International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Networks Cyber Security Model by Using Machine Learning Techniques

Farah Abbas Obaid Sari[1, *], Ali Abdulkarem Habib Alrammahi[1,*], Asaad Shakir Hameed[2,3], Haiffa Muhsan B. Alrikabi[4], Abeer A. Abdul–Razaq[5], Huda Karem Nasser[3], Mohammed F. AL-Rifaie[6]

Submitted: 19/09/2022 **Accepted:** 24/12/2022

Abstract: Since artificial intelligence relies on learning just like humans, it is useful to use these algorithms to address cyber-attacks, which represent the greatest concerns for network users, especially companies and institutions, as a result of the dire consequences of these attacks such as large material losses and the leakage or falsification of important data. The methods used to detect cyber-attacks are slow or they detect attacks after their occurrence and then analyze them and issue reports. In this research, we propose a conceptual framework that contains new rules that are used along with the previous rules for the purpose of creating a network monitoring tool in order to counteract cyber-attacks in real-time by relying on artificial intelligence algorithms such as classification and prediction based on user behavior. GMM algorithm has been suggested in this paper because of its efficiency in comparison with the commonly used algorithms in this sector like k-means it depends on behavioral similarities, not on distance.

Keywords: *Cyber security, Network Monitoring, GMM, AI in cyber security, cyber-attack*

1. Introduction

The great technological development in the current era explains the great need to increase network communications in a wider manner, and in view of this, one of the most important challenges that network users face is ensuring cybersecurity [1].

Cyber-attacks are terrorist attacks that target a wide variety of networks, such as banks, government websites, and even educational sites in universities. These attacks may lead to financial losses or the forgery and leakage of sensitive data [2]. Therefore, ensuring cyber security is a priority for all specialists, which explains the relentless efforts of researchers in this topic. In general, common methods of countering cyber-attacks depend on identifying the source of the attack and intercepting it at a specific point. However, these methods have become primitive as a result of the development of types of attacks and the use of artificial intelligence techniques in attack operations [3].

With the entry of the Internet of Things into the arena, the increase in the use of networks leads to an increase in cyber-attacks with a positive relationship. Despite attempts to control cybersecurity, there is no sufficiently effective method. As the methods of protecting cyber security relied on different techniques based on establishing a buffer line called the first line of protection based on techniques such as access control and authentication [4].

Artificial intelligence technologies are entering the cybersecurity industry and market. This is evident by its entry into offensive operations as well as defense and penetration detection. Several artificial intelligence algorithms have been used in this regard, especially machine learning algorithms [5] . For example, classification algorithms are used to sort attacks, classify them, and store them in databases in order to compare them to new attacks and find out harmful data before the attack occurs. In addition, data mining algorithms are used in order to discover and deal with user behavior before an attack occurs [6].

Threat response time is among the most pivotal measures of the effectiveness of cybersecurity teams. From time of use to time of spread, malicious attacks are known to move very quickly. In the past, attackers would scrutinize network

[1] *Department of Computer Sciences, Faculty of Computer Science and Mathematics, University of Kufa, Iraq;*
[2] *Performance Quality Department, Mazaya University College, Iraq;*
[3] *Department of Mathematics, General Directorate of Thi-Qar Education, Ministry of Education, Iraq;*
[4] *Department of Mathematics, College of Education for Pure Sciences, Thi-Qar University, Iraq;*
[5] *Department of Mathematics, College of Computer Science and Mathematics, Thi-Qar University, Iraq;*
[6] *Department of Information and Communications, Basra University College of science and technology, Iraq*
Corresponding author's Email: faraha.altaee@uokufa.edu.iq; alia.alramahi@uokufa.edu.iq

permissions and sideways disable the security weapon, sometimes for weeks on end, before launching their attack.

Unfortunately, cyber defense experts are not the only ones benefiting from technological innovations; Automation has since become more common in cyberattacks. Machine learning-based programming can help highlight commonalities between a new threat and pre-selected threats to aid in attack detection. This is something that humans cannot do effectively in a timely manner, and it further highlights that adaptive security models are essential. From this point of view, machine learning can also make it easier for teams to anticipate new threats and reduce latency due to increased threat awareness [7].

