CYSP 2023

THE SECOND CONFERENCE ON

CYBERSPACE

Proceedings

دهمین سال

تأسیس دانشکده مهندسی دانشکدگان فارابی دانشگاه تهران

قم، ابتدای جاده قدیم تهران، دانشکدگان فارابی دانشگاه تهران، دانشکده مهندسی

تلفن: ۰۲۵-۳۶۱۶۶۶۵۱

Faculty of Engineering, College of Farabi, University of Tehran, Old Qom-Tehran Road, Qom, Iran

Phone Number:(+98-25)36166651

Website: http://cysp2023.ut.ac.ir

E-mail: cysp.conf@ut.ac.ir

Instant Messengers: @cysp_conf

بسم الله الرحمن الرحیم

IN THE NAME OF ALLAH

CYSP 2023

The Second Conference on Cyberspace

Proceedings

ISC Islamic World Science Citation Center 02230-21686

University of Tehran

31 Oct., 1-2 Nov. 2023

.

Proceedings of
The First Conference on Cyberspace (CYSP 2023)
Organizer: University of Tehran

Preparation: *Kazim Fouladi-Ghaleh*
with: *Hussein Azimi, Alireza Zeini*

Publisher: Faculty of Engineering, College of Farabi, University of Tehran
Printing: Matris Publishing Co.
Publishing Year: 2023

Secretariat Address: Faculty of Engineering, College of Farabi, University of Tehran, Old Qom-Tehran Road, Qom, Iran, Postal Code: 3718117469.
Telephone: 025-36166651
Fax: 025-36166652
E-mail: `cysp.conf@ut.ac.ir`
Web: `http://cysp2023.ut.ac.ir`
Messengers: `@cysp_conf`
All links in one: `http://conf.cysp.ir/links`

Table of Papers

Table of Authors

.

.

English Full Papers

Improving Cybersecurity Systems Using Artificial Intelligence Techniques

Ismail A. Humied[1]

[1]*Associate Professor, Faculty of Police, Policy Academic, Ministry of Interior, Sana'a, Yemen*
dr.ismail_humied@yahoo.com

Abstract

Cyberattacks and threats are in complexity every day. With the number of attacks growing every day around the world, there are a variety of methods and strategies for penetrating business systems and individual devices. Individual attackers, organizations, and whole nations are responsible for these attacks. Attack resources are growing and cyberattacks can have severe global impact and consequences. These variables make it difficult for security teams to keep pace, so smarter solutions are needed. This research provides an overview of how two areas of artificial intelligence (AI), machine learning and deep learning, can be used to meet cybersecurity challenges. The research can also show how existing AI technologies can improve cybersecurity. Security systems can use AI to automate time-consuming manual security operations and make detection and response more proactive and predictive. As part of this research, a workshop was held with official from government and private agencies, who are responsible for network monitoring and security. During this workshop, a discussion occurred what problems individuals had at work. Thus, this research includes potential AI-based remedies for identifying issues.

Keywords: *Artificial Intelligence, Machine Learning, Deep Learning, Cybersecurity, Cyber-attacks, Cybercrime, Detect.*

1 Introduction

Computers are indispensable in today's workplace and daily life. The development and spread of modern technology have increased the need for information security. The amount of data collected is enormous and is constantly being driven by commercial, military, financial, medical, and government applications. This highlights the importance of cyber security. The term "cybersecurity" is widely used and is an issue that we must constantly address and move forward. Cybersecurity is the process of protecting data and information on networks, mobile devices, computers, and other electronic devices

or connections. It is sometimes called computer security or information technology security. In addition to preventing harmful attacks, unauthorized access, and other harm, the objectives are to guarantee the data's integrity, availability, and confidentiality [1, 2].

Most of our computer systems and network infrastructure are connected via the Internet. Virtually all businesses, governments, and even individuals today rely on cybersecurity to protect their data, grow their businesses, and protect their personal information. People send and receive data over the network infrastructure. Routers are vulnerable to external hacking and tampering. Big data has emerged as a result of increased data volume and complexity due to increased internet usage. Due to the constant growth of the Internet and the amount of data it contains, it has become necessary to develop a reliable intrusion detection system. The network security subset of cybersecurity protects networked systems from unwanted activity. Networked computers become available to ensure data confidentiality, integrity, and accessibility. Current cybersecurity research focuses on developing reliable intrusion detection systems that can detect both known and new attacks and threats with high accuracy and minimal false alarm rates [3].

The most commonly used rule-based systems today are used to implement cybersecurity. The method used to teach a system rule to understand how to process, store, and sort data is called a rule-based system. The system manufacturer or vendor implements the rules. Rules can often be changed by updates. In the event of an attack, the system searches through a set of rules to find the correct answer. If no action is taken to counter a particular attack, the system should be shut down. The designer then has to detect the problem and manually fix the system with a patch or software upgrade [4, 5].

Rule-based systems are not amenable to modification and customization, resulting in a very high rate of new attacks and variants of the same attack. The tedious remediation process of these stubborn attacks takes a lot of time and effort, reducing production and efficiency. Addressing these issues requires designing systems that can adapt to their environment, learn from experience, and change rules to counter future adversarial attempts. This means that the system can patch itself and figure out how to fix vulnerabilities on its own. Additionally, the system can track past attacks and rebuild the system based on newly created rules [6].

To solve the above problems, cybersecurity can use artificial intelligence (AI). Cybersecurity products based on artificial intelligence have become more popular and have evolved rapidly over the past decade. The expansion of these technologies will improve the effectiveness of cybersecurity-related tasks and reduce the frequency and risk of security breaches [7, 8].

The purpose of this research is to show how cybersecurity can be advanced by systems and techniques based on artificial intelligence. There are many facets to AI, and this study does not explore them all. Moreover, there are so many types of attacks that it's impossible to cover them all here. The main purpose is to see real systems already

in place and how they can be used for network security. Intrusion detection systems are getting better at detection and remediating anomalous activities and potential threats as more AI techniques are integrated into computer security. Finally, the study seeks to shed light on the obstacles hindering the continued growth of the use of AI tools in cybersecurity and the prospects for AI in this area.

This research starts by descript machine learning techniques and deep learning techniques. In the next section, the presents analysis, possible solutions and current techniques from the workshop conducted and the individual meetings of participants after that are presented. Finally presents the discussion and conclusion.

2 Methodology

Literature research and workshop results were used as research methods in this research. A general literature review was conducted to learn more about the field of AI and cybersecurity. The literature survey first aims to provide an overview of artificial intelligence and various cybersecurity practices. Then, in the background, we discuss the theory of machine learning and deep learning; and how these techniques can be used to detect intruders and attacks. On November 3, 2020, in the meeting room of the Ministry of Communications and Information Technology, a workshop was held with the participation of thirteen responsible and specialists from several government and private agencies, which its results were used as a second technique in this research. During this workshop and the individual meetings of the participants after that, a discussion occurred to detecting concerns about network security and monitoring; and suggestions for improvement. Thus, this research includes potential AI-based remedies for identifying issues.

This study used qualitative research as a methodology. The focus is on individual and group actions and perspectives. Subjective opinions and observations often form the basis of qualitative research, there may not be many participants at the workshop, but a lot of information flows and enough data are collected to answer the questions. It is important that participants are suitable for the study and are carefully selected [9].

In qualitative research, a deductive technique involves a thoughtful, predefined rationales with clear connections between research and meetings. Thematic content analysis and narrative analysis are her two subsets of the inductive technique. looking for recurring themes and patterns in the material collected from participants. Participant comments are analyzed and the most important findings are highlighted [10]. A thematic content analysis technique was chosen for this study.

The first part of the workshop focused on the current network and security monitoring tasks and the types of activities each participant performs on a regular basis. Participants were then asked about the types of problems they encountered in performing their duties and how those problems affected the monitoring and security elements. Suggestions for improvement were also considered. I wrote down the problems and suggestions for improvement, then typed it up. It is difficult to have a general dis-

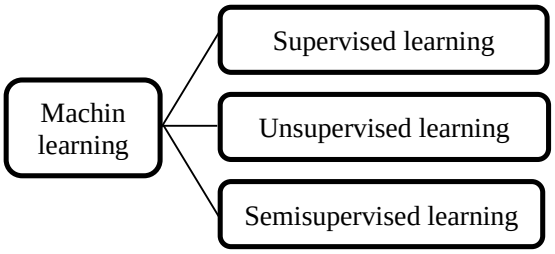

Figure 1: a taxonomy diagram of the machine learning algorithms [8]

cussion about AI due to the lack of specific AI expertise among most of participants. In hindsight, this was not ideal, as a more complete explanation of AI technology and how to use it would have been very helpful. I had no access to or training in network monitoring tools, this was another limitation for the study.

3 Machine learning techniques

The question whether a general-purpose computer should look at data and generate rules rather than relying on humans to generate them was raised by Alan Turing's assertion that computers can learn and determine their uniqueness. Techniques that can learn from data and adapt are called machine learning techniques. Machine learning techniques are created to generate results based on what is learned from the data and examples. For example, such techniques will allow a computer to select and perform certain innovative traffic detection tasks without explicit instructions [11].

Machine learning can be effectively used to perform automated evaluation of attacks and security events such as spam emails, user identification, social media analysis, and attack detection [3]. Supervised, unsupervised, semi-supervised, and reinforcement learning are the three primary methods used in machine learning. Supervised learning is based on labeled data, unsupervised learning is based on unlabelled data, and semisupervised learning is based on both [12], as shown in Figure 1.

4 Deep learning techniques

The main difference between these two methods is feature selection. Unlike ML, where selection must be done manually, feature selection in DL is automated. The goal of DL techniques is to better understand the input data and extract data qualities that are difficult for humans to perceive [13]. Machine learning methodologies are also utilized in deep learning. owever, other ways are employed in deep learning, such as Transfer Learning [8], as shown in Figure 2.

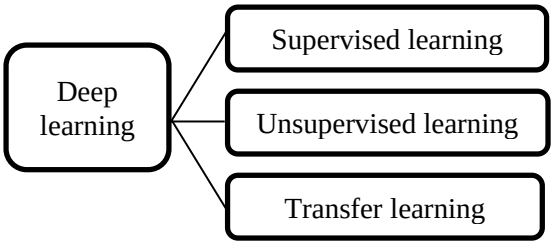

Figure 2: a taxonomy diagram of the deep learning algorithms [8]

5 Analysis, solutions and current techniques

This section presents analysis of workshop discussions, and possible solutions and current techniques which we can used.

