

Utilizing discrete structure and applied algebra to model and analysis network resilience against cyber attack

Shilpa Sharma [‡]

Monika Sharma [†]

Vaibhav Bhavnagar [§]

Rekha Jain ^{*}

Department of Computer Applications

Manipal University Jaipur

Jaipur 303007

Rajasthan

India

Abstract

This paper offers a novel method for modeling and assessing network resilience against cyber-attacks by utilizing discrete structures and algebraic techniques. Utilizing edge and node representations taken from actual intrusion detection datasets build a network graph and examine its structural aspects to identify attack trends and resilience traits. The selection of key characteristics via correlation analysis, supervised machine learning classifiers—Support Vector Machine (SVM), Random Forest (RF), and K-Nearest Neighbors (KNN)—are used to predict and categorize attack types. Confusion matrices and common classification metrics, accuracy, precision-, recall, and F1-score-0.99, are used to assess the models. This research supports SDG 9 by advancing secure and resilient infrastructure through innovative use of algebra and discrete structures. It also aligns with SDG 16 Peace Justice and strong institutions by contributing to the development of secure institutions.

Subject Classification: *Primary 68M250, Secondary 68M99.*

Keywords: *Wavelet cryptography, Adaptive encryption, Cybersecurity, SDG-aligned security, Quantum-resistant algorithms.*

[‡] E-mail: shilpa.sharma@jaipur.manipal.edu

[†] E-mail: monika.sharma23fs30sbs00078@mujaipur.manipal.edu

[§] E-mail: vaibhav.bhatnagar@jaipur.manipal.edu

^{*} E-mail: rekha.jain@jaipur.manipal.edu (Corresponding Author)

1. Introduction

In such a world where big data, social networks, online transfers, information stored or manipulated online, and automated process conducted by IT systems are increasingly becoming common, the element of information security and privacy of data is also continuously under threat. Cyber-crime is on the constant increase in regards to the magnitude of attacks and extent of damage to its targets. The attackers aim to attack the confidentiality, integrity, and availability of the information through the development of new tricks to have unauthorized access to the networks, programs, and information. They are not satisfied attacking individuals only but small and medium-sized enterprises and even great corporate companies. During nearly over two decades, the Internet was a crucial feature of international communications and became increasingly integrated in everyday life of people.

Currently, the Internet counts more than 3 billion of users worldwide due to the progress and low prices in the sphere which greatly contributed to the spread of the Internet, the number of people who use it and its functioning [1]. Each year, the global economy makes billions of dollars in result of the extensive worldwide network the Internet has created. Jasim et al. introduced a novel secant-method for minimizing functions, which can be applied to optimize network configurations for maximum resilience [2].

Cyber-attacks harm essential services and data, and they have become more common in modern computer networks. While detection and prevention are the primary objective of traditional cyber security measures, it's equally critical to understand a network's structural resilience or its capacity to tolerate and bounce back from attacks. In order to model and improve network resilience, this research investigates the application of discrete and algebraic mathematical methods. Discrete structures and algebra provide excellent resources for modeling finite Systems, including communication networks.

The objective is to use these tools to offer formal frameworks that facilitate quantitative evaluation of network resilience and vulnerability. Machine learning is studied as computer programs that use statistical models and algorithms to learn or ferret out inference and patterns in data without being explicitly programmed [3]. Machine learning algorithms are categorized into five groups [4]: supervised, unsupervised, reinforcement, deep learning and deep reinforcement. In supervised machine learning, an algorithm is given a set of labeled datasets to train.

The other attributes have an impact on the target attribute of supervised machine learning.

2. Related Work

The vulnerability analyzes (VA) Traditional activities have grown increasingly challenging due to the development and spread of cyber-physical systems (CPS), which have highlighted additional access methods and unforeseen interdependency amid the cyber and physical infrastructure. VAs tend to be done on physical or cyber infrastructure. Such tests do not consider complex connections that are formed due to interface in systems. Physical protection methods are now used in the study of physical systems e.g. the timeliness of detecting methods. Since cyber tools have been continuously developed to exploit further vulnerabilities in networks, so too has their corresponding methodologies of cyber VA been continuously researched upon and enhanced.

