Sequential Pattern Mining: A Proposed Approach for Intrusion Detection Systems

Moemedi Lefoane
University of Bradford
Bradford, United Kingdom
m.lefoane@bradford.ac.uk

Ibrahim Ghafir
University of Bradford
Bradford, United Kingdom
i.ghafir@bradford.ac.uk

Sohag Kabir
University of Bradford
Bradford, United Kingdom
s.kabir2@bradford.ac.uk

Irfan-Ullah Awan
University of Bradford
Bradford, United Kingdom
i.u.awan@bradford.ac.uk

ABSTRACT

Technological advancements have played a pivotal role in the rapid proliferation of the fourth industrial revolution (4IR) through the deployment of Internet of Things (IoT) devices in large numbers. COVID-19 caused serious disruptions across many industries with lockdowns and travel restrictions imposed across the globe. As a result, conducting business as usual became increasingly untenable, necessitating the adoption of new approaches in the workplace. For instance, virtual doctor consultations, remote learning, and virtual private network (VPN) connections for employees working from home became more prevalent. This paradigm shift has brought about positive benefits, however, it has also increased the attack vectors and surfaces, creating lucrative opportunities for cyberattacks. Consequently, more sophisticated attacks have emerged, including the Distributed Denial of Service (DDoS) and Ransomware attacks, which pose a serious threat to businesses and organisations worldwide. This paper proposes a system for detecting malicious activities in network traffic using sequential pattern mining (SPM) techniques. The proposed approach utilises SPM as an unsupervised learning technique to extract intrinsic communication patterns from network traffic, enabling the discovery of rules for detecting malicious activities and generating security alerts accordingly. By leveraging this approach, businesses and organisations can enhance the security of their networks, detect malicious activities including emerging ones, and thus respond proactively to potential threats.

CCS CONCEPTS

• **Security and privacy → Intrusion detection systems**.

KEYWORDS

Scanning Detection, Sequential Pattern Mining, Unsupervised Learning, Intrusion Detection System, Network Security

ICFNDS 2023, December 21–22, 2023, Dubai, United Arab Emirates

ACM ISBN 978-1-4503-8734-7/21/12…$15.00
https://doi.org/10.1145/3508072.3508102

ACM Reference Format:
Moemedi Lefoane, Ibrahim Ghafir, Sohag Kabir, and Irfan-Ullah Awan. 2023. Sequential Pattern Mining: A Proposed Approach for Intrusion Detection Systems. In *The 7th International Conference on Future Networks & Distributed Systems (ICFNDS 2023), December 21–22, 2023, Dubai, United Arab Emirates.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3508072.3508102

1 INTRODUCTION

4IR has played a pivotal role in the digital transformation of businesses and industries. The COVID-19 pandemic has further accelerated this trend, forcing us to rely more heavily on technology for daily activities such as accessing government services and transportation. This paradigm shift has revolutionised how employees work, promoting remote work and increasing the use of online communication platforms such as Teams and Zoom. This shift has brought numerous benefits, including cost savings, increased productivity and efficiency. However, it has also increased the attack surface for adversaries, creating a lucrative opportunity for cyber-attacks due to the deployment of a large number of smart technologies that operate without human intervention. These technologies have also increased the risk of sophisticated attacks, such as multi-stage attacks [2, 3, 14, 21] and Distributed Denial of Service (DDoS) attacks [8, 9], which have been a serious threat in recent years. As businesses and industries continue to embrace digital transformation, it is important to remain vigilant and take proactive measures to mitigate the risks of cyber-attacks.

The stages of a cyber attack typically begin with reconnaissance, which involves gathering information about the target organisation to map its security posture. This is followed by a scanning attack, which is a pre-attack stage that adversaries use to identify potential attack vectors that can be exploited to gain access to the network. During this stage, port scanners, ping scanners, and related tools are employed to discover open ports and obtain information about the network services running on them, as well as details about the operating systems and versions in use. The output of this stage usually consists of a list of attack vectors that can be used to penetrate the target organisation's defences.