5.1 Analysis of workshop discussions

In this subsection, a description of the discussion and notes taken during the workshop and the individual meetings of participants after that.

Director of Information Security - Yemen Mobile Company: this participant claimed that lack of visibility of network data was the biggest problem with network monitoring. Although no specific solutions were presented in this workshop, participants were intrigued by the potential application of AI to detect anomalies. We also explored what changes and adjustments should be made to existing networks to facilitate automation of anomaly detection anomalies.

Director of Information Systems and his two employees - Ministry of Communications and Information Technology: also, these respondents raised the concern of data visibility. Latitudinal mobility and methods for spotting potential assaults and threats inside the network were the specific types of network visibility that were covered. After they have gained access to the network, attackers exploit lateral movement to penetrate farther. One respondent mentioned a number of technologies, including security operations, automation and response technology, and Network Detection and response systems. According to the three respondents, these kinds of technologies were important for the ministry 's upcoming network security and monitoring goals. Moreover, skill certification was listed as a potential upgrade for the future.

Assistant Undersecretary for Technical Affairs at the Ministry of Communications and Information Technology: the topic of data visibility was also brought up by this participant. He claimed that the technology used does not provide real-time insight and alerts on network anomalies and risks. Endpoint detection technology and the potential to improve the monitoring process were mentioned by the respondent. Interesting topic he is one. He said researching techniques to increase the visibility of lateral movement and the hazards it poses will be another important feature in the future. Members of the Teaching Staff - Higher Military Academy: the issue of network visibility was also

raised by the participants. Respondents said that additional monitoring and carefully selected tools designed for the specific problem would be the answer. The idea behind the recommended technique was to streamline surveillance activities and increase their effectiveness.

Director of Information Security - General Post Authority: a key point for the participant in this conversation, also related to data visibility. According to the person, data about anomalies under investigation may lack relevant information, so one remedy is to learn more about the causes and underlying problems of anomalies. Custom development is another key component of network monitoring that the participant pointed out. According to the participant, this makes it easier to match the infrastructure with the authority's needs.

Director of Monitoring and Control - Public Telecommunications Corporation: this participant raised the possibility of improved alarm handling and alarm accuracy. Participants hope that in the future, incident managers will take a more active role in building and improving surveillance systems, resulting in more reliable system output and more accurate warnings. This participant suggested intelligent monitoring and creating easy-to-understand dashboards that monitoring teams could use as tools.

Director of the College of Postgraduate Studies - Police Academy and director of the Command and Staff College - Higher Military Academy: These two participants emphasized the management aspect. They noted the importance of clearly assigning and delegating responsibilities for various aspects of the monitoring and security process. In-house skills were also a factor, they said, and felt that this strategy should be promoted in the future. It was determined that some planned software improvements would require major hardware refreshes, which is neither feasible nor beneficial from a cost-effectiveness standpoint. Another technical issue was the difficulty or impossibility of linking the colleges' specific software to other system security colleges.

Director General of Cisco Academy - General Institute of Communications: this participant pointed out that some monitoring tasks and root cause analysis are still done manually. According to this participant, more automated techniques for detection of anomalies would be the solution to this problem. Participants specifically recommended that automatic feature selection can reduce some human effort. Participants also expressed a desire for staff to receive additional training on surveillance-related responsibilities in the future. According to the participant, one of the possible remedies, he said is skill certification, which will enhance the overall oversight process.

Director of Information Systems - TeleYemen Company: according to this participant, the existing system is too passive. This individual was interested in how AI could increase the aggressiveness of network monitoring systems and how AI-based systems could provide more real-time data and information. Attendees were also interested in AI-based products on the market and how they can be used to help companies.

Information Systems Consultant - General Post Authority: this participant found that some devices connected to the internal network did not have enough information during their daily work. Future solutions this participant hopes will include richer

endpoint data and more real-time network traffic statistics. Participants made these suggestions to increase the effectiveness of their Workshop outcomes.

The first four participants brought up the visibility of network data. The data relevance and relevance of pertinent information about network traffic and networked devices are closely tied to the visibility issue. Directors of Information Systems bring up these kinds of problems. The collecting and analysis of data traveling inside the network as well as into and out of the network constitutes visibility in terms of network monitoring. East-west traffic is a term used to describe traffic that travels from one server to another inside a network.

Director of Information Systems and his two employees and Assistant Undersecretary for Technical Affairs at the Ministry of Communications and Information Technology raised concerns about lateral or east-west traffic monitoring. Around half of the participants brought up visibility-related issues. The fact that participants from other teams or agencies also mentioned this issue demonstrates the broad impact the visibility issue has. This makes it quite evident that in order to address this in the future, suitable solutions and action must be taken. An organizational network's visibility is essential in many ways. Visibility of network data is essential in a setting where the frequency of cyberattacks is rising.

Real-time network data and information were the second most common issue of the workshop. Real-time data is crucial because it enables monitoring and security teams to respond to threats and anomalies more quickly. Director of Information Security - General Post Authority and Assistant Undersecretary for Technical Affairs at the Ministry of Communications and Information Technology raised this issue. This problem is especially worth detection and recognizing because they have diverse job descriptions and duties.

Other top priorities are problems with alert processing and alert accuracy. The importance of alert accuracy makes issues with it a typical occurrence in network monitoring. More automated detection techniques have a byproduct of false positives and false negatives. This topic was only mentioned by one person (Director of Monitoring and Control - Public Telecommunications Corporation). Yet, this does not imply that the problem is unimportant, and the accuracy of the alerts is a crucial component of any security monitoring system. User and Entity Behavior (UEBA) can offer solutions to improve alert, confidence, and accuracy.

The monitoring procedure, including feature selection, still involves several manual chores, according to Director General of Cisco Academy. The application and use of DL to this problem is possible.

A few more themes that cannot be resolved by AI were also raised by the attendees. Finding solutions that are carried out internally was a prevalent subject that came up in the workshop. This gives the business more control over the design and system infrastructure. Regarding monitoring, this helps the agency to find solutions to issues more quickly without having to wait for input from providers. Director of the College of Postgraduate Studies - Police Academy and director of the Command and

Staff College - Higher Military Academy discussed the administration of the system. Clear management and well-defined goals are necessary for the continued growth of the monitoring and security system in order to enhance the system.

5.2 Solutions and current techniques

As a part of my task at the workshop, I had to find and present AI-based solutions and techniques. The purpose of this section is to propose solutions and current techniques to the problems brought up by the participants. This section describes the design principles needed to create an ML-based recognition system and the factors that need to be considered. One of the main themes of the workshop was how to move from reactive to proactive action, so planning for systems should be emphasized. Moreover, the role of intrusion detection systems (IDS) and intrusion prevention systems (IPS) in network security and monitoring is discussed. Some typical surveillance systems use AI. IPS and IDS concepts and technologies are moving towards so-called Network Detection and Response NDR systems. Therefore, in the next subsections, we will discuss how to develop detection, response systems and AI-based strategies to address many of the issues raised at the talks. Together with AI-enhanced security operations, this section also discusses unsupervised learning techniques.

5.2.1 Machine Learning Anomaly Detection Systems

Today, any agency must regularly analyze its network data for trends. A large part of network monitoring is finding anomalies. A data analysis process called anomaly detection looks for unusual and irregular patterns in data. Deviations from standard network behavior and their identification are essential so that network administrators can take action and respond to attacks and hacking attempts. AI can be used in this process. Anomalies can be classified in different ways, depending on the type of anomaly and the context in which it is examined. A particular data point or instance that appears to deviate from the typical metric value of the dataset is called a point anomaly. Contextual anomalies are anomalies caused by the specific context and facts of a situation. For example, some periods or seasons are expected to deviate from typical data behavior. However, if this occurs outside the expected time frame, it is considered an anomaly of context. Groups of data instances that collectively deviate from normal network behavior are collectively referred to as collective anomalies. A single anomalous result is not considered an anomaly in this scenario [14].

Manual techniques are still commonly used to detect anomalies. A weekly reviewed dashboard is a popular way to monitor a network. Personnel monitoring the network watch for spikes and dips in traffic, her activities and see if they are abnormal. This method is limited to the original collection of measurements and is difficult to scale. This method can detect large anomalies, but may not easily detect small anomalies. Many attacks today are more scattered and short-lived, so the system may not even know what to look for. Reviewing the dashboard after an anomaly has already occurred

significantly slows the incident response time. Manually setting thresholds on selected metrics is another way to detect anomalies. This technique relies on alerts being sent when metric measurements deviate from thresholds. Setting the thresholds correctly is essential for this strategy to work, but it can be difficult even for small networks as there are hundreds of measurements and functions to consider. Setting the threshold too high or too low will increase the number of false positives. Anomalies can also go unnoticed if the limit constraints are not specified correctly [15].

Several participants pointed out problems with real-time information and expressed a desire to further automate the anomaly detection process in the future. There are five key design elements that must be considered throughout the planning phase of an ML-based automated anomaly detection system [15]: Event timeliness, scale, rate of change, conciseness, and definition of incident.

The Event timeliness deals with the question of how quickly anomalies must be discovered. Alerts for attacks, threats, and identified dangerous network activity are part of real-time anomaly identifying. Enterprises should be aware of this factor as it influences the choice of which type of ML technique to apply and the tasks and goals for which it is used.

The infrastructure and execution of detection systems are highly dependent on scale and data amount. Different monitoring activities use different record sizes. In this case, it is important to examine whether the system works with large amounts of data or small amounts of information. ML and DL techniques react differently to data volume and labeling, or lack thereof, so the planning of the system is affected.

Rate of change is the rate of change of the measured data. Change rate affects the ML or DL technique used for a task, depending on the metric being monitored. Systems may experience frequent changes in measurements during network monitoring, so to work effectively, techniques must include adaptive properties.

