An AI Based Study of IOT Cyber Attacks

Chandan Kumar , Research Scholar, Department of Computer Science & Engineering, Kalinga University Raipur (C.G.)

Dr. Nidhi Mishra, Department of Computer Science & Engineering, Kalinga University Raipur (C.G.)

Dr. Brijesh Kumar Bharadwaj, Associate Professor MCA, Dr. Rammanohar Lohia Avadh University Ayodhya U.P.

Abstract

The number of Internet of Things (IoT) devices is increasing day by day. This growing of IoT devices involves a big challenge in the field of security, especially for network and telecom operators, IoT service providers and also for the users. To detect and prevent these types of attack in IoT devices that use the mobile network, it's needed to have a proper overview of the existing threats and vulnerabilities. The main prospective of this paper is to present and compare different machine learning algorithms. Supervised machine learning classification methods are used in this study, where five machine learning algorithms are tested and evaluated by their performance.

Index terms Internet of Things, Threats and Vulnerabilities, machine learning classification methods

I Introduction

Concerns over security and privacy regarding computer networks are increasing in the world, and computer security has become a requirement as a result of the spread of information technology in daily life. The raise in the amount of Internet applications and the appearance of modern technologies such as the Internet of Things (IoT) are followed with new and recent efforts to invade computer networks and systems[1]. The Internet of Things (IoT) is a set of interrelated devices where the devices have the ability to connect without the need for human intervention. With IoT, many things that have sensors (such as coffee makers, lights, bicycles, and many others) in areas like healthcare, farming, transportation, etc. can connect to the Internet. By saving time and resources, IoT applications are changing our work and lives. It also has unlimited advantages and opens numerous opportunities for the exchange of knowledge, innovation, and growth. Every security threat within the Internet exists within the IoT as well because the Internet is the core and center of the IoT[2]. Compared to other traditional networks, IoT nodes have low capacity and limited resources, and do not have manual controls. Also, the rapid growth and broad daily life adoption of IoT devices makes IoT security issues very troublesome, raising the need to develop security solutions based on networks. While current systems perform well in identifying some attacks, it is still challenging to detect others. As network attacks grow, along with a massive increase in the amount of information present in networks, faster and more effective methods of

detection of attacks are required and there is no doubt that there is scope for more progressive methods to improve network security. In this context, in order to provide embedded intelligence in the IoT environment, we can consider Machine Learning (ML) as one of the most effective computational models. Machine learning approaches have been used for different network security tasks such as network traffic analysis, intrusion detection, and botnet detection. Many researchers have examined the risks imposed by the Internet of Things (IoT) devices on big companies and smart towns. Due to the high adoption of IoT, their character, inherent mobility, and standardization limitations, smart mechanisms, capable of automatically detecting suspicious movement on IoT devices connected to the local networks are needed. With the increase of IoT devices connected through internet, the capacity of web traffic increased. Due to this change, attack detection through common methods and old data processing techniques is now obsolete. Detection of attacks in IoT and detecting malicious traffic in the early stages is a very challenging problem due to the increase in the size of network traffic. The Internet of Things (IoT) integrates billions of smart devices that can communicate with one another with minimal human intervention[3]. It is one of the fastest developing fields in the history of computing, with an estimated 50 billion devices by the end of 2020. On the one hand, IoT technologies play a crucial role in enhancing several real-life smart applications that can improve life quality. On the other hand, the crosscutting nature of IoT systems and the multidisciplinary components involved in the deployment of such systems have introduced new security challenges. Implementing security measures, such as encryption, authentication, access control, network security and application security, for IoT devices and their inherent vulnerabilities is ineffective.