Attackers often aim to collect information in order to gain insights into the organisation's security posture. This information-gathering process is crucial for identifying vulnerabilities that can be exploited in the next stage of the attack. One common method of information gathering is port scanning, which involves sending

packets to the target host to initiate a TCP connection through a three-way handshake. Through this process, a scanner can determine the state of the port on the target network hosts by sending a packet with the SYN flag set and analysing the response from the host being scanned. There are various types of port scans, such as the syn scan, TCP connect scan, and stealth scan. The stealth scan is particularly effective as it limits the noise generated during the scan by not completing the full three-way handshake, thus making it more difficult to detect.

Network monitoring tools, such as Zeek [27] and Snort [7, 13, 17], are equipped with pre-defined rules and signatures that enable the detection of common scanning attacks. On the other hand, firewalls are typically deployed to secure networks, employing different sets of rules to filter out malicious traffic while allowing only legitimate traffic into the network. Given the availability of these tools and technologies, the likelihood of successful execution of scanning techniques by attackers is considerably low.

As technology continues to advance and security measures become more sophisticated, attackers are constantly developing new techniques to gain access to target networks. In addition to standard scanning methods, adversaries create custom scans that involve sending packets with combinations of TCP flags that are not typically used in normal communication. Firewalls are commonly employed to secure networks, and they typically include rules that filter out malicious packets, thereby preventing malicious traffic from penetrating the network. However, more sophisticated attackers aim to map firewall rules by understanding the traffic filtration rules implemented on the firewall. This helps them develop attack strategies that allow them to send traffic that probes the network in a manner that evades the implemented rules. By doing so, attackers can identify attack vectors and exploit them to gain unauthorised access to the target network. It is imperative for organisations to detect these malicious activities at an early stage to prevent ultimate attacks. Timely detection and appropriate countermeasures can protect organisations from severe financial and reputation damage.

This paper proposes an approach for intrusion detection of malicious activities in network traffic that utilises SPM techniques. As a proof of concept, this work focuses on detecting the second phase of a typical attack life cycle, which is the scanning phase. SPM is an unsupervised learning technique that extracts intrinsic communication patterns from network traffic. The patterns discovered through SPM are then used to detect scanning activities on the monitored network. Additionally, a rule-based approach is proposed as part of the system for the classification of scanning traffic based on the discovered sequential patterns.

The rest of this paper is organised as follows: Section 2 discusses related work, Section 3 presents the proposed methodology, The experimental setup, dataset used and results are discussed in Section 4 and finally, the conclusion of the paper is provided in Section 5.

2 RELATED WORK

Ananin et al. [1] conducted a comprehensive review of various port scan types, including scanning attacks, and developed a mathematical model for detecting anomalies related to these attacks. They evaluated their approach by implementing an algorithm derived from the mathematical models to test their detection model.

Birkinshaw et al. [5] proposed an Intrusion Detection and Prevention System (IDPS) designed to detect port scanning attacks and Denial of Service (DoS) attacks. The authors stress the importance of early detection, such as during port scanning, to prevent the potentially devastating impact of ultimate attacks such as DoS. Their proposed approach utilises Software Defined Network (SDN) technology and is capable of real-time detection. Moreover, the approach can be extended to include the detection of other types of malicious activities. The authors reported a low False Positive Rate (FPR) for their approach.

Husák et al. [18] conducted a study highlighting the underutilisation of data mining techniques in the cybersecurity domain. They provided an in-depth discussion of rule mining and SPM use cases, particularly in the context of cyber alert analysis. Moreover, they conducted a survey on alert correlation and attack prediction. The authors evaluated pattern mining techniques, considering speed, using a real dataset of alerts. Finally, they presented a comparison of different methods and shared valuable lessons learned, and thus demonstrated the importance of exploring the full potential of data mining techniques in the cybersecurity domain.

Tıktıklar et al. [29] conducted a study that investigated the existing SPM algorithms. The study analysed the underlying principles of the algorithms and performed a comparative analysis across various domains such as cybersecurity, telecommunications, air quality monitoring, and user behaviour analysis. The evaluation of the algorithms was based on a real-life telecommunications dataset. The study compared three SPM algorithms, namely GSP, Prefix Span, and CMRules, and concluded that their performance may vary depending on the dataset analysed.

