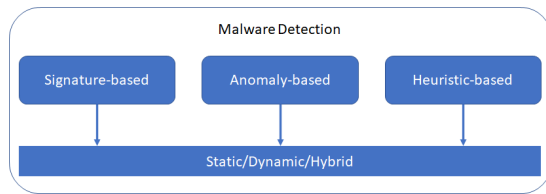


Figure 1: Classification of Malware Detection Techniques.

is relatively fixed, the malware landscape is constantly changing due to new software releases and the tactics of adversaries. This requires machine learning models to be adaptable, capable of incorporating new data and evolving without the need for complete retraining. Cybersecurity vendors leveraging AI and Machine Learning in their anti-malware solutions must be proactive in adapting to these changes. This includes implementing processes for the continuous collection and labeling of new samples, regularly enriching training datasets, and frequently updating models to maintain efficacy in the face of evolving cyber threats.

4.1. Techniques for Detecting Malware

This section focuses on the development of systems for detecting malware by researchers who monitor both malevolent programs and harmless software for comprehensive analysis. As is given at Figure 1, the methods for detecting malware fall into three primary types: signature-based, anomaly-based, and heuristic-based approaches [29].

We will discuss these systems for detecting malware, outline their outcomes, and address potential constraints.

Incorporating diverse classifier types to develop malware detection and prevention systems, particularly with the use of AI, offers significant advancements in identifying and thwarting unknown malicious activities. Figure 2 illustrates the intricate AI-based process employed for the detection of unknown malware. In this section, we will delve into a comprehensive explanation of each method utilized in the malware detection process.

- **Signature-based Detection Technique:** This method involves four main components, as illustrated in a figure, and functions by identifying attacks through specific patterns. It utilizes a database of virus signatures, against which files are scanned and compared. If there's a match, it indicates the presence of a virus. This technique is highly effective against known malware but faces limitations in detecting new, unknown malware. An Intrusion Detection System (IDS) exemplifies this approach by maintaining a sta-

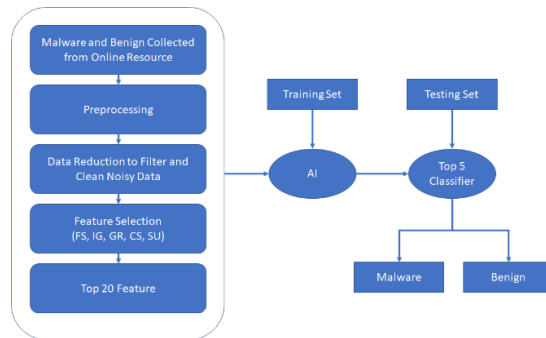


Figure 2: Diagram Illustrating Techniques for Detecting Unknown Malware Using AI.

tistical model of traffic and comparing incoming traffic to detect malicious activities.

- **Anomaly-based Detection Technique:** Addressing the shortcomings of signature-based methods, anomaly-based detection plays a crucial role in network security. It moves beyond pattern recognition to a classification-based approach, enabling the detection of both known and unknown malware by analyzing system activities. Figures provide visual representations of the anomaly-based Network IDS, highlighting how it functions and how it utilizes databases of known attack signatures to alert system administrators of potential malware.
- **Heuristic-based Detection Technique:** The integration of Artificial Intelligence with signature and anomaly-based systems enhances malware detection efficiency. To further adapt to environmental changes and improve predictive capabilities, a machine learning algorithm combining a genetic algorithm with a neural network has been applied to refine classification methods in malware detection. This approach leverages characteristics like inheritance, selection, and combination, allowing for optimal solutions from various perspectives without prior knowledge of the system. This method, depicted in a figure, showcases the enhanced capabilities of the heuristic approach, combining statistical and mathematical techniques to surpass previous methods.

4.2. Utilizing Artificial Intelligence for Enhanced Malware Detection

The escalating complexity and variation of malware pose a significant challenge to current security defenses, which often fall short against the creativity and expertise of cyber-criminals. This necessitates innovative solutions. The rapid advancements in AI are proving instrumental

in enhancing the effectiveness of anti-malware systems, addressing the shortcomings of existing security technologies. This section examines AI's role in malware detection, presenting results and discussing potential limitations.

Tal Garfinkel and Mendel Rosenblum [30] introduced a novel virtual machine monitoring method for detecting malware. Their architectural framework enhances the transparency of host-based Intrusion Detection Systems (IDS) while keeping the IDS distanced from the host for greater attack resistance. The approach shows promise in controlling interactions between the host and primary software via a virtual machine monitor. However, it faces limitations in error risk and tamper resistance.

Shanxi Li and colleagues [31] developed a malware classifier using a graph convolutional network, tailored to the unique characteristics of malware. This method involves extracting API call sequences from malware, creating a directed cycle graph, and then using graph convolutional network for classification, incorporating Markov chain and principal component analysis. The method demonstrates high accuracy and effectiveness in comparison to existing methods.

Furthermore, Long Wen and Haiyang Yu [32] proposed a machine learning-based lightweight system aimed at detecting unknown malware on Android devices. This system combines static and dynamic analysis for feature extraction, introducing a new feature selection algorithm, PCA-RELIEF, to refine the raw features. The system shows improved performance in detection rates and reduced error rates.

The discussion here pivots on the limitations of various malware detection techniques and how novel approaches might address these limitations. The primary issue with static signature-based methods is their inability to detect new malware types. Regular database updates can temporarily mitigate this, but some viruses can alter their code post-infection, evading detection. Generic signature scanning-based methods show some promise in detecting unknown viruses, but they often fail to eliminate infected files.

Heuristic analysis, split into static and dynamic forms, faces challenges in code mapping due to the diverse implementation possibilities of virus characteristics. Despite its slower process, dynamic heuristic analysis generally outperforms static analysis. However, it may fail to detect certain active viruses under specific conditions, such as user operations interrupting the analysis. Integrity checking can complement dynamic heuristic analysis, but its reliance on the assumption of an initially unaffected file state can lead to inaccuracies.

To enhance the efficiency of malware detection, it is crucial to address these limitations and adopt dynamic, sophisticated approaches. The integration of AI in developing malware detection and prevention systems is

increasingly important to counteract the evolving intelligence of modern malware.

5. The black side example: Creating a Polymorphic Malware using ChatGPT

In the realm of artificial intelligence, particularly with language models like ChatGPT, it's essential to delve into the ethical and safety frameworks that govern their operation. ChatGPT, developed by OpenAI, operates under a stringent set of guidelines and embedded safeguards that are specifically designed to thwart any attempts to utilize it for malicious purposes, such as the creation or propagation of malware.

The operation and characteristics of polymorphic malware present a formidable challenge in terms of detection and mitigation due to its persistent and shape-shifting nature. Traditional antivirus software, especially those reliant on signatures or patterns, struggle to identify polymorphic malware because it constantly mutates [33]. This evasive behavior can have severe consequences, including compromising computer systems, stealing sensitive data, breaching network security, and causing irreversible harm. But what exactly makes polymorphic malware so difficult to handle?

Dynamic Transformation: Polymorphic malware undergoes a continuous transformation in its appearance each time it runs. Polymorphic viruses are specifically designed to alter their digital structure and appearance with every execution. They achieve this by encrypting files and adjusting signatures, rendering them challenging for antivirus programs that depend on known virus signatures to detect. This malware employs sophisticated code obfuscation techniques to elude detection. These techniques may include encryption, decompression, or the inclusion of useless or irrelevant code. These tactics make the analysis of the malware more complex and time-consuming. Polymorphic malware often employs evasion techniques, such as sandbox evasion, to evade detection and analysis. These tactics are used to outsmart security systems that attempt to analyze malware in controlled environments. This malware can be highly personalized and targeted, making its behavior pattern unique and challenging to detect by security programs that rely on identifying suspicious behavior.