The Author provides a compact list of the criteria that could be used to determine the risk of a particular vulnerability of an infrastructure [5]. These involve determining the plausibility of the threat or assault situation, developing schemes of how interdependent variables in the system may respond and the estimation of attack aftermath. Ezell [6] provides the infrastructure vulnerability assessment model (I-VAM) basing the measure of the vulnerability on the investigation task. I-VAM is a value model that considers some aspects of deterrent, delay, detection and response to analyze the effectiveness of security. Abdalbaqi, Al-Swidi, and Hadi [7] Graph theory, algebraic structures, and topological methods allow for rigorous analysis of network robustness and vulnerability patterns. The technique allows greater fidelity in protective system tests unlike the alternative that existed previously. The final outcome is more knowledge of the associations of the physical infrastructure systems. Evaluation of SCADA (supervisory control and data acquisition) systems is done using attack trees [8].

Ingols et al. [9] extend the attack tree methodology to consider specifically the attacks on network, detection systems and protection systems. The network reachability metric and a list of vulnerabilities are used creating the set of attack graphs in this technique.

The results offer an insight into the success of implementing defense and countering mechanisms to prevent cyber-attacks which are disruptive to such networks. Shi et al. [10] provide the distinction between characteristics of cyber-physical systems. The authors give examples of

ITS, the electric power grid, healthcare and medical technology systems as the examples of CPS. An attack to any of these systems, all of which are examples of the Internet of Things, might have real-life consequences when hackers intrude into a hostile information domain.

3. Proposed Methodology

The Proposed Work Used machine learning classifiers to predict and assess the robustness of IoT networks. To manage heterogeneous data types and standardize input characteristics, the dataset first loaded and preprocessed using feature scaling and encoding. The data is divided into testing and training sets. The classifier of machine learning uses Statistical Properties that are trained based on the flow of traffic in order to classify and locate types of attacks. After model training performance reports in terms of matrices of classification and confusion are provided. To forecast network defense against prospective cyber assaults in actual-time IoT conditions and understand the trends of the classifier, the results are lastly checked.

3.1 Dataset

The data set RTIOT2022 [11] includes 85 detailed characteristics that describe a network flow, consisting of different types of IoT traffic in various applications and protocols, packet lengths, time differences, the number of flags and types of attacks. The RT-IoT 2022 dataset Kaggle sourced includes a supervised and diverse set of real-time industrial Internet of Things network traffic with malicious and benign activity on many protocols including Modbus and MQTT. This is a very important dataset when modeling resilience of networks in both algebraic and discrete structures. Particularly, matrices, graphs, and Boolean functions may be employed to capture the complex flow-based details, such as the duration of the time, the number of flags, and packet information. Such representations allow the determination of strong sub networks, the structural consequences of cyber-attack scenarios and simulation of the same. Whereas algebraic approaches make it possible to abstract and formally verify defense mechanisms, other discrete mathematical tools will assist in modeling node structure and interaction, such as graph theory.

3.2 Method

Figure 1 presents the correlation heat map of the selected top 10 features which are determined in terms of their correlation values on the dataset.

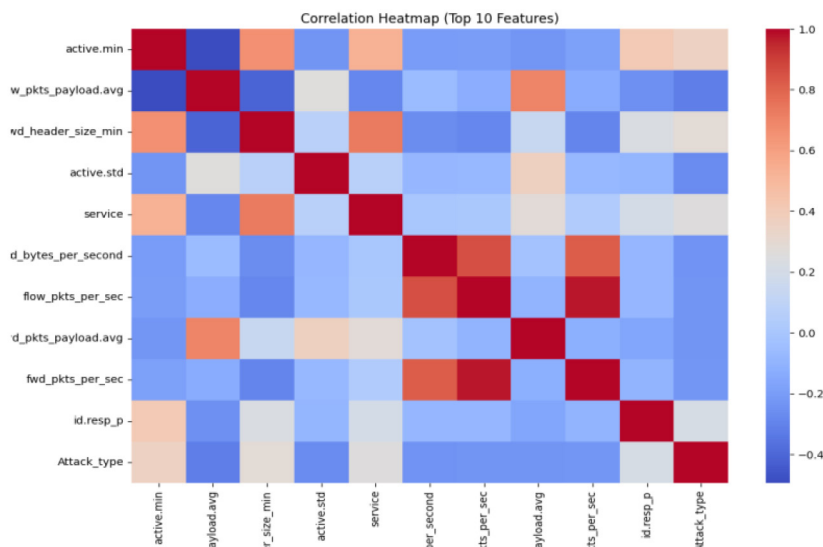


Figure 1
Correlation heatmap

The degree of correlation between two variables is shown by each cell in the heatmap; red denotes a high positive connection, blue a negative correlation, and white little to no association. Notably, there is a positive correlation between features flow pkts per sec, fwd pkts payload. avg, and fwd pkts per sec, indicating that they behave similarly in specific network situations. However, there is a significant negative association between features like fwd pkts payload.avg and active.min. A few attributes also exhibit clear correlations with the Attack type variable, suggesting that these traits are important for identifying or categorizing network attacks. This research aids in determining which features are most useful for predicting and spotting trends in cyber-attacks.