Fournier-Viger [11] conducted a comprehensive survey on SPM and identified its trends for discovering patterns in sequential data. SPM algorithms have found numerous applications in different domains ranging from bioinformatics to e-commerce. One of the prominent applications of SPM is natural language processing, particularly in text analysis. In addition, SPM algorithms have been used in market analysis to analyse customers' purchasing patterns, which helps in recommending products to customers. The study discusses some popular SPM algorithms such as PrefixSpan, highlighting their strengths and weaknesses.

Jafarian et al. [19] proposed a DNS-based technique for detecting network scanning attacks aimed at enterprise networks, both internal and external. Their approach involves monitoring the network subnet's ingress and egress flow and correlating it with the preceding DNS query/response. This method has been shown to effectively detect scans with less overhead.

In their study, Yue et al. [30] analysed the Train Ethernet Consist Network (ECN), which is responsible for transmitting train control signals. They identified intrusion threats to the data security of railway vehicles due to the increased interaction between the train network and the external environment. To address these challenges, they proposed an ensemble-based IDS that can detect ECN attacks such as IP Scan, Port Scan, DoS and Man-in-the-Middle (MITM) attacks. Their proposed IDS employs Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) to detect such attacks. The authors evaluated their IDS on an ECN testbed and reported a high accuracy of 0.975.

In their study, Sagatov et al. [26] emphasised the significance of protecting networks against scanning attacks, which are often the first steps in exploiting network vulnerabilities. These attacks exploit protocol behaviour to gather information about open ports and the services running on a target network, which can then be used to exploit any discovered vulnerabilities. The researchers proposed a method to detect the initial stages of attacks in TCP and UDP, which could help address the challenges of defending against these attacks. They tested their method on a testbed they created and evaluated its effectiveness.

Aparicio-Navarro et al. [4] proposed an IDS that uses Fuzzy Cognitive Map (FCM) and Pattern-of-Life (PoL) techniques to detect malicious activities. The IDS is designed to address the increasing complexities of cyber attacks. In their evaluation, the team reported a high detection rate of 99.76% with a low FPR of 6.33%. Other intrusion detection approaches for detection some of malicious activities that have been a serious threat recently include machine learning approaches [6, 15, 16, 31]. Specifically, feature selection approaches contribute to improved performance evidenced by high True Positive Rate (TPR) and low FPR [20, 22].

3 PROPOSED METHODOLOGY

The proposed methodology is illustrated in Fig.1. The process involves taking network traffic as input into the network traffic filter module. This module processes and extracts sequences of network traffic between hosts, which are then passed on to the sequential pattern miner module. The sequential pattern miner module is responsible for mining frequent sequential patterns. The frequent sequential patterns are utilised for two important tasks. One is the detection module, where malicious activities are detected. The other task is the generation of detection rules. This involves the performance evaluation module and the network security team. The remaining part of this section provides more elaboration on the modules of the proposed approach.

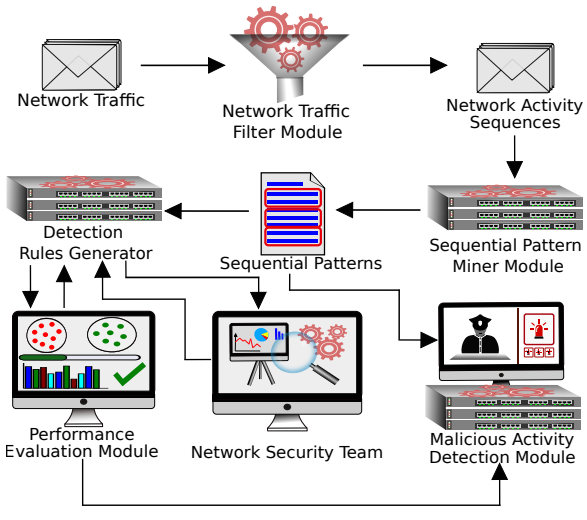

Figure 1: Proposed Methodology for the detection of malicious activities.

Sequential pattern mining is a technique used to extract valuable insights from sequential data in various domains. For instance, it is used in recommender systems to analyse sequences of products purchased together or subsequently, revealing crucial insights about customer buying behaviour. The discovered sequential patterns are then used to recommend products to customers based on their purchasing patterns [28]. Apart from the retail domain, sequential pattern mining has also been successfully employed in other domains, such as cybersecurity [18], to analyse sequential patterns.