In this paper we will implement dynamic programming techniques with clustering and using multiple algorithms. The goal is to discover the best methods capable of identifying and detecting Malware. The larger the database of attacks, the greater the opportunity to protect the network. Gaussian Mixture Model (GMM) sets are implemented. AI algorithms have the ability to enhance cybersecurity in order to reduce network attacks before they occur.

2. Related works

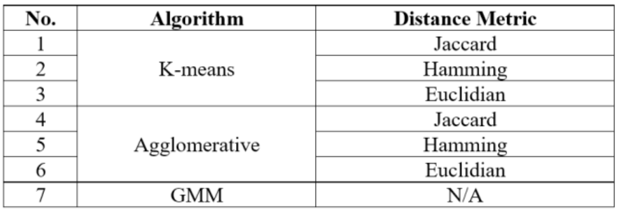
Several techniques are used to test the ability of network protection and cybersecurity systems by executing simultaneous attacks for the purpose of testing or by designing simulated attacks. The design of these simulators makes great use of the system in terms of identifying vulnerabilities as well as the time used for the purpose of detecting the attack and training the system [8]. These processes are essential in the design of any protection system and are considered as a test of the efficiency of the system before it is actually issued and used [9].

There are many studies that have investigated the use of artificial intelligence algorithms in the face of cyber-attacks for networks as fast algorithms and are distinguished by their ability to learn in order to develop themselves in the face of new attacks. The study [9] proposed a system for monitoring data security on the network by presenting a tool based on the use of genetic algorithms and chaotic neural networks by generating random numbers that are used as usage parameters for the purpose of encrypting data and then reversing the process for decryption.

The study [11] has introduced a tool to ensure the cybersecurity of the network by relying on machine learning algorithms in order to detect hackings and attacks on the network and issue reports on users and attackers. Alberto Perez [12] proposed a system that relied on machine learning algorithms to counter cyber-attacks on networks. As these algorithms were used to give greater flexibility to the system in order to learn and develop methods of detecting and thwarting attacks. The literature [13] suggested a network defense system that supports machine learning using neural network algorithms. Usually, neural networks are not used in cybersecurity, so other techniques are recommended. The study [14] proposed a defense system for cybersecurity based on artificial intelligence techniques, where a tool was designed to examine the network based on the application of the SVM algorithm by classifying attacks and usual methods and storing them in the form of classifications based on behavioral similarity and comparing them with current behavior and making the decision based on distinguishing commonalities.

In this paper, we propose a tool for monitoring and securing the system from cyber-attacks, depending on the GMM algorithm, by creating models that contain groups representing the previous attacks that were recorded and compared with the current behavior in order to protect the network. We propose to use the GMM algorithm because it does not depend on distance like other popular algorithms in this field such as k-means, but rather it depends on the similar behavior of samples as it is shown in table 1.

Table 1. Selected clustering algorithms and corresponding distance metrics [15]



No.	Algorithm	Distance Metric
1	K-means	Jaccard
2		Hamming
3		Euclidian
4	Agglomerative	Jaccard
5		Hamming
6		Euclidian
7	GMM	N/A

3. Cyber Security

Processing data in cybersecurity is a complex process in most cases, so human supervision is used for this. For example, some e-mail messages are filtered in the form of malicious messages and are archived directly in Spam with the ability for the user to read them and transfer them to the incoming mail if it is not Harmful [16].

Among some of the most well-known cybersecurity applications is what appears in browsers in the URL portion, where websites are filtered through algorithms that mark the machine into malicious sites that are indicated in red and natural ones that are green. Also, cybersecurity data processing appears in word processing in social networking sites by training systems to discover some phrases that incite hatred or racist discourse, using machine learning algorithms such as Naive Bayes [17].

Machine learning excels at tedious tasks such as identifying and adapting to a data pattern; Humans are not well suited to these types of tasks due to fatigue from tasks and a general lack of tolerance for monotony. So, while the interpretation of data analysis is still under human management, machine learning can help frame the data into a readable, breakdown-ready presentation. Machine learning cybersecurity comes in several different forms, each with its own unique benefits.

3.1 Data Classification

Data classification works by using predefined rules to assign categories to data points. Categorizing these scores is an important part of building a profile of attacks, vulnerabilities, and other aspects of proactive security. This is central to the fusion between machine learning and cybersecurity [18].

