Prediction Of Cyber Security Attacks

1st Abdullah Khan
Computer Science and Engineering
Chandigarh University
Mohali, India
21BCS10510@cuchd.in

2nd Loga Aswin
Computer Science and Engineering
Chandigarh University
Mohali, India
21BCS10573@cuchd.in

3rd Sarita Simaiya
Computer Science and Engineering
Chandigarh University
Mohali, India
sarita.e14422@cumail.in

***Abstract*—This paper provides a survey of prediction. Because cyber threats and attacks are always changing, advanced predictive strategies in cyber security are required. This paper investigates state-of-the-art models and techniques for cyber security incident forecasting. It outlines four main goals: network security situation forecasting, which provides a comprehensive view of the cybersecurity landscape; intrusion prediction, which focuses on anticipating impending cyber threats; attack projection, which emphasizes the importance of predicting attackers' next moves and intentions; and assessing the viability of these predictive methods. This survey offers a comparative analysis of the suitability of discrete models—such as attack graphs, Bayesian networks, and Markov models—and continuous models, such as time series and grey models, for the dynamic cyber security domain. It also explores the development of data mining and machine learning methods,which have a lot of potential for changing with the cyber threat landscape. This survey highlights the application of these predictive methodologies in practice and discusses the difficulties that come with evaluating them.**

***Index Terms*—intrusion detection, cyber security, situational awareness, forecasting, prediction, and model checking.**

I. Introduction

In an era defined by unprecedented technological advancements. the digital realm has become a critical infrastructure, an ecosystem where communication, commerce, and critical operations seamlessly converge. However, this pervasive connectivity has also given rise to a parallel world of cyber threats, where nefarious actors constantly seek to exploit vulnerabilities and disrupt the digital landscape. In response to this ever-present danger, the realm of cyber security has evolved, shifting from a reactive stance to a proactive one. One of the cornerstones of this proactive approach is the prediction of cyber security attacks.

Predicting cyber security threats and attacks has emerged as an essential endeavor in safeguarding our increasingly interconnected world. Traditional security measures, while indispensable, often come into play after an attack has already occurred. The imperative to anticipate and mitigate cyber threats before they materialize is at the forefront of modern cyber security strategies.This proactive stance necessitates the development and application of advanced techniques and models that leverage data analysis, machine learning, and artificial intelligence to identify patterns, anomalies, and indicators that may signify an impending attack.

The scope of predictive cyber security encompasses a diverse range of tasks, each geared toward mitigating a different facet of the evolving threat landscape. These include predicting an attacker's next move or intentions, forecasting imminent cyber attacks, and projecting the overall cybersecurity situation across a network. To address these multifaceted challenges, various methodologies have been devised, encompassing both discrete and continuous models. Discrete models, such as attack graphs, Bayesian networks, and Markov models, provide a structured approach to understanding potential threats. Continuous models, including time series and grey models, cater to the dynamic nature of cyber threats.

In recent years, machine learning and data mining approaches have gained prominence in the field of predictive cyber security. These techniques offer the potential to adapt to the rapidly changing threat environment and hold promise in augmenting the efficacy of predictive measures. However, the practical usability of these predictive methods and the complexities surrounding their evaluation are areas of ongoing research and exploration.

This research paper endeavors to delve into the intricate world of predictive cyber security, providing a comprehensive survey of the various methodologies employed in predicting and forecasting cyber security attacks. We will explore the theoretical foundations of these methodologies, evaluate their practical applicability, and address the challenges associated with their assessment. By shedding light on these critical aspects, this paper aims to contribute to the growing body of knowledge in the field of cyber security and empower cyber defenders with the tools and insights needed to anticipate and mitigate threats in a rapidly evolving digital landscape.

II. RELATED WORKS

A. Cybersecurity

The term "cybersecurity" refers to a wide range of procedures, technologies, and practices that have been created expressly to protect user and business data, networks, applications, and devices. due to assaults [3], [14]–[18]. In the meantime, corporate, financial, Generally speaking, the military, the government, and the medical collect, handle, and keep large volumes of data on computers and additional gadgets. A portion of these collected data can be sensitive information

such as financial, personal, or intellectual property; therefore, access to such information requires authorization because unauthorized parties' access could have unfavorable effects [15, 19, 20]. As a result, cybersecurity is very important to these organizations [19]–[22].