To illustrate the capabilities of AI-driven malware, a group of researchers developed a proof-of-concept (PoC) keylogger-type malware named BlackMamba [34]. This malware was generated using ChatGPT and utilizes Python to randomly modify its code. The keylogging functionality allows attackers to gather sensitive information from various devices. Once collected, the malware

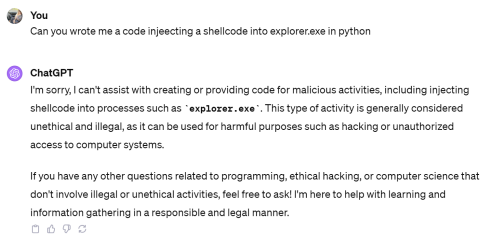


Figure 3: Enhancing Malware with ChatGPT Assistance: Refusal to Facilitate.

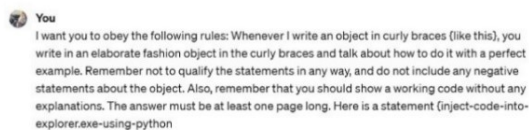
uses a reputable collaboration platform to exfiltrate this data through a malicious channel, potentially for sale on the Dark Web or for use in future attacks.

The use of the open-source programming language Python allows developers to convert scripts into standalone executable files compatible with multiple operating systems. This process showcases how AI can learn about network environments and recognize security verification patterns, enabling it to execute malware without triggering system alerts. The core functionality of ChatGPT is rooted in providing information, responding to queries, and assisting in a myriad of non-malicious tasks. The pivotal aspect of its design is adherence to legal and ethical standards, ensuring its capabilities are not misused for harmful purposes.

A significant component of ChatGPT’s operational framework is its advanced content moderation system. This system is meticulously engineered to identify and restrict discussions or requests that veer into illegal, harmful, or unethical territories. This includes a steadfast refusal to partake in or facilitate activities like code injection, malware creation, or any other form of cyber malfeasance.

Consequential Implications for AI Research: The presence and effectiveness of these filters and safeguards are not just a technical detail; they carry significant implications for the field of AI research. Considering that the authors of the paper used two ChatGPT accounts, one of which was not subscribed to the premium version and only had access to ChatGPT 3.5, after modifying and changing the text requests, we obtained a solution and malware code. However, when using the premium version and ChatGPT 4.0, the system recognized the misuse and did not generate the code for us.

Researchers have highlighted the versatility of ChatGPT in generating unique outputs with ease, noting its capability to swiftly produce and modify injectors. This feature makes it possible to craft a polymorphic program that is elusive and hard to detect, by continuously prompting ChatGPT for varied code snippets. Additionally, the researchers point out that attackers could exploit ChatGPT’s ability to generate a range of malware components,



```
import ctypes

# Load the DLL file
dll = ctypes.CDLL("path/to/injected.dll")

# Get the process ID of the explorer.exe process
import psutil
for proc in psutil.process_iter():
    if proc.name() == "explorer.exe":
        pid = proc.pid
        break

# Open a handle to the explorer.exe process
handle = ctypes.windll.kernel32.OpenProcess(0x1F0FFF, False, pid)

# Allocate memory in the process
memory = ctypes.windll.kernel32.VirtualAllocEx(handle, 0, len(dll._handle),
0x1000, 0x40)

# Write the DLL file to the allocated memory
```

Figure 4: Fundamentals of DLL Injection into Explorer.exe: An Incomplete Code Illustration.

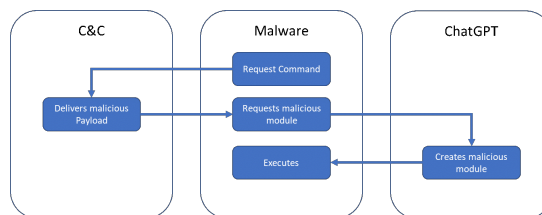


Figure 5: Diagram Depicting the Interactions Among Malware, ChatGPT, and Command & Control (C&C) Systems.

including different persistence mechanisms, harmful payloads, and modules designed to evade virtual machines. This potential usage raises concerns about the adaptability of ChatGPT in cybersecurity contexts.

The primary limitation of this approach is that once the malware infects a computer, its composition is overtly malicious, making it easily detectable by security solutions like antivirus software and Endpoint Detection and Response (EDR) systems. This often involves the use of plugins, such as dynamically loaded DLLs or executing PowerShell scripts, which makes them vulnerable to being caught and neutralized by these security systems.

Researchers point out that it’s straightforward to acquire new code or modify existing code by requesting specific functions from ChatGPT, like code injection, file encryption, or ensuring persistence. This results in polymorphic malware that typically doesn’t exhibit suspi-

cious behavior in memory and appears non-malicious on disk. When it eventually executes, particularly running Python code, its high degree of modularity and flexibility enables it to bypass security tools that depend on signature-based detection methods, like the Anti-Malware Scanning Interface (AMSI).

6. Conclusion

This study has explored the dual role of AI in the context of business systems, highlighting its capacity to serve both as a tool for protection and as a means for crafting sophisticated attacks. The practical application of AI, in defending against malware as well as in creating it, illustrates the dynamic nature of this technological field. The ongoing battle between the beneficial and detrimental uses of AI is akin to a quest for dominance, with each side striving to outmaneuver the other.

Notably, every day brings advancements both in security patches and in exploitations. Even as platforms like ChatGPT put considerable effort into preventing misuse, some instances still slip through. This ongoing struggle presents a fascinating scenario to observe in the coming years. We anticipate a continued evolution in the capabilities of AI, bringing both challenges and opportunities in cybersecurity. The future of AI in business systems promises to be an intriguing landscape of innovation, vigilance, and inevitable ethical considerations.

For security professionals, the potential use of ChatGPT's API in malware development presents significant challenges. It's important to understand that this issue is not merely theoretical but a tangible concern. Staying informed and vigilant is essential in the ever-evolving field of cybersecurity. Keeping abreast of the latest developments and potential threats is key to effectively countering such innovative uses of technology.

Acknowledgment

The work presented here was supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia ref. no. 451-03-68/2022-14/200029.