Conciseness is the idea that a large number of measurements and factors should be considered when detection anomalies. This concept helps provide users and systems with a comprehensive view of the root cause of anomalies. Often, looking at just one measure doesn't tell the whole picture. For example, a system upgrade in one region may cause delays in another region. The conciseness of ML anomaly detection system can be a technique in three different ways. Each system metric is used by Univariate Anomaly Detection to create a map of typical behavior.

This technique scales easily because each instance is handled independently. Root cause analysis is not possible with this technique, and many anomalies are generated by one unexpected event affecting multiple metrics simultaneously. Multivariate anomaly detection takes multiple inputs and analyzes them together. We combine these metrics to create a virtual event model. This has the disadvantage that the anomaly output does not indicate the parameter that caused the anomaly. In order for the system to have the required computing power, the input signal types must be identical. The hybrid technique of these two methods takes a univariate single metric technique, but explores the relationships between anomalies rather than grouping many of them in a black box.

This gives better results when conducting root cause analysis.

Definition of incident is the final design principle to consider. A fully automated system for detection of intruders and anomalies is not yet feasible, so it remains necessary to determine what defines an event. A complete description of the event requires detection of all, or at least most, causes of the anomaly. This works for systems with a limited set of parameters and metrics, which is usually not the case. At this stage of system design, the ML and DL concepts of supervised, unsupervised, and semisupervised learning should be considered. Supervised methods are appropriate when some measures are clearly specified and the goal of the system is to provide classification or regression. For these techniques, to establish a baseline of typical behavior, the training data needs to be tagged. This technique is not very effective at detection new intruders or anomalies due to supervised learning-based anomaly detection training models.

Unsupervised learning is a valid strategy in most cases because it can be difficult to predict all possible situations. An unsupervised procedure identifies anomalies whenever the investigated data deviates from the trained model, and the system gradually learns typical system behavior. The training dataset is unlabeled, so the system can detect new incidents, whether known or unknown. The success of unsupervised learning depends on how accurately and completely the behavior of a typical system is described. This makes the training phase essential as it sets the thresholds and the framework within which the technique operates. Combining these two techniques creates a better methodology.

This is achieved through semisupervised learning. Here, the technique trains on a sparse set of labeled data to understand the input structure. Combining supervised and unsupervised techniques often gives the best results for anomaly and intruder identifying. Intrusion and anomaly automation fundamentally relies on the notion of normal or abnormal behavior. Any approach for detecting anomalies generally works by gathering data, determining what is normal, and then using a statistical test to determine if any subsequent data point in the same time series is normal or abnormal. Such as the shaded region in Figure 3. Below was created as a result of such statistical analysis. So, we might use statistical tests to identify any data point outside of the shaded region as abnormal and any data point of it as normal.

The input data for the network is often very large, but it can also be unpredictable and nonlinear. Combining several different models is a more reliable option to create a correct baseline, as it is impossible to simulate typical behavior. Seasonality is an important factor to consider when creating baselines, and failure to do so can result in contextual anomalies. Second, because ML and DL techniques periodically update new normal values based on incoming data, anomalous data points are subject to the fact that they change the baseline and to prevent or produce erroneous results in the future, it should be given less weight [16].

Output rendering in anomaly and intrusion detection is often problematic. The significance of an anomaly should be expressed in some way when designing an detection system to determine how far an anomaly deviates from the normal and how the system should respond. Scoring and binary metrics can usually be used to label anomalies.

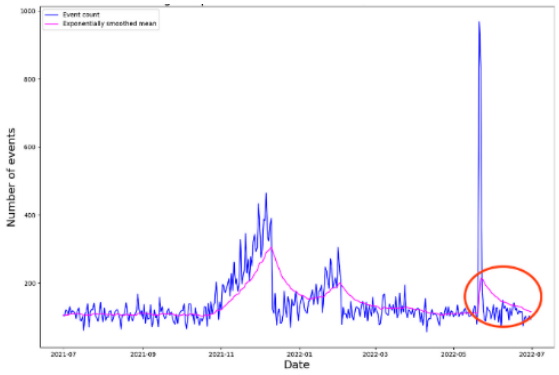

Figure 3: The data points outside of the shaded region are anomaly [15]

Score-based detection assigns a score to the amount by which an identified anomaly deviates from a specified normal norm. Anomalies are ranked according to their scores and can be further investigated. The second method, binary classifies instances as either normal or abnormal [14]. In other words, the system should be able to assess the relevance of anomalies to intruders and react only when necessary. The result is fewer pointless alerts and less manual work for network analysts.

Understanding how metrics are related and how they could stem from the same root cause is a challenge related to the problem of many alarms being active at the same time. Therefore, learning behavioral topology is of great importance in the development of automated systems. Investigating the occurrence of anomalies and determining whether there is a causal relationship between them is one way to detect commonalities in anomalies. The use of clustering techniques is a machine learning technique for determining causality [17]. Clusters can then be represented using the two techniques described earlier (scoring and binary labeling). The Latent Dirichlet Assignment (LDA) technique is a popular candidate for clustering big data. To find the underlying patterns in the data and understand how different sections of the data are related, thematic modeling uses his LDA, a machine learning technique [18]. Often data points are classified as belonging to only one class using clustering techniques. LDA has the advantage that measurements can belong to more than one class and commonalities can be found between them [17]. A strategy for intruder detection using the LDA technique was proposed by Huang et al. [18].

5.3 Intrusion Detection System

Intrusion detection systems (IDS) are one of the best known areas of cybersecurity using ML techniques. This section describes how IDS works and how to create an IDS taxonomy. IDS is a security software that automatically notifies network administrators when hostile activity, system compromises, and security policy violations occur [19]. IDS uses hardware solutions or software embedded in firewalls to monitor networks and

detect malicious network traffic. Security information and event management systems often combine the output of network resources into hostile activity reports, Security information, and event management (SIEMs). An IDS is one part of a system that includes firewalls and intrusion prevention systems (IPS). However, there are significant differences in the functionality of these components. IPS is similar to IDS, except that the system has the ability to disconnect the network. An IPS is an active, inline threat protection system, often placed right behind a firewall. In contrast to IPS, IDS is a passive system that observes network packets and identifies potentially malicious behavior by comparing signature patterns to predefined typical patterns [20, 21].

There are two types of intrusion detection systems [20, 19]. Both host-based intrusion detection systems (HIDS) and network intrusion detection systems (NIDS) monitor operating system files and network traffic, respectively. NIDS works by distributing sensors throughout the network to monitor network activity. HIDS is a host or device-specific implementation and only tracks traffic on that host or device. Both NIDS and HIDS use two techniques to detect malicious network traffic: signature-based attack detection and Anomaly-based attack identifying.

The basis of signature-based intrusion detection is the identification of identified threats. This technique requires a predefined set of indicators of compromise to scan the traffic for. Byte sequences, file hashes, and email subjects are examples of indicators. When the system identifies a suspicious network activity, it observes the packet, compares it to a database of threat indicators, and flags the packet [22].

The second technique, known as anomaly-based intrusion identifying, uses machine learning techniques. Here, the system uses ML techniques to train the detection engine to detect typical behaviors and create a baseline. This strategy compares network behavior to the norm rather than looking for threats. This means that the system will sound an alarm if the behavior deviates from the norm. This method has advantages over signature-based strategies in that it can detect unknown threats [22]. The majority of surveillance infrastructure and systems use IDS and IPS. Nonetheless, there are some issues that hinder some components of the monitoring task. The primary purpose of IDS is to track north-south network traffic. Client/server traffic entering or exiting a network or data center is referred to as north-south traffic. In other words, IDS can detect risks and anomalies originating from networks outside the organization network. IDS cannot detect attacks inside the network [23].

Lack of visibility into internal network traffic such as Server-to-server or called east-west transmissions are problematic. Moreover, the IDS technique relies on a signing library, so it cannot detect new attack types. Moreover, automated analysis and investigations require human involvement or cooperation with IPS [23]. Like IDS products, intrusion prevention systems have some disadvantages. A large data set generates many meaningless alarms and IPS cannot separate the relevant signal from the noise. To effectively detect threats and anomalies, IPS solutions use signature models that need to be updated frequently [24].

While IDS and IPS technologies are essential components of any security operation

and cannot be replaced, there are techniques that can address the above issues. Gartner's security operations center visibility triad represents a new development in network security. In the next section, we'll take a look at it.

5.4 Gartner's security operations center Visibility Triad

The SOC Visibility Triad was first mentioned by Augusto Barros, Anton Chuvakin, and Anna Belak in their March 18, 2019 Gartner research paper titled "Applying Network-Centric Approaches for Threat Detection and Response." The rising sophistication of threats necessitates companies to leverage many sources of data for threat detection and response, Gartner says in this note. Composed of three components called SIEM, EDR, and NDR, this idea enhances visibility of network traffic, reduces attack response time, and helps to detect various types of East-West attacks [25, 26].

5.4.1 Security information and event management

A software program called Security Information and Event Management (SIEM) is used to log and collect network data. SIEM brings this data gathered from applications, endpoints, cloud services, and security devices, and collects them in an integrated manner. Threats and events can be classified based on this aggregated data and alerts can be generated based on defined criteria. AI-based UEBA technology is incorporated into SIEM methodology. The process of creating network baselines is more automated using UEBA [27].

5.4.2 Endpoint detection and response

The second component of the visibility triad is endpoint detection and response (EDR) technology. The role of this technology is to detect malicious data and traffic on endpoints such as servers and laptops. EDR uses behavioral analysis on top of traditional antivirus software to detect risky activity [27]. EDR is a complementary technology to SIEM, and the two parts have long served as the foundation of security operations. An endpoint system can detect this behavior and flag it as an anomaly, but the log could be corrupted by an attacker, resulting in the threat not being reported by her SIEM [27]. Security operations personnel can use EDR technology to isolate infected endpoints from the network and prevent lateral spread of threats.

