DOI : https://doi.org/10.56452/7-2-548

ISSN: 0974-5823 Vol. 7 No. 2 February, 2022

International Journal of Mechanical Engineering

Automated Cyber Attack Prediction Workflow

Poonam Verma

Asst. Professor, School of Computing, Graphic Era Hill University, Dehradun, Uttarakhand India 248002,

Abstract:

Data transmission rates have increased due to the widespread use of technologically driven services and applications. Because networks are always vulnerable to attack, the increasing quantity of network throughput has also increased the security risks, making cybercrime more likely. To combat this, we suggest a cyber security strategy that makes use of GA and neural network structure. Features of DDoS and malware attacks are represented by characteristics, and the GA is used to pick and optimise them. In order to train and classify the retrieved data, it is fed into a neural network. Precision, recall, and f-measure analysis were used to the detection of DDoS and malware nodes to determine the efficacy of the suggested cyber security strategy. Based on the results of the comparison, it was clear that the cyber security method was superior in detecting network threats. The authors then provide a method that blends SI with methods taken from nature. The GA is used to narrow down feature candidates and shrink down data sets. After this, DWT is used in conjunction with ABC to further eliminate superfluous data points. Finally, the training and classification stages employ the ANN-SVM hybrid, with ANN handling the former and SVM handling the latter. The hybrid method's high TPR, detection rate, accuracy, and f-measure with low FPR are the result of a simulation study that included hundreds of simulation rounds. Throughput and PDR are shown by a suggested preventative strategy, which is then tested by altering the number of nodes in comparison to prior efforts.

Keywords: ANN-SVM, TPR, DDoS, detection rate, precision, DWT

Introduction

The goal of machine learning is forward prediction based on historical data. With the use of machine learning (ML), computers may acquire new skills and knowledge without being explicitly taught them. The primary goals of machine learning are (1) the creation of data-driven computer programmes and (2) the implementation of a basic machine learning algorithm in a programming language like Python. The employment of tailored algorithms is crucial to the training and forecasting process[1]. It utilises an algorithm to make predictions based on the training data and the fresh test data. The field of machine learning may be loosely broken down into three subfields. The three main types of machine learning are supervised, unsupervised, and reinforcement. In order for a supervised learning programme to acquire knowledge, human beings must first label the incoming data. Without labels, unsupervised learning cannot occur[2]. In doing so, it aided the algorithm for learning. The input data to this method must be clustered. Last but not least, Reinforcement learning improves via both positive and negative feedback from its environment[3,4]. Python is a popular tool among data scientists, who use it to explore datasets in search of patterns that may then be used to draw

conclusions and take action. Based on how they "learn" from data to produce predictions, these algorithms may be roughly divided into two categories: supervised and unsupervised learning. Predicting what category a set of data points belongs to is called classification. Classes may also be referred to as aims, labels, or groups. Estimating a mapping function from continuous input variables (X) to discrete output variables (y) is the goal of classification predictive modeling[5]. Classification is a supervised learning method in machine learning and statistics that allows a computer programme to learn from its input data and then apply that knowledge to the classification of new observations[6]. It's possible that this piece of data just contains binary information (such as whether the subject is male or female or whether the email is spam or not), or it might include multi-class information as well. Speech recognition, handwriting recognition, biometric identification, document categorization, etc. are all instances of classification difficulties. Most businesses respond defensively and reactively to cyberattacks[7]. When threats are first identified after they have already caused damage, it is too late to take any preventative measures. Most businesses use standard technologies and practises for intrusion detection and prevention, such as anti-virus and firewall software, and access restrictions like passwords[9]. However, given the growing sophistication and complexity of cybercrime in recent years and the sheer volume of assaults that seldom make headlines, it's reasonable to argue that reactive actions are, at best, damage control tactics that rarely succeed. However, the outlook for cyber security isn't quite as gloomy as we've made it sound[10]. The proliferation of AI, ML, and quantum encryption has led to the creation of several novel strategies for countering cyber-attacks. Industry and government have a lot to gain from investing in cyber security. Based on this knowledge, several of our brightest brains are working to solve this issue, with encouraging results thus far. In order to foretell the future, analysts use historical data and state-of-the-art technology to probe into complex intermediary processes that have yet to be fully uncovered. An early warning system's primary focus in the network offence and defence duty is on providing precise DoS attack forecasts. Abnormity-based detection is useful for identifying Denial of Service attacks. DoS attacks have been the subject of a number of research from a variety of perspectives[11]. These techniques, however, need previous information and had trouble telling the difference between regular burst traffic and the ebb and flow of DoS assaults. Furthermore, they needed a lot of historical facts and were unable to forecast such assaults effectively. This paper proposes a prediction model of the discrete probability distribution of a denial-of-service (DOS) attack using data from flux inspecting and intrusion detection, and then uses the genetic algorithm to implement the optimisation of clustering methods. Using optimised clustering on the sample data, we classify the relationship between traffics and attack quantities into a number of categories and subsequently construct numerous prediction sub-models for DoS attacks. Also, the Bayesian approach is used to derive discrete probability calculations for each sub-model, and then the DoS attack prediction model is obtained from this distribution. Starting with the observation that DoS attacks are proportional to the volume of network traffic data, this study suggests using a genetic algorithm-based clustering technique to classify such information. Based on optimised clustering, this technique first obtains the appropriate division of the relationship between network traffic and the quantity of DoS attack, and then constructs the DoS attack prediction sub-models. Meanwhile, the Bayesian approach derives the distribution of the quantity of DoS attack in the future from the computation of the output probability corresponding to each sub-model. Social networking sites, blogs, opinions, ratings, reviews, serial bookmarking, social news, media sharing, and even Wikipedia have all contributed to the rapid dissemination of knowledge with the advent of mobile internet. Thorough study of these patterns may reveal a wealth of previously unknown information, including whether or whether the individual in question is engaging in malicious or innocuous conversations with a certain user, which might then serve as a basis for sociotechnical assaults. [12-15].