The proposed system takes in network traffic as input, which is then processed by the Network Traffic Filter Module. This module extracts key features from the packets transmitted between hosts such as ICMP type and code IDs or TCP header flags. These key features are then organised into a sequence that accurately represents the activities between the hosts. The extracted features can be related to different communication activities between two hosts communicating through TCP, User Datagram Protocol (UDP) or any other protocol. The output of this module is traffic filtered with only relevant features organised as a database of sequences. This sequence database is passed to the Sequential Pattern Miner Module for further processing.

The generation and analysis of sequences are a crucial part of the proposed system. Sequential Pattern Mining Framework (SPMF), a data mining software, is used to extract sequential patterns from the sequences [10, 12]. The sequences generated from network traffic are preprocessed to transform them into a format compatible with SPMF. Since SPMF only takes integer values as input, the preprocessing includes converting feature values for the sequences into integers. Once the sequences of activities are ready, they are passed into the Sequential Pattern Miner Module, which extracts sequential patterns from the network traffic sequences for further analysis using SPMF revealing insights and patterns of network malicious activities taking place on the network.

The sequential pattern mining algorithm utilised and the one implemented on SPMF is PrefixSpan [11, 25]. PrefixSpan uses two techniques: database projection of subsequences within the databases and depth-first search for traversing the entire sequence database for mining frequent sequential patterns. This process of finding different sequential patterns is done recursively. To mine different sequential patterns the algorithm requires an input sequence database and minimum support, where minimum support means the frequency of occurrence of sequential patterns or how many sequences contain a particular sequential pattern.

Upon receiving input, the algorithm scans the entire database and counts the minimum support of each sequential pattern in the set of sequences. The minimum support for each of the sequential patterns is then evaluated against the minimum support. Any sequential pattern with support less than the minimum support is considered infrequent and is consequently eliminated. The process is repeated to find the next sequential pattern comprising of occurrence on one item followed by another item, this is performed for each of the sequences in the sequence database. Again, the minimum support is compared against the support of subsequences, and those found to be less frequent are eliminated. This process is continued until even longer and more frequent item sets are discovered [25]. One of the benefits of PrefixSpan algorithm is that it considers only the

observed sequences database as opposed to creating new ones like other algorithms do and is easy to extend.

The Sequential Pattern Miner Module produces sequential patterns that are passed to the Detection Rule Generator (DRG) module and the deployed Malicious Activity Detection module. This module is responsible for generating and evaluating detection rules. During the generation of detection rules, these patterns are passed to the rule generator for detecting different malicious activities based on the identified patterns. If the rule generator encounters a pattern that does not match any existing rules, the unknown malicious pattern is forwarded to the network security team. The team performs an in-depth analysis of the activities and generates new detection rules to update the detection rules to incorporate new rules for previously unknown malicious activities. The purpose of the Malicious Activity Detection module is to detect malicious activities by matching the sequential patterns with the predefined detection rules. Every now and then when new patterns are discovered, the network security team may update the detection rules and feed them back into the rule generator.

4 EVALUATION RESULTS

This section provides a discussion on the evaluation of the proposed approach. It is split into two subsections. The first is Sec 4.1, covering dataset description as well as steps followed and the second Sec 4.2 on the analysis and discussion.

4.1 Experimental Setup

To evaluate the effectiveness of the proposed system, the reconnaissance dataset [23] consisting of port scanning activities is utilised. Specifically, TCP three-way handshake traffic relating to TCP flags for setting up communication connections is derived from this dataset. As a proof of concept for the performance evaluation of the proposed system, a publicly available dataset by the Canadian Institute for Cybersecurity based at the University of New Brunswick is utilised [23]. This dataset is network traffic generated from 105 IoT devices and 33 different attacks have been executed including reconnaissance activities and more specifically port scanning. The experiment was performed following the steps illustrated in Fig. 1, the process begins with extraction of relevant features. The feature extraction process focuses on the TCP three-way handshake negotiation process between two hosts communicating through TCP. This approach provides a detailed evaluation of the network traffic and its patterns. Specifically, for each TCP connection setup, a sequence is generated for network packets, with a particular emphasis on TCP flags for each connection setup. This enables an in-depth and critical analysis which then leads to gaining insights on these malicious activities in terms of how they work and target goals. This then results in the development of countermeasures to combat these malicious activities.