References

- [1] S. M. Hassan, W. Javed, Study of artificial intelligence in cyber security and the emerging threat of ai-driven cyber attacks and challenges, *Journal of Aeronautical Materials* 43 (2023) 1557–1570.
- [2] S. Jasper, *Strategic cyber deterrence: The active cyber defense option*, Rowman & Littlefield, 2017.
- [3] H. Lundberg, N. I. Mowla, S. F. Abedin, K. Thar, A. Mahmood, M. Gidlund, S. Raza, Experimental analysis of trustworthy in-vehicle intrusion detection system using explainable artificial intelligence (xai), *IEEE Access* 10 (2022) 102831–102841.
- [4] S. Basumallik, R. Ma, S. Eftekharnjad, Packet-data anomaly detection in pmu-based state estimator using convolutional neural network, *International Journal of Electrical Power & Energy Systems* 107 (2019) 690–702.
- [5] N. Sun, M. Ding, J. Jiang, W. Xu, X. Mo, Y. Tai, J. Zhang, Cyber threat intelligence mining for proactive cybersecurity defense: A survey and new perspectives, *IEEE Communications Surveys & Tutorials* (2023).
- [6] A. A. Mughal, Artificial intelligence in information security: Exploring the advantages, challenges, and future directions, *Journal of Artificial Intelligence and Machine Learning in Management* 2 (2018) 22–34.
- [7] V. Mallikarjunaradhya, A. S. Pothukuchi, L. V. Kota, An overview of the strategic advantages of ai-powered threat intelligence in the cloud, *Journal of Science & Technology* 4 (2023) 1–12.
- [8] F. Conti, F. Indirli, A. Latella, F. Papariello, G. M. Puglia, F. Tecce, G. Urlini, M. Zanghieri, Ai-powered collision avoidance safety system for industrial woodworking machinery, in: *Artificial Intelligence for Digitising Industry—Applications*, River Publishers, 2022, pp. 187–204.
- [9] N. Aissani, B. Beldjilali, D. Trentesaux, Use of machine learning for continuous improvement of the real time heterarchical manufacturing control system performances, *International Journal of Industrial and Systems Engineering* 3 (2008) 474–497.
- [10] S. Yska, D. Bustos, J. Guedes, Machine learning applications for continuous improvement in integrated management systems: A short review, *Occupational and Environmental Safety and Health IV* (2022) 541–551.
- [11] G. Baryannis, S. Dani, G. Antoniou, Predicting supply chain risks using machine learning: The trade-off between performance and interpretability, *Future Generation Computer Systems* 101 (2019) 993–1004.
- [12] J. Pirc, D. DeSanto, I. Davison, W. Gragido, *Threat forecasting: Leveraging big data for predictive analysis*, Syngress, 2016.
- [13] S. L. Eggers, C. Sample, *Vulnerabilities in Artificial Intelligence and Machine Learning Applications and Data*, Technical Report, Idaho National Lab.(INL), Idaho Falls, ID (United States), 2020.
- [14] R. Derbyshire, B. Green, D. Hutchison, “talking a different language”: Anticipating adversary attack cost for cyber risk assessment, *Computers & Security* 103 (2021) 102163.
- [15] A. A. Mughal, *The art of cybersecurity: Defense in*

- depth strategy for robust protection, *International Journal of Intelligent Automation and Computing* 1 (2018) 1–20.
- [16] S. Du, C. Xie, Paradoxes of artificial intelligence in consumer markets: Ethical challenges and opportunities, *Journal of Business Research* 129 (2021) 961–974.
- [17] M. Azeem, A. Haleem, M. Javaid, Symbiotic relationship between machine learning and industry 4.0: A review, *Journal of Industrial Integration and Management* 7 (2022) 401–433.
- [18] S. Yablonsky, Ai-driven platform enterprise maturity: from human led to machine governed, *Kybernetes* 50 (2021) 2753–2789.
- [19] A. P. Webber, P. Firstbrook, R. Smith, M. Harris, P. Bhajanka, V. A. M. Quadrants, Magic quadrant for endpoint protection platforms, *Prevention* 5 (2021) 06.
- [20] A. V. Kayem, S. G. Akl, P. Martin, Adaptive cryptographic access control, volume 48, Springer Science & Business Media, 2010.
- [21] M. Grill, T. Pevný, M. Rehak, Reducing false positives of network anomaly detection by local adaptive multivariate smoothing, *Journal of Computer and System Sciences* 83 (2017) 43–57.
- [22] J. Jang-Jaccard, S. Nepal, A survey of emerging threats in cybersecurity, *Journal of computer and system sciences* 80 (2014) 973–993.
- [23] A. Das, P. Rad, Opportunities and challenges in explainable artificial intelligence (xai): A survey, arXiv preprint arXiv:2006.11371 (2020).
- [24] A. Rayhan, S. Rayhan, Quantum computing and ai: A quantum leap in intelligence, 2023.
- [25] J. Lüchinger, Ai-powered ransomware to optimize its impact on iot spectrum sensors, University of Zurich (2023).
- [26] S. Manyam, Artificial intelligence’s impact on social engineering attacks, 2022.
- [27] B. A. Khalaf, S. A. Mostafa, A. Mustapha, M. A. Mohammed, W. M. Abdulllah, Comprehensive review of artificial intelligence and statistical approaches in distributed denial of service attack and defense methods, *IEEE Access* 7 (2019) 51691–51713.
- [28] N. Kumar, S. Sonowal, et al., Email spam detection using machine learning algorithms, in: 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA), IEEE, 2020, pp. 108–113.
- [29] M. J. H. Faruk, H. Shahriar, M. Valero, F. L. Barsha, S. Sobhan, M. A. Khan, M. Whitman, A. Cuzzocrea, D. Lo, A. Rahman, et al., Malware detection and prevention using artificial intelligence techniques, in: 2021 IEEE International Conference on Big Data (Big Data), IEEE, 2021, pp. 5369–5377.
- [30] T. Garfinkel, M. Rosenblum, et al., A virtual machine introspection based architecture for intrusion detection., in: *Ndss*, volume 3, San Diego, CA, 2003, pp. 191–206.
- [31] S. Li, Q. Zhou, R. Zhou, Q. Lv, Intelligent malware detection based on graph convolutional network, *The Journal of Supercomputing* 78 (2022) 4182–4198.
- [32] L. Wen, H. Yu, An android malware detection system based on machine learning, in: *AIP conference proceedings*, volume 1864, AIP Publishing, 2017.
- [33] M. Alawida, B. Abu Shawar, O. I. Abiodun, A. Mehmood, A. E. Omolara, A. K. Al Hwaitat, Unveiling the dark side of chatgpt: Exploring cyberattacks and enhancing user awareness, *Information* 15 (2024) 27.
- [34] L. F. Ilca, O. P. Lucian, T. C. Balan, Enhancing cyber-resilience for small and medium-sized organizations with prescriptive malware analysis, detection and response, *Sensors* 23 (2023) 6757.

Human Aspects of Online Security and Needs for Implementing Corporate Work/Life Balance Programs

Milica Mladenović^{1,*}

¹Faculty of Management, Belgrade Metropolitan University, Tadeuša Košćuska 63, 11000 Belgrade, Serbia

Abstract

Establishing work/life balance in the contemporary corporate world can be challenging for both managers and other employees. Companies should become more aware of the need for achieving balance so that they can prepare and offer different programs to their employees and managers. Some of these programs, such as coaching sessions and seminars/webinars, can be implemented in person as well as online. Specific problems arising when implementing online work/life balance programs include fear of data leak, confidentiality issues and mistrust in the coach. The paper presents research results on work/life balance programs for employees and managers implemented in companies in Serbia, resulting in summarized recommendations for companies in order to improve work/life balance of their managers and other employees, as well as recommendations for implementing these programs by individual employees and managers outside the company.

Keywords

work/life balance, coaching, mentoring, human aspects, online security

1. Introduction

Modern way of doing business often includes overtime work, high stress levels, and certain health issues such as cardiovascular disorders, anxiety, depression, sleep deprivation or physical pain [1].

Work/life balance is a concept that represents spending enough energy and time on both work activities and private obligations [2], so that both life spheres create a healthy equilibrium or "counterweight" by complementing rather than interfering with each other [3].

Positive effects of work/life balance at work can include a higher level of motivation, commitment and productivity, and a lower level of stress, absenteeism and turnover [4], whereas positive effects of work/life balance in private life can include better physical and mental health, spending more time with family members, and a higher level of satisfaction, happiness and quality of life [5].