II Internet of Things and Threats

IOT refers to a large number of heterogeneous sensing devices communicating with each other, either in a LAN or over the Internet. IoT threats are different from conventional networks, significantly due to the available resources of end devices. IoT devices have limited memory and computational power, whereas the conventional Internet comprises powerful servers and computers with plentiful resources. Due to this, a traditional network can be secured by multifactor security layers and complex protocols, which is what a real-time IoT system cannot afford. In contrast to traditional networks, IoT devices use less secure wireless communication media. Lastly, due to application-specific functionality and lack of common OS, IoT devices have different data contents and formats, making it challenging to develop a standard security protocol. All these limitations make IoT prone to multiple security and privacy threats, thus opening venues for various types of attacks. The probability of an attack in a network increase with the network size. Therefore, the IoT network has more vulnerabilities than a traditional network, for example, a company office. Additionally, IoT devices communicating with each other are usually multi-vendor devices with different standards and protocols. The communication between such devices is a challenge, which requires a trusted third party to act as a bridge. Moreover, several studies have raised the concern of regular software updates to billions of smart devices. The computational resources of an IoT device are limited, so the capabilities of dealing with advanced

threats are degraded. To summarize, IoT vulnerabilities can be categorized as specific and common. For example, vulnerabilities like battery-drainage attack, standardization, and lack of trust are specific to IoT devices, and Internet-inherited vulnerabilities can be regarded as common vulnerabilities[4]. Several IoT threats and their categorization have been introduced in the past. We discuss the most common threats in IoT reported in the past decade and attempt to classify them into security and privacy categories.

Figure 1: Security threats

III Cyber Security

In recent days, the demand for cyber security and protection against various types of cyber-attacks has been ever increasing. The main reason is the popularity of Internet-of-Things (IoT), the tremendous growth of computer networks, and the huge number of relevant applications that are used by individuals or groups for the purpose of either personal or commercial use. Cyber-attacks such as the denial-of-service (DoS) attack, computer malware, or unauthorized access led to irreparable damage and financial losses in large-scale networks. For instance, according to experts, in May 2017, one ransomware virus brought huge losses to many organizations and industries, including finance, medical care, energy, and universities as well, causing a loss of 8 billion dollars[5].

A cyber security system typically consists of a network security system and a computer security system (Chakraborty, 2017). Although various systems, such as firewall and encryption, are designed to handle Internet-based cyber-attacks, an intrusion detection system (IDS) is more capable of resisting the computer network from external attacks. Thus, the main purpose of an IDS is to detect various types of malicious network communications and computer systems usage for prevention. The conventional solutions, e.g., firewalls, are unable to perform the tasks well. An IDS conducts the process of identifying malicious cyber-attack behavior on a network while monitoring and evaluating the daily activities in a network or computer system to detect security risks or threats such as denial-of-service (DoS)[6].

An intrusion detection system also helps to discover, determine and identify unauthorized system behavior such as unauthorized access, or modification and destruction. Therefore, detecting various types of cyber-attacks and anomalies in a network and to build an effective IDS that performs an essential role in today's network security is needed to facilitate a system's security.

Intrusion detection systems could be different categories according to the usage scope. For instance, the most common types of intrusion detection systems are host-based and network-based, which are in the scope of single computers to large networks. A host-based intrusion detection system (HIDS) relies on an individual system and monitors important operating system files for suspicious or malicious activities, which is very limited to detect unknown malicious code.

An intrusion detection system is typically used to identify malicious cyber-attack behavior on a network while monitoring and evaluating the daily activities in a network or computer system to detect security risks or threats. A number of researches has been conducted in the area of cyber security with the capability of detecting and preventing cyber-attacks or intrusions. Signature-based network intrusion detection is one of the well-known systems used in the cyber industry. This system takes into account a known signature and has seen widespread adoption including commercial success in recent time[7]. On the other hand, the anomaly-based approach has advantage over the signature-based approach for detecting unseen attacks, including the ability to identify unknown or zero-day attacks. This approach monitors network traffic and finds attack behavioral patterns by analyzing the relevant security data. Various data mining and machine learning techniques are used to analyze such security incident patterns for making useful decisions. The main drawback of the anomaly-based approach is that it may produce high false alarm rates as it may categorize the previously unseen system behaviors as anomalies. Thus, limiting the false positive rates of an intrusion detection system must be a top priority. Therefore, a machine learning-based effective detection approach is needed to minimize these issues.