4.2 Analysis and Discussion

This section provides a detailed discussion of the results of the experimental setup. Fig. 2 provides a sample of sequences generated for each pair of source IP address & source port and destination IP address & destination port, while setting up network communication. For example, for two hosts communicating, namely the scanner and the target hosts, the sequence generated might be 0x0002 - 0x0012 - 0x0010 - 0x0004. This communication sequence would translate to [SYN] –> [SYN, ACK] –> [ACK] –> [RST]. This communication sequence translates to a type of scanning activity known as a stealth scan or a half-open scan. Fig. 3 shows a sample of frequent sequential patterns within the port scanning traffic uncovered by the SPM process. While existing signature-based detection approaches can already detect this type of scan and related, advanced cyber attackers do not confine themselves to the standard communication patterns, they however, experiment with different custom scans that are not necessarily aimed to determine whether a particular port is open but instead the goal is mapping firewall rules [24]. Once the firewall rules on the target network are well understood, the adversaries can then develop a successful strategy to breach firewalls and further probe the network for running services. This leads to the discovery of version numbers of these services and ultimately vulnerabilities which are exploited to gain access. With the proposed SPM system, these custom patterns will be detected by rules generated for the detection of such malicious activities.

```
     1 - 0x0002 - 0x0012 - 0x0010 - 0x0018 - 0x0010 - 0x0010 - 0x0010 - 0x0010 -
         0x0018 - 0x0018 - 0x0010 - 0x0018 - 0x0011 - 0x0010 - 0x0010
     2 - 0x0002 - 0x0012 - 0x0010 - 0x0018 - 0x0010 - 0x0010 - 0x0018 - 0x0010 -
         0x0018 - 0x0018 - 0x0018 - 0x0018 - 0x0010 - 0x0010 - 0x0018 - 0x0011 - 0x
     3 - 0x0018 - 0x0018 - 0x0010 - 0x0018 - 0x0018 - 0x0010 - 0x0018 - 0x0018 -
         0x0018 - 0x0010 - 0x0018 - 0x0018 - 0x0010 - 0x0018 - 0x0018 - 0x0010 - 0x
         0x0018 - 0x0018 - 0x0010 - 0x0018 - 0x0018 - 0x0018 - 0x0010 - 0x0018 - 0x
         0x0018 - 0x0010 - 0x0018 - 0x0018 - 0x0010 - 0x0018 - 0x0018 - 0x0010 - 0x
         0x0018 - 0x0010 - 0x0018 - 0x0018 - 0x0018 - 0x0010 - 0x0018 - 0x0010 - 0x
         0x0018 - 0x0010 - 0x0018 - 0x0018 - 0x0010 - 0x0018 - 0x0010 - 0x0018 - 0x
         0x0010 - 0x0018 - 0x0010 - 0x0018 - 0x0018 - 0x0010 - 0x0018 - 0x0018 - 0x
                                          ⋮
  5897 - 0x0002 - 0x0012 - 0x0010 - 0x0018 - 0x0010 - 0x0018 - 0x0011 - 0x0010
  5898 - 0x0002 - 0x0012 - 0x0010 - 0x0018 - 0x0010 - 0x0018 - 0x0011 - 0x0010
  5899 - 0x0002 - 0x0012 - 0x0010 - 0x0018 - 0x0010 - 0x0018 - 0x0011 - 0x0010
                                          ⋮
160145 - 0x0014
160146 - 0x0014
```

Figure 2: Samples of Network Activity Sequences

```
[PSH, ACK]
[SYN]-->[ACK]-->[PSH, ACK]-->[ACK]-->[FIN, ACK]
[PSH, ACK]-->[ACK]
[PSH, ACK]-->[FIN, ACK]-->[ACK]
[SYN]-->[SYN, ACK]-->[FIN, ACK]
[SYN]-->[SYN, ACK]-->[PSH, ACK]
[SYN]-->[SYN, ACK]-->[ACK]-->[FIN, ACK]-->[ACK]
[ACK]-->[PSH, ACK]
[SYN]-->[SYN, ACK]-->[ACK]
[PSH, ACK]-->[ACK]-->[PSH, ACK]-->[ACK]
[ACK]-->[PSH, ACK]-->[ACK]-->[PSH, ACK]
```