5.4.3 Network Detection & Response systems

The Network Detection and Response systems (NDR) represent a new development in network security and it is the latest addition to the Triad. NDR technology complements here, EDR, and SIEM methods by enabling detection without the use of rules or signatures that govern system operation [28]. By using carefully placed sensors, NDR can monitor east-west communications in addition to monitoring north-south traffic at the network perimeter. Advanced NDR systems can detect novel unknown attacks. The

NDR system uses AI technology to provide businesses with enhanced detection capabilities. Furthermore, ML has given the NDR system the ability to more accurately assess threat levels [29]. It is important to emphasize the complementary nature of triads. Endpoint detection systems require continuous monitoring and maintenance, reducing internal network visibility. NDR uses behavioral analytics supported by ML and AI-based techniques to detect threats propagating and communicating between intranet devices. A more reliable security architecture is created by NDR's real-time monitoring capabilities and EDR's signature-based technology. Additionally, the NRD system uses a cloud-based ML technique to reduce tedious modeling work and reduce system load. In addition, NDR introduces the ability to automatically update its signature database and detection models using machine learning techniques. Real-time network information provided by the NDR component can improve the EDR system's ability to isolate specific endpoints from the network [25].

SIEM systems are a popular method used by companies for security operations because they are effective at recording data. With well-defined parameters for malicious behavior, SIEMs are typically able to detect threats early. Data analysis by SIEM systems leads to a large number of false alarms. This is an undesirable result when using security systems. Due to the large amount of data generated in large organizations, network traffic logging is disabled at certain times of the day or week. As a result, there is a window for attackers to launch attacks and delete or modify previous logs, making attacks difficult to detect. Wire data (network packets) is used in the NDR system to show network communication and paths through the network. Wire data, unlike log data, is immutable, providing SIEM systems with complete and reliable metadata to perform their functions [25].

5.5 Unsupervised learning techniques

In this subsection, we shall present three patented unsupervised learning security techniques.

5.5.1 Enterprise Immune System

The system uses unsupervised learning techniques to detect and block known and unconfirmed malicious network traffic. This method addresses the problem of rule-based systems by learning from data to create models of both normal and deviant behavior [30]. An example shows how the AI component of Darktrace's company solution can detect ransomware attacks in real time and react to them [31]. In one case, an employee accessed a malicious word document via a work email, allowing ransomware to enter the network. The employee's computer started connecting to suspicious external sites and searching for SMB shares. SMB or Server Message Block is a network protocol that allows files to be shared between devices connected to the same network [32]. Additionally, the ransomware can start encrypting SMB shares and have a detrimental effect on the entire corporate network. There were no employees on site as the incident occurred

after business hours. The enterprise immune system detected this within 9 seconds, and after 24 seconds the engine stopped cryptographic operations, fixing the problem without human intervention. Due to the fast response time of the system, only a small portion of the network was slightly damaged. Detection of these threats is easier using unsupervised learning techniques, as demonstrated by the Darktrace case.

5.5.2 Vectra Cognito

Vectra AI company develops an NDR cybersecurity platform with AI-driven attack behavior detection [33]. Vectra Cognito is the name of the proposed system, and it claims that, similar to the enterprise immune system, Vectra AI can learn the system's behavior and capabilities in detecting threats. Instead of using all available network data, Vectra Cognito collects useful enriched metadata to give you a clearer picture of your system. This feature also reduces noise in the data using SIEM and EDR systems.

5.5.3 Security operations, automation, and response technology

A relatively new tool and implementation idea in network and cyber security is Security operations, automation, and response (SOAR) technology. SOAR technology aims to tackle three key aspects of modern cybersecurity. These three aspects include threat and vulnerability management (orchestration), automation of security operations (automation), and incident response (response) [34]. Input from internal and external sources can be accepted into the SOAR system, also improving visibility of network events and traffic flows. The system can automatically respond to situations using SOAR and AI, as well as recommending actions to security operations teams. The task of the orchestration function is to combine and coordinate the automated and manual processes of the information system [35].

6 Results and Discussion

Cybersecurity has made incredible progress over the past few decades, and there are solid examples of how AI has been applied to it. The results of this study show that AI techniques are suitable for cybersecurity. Integrating AI into security and surveillance systems can increase the effectiveness of anomalies and threat detection. It also reduces the number of daily repetitive tasks that work security teams have to perform.

Before adopt an AI technique, companies should assess their current systems to see what problems they face. There were recurring themes in this research that emerged from multiple sources, and most of these issues could be resolved or mitigated using AI techniques. But the term "AI" is used both as a commercial ploy and as a promising hypothesis of a panacea for any problem. Certain artificial intelligence techniques such as NDR have shown promising results in security tasks, but these are mostly machine learning techniques that are not as complex as could be possible. For example,

these systems are not intelligent in the sense that they have human-like awareness and knowledge in addressing problems.

Most ML techniques still rely heavily on manual oversight and human interaction. The unsupervised learning techniques have the potential to learn on their own, whereas supervised techniques are limited to certain tasks in the security architecture. The aspirations for a self-learning and decision-making computer have been partially satisfied by unsupervised methods and deep learning techniques, and as previously said, DL techniques have elevated the field of AI to a whole new level.

Future work on AI in cybersecurity should allocate more resources to testing unsupervised methods. It's important to remember that artificial intelligence is still a tool for security teams, not a replacement for human reasoning and problem-solving skills. This can improve the Preventive and automated aspects of security systems, so it is important to further develop AI techniques that understand context.

The AI techniques used in attacks may be more sophisticated than defense, so the challenge going forward will be keeping up with the latest techniques from attackers. Most datasets used to train and test AI systems are not up to date and do not contain enough malicious data points. In order for the chosen AI system to be able to learn what a typical activity actually looks like, and vice versa, the dataset needs to be updated [36]. AI techniques can also be compromised, so it is important to protect the integrity of AI techniques. Compromised data collection combined with compromised AI techniques can lead to erroneous results and prevent security systems from functioning properly [37].

Advances in cybersecurity are driven by knowledge and expertise in the field. These talented workers are in great demand and there is a skill gap in cybersecurity. More people need cybersecurity education and training to create identification systems, improve AI techniques, and develop security measures [38]. This is related to the interpretability of AI features, as the analyst needs to understand how the AI component of the system reaches its conclusions. Developing improved AI also requires a basic understanding of current AI techniques, but progress is hampered by widespread ability deficits. By making AI security solutions intuitive to use and visually compelling for companies and enterprises will adopt AI security solutions on a larger scale. The term "AI" is so widely used that there can be confusion about what a particular AI feature actually does. AI-controlled machines are different from AI-enabled machines.

Fully automated AI-driven solutions are still a concept of the future. The primary goal of most AI solutions is to help analysts make better decisions. Industry standards can be established to measure system autonomy and evaluate AI systems to prevent blind reliance on vendor solutions [37]. Product testing transparency is another issue when implementing AI. There may be human resistance to the paradigm shift towards AI, as there are not many industry norms and standards to rely on.

7 Conclusion

The concept of using AI techniques to defend against future cyberattacks has its pros and cons. As attackers and malicious actors continue to improve their attack methods, there is an urgent need for action. IDS, IPS, SOC, and SOAR systems in use today still rely heavily on human intervention. But with the number of attacks outstripping the effectiveness of traditional reactive rule-based defense techniques, the use of AI technology is the ideal way. AI, especially machine learning (ML), gives detection and response systems the opportunity to be more proactive and form real-time actions. It also improves the collection and analysis of network traffic data to improve the effectiveness of security operations and security teams. Today's AI tools are only useful for a limited number of network and cybersecurity related activities. Future research should focus on ways to facilitate automation of AI solutions while reducing the need for human interaction. Achieving this goal requires more accurate assessments, training datasets, and industry standards. ML and DL techniques also need to better understand the context within datasets in order to make decisions similar to humans and have a low rate of false outcomes.

References

[1] K. Thakur and A.-S. K. Pathan, Cybersecurity Fundamentals: A Real-World Perspective. CRC Press is an imprint of Taylor & Francis, 2020. Available: https://tinyurl.com/3kc6v6f8.

[2] "Cyber Security Education: Principles and Policies," Routledge & CRC Press, 2020. https://tinyurl.com/3dy44cua (accessed Apr. 15, 2023).

[3] D. Edeh, "Network Intrusion Detection System using Deep Learning Technique," www.utupub.fi, Aug. 2021, Accessed: Apr. 15, 2023. [Online]. Available: https://tinyurl.com/ywhvx2sy.

[4] "Computer Security Fundamentals, 3rd Edition | Pearson IT Certification," www.pearsonitcertification.com, 2016. https://tinyurl.com/4sx2xsdd (accessed Apr. 15, 2023).

[5] D. Franke, "Amazon.com: Cyber Security Basics: Protect your organization by applying the fundamentals: 9781522952190: Franke, Don: Books," Amazon.com, 2023. https://www.amazon.com/Cyber-Security-Basics-organization-fundamentals/dp/1522952195.

[6] C. Sjöblom, "Artificial Intelligence in Cybersecurity and Network security," 2021. Available: https://www.doria.fi/bitstream/handle/10024/181168/sjoblom_christoffer.pdf

[7] R. Mark and R.-O. Bsc, 2022. Accessed: Apr. 15, 2023. [Online]. Available: https://tinyurl.com/mr2a9myf

[8] S. A. Salloum, M. Alshurideh, A. Elnagar, and K. Shaalan, "Machine Learning and Deep Learning Techniques for Cybersecurity: A Review," Advances in Intelligent Systems and Computing, pp. 50–57, 2020, doi: https://doi.org/10.1007/978-3-030-44289-7_5.

[9] E. Fossey, C. Harvey, F. Mcdermott, and L. Davidson, "Understanding and Evaluating Qualitative Research," Australian and New Zealand Journal of Psychiatry, vol. 36, no. 6, pp. 717–732, Dec. 2002, doi: https://doi.org/10.1046/j.1440-1614.2002.01100.x.

[10] Rev, "How to Analyze Interview Transcripts in Qualitative Research," Rev, Mar. 30, 2022. https://www.rev.com/blog/transcription-blog/analyze-interview-transcripts-in-qualitative-research.

[11] G. Fernandez, Deep learning approaches for network intrusion detection, M.S. thesis, Dept. Comput. Sci., Univ. Texas at San Antonio, San Antonio, TX, USA, 2019. https://tinyurl.com/3p32m3xt.

[12] P. Uppamma and S. Bhattacharya, "Deep Learning and Medical Image Processing Techniques for Diabetic Retinopathy: A Survey of Applications, Challenges, and Future Trends," Journal of Healthcare Engineering, vol. 2023, pp. 1–18, Feb. 2023, doi: https://doi.org/10.1155/2023/2728719.