IV IoT and Artificial Intelligence

According to experts, the enormous and bulk presence of IoT devices has brought a new dimension and paradigm shift in the computing world. The scenario of interconnected devices at every household is very likely at an alarming rate, and the needs for having a more reliable cyber security infrastructure to handle and mitigate the risk against the data at rest, data in use, and data in motion, has been one of the major critical security needs. This research provides a detailed review of ML algorithms employed to protect IoT applications from security and privacy attacks. Based on the review, we highlight that a combination of ML algorithms can offer more effective solutions to security and privacy challenges in the IoT environment. To the best of our knowledge, this is the first innovative mechanism that presents a review of security and privacy vulnerabilities in the IoT environment and their countermeasures based on ML algorithms. Machine learning is a sub-field of artificial intelligence that aims to empower systems with the ability to use data to learn and improve without being explicitly programmed**.** It relies on mathematical models derived from analyzing patterns in datasets, which are then used to make predictions on new input data[8].

Applications of machine learning span across a vast set of domains including e-commerce, where machine learning applications are used to make recommendations based on customer behavior and preferences, and health care, where machine learning is used to predict epidemics or the likelihood of a patient having certain

diseases, such as cancer, based on their medical records**.** Machine learning algorithms can be categorized as Predictive (Supervised Learning) or Pattern Discovery (Unsupervised Learning).

In supervised learning, there is always a target variable, the value of which the machine learning model learns to predict using different learning algorithms e.g., based on an IP address location, frequencies of Web requests and times of request, a machine learning model can predict if a given IP address was part of a Distributed Denial of Service (DDOS) attack[9].

V Model Development

A number of experiments were performed to evaluate various aspects of the implementation, which are shown and described here. For all these experiments, training data and validation data were always distinct halves of the same dataset. Each experiment graph herein consists of three displays: the overall accuracy, the false positive percentage and the true positive percentage.

In figure 2 show the model's ability to correctly classify readings, to not generate more alarms than necessary and its ability to detect attacks, respectively. Each of the following experiments generates a collection of twenty resulting data points. These data points are generated with the same values excepting the examined value, and are based on eleven distinct training sessions each. In total, an experiment is thus the result of 24862 instances distinct runs.