3.2 Data Collection

Data aggregation takes outliers to classify predefined rules, and puts them into "clusters" of data with common features or individual advantages. For example: this can be used when analyzing attack data for which the system is not already trained. These groups can help determine how an attack occurred, as well as what was accessed and detected [19].

3.3 Predictive Estimation

Predictive estimation is the most forward-thinking process of the machine learning component. This feature is achieved by anticipating possible outcomes by evaluating existing datasets. This can be used primarily to build threat models, identify fraud prevention, and protect against data breaches, and is a key component of many predictive endpoint solutions [20].

3.4 Recommended Courses of Action

Recommended workflows increase the proactivity of a machine learning security system; These warnings are based on past behavior patterns and decisions, and provide naturally suggested courses of action [21]. It is important to reiterate here that this is not true autonomous AI intelligent decision making. Instead, it is an adaptive inference framework that can access through pre-existing data points to conclude logical relationships. This type of tool can be of great help in responding to threats and mitigating risks.

4. Methodology

In this section we present the implementation of dynamic programming and assembly with different artificial intelligence techniques. They are compared among themselves to find out the best performance in the ability to detect malware and therefore this technology can be relied on in our work.

4.1 Dynamic Programming Implementation

This work focuses on the representation of dynamic program behavior based on the system-call co-occurrence matrix proposed by Shu et al. [22]. This representation was used to determine the similarity of behavior between the two malware samples. was extended to identify malware from benign applications using only this representation. Moreover, this programmatic representation could be easily transformed into a set of functions for machine learning algorithms and used as input for GMM clustering. Program presentation by Shu et al. It consists of two matrix primitives that can be computed from the "call" trajectory. A "call" here can be at any level of abstraction, from a specific function called by a program to a specific operating system API function to a system call. A co-occurrence matrix is defined as an $m\times m$ binary matrix O, where

OI,j ,it is $True$ if i occurred before j, else it will be $False$

Then an occurrence frequency matrix is define as an $m\times m$ matrix F where

fi,j = count(occurrences of i before j)

Suppose for the example that the system made the calls as follows

NTOpenFile

NtSetInformationFile

NTReadFile

NTReadFile

NTReadFile

NTWriteFile

NTCloseFile

The co-occurrence matrix will be as follows,

Calls	NTOpenFile	NtSetInformationFile	NTReadFile	NTWriteFile	NTCloseFile
NTOpenFile	0	1	0	0	0
NtSetInformationFile	0	0	1	0	0
NTReadFile	0	0	1	1	0
NTWriteFile	0	0	0	0	1
NTCloseFile	0	0	0	0	0

However, the occurrence frequency matrix will be as follows,

Calls	NTOpenFile	NtSetInformationFile	NTReadFile	NTWriteFile	NTCloseFile
NTOpenFile	0	1	0	0	0
NtSetInformationFile	0	0	1	0	0
NTReadFile	0	0	2	1	0
NTWriteFile	0	0	0	0	1
NTCloseFile	0	0	0	0	0

The idea is to create an array of system calls from which to attack and hack. The calls are grouped into an array and then the iterations of the call into another array. By performing simple arithmetic operations, the most frequent calls will be identified and compared to each other.

4.2 Clustering Implementation

Three main methods were evaluated in this work through their application in the representation of the mentioned program. These methods are:

1. K-Means Clustering: Calculates the lowest value for the distance from the mean or midpoint of the identified samples. And then recalculating the midpoint of the cluster until convergence is reached [23].

2. Agglomerative Clustering is the act of categorizing similar data points into specific groups (clusters). Whereas, data points classified in the same group have similar characteristics or features, while data points classified in different groups have very different characteristics or features [24]. Where the groups are merged repeatedly depending on reducing the distance measurements until the desired number of groups is achieved [25].

3. Gaussian mixture model (GMM): One of the main drawbacks of the K-Means algorithm is its simple use of the group mean as its center. We can see that this method is not the best way to do the aggregation for some data, but in GMM it gives us more flexibility compared to the aggregation algorithm (K-Means). Where in the Gaussian mixed model (GMMs), we assume that the data points follow a Gaussian distribution; That is, the normal distribution [26,27].