[13] J. Li, "Cyber security meets artificial intelligence: a survey," Frontiers of Information Technology & Electronic Engineering, vol. 19, no. 12, pp. 1462–1474, Dec. 2018, doi: https://doi.org/10.1631/fitee.1800573.

[14] M. Ahmed, A. Naser Mahmood, and J. Hu, "A survey of network anomaly detection techniques," Journal of Network and Computer Applications, vol. 60, pp. 19–31, Jan. 2016, doi: https://doi.org/10.1016/j.jnca.2015.11.016.

[15] "Gmail," accounts.google.com, 2019. https://tinyurl.com/4prmaw9c (accessed Apr. 15, 2023).

[16] K. Corrie, "Building a Large Scale Machine Learning Based Anomaly Detection System Part 2 Normal Behavior of Time Series Data," www.academia.edu, 2019, Accessed: Apr. 15, 2023. [Online]. Available: https://tinyurl.com/yc34t6ac.

[17] K. Corrie, "ULTIMATE GUIDE TO BUILDING A MACHINE LEARNING ANOMALY DETECTION SYSTEM PART 1: DESIGN PRINCIPLES," www.academia.edu, 2019, Accessed: Apr. 15, 2023. [Online]. Available: https://tinyurl.com/mswk4br8.

[18] J. Huang, Z. Kalbarczyk, and D. M. Nicol, "Knowledge Discovery from Big Data for Intrusion Detection Using LDA," IEEE Xplore, Jun. 01, 2014. https://ieeexplore.ieee.org/document/6906855 (accessed Apr. 15, 2023).

[19] M. Amrollahi, S. Hadayeghparast, H. Karimipour, F. Derakhshan, and G. Srivastava, "Enhancing Network Security Via Machine Learning: Opportunities and Challenges," Springer Link, 2020. https://tinyurl.com/bddar2tx (accessed Apr. 15, 2023).

[20] I. Humied, Cybersecurity Amazon, 2023. Accessed: Apr. 15, 2023. [Online]. Available: https://a.co/d/44cfZbK.

[21] R. Bhardwaj, "IDS vs IPS vs Firewall - Know the Difference - IP With Ease," ipwithease.com, Sep. 10, 2020. https://tinyurl.com/24pxy4jx (accessed Apr. 15, 2023).

[22] N-able, "Intrusion Detection System (IDS): Signature vs. Anomaly-Based," N-able, Mar. 15, 2021. https://tinyurl.com/37de7vn9 (accessed Apr. 15, 2023).

[23] C. Snyder, "NDR vs. IDS for Intrusion Detection - ExtraHop | ExtraHop," www.extrahop.com, Jan. 23, 2019. https://www.extrahop.com/company/blog/2019/network-detection-response-vs-intrusion-detection-systems/..

[24] C. Snyder, "NDR vs. IPS for Intrusion Prevention, Detection, and Response - ExtraHop | ExtraHop," www.extrahop.com, Feb. 07, 2019.

https://www.extrahop.com/company/blog/2019/network-detection-response-vs-intrusion-prevention-systems/ (accessed Apr. 15, 2023).

[25] "NDR and the SOC Visibility Triad | ExtraHop | ExtraHop," www.extrahop.com, 2021. https://tinyurl.com/5n8ejpek (accessed Apr. 15, 2023).

[26] "Point of View." Available: https://www.crowdstrike.com/wp-content/uploads/2021/05/soc-triad-solution-brief.pdf

[27] Nettitude, "The SOC Visibility Triad – SIEM, EDR & NDR | Nettitude," blog.nettitude.com, 2020. https://tinyurl.com/mvuy4mme (accessed Apr. 15, 2023).

[28] I. Cybersecurity Inc., "Dynamic detection for dynamic threats," www.ironnet.com, 2020. https://tinyurl.com/5n89rews (accessed Apr. 15, 2023).

[29] "What is Network Detection and Response?," www.ironnet.com, 2021. https://tinyurl.com/37cct6mh (accessed Apr. 15, 2023).

[30] K. Bissinger, "Darktrace Immune System. Self-learning Detection & Response,"2020. https://tinyurl.com/yzr7u6z4 (accessed Apr. 15, 2023).

[31] K. Bissinger, "Fighting Ransomware with AI," www.n3t.com, 2022. https://www.n3t.com/about-us/blog/fighting-ransomware-with-ai (accessed Apr. 15, 2023).

[32] TechTerms, "SMB (Server Message Block) Definition," techterms.com, 2021. https://tinyurl.com/3mnsa2ft (accessed Apr. 15, 2023).

[33] "About Vectra - AI Driven Cybersecurity Company | Vectra AI," www.vectra.ai. https://tinyurl.com/bhxe967t (accessed Apr. 15, 2023).

[34] FireEye, "What Is SOAR? | Definition & Benefits | Trellix," www.trellix.com, 2019. https://tinyurl.com/muzap5zh (accessed Apr. 15, 2023).

[35] Y. Bari, "Infosys Knowledge Institute | The Future of Tomorrow: Automation for Cybersecurity," www.infosys.com, 2019. https://www.infosys.com/iki/perspectives/future-tomorrow.html (accessed Apr. 15, 2023).

[36] E. Segal, "The Impact of AI on Cybersecurity | IEEE Computer Society," Computer.org, 2020 https://www.computer.org/publications/tech-news/trends/the-impact-of-ai-on-cybersecurity.

[37] P. Donegan, "'Trusted Research, Analysis and Insight in IT & Telecom Security' AI in Cyber Security: Filtering out the Noise," 2019. Accessed: Apr. 15, 2023. [Online]. Available: https://tinyurl.com/ycx5nvwj.

[38] Vectra AI, "E-book Prevention Phase Active Phase Clean-up Phase Initial Infection Minding the cybersecurity gap," 2017. Accessed: Apr. 15, 2023. [Online]. Available: https://tinyurl.com/nbjrremc.

Vulnerability Analysis of Social Media Accounts Against Cyber Attacks

Ahamd Farid Aseel[1], Yashar Abri[1]

[1]*M.Sc. Student, Information Technology Engineering, Faculty of Engineering, College of Farabi, University of Tehran*
faridaseel.4all@gmail.com, yasharabri@ut.ac.ir

Abstract

This article examines the vulnerability of social media accounts against cyber attacks. With the increasing number of users on these platforms and the accumulation of sensitive information in user accounts, the security of these networks is facing cyber threats. The article analyzes phishing attacks and brute-force attacks as penetration methods and investigates the weaknesses of social media networks against these attacks. Additionally, social engineering, as a deception-based and psychological attack on individuals within social networks, is analyzed. Through a questionnaire, users' attitudes and behaviors regarding security measures are evaluated. The results indicate that some users use strong passwords but do not change them regularly, and users' awareness of security practices is inadequate. The security of social media accounts is of utmost importance, and this article emphasizes that increasing users' and administrators' awareness of cyber threats and countermeasures is crucial to strengthening the security of these platforms. The information presented in this article can contribute to enhancing the security of accounts and protecting users' personal information against cyber attacks.

Keywords: *Cyber Attacks, Social Networks, Phishing, Brute Force, Penetration.*

1 Introduction

In the world of digital communications and the widespread adoption of social networks, individuals increasingly utilize these platforms for communication, information sharing, and social interactions. However, with the continuous expansion of user numbers on these networks and the accumulation of sensitive information and user credentials in their accounts, the issue of security and vulnerabilities has become one of the most critical and prevalent concerns. Furthermore, as technology advances and cyber threats escalate, social networks are constantly facing various types of attacks. Attacks such as phishing, brute force, and social engineering are among the methods exploited by malicious actors to infiltrate user accounts and misuse sensitive information [1]. The main

objective of this scientific article is to examine the vulnerability of social media accounts against cyber-attacks. To achieve this goal, we delve into the analysis of phishing and brute force attacks as two important penetration methods on user accounts, evaluating the security weaknesses of these networks against such attacks. Additionally, we investigate and analyze social engineering as an attack based on deception and psychological manipulation of individuals. Through a comprehensive and reliable questionnaire, we assess users' attitudes and behaviors regarding security measures on social networks. This article emphasizes that users' awareness of cyber threats and countermeasures is key to strengthening the security of social media accounts and safeguarding their personal and confidential information. Given the increasing significance of social networks in modern society and their prominent role in human interactions, we hope that this article contributes to enhancing users' and administrators' awareness of information security and their accounts, ultimately leading to an overall improvement in the security of these platforms.

2 Hack

Hacking is a computer crime that involves using social media websites to gain unauthorized access to computers or digital devices. Cybercriminals can employ various attack methods to gain access to the targeted users' digital devices [2]. They send emails or messages to users of social media sites, and when the user clicks on a suspicious link, the hackers gain unauthorized access to information through hacking [3]. Two types of attacks are commonly used by cybercriminals: targeted attacks and opportunistic attacks. In targeted attacks, hackers use specific tools to attack a particular target, while opportunistic attacks utilize viruses and worms. This type of attack is especially carried out by hackers, spammers, and cybercriminals [4].

2.1 Phishing

Phishing refers to a type of cyber attack where the goal is to gain access to sensitive and important user information through deception and forgery and exploit it for malicious purposes. In this type of attack, attackers encourage users to provide sensitive information such as usernames, passwords, banking details, and similar information by sending deceptive messages or fake websites [4]. Social networks as the primary target of phishing attacks: Given the large number of users and the various personal information shared on social networks, these platforms are considered the primary target of phishing attacks. Attackers typically attempt to persuade users to enter their sensitive information on various pages by sending deceptive links or pages [5]. Consequences of phishing attacks on social networks: Phishing attacks result in the misuse of users' sensitive information. Attackers may use the obtained information for targeting, fraud, identity theft, and other malicious activities [6].

2.2 Combation Phishing

1. Awareness and Education:

   Raising awareness among users about different phishing attacks and educating them on methods to detect deceptive messages and pages can significantly improve the security of their accounts.

2. Detection of Suspicious Links:

   Using tools and software that detect suspicious links and inform users to refrain from accessing suspicious pages [5].

3. Verifying Website Identities:

   Users should carefully verify the identity of websites and only enter sensitive information on official and reputable pages.

4. Software Updates:

   Regularly updating software and operating systems to patch vulnerabilities and potential security weaknesses [7].

