World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(REVIEW ARTICLE)

Check for updates

Leveraging AI/ML for anomaly detection, threat prediction, and automated response

Olakunle Abayomi Ajala [1,*] and Olusegun Abiodun Balogun [2]

[1] *Department of Management, Indiana Wesleyan University USA.*
[2] *Department of Mechanical Engineering, Jomo Kenyatta University of Agriculture and Technology, Kenya.*

World Journal of Advanced Research and Reviews, 2024, 21(01), 2584–2598

Publication history: Received on 14 December 2023; revised on 22 January 2024; accepted on 25 January 2024

Article DOI: https://doi.org/10.30574/wjarr.2024.21.1.0287

Abstract

The rapid evolution of information and communication technologies, notably the Internet, has yielded substantial benefits while posing challenges to information system security. With an increasing frequency of cyber threats—from unauthorized access to data breaches—the digital landscape's vulnerability is evident. Addressing the financial impact of cybercrime, this study delves into the role of Artificial Intelligence (AI) and Machine Learning (ML) technologies in cybersecurity. Analyzing advancements and outcomes, the research explores practical techniques for anomaly detection, threat prediction, and automated response. By investigating prior research and real-world implementations, the study provides valuable insights into the potential of AI/ML, uncovering current trends, challenges, and prospects in enhancing cybersecurity tactics amid a dynamically changing threat landscape.

Keywords: Information security; Cybercrime; Cybersecurity; Artificial Intelligence (AI); Machine Learning (ML); Anomaly detection; Threat prediction; Automated response

1. Introduction

The swift advancement of information and communication technologies, notably the Internet, has ushered in favorable outcomes for both organizations and individuals. The Internet serves as a vital platform for fostering communication and networking, facilitating knowledge exchange [1], and enabling social interaction [2], all of which play crucial roles in human progress. However, alongside these advantages, there exists a shadowy facet. The growing dependence on third-party, backend users and cloud-based data storage and applications within the digital landscape has rendered it exceedingly challenging and vulnerable for organizations to ensure comprehensive security for their information systems.

In recent years, there has been a dramatic surge in information security incidents, encompassing unauthorized access [3], denial of service (DoS) attacks [4], malware intrusions [5], zero-day exploits [6] , data breaches [7], and social engineering tactics like phishing [8]. This surge can be attributed to the ever-increasing significance of information technology in our lives. To illustrate, consider that in 2010, the security industry was aware of fewer than 50 million unique malware executables. By 2012, this number had doubled to surpass 100 million. More recently, in 2019, the security sector identified a staggering 900 million malicious executables, and this figure continues to grow, as indicated by AV-TEST statistics [9].

The repercussions of cybercrime and network attacks can be financially devastating for both businesses and individuals. For instance, research reveals that the average global cost of a data breach is USD 8.19 million, with a cost of USD 3.9 million in the United States [10]. Moreover, cybercrime exacts a hefty toll on the global economy, amounting to USD 400 billion annually [11]. Security experts project that the number of breached records is poised to triple in the next five

* Corresponding author: Ajala Olakunle Abayomi

years [12]. Consequently, organizations must develop and implement comprehensive cybersecurity strategies to mitigate potential future losses.

It is quite important that the security of countries, nations, depends vigorously on people, government substances, and those with admittance to information, applications, and high-exceptional status apparatuses, as demonstrated by late financial exploration discoveries [13].

Besides, endeavors are being constrained to look for creative arrangements that go past customary safety efforts because of the sharp expansion in computerized dangers and weaknesses in the always changing online protection scene. Irregularity identification, danger expectation, and robotized response have become fundamental components in the fight against digital assailants. In this unique situation, there has been a great deal of interest in the capability of artificial intelligence and ML innovations to reinforce the security stance of associations and organizations. [14].

ML offers a pragmatic method for propelling online protection endeavors in view of their expedient examination of huge datasets and capacity to recognize patterns. This study examines the viability of network protection by consolidating artificial intelligence/ML into the spaces of mechanized reaction, danger forecast, and peculiarity discovery. Utilizing computer-based intelligence/ML calculations' capacities, associations might have the option to expect and defeat hurtful action, answer immediately to arising dangers, and invigorate their protections [15].

This article investigates the province of simulated intelligence/ML applications in online protection, displaying striking turns of events and their results. It likewise analyzes the strategies, calculations, and models that have been effective in distinguishing irregularities and anticipating risks, illustrating both their benefits and burdens.

This examination intently looks at robotized reaction frameworks and the Normal Shortcoming Specification (CWE) Classification of Weaknesses, with an emphasis on how computer-based intelligence/ML can be utilized to rapidly and precisely foster countermeasures against cyberthreats. through an exhaustive investigation of the corpus of past examination, contextual investigations, and genuine executions. The motivation behind this study is to introduce astute perspectives on the changing field of network safety empowered by man-made intelligence and ML. Eventually, it adds to a superior comprehension of how simulated intelligence/ML can be utilized in contemporary online protection methodologies for danger expectation, automated reaction, and irregularity recognition.

2. Literature Review

2.1. An Outline of Artificial intelligence and ML Approaches for Peculiarity Location in Network protection

The utilization of AI in network protection has drawn in a ton of consideration of late. An exhaustive review that tended to potential exploration bearings, gave top to bottom measurements and conveyances of the studied drives, and analyzed the ongoing Artificial intelligence applications in network safety was done. An alternate examination project has offered a careful comprehension of Artificial intelligence driven network protection, including the speculations and ideas that help savvy and mechanized network safety administrations and organization. Data security occurrences have changed decisively after some time. Throughout the course of recent years, there has been a perceptible ascent in unapproved access [16], disavowal of administration (DoS) assaults [17], malware interruptions [18], zero-day takes advantage of [19], information breaks [20], social designing or phishing endeavors [21], and other related episodes.

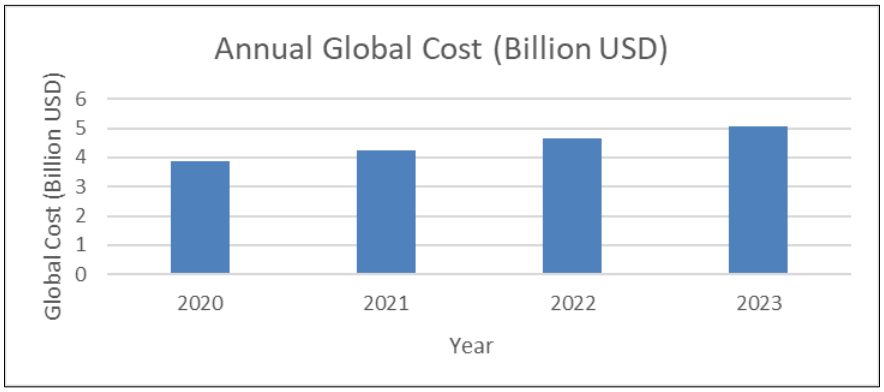

Figure 1 The cost implication of recovering from data breaches as increased globally over the years.

This rise is primarily attributable to the increasing importance of information technology.

To put things in perspective, the security industry documented fewer than 50 million unique malware executables in 2010. This figure doubled to over 100 million by 2012, and the security industry identified over 900 million malicious executables by 2019 [22]. Businesses and individuals alike will suffer grave financial consequences as a result of this concerning increase in cyber-incidents. According to studies, the average cost of a data breach can reach USD 8.19 million worldwide and USD 3.9 million in the US [23]. More broadly, the annual cost of cybercrime to the world economy is estimated to be $400 billion [24]. Regretfully, forecasts from the security community suggest that over the next five years, the quantity of compromised records is likely to triple [25].