The number of groups is considered the number of malware families in the test groups.

Jaccard Distance: to measure the distance between any two groups which is the inverse of the similarity of jaccard which is calculated as,

$$J(A,B) = \frac{(|A \cap B|)}{(|A \cup B|)} \qquad (1)$$

Hamming Distance: To measure the distance between any two groups according to the order condition, which is calculated as,

$$d(p,q) = \sqrt{(q_1 - \rho_1)^2 + (q_2 - \rho_2)^2} \qquad (2)$$

Euclidean distance: to measure the distance between two sets of n dimensions.

First, the performance of distance-based clustering algorithms will be compared and then GMM is applied to show the difference and determine which algorithm has the best performance. In order to evaluate the mentioned algorithms, the distance between the co-occurrence matrices was relied on using different samples. As for GMM, it can be applied directly to the sample occurrence frequency matrix, but after the calls are processed in pairs and used as a feature until the repetition frequency matrix turns into a full vector of features.

4.3 Performance Evaluation

In order to evaluate the performance of all algorithms, the distance scale will be based on the above, so we need a set of malware data to test the algorithms on. In order to obtain better reliability, we will collect the data ourselves by relying on VirusShare. First of all, we filter the data in order to obtain data that is suitable for the purpose of our project and is compatible with the analysis environment. After filtering the data and deleting the incompatible ones, we run the remaining 107 samples in the sandbox environment. As for the system calls, they were collected by relying on the DrStrace program and converted into an iteration matrix. Table 2 shows the most important families that emerged for malware.

Table 2: Malware families by antivirus label.

No	Antivirus Label	Count
1	Trojan.Jevafus.A	31
2	Trojan.SMSHoax.X	29
3	Backdoor.Turkojan.DQ	23
4	Backdoor.Optix.Pro.1.3.Dam.2	16
5	Trojan.Spy.BZub.NHN	8

Two scales are used to compare the algorithms:

1. The Adjusted Rand Index (ARI): It measures the similarity between two groups by comparing the number of pairs in the sample chosen in the same group or in different groups. The range on this scale is between -1 and 1 where the higher the value, the greater the similarity.

2. Adjusted Mutual Information Index (AMI): It measures the extent of similarity between two groups, and the range in this scale is between -1 and 1, where the higher the value, the greater the similarity.

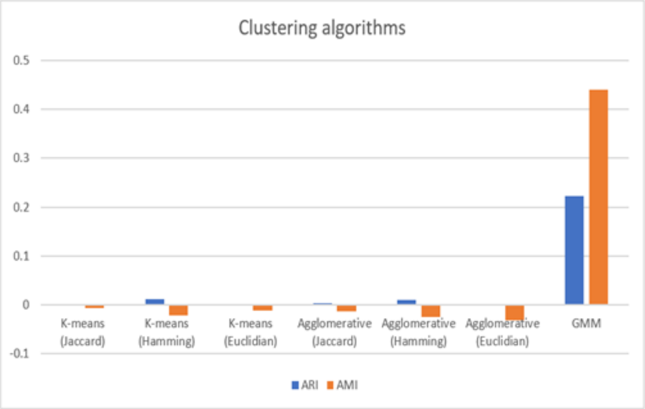


Figure 1: The clustering performance

The implementation of the representation in this work relies on the aforementioned foundation of iteration arrays and system calls. Algorithms analyze calls that have a high probability of occurring together or sequentially. For example, when the system calls NtOpenFile, it is very likely that the next call will be either NtReadFile or NtWriteFile. GMM estimates the probability density of this sequence of calls occurring, and k-means and agglomerative clustering measures each feature equally as it can negatively affect clustering. The computational results indicate a clear superiority of the k-means and agglomerative clustering algorithms over GMM in the data collection process due to the high dimensionality of the data. 314 calls were observed in the system, resulting in $(314)^2$ features. However, the minimum variance was used in order to reduce the dimensions of the data in order for GMM clustering to be possible.