2.3 Brute Force

In a brute force attack, the attacker attempts to gain access to the target user's account by using automated methods and extensive trial and error. The attacker tries all possible combinations to discover the user's password. If the account's password is weak and easily predictable, the attacker can easily penetrate the desired account and access the user's personal information, images, and content. Brute force attack is one of the most commonly used methods to infiltrate systems and user accounts and has been employed extensively in the past. From a technical perspective, this attack takes two main forms: brute force attack, which involves trying different combinations word by word, and dictionary-based brute force attack, which uses a list of words to guess the password [8].

2.4 Combatting Brute Force

To prevent brute force attacks, users should use strong and complex passwords that include a combination of uppercase and lowercase letters, numbers, and symbols. Additionally, enabling security features such as two-factor authentication can significantly enhance the security of user accounts. Given the increasing importance of social networks and the information stored in user accounts, it is essential for users and administrators of these networks to be aware of the vulnerabilities and threats posed by brute force attacks and take necessary actions to strengthen the security of these platforms [9].

2.5 Social Engineering

Social engineering is one of the most complex and popular methods of attacking the security of social networks. In this method, the attacker utilizes psychological tricks and social knowledge to prompt users to provide sensitive information and their credentials. With the increasing use of social networks and the importance of personal information in these environments, social engineering has become one of the biggest security threats in these networks [10].

As a deception-based attack, social engineering seeks to persuade individuals to perform inappropriate actions and disclose their sensitive information by exploiting their motivations, needs, and fears. Common social engineering techniques include phishing emails, enticing messages, and unknown phone calls, with the aim of extracting personal information and gaining access to user accounts [11].

To prevent social engineering attacks, users should consider the following:

- Do not trust unfamiliar and suspiciously sent information.
- Avoid sharing sensitive information and personal credentials with unknown individuals.
- Enable two-factor authentication and other security features in user accounts.
- Ensure the validity and reliability of received website sources and messages.

3 Methods

In this study, a questionnaire was used as a tool for data collection. This questionnaire consists of ten main questions presented with options A, B, and C for participants to respond to. The responses to this questionnaire were completed by individuals from the Persian-speaking community in various countries, including Australia, Iran, Afghanistan, and some European countries. The participants in this questionnaire include individuals with different educational levels, including Ph.D., Master's, Bachelor's, and lower levels. Furthermore, the design and collection of responses were conducted online.

In this research, responses from 80 participants were obtained. The questionnaire was distributed randomly among users of social networks such as Facebook, Instagram, and Telegram. After data collection, statistical analysis was performed, and the results were interpreted statistically. The questionnaire used in this paper is available in appendix 1 .

4 Results

The results showed that 53.3% of users always use strong passwords (including uppercase and lowercase letters, numbers, and symbols) for their accounts, and 46.7% do use this

type of password, but not always. Additionally, 80% of users update their passwords regularly, but not consistently, while 13.3% never change their passwords.

Regarding sharing sensitive information, 20% of users always share sensitive information, and 40% do so cautiously, while 40% do not share sensitive information at all.

The results indicate that 40% of users always check attachments before opening them, 26.7% do so cautiously, and 33.3% ignore this step.

Regarding the use of security apps and extensions, 13.3% of users always use them, 26.7% use them cautiously, and 60% do not use these tools at all.

Results show that 47.6% of users have experienced phishing or attempted unauthorized access to their social media accounts, while 26.7% of users have not experienced such attacks. Additionally, 26.7% of users are not familiar with phishing and lack sufficient awareness.

Regarding actions taken after experiencing phishing, the results indicate that 53.3% of users reported the incidents and took necessary measures to protect their accounts, while 46.7% of users did not take any actions.

Regarding awareness of security measures related to social media accounts, 53.3% of users have sufficient awareness, while 46.7% lack sufficient awareness.

Regarding the use of public social networks or public accounts for communication and sharing personal information, 46.7% of users use these networks, while 53.3% do not use them.

Regarding awareness of cyber threats and ways to counter them, 20% of users have sufficient awareness, while 80% lack sufficient awareness. These results indicate the need for promoting cybersecurity awareness and increasing users' knowledge about cyber threats.

5 Conclusion

In conclusion, the results indicate that a significant portion of users still lack sufficient awareness regarding cyber threats and ways to counter them. Therefore, increasing user awareness about cyber threats and protective measures on social media platforms requires further efforts from organizations and security-related entities.

By implementing necessary changes in user behavior and attitudes and promoting the importance of security for social media accounts, a considerable improvement in the security of these platforms and a reduction in security breaches and abuses can be achieved. As an initial study, these findings can serve as a motivational factor for conducting further research and taking more practical actions to enhance the security of user accounts on social media platforms.

References

[1] S. Y. A. A. M. subhi R. M. Zeebaree, "Social Media Networks Security Threats, Risks and Recommendation: A Case", in International Journal of Innovation, Creativity and Change. www.ijicc.net, 2020.

[2] A. Power, "What is social media?", British Journal of Midwifery, vol. 22, pp. 896-897, 2014.

[3] J. C. Bertot, P. T. Jaeger, and D. Hansen, "The impact of polices on government social media usage: Issues, challenges, and recommendations", Government information quarterly, vol. 29, pp. 30-40, 2012.

[4] S. Norden, "How the internet has changed the face of crime", 1554411 M.S., Florida Gulf Coast University, Ann Arbor, 2013.

[5] C. Konradt, A. Schilling, and B. Werners, "Phishing: An economic analysis of cybercrime perpetrators", Computers & Security, vol. 58, pp. 39-46, 5, 2016.

[6] K. J. Mitchell, D. Finkelhor, L. M. Jones, and J. Wolak, "Use of Social Networking Sites in Online Sex Crimes Against Minors: An Examination of National Incidence and Means of Utilization", Journal of Adolescent Health, vol. 47, pp. 183-190, 2010.

[7] M. Omar Saeed Al, "Threats and Anti-threats Strategies for Social Networking Websites", International Journal of Computer Networks & Communications, vol. 5, pp. 53-61, 2013

[8] Jan Vykopal. A Flow-Level Taxonomy and Prevalence of Brute Force Attacks. In Advances in Computing and Communications, pages 666–675, Kochi, India, 2011. Springer.

[9] A. P. a. M. T. Enrico Franchi, "Information and Password Attacks on Social Networks", in JITR, Italy, 2015.

[10] F. Mouton, L. Leenen, and H. S. Venter, "Social engineering attack examples, templates and scenarios", Computers & Security, vol. 59, pp. 186- 209, 6, 2016.

[11] E. U. Osuagwu, G. A. Chukwudebe, T. Salihu, and V. N. Chukwudebe, "Mitigating social engineering for improved cybersecurity", in Cyberspace (CYBERAbuja), 2015 International Conference on, 2015, pp. 91-100.

Appendix

sample questionnaire: Below is a sample questionnaire used in this article.

1. Do you use a strong password (including uppercase and lowercase letters, numbers, and symbols) for your social media accounts?
A.Yes, always
B.Yes, but not always
C.No, I don't use one at all

2. Is your password for social media accounts up to date and changed periodically?
A.Yes, I regularly change my password
B.Yes, but not regularly
C.No, I never change the password

3. Do you share sensitive information (such as personal information, phone number, home address, etc.) in your social media posts and profile?
Yes, always
Yes, but with caution
No, I don't share sensitive information

4. Do you review attachments you receive on social media (such as files, links, etc.) before opening them?
A.Yes, always
B.Yes, but with caution
C.No, they are ignored

5. Do you use security-related apps and extensions to protect your social media accounts?
A.Yes, always
B.Yes, but with caution
C.No, I don't use them at all

6. Have you ever experienced phishing or attempts to infiltrate your social media accounts?
A.Yes, I have experienced it
B.No, I have never experienced it
C.I don't know what phishing is

7. If you have experienced phishing or intrusion attempts, have you reported it and taken necessary actions to protect your account?
A.Yes
B.No

8. Do you have sufficient knowledge of security measures related to social media account protection?
A.Yes
B.No

9. Do you use public social networks or public accounts for communication and sharing personal information?
A.Yes
B.No

10. Are you knowledgeable about cyber threats and ways to combat them?
A.Yes
B.No

Image Hashing Algorithm: An Analysis and Improvement

Seyed Amirhossein Tabatabaei[1]

[1]*Department of Computer Science, Faculty of Mathematical Sciences, University of Guilan, Rasht, Iran*
amirhossein.tabatabaei@guilan.ac.ir

Abstract

Image authentication code algorithms are schemes wherein the authenticity of an image is considered in a robust war. As the images are mainly subjected to some authorized modifications, such schemes must be able to accept the authorized changes while rejecting the malicious ones. This article analyzes an image authentication algorithm based on error detection code. The image authentication scheme utilizes a type of error detection code in order to encode the mixed most significant bits (MSB) intensity value of each image pixel. The secrecy of system is based on two secret keys. The algorithm provides an acceptable robustness and elaborate design however, it suffers from some security features that leave a gap for improvement. These features are analyzed and will be employed to show the vulnerabilities of the scheme. Also, some solutions are proposed to elaborate the algorithm more. The solutions will increase the strength of the scheme while keeping the robustness.

Keywords: *Image Authentication, Robustness, Image Hashing, Error detection and correction.*

1 Introduction

Appearance of the advanced technologies in multimedia processing techniques like image processing softwares facilitates performing the illegal actions on the digital multimedia object. Violation of image ownership, unauthorized duplication and redistribution and malicious copying and manipulating acts are examples which indicate high demand for image authentication. Image authentication aims to verify the authenticity of images which are subjected to some modification and/or unauthorized manipulation. Image authentication schemes are mainly based on the perceptual image hashing or watermarking techniques. A shared secret key is used in hash generation or watermark embedding process and verification to provide security of the scheme. A perceptual image hashing algorithm differs basically from a cryptographic hash function: the generated hash

corresponding to the images with semantically identical content are equal or very close to each other. This property addresses the robustness of perceptual image hashing. However, the independence of the hash output for perceptually different images must be ensured [1]. The concept of security in a hash-based image authentication is a challenging issue. It refers to the ability of the attacker to find perceptually different images with almost equal hashes after observing sufficient number of image-hash pairs. Also, it ensures that the secret key can not be compromised and no image hash or watermark can be generated without knowledge about the secret key.