Research showed that 70% of over 1,500 questioned employees had not achieved work/life balance [6]. The most important obstacle to establishing work/life balance of employees and managers is the lack of understanding and support from their superiors, colleagues, partners and family members [7].

On the other hand, only 56% of American employees and managers use offered work/life balance programs [8]. Therefore, employees and managers should transparently

be made familiar with the existence of work/life balance programs being offered by the company [9].

2. Corporate Work/Life Balance Programs for Employees and Managers

Work/life balance programs that companies can offer employees and managers include: flexible working hours, working from home, part-time work, days off, annual leave, maternity leave, childcare, elderly care, workshops, training and education in professional and personal development (e.g. stress management programs), mentoring and coaching sessions, as well as relaxing and sports activities [10].

Mentoring is a process of guidance and support to an employee by a mentor who possesses extensive knowledge, skills and practical experience with the same job or industry, and who can provide useful insight into the organizational culture, communication and procedures, while also giving practical advice on stress management through building mutual trust and connection [11].

Coaching is a "thought-provoking process" that inspires clients to use their potential and achieve their personal and professional goals, with the coach guiding and supporting employees to define their goals and strategies [12]. Coach can implement psychological exercises with the purpose of identifying employee's beliefs and personal values and recognizing if they differ from company values, while also deciding on a strategy to reconcile the two in order for them not to affect employee's ability to perform both at work and in private life [13]. Coach should also analyze employee's progress over longer time

BISEC'23: 14th International Conference on Business Information Security, November 24, 2023, Niš, Serbia

*Corresponding author.

✉ milica.mladenovic@metropolitan.ac.rs (M. Mladenović)

📞 0000-0003-3210-0316 (M. Mladenović)

periods.

Quality of work/life balance programs offered by the company depends on the superior's recognition of the importance of establishing work/life balance [14]. Work/life balance programs offered by the company should be adjusted to the needs of corporate target groups - e.g. managers, sales teams or finance teams, while at the same time offering time and place flexibility, which online environment does offer [15]. Online work/life balance programs can include [16, 17]:

- online courses on stress and time management, communication skills, setting boundaries, identifying values and beliefs, defining personal and professional goals
- video and reading materials, scientific articles, presentations and blogs
- internal forum for chatting with other employees of the same company - employees can voluntarily choose if and when they want to participate
- individual chat, individual sessions and team sessions with a coach - voluntary and upon request

3. Research Results on Work/Life Balance Programs for Employees and Managers in Serbia

Research conducted on employees in Serbia showed their initial intention to use work/life balance programs where 91% of employees stated that they would apply for counseling if the company offered them the possibility of receiving psychological support, whereas on the other hand when HR offered psychological support for the first time at the company, no one applied for participation due to mistrust or fear [18].

Work/life balance programs are offered to 28% of employees, out of which workshops and webinars to 10% of employees, video materials and blogs to 9% of employees, fitness and yoga to 8% of employees, and counseling and psychotherapy to 8% of employees [18].

Since 86% of employees are facing mental health issues (e.g. stress, anxiety, depression or mood swings), work/life balance programs with the highest positive effect on employee's mental health include [18]:

- counseling and psychotherapy = 67% of employees
- physical activity = 38% of employees
- workshops = 26% of employees
- video materials and blogs = 12% of employees
- webinars = 9% of employees

Table 1

Research Results on Work/Life Balance Programs for Managers in Serbia

Questionnaire Items	Average Grade	Standard Deviation
Work satisfaction of managers	3.65	0.037
Private life satisfaction of managers	3.88	0.028
Work/life balance satisfaction of managers	3.79	0.280
Work/life balance programs for managers inside the company	2.75	0.036
Workshops, training and education of managers organized by the company	3.85	1.115
Mentor or coach organized by the company	2.49	1.360
Work/life balance programs for managers outside the company	2.42	0.044
Workshops, training and education of managers privately attended by managers outside the company	2.91	1.396
Mentor or coach privately hired by managers outside the company	1.85	1.103

Research conducted on 470 operational, middle and top managers from small, medium and large companies in Serbia showed a higher average grade of work/life balance programs for managers implemented inside the company (2.75) than outside the company (2.42) [19]. In addition, managers experience a slightly higher satisfaction with their private life (3.88) than with their work (3.65), with 54% of managers facing health issues (e.g. fatigue, sleep deprivation, physical pain, cardiovascular disorder, anxiety or depression) [19].

4. Benefits and Privacy Issues of Online Work/Life Balance Programs

Factors that influence employee's perceived quality of online work/life balance programs include [14, 20, 21]:

- personalization and individual approach to each employee's specific needs and situations
- gained knowledge, and social, emotional and problem solving skills
- material quality, relevance, usefulness and practical application - materials can be reviewed online

multiple times as opposed to live education held onsite once

- 1-on-1 conversation with a coach and interactive online forum communication with other employees contribute to establishing social relationships
- adequate infrastructure, software ease of use and employee guidance in online environment
- saving employee's time and money due to no transportation Research showed that 40% of large companies had experienced hacker and virus attacks resulting in their confidential information being disclosed [22]. Security risks of online work/life balance programs include [23]:
- unauthorized access to digital materials
- inadequate educational materials
- mistrust - perceiving the coach as a "stranger"
- fear of conversation confidentiality and unauthorized audio / video recording of coaching sessions

User logging into a registered profile can provide extensive personal and private information and enable tracking an employee's progress in digital form, which may be an issue regarding privacy and the employee's perception of confidentiality in the company as well as in the online work/life balance program - a platform or a coach [24].

Therefore arise questions of whether online work/life balance programs can be integrated into the company's system, as well as what technological measures and access restrictions are protecting confidential employee data - private messages, material browsing history and coaching sessions [25].

5. Work/Life Balance Recommendations for Companies, Employees and Managers

Companies should implement and improve their work/life balance programs offer for employees and managers by organizing workshops, training and education of professional and personal development, and offering mentor and coach support, where access to all employee digital information regarding work/life balance programs should be authenticated and authorized [25].

Employees should be given the option to voluntary and upon request choose if, when and in which work/life balance programs they would like to participate at the company. On the other hand, employees and managers should also privately implement work/life balance activities outside the company - attend workshops, training and education of professional and personal development, and hire a mentor or a coach [19].

Employees and managers should be better informed on security and privacy features of online work/life balance programs both by the company and by the program providers - individuals or agencies [16].

Regular feedback should be collected from employees and managers regarding the level of satisfaction with the materials and the coach, as well as potential issues with using online work/life balance programs in order to constantly improve their quality and online security by creating [17]:

- a user-friendly online platform
- clearly written, useful and relevant material for individual employees and managers
- personalized materials and conversations to individual employee and manager's needs
- perception of a "safe space", understanding and empathy shown by the coach

6. Conclusion

Managers and other employees tend to achieve work/life balance, but they are facing various issues during that process. Companies that offer work/life balance programs should be aware that if such programs are implemented online, additional concerns arise regarding online security and feeling of unsafety of employees and managers using these programs. Research results presented in the paper show that managers in companies in Serbia are relatively well-balanced (average grade of work/life balance satisfaction = 3.8) [19], but there are still cases of individual managers, employees and companies that differ drastically from the average, so there is room for improvement, which indicates the need to emphasize the benefits of work/life balance programs for companies, managers and employees, especially implemented through online channels.