Hence, organizations must establish comprehensive cybersecurity strategies to mitigate impending losses proactively. The nation's security, as underscored by the latest socioeconomic research, relies heavily on safeguarding data, applications, and instruments demanding high-security clearance [26]. Equally significant is the role of businesses in providing their staff with the necessary tools and knowledge to swiftly recognize and address cyber threats. Thus, the critical challenge is the intelligent identification of a spectrum of known and unknown cyber events to safeguard vital systems against cyberattacks.

This involves the application of cybersecurity technologies and protocols to shield computer programs, networks, computers, and data from harm, attacks, or unauthorized access by external parties [27]. Cybersecurity is a multifaceted domain encompassing various categories that span diverse contexts, such as business and mobile computing. These categories encompass network security (focused on preventing unauthorized access to computer networks), application security (aimed at averting cyber threats from compromising hardware and software), operational security (pertaining to the safeguarding of data assets through defined protocols), and information security, which primarily concerns the privacy and security of critical data. Conventional cybersecurity solutions, including firewalls, antivirus software, and intrusion detection systems, have been the norm.

However, a notable transformation is underway, driven by data science. Machine learning, a pivotal facet of "Artificial Intelligence," is emerging as a potent tool for unearthing concealed patterns within data. This transformation holds significant implications for the cybersecurity landscape and heralds a new era in science [28]. As elaborated in the referenced article [29], advancements in cyber threat-related technologies are continuously pushing the boundaries, with attackers demonstrating increasing sophistication, necessitating an adaptable approach to security.

In summary, within the context of AI and ML approaches for anomaly detection in cybersecurity, the escalating challenges of safeguarding digital domains underscore the critical role of advanced technologies and proactive strategies. Cybersecurity is evolving to embrace machine learning and data science, heralding a transformative era in combating emerging threats.

In this study, concentration is made on cybersecurity machine learning, which is strongly related to these fields in terms of security, intelligent decision-making, and the data processing techniques to be used in practical applications. The overall goal of this project is to predict cyber-hazards and improve cybersecurity procedures utilizing security data and machine learning techniques. Researchers from academia and industry who are interested in researching and creating machine learning-based data-driven smart cybersecurity models can also benefit from this project.

Cybersecurity data analysis and tool development are necessary to successfully process cybersecurity data to prevent attacks beyond basic functional requirements and awareness of risks, threats, or vulnerabilities. To efficiently extract the insights or patterns of security occurrences, a variety of machine learning techniques, including but not limited to feature reduction, regression analysis, unsupervised learning, identifying connections, or neural network-focused deep learning techniques, can be applied. The "Machine learning techniques in cybersecurity" section briefly touches on this. These learning approaches can identify irregularities, malicious behavior, and data-driven patterns of associated security risks. They can also make wise decisions to stop cyberattacks. Traditional well-known security solutions, such as user authentication and access control, firewalls, and cryptography systems, may or may not be effective in fulfilling today's cyber business needs. Machine learning is a partial but considerable departure from these solutions [30]. When ad hoc data management is required, the primary challenge is that domain experts and security analysts manually solve these [31]. Traditional solutions, however, have shown to be ineffectual in addressing these cyber hazards as a growing number of cybersecurity incidents in diverse formats emerge over time. As a result, several fresh, intricate attacks surface and spread quickly throughout the network. To create cybersecurity models, as discussed in the section "Machine learning techniques in cybersecurity," several academics use various data analytic and knowledge extraction models. These models are based on the effective identification of security insights and the most recent security trends that may be more pertinent. According to studies, solving the cyber problem requires the creation of more adaptable

and effective security systems that can respond to attacks and intelligently update security rules to stop them in their tracks. To accomplish this, it is necessary to examine a sizable amount of pertinent cybersecurity data gathered from many sources, including network and system sources. Furthermore, these methods ought to be used with little to no human involvement, increasing automation.

2.2. Challenges and Solutions in Implementing AI-based Automated Response Systems for Cyber Threats

The protection of theater-wide networks from cyber threats involves the cooperative tracking of intrusions and the sharing of attack-related information and response strategies across interconnected domains. Smith (2002) [32] expanded on the Intruder Detection and Isolation Protocol (IDIP), which combines intrusion detection systems and cooperative boundary controllers within a single administrative domain to trace network intruders back to their source and dynamically adjust network-level access controls to thwart real-time attacks.

This document was created to provide guidance on ensuring secure data transfer within the intricate computational infrastructure representative of the electric power, oil, and natural gas sectors, including the control systems they employ [33]. Over the past two decades, the focus of the cybersecurity community has predominantly centered on preventive measures aimed at fortifying systems with a robust outer layer that resists penetration.

The text introduces key concepts such as risk assessment, threat analysis, and network forensics, complemented by online access to a wealth of supplementary materials, including labs, Cisco challenges, test questions, and web-based videos. It covers various technologies like Wireless LAN Controllers (WLC), Wireless ADUs, ASDMs, SDMs, Juniper, and more. Additionally, it delves into advanced topics, including data hiding, obfuscation, web infrastructures, and cloud and grid computing [34].

The evolution of Industry 4.0 and the Industrial Internet Consortium, coupled with the advent of the Internet of Things and Cyber Physical Systems, has been instrumental in driving this transformation, as explored by Lesjak et al. (2015) in securing smart service use cases for industrial maintenance scenarios. Lakhno (2016) [35] investigates the creation of an adaptive cyber threat detection system based on fuzzy feature clustering. A categorical model for constructing the adaptive intelligent cyber threat detection system (ICTDS) is proposed, addressing complex issues in CIS cyber defense control and software solutions for cyber defense systems.

In some instances, insufficient attention to security during design and implementation can leave systems vulnerable to cyber-attacks, a concern tackled by [36] with the introduction of SoftGrid, a software-based smart grid testbed for assessing security solutions aimed at safeguarding remote control interfaces of substations.

Cyber deception strategies come into play for gathering data on botnets, spreading worms, and detecting hidden persistent external attackers and insider threats. Väisänen (2017) [37] presents an automated categorization approach for assessing the severity of information from decoys, offering solutions for these challenges. For an automated low altitude, small unmanned aircraft traffic management system, Ong et al. (2017) [38] proposed a short-term conflict avoidance algorithm that generates advisories for individual aircraft based on a multiagent Markov decision process.

Understanding the role that humans play in Organization Cyber-Physical Systems (OCPS), [39] explore the functionality of user and entity behavior analytics (UEBA). In order to function, UEBA gathers a variety of data, including user roles and titles, access, accounts, and permissions; user activity and location; and security alerts. This information can be gathered from past and present activity, and the analysis compares anomalous behavior to peer group activity by taking into account elements including the resources consumed, the length of sessions, connectivity, and peer group activity [40]. When modifications are made to the data, such as promotions or new permissions, it also automatically updates.

Not every anomaly is necessarily reported as dangerous by UBA and UEBA systems. Instead, they consider the consequences of the behavior. The conduct has a low effect score if it involves fewer sensitive resources. A greater impact score is given if it concerns more delicate information, including personally identifiable data. As the UBA system immediately restricts or makes it more difficult to authenticate the user exhibiting suspicious behavior, security teams can prioritize what to investigate. Behaviors that UBA and UEBA systems typically watch for are those linked to certain assaults, including brute-force attacks [41], improper data access, data loss [42], transfer of data by an unauthorized user.