In this paper, the analysis of an interesting proposed code-based image authentication scheme [2] is presented. The proposed scheme is based on the initial work given in [3] whose perceptual hashing algorithm uses Hamming code technique to generate and embed parity bits in the pixels. To provide the security two secret keys are involved in the embedding process. The main contribution of this paper is the analysis of the security aspects of the authentication scheme for further enhancement. In fact it is shown that the engaged keys do not strengthen sufficiently the security to be met by an image authentication mechanism. Furthermore a solution will be proposed and discussed in details. The rest of this paper is organized as follows. A short description on the related works is given in Section II. Section III describes the subjected coding-based image authentication scheme. The analysis of the scheme is given in Section IV followed by the proposed solution in Section V and security discussion in Section VI. Section VII concludes the paper.

2 Related Works

There is a large body of research works in the literature in the field of image authentication. They are categorized according to their construction technique and application. Classification based on the robustness of the scheme is mostly used in the taxonomy of image authentication methods. According to this classification they are categorized into two major groups performing hard authentication and soft authentication respectively [4]. The main techniques in the first group where the robustness and the number of acceptable modifications is limited are based on standard cryptography and fragile watermarking. In the second group of image authentication schemes, the level of robustness is higher than the first group and a wider range of authorized modifications is accepted while the malicious manipulations are supposed to be detected. The main technique in this group is based on the semi-fragile watermarking and content-based signature wherein the semantic content of an image is extracted as a feature in order to generate a digital signature [4, 5, 6, 7, 8]. Despite of a fast progress in extracting and analyzing the image data used in the image authentication scheme, there are relatively less attention to the corresponding security analysis as one of the main pillars in this design. As one of the leading works regarding to this matter, Swaminathan, et.al., [9] proposed to use differential entropy to evaluate the security of some existing image hashing schemes in the literature. Although it was shown later that the proposed ap-

proach does not justify due to existing scale variant property [10]. In another approach unicity distance was used to determine the maximum number identical used keys for an image authentication scheme [11]. The fundamental works which consider the generic security of perceptual hashing in information theory viewpoint emerge in [1, 12]. Some works also analyze the security problems of the existing schemes individually [13, 14]. A secure framework for general perceptual image hashing also has been proposed in [15].

3 Image Authentication Method based on Applying Hamming Code on Mixed Bits

The proposed method by Chan consists of three procedures including the embedding, the detecting and recovery procedure [2]. It is based on the initial scheme firstly introduced in 2007 [3] where some improvements have been applied. In this section just the recent scheme [2] is recalled and described. The embedding procedure generates the parity check from each pixel intensity value and embed it into another pixel intensity value. In this procedure a Hamming code scheme (Hamming(7, 4)) is used to generate three parity check bits from four most-significant bits of a pixel. To generate the parity check bits, at first the order of the first four most-significant bits of each pixel indicating the data bits is reversed and then a three-bit parity check value is produced. The three-bit parity check value is rotated and embedded into another pixel specified by a transformation. The rotation operation is performed by the use of a sequence of random numbers extracted from a random number generator based on a secret key k_1. Let the image size be $N \times N$, the aforementioned parity check of the pixel P_i with the original bit order J and new bit order J' goes as $J' = (J + R_i) \bmod 3$. For the sake of simplicity the rotated parity check of each pixel is called authentication data. The authentication data bits are embedded into another pixel using the Torus automorphism and Modulus function as follows [2]. Let $P_i = (x_i; y_i)$ be the first pixel to be processed. The next pixel to be processed is specified by Torus automorphism:

$$\begin{bmatrix} x'_i \\ y'_i \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ k_2 & k_2 + 1 \end{bmatrix} \times \begin{bmatrix} x_i \\ y_i \end{bmatrix} \tag{1}$$

In the above equation k_2 is the second secret key. The authentication data of pixel P_i is modified using the Modulus function before embedding into the new pixel position $P'_i = (x'_i; y_i)$. In the Modulus function a secret value s of k bits length is embedded in a pixel value y. Let $d = s - (y \bmod 2^k)$, the new value of y denoted by y_0 is computed according to $y_0 = d_0 + y$ where d_0 is defined as follows.

$$d_0 = \begin{cases} d & , \text{ if } -\lfloor (2^k - 1)/2 \rfloor \le d \le \lfloor (2^k - 1)/2 \rfloor \\ d + 2^k & , \text{ if } -2^k + 1 \le d \le -\lfloor (2^k - 1)/2 \rfloor \\ d - 2^k & , \text{ if } \lceil (2^k - 1)/2 \rceil \le d \le 2^k \end{cases} \tag{2}$$

This procedure is continuing by embedding the authentication data of pixel P'_i into another pixel specified by the Torus automorphism and the Modulus function until the

cycle is completed by reaching the first pixel P_i. Then the process is repeated to process the whole pixels set. At the end of this procedure the authentication data of each pixel is embedded into another pixel. The detecting procedure localizes and marks the tamper area for the sake of recovery in the recovery procedure. In this phase the authentication data of each pixel is extracted firstly and compared with the regenerated value. The details of this procedure are given in [2].

4 Analysis of Chan's Scheme

The presented image authentication scheme uses two main secret keys in the embedding phase. This first key is used as a seed to generate a sequence of random numbers for the rotation operation and the second key is used in the Torus automorphism to provide ambiguity in locating the host pixel. According to [2, 3], the purpose of employing two secret keys is to provide the security against cases when one key is recovered by the attacker. In fact, if the keys are disclosed, the attacker can design different attack scenarios like impersonation or substitution attack. In this section, we show that when the Chan's image authentication scheme is applied on the chosen images with specified pattern (chosen plaintext attack model) then the second secret key can be recovered trivially followed by an exhaustive search to recover the first key. The second key can be recovered by inquiring two authenticated images I, I' differing in one pixel $(x_i; y_i)$, by solving the Equation (1) for k_2. However, it might happen that more than one pixel in the row $i+j$ in I' will be modified. In this case the key k_2 cannot be determined uniquely. The main security issue of the discussed image authentication scheme is the lack of enough diffusion which is required for a robust image authentication scheme. In fact, the required diffusion in image authentication schemes is not evaluated as in classical cryptographic primitives. It should not hinder the robustness without compromising the security. This issue is used to recover the value of the second key uniquely. To explain this property, let the image I consisting of the pixels all with the same intensity value and its authentication embedded version J are given. The attacker manipulates a pixel value P_i at a chosen position $(x_i; y_i)$ such that $P'_i = P_i + d_i$. The choice of d_i and the constant pixel intensities will be explained later. To bypass the influence of the first key, it is supposed that the parity check (and therefore the authentication data) of the pixel p_i is $(000)_2$ or $(111)_2$. With this assumption reordering the bit positions does not influence on the recovered value. To satisfy the latter assumption, the four MSB values of P_i must be one of the possible values $(0000)_2$, $(0111)_2$, $(1000)_2$ and $(1111)_2$ based on the Hamming code generator matrix. The strategy of the attack is to manipulate one-pixel intensity value such that its authentication data just modifies the host pixel while embedding. To observe how the attack works, the simplest case is considered. Let $P_i = 0$ and $d_i = 96 = (01100000)_2$. Then the authentication data of P'_i is calculated as $s'_i = (101)_2$ or $s'_i = (011)_2$ or $s'_i = (110)_2$ depending on the corresponding random number R_i. Let the host pixel wherein s_i is embedded be denoted by P_j. According to Equation (2), the updated value of P_j indicated by P'_j would be one of the intensity values $(00000101)_2$,

$(00000011)_2$, $(00000110)_2$ receptively. It is observed trivially that as the four MSBs of P'_j are left unchanged, its authentication data will not be affected as well. This indicates the termination of the diffusion in the embedding process. Now, with the knowledge on the position of modified pixel (j), Equation (1) and aforementioned weakness property the value of the second key can be extracted. This attack in a chosen plaintext model can be easily generalized. It is supposed that the attacker has a black-box oracle access to the image authentication scheme indicating that she can request the watermarked image on any chosen input image. The attack is described in the following steps:

- Step One (offline step): The attacker chooses an image I whose pixels have the same intensity value $P_i = p$. To skip the impact of the first secret key (to ease the attack scenario as mentioned before), p can be selected from the set $[0; 15] \cup [112; 143] \cup [240; 255]$.
- Step Two: The watermark image I' is calculated using the image authentication scheme. It is easy to verify analytically that for some values like $p = 0, 8, 112$, $I = I'$.
- Step Three: The attacker modifies a pixel value P_i as $P'_i = p + d_i$. The choice of d is of great importance and plays a crucial role. d_i is selected such that $d > 0$ and the MSBs of the host pixel will not be influenced however LSBs will be modified.
- Step Four: The attacker calculates the watermarked image I'' using the image authentication scheme.
- Step Five: A binary difference matrix indicating the difference between I' and I'' is calculated as $[Dij] = 1 - d_{(I'ij - I''ij)}$.

The elements corresponding to '1' in the above difference matrix is used to solve the system of linear equations 1 to recover the unknown value k_2. The secret permutation generated by the first secret key using a sequence of random numbers is extracted by an exhaustive search using some chosen MSBs which generate the required parity check values. To launch the attack, the attacker just requires two calls to the image authentication scheme for recovering the second key. Also, secret permutation used before embedding can be extracted with $2N^2$ calls resulting in maximum $2N^2 + 2$ total calls with chosen images. Considering the typical image sizes corresponding to N = 512, 1024, the complexity is of order 219 or 221 which is negligible in cryptanalysis.

This attack can be extended to the earlier version of the Chan's image authentication scheme presented in [3] trivially. In the initial scheme, the LSB replacement is used instead of the Modulus function and the pixels are being processed one by one from left to right according to the Torus automorphism. Also the bit reversion on the MSBs are not applied. The attack conditions are met much easier in this case. In fact the choice of difference value di in the third step is less restricted: all of the values of d which keep the encoded value of $(p + d) >> 4$ unchanged (while changing the LSBs of $p + d$) would be valid for this attack. However the complexity of the attack remains unchanged.

5 Improving and Enhancing the Authentication Model

5.1 Enhancing the Scheme

To enhance the Chan's scheme and increase the security against the aforementioned attack while keeping its original elaborated robustness, the following modifications will be suggested on two steps of the scheme.