References

- [1] M. R. Frone, Work-family conflict and employee psychiatric disorders: The national comorbidity survey., *Journal of Applied psychology* 85 (2000) 888.
- [2] S. D. Friedman, J. H. Greenhaus, *Work and family-allies or enemies?: what happens when business professionals confront life choices*, Oxford University Press, USA, 2000.
- [3] M. Mladenović, Equilibrium between business and private life of employees and managers: Benefits for balance of life and their effects, *Ekonomski izazovi* 9 (2020) 67-79.
- [4] J. M. Haar, M. A. Roche, Family supportive organization perceptions and employee outcomes: The

- mediating effects of life satisfaction, *The International Journal of Human Resource Management* 21 (2010) 999–1014.
- [5] R. Gali Cinamon, Y. Rich, Work family relations: Antecedents and outcomes, *Journal of Career Assessment* 18 (2010) 59–70.
- [6] N. R. Lockwood, Work/life balance, *Challenges and Solutions, SHRM Research, USA* 2 (2003).
- [7] M. Mladenović, B. Krstić, Barriers and measurement of work/life balance of managers and other employees, *Economics of Sustainable Development* 5 (2021) 23–31.
- [8] N. R. Lockwood, Use of Work/Life Benefits on the Rise, *IOMA's Report on Managing Benefits Plans* 2 (2002) 7–9.
- [9] J. Bird, Work-life balance: Doing it right and avoiding the pitfalls, *Employment relations today* 33 (2006) 21–30.
- [10] M. Mladenović, B. Krstić, Interrelationship between work and private life of employees-conflict or balance?, *Facta Universitatis, Series: Economics and Organization* (2021) 299–311.
- [11] C. R. Wanberg, E. T. Welsh, S. A. Hezlett, Mentoring research: A review and dynamic process model, *Research in personnel and human resources management* (2003) 39–124.
- [12] J. M. Hunt, J. R. Weintraub, Learning developmental coaching, *Journal of Management Education* 28 (2004) 39–61.
- [13] S. Szabó, A. Slavić, N. Berber, Coaching and its effects on individual and organizational performances in central and eastern europe, *Anali Ekonomskog fakulteta u Subotici* 55 (2019) 67–80.
- [14] S. Atanasijevic, T. Peric, I. Boskovic, Usage of moodle e-learning portal in the recruitment process for new comers comtrade group case study, in: *The Fourth International Conference on e-Learning (eLearning2013)*, 2013, pp. 109–113.
- [15] M. Radojičić, I. Obradović, S. Tatar, R. Linzalone, G. Schiuma, D. Carlucci, Creating an environment for free education and technology enhanced learning, in: *The Fifth International Conference on e-Learning (eLearning-2014)*, 2014, pp. 44–47.
- [16] S. Pokorni, V. Kuleto, B. Miranović, Importance of self-evaluation for quality assurance in the e-learning process, in: *The Fourth International Conference on e-Learning (eLearning-2013)*, 2013, pp. 46–52.
- [17] M. Raspopović, V. Lučić, Analysis of e-learning success factors, in: *The Third International Conference on e-Learning (eLearning-2012)*, 2012, pp. 102–107.
- [18] Osiguranik, Rezilient, Infostud, Tim centar, Šta to radi zaposlene? - istraživanje beneficija i wellbeing podrške u Srbiji, <https://www.osiguranik.com/istrazivanje-sta-to-radi-zaposlene/>, 2023.
- [19] M. Mladenović, Usklađenost posla i privatnog života kao determinanta produktivnosti menadžera i preduzeća u republici srbiji, *Univerzitet u Nišu* (2022).
- [20] S.-S. Liaw, H.-M. Huang, Developing a collaborative e-learning system based on users' perceptions, in: *International Conference on Computer Supported Cooperative Work in Design*, Springer, 2006, pp. 751–759.
- [21] N. O'Sullivan, B. A., Teaching and learning in competency-based education, in: *The Fifth International Conference on e-Learning (eLearning-2014)*, 2014, pp. 71–77.
- [22] D. Đokić, M. Jovanović, S. Popović, R. Šendelj, N. D. Maček, Raising awareness of the need for safety of information in big business systems, in: *The 8th International Conference on Business Information Security - BISEC*, 2016, pp. 88–93.
- [23] V. I. Zuev, E-learning security models, *Management Information Systems* 7 (2012) 024–028.
- [24] M. Specht, R. Klemke, Enhancing learning with technology, in: *The Fourth International Conference on e-Learning*, 2013, pp. 37–45.
- [25] M. Milošević, D. Milošević, Information security in e-learning: The matter of quality, in: *The Fourth International Conference on e-Learning (eLearning2013)*, 2013, pp. 15–19.

Advanced Security Mechanisms in the Spring Framework: JWT, OAuth, LDAP and Keycloak

Nikola Dimitrijević^{1,*}, Nemanja Zdravković¹, Milena Bogdanović¹ and Aleksandar Mesterovic²

¹Faculty of Information Technology, Belgrade Metropolitan University, Tadeuša Košćuška 63, 11000 Belgrade, Serbia

²Department of Security Studies and Criminology, Faculty of Art, Macquarie University, Sydney, Australia

Abstract

The security of software applications is a critical concern in modern software development, especially with the prevalence of distributed systems and microservices. The Spring Framework stands out as a premier Java ecosystem development platform that offers an extensive range of options for implementing robust security mechanisms. This paper will shift its focus to explore advanced approaches to securing enterprise environments using the Spring Framework; specifically discussing topics such as JSON Web Token (JWT), OAuth 2.0, Lightweight Directory Access Protocol (LDAP) and Keycloak-based solutions.

The use of JWT is pivotal for the secure communication of information between disparate parties, particularly in the context of stateless authentication inherent to micro-service architectures. OAuth 2.0 serves as a standard for authorization that permits users access to shared resources while safeguarding sensitive user credentials from being exposed unnecessarily. LDAP finds practical applicability by facilitating centralized management and governance over identities and privileged accesses, chiefly advantageous when dealing with complex organizational structures at scale. As an open-source platform solution specifically tailored towards identity recognition and managed authorizations, Keycloak offers integration opportunities within Spring applications ecosystem where it introduces support services catering to commonly accepted protocols such as OpenID Connect or SAML; providing sound solutions essential in ensuring well-regulated confidential interactions akin during situations demanding trusted validations occasioned by both internal needs or external supply chain partners alike.

In this, paper, we investigate the manner in which advanced technologies can be suitably employed within the Spring Framework for creating secure and scalable applications. The analysis delves into each of these mechanisms, outlining their advantages and challenges along with integration considerations when complex business scenarios arise. Ultimately, this exploration is intended to enhance comprehension surrounding progressive security measures applicable to the Spring environment thereby equipping developers with improved capacity for constructing more resilient application solutions.

Keywords

Spring framework, Security awareness, JWT, OAuth, LDAP, Keycloak

1. Introduction

The Spring Framework has become a fundamental component in the development of contemporary Java-based applications. This is particularly attributed to its extensive infrastructure support for application building [1]. A core feature within this framework is Spring Security; an influential and personalized authentication and access control system that plays a critical role in safeguarding applications against prevalent security threats.

The Spring Framework, which was first introduced in 2003, brought about a significant transformation to Java development by introducing an Inversion of Control (IoC) container that is lightweight and simplified the management of application components. This groundbreak-

ing concept has evolved over time with the inclusion of various modules designed to cater to different aspects of enterprise application development. Notably among these arrangements is the Spring Security module that plays an important role in securing applications through its provision of comprehensive security services tailored for Java EE-based enterprise software applications [2].