DOI : https://doi.org/10.56452/7-2-548

NETWORK ARCHITECTURE

In order to monitor hostile actions in cyber security against a wide variety of assaults including DDoS, malware, and spoofing, the above-mentioned graphic depicts the architecture of the suggested model. For network security, the proposed design aids in the detection of threats or assaults that might undermine the network's integrity, privacy, or availability. Many users' devices communicate with the data management system, and the server uses a pre-existing database to interpret the data it receives from those devices in order to monitor their private lives. The second aspect is the so-called "outside firewall," which takes into account online activity via the cloud. Both types of networks rely on the IDS's monitoring to detect malicious activity according to their respective train structures, with the use of AI and ML concepts and techniques. [16]

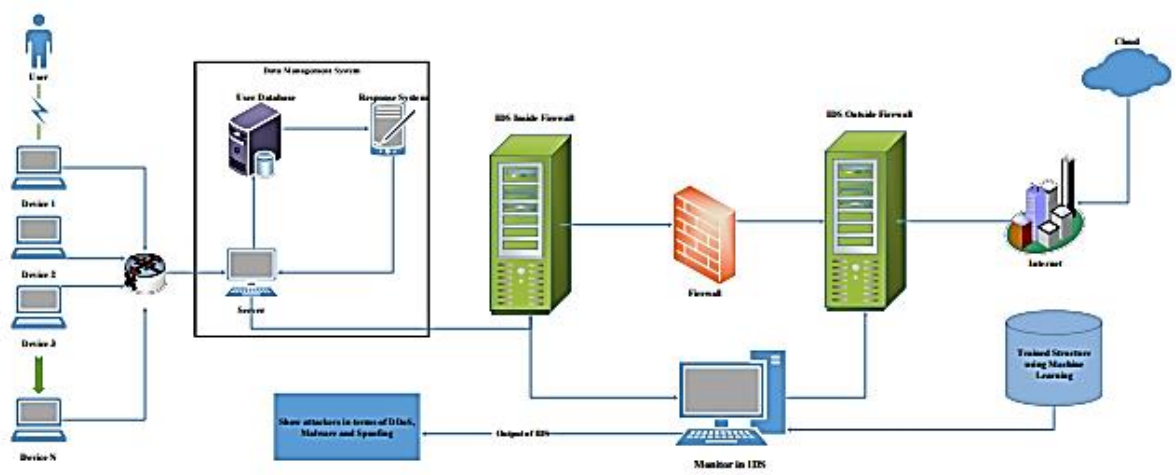

Figure 1 Network Architecture

Types of Cyber-Attacks

A cyber-attack is any kind of hostile action that uses several sophisticated ways to steal or manipulate sensitive information from computer information systems, infrastructures, and personal computer devices. Cyber-attacks pose a threat to any wirelessly connected electronic device, including smartphones, laptops, and tablets. This contributes to a rise in data breaches, the disclosure of sensitive information, and damage to the company's brand. Therefore, it is crucial to understand the many forms of cyber-attacks and the measures you may take to safeguard your network..