Figure 3: Samples of Frequent Sequential Patterns

The proposed approach is evaluated by analysing the TCP handshake traffic and labelling it for horizontal scanning. To create ground truths for horizontal port scans, the approach considers a scenario where a source IP address scans multiple IP addresses N on the same port. The number of IP addresses scanned N can be

set to a sufficiently large enough value to constitute a horizontal scan. A detection rule is then developed to identify similar patterns across multiple devices, which is indicative of the same type of malicious activity. Beyond just detecting a scan, these frequent sequential patterns detected on multiple hosts are forwarded to the network security team for further insights into the goals of the malicious activity. This approach can reveal firewall rules that the malicious activity is attempting to circumvent. Once the goals of the sequential patterns are determined, specific rules can be developed to detect similar patterns more quickly. The confusion matrix in Fig. 4 shows that the proposed approach has a high true positive rate (TPR) and a low false positive rate (FPR), with values of 99.12% and 0%, respectively.

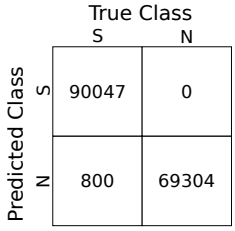

Figure 4: Confusion Matrix of the Scanning Detection

5 CONCLUSION

This paper presents an approach for IDSs that utilises SPM for detecting malicious activities in network traffic. The proposed system uses SPM to identify sequential patterns from the network traffic, which are then utilised to detect malicious traffic using a rule-based engine. The system is evaluated on a publicly available reconnaissance dataset for detecting port scanning activities and is capable of detecting advanced custom scans and stealth scans. The proposed system also facilitates the generation of security rules by forwarding unknown sequential patterns related to new advanced custom scans to the network security team for further analysis. This approach provides efficient and realistic labelling of the scanning attack and improves network security. Future work will focus on developing and adding more rules to the system to enable the detection of other malicious activities in addition to port scanning.

REFERENCES

[1] Evgeny V. Ananin, Arina V. Nikishova, and Irina S. Kozhevnikova. 2017. Port scanning detection based on anomalies. In *2017 Dynamics of Systems, Mechanisms and Machines (Dynamics)*. 1–5. https://doi.org/10.1109/Dynamics.2017.8239427

[2] Francisco J Aparicio-Navarro, Timothy A Chadza, Konstantinos G Kyriakopoulos, Ibrahim Ghafir, Sangarapillai Lambotharan, and Basil AsSadhan. 2019. Addressing multi-stage attacks using expert knowledge and contextual information. In *2019 22nd Conference on Innovation in Clouds, Internet and Networks and Workshops (ICIN)*. IEEE, 188–194.

[3] Franciso J Aparicio-Navarro, Konstantinos G Kyriakopoulos, Ibrahim Ghafir, Sangarapillai Lambotharan, and Jonathon A Chambers. 2018. Multi-stage attack detection using contextual information. In *MILCOM 2018-2018 IEEE Military Communications Conference (MILCOM)*. IEEE, 1–9.

[4] Francisco J. Aparicio-Navarro, Konstantinos G. Kyriakopoulos, Yu Gong, David J. Parish, and Jonathon A. Chambers. 2017. Using Pattern-of-Life as Contextual Information for Anomaly-Based Intrusion Detection Systems. *IEEE Access* 5 (2017), 22177–22193. https://doi.org/10.1109/ACCESS.2017.2762162

[5] Celyn Birkinshaw, Elpida Rouka, and Vassilios G. Vassilakis. 2019. Implementing an intrusion detection and prevention system using software-defined networking: Defending against port-scanning and denial-of-service attacks. *Journal of Network and Computer Applications* 136 (2019), 71–85. https://doi.org/10.1016/j.jnca.2019.03.005

[6] Andrew Carlin, Mohammad Hammoudeh, and Omar Aldabbas. 2015. Intrusion detection and countermeasure of virtual cloud systems-state of the art and current challenges. *International Journal of Advanced Computer Science and Applications* 6, 6 (2015).