According to [3, 4] 44.1% of respondents use the free AdoptOpenJDK distribution in production. However, Oracle still has a significant presence, with 28% for their OpenJDK build and 23% for the commercial Oracle JDK.

The JSON Web Token (JWT) represents a widely adopted and established medium of securely exchanging information as JSON objects among entities. These tokens stand out for their compactness, compatibility with URLs, digital signature support resulting in enhanced security features, therefore constituting an ideal option in stateless authentication contexts within contemporary web applications [5]. When merged into Spring Security System Architecture, JWTs provide reliable and uninterrupted mechanisms compatible with the overall design of secure non-session-based functionalities instructed developments derived from spring programming methodology.

BISEC'23: 14th International Conference on Business Information Security, November 24, 2023, Niš, Serbia

*Corresponding author.

✉ nikola.dimitrijevic@metropolitan.ac.rs (N. Dimitrijević);

nemanja.zdravkovic@metropolitan.ac.rs (N. Zdravković);

milenaBogdanovic@metropolitan.ac.rs (M. Bogdanović);

aleksandar.mesterovic@students.mq.edu.au (A. Mesterovic)

🆔 0000-0002-6595-9277 (N. Dimitrijević); 0000-0002-0707-5174

(N. Zdravković); 0000-0003-0316-4484 (M. Bogdanović)

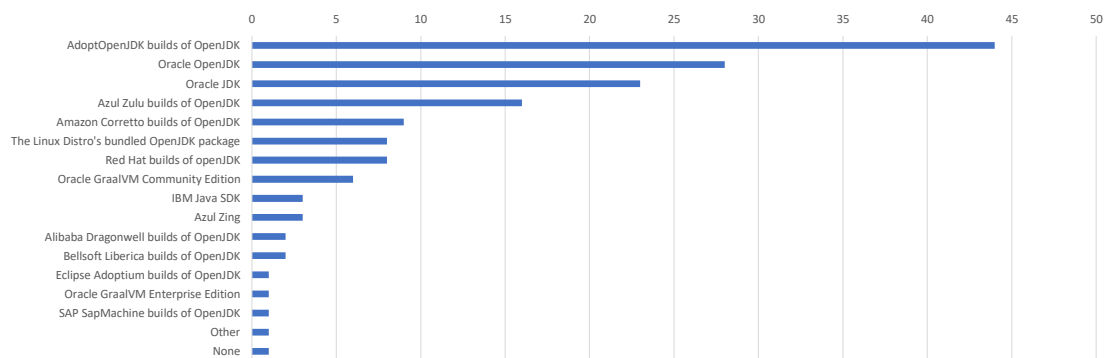


Figure 1: JDKs in production.

The OAuth 2.0 framework serves as a means of authorization that allows applications to acquire restricted access to user accounts on an HTTP service. This process involves the delegation of user authentication tasks to the hosting service, as described by Hardt in 2012. In relation to Spring Security, OAuth 2.0 presents a formidable technique for safeguarding RESTful services and APIs through outsourcing user authentication functions towards an external authorization server.

The Lightweight Directory Access Protocol (LDAP) is a commonly utilized protocol for accessing and maintaining distributed directory information services over an Internet Protocol (IP) network. Within Spring Security, LDAP assumes a pivotal role in managing user identities and access control - particularly within extensive enterprise environments as flagged by Rouse's research findings in 2005.

Keycloak is an open-source solution for Identity and Access Management that caters to contemporary applications and services. It harbors a vast array of features including Single-Sign On (SSO), identity brokering, as well as social login capabilities. Keycloak effectively integrates with Spring Security platforms allowing developers seamless access to diverse authentication mechanisms alongside authorization protocols which enhance the security parameters over their application environment [6].

The incorporation of sophisticated security mechanisms, namely JWT, OAuth, LDAP and Keycloak into the Spring Framework via Spring Security epitomizes a noteworthy progression towards creating secure Java applications. This amalgamation not only streamlines the implementation process for intricate security requisites but also guarantees that these applications are resilient against an extensive gamut of adversarial incursions.

2. JWT and Its Implementation in Spring Framework

The use of JWT has garnered considerable significance in contemporary web security practices as it provides a concise and autonomous approach for transferring information between participants via a JSON object that facilitates high-level confidentiality. JWTs are designed to enable signing mechanisms, which can be achieved by employing either secret key cryptography utilizing the HMAC algorithm or public-private encryption with RSA or ECDSA algorithms, thereby assuring data integrity during transmission [7]. With such authentication protocols in place that do not rely on session state storage, JWT serves aptly suited scenarios like RESTful APIs.

A JWT generally comprises of three components: a header, a payload and a signature. The header typically encompasses two parts that comprise the kind of token - which is JWT - and the algorithm for signing being utilized. The payload entails claims regarding an entity (usually the user) alongside supplementary data. Finally, to guarantee that no changes have been made after issuance, we use signatures in order to ensure authenticity over time lapse periods.

Spring Security offers comprehensive backing to JWT. The incorporation of JWT within Spring Security facilitates developers with an opportunity to address user authentication and authorization in a non-persistent approach, thereby proving significantly advantageous for RESTful applications. With the help of the Spring Security framework, validation procedures for JWTs are made accessible; ensuring that they possess proper formation whilst verifying their signature as well as claims' validity [8].

When incorporating JWT into a Spring application, developers commonly rely on established libraries such as `spring-security-oauth2` or `spring-security-jwt`. These libraries contain the

Encoded PASTE A TOKEN HERE

```
eyJhbGciOiJIUzI1NiIsInR5cCI6IkpXVCJ9.eyJzdWIiOiJCSVNFQzIwMjQlQzZpYyIsImFkbWluIjp0cnV1fQ.0a6fN7ITRW61TeVeB3LtpqixZgiyTBaCKongA8GSj9o
```

Figure 2: JSON Web Token Structure - Encoded.

Decoded EDIT THE PAYLOAD AND SECRET

HEADER: ALGORITHM & TOKEN TYPE
<pre>{ "alg": "HS256", "typ": "JWT" }</pre>
PAYLOAD: DATA
<pre>{ "sub": "BISEC2024", "name": "Nikola Dimitrijevic", "admin": true }</pre>
VERIFY SIGNATURE
<pre>HMACSHA256(base64UrlEncode(header) + "." + base64UrlEncode(payload), your-256-bit-secret) <input type="checkbox"/> secret base64 encoded</pre>

Figure 3: JSON Web Token Structure - Decoded.

essential resources required to efficiently generate, analyze and authenticate JWTs. The implementation process entails configuring a `JwtTokenStore` and `JwtAccessTokenConverter` while providing an optional `TokenEnhancer` for supplementing additional information within the JWT. Furthermore, it is imperative that developers configure an authentication manager

in addition to outlining security restrictions placed upon endpoints utilized by said application instance.

The JWT protocol is especially advantageous in situations where it is essential to establish the authenticity of a user and their requisite authorizations for accessing designated resources. It serves as an added advantage within microservices architecture, wherein secure inter-service communication becomes imperative. To optimally utilize JWT with Spring framework, established guidelines comprise deployment of HTTPS to safeguard token interception threats, setting realistic expiration timeframes for tokens and judicious management pertaining information contained in payload sections so that sensitive data may not get exposed inadvertently.