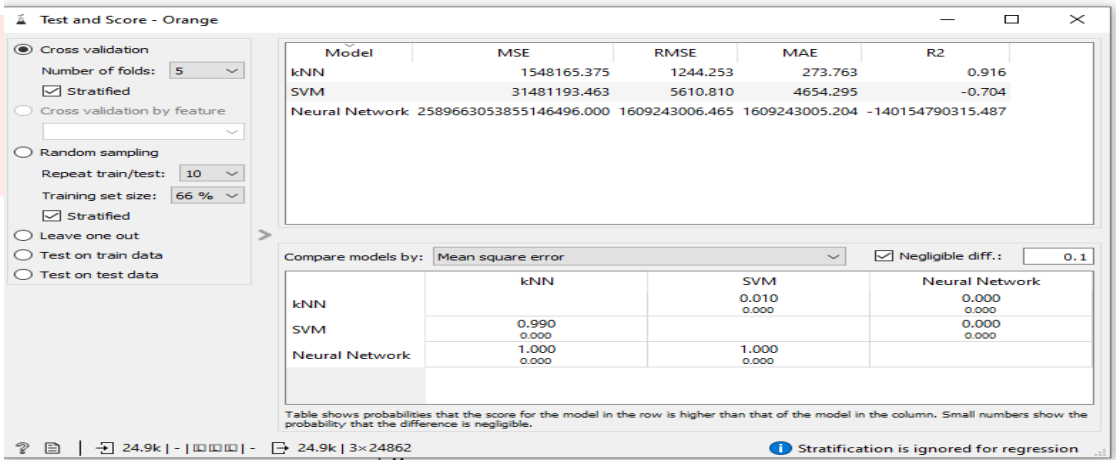

Figure 2: Mean Square Error

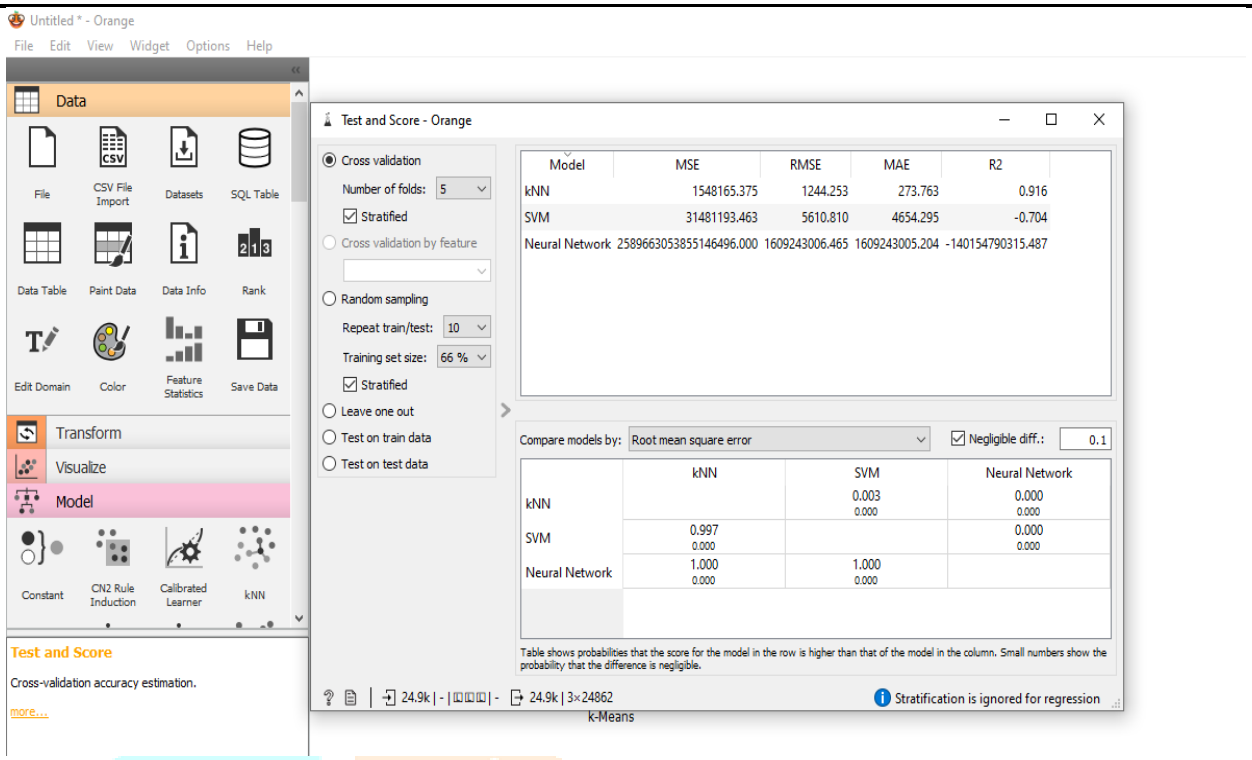

Figure 3: RMSE Impact

As a reminder, only 5 were fold data, in order to reduce the time the experiments took to run and to limit the extent to which bad learning data could enter the data set. One experiment examined the relationship between the ratios of attack and no-attack data in terms of the model's accuracy important background information. Crude expectations would be for true positive ratings to benefit from more attack data, whereas false positives benefit from more non-attack data inaccurate when less than 10% of their data is included. The ratings stabilize surprisingly quickly, whereas the changes in true positive ratings turned out to be more gradual, better matching the naive assumption of a linear relationship. This may hint at the conclusion that the best ratio is one where true- and false positive data occur in the same quantity.