The Author uses discrete structures and algebraic foundations within a machine learning framework to offer a useful method of modeling network resilience against cyber-attacks. Encoding of categorical variables is done with one-hot encoding which is a technique within discrete

mathematics based on set theory and binary representation when assigned the values of the protocol and service type encoded using a labeled dataset of network traffic data and type of attack. The mileage of the code is developed under the machine learning scheme based on discrete structures and algebraic foundations to provide a helpful way of modeling network resilience with respect to cyber-attacks. The target variable that represents types of assaults is label-encoded using algebraic structures such as mappings and relations in order to transform symbolic classes into numerical classes.

4. Results and Discussion

To evaluate to what extent various machine learning classifiers are able to identify cyber-attacks in industrial IoT deployed within an industrial IoT setting, we tested four popular models on the RT-IoT 2022 dataset: Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Random Forest (RF), and a Multi-layer Perceptron Neural Network. To train each of the classifiers correctly and be able to predict the appropriate type of attack, the data of normalized and encoded network traffic was used. The categorization report returns such valuable measures as precision, recall, and F1-score and indicates the competence levels of the algorithm on a wide selection of benign-traffic and cyber-attack classes. These are estimates of the ability of the model to differentiate and accurately detect various types of attacks as well as the performance of the model of identifying various type of attacks an application of discrete structures such as discrete sets and logical classification, which are important measures of the model performance in terms of its robustness.

A Summary of results as determined using each model is given in Table 1.

RF classifier performed the best and got the best values in all measures of evaluation compared to the others. It is a good performance that can be attributed to its ensemble character and thus being capable of processing complicated high-dimensional patterns in the examined network traffic data. A deep learning model with several hidden layers, MLP, showed the good results as well, performing closely to Random Forest in all measures. It can be used to model non-linear relationships, thus having the capability to distinguish slight differences in the type of attacks.

The SVM classifier performed on a moderate level. Although its accuracy and precision were satisfactory, its recall was a bit lower, which means that it was not as efficient at correctly identifying all examples of

each attack type - perhaps, because the dataset was multiclass in nature, and feature distributions overlapped. The models showed KNN with the lowest scores. Being a distance-based learner, KNN is less effective in the high-dimensional space and could be prone to the curse of dimensionality, which probably was one of the reasons of its decreased performance.

Table 1
Classifier Results

Model	Accuracy	Precision	Recall	F1-Score
SVM	0.97	0.9732	0.98	0.89
KNN	0.96	0.9654	0.95	0.96
RF	0.99	0.9823	0.97	0.90
Neural Network	0.95	0.9533	0.96	0.95

Table 1 findings indicate that Random Forest and MLP neural networks are the most suitable choices for deploying intrusion detection systems (IDS) in real-time IoT environments. Their ability to deliver high accuracy along with balanced precision and recall ensures they not only identify attacks effectively but also minimize false positives and negatives - which is critical in security-sensitive domains.

Confusion matrices can be instrumental to determine the practicability and stability of classification models. Confusion matrices provide a methodological and clear representation of the performance of each model in categorizing the various forms of network attacks.

The confusion matrices in Figure 2 show the performance of the different classification models where Random forest had the best accuracy (100%), and the fewest misclassifications. Conversely, SVM and KNN classifiers were slightly confused especially in the classes that have similar names such as Thing_Speak and ARP_poisoning. In general, Random Forest proves the best and accurate classification of all models tested.