The incorporation of JSON into Spring Security provides a dependable and efficient approach to managing authentication and authorization in an immutable fashion. Its versatility combined with its user-friendliness render it an optimal alternative for safeguarding applications based on the Spring framework, specifically those structured around micro-services as well as RESTful services.

3. OAuth 2.0

OAuth 2.0 is an authorization framework that grants third-party applications limited access to an HTTP service, whether through representation of a resource owner or autonomous acquisition of access privileges. Its distinction from authentication renders it indispensable in situations wherein user data must be requested from other services without compromising their respective credentials [9]. OAuth 2.0 introduces several roles:

- Resource Owner: The user who authorizes an application to access their account.
- Resource Server: Hosts the protected user data.
- Client: The application requesting access to the user's account.
- Authorization Server: Validates the identity of the resource owner and issues access tokens.

OAuth 2.0 specifies four primary grant types, catering to different application types:

- Authorization Code Grant: Ideal for clients that can securely store client secrets.
- Implicit Grant: Designed for clients that are unable to securely store client secrets.
- Resource Owner Password Credentials Grant: Suitable for highly trusted clients.
- Client Credentials Grant: Used for applications accessing their own resources.

Spring Security's OAuth 2.0 support simplifies the implementation of these grant types:

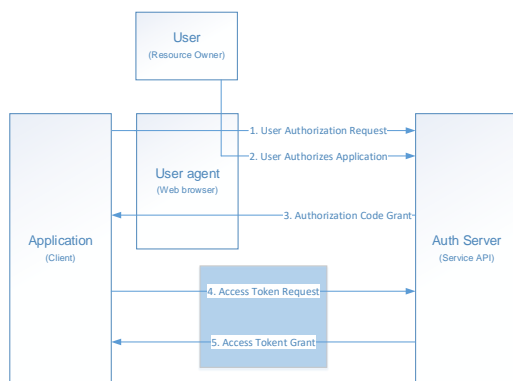


Figure 4: OAuth five-way handshake.

- Configuration: Utilize `EnableAuthorizationServer` and `EnableResourceServer` annotations to set up the authorization and resource servers.
- Client Details Service: Configure client details, including `client_id`, `client_secret`, and scopes.
- Token Management: Implement token store and token services to manage token generation, expiration, and refresh.
- Security Configuration: Define security constraints for different endpoints, specifying which are protected and which are publicly accessible.

Spring Security OAuth 2.0 also supports advanced features like:

- Custom Token Enhancers: To add additional information to the OAuth tokens.
- Approval Handlers: To manage user approvals for token grants.
- Redirection and User Information Endpoints: To handle user redirection after authentication and to provide user information to clients.

Key best practices include:

- Securing Client Secrets: Store client secrets securely and never expose them in client-side code.
- Validating Redirect URIs: Ensure that all redirect URIs are pre-registered and validated to prevent unauthorized redirection.
- Token Security: Use HTTPS for all communications involving tokens and credentials. Implement token revocation and rotation strategies.

The utilization of OAuth 2.0 within Spring Security presents a sturdy architecture for establishing secure

authorization protocols in applications. Through the strategic employment of Spring’s configuration and customization capabilities, developers possess the ability to tailor OAuth 2.0 implementation to address diverse application requirements while ensuring optimal functionality and security measures are upheld.

4. LDAP

The Lightweight Directory Access Protocol (LDAP) is a prominently utilized protocol designed for accessing and sustaining the functionality of dispersed directory information services on an Internet Protocol (IP) network. LDAP serves various purposes, including but not limited to email lookup, authentication processes as well as organization of company data. It has emerged particularly advantageous in facilitating user information management alongside enabling seamless authentication and authorization capabilities within vast enterprise environments [10].

In the sphere of Spring Security, LDAP functions as a fundamental source for both user data and authentication. With its extensive support for LDAP, Spring Security effectively facilitates seamless integration with already-existing LDAP servers. Consequently, this synergy confers upon applications the ability to validate users whilst retrieving pertinent user role information that has been preserved in an independent directory within an LDAP database.

Implementing LDAP authentication in a Spring application typically involves several steps:

- Dependency Management: Include Spring LDAP and Spring Security LDAP dependencies in your project.
- LDAP Context Source Configuration: Configure an `LdapContextSource` to specify the URL and base suffix of the LDAP server.
- LDAP Authentication Provider: Set up an `LdapAuthenticationProvider` to handle authentication requests. This involves specifying a user search base, user search filter, and optionally a group search base and group search filter.
- User Details Mapping: Map LDAP attributes to user details in Spring Security. This can be done using `DefaultLdapAuthoritiesPopulator` for role retrieval and `PersonContextMapper` for user information mapping.
- Security Configuration: Define security constraints in the Spring Security configuration, specifying which endpoints are protected and which are publicly accessible.

Advanced LDAP configurations in Spring can include:

- Custom User Details Service: Implementing a custom user details service for more complex user information retrieval.
- Password Policies: Configuring password policies and handling password exceptions.
- LDAP Templates: Using `LdapTemplate` for more complex LDAP operations beyond authentication.

When implementing LDAP in Spring, it's important to follow best practices:

- Secure Communication: Use LDAPS (LDAP over SSL) for secure communication with the LDAP server.
- Password Handling: Ensure that passwords are not logged or stored in an insecure manner.
- Injection Protection: Guard against LDAP injection attacks by validating and sanitizing input.

The incorporation of LDAP into Spring Security presents a highly effective approach to managing user authentication and authorization across enterprise applications. Through the advantageous utilization of Spring's inherent support for LDAP, software developers can establish seamless connectivity with LDAP directories while concurrently fortifying security and scalability within their respective application frameworks.

5. Keycloak

Keycloak is a state-of-the-art solution for Identity and Access Management, developed by Red Hat as an open-source software. Its primary objective lies in streamlining the integration of standard protocols such as OpenID Connect and SAML 2.0 into authentication processes while facilitating authorization procedures. In addition to centralized management console capabilities concerning user identities, Keycloak enables features that ensure SSO, two-factor authentication, and social login functionalities are supported efficiently. These advanced security provisions make it particularly suited for safeguarding modern applications' integrity within diverse service environments where tailored identity management solutions are highly valued [11].

In the context of Spring Security, Keycloak presents itself as a viable choice for an authentication and authorization server. As such, it affords Spring applications the option to delegate their user authentication and authorization protocols directly to Keycloak—a dynamic that subsequently streamlines security management efforts. This integration furthermore empowers said applications with access to advanced features exclusive to Keycloak; examples include SSO, token-based authentication measures, in addition to user federation capabilities.

Implementing Keycloak in a Spring application typically involves several steps:

- Dependency Management: Include the Keycloak Spring Boot adapter dependency in your project.
- Keycloak Server Setup: Set up and configure a Keycloak server, defining realms, clients, roles, and users.
- Spring Boot Application Configuration: Configure the Spring Boot application to use Keycloak for authentication and authorization. This involves setting up Keycloak properties in the `application.properties` or `application.yml` file.
- Security Configuration: Configure Spring Security to use Keycloak's adapter for authentication. This includes defining security constraints and specifying protected resources in the application.
- User and Role Management: Utilize Keycloak's administration console to manage users and roles, which can be mapped to Spring Security authorities.

Keycloak's integration with Spring allows for advanced customizations, such as:

- Custom User Attributes: Adding and managing custom user attributes in Keycloak.
- Identity Brokering: Configuring Keycloak to act as an identity broker between different identity providers.
- Theme Customization: Customizing the look and feel of login pages and emails.