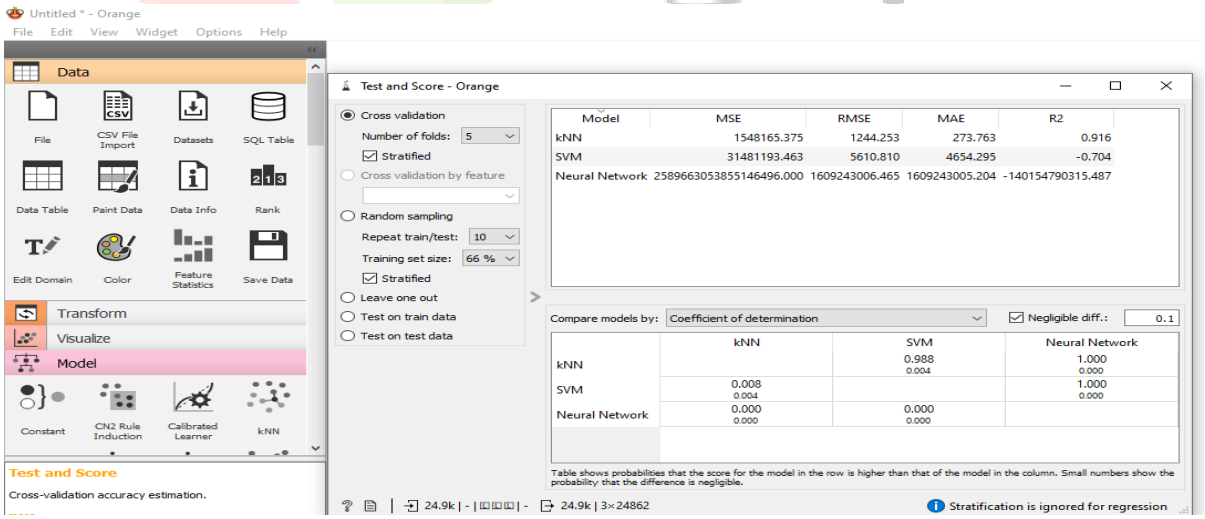

Figure 4: Coefficient of Determinant

However, it is much worse to get a false positive than a false negative. Many false negatives will cause attacks to be less reliably detected, but more false positives will reduce trust in the system as a whole as an administrator will be falsely alerted of attacks. If attacks do not occur often, this is likely to reduce their trust in the detection system and may cause them to ignore true attack alerts entirely. At equal ratios, true positive and true negative readings both reach 90%.

This paper has aimed to detect IoT network attacks by using machine learning methods. In this context, the backdoor attack data (Kaggle) was used as a dataset because of its regular updates, wide attack diversity, and various network protocols. During the implementation, the importance of weight calculations was made with the algorithm to decide which of the features would be used in the machine learning methods. Due paper limitation, we cannot show all figures and tables of quantification. Three approaches were used when making these calculations. In the first approach, the importance weights were calculated separately for each attack type, and in the second approach, all the attacks were collected in a single group and the importance weights for this group were calculated; i.e., the common properties that were important for all attacks were determined. Finally, three machine learning algorithms which are widely used and have different qualities were applied to the data.

References

1. B. I. P. Rubinstein, B. Nelson, L. Huang, A. D. Joseph, S. h. Lau, S. Rao, N. Taft, and J. D. Tygar. "Antidote: understanding and defending against poisoning of anomaly detectors". In Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference, IMC '09, pages 1–14, New York, NY, USA, 2009. ACM.
2. B. Biggio, B. Nelson, and P. Laskov. "Poisoning attacks against support vector machines". In J. Langford and J. Pineau, editors, 29th Int'l Conf. on Machine Learning. Omnipress, 2012.
3. R. N. Rodrigues, L. L. Ling, and V. Govindaraju. "Robustness of multimodal biometric fusion methods against spoof attacks". J. Vis. Lang. Comput., 20(3):169–179, 2009.
4. Apruzzese, G., Ferretti, L., Marchetti, M., Colajanni, M., & Guido, A. (2018). On the Effectiveness of Machine and Deep Learning for Cyber Security, Conference paper, 1-17.
5. Pierazzi, F. Apruzzese, G. Colajanni, M. Guido, A. & Marchetti, M. (2017). Scalable architecture for online prioritization of cyber threats. International Conference on Cyber Conflict (CyCon), 1-23.
6. Chakraborty, T., Pierazzi, F., & Subrahmanian, V. (2017). Ec2: Ensemble clustering and classification for predicting android malware families. IEEE Transactions on Dependable and Secure Computing, 1-17.
7. Ucci, D., Aniello, L., & Baldoni, R. (2018). Survey of machine learning techniquesfor malware analysis. arXiv:1710.08189v3 [cs.CR], 1-56.
8. Zhao, H., Li, M., Wu, T., & Yang, F. (2018). Evaluation of supervised machine learning techniques for dynamic malware detection. International Journal of Computational Intelligence Systems, 11, 1153-1169.
9. Gibert, D., Mateu, C., & Planes, J. (2020). The rise of machine learning for detection and classification of malware: Research developments, trends and challenges. Journal of Network and Computer Applications, 156, 1-22.