DOI : https://doi.org/10.56452/7-2-548

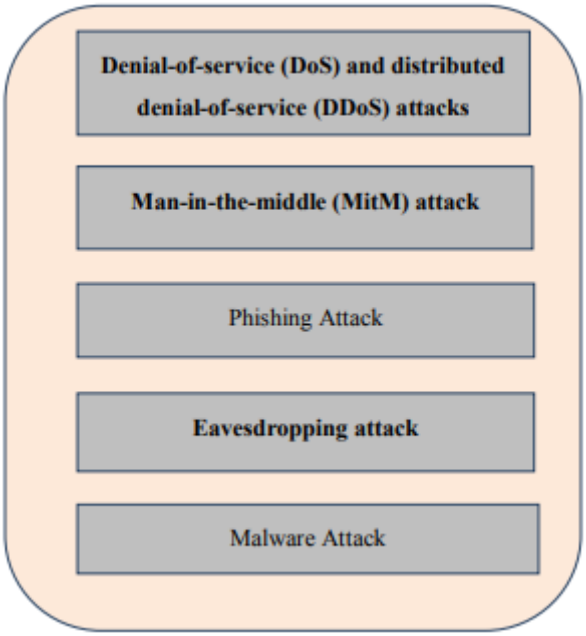

Figure 2 Classification of Cyber-attack

Source:
https://info.totalprosource.com/hubfs/6%20Common%20Types%20of%20Cyber%20Attacks%20Infographic.pdf

Literature Review

Md Anisur Rahman et.al.,(2020) The proliferation of cyberattacks is keeping pace with the rapid development of new technologies. Every company that uses private information in its operations must thus prioritise the ability to identify and anticipate cyber threats. In this research, we provide a cyber security architecture that utilises a data mining approach to anticipate cyber-attacks, which may aid in implementing effective countermeasures. Cyberattack detection and prediction are the two primary functions of the system. To anticipate future cyber-attacks, the approach initially uses a J48 decision tree algorithm to uncover patterns associated with past ones. The Canadian Institute for Cybersecurity has made its cyber security statistics accessible to the public, and we use them to test the framework. DDoS, Port Scan, Bot, Brute Force, SQL Injection, and Heartbleed are only few of the cyberattacks that are documented in the datasets. The suggested system successfully recognises cyberattacks and delivers cyberattack pattern information. The suggested prediction model has an almost 99% accuracy rate in identifying cyber threats. Future cyberattacks may be predicted using the prediction model's extracted patterns based on past data. Prediction model experimental findings show model's superiority in detecting future cyberattacks.

Martin Husák et.al.,(2019) This study offers a review of several cyber security prediction and forecasting techniques. We begin by discussing four main tasks: attack projection and intention recognition, in which we try to anticipate the next move or intentions of the attacker; intrusion prediction, in which we try to anticipate upcoming cyber-attacks; and network security situation forecasting, in which we attempt to anticipate the cyber security situation in the entire network. In many cases, the theoretical foundations shared by the various approaches to these objectives also make them mutually beneficial. Methods using continuous models, such time series and grey models, are also reviewed and contrasted with those using discrete models, including attack graphs, Bayesian

networks, and Markov models. We then go on to talk about how machine learning and data mining techniques are getting a lot of attention these days since they show promise in the ever-evolving field of cyber security. The survey also delves into the assessment issues and practical applicability of the methodologies.

Sina Pournouri et.al.,(2020) In order to effectively prepare for future cyber disasters, it might be helpful to have some idea of the kind of targets that will likely be attacked. Some forms of cyber-attack are more likely to target particular industries because of the unique characteristics of their operations. In order to be ready for future cyber breaches, many firms need to learn from their experiences and identify the dangers and vulnerabilities they face. The purpose of this research is to look at target-type prediction using O.S.I. and classification methods.

Artificial Intelligence Techniques

These days, the term "artificial intelligence" is used to describe any system that can do complex tasks as well as a person, or even better. This kind of work is all around us. Security systems at subways and airports can single out potential threats from the crowd, and overhead cameras can determine a car's speed, identify its make and model, and issue a fine. All of this is now generally understood to be AI, even if the algorithms behind individual technologies vary widely. Only a select handful make advantage of machine learning.