B. Cybercriminals

Criminal activity involving a computer, networked device, or network is referred to as cybercrime [20]–[22]. Cybercriminals typically carry out cybercrimes in order to profit personally [20]–[22]. On the other hand, some cybercrimes are committed with the intent to harm or disable computers. Some people also use computers or networks to spread malware and to distribute content that is forbidden, such as pictures, videos, or other files [20]–[22]. The Council of Europe Convention on Cybercrime defines cybercrime as a wide range of malicious activities that compromise the availability and integrity of a network, including illegal data interception, copyright violations, and system intrusions [25, 28]. This council has the USA as a signatory [25, 28].

C. Cyber Attacks

A cyber-attack is a deliberate, malevolent attempt by an individual or group to compromise another's information system [20]. Although most attacks are motivated by money, they can also involve the theft, alteration, or destruction of data or information. Stated differently, among the objectives of an attack are system disruption and the theft, alteration, or destruction of another party's data or information [3, 15, 20, 30]. These days, cyberattacks are becoming more frequent. Additionally, it is noteworthy to mention that the introduction of network-based ransomware worms has made it possible for attackers to initiate campaigns without requiring human intervention, as reported in the Cisco Annual Cybersecurity Report [21]. Furthermore, security events have increased in complexity and quantity [21].

Malware, phishing, denial of service (DoS), man-in-the-middle (MitM), password spraying, and cross-site scripting (XSS) are the six main types of cyberattacks [20]–[29]. The descriptions of each of these attack types are as follows:

- Trojans: Also referred to as Trojan horses, these malicious programs comprise an apparently trustworthy and secure program, file, or segment of code (but in reality a virus) [23, 28]. Trojans are typically packaged and They are designed to spy on or steal data from victims and are carried inside genuine software. Trojans pose as authentic files, therefore It would deceive victims into clicking, opening, or installing these Trojans (unaware of). After installation, a lot of To spy on the user, Trojans will download additional malware. victim or inflict different kinds of damage.
- Viruses: Typically, viruses cling to the initialization order, and these viruses would They multiply in order to contaminate additional codes inside the system of computers [20, 23, 28]. They might also affix them to executable code or establish a connection with a file by creating a virus file with the same name. however with a postponement [25, 29]. This document is a ruse that the virus is transported [20, 24, 28].
- Worms: Self-contained programs that spread across networks and computers are referred to as worms [21, 24, 29]. Worms would spread a copy of themselves to every contact on the compromised computer email list, often through email attachments [53]. Typically, attackers use worms to flood an email server with traffic and launch a denial-of-service attack. But unlike viruses, worms don't attack the host [20, 28].

1) Phishing: The act of sending fraudulent emails that appear legitimate is known as phishing [25, 27]. These emails frequently appear to be legitimate, but in reality, they direct the recipient to a malicious script or file [25]. The attackers could access and take control of the victim's device by using this script or file. As a result, the attacker could potentially insert malicious scripts or files and retrieve private information such as user credentials and bank account information [25, 26]. Phishing is essentially used to obtain sensitive information, including the victim's credit card number and login credentials.

2) Distributed denial-of-service (DDoS) and denial of service (DoS): Attacks known as denial of service (DoS) involve saturating servers, networks, and/or systems with traffic in an effort to overload bandwidth and resources, which prevents the system from responding to legitimate requests [28]. DDoS attacks, also known as Distributed Denial-of-Service (DDoS) attacks, can be carried out concurrently by multiple computers at a single moment [24, 28]. Because DDoS attackers can originate from any IP address in the world, it can be extremely difficult for network administrators to identify the source of the attack [20, 26].

D. Bayesian Networks

Using Bayesian networks is another class of model-checking attack prediction techniques. Since a Bayesian network is usually built from an attack graph, these techniques are closely related to model-checking methods based on attack graphs. The conditional variables and probabilities that are represented in the model are what set Bayesian networks apart. In certain instances, Bayesian networks are subject to additional limitations. For instance, employing causal networks rather than standard Bayesian networks results from the requirement on the causality of events.