4.4 Gaussian Mixture Model (GMM)

In this paper, the GMM algorithm is used for the purpose of clustering samples of groups that represent parameters of cyberattacks. The data is divided into interconnected families of cyber-attacks and malware that were previously exposed, and based on this division, the events, and behaviors that the user performs are compared in order to ensure his behavior and distinguish it whether it was a cyber-attack or a natural state. We suggest using the GMM algorithm due to its high efficiency in collecting samples, as it does not depend on distances such as k-means, but rather it depends on the relative performance and the classification of operations in the form of groups.

4.5 Conceptual Framework

In this section we present a system architecture to illustrate the tool components presented in this paper. In the first stage, samples from previous cyber-attacks and malware are collected and stored in the system in the form of groups using artificial intelligence techniques in order to create models that are later trained and used in the process of comparison with user behavior as it is shown in Figure 2.

In this research, work is divided into phases consisting of three basic stages: collection, verification and testing. As a first step, data necessary for the system's operation such as IP address, Port, Gate, etc. are collected. And then create a test model where the model is trained based on artificial intelligence algorithms such as GMM, where the results are classified at the beginning and then the model is trained and stored. This process can be performed by the user to monitor the activity of a specific user by entering his IP address into the scan tool in order to collect data about that user only. In addition, the system user can specify the required time period to check the network by specifying the period within the form. After creating a set of models, they are tested on demand, for example General Form to work in general and to check the network as a whole or a special form to check the user 192.168.50.90.

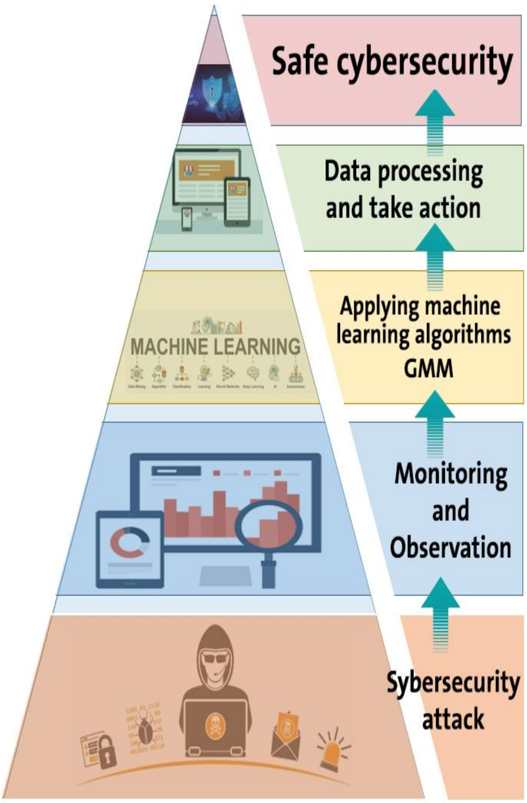

Safe cybersecurity
Data processing and take action
MACHINE LEARNING
Applying machine learning algorithms GMM
Monitoring and Observation
Sybersecurity attack

Figure 2. Conceptual frame work of the system

In the second stage, the system is monitored and observed, as this stage is based on verification and testing in order to record user behavior and verify it in the next stage. In the third stage, the user behavior that was observed in the previous stage is compared with the models stored in the system by applying artificial intelligence algorithms such as the GMM algorithm, where the data is divided into groups and families based on the similar behavior of the previous attacks. This stage is the main stage in this proposed system, as the user's behavior is compared with these models to determine whether it was a cyber-attack or normal behavior.

In the data processing stage, the decision about the user is made based on the outcome of the previous stage.

Despite all the fruitful dialogue about the future of this form of security, there are still limitations to note:

Machine learning requires datasets but may conflict with data privacy laws. Training software systems require a lot of data points to build accurate models, which doesn't mix well with the "right to be forgotten". Human identifiers for some data may cause breaches, so potential solutions should be considered. Possible fixes include making it nearly impossible for systems to access the original data once the software is trained.

5. Conclusion

Data protection is the main challenge facing all network users, and with the advent of IoT use, this has led to an increase in cyber-attacks on the other side. As indicated in previous studies, the most severe cyber-attacks are those that rely on artificial intelligence techniques. Therefore, in this paper, a model is presented for monitoring the network and examining attacks based on artificial intelligence techniques. It was suggested to use the GMM algorithm to collect the data samples and classify them into groups that would be stored in the system in the form of models. These models are created and trained to compare user behavior with the model samples that represent cyber-attacks and based on this decision is made in order to preserve the security of the cyber network.