Machine Learning

Supervised learning and unsupervised learning are two popular methods in machine learning. Both allow the user to feed extensive datasets into the computer for analysis and connection building. A "feature vector" describes the compiled information.

Supervised

There are two forms of supervised learning: regression and classification. The essence of their production vector is the distinction between the two. If the output variable found in form of any real value, it is referred to as regression. For example height, size, weight, economic value, etc. Classification is done, when the output variable is in the form of a class or category. Choosing between “pink”, “white”, “high” or “short” falls under the category of classification. Labelling the input data into exactly two categories is a binary classification. Marking in more than two classes is a multi-class classification.

Un-supervised

This machine learning method use an unlabeled data set. Without any training, the system must identify the proper target by seeing previously unseen patterns in the data. Algorithms need to be built in a manner that allows them to discover these patterns and structures in the data on their own. Unsupervised learning includes techniques like clustering and association. Developers use clustering to identify patterns and establish hierarchy among otherwise disorganised data. Association is utilised while trying to establish connections between disparate data points in large datasets.

DOI : https://doi.org/10.56452/7-2-548

Preventive Mechanisms of Cyber Attack

Over the last decade, several technologies for detecting and evading cyber-threats have emerged. In order to choose the most appropriate mitigation mechanism, businesses with varying network infrastructures and security needs must be aware of the existence of these systems. In this article, we'll describe the tools used to identify malicious software and counteract cyberattacks. In order to breach a network of computers, attackers often exploit its weaknesses. Malicious programmes have a better chance of succeeding once they initiate an attack because they only launch assaults if certain conditions are met in the target network or system. This highlights the significance of finding and removing malware from online systems. The greatest security mechanisms either block malicious software from accessing particular networks or render them incapable of harming computer network performance. Modern computer networks are very dynamic and need interaction with a wide variety of people and cyber systems, making this objective challenging to realise. The optimal protection alternative for computer networks may be thought of as prevention mechanisms like policy, awareness, vulnerability reduction, and threat mitigation..

Machine Learning

Phishing assaults enabled by ML-based methods are particularly effective since they stand out in a blacklist. By analysing the steps involved in large-scale data processing, the ML framework is able to build its own aggregation model. Among the benefits of ML approaches are:

- It is feasible to examine a presenting model without physically analysing the data to understand complicated connections when testing is conducted on huge data collections.

- ANN is a set of operations on a large number of elements. The network is trained and evaluated based on the input data to determine its accuracy. The neural network is taught by feeding it problem-specific features drawn from both phishing and authentic data. To determine if newly posted content is legitimate or malicious, ANN compares it to its database of previously posted content.

Training and Classification

At this point, artificial neural network (ANN) training is followed by SVM-based classification. Artificial neural networks (ANNs) are a kind of supervised learning that take input data and compare it to a target sample. If an error occurs at the output, the value of the error is sent to the hidden layer, where it is used to adjust the weight matrix. Figure 3 provides a high-level overview of the ANN architecture.