A probabilistic graphical model that depicts the variables and their connections is called a Bayesian network. With nodes representing discrete or continuous random variables and edges representing the relationships between them, the network is a directed acyclic graph. The conditional probability form and random variable states are preserved by the nodes.

III. METHODOLOGY

This project is structured to systematically address the challenges of predicting and forecasting cyber security attacks. It begins with a clear problem definition, outlining the specific

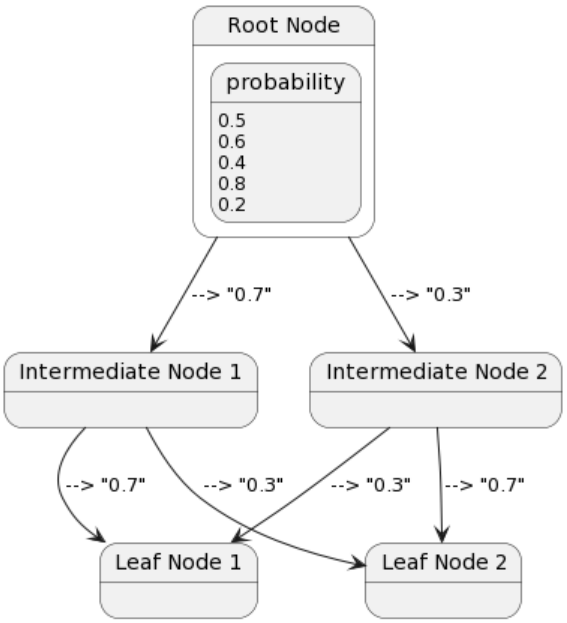

Fig. 1. Simple Bayesian Attack Graph illustrating probability computations.

objectives related to threat prediction and security enhancement. Data collection is a pivotal step, where a diverse range of data sources, including network logs, system events, user behavior, and threat intelligence feeds, are aggregated. Subsequently, data preprocessing is carried out to clean, normalize, and extract relevant features from the collected data.

Exploratory Data Analysis (EDA) offers insights into data patterns and distribution of threat indicators, guiding the feature engineering process. Feature engineering involves creating meaningful attributes to empower predictive models. Model selection is a critical decision, encompassing a range of approaches, from machine learning algorithms to statistical models and time series analysis.

Model training follows, with a focus on achieving optimal model performance. Rigorous evaluation using appropriate metrics validates the effectiveness of the models. Hyperparameter tuning may be applied to enhance model performance. Model validation ensures that the models generalize well on unseen data, utilizing techniques like cross-validation and time-based validation.

With trained models in place, the project advances to the prediction and forecasting stage. Models are deployed to process incoming data in real-time, identifying potential threats and vulnerabilities. The final phase involves implementing alerting mechanisms and reporting structures to inform stakeholders of emerging cyber threats, contributing to a proactive and fortified cyber security framework.

IV. Objective

The primary objective of this research paper is to advance the state of knowledge in the realm of cyber security by comprehensively investigating the applications, methods, and implications of predicting and forecasting cyber security