[7] Cisco. 2021. Snort. https://www.snort.org/. Accessed: 2021-10-20.

[8] Diab M Diab, Basil AsSadhan, Hamad Binsalleeh, Sangarapillai Lambotharan, Konstantinos G Kyriakopoulos, and Ibrahim Ghafir. 2019. Anomaly detection using dynamic time warping. In *2019 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC)*. IEEE, 193–198.

[9] Diab M Diab, Basil AsSadhan, Hamad Binsalleeh, Sangarapillai Lambotharan, Konstantinos G Kyriakopoulos, and Ibrahim Ghafir. 2021. Denial of service detection using dynamic time warping. *International Journal of Network Management* 31, 6 (2021), e2159.

[10] Philippe Fournier-Viger. 2021. SPMF An Open-Source Data Mining Library. https://www.philippe-fournier-viger.com/spmf/index.php/. Accessed: 2023-11-24.

[11] Philippe Fournier-Viger, Jerry Chun-wei Lin, Rage Uday Kiran, Yun Sing Koh, and Rincy Thomas. 2017. A Survey of Sequential Pattern Mining. https://api.semanticscholar.org/CorpusID:9784038

[12] Philippe Fournier-Viger, Jerry Chun-Wei Lin, Antonio Gomariz, Ted Gueniche, Azadeh Soltani, Zhihong Deng, and Hoang Thanh Lam. 2016. The SPMF Open-Source Data Mining Library Version 2. In *Machine Learning and Knowledge Discovery in Databases*, Bettina Berendt, Björn Bringmann, Élisa Fromont, Gemma Garriga, Pauli Miettinen, Nikolaj Tatti, and Volker Tresp (Eds.). Springer International Publishing, Cham, 36–40.

[13] RaviTeja Gaddam and M. Nandhini. 2017. An analysis of various snort based techniques to detect and prevent intrusions in networks proposal with code refactoring snort tool in Kali Linux environment. In *2017 International Conference on Inventive Communication and Computational Technologies (ICICCT)*. 10–15. https://doi.org/10.1109/ICICCT.2017.7975177

[14] Ibrahim Ghafir, Mohammad Hammoudeh, Vaclav Prenosil, Liangxiu Han, Robert Hegarty, Khaled Rabie, and Francisco J. Aparicio-Navarro. 2018. Detection of advanced persistent threat using machine-learning correlation analysis. *Future Generation Computer Systems* 89 (2018), 349–359. https://doi.org/10.1016/j.future.2018.06.055

[15] Mohammad Hammoudeh, Gregory Epiphaniou, Sana Belguith, Devrim Unal, Bamidele Adebisi, Thar Baker, ASM Kayes, and Paul Watters. 2020. A service-oriented approach for sensing in the Internet of Things: Intelligent transportation systems and privacy use cases. *IEEE Sensors Journal* 21, 14 (2020), 15753–15761.

[16] Mohammad Hammoudeh and Robert Newman. 2015. Information extraction from sensor networks using the Watershed transform algorithm. *Information Fusion* 22 (2015), 39–49.

[17] Xiaojin Hong, Changzhen Hu, Zhigang Wang, Guoqiang Wang, and Ying Wan. 2012. VisSRA: Visualizing Snort Rules and Alerts. In *2012 Fourth International Conference on Computational Intelligence and Communication Networks*. 441–444. https://doi.org/10.1109/CICN.2012.207

[18] Martin Husák, Jaroslav Kašpar, Elias Bou-Harb, and Pavel Čeleda. 2017. On the Sequential Pattern and Rule Mining in the Analysis of Cyber Security Alerts. In *Proceedings of the 12th International Conference on Availability, Reliability and Security* (Reggio Calabria, Italy) *(ARES '17)*. Association for Computing Machinery, New York, NY, USA, Article 22, 10 pages. https://doi.org/10.1145/3098954.3098981

[19] Jafar Haadi Jafarian, Masoumeh Abolfathi, and Mahsa Rahimian. 2023. Detecting Network Scanning Through Monitoring and Manipulation of DNS Traffic. *IEEE Access* 11 (2023), 20267–20283. https://doi.org/10.1109/ACCESS.2023.3250106

[20] Moemedi Lefoane, Ibrahim Ghafir, Sohag Kabir, and Irfan-Ullah Awan. 2021. Machine Learning for Botnet Detection: An Optimized Feature Selection Approach. In *The 5th International Conference on Future Networks & Distributed Systems* (Dubai, United Arab Emirates) *(ICFNDS 2021)*. Association for Computing Machinery, New York, NY, USA, 195–200.