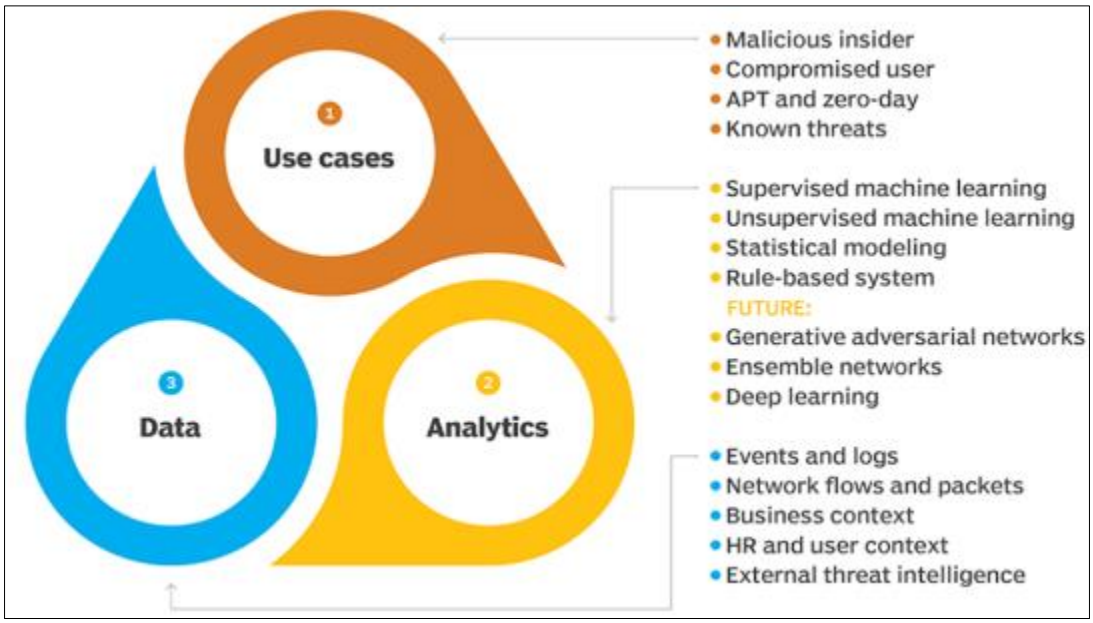

Figure 2 User and entity behavior analytics work cycle [43]

2.3. Issues with Scalability in AI-Powered Threat Recognition and Reaction Systems

Through AI-driven threat detection and response (TDR) systems, organizations of all sizes can enhance their capacity to recognize and address cyber threats, thereby transforming the cybersecurity field [44]. But as these systems proliferate, they are encountering new difficulties, especially with regard to scalability because of the enormous volume and complexity of data that they have to handle. AI-driven TDR systems collect data from a variety of sources, such as endpoint devices, network traffic, and security logs, which can be complex and challenging to manually analyze [45]. Additionally, since cyber theft are frequently creating modern technological methods and software to intrude into detection, AI-powered TDR systems must keep up with the constantly shifting threat landscape [46]. The effectiveness of AI-driven threat detection systems is greatly influenced by the quantity and quality of the data used in model training [47], which also provides guidance on how to train deep learning models with large amounts of data.

3. AI-Driven Data Collection and Analysis

Threat intelligence is the process of gathering and analyzing information about actual or possible attacks that might jeopardize an organization. Threat intelligence is centered on the analysis and interpretation of data to detect threats, pinpoint early warning signs, and put defensive measures in place. Artificial intelligence (AI) is a major factor in threat intelligence since it helps to automate data storage, analysis, and collection. Handling the enormous volumes of data produced in today's digital environment is made possible by this.

Machine learning (ML) and other AI techniques are used in AI-powered threat intelligence to find patterns and anomalies that may be signs of approaching danger. It excels at spotting patterns and trends in sizable datasets, foreseeing impending attacks, and offering practical suggestions for lowering risk. AI-driven threat intelligence has several advantages, such as enhanced predictive power, speedier threat identification, and effective handling of massive data volumes. A Comprehensive case study on cybersecurity risks that centers on vulnerabilities found in 2022 after a sizable dataset from the Cybersecurity and Infrastructure Security Agency (CISA) was analyzed. Comprehending the current weaknesses is essential for formulating efficient defensive tactics and eliminating possible hazards. The Common Vulnerabilities and Exposures (CVE), vendor and product details, vulnerability names, dates of discovery, severity ratings, and suggested actions are just a few of the many details about multiple vulnerabilities that can be found in the dataset that was retrieved from www.kaggle.com. examining this data to look for important trends, patterns, and insights. Throughout this study, various methods and approaches were employed to capitalize on the potent data analysis and visualization features of the R programming language, including clustering, network analysis, and Bayesian analysis.

3.1. AL Libraries

The R programming language loads a set of R packages that are a part of the Tidyverse database using the library(tidyverse) statement. Popular tools for data analysis, visualization, and manipulation like ggplot2, tidyr, and dplyr are included in these packages.

3.2. Dataset Preparation and Correlation

A correlation plot is a graphical representation in R that shows the connections or relationships between multiple variables. It is commonly employed in exploratory data analysis to determine the relationships between various variables within a dataset. When dealing with numerical data, correlation graphs are especially helpful for spotting patterns, connections, and interdependencies between variables. In R-Programming, correlation plots are useful tools that help with feature selection, data analysis, and model building decisions. They can be used to comprehend correlations within a dataset. As seen in figure 3, they support modeling and data exploration activities and offer a succinct synopsis of the relationships between the variables. The correlation coefficients between two variables, like severity, can be seen visually with correlation plots. The strength and direction of the linear relationship between the variables are thus shown by the correlation coefficients. A high positive correlation (near 1) means that the variables represented by the combination of severity and sevrity move in the same direction, whereas a high negative correlation (near -1) means that the variables move in opposite directions, as can be seen in the legend in figure 3 below. A helpful tool for identifying which variables have a strong association and which do not is a correlation plot. For instance, there is a clear correlation between the severity and the variables. This is important when selecting features for machine learning, as you may want to choose the variables that are most relevant to a prediction model. Correlation charts are often represented by heatmaps, where the intensity of each color represents the strength of the correlation; dark blue, for example, is typically associated with high positive correlations; dark red, for example, is associated with large negative correlations; and neutral colors, like white or light gray, are associated with weak correlations. The distribution of vulnerability severity levels is shown in Figure 4, which can help reduce dimensionality and facilitate the interpretation of complex datasets.