The GMM algorithm has often proved to be highly efficient in classifying samples versus other classification algorithms because it does not depend on distance but rather on the sampling behavior of the samples. In addition, most of the previous studies provided solutions based on algorithms subject to human supervision, such as Naive Bayes, SVM, Nearest Neighbor and etc. Therefore, this paper is a basic structure for developing work in cybersecurity using unsupervised algorithms.

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

The theoretical work was built by the first and second author, who also carried out the analytic computations and numerical implementation. The second and third authors finished the literature review and the conceptual framework. The fourth, fifth and sixth authors analyzed the literature review and the clustering. The editing and final revision was done by the third and the sixth authors.

References

[1] M. Harini and M. D. Reddy, “Optimal location and sizing of distributed generation units for maximum loss reduction using Teaching Learning Based Optimization through Matlab GUI,” Int. Conf. Electr. Electron. Signals, Commun. Optim. EESCO 2015, 2015, doi: 10.1109/EESCO.2015.7253906.

[2] M. Qiu, K. Gai, B. Thuraisingham, L. Tao, and H. Zhao, “Proactive usercentric secure data scheme using attribute-based semantic access controls for mobile clouds in financial industry,” Futur. Gener. Comput. Syst., 2016.

[3] W. Ahmad, J. Sunshine, C. Kastner, and A. Wynne, “Enforcing Fine-Grained Security and Privacy Policies in an Ecosystem within an Ecosystem,” 3rd Int. Work. Mob. Dev. Lifecycle, pp. 28 – 34, 2015.

[4] T. M. Mbelli and B. Dwolatzky, “Cyber Security, a Threat to Cyber Banking in South Africa: An Approach to Network and Application Security,” Proc. - 3rd IEEE Int. Conf. Cyber Secur. Cloud Comput. CSCloud 2016 2nd IEEE Int. Conf. Scalable Smart Cloud, SSC 2016, pp. 1–6, 2016, doi: 10.1109/CSCloud.2016.18.

[5] M. F. Alrifaie, Z. H. Ahmed, A. S. Hameed, and M. L. Mutar, “Using Machine Learning Technologies to Classify and Predict Heart Disease,” Int. J. Adv. Comput. Sci. Appl., vol. 12, no. 3, pp. 123–127, 2021, doi: 10.14569/IJACSA.2021.0120315.

[6] M. F. Alrifaie, O. A. Ismael, A. S. Hameed and M. B. Mahmood, "Pedestrian and Objects Detection by Using Learning Complexity-Aware Cascades," 2021 2nd Information Technology To Enhance e-learning and Other Application (IT-ELA), 2021, pp. 12-17, doi: 10.1109/IT-ELA52201.2021.9773589.

[7] B. S. Sagar, S. Niranjan, N. Kashyap, and D. N. Sachin, “Providing cyber security using artificial intelligence - A survey,” Proc. 3rd Int. Conf. Comput. Methodol. Commun. ICCMC 2019, no. Iccmc, pp. 717–720, 2019, doi: 10.1109/ICCMC.2019.8819719.

[8] U. Adhikari, T. Morris, and S. Pan, “Wams cyber-physical test bed for power system, cybersecurity study, and data mining,” IEEE Trans. Smart Grid, vol. 8, no. 6, pp. 2744–2753, 2017.

[9] M. Wu, Z. Song, and Y. B. Moon, “Detecting cyber-physical attacks in cybermanufacturing systems with machine learning methods,” J. Intell. Manuf., pp. 1–13, 2017.

[10] K. V. K. V. Mohana and S. H.N., “Data Security using Genetic Algorithm and Artificial Neural Network,” Int. J. Sci. Eng. Res., vol. 5, no. 2, pp. 543–548, 2014.

[11] R. Mittu and Willliam F. Lawless, “Human Factors in Cybersecurity and the Role for AI,” Found. Auton. Its Threat. From Individ. to Interdepend. Pap. from 2015 AAAI Spring Symp., pp. 39–43, 2015.