[21] Moemedi Lefoane, Ibrahim Ghafir, Sohag Kabir, and Irfan-Ullah Awan. 2022. Multi-stage Attack Detection: Emerging Challenges for Wireless Networks. In *2022 International Conference on Smart Applications, Communications and Networking (SmartNets)*. 01–05. https://doi.org/10.1109/SmartNets55823.2022.9994027

[22] Moemedi Lefoane, Ibrahim Ghafir, Sohag Kabir, and Irfan-Ullah Awan. 2023. Unsupervised Learning for Feature Selection: A Proposed Solution for Botnet Detection in 5G Networks. *IEEE Transactions on Industrial Informatics* 19, 1 (2023), 921–929. https://doi.org/10.1109/TII.2022.3192044

[23] Euclides Carlos Pinto Neto, Sajjad Dadkhah, Raphael Ferreira, Alireza Zohourian, Rongxing Lu, and Ali A. Ghorbani. 2023. CICIoT2023: A Real-Time Dataset and Benchmark for Large-Scale Attacks in IoT Environment. *Sensors* 23, 13 (2023). https://doi.org/10.3390/s23135941

[24] NMAP. 2023. Custom Scan Types with –scanflags. https://nmap.org/book/scan-methods-custom-scanflags.html/. Accessed: 2023-11-26.
[25] Jian Pei, Jiawei Han, B. Mortazavi-Asl, Jianyong Wang, H. Pinto, Qiming Chen, U. Dayal, and Mei-Chun Hsu. 2004. Mining sequential patterns by pattern-growth: the PrefixSpan approach. *IEEE Transactions on Knowledge and Data Engineering* 16, 11 (2004), 1424–1440. https://doi.org/10.1109/TKDE.2004.77
[26] E.S. Sagatov, S. Mayhoub, A.M. Sukhov, F. Esposito, and P. Calyam. 2021. Proactive Detection for Countermeasures on Port Scanning based Attacks. In *2021 17th International Conference on Network and Service Management (CNSM)*. 402–406. https://doi.org/10.23919/CNSM52442.2021.9615577
[27] The-Zeek-Project. 2021. Zeek. https://zeek.org/. Accessed: 2021-10-20.
[28] Ridho Trivonanda, Rahmad Mahendra, Indra Budi, and Rani Aulia Hidayat. 2020. Sequential Pattern Mining for e-Commerce Recommender System. In *2020 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*. 393–398. https://doi.org/10.1109/ICACSIS51025.2020.9263192
[29] Doruk Tıktıklar, Gürsel Baltaoğlu, Efsa Çakır, Zeynep Küçük, and Mehmet S. Aktas. 2021. On the Comparative Analysis of Sequence Mining Algorithms: Case Study in Telecommunications. In *2021 6th International Conference on Computer Science and Engineering (UBMK)*. 145–150. https://doi.org/10.1109/UBMK52708.2021.9558935
[30] Chuan Yue, Lide Wang, Dengrui Wang, Ruifeng Duo, and Xiaobo Nie. 2021. An Ensemble Intrusion Detection Method for Train Ethernet Consist Network Based on CNN and RNN. *IEEE Access* 9 (2021), 59527–59539. https://doi.org/10.1109/ACCESS.2021.3073413
[31] Yuan Zhang, Qinghai Yang, Sangarapillai Lambotharan, Konstantinos Kyriakopoulos, Ibrahim Ghafir, and Basil AsSadhan. 2019. Anomaly-based network intrusion detection using SVM. In *2019 11th International conference on wireless communications and signal processing (WCSP)*. IEEE, 1–6.