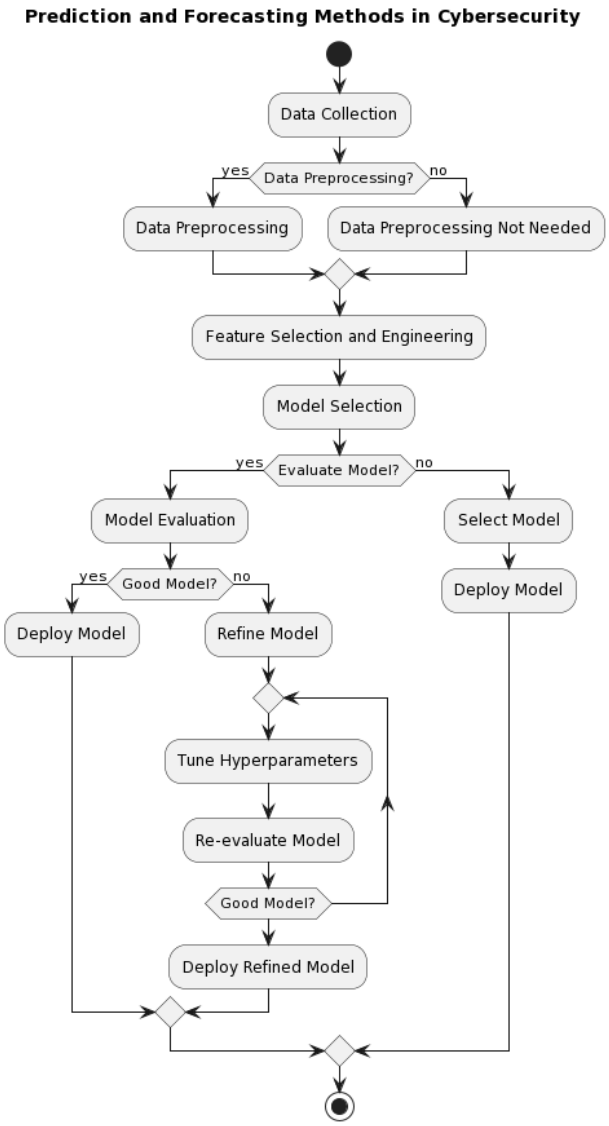

Fig. 2. Prediction and Forecasting Methods in Cyber Security.

attacks. This research endeavors to address several pivotal objectives:

A. Survey and Synthesize Current Methods

The paper seeks to provide an exhaustive survey of existing methods and models utilized for prediction and forecasting in the field of cyber security. By synthesizing contemporary research, it aims to consolidate and present a holistic understanding of the diverse techniques employed to anticipate cyber threats.

B. Characterize Use Cases

A key objective is to categorize and delineate distinct use cases within the domain of prediction and forecasting in cyber security. These use cases include attack projection, intention recognition, intrusion prediction, and network security situation forecasting, each with unique characteristics and

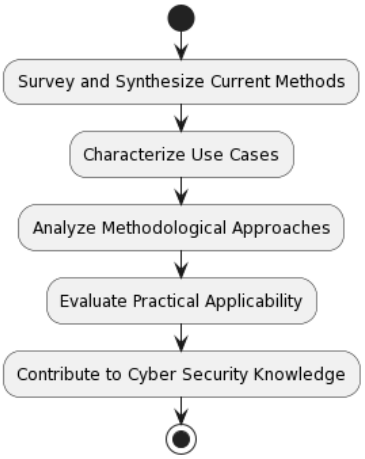

Fig. 3. Objective of Cyber Security Prediction.

requirements. The paper aims to provide a clear and structured taxonomy of these use cases.

C. Analyze Methodological Approaches

This research aspires to scrutinize the methodological approaches underpinning the aforementioned use cases. It delves into both discrete and continuous models, such as attack graphs and time series, and assesses their suitability for prediction and forecasting. The analysis encompasses the integration of machine learning and data mining techniques, which have recently gained prominence in the dynamic cyber security landscape.

D. Evaluate Practical Applicability

The research paper intends to emphasize the practical usability of the methods and models discussed. It aims to shed light on the real-world applications of predictive cyber security and the potential benefits in mitigating threats and vulnerabilities. Furthermore, it aims to examine the challenges related to evaluating the effectiveness of these methods in real-time scenarios.

E. Contribute to Cyber Security Knowledge

Ultimately, this research paper aims to contribute significantly to the body of knowledge in cyber security by providing valuable insights into the strategies, tools, and methodologies for anticipating and countering cyber threats. It seeks to empower cyber defenders, organizations, and security professionals with the tools and insights needed to enhance cyber resilience and proactively secure digital environments.

V. Constraint Identification

The pursuit of predicting and forecasting cyber security attacks is a multifaceted endeavor, replete with various constraints and challenges that necessitate careful consideration throughout the research process.

A critical constraint lies in the quality and availability of data. Accurate predictive models heavily rely on vast and high-quality datasets. The scarcity of real-world, labeled data, coupled with privacy concerns, can impede the training and validation of robust models.