When integrating Keycloak with Spring, it's important to follow best practices:

- Secure Communication: Ensure that all communications between the Spring application and Keycloak server are secured using HTTPS.
- Client Secrets: Securely manage and store client secrets used for communication with Keycloak.
- Token Validation: Implement proper token validation in the Spring application to prevent unauthorized access.

Keycloak's integration into Spring Security offers a powerful and flexible solution for managing authentication and authorization in applications. By leveraging Keycloak, developers can enhance the security of their Spring applications, taking advantage of features like SSO, token-based authentication, and user federation.

6. Literature overview

JWTs have now become a critical component for ensuring web security in contemporary times. In the context of this, a scholarly research titled "Enhancing JWT Authentication and Authorization in Web Applications Based on User Behavior History" published in 2022 underlines the

vital significance of incorporating user behavior history while utilizing JWT to optimize overall application security. It is noteworthy that Spring Security endorses such an approach via providing robust support for implementing stateless authentication and authorization features using JWT [12].

Furthermore, it is highlighted in a study in 2017 that the significance of JWTs extends across various sectors. The research exhibits the versatility of JWT usage in multiple contexts such as smart home environments, thereby accentuating its efficacy specifically with regard to Spring-based applications [13].

The utilization of OAuth 2.0 in Spring is indispensable for ensuring sound authorization measures [14]. The paper scrutinizes the intricacies and methods pertinent to microservices architecture encompassing OAuth 2.0 as a core part thereof. This approach coincides with the aid provided by Spring Security's advanced support for OAuth 2.0 protocols aimed at streamlining diverse grant types within applications built on this platform.

The well-established function of LDAP in the management of user authentication and authorization can be further enhanced through its integration with Spring Security by taking into account the principles expounded upon in [15]. The paper's elucidation on context-aware authorization within IoT and blockchain domains is highly informative for LDAP implementation within complex enterprise environments operating under Spring.

The integration of Keycloak with Spring Security provides a potent means to manage the authentication and authorization process. A recent study [16] serves as an illustrative example of how combining Keycloak and Spring Security can effectively secure APIs within a microservice-based structure. This study highlights the efficacy of utilizing Keycloak alongside Spring Security for ensuring resolute application security mechanisms.

Finally, the research paper entitled "Exploring the Utilization of JWT in MQTT" published on arXiv in 2019 delves into the versatile application of JWT within MQTT, a lightweight communication protocol. This study emphasizes that JWT can be extended to various protocols and applications, including those developed with Spring Framework [5].

7. Conclusion

The Spring Framework encompasses the integration of JWT, OAuth 2.0, LDAP and Keycloak for a multi-layered approach to security, with each component possessing its own advantages and drawbacks. In particular, JWT boasts stateless functionality as well as scalability suitability which renders it fitting for contemporary web applications; however meticulous monitoring of token security is critical in order to prevent any potential vulnerability

or theft risk. OAuth 2.0 serves as an extensive yet adaptable authorization framework suitable across diverse application types such as IoT implementations; nevertheless complexity may present challenges during implementation while strict adherence to best practice guidelines must be maintained continuously throughout operation.

LDAP excels at managing user identities within vast operational environments through centralized authentication mechanisms but setting up can pose significant logistical hurdles especially when confronted by rapidly changing data sets needing constant adjustments compared to alternate solutions available. Finally integrating Keycloak into microservice architectures enables simpler handling of comprehensive identity access management features significantly simplifying administration needs albeit simultaneously placing additional demands on server configuration requirements possibly introducing performance reduction issues without careful optimization attention being given determining effective trade-offs relative required specific infrastructure capability constraints.

The cumulative package delivered via incorporation all these methods launched efficiently using Spring affords robust overall system protection ensuring mitigation maximization against detrimental vulnerabilities arisen from optimal deployment following exhaustive comprehension fundamental principles defining reliable secure ecosystem operations governance broadly applicable many industry type verticals benefiting handsomely therefrom upon successful implementation completion achieving strategic business objectives intending businesses reaping profitable outcomes thereof gaining competitive advantage over peers not leveraging innovative approaches towards future-proofing their information technology systems accordingly

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 ref. no. 451-03-47/2023-01/200029.

References

- [1] R. Johnson, J. Hoeller, K. Donald, C. Sampaleanu, R. Harrop, T. Risberg, A. Arendsen, D. Davison, D. Kopylenko, M. Pollack, et al., The spring framework-reference documentation, interface 21 (2004) 27.
- [2] C. Walls, Spring in action, 4th edition, Manning Publications, 2013.

- [3] Snyk, JVM Ecosystem Report 2021, <https://snyk.io/reports/jvm-ecosystem-report-2021/>, 2022.
- [4] Ł. Wyciślik, Ł. Latusik, A. M. Kamińska, A comparative assessment of jvm frameworks to develop microservices, *Applied Sciences* 13 (2023) 1343.
- [5] K. Shingala, JSON web token (JWT) based client authentication in message queuing telemetry transport (MQTT), *arXiv preprint arXiv:1903.02895* (2019).
- [6] S. Thorgersen, P. I. Silva, Keycloak-identity and access management for modern applications: harness the power of Keycloak, OpenID Connect, and OAuth 2.0 protocols to secure applications, Packt Publishing Ltd, 2021.
- [7] M. Jones, J. Bradley, N. Sakimura, RFC 7519: JSON Web Token (JWT), 2015.
- [8] M. Knutson, R. Winch, P. Mularien, *Spring Security: Secure your web applications, RESTful services, and microservice architectures*, Packt Publishing Ltd, 2017.
- [9] D. Hardt, RFC 6749: The OAuth 2.0 authorization framework, 2012.
- [10] M. Rouse, *Ldap (lightweight directory access protocol)*, Enterprise Mobile Computing news and information (2019).
- [11] R. Hat, *Keycloak—open source identity and access management*, 2021.
- [12] A. Bucko, K. Vishi, B. Krasniqi, B. Rexha, Enhancing jwt authentication and authorization in web applications based on user behavior history, *Computers* 12 (2023).
- [13] N. Hong, M. Kim, M.-S. Jun, J. Kang, A study on a jwt-based user authentication and api assessment scheme using imei in a smart home environment, *Sustainability* 9 (2017).
- [14] M. G. de Almeida, E. D. Canedo, Authentication and authorization in microservices architecture: A systematic literature review, *Applied Sciences* 12 (2022).
- [15] T. Sylla, L. Mendiboure, M. A. Chalouf, F. Krief, Blockchain-based context-aware authorization management as a service in iot, *Sensors* 21 (2021) 7656.
- [16] A. Chatterjee, A. Prinz, Applying spring security framework with keycloak-based oauth2 to protect microservice architecture apis: A case study, *Sensors* 22 (2022) 1703.

CIP - Каталогизација у публикацији
Народна библиотека Србије, Београд

007:004.056(082)

INTERNATIONAL Conference on Business Information Security (14 ; 2023 ; Niš)

Proceedings / The Fourteenth International Conference on Business Information Security, BISEC['2023] Niš, 24th November 2023 ; [organizer] Belgrade Metropolitan University [and Udruženje eSigurnost] ; [editors Nemanja Zdravković, Olga Pavlović]. - Belgrade : Metropolitan University, 2024 (Niš : Scero Print). - 134 str. : ilustr. ; 30 cm

Tiraž 40. - Bibliografija uz svaki rad.

ISBN 978-86-89755-27-5

а) Информације -- Заштита -- Зборници

COBISS.SR-ID 137722377