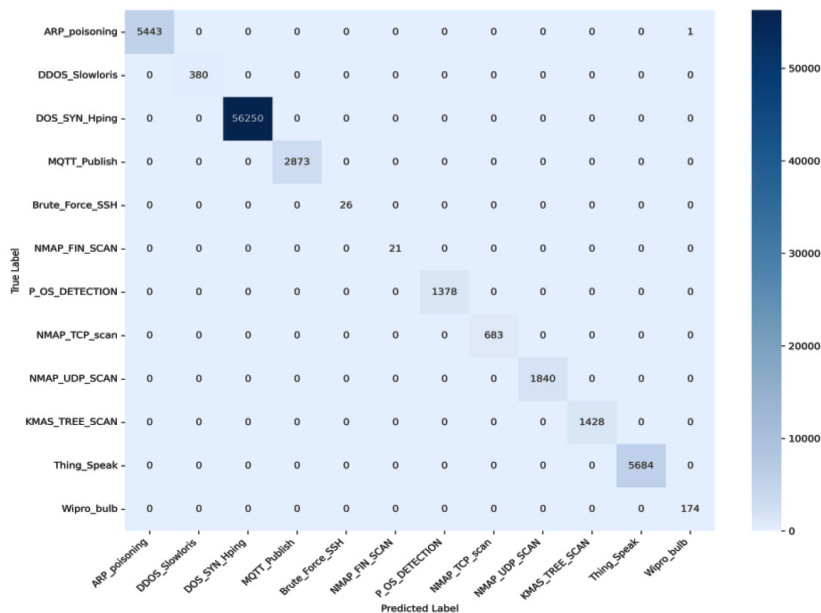


Figure 2
Confusion matrix

5. Conclusion

It is the performance of the four supervised machine learning algorithms (Support Vector Machine (SVM) with 97%, K-Nearest Neighbors (KNN) 0.96%, Random Forest (RF) with 0.99% and neural network with 0.95% to detect and classify network attacks in the RT-IOT2022 data set that was evaluated. The preprocessing of the data was performed by using one-hot encoding of categorical variables and label encoding of attack types targets. The classifiers were also trained by the 70/30 train /test ratio and evaluated based on the results on the training and test accuracy and confusion matrices which gave much information about the classification performance of each model.

References

[1] L. Ma, Y. Zhang, C. Yang, and L. Zhou, "Security control for two-time-scale cyber physical systems with multiple transmission channels under DoS attacks: The input-to-state stability," *Journal of the Franklin Institute*, vol. 358, no. 12, pp. 6309–6325 (2021).

- [2] A. M. Jasim, Y. J. Subhi, and B. A. Hassan, "On new secant-method for minimum functions of one variable," *Journal of Interdisciplinary Mathematics*, vol. 28, no. 1, pp. 291–296 (2025).
- [3] V. Bhatnagar and R. C. Poonia, "Comparative study of supervised machine learning algorithms for healthcare dataset using Orange," in Congress on *Intelligent Systems*, Springer, pp. 195–208 (2020).
- [4] T. O. Ayodele, "Types of machine learning algorithms," *New Advances in Machine Learning*, vol. 3, no. 5–1, pp. 19–48 (2010).
- [5] Y. Y. Haimes, "On the definition of vulnerabilities in measuring risks to infrastructures," *Risk Analysis: An International Journal*, vol. 26, no. 2 (2006).
- [6] B. C. Ezell, "Infrastructure vulnerability assessment model (I-VAM)," *Risk Analysis: An International Journal*, vol. 27, no. 3, pp. 571–583 (2007).
- [7] L. S. Abdalbaqi, L. A. Al-Swidi, and M. H. Hadi, "On condensed set in ideal topological spaces," *Journal of Interdisciplinary Mathematics*, vol. 25, no. 5, pp. 1445–1451 (2022).
- [8] E. J. Byres, M. Franz, and D. Miller, "The use of attack trees in assessing vulnerabilities in SCADA systems," in Proceedings of the International Infrastructure Survivability Workshop, vol. 202, Citeseer (2004).
- [9] K. Ingols, M. Chu, R. Lippmann, S. Webster, and S. Boyer, "Modeling modern network attacks and countermeasures using attack graphs," in 2009 Annual Computer Security Applications Conference, *IEEE*, pp. 117–126 (2009).
- [10] J. Shi, J. Wan, H. Yan, and H. Suo, "A survey of cyber-physical systems," in 2011 International Conference on Wireless Communications and Signal Processing (WCSP), *IEEE*, pp. 1–6 (2011).
- [11] A. Abdo, "RT-IoT 2022: Real-time IoT network dataset for cybersecurity research," Kaggle (2022). [Online]. Available: <https://www.kaggle.com/code/adhamabdo/rt-iot-2022> [Accessed: Jun. 1, 2025].

Received May, 2025