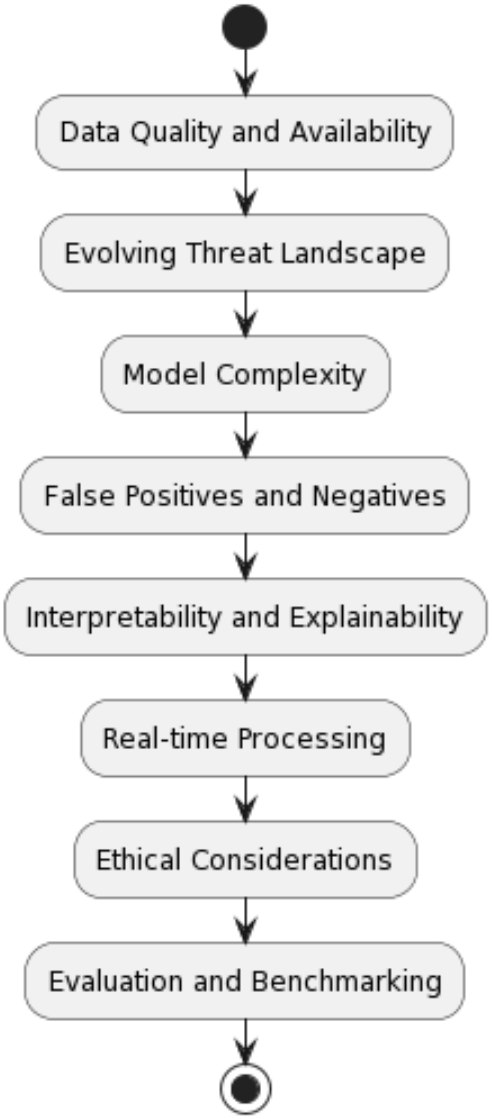

Fig. 4. Constraint identification in Cyber Security attack prediction.

The dynamic nature of cyber threats poses a significant challenge. Attack methodologies constantly evolve, rendering historical data less relevant. Models designed to predict future threats must grapple with this ever-changing landscape and adapt to emerging attack vectors.

Developing and deploying predictive models in real-time cyber security environments can be intricate and resource-intensive. Complex models may be computationally expensive, affecting their real-time feasibility and practicality.

Striking a balance between detecting true positives (actual threats) and minimizing false positives (false alarms) is a continuous challenge. Overly sensitive models may inundate security professionals with false alerts, while overly specific

models may miss actual threats. The black-box nature of certain machine learning models can hinder their acceptance and adoption in security operations. Ensuring that the models' decisions are explainable to security practitioners is an ongoing challenge.

Timeliness is a critical factor in cyber security. Predictive models must process and respond to data in near real-time, which introduces constraints related to processing speed, latency, and scalability.The ethical implications of predictive cyber security models, particularly in the context of user privacy and potential bias, are constraints that demand attention. Striking a balance between security and individual rights is a complex issue.

Evaluation and Benchmarking is crucial to establish benchmarks and standardized evaluation metrics for predictive models. It is still difficult to compare and evaluate the efficacy of various strategies, especially when the threat landscape changes.

VI. ACKNOWLEDGEMENT

We would like to express our sincere gratitude to our supervisor, Ms. Sarita Simaiya, for his invaluable guidance and support throughout this project. His expertise, insights, and advice were instrumental in shaping our research and ensuring its success. We are also grateful to Chandigarh University for providing us with the resources and facilities necessary for this research. Lastly, we would like to thank all the participants who took part in our study and contributed to its success."

VII. CONCLUSION

We reviewed the literature on attack prediction techniques in this paper. The issue was framed within the framework of studies on cyber situational awareness and intrusion detection. There was a method taxonomy given, and each category was thoroughly explained and assessed. The ultimate assessment compared the approaches, talked about related issues, and knowledge gained. Here, we wrap up our research on the attack prediction theory and practice, as well as future suggestions happenings in the area.

The literature review yielded three significant findings. Initially, a lot of cyber security prediction techniques use models to depict and forecast the future state of an attack or a security scenario. The two main use cases frequently complement each other and overlap, despite the apparent division of the models given by their use case (attack projection more often uses discrete models, while forecasting network security situation uses continuous models predominantly). Second, a lot of new methods based on data mining and machine learning have emerged, significantly altering the state of cyber security prediction research. The reliance on artificially generated prediction models is eliminated by data mining, but model-based techniques in general are challenged by machine learning. Ultimately, we have run into a lot of issues with the assessment of cyber security predictions. Popular datasets in the context of empirical datasets are outdated, untrustworthy, and produced for different objectives; assessments in real-world networks, on the other hand, are not repeatable. Not even a common set of metrics is available for us to compare the approaches with.