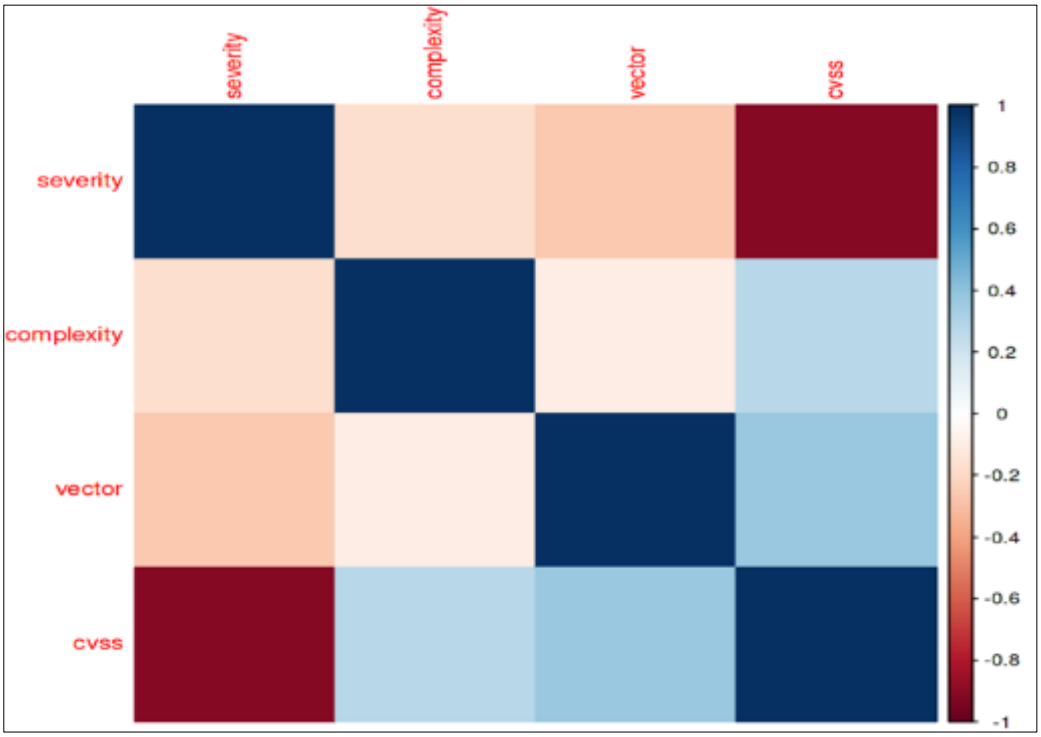

Figure 3 Convolution Matrix

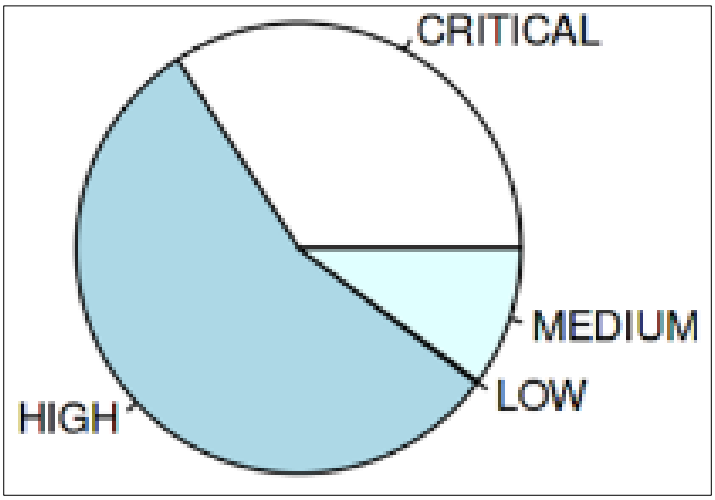

Figure 4 Distribution of Vulnerability Severity Levels

3.3. Examination of Image Trend and Number of Vulnerabilities

Figure 5 illustrates the vulnerability risk over time, offering a concise summary of the escalating number of disclosed vulnerabilities in the past two years. This upward trend is likely influenced by the growing usage of technology, the intricate nature of software and systems, and the increasing sophistication of attackers. Such a pattern holds significant implications for organizations, as attackers can leverage vulnerabilities for launching attacks, accessing systems and data, or pilfering information. As a result, businesses may experience a range of issues, such as monetary losses, harm to their brand, and operational pauses. Figure 6 below, which displays the Trend Vulnerability Counts by Severity, illustrates that over the past year, there have been more vulnerabilities across all severity categories. The greatest increase in vulnerabilities has been seen in those with high and critical severity, with medium and low severity coming in close second.

The growing complexity of software and systems, the rise in the number of people utilizing technology, and the growing expertise of attackers are all likely contributing factors to this trend. Organizations are significantly impacted by this trend. Attackers can launch attacks, obtain access to systems and data, or steal information by using vulnerabilities of any severity. As a result, businesses may experience a range of issues, such as monetary losses, harm to their brand, and operational pauses. Companies need to be proactive in protecting themselves from vulnerabilities that range in severity. This entails spotting vulnerabilities and fixing them as soon as it's practical, putting security measures in place to lessen the chance that vulnerabilities will be exploited, and training staff members on security best practices.

3.4. Effect on organizations

For companies of all kinds, the growing number of vulnerabilities is a serious concern. In order to safeguard themselves against vulnerabilities, companies must implement these measures:

- Identifying weaknesses and fixing them: As soon as system and software vulnerabilities are discovered, businesses must find them and fix them. Numerous tools and methods, including vulnerability scanning, penetration testing, and code review, can be used to achieve this.
- Putting security controls in place: In order to lessen the chance that vulnerabilities will be exploited, organizations must put security controls in place. Firewalls, access control systems, and intrusion detection systems are a few of these controls.
- Employee education: Organizations must provide training to staff members on security best practices and how to fend off attacks that take advantage of weaknesses.

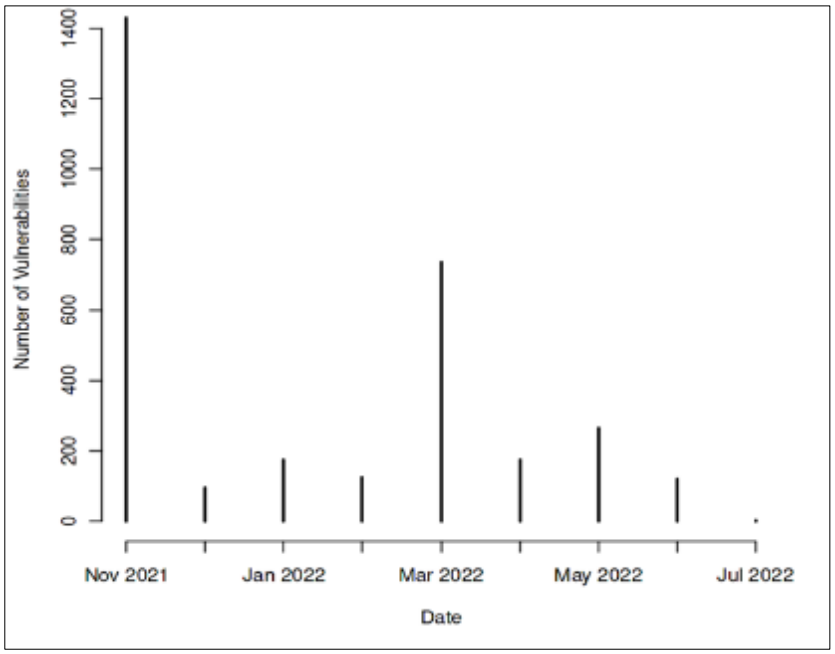

Figure 5 Trends Vulnerability Counts

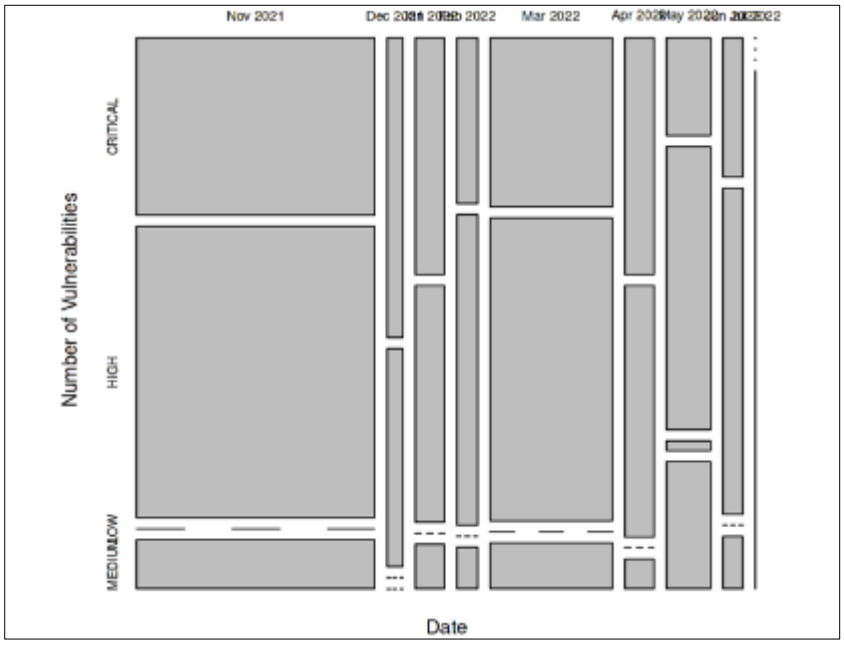

Figure 6 Trends Vulnerability Counts by severity.