DOI : https://doi.org/10.56452/7-2-548

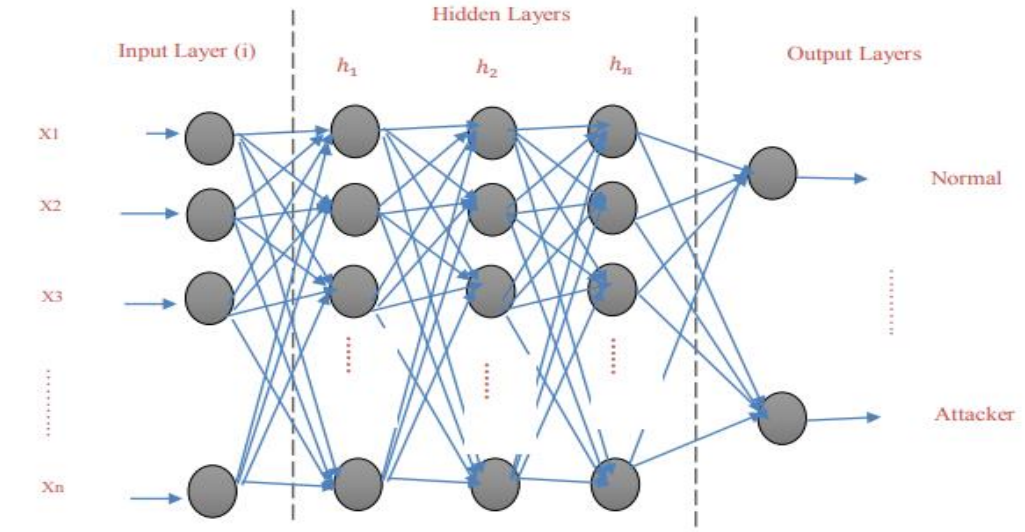

Figure 3. ANN architecture

Machine learning was the only approach that could have been conceived to reduce the computational complexity of the identification network, which is essential for protecting the system from any future risks. Both statistical methods and practical application may be used to train a machine learning system. Whereas the experience method seeks to learn from decision making, the statistical approach seeks to address the issue via quantitative representation. The suggested mechanism incorporates both methods, but does so in two distinct phases. The first part makes advantage of network simulation for DDoS and malware data classification. A label set (Ls) acpt, n-acpt n where n is the total number of identities used to train the system to recognise the network assault labels the data as "acceptable" (acpt) or "non-acceptable" (n-acpt). As can be seen in Figure 4, we first activate a ten-ten-two-layer sigmoid-propagating neural network.

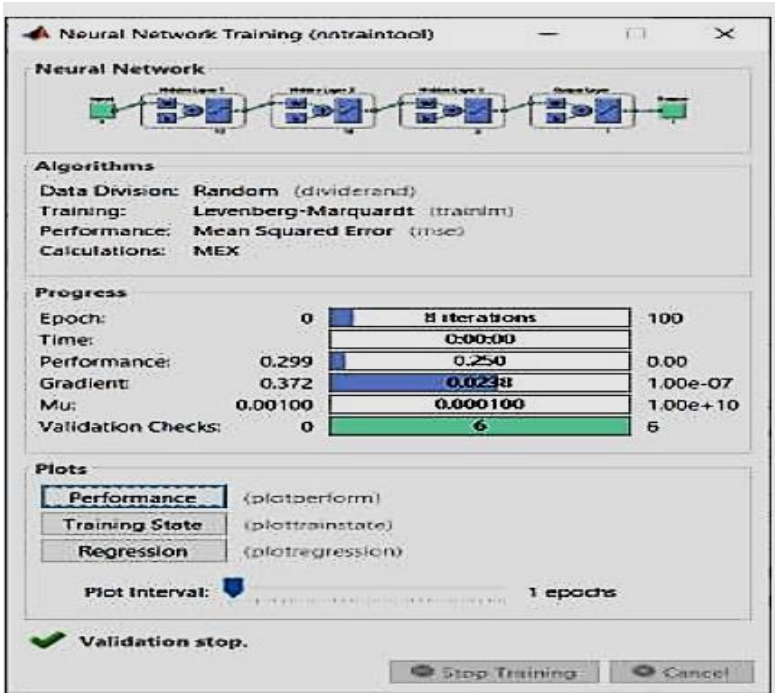

Figure 4 Multilayer propagation Mechanism

Conclusion

Data cleansing and processing came first, followed by missing value analysis, then exploratory analysis, model construction, and lastly assessment. Each algorithm will be compared with the various types of network assaults to see which will provide the most accurate prediction results in the future. The following are some of the conclusions we may draw regarding diagnosing network attacks for each new connection that results from this. To use AI to offer a prediction model that can outperform humans and expand early detection opportunities. From this model, we can deduce that combining geographic information systems analysis with machine learning techniques yields accurate predictive models that speed up the diagnostic process in the networking industry while eliminating human error.