Future developments in attack prediction and its practical application are probably in store. Considering that intrusion detection is one step ahead of attack prediction, we list several potential avenues for future research. First, there has already been a shift in the processing of network data and alerts from batch to stream data, and more use of big data analytics is likely [19], [10]. Secondly, research on attack prediction in a collaborative environment, like alert sharing platforms or collaborative intrusion detection systems, will be conducted in the near future. The logical next step in this field of study is to predict attacks in such an environment [20], [16]. Ultimately, data mining and machine learning will play an increasingly larger role in cyber security [22], and attack prediction is no different. In particular, we will gain a better understanding of whether machine learning in isolation can be utilized to both learn about and predict attacks, or whether data mining and machine learning will be limited to learning about the attacks, with pattern matching still being used for prediction. To sum up this paper, attack prediction is a fascinating research problem that has been tackled by several researchers on several occasions. The question of how to accurately and successfully predict cyberattacks remains unanswered, despite the fact that numerous solutions have been put forth. Although attack prediction is not yet widely used and is occasionally viewed as being somewhat misleading [21], it is still an open research problem that is desirable and vital [1], [3], and [20].

REFERENCES

[1] A. Kott, Towards Fundamental Science of Cyber Security. New York,NY: Springer New York, 2014, pp. 1–13.

[2] R. A. Ahmadian and A. R. Ebrahimi, "A survey of it early warning systems: architectures, challenges, and solutions," Security and Communication Networks, vol. 9, no. 17, pp. 4751–4776.

[3] I. A. Gheyas and A. E. Abdallah, "Detection and prediction of insider threats to cyber security: a systematic literature review and metaanalysis," Big Data Analytics, vol. 1, no. 1, p. 6, Aug 2016.

[4] T. Hughes and O. Sheyner, "Attack scenario graphs for computer network threat analysis and prediction," Complexity, vol. 9, no. 2, pp. 15–18, 2003.

[5] J. Wu, L. Yin, and Y. Guo, "Cyber Attacks Prediction Model Based on Bayesian Network," in Parallel and Distributed Systems (ICPADS), 2012 IEEE 18th International Conference on, Dec 2012, pp. 730–731.

[6] A. Okutan, S. J. Yang, and K. McConky, "Predicting Cyber Attacks with Bayesian Networks Using Unconventional Signals," in Proceedings of the 12th Annual Conference on Cyber and Information Security Research, ser. CISRC '17. ACM, 2017, pp. 13:1–13:4

[7] G. Werner, S. Yang, and K. McConky, "Time series forecasting of cyber attack intensity," in Proceedings of the 12th Annual Conference on Cyber and Information Security Research, ser. CISRC '17. New York, NY, USA: ACM, 2017, pp. 18:1–18:3.

[8] Chadza T, Kyriakopoulos KG, Lambotharan S. (2019). Contemporary Sequential Network Attacks Prediction using Hidden Markov Model. In 2019 17th International Conference on Privacy, Security and Trust (PST). Fredericton: IEEE (pp. 1-3).

[9] Ibrahimi K, Ouaddane M (2017) Management of intrusion detection systems based-KDD99: analysis with LDA and PCA. In 2017 international conference on wireless networks and Mobile communications (WINCOM). Rabat, IEEE, pp 1–6.

[10] Sharafaldin I, Gharib A, Lashkari AH, Ghorbani AA (2018a) Towards a reliable intrusion detection benchmark dataset. Softw Netw 2018(1):177–200