Figure 7 below displays a pie chart that illustrates the distribution of vulnerabilities by vector. The pie chart indicates that remote attack vectors account for 65 percent of vulnerabilities. This implies that physical access to the target computer is not necessary for an attacker to take advantage of these vulnerabilities.

Adjacent network attack vectors rank second in terms of vulnerability severity, accounting for 20% of all vulnerabilities. This implies that attackers with access to the same network as the target system may be able to take advantage of these vulnerabilities. 15% of vulnerabilities are caused by local attack vectors. This implies that attackers with physical access to the target computer are the only ones who can take advantage of these vulnerabilities. It is important to keep in mind that the majority of vulnerabilities originate from remote attack vectors. As a result, companies should concentrate their security efforts on eliminating distant threats. Security measures like firewalls, intrusion detection systems, and access

control systems can be put into place to achieve this. Companies also need to be mindful of the weaknesses that hackers, gaining entry to the same network as the target computer, may exploit against them. Businesses that operate in high-risk sectors like finance and healthcare should pay particular attention to this.

Lastly, businesses must take precautions against deliberate assaults. This entails setting up both technological and physical security measures, such as entry control and CCTV cameras.

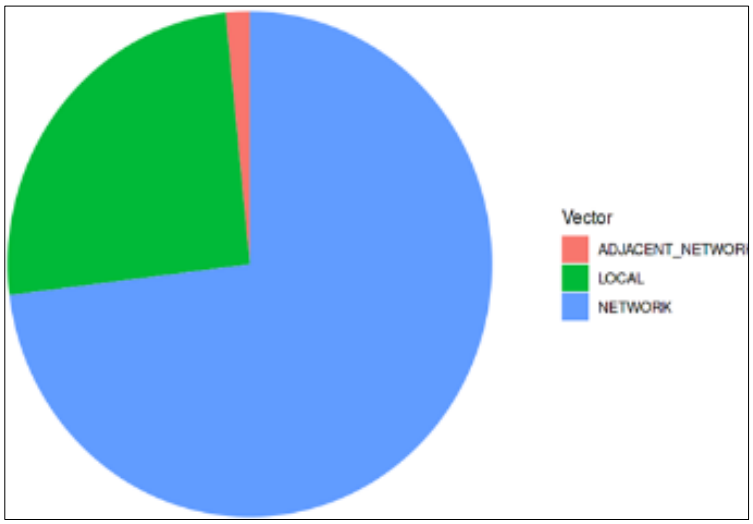

Figure 7 Distribution of Vulnerability by Vector**.**

Based on information from the National Institute of Standards and Technology (NIST), Figure 9 enumerates the top 10 vendors with the greatest number of vulnerabilities. Citrix, Oracle, Apple, Google, Cisco, Microsoft, VMware, Adobe, and Pulse Secure are the top ten vendors.

This information is important because it demonstrates that vulnerabilities exist even in the biggest and most reputable technology companies. Businesses that utilize the software from these providers are required to be aware of the vulnerabilities and take the appropriate action to address them.

Vulnerabilities might make a big difference. Attackers can use vulnerabilities to initiate attacks, obtain access to systems and data, or steal data. Organizations may face numerous issues as a result of this, such as financial losses, damage to their brand, and interruptions to business operations. Organizations can take a number of steps to reduce the possibility that their vulnerabilities will be exploited, including:

- Determining and addressing vulnerabilities: Companies need to identify and address system and software vulnerabilities as soon as they are found. To achieve this, a variety of instruments and techniques can be applied, such as code review, vulnerability scanning, and penetration testing.
- Implementing security controls: Organizations need to implement security controls to reduce the likelihood that vulnerabilities will be exploited. Among these controls are intrusion detection systems, firewalls, and access control systems.
- Employee education: Companies need to train employees on security best practices and how to defend against exploitative attacks.

Companies using software from the top ten vendors listed above should be particularly mindful of the software's vulnerabilities. To do this, businesses can visit the NIST website and search for software vulnerabilities that affect their software. Companies can also register to receive security alerts from the vendors they collaborate with.

By taking these precautions, organizations can reduce the risk that their vulnerabilities will be exploited and protect themselves from the negative consequences of security breaches.

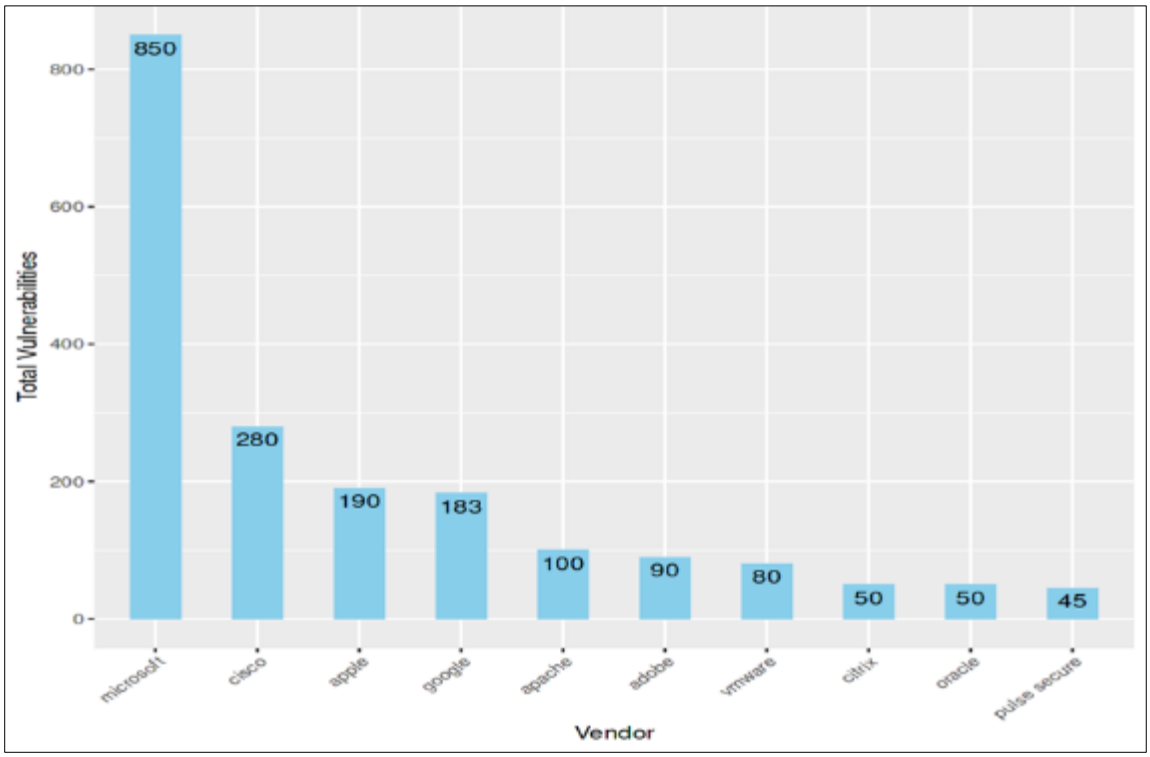

Figure 8 Top Ten Vendor with highest number of vulnerabilities.