References

1. M. A. Rahman, Y. Al-Saggaf and T. Zia, "A Data Mining Framework to Predict Cyber Attack for Cyber Security," *2020 15th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, Kristiansand, Norway, 2020, pp. 207-212, doi: 10.1109/ICIEA48937.2020.9248225.
2. M. Husák, J. Komárková, E. Bou-Harb and P. Čeleda, "Survey of Attack Projection, Prediction, and Forecasting in Cyber Security," in IEEE Communications Surveys & Tutorials, vol. 21, no. 1, pp. 640-660, Firstquarter 2019, doi: 10.1109/COMST.2018.2871866**.**
3. S. Pournouri, S. Zargari and B. Akhgar, "An Investigation of Using Classification Techniques in Prediction of Type of Targets in Cyber Attacks," *2019 IEEE 12th International Conference on Global Security, Safety and Sustainability (ICGS3)*, London, UK, 2019, pp. 202-212, doi: 10.1109/ICGS3.2019.8688266.
4. E. Mousavinejad, X. Ge, Q. -L. Han, F. Yang and L. Vlacic, "Detection of Cyber Attacks on Leader-Following Multi-Agent Systems," *IECON 2019 - 45th Annual Conference of the IEEE Industrial Electronics Society*, Lisbon, Portugal, 2019, pp. 6243-6248, doi: 10.1109/IECON.2019.8927195.
5. A. V. Uzunov, S. Nepal and M. Baruwal Chhetri, "Proactive Antifragility: A New Paradigm for Next-Generation Cyber Defence at the Edge," *2019 IEEE 5th International Conference on Collaboration and Internet Computing (CIC)*, Los Angeles, CA, USA, 2019, pp. 246-255, doi: 10.1109/CIC48465.2019.00039.
6. X. He, "Internal Threat Prediction Algorithm Based on VASG Model," *2019 IEEE 4th International Conference on Signal and Image Processing (ICSIP)*, Wuxi, China, 2019, pp. 606-610, doi: 10.1109/SIPROCESS.2019.8868483.
7. G. Ioannou, P. Louvieris and N. Clewley, "A Markov Multi-Phase Transferable Belief Model for Cyber Situational Awareness," in *IEEE Access*, vol. 7, pp. 39305-39320, 2019, doi: 10.1109/ACCESS.2019.2897923.
8. M. Kadoguchi, S. Hayashi, M. Hashimoto And A. Otsuka, "Exploring the Dark Web for Cyber Threat Intelligence using Machine Leaning," *2019 IEEE International Conference on Intelligence and Security Informatics (ISI)*, Shenzhen, China, 2019, pp. 200-202, doi: 10.1109/ISI.2019.8823360.
9. S. Yoginath *et al*., "On the Effectiveness of Recurrent Neural Networks for Live Modeling of Cyber-Physical Systems," *2019 IEEE International Conference on Industrial Internet (ICII)*, Orlando, FL, USA, 2019, pp. 309-317, doi: 10.1109/ICII.2019.00062.
10. Z. Li, L. Zhu, M. Huang, Z. Chen, S. Chen and B. Li, "Racing APUF: A Novel APUF against Machine Learning Attack with High Reliability," *2019 IEEE 4th International Conference on*

Signal and Image Processing (ICSIP), Wuxi, China, 2019, pp. 722-726, doi: 10.1109/SIPROCESS.2019.8868387.

11. N. Bezzo, "Predicting Malicious Intention in CPS under Cyber-Attack," *2018 ACM/IEEE 9th International Conference on Cyber-Physical Systems (ICCPS)*, Porto, Portugal, 2018, pp. 351-352, doi: 10.1109/ICCPS.2018.00049.

12. E. Mousavinejad, F. Yang, Q. -L. Han and L. Vlacic, "A Novel Cyber Attack Detection Method in Networked Control Systems," in *IEEE Transactions on Cybernetics*, vol. 48, no. 11, pp. 3254-3264, Nov. 2018, doi: 10.1109/TCYB.2018.2843358.

13. Q. Deng and J. Sun, "False Data Injection Attack Detection in a Power Grid Using RNN," *IECON 2018 - 44th Annual Conference of the IEEE Industrial Electronics Society*, Washington, DC, USA, 2018, pp. 5983-5988, doi: 10.1109/IECON.2018.8591079.

14. P. Yermalovich and M. Mejri, "Formalization of Attack Prediction Problem," *2018 IEEE International Conference "Quality Management, Transport and Information Security, Information Technologies" (IT&QM&IS)*, St. Petersburg, Russia, 2018, pp. 280-286, doi: 10.1109/ITMQIS.2018.8525128.

15. E. Mousavinejad, F. Yang, Q. -L. Han, Q. Qiu and L. Vlacic, "Cyber Attack Detection in Platoon-Based Vehicular Networked Control Systems," *2018 IEEE 27th International Symposium on Industrial Electronics (ISIE)*, Cairns, QLD, Australia, 2018, pp. 603-608, doi: 10.1109/ISIE.2018.8433814.

16. E. Unal, S. Sen-Baidya and R. Hewett, "Towards Prediction of Security Attacks on Software Defined Networks: A Big Data Analytic Approach," *2018 IEEE International Conference on Big Data (Big Data)*, Seattle, WA, USA, 2018, pp. 4582-4588, doi: 10.1109/BigData.2018.8622524.