[12] A. P. Veiga, “Application of Artificial Intelligence (AI) to Network Security,” ITEC 625 – Inf. Syst. Infrastruct., 2018.

[13] T. Enn, “Conflict, Artificial Intelligence in Cyber Defense,” 2011 3rd Int. Conf. Cyber, 2011.

[14] S. Xiaokui, Y. Danfeng, and R. Naren, “Unearthing Stealthy Program Attacks Buried in Extremely Long Execution Paths,” Proc. 22nd ACM SIGSAC Conf. Comput. Commun. Secur., 2015.

[15] J. Yu, Q. Wan, Q. Liu, X. Chen, and Z. Li, A novel ship detector based on gaussian mixture model and K-means algorithm, vol. 842. Springer International Publishing, 2019. doi: 10.1007/978-3-319-98776-7_72.

[16] C. Chen, “A performance evaluation of machine learning-based streaming spam tweets detection,” IEEE Trans. Comput. Soc. Syst., vol. 2, no. 3, pp. 65–76, 2015.

[17] J. Zhang, Y. Xiang, Y. Wang, W. Zhou, Y. Xiang, and Y. Guan, “Network traffic classification using correlation information,” IEEE Trans. Parallel Distrib. Syst., vol. 20, no. 1, pp. 104–117, 2013.

[18] S. Laazizi, J. Ben Azzouz and A. Jemai, "cybclass: classification approach for cybersecurity in industry 4.0," 2022 IEEE 9th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT), 2022, pp. 378-384, doi: 10.1109/SETIT54465.2022.9875643.

[19] J. C. Acosta, S. Medina, J. Ellis, L. Clarke, V. Rivas and A. Newcomb, "Network Data Curation Toolkit: Cybersecurity Data Collection, Aided-Labeling, and Rule Generation," MILCOM 2021 - 2021 IEEE Military Communications Conference (MILCOM), 2021, pp. 849-854, doi: 10.1109/MILCOM52596.2021.9653049.

[20] C. Latinopoulos, N. Daina and J. W. Polak, "Trust in IoT-enabled mobility services: Predictive analytics and the impact of prediction errors on the quality of service in bike sharing," Living in the Internet of Things: Cybersecurity of the IoT - 2018, 2018, pp. 1-7, doi: 10.1049/cp.2018.0044.

[21] B. Bokan and J. Santos, "Managing Cybersecurity Risk Using Threat Based Methodology for Evaluation of Cybersecurity Architectures," 2021 Systems and Information Engineering Design Symposium (SIEDS), 2021, pp. 1-6, doi: 10.1109/SIEDS52267.2021.9483736.

[22] Xiaokui Shu, Danfeng Yao, and Naren Ramakrishnan. “Unearthing Stealthy Program Attacks Buried in Extremely Long Execution Paths.” Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security, 2015.

[23] Y. Rong and Y. Liu, "Staged text clustering algorithm based on K-means and hierarchical agglomeration clustering," 2020 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA), 2020, pp. 124-127, doi: 10.1109/ICAICA50127.2020.9182394.

[24] A. Nugraha, M. Arista Harum Perdana, H. Agus Santoso, J. Zeniarja, A. Luthfiarta and A. Pertiwi, "Determining The Senior High School Major Using Agglomerative Hierarchial Clustering Algorithm," 2018 International Seminar on Application for Technology of Information and Communication, 2018, pp. 225-228, doi: 10.1109/ISEMANTIC.2018.8549834.

[25] S. H. Shihab, S. Afroge and S. Z. Mishu, "RFM Based Market Segmentation Approach Using Advanced K-means and Agglomerative Clustering: A Comparative Study," 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), 2019, pp. 1-4, doi: 10.1109/ECACE.2019.8679376.

[26] Y. Zhou, A. Rangarajan and P. D. Gader, "A Gaussian mixture model representation of endmember variability for spectral unmixing," 2016 8th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), 2016, pp. 1-5, doi: 10.1109/WHISPERS.2016.8071802.

[27] Z. Lei, H. Yan, C. Liu, M. Ma and Y. Yang, "Two-Path GMM-ResNet and GMM-SENet for ASV Spoofing Detection," ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2022, pp. 6377-6381, doi: 10.1109/ICASSP43922.2022.9746163.