1) At first as a preprocessing step, the image undergoes some preprocessing steps including a bilinear interpolation and mapping to a fixed-size square image. An optional low-pass filtering can be applied on the image to provide minor robustness.

2) To start processing the image pixels, four pixels are chosen at random. The main modification applies in calculating the parity check parts of the pixels in the processing step. Similar to the original scheme a Hamming code with parameters $n = 7$ and $k = 4$ is used for encoding purpose. However all four pixels contribute in generating each parity check value in the proposed scheme. The details are as follows: let the four selected pixels are denoted by P_1, P_2, P_3, and P_4 whose first four MSBs are indicated by $p_{i1}p_{i2}p_{i3}p_{i4}$ for $i = 1, 2, 3, 4$.

$$W_1 = p_{11}p_{22}p_{33}p_{44}, W_2 = p_{21}p_{12}p_{43}p_{34}, W_3 = p_{31}p_{42}p_{13}p_{24}, W_4 = p_{41}p_{32}p_{23}p_{14} \quad (3)$$

By this arrangement of the MSBs, each generated authentication bit would be a function of all processed pixels. Also each new nibble W_1, W_2, W_3, W_4 contains all significant values. In the proposed modified scheme four pixels are used in each step for extracting the authentication data. To increase the entropy of the intermediate bits, the nibbles are added via XOR operation with random words (R_1, R_2, R_3, R_4) generated from a key-based pseudo random number generator PRNGk as follows.

$$W'_i = W_i \oplus R_i \quad , i = 1, 2, 3, 4 \quad (4)$$

Some proper Boolean functions can be used instead to introduce some nonlinearity into the scheme.

3) Four 3-bit parity check values as authentication data of four pixels are computed by applying the Hamming code on W'_1, W'_2, W'_3, W'_4 respectively. The authentication bits are firstly concatenated and then permuted using a key-based random permutation before embedding. Finally the next four pixels to be processed are determined by Torus transformation wherein the permuted parity check bits are embedded.

The recovery procedure in the modified image authentication is almost the same as recovery procedure in the original Chan's scheme. In the recovery phase, the original pixel values (MSBs) are recovered by applying reverse permutation and decoding operation followed by XORing with the random nibble words. However in the modified scheme, four pixels are recovered at the same time. Hamming codes are able to detect up to two erroneous bits whereas are able to correct one bit without detection. The possible tampered or manipulated areas of the image are localized further and tried to

be corrected using the correction capability of the error-correcting code. Whenever the occurred errors are beyond the capability of Hamming code (more than one bits) the correction attempt is further continued using adjacent pixels. If the latter trial fails, the erroneous pixels are marked as tampered or manipulated area of the image.

5.2 Extension to 16-bit Grayscale Images

The scheme has been so far set for the cases wherein images are grayscale and the pixel depth is an 8-bit value. However, there are many applications specially in medical images in which the pixel depth is 16 bits. The proposed scheme with the existing coding structure is not applicable in such cases due to the code word length. To extend the scheme, the Hamming error correcting code with parameters $n = 15$, $k = 11$ ($[15, 11, 3]_2$) with higher coding rate can be used. In the latter code, four parity check bits for eleven MSBs of each pixel are generated and embedded into the four LSBs of other pixels values corresponding to the revised algorithm described in Section V. However the embedding procedure is modified slightly and eleven pixels are selected in each step to generate four parity check words for embedding in the eleven target pixels determined by Torus transformation (the value of k is set to 4). When the image size number is not a multiple of 11 which is in the most of the cases, the last pixels to be processed are processed naively one by one which does not significant impact on the security. Similar to the initially proposed scheme, each selected pixel will contribute in generating each parity check word. However to avoid further computational cost and size adjustment problem one can update four pixels in each step like the latter enhanced scenario as follows.

$$
\begin{aligned}
W_1 &= p_{11}p_{22}p_{33}p_{44}p_{15}p_{16}p_{17}p_{18}p_{19}p_{110}p_{111} \\
W_2 &= p_{21}p_{12}p_{43}p_{34}p_{25}p_{26}p_{27}p_{28}p_{29}p_{210}p_{211} \\
W_3 &= p_{31}p_{42}p_{13}p_{24}p_{35}p_{36}p_{37}p_{38}p_{39}p_{310}p_{311} \\
W_4 &= p_{41}p_{32}p_{23}p_{14}p_{45}p_{46}p_{47}p_{48}p_{49}p_{410}p_{411}
\end{aligned}
$$

Further processing is the same as the enhanced scheme in the latter subsection in which the 11-bit words are added via XOR operation with random words of the same size before encoding. The generated parity check bits are permuted and then embedded into the LSBs accordingly. Like the original scheme, erroneous bits can be detected or one can be corrected without detection of other erroneous bits. However the security of the scheme will be strengthened due to longer parity check values. This point is addressed in the following section.

6 Revisiting the Analysis of the Modified Scheme

The main security aspect of the modified scheme is that any significant changes in a pixel (MSBs) will be reflected to up four host pixels which was one pixel at max in the original Chan's scheme. Also expanding the pixels processing to four branches simultaneously

instead of one in the original scheme will increase the diffusion. However, it is worth to mention that the robustness of the scheme is not influenced in the proposed modification due to using the same coding and embedding building blocks. Here it is shown that the proposed modification strengthens the security of the original scheme significantly against the mentioned key recovery attack. Let the image be of size $N \times N$. To recover the first secret key corresponding to Torus automorphism uniquely, the attacker does not have difficulties due to Equation 1 and negligible image size in this sense. However the cost of finding the positions is of order $O(N^6)$. Also the cost of recovering the secret random permutation is approximately 12!. So the whole computational complexity required for recovery of secret components excluding the generated random numbers would be $C = O(N^6) + 16 \times 12!$. The generated random numbers must be disclosed via exhaustive search as well which increases the complexity.

The complexity would be higher for the extended version as well. In fact, the cost of recovering secret random permutation used in the parity check part is 16! which simply affects the second term of the equation. Concerning the commonly used values for N and pixel depth, the attack complexity C has increased significantly however some low thresholds on N and pixel depth can be set to satisfy desired security strength according to the existing computational power. To enhance the security more, a keyed pseudo-random function whose label is dependent to the image can be used instead to embed the generated random output into the MSBs. In this case the labeled function provides the intrinsic security strength of a cryptographic authentication function. Further improvement in the accuracy of the tamper localization is possible as well by using another code structure with higher detection and correction capability. The latter suggestions are considered as future work since there is not enough room left in this paper.

7 Conclusion

An image authentication scheme based on Hamming code and rearrange MSBs of pixel intensity values has been analyzed. Some modifications in preprocessing step and computation of authentication data have been proposed to enhance the security and accuracy of the scheme. The proposed modifications strengthen the scheme against the aforementioned flaw and generalize the scheme to be used on grayscale images with 16 bits pixel intensity values. However there is still some room left for further enhancement to elaborate the design more. Using different types of systematic error-correcting schemes which increases the detection and correction capability and engaging a keyed labeled cryptographic pseudorandom function would be a potential track for the improvement and further secure designs.

References

[1] O. Koval, S. Voloshynovskiy, F. Beekhof, and T. Pun, “Security analysis of robust perceptual hashing,” in Electronic Imaging 2008, pp. 681906–681906–10, International Society for

Optics and Photonics, 2008.

[2] C.-S. Chan, "An image authentication method by applying hamming code on rearranged bits," Pattern Recognition Letters, vol. 32, no. 14, pp. 1679 – 1690, 2011.

[3] C.-S. Chan and C.-C. Chang, "An efficient image authentication method based on hamming code," Pattern Recogn., vol. 40, pp. 681–690, Feb. 2007.

[4] A. Haouzia and R. Noumeir, "Methods for image authentication: a survey," Multimedia Tools and Applications, vol. 39, no. 1, pp. 1–46, 2008.

[5] Ling Du, Anthony T.S. Ho, Runmin Cong, Perceptual hashing for image authentication: A survey, Signal Processing: Image Communication, Volume 81, 2020, pp. 115713.

[6] M. Sajjad, I. U. Haq, J. Lloret, W. Ding and K. Muhammad, "Robust Image Hashing Based Efficient Authentication for Smart Industrial Environment," in IEEE Transactions on Industrial Informatics, vol. 15, no. 12, pp. 6541-6550, Dec. 2019, doi: 10.1109/TII.2019.2921652.

[7] Karsh, R.K. LWT-DCT based image hashing for image authentication via blind geometric correction. Multimed Tools Appl 82, pp. 22083–22101, 2023.

[8] Thabit, R. Review of medical image authentication techniques and their recent trends. Multimed Tools Appl 80, 13439–13473, 2021.

[9] A. Swaminathan, Y. Mao, and M. Wu, "Robust and secure image hashing," Information Forensics and Security, IEEE Transactions on, vol. 1, pp. 215 – 230, 2006/06// 2006

[10] G. Zhu, J. Huang, S. Kwong, and J. Yang, "A study on the randomness measure of image hashing," IEEE Transactions on Information Forensics and Security, vol. 4, pp. 928–932, Dec 2009.

[11] Y. Mao and M. Wu, "Unicity distance of robust image hashing," IEEE Transactions on Information Forensics and Security, vol. 2, pp. 462–467, Sept 2007.

[12] O. Koval, S. Voloshynovskiy, P. Bas, and F. Cayre, "On security threats for robust perceptual hashing," in Media Forensics and Security, vol. 7254, p. 72540H, feb 2009.

[13] T. Uehara and R. Safavi-Naini, "On (In)security of A Robust Image Authentication Method", pp. 1025–1032. Berlin, Heidelberg: Springer Berlin Heidelberg, 2002.

[14] M. Heidari, S. Samavi, S. M. R. Soroushmehr, S. Shirani, N. Karimi, and K. Najarian, "Framework for robust blind image watermarking based on classification of attacks," Multimedia Tools and Applications, Nov 2016.

[15] D. Hu, B. Su, S. Zheng, and Z. Zhang, "Secure architecture and protocols for robust perceptual hashing," in Computational Intelligence and Security (CIS), 2013 9th International Conference on, pp. 550–554, Dec 2013.