The figure 8 above shows the top 10 vendors with the highest number of vulnerabilities in 2022, according to CVE Details shown in figure 8 above.

With 850 vulnerabilities, Microsoft has the most, according to the graph. With 800 vulnerabilities, Cisco comes in second, followed by Apple with 600 vulnerabilities. Both Google and Apache have 400 vulnerabilities, which ties them for fourth place. There are 380 vulnerabilities with Adobe, 280 with VMware, 200 with Citrix, 190 with Oracle, and 183 with Pulse Secure.

There are various cybersecurity concerns for this graph. Firstly, it demonstrates that the most vulnerable software suppliers are also the biggest and most recognizable ones. This is because there are several ways for attackers to discover and take advantage of software vulnerabilities due to these companies' extensive attack surface. Second, the graph indicates that there are more vulnerabilities every year. This can be attributed to various variables, such as the expanding quantity of software applications, the complexity of software, and the intelligence of attackers.

Cybersecurity is negatively impacted by the growing number of software vulnerabilities. It first makes maintaining the security of an organization's systems more challenging. Companies must promptly fix vulnerabilities, but it can be challenging to stay on top of the ever-increasing number of vulnerabilities.

Second, attackers find it simpler to exploit systems as a result of the growing number of vulnerabilities. Vulnerabilities can be used by attackers to break into systems, steal information, or initiate attacks. Third, maintaining a secure posture costs more for enterprises due to the growing number of vulnerabilities. For their systems to remain secure, organizations must spend in security staff and equipment.

To lower their risk exposure, organizations should also think about utilizing software from a range of providers. A software vulnerability in a single vendor's product is more likely to impact an organization that uses only that vendor's software. Utilizing software from many manufacturers can help organizations lower their vulnerability to a single flaw.

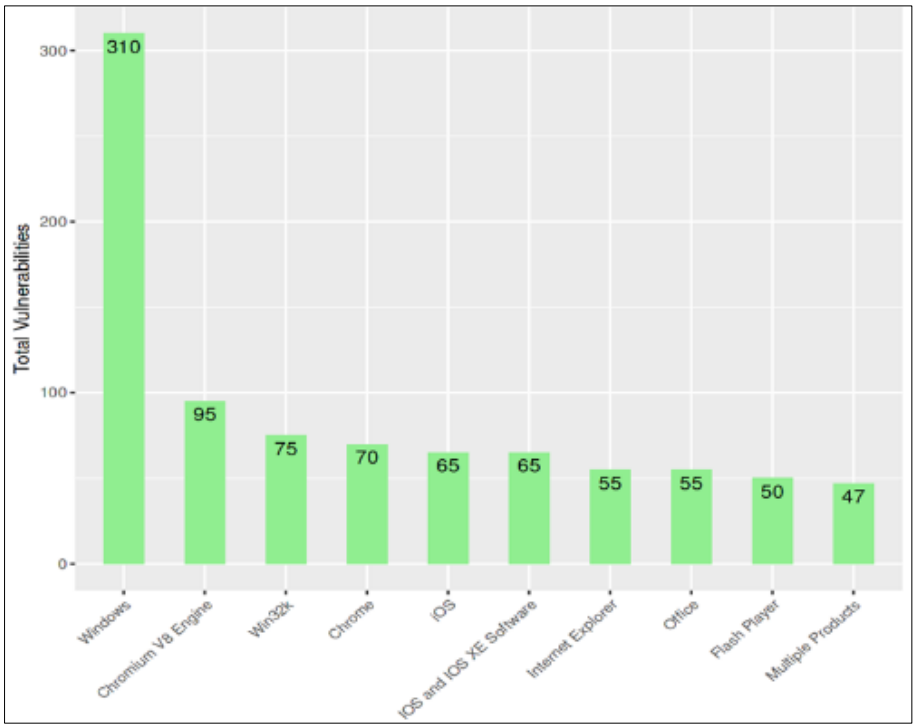

Figure 9 Top Ten Applications with highest number of vulnerabilities.

3.5. Top Ten Applications with highest number of vulnerabilities

The top 10 apps with the most vulnerabilities in 2022 are depicted in figure 9 above, which has many ramifications for cybersecurity. Initially, it demonstrates that the most extensively used and well-liked programs are also the most susceptible. This is a result of the enormous attack surface these apps possess, which gives hackers a wide range of options for locating and taking advantage of software flaws.

Secondly, the graph indicates that there are more vulnerabilities in applications every year. This can be attributed to various variables, such as the expanding quantity of software applications, the complexity of software, and the intelligence of attackers. Numerous detrimental implications on cybersecurity result from the growing number of vulnerabilities in applications. It first increases the difficulty of keeping an organization's systems secure. Businesses need to address vulnerabilities as soon as possible, but it can be challenging to keep up with the increasing number of vulnerabilities.

Second, attackers find it simpler to exploit systems as a result of the growing number of vulnerabilities. Vulnerabilities can be used by attackers to break into systems, steal data, or initiate attacks.

Third, it costs more for businesses to maintain a secure posture because there are an increasing number of vulnerabilities. For their systems to continue to be secure, organizations must invest in security personnel and equipment.

In addition to the aforementioned, companies ought to think about the following:

- To prevent intrusions into their web applications, use a web application firewall (WAF). Web application vulnerability-exploiting attacks can be detected and prevented with the aid of a WAF.
- To safeguard their cloud-based apps, use a cloud security platform. The use of a cloud security platform can help detect and prevent attackers trying to take advantage of holes in cloud-based applications.
- Set up a security information and event management (SIEM) system to keep an eye out for any odd activity on their systems.

The identification of attacks that take advantage of application vulnerabilities can be aided by a SIEM system. Organizations can lower the risks associated with the rising number of application vulnerabilities by implementing these measures.

