Towards Development of Machine Learning Framework for Enhancing Security in Internet of Things

Mutyalaiah Paricherla, Sallagundla Babu, Khongdet Phasinam, Harikumar Pallathadka, Abu Sarwar Zamani, Vipul Narayan, Surendra Kumar Shukla, and Hussien Sobahi Mohammed

1N.B.K.R. Institute of Science & Technology, Vidyaganj, Andhra Pradesh, India
2CSE Department, V.R Siddhartha Engineering College, Vijayawada, India
3Faculty of Food and Agricultural Technology, Pibulsongkram Rajabhat University, Phitsanulok, Thailand
4Manipur International University, Imphal, Manipur, India
5Department of Computer and Self Development, Preparatory Year Deanship, Prince Sattam Bin Abdulaziz University, Al-Kharj, Saudi Arabia
6Department of Computer Science and Engineering, Madan Mohan Malaviya University of Technology, Gorakhpur, India
7Department of Computer Science & Engineering, Graphic Era Deemed to Be University, Dehradun, Uttarakhand, India
8University of Gezira, Wad Medani, Sudan

Correspondence should be addressed to Hussien Sobahi Mohammed; hussiensobahi@uofg.edu.sd

Received 19 March 2022; Revised 12 April 2022; Accepted 23 April 2022; Published 17 May 2022

An IoT system is a smart network that connects all items to the Internet and exchanges data using Internet Engineering Task Force established protocols. As a consequence, everything is instantly accessible from any place and at any time. The Internet of Things (IoT) network is built on the backbone of tiny sensors embedded in common objects. There is no need for human intervention in the interactions of IoT devices. The Internet of Things (IoT) security risk cannot be ignored. Untrusted networks, such as the Internet, are utilized to provide remote access to IoT devices. As a result, IoT systems are susceptible to a broad range of harmful activities, including cyberattacks. If security problems are not addressed, critical information may be hacked at any time. This article describes a feature selection and machine learning-based paradigm for improving security in the Internet of Things. Because network data are inherently abundant, it must be reduced in size before processing. Dimension reduction is the process of constructing a subset of an original data collection that removes superfluous content from the essential data set. Dimension reduction is a data mining approach. To minimize the number of dimensions in a dataset, linear discriminant analysis (LDA) is used. Following that, the data set with fewer dimensions is put into machine learning predictors as a training set. The effectiveness of machine learning approaches has been assessed using a range of criteria.

1. Introduction

A massive quantity of data is being generated by the Internet of Things (IoT) [1] as it expands fast. There are a lot of sensors, such as those that measure the quality of the air, the temperature, the GPS, the pressures, and the motions. Massive volumes of data may be stored on a variety of devices. Data analysis technologies that operate in real time were required because of the enormous amount of data. IoT-generated data overwhelm current big data platforms at every step.

IoT security involves an awareness of risks, vulnerabilities, and attacks [2]. An advantage might be jeopardised if a future revengeful act occurs. You can see the danger because of its flaws. This might be attributed to poor design, incorrect setups, or inadequate and inaccurate coding techniques.
Taking advantage of vulnerability is referred to as an "attack." For example, an attack can include submitting a malicious program contribution or flooding a device with data in an effort to thwart help.

An IoT system is a smart network that links all objects to the Internet and exchanges information using protocols established by the Internet Engineering Task