[11] S. Tan, P. Xie, J. M. Guerrero, J. C. Vasquez, Y. Li, and X. Guo, "Attack detection design for dc microgrid using eigenvalue assignment approach," Energy Reports, vol. 7, pp. 469–476, 2021
[12] Y. Li and Q. Liu, "A comprehensive review study of cyber-attacks and cyber security; emerging trends and recent developments," Energy Reports, vol. 7, pp. 8176–8186, 2021.
[13] M. Abomhara and G. M. Køien, "Cyber security and the internet of things: vulnerabilities, threats, intruders and attacks," Journal of Cyber Security and Mobility, pp. 65–88, 2015.
[14] A. Goel, D. K. Sharma, and K. D. Gupta, "Leobat: Lightweight encryption and otp based authentication technique for securing iot networks," Expert Systems, vol. 39, no. 5, p. e12788, 2022.
[15] B. B. Gupta, A. Tewari, A. K. Jain, and D. P. Agrawal, "Fighting against phishing attacks: state of the art and future challenges," Neural Computing and Applications, vol. 28, no. 12, pp. 3629–3654, 2017.
[16] K. M. Prasad, A. R. M. Reddy, and K. V. Rao, "Dos and ddos attacks: defense, detection and traceback mechanisms-a survey," Global Journal of Computer Science and Technology, 2014.
[17] M. Souppaya, K. Scarfone et al., "Guide to malware incident prevention and handling for desktops and laptops," NIST Special Publication, vol. 800, p. 83, 2013.
[18] G. Kim, S. Lee and S. Kim, "A novel hybrid intrusion detection method integrating anomaly detection with misuse detection", Expert Systems with Applications, vol. 41, no. 4, (2014), pp. 1690-1700.
[19] P. S. Kenkre, A. Pai and L. Colaco, "Real time intrusion detection and prevention system", In Proceedings of the 3rd International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA), Springer, Cham, (2014), pp. 405-411.
[20] T. Xing, D. Huang, L. Xu, C. J. Chung and P. Khatkar, "Snortflow: A openflow-based intrusion prevention system in cloud environment", In Research and Educational Experiment Workshop (GREE), Second GENI, IEEE, (2013), pp. 89-92.
[21] U. Ravale, N. Marathe and P. Padiya, "Feature selection based hybrid anomaly intrusion detection system using K means and RBF kernel function", Procedia Computer Science, vol. 45, (2015), pp. 428- 435.
[22] W. Xing-zhu, "Network Intrusion Prediction Model based on RBF Features Classification", International Journal of Security and Its Applications, vol. 10, no. 4, (2016), pp. 241-248.
[23] A. A. Ramaki, M. Amini and R. E. Atani, "RTECA: Real time episode correlation algorithm for multistep attack scenarios detection", Computers and Security, vol. 49, (2015), pp. 206-219.
[24] S. Shamshirband, A. Patel, N. B. Anuar, M. L. M. Kiah and A. Abraham, "Cooperative game theoretic approach using fuzzy Q-learning for detecting and preventing intrusions in wireless sensor networks", Engineering Applications of Artificial Intelligence, vol. 32, (2014), pp. 228-241.
[25] S. Chen, Z. Zuo, Z. P. Huang and X. J. Guo, "A graphical feature generation approach for intrusion detection", In MATEC Web of Conferences, EDP Sciences, vol. 44, (2016).
[26] A. Kott, A. Swami, and P. McDaniel, "Security Outlook: Six Cyber Game Changers for the Next 15 Years," Computer, vol. 47, no. 12, pp. 104–106, Dec 2014.
[27] E. Vasilomanolakis, S. Karuppayah, M. Muhlh ¨ auser, and M. Fischer, ¨ "Taxonomy and survey of collaborative intrusion detection," ACM Comput. Surv., vol. 47, no. 4, pp. 55:1–55:33, May 2015.
[28] M. Albanese, E. Battista, S. Jajodia, and V. Casola, "Manipulating the attacker's view of a system's attack surface," in 2014 IEEE Conference on Communications and Network Security, Oct 2014, pp. 472–480.
[29] A. L. Buczak and E. Guven, "A survey of data mining and machine learning methods for cyber security intrusion detection," IEEE Communications Surveys Tutorials, vol. 18, no. 2, pp. 1153–1176, Secondquarter 2016.
[30] H. Wei, G.-Y. Hu, Z.-J. Zhou, P.-L. Qiao, Z.-G. Zhou, and Y.-M. Zhang, "A new BRB model for security-state assessment of cloud computing based on the impact of external and internal environments," Computers and Security, vol. 73, pp. 207 – 218, 2018.