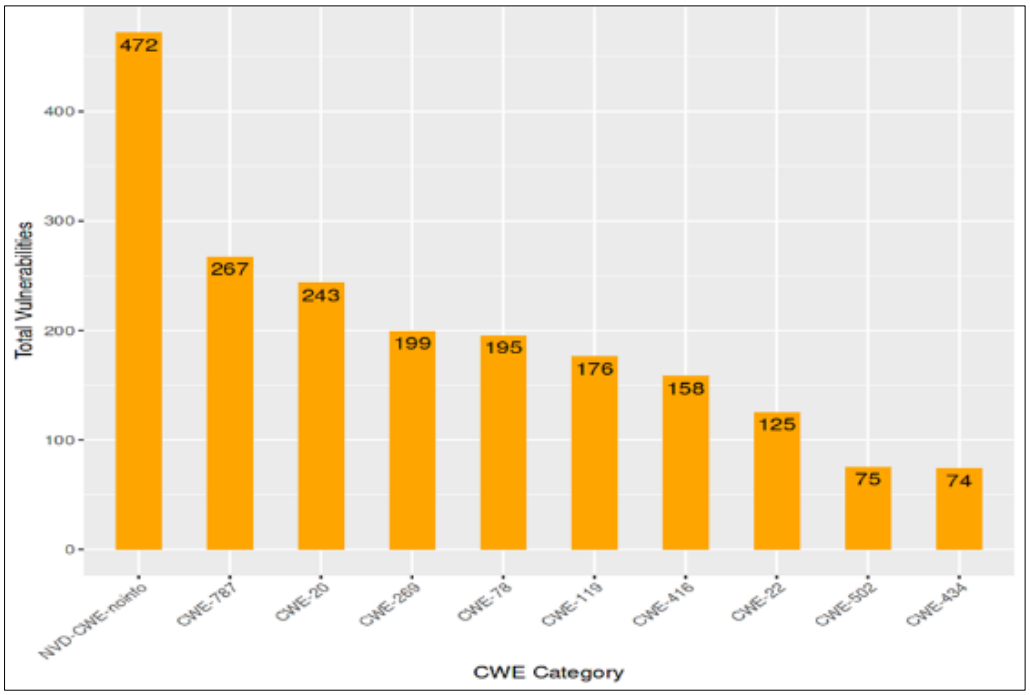

Figure 10 Most Common CWE categories among Vulnerabilities**.**

The most prevalent CWE (Common Weakness Enumeration) categories for vulnerabilities are displayed in figure 10. CWE is a list of software and hardware vulnerability types that the community has developed. The following are the top ten categories:

- CWE-119: Improper Input Validation
- CWE-78: Improper Neutralization of Special Elements used in Output
- CWE-20: Improper Authentication
- CWE-269: Improper Privilege Management
- CWE-787: Out-of-bounds Read
- CWE-416: Use of Hard-coded Credentials
- CWE-22: Path Traversal
- CWE-502: Deserialization of Untrusted Data
- CWE-200: Information Exposure
- CWE-400: Unrestricted Uploading of File with harmful File Type

These categories are all related to common programming errors that could lead to software vulnerabilities. For example, CWE-119 explains how insufficient user input validation exposes a system to the risk of malicious code injection by attackers. Attackers may use cross-site scripting attacks (CWE-78) if special characters in output are not properly neutralized.

Because these vulnerabilities are common, cybersecurity is negatively impacted in a number of ways. It first makes it easier for attackers to exploit systems. Attackers may use these vulnerabilities to gain access to systems, steal data, or launch attacks. Second, it increases the difficulty of keeping an organization's systems secure. Companies must quickly fix these vulnerabilities, but it can be challenging to stay on top of the constantly expanding list of flaws.

The following recommendations address how businesses can lessen the risks associated with the most prevalent CWE categories of vulnerabilities:

- Implement the Security Development Lifecycle (SDL) program. An SDL program is the process of incorporating security into the software development life cycle.
- Make use of the static application security testing (SAST) technology. An SAST tool can be used to identify software vulnerabilities before they are deployed.
- Make use of DAST (dynamic application security testing) software. It is possible to identify software vulnerabilities while it is operating by using a DAST tool.

- Provide secure coding techniques training to developers.

The most prevalent CWE categories for vulnerabilities should be known to developers, as well as how to prevent them. By taking these steps, businesses can help lower the risks connected to the most common CWE categories of vulnerabilities. The following frequently occurring CWE categories contain exploitable vulnerabilities:

Table 1 CWE Categories among other vulnerabilities.

CWE Category	**vulnerabilities**
CWE-119	Inadequate input validation can allow an attacker to insert SQL code into a database and steal data or perform other malicious tasks.
CWE-78	Cross-site scripting attacks can be leveraged by improper neutralization of special elements used in output, giving an attacker the ability to steal session cookies.
CWE-20	Unauthorized access to a system or application can be obtained by abusing improper authentication.
CWE-269	An attacker may be able to carry out harmful activities on a system by elevating privileges through the use of improper privilege management.
CWE-787	By using read to read data outside of the allocated memory buffer and breaking the limits, an attacker might be able to obtain private data.
CWE-416	Passwords and other sensitive data can be stolen through the use of hard-coded credentials.
CWE-22	An attacker may be able to obtain files outside of the intended directory by using path traversal, which gives them access to sensitive information.
CWE-502	It is possible to use the deserialization of untrusted data to run arbitrary code on a computer.
CWE-200	Information Exposure can be used to steal private information, including financial or consumer data.
CWE-400	Malicious files could be uploaded to a system by an attacker, who could then use them to steal data or run arbitrary code. Due to this vulnerability, files containing risky file types can be uploaded without restriction.

4. Conclusion

Entities have various options to lessen the impact of the growing number of vulnerabilities. Using a risk-based strategy in vulnerability management is one efficient method. This entails ranking every vulnerability according to its susceptibility to exploitation and its possible outcomes. Businesses can also use automation to handle critical vulnerability management tasks like patching and vulnerability scanning. Another workable solution is to outsource vulnerability management duties to managed security service providers (MSSPs), which enables organizations to assign these duties entirely or in part. The most Common Weakness Enumeration (CWE) categories linked to vulnerabilities must be known by organizations. Reducing risk requires taking proactive steps to address vulnerabilities in these categories.

Organizations can protect themselves from the negative effects of security breaches and reduce the possibility of exploitation. This all-inclusive vulnerability management strategy, which incorporates automation, outsourcing, and risk-based tactics, enables organizations to successfully traverse the challenging terrain of cybersecurity.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Brakerski, Z., Gentry, C. and Vaikuntanathan, V., (Leveled) Fully Homomorphic Encryption without Bootstrapping, ACM Trans. Comput. Theory, vol. 6, no. 3, pp. 1-36, 2014.

[2] Ben Neria, M., Yacovzada, N.-S. and Ben-Gal, I., A Risk-Scoring Feedback Model for Webpages and Web Users Based on Browsing Behavior, ACM Trans. Intell. Syst. Technol., vol. 8, no. 4, pp. 1-21, 2017.

[3] Li, S., Da Xu, L., Zhao, S., The internet of things: A survey, Inf. Syst. Front, vol. 17, p. 243–259, 2015.

[4] Sun, N., Zhang, J., Rimba, P., Gao, S., Zhang, L.Y., Xiang, Y., Data-driven cybersecurity incident prediction, A survey. IEEE Commun. Surv. Tutor, vol. 21, p. 1744–1772, 2018.

[5] McIntosh, T., Jang-Jaccard, J., Watters, P., Susnjak, T, The inadequacy of entropy-based ransomware detection, Springer, pp. 181-189, 12–15 December 2019.

[6] Alazab, M., Venkatraman, S., Watters, P., Alazab, M., Zero-day malware detection based on supervised learning algorithms of API call signatures, In Proceedings of the Ninth Australasian Data Mining Conference (AusDM'11), Ballarat, Australia, 1–2 December 2011.

[7] Shaw, A, Data breach: From notification to prevention using PCI DSS, Colum. JL Soc. Probs., vol. 517, no. 43, 2009.

[8] Gupta, B.B., Tewari, A., Jain, A.K., Agrawal, D.P., Fighting against phishing attacks: State of the art and future challenges, Neural Comput. Appl., vol. 28, p. 3629, 2017.

[9] Geer, D., Jardine, E., Leverett, E., On market concentration and cybersecurity risk., J. Cyber Policy, vol. 5, pp. 9-29, 2020.

[10] Buecker, A., Borrett, M., Lorenz, C., Powers, C., Introducing the IBM Security Framework and IBM Security Blueprint to Realize Business-Driven Security, International Technical Support Organization: Riyadh, 2010.

[11] Fischer, E.A, Cybersecurity Issues and Challenges: In Brief; Library of Congress, 2014.

[12] Chernenko, E., Demidov, O., Lukyanov, F., Increasing International Cooperation in Cybersecurity and Adapting Cyber Norms, Council on Foreign Relations.

[13] Papastergiou, S., Mouratidis, H., Kalogeraki, E.M., Cyber security incident handling, warning and response system for the european critical information infrastructures (cybersane)., In Proceedings of the International Conference on Engineering Applications of Neural Networks, Crete, Greece, 24–26 May 2019; Springer: Berlin/Heidelberg, p. 476–487, 2019.

[14] Nicholas Jeffrey, Qing Tan and José R. Villar, A Review of Anomaly Detection Strategies to Detect Threats to Cyber-Physical Systems, Electronics, vol. 12, p. 3283, July 2023.

[15] Jiang, M., Zhang, Q. and Kong, J., Multiformer-based hybrid learning with outlier re-assignment for unsupervised person re-identification, International Journal of Machine Learning and Cybernetics, pp. 1-18, 2023.

[16] Li, S.; Da Xu, L.; Zhao, S., The internet of things: A survey., Inf. Syst. Front., vol. 17, p. 243–259, 2015.

[17] Sun, N.; Zhang, J.; Rimba, P.; Gao, S.; Zhang, L.Y.; Xiang, Y., Data-driven cybersecurity incident prediction: A survey., IEEE Commun. Surv. Tutor, vol. 21, p. 1744–1772, 2018.

[18] McIntosh, T.; Jang-Jaccard, J.; Watters, P.; Susnjak, T, The inadequacy of entropy-based ransomware detection, in In Proceedings of the International Conference on Neural Information Processing, , Sydney, Australia, 2019.

[19] Alazab, M.; Venkatraman, S.; Watters, P.; Alazab, M., Zero-day malware detection based on supervised learning algorithms of API call signatures, in In Proceedings of the Ninth Australasian Data Mining Conference (AusDM'11), Ballarat, Australia, 2019.

[20] Shaw, A., Data breach: From notification to prevention using PCI DSS. Colum, JL Soc. Probs., vol. 43, p. 517, 2009.

[21] Gupta, B.B.; Tewari, A.; Jain, A.K.; Agrawal, D.P., Fighting against phishing attacks: State of the art and future challenges, Neural Comput. Appl., vol. 28, p. 3629–3654, 2017.

[22] Geer, D.; Jardine, E.; Leverett, E., On market concentration and cybersecurity risk, J. Cyber Policy, vol. 5, p. 9–29, 2020.

[23] Buecker, A.; Borrett, M.; Lorenz, C.; Powers, C, Introducing the IBM Security Framework, and IBM Security Blueprint to Realize Business-Driven Security, International Technical Support Organization, 2010.

[24] Fischer, E.A, Cybersecurity Issues and Challenges: In Brief, Library of Congress, 2014.

[25] Chernenko, E., Demidov, O., Lukyanov, F, Increasing International Cooperation in Cybersecurity and Adapting Cyber Norms, Council on Foreign Relations, 2018.

[26] Papastergiou, S.; Mouratidis, H.; Kalogeraki, E.M., Cyber security incident handling, warning and response system for the european critical information infrastructures (cybersane), in In Proceedings of the International Conference on Engineering Applications of Neural Networks, Berlin/Heidelberg, Germany, 2019.

[27] O'Connell, M.E., Cyber security without cyber war, J. Confl. Secur. Law, vol. 17, p. 187–209, 2012.

[28] Tolle, K.M.; Tansley, D.S.W.; Hey, A.J, The fourth paradigm: Data-intensive scientific discovery, Proc. IEEE, 2011.

[29] Cost of Cyber Attacks vs. Cost of Cybersecurity in 2021|Sumo Logic, 2021. [Online]. Available: www.sumologic.com/blog/costof-cyber-attacks-vs-cost-of-cyber-security-in-2021/.

[30] Anwar, S.; Mohamad Zain, J.; Zolkipli, M.F.; Inayat, Z.; Khan, S.; Anthony, B.; Chang, V., From intrusion detection to an intrusion response system: fundamentals, requirements, and future directions, Algorithms, vol. 10, no. 39, 2017.

[31] Saxe, J.; Sanders, H, Attack Detection and Attribution, Malware Data Science, 2018.

[32] Randall Smith, Multi-Community Cyber Defense (MCCD), 2002.

[33] Robert E. Mahan; Jerry D. Fluckiger; Samuel L. Clements; Cody W. Tews; John R. Burnette; Craig A. Goranson; Harold Kirkham, Secure Data Transfer Guidance for Industrial Control and SCADA Systems, 2011.

[34] William J Buchanan, Introduction to Security and Network Forensics, 2011.

[35] Christian M. Lesjak; Daniel M. Hein; Johannes Winter, Hardware-security Technologies for Industrial IoT: TrustZone and Security Controller, in IECON 2015 - 41ST ANNUAL CONFERENCE OF THE IEEE INDUSTRIAL, 2015.

[36] Prageeth Gunathilaka; Daisuke Mashima; Binbin Chen, SoftGrid: A Software-based Smart Grid Testbed for Evaluating Substation Cybersecurity Solutions, 2016.

[37] Teemu Väisänen, Categorization of Cyber Security Deception Events for Measuring The Severity Level of Advanced Targeted Breaches, in PROCEEDINGS OF THE 11TH EUROPEAN CONFERENCE ON SOFTWARE, 2017.

[38] Hao Yi Ong; Mykel J. Kochenderfer, Markov Decision Process-Based Distributed Conflict Resolution for Drone Air Traffic Management, JOURNAL OF GUIDANCE CONTROL AND DYNAMICS, 2017.

[39] Jeffrey Wermann; Armando Walter Colombo; Agnes Pechmann; Maximilian Zarte, Using an Interdisciplinary Demonstration Platform for Teaching Industry 4.0, PROCEDIA MANUFACTURING, 2019.

[40] Peter Sullivan, Network anomaly detection: The essential antimalware tool, 2023.

[41] Katie Terrell Hanna, Brute- force attack.

[42] Erin Sullivan, Data Loss, 2023.

[43] Peter Loshin, user behavior analytics (UBA), Tech Target.

[44] Quadrant, M., Magic quadrant for security information and event management, Magic Quadrant, 2016.

[45] Forrester Research, The Forrester Wave™: Security Threat Intelligence Services, 2022.

[46] IDC, Demand Driven by the Need for Security Automation to Address Growing Complexity and Scale of Threats, Worldwide Security Orchestration, Automation, and Response (SOAR) Software Market Shares, 2022.

[47] Moncada-Torres, A., van Maaren, M.C., Hendriks, M.P., Siesling, S. and Geleijnse, G, Explainable machine learning can outperform Cox regression predictions and provide insights in breast cancer survival, Scientific reports, vol. 11, no. 1, p. 6968, 2021.

[48] Najafabadi, M.M., Villanustre, F., Khoshgoftaar, T.M., Seliya, N., Wald, R. and Muharemagic, E., Deep learning applications and challenges in big data analytics, Journal of big data, vol. 2, no. 1, pp. 1-21, 2015.

[49] Pandey, S., Srivastava, A.K. and Amidan, B.G., A real time event detection, classification and localization using synchrophasor data., IEEE Transactions on Power Systems, vol. 35, no. 6, pp. 4421-4431, 2020.

[50] Meng, Z., Xu, H., Chen, M., Xu, Y., Zhao, Y. and Qiao, C., Learning-driven decentralized machine learning in resource-constrained wireless edge computing., in In IEEE INFOCOM 2021-IEEE Conference on Computer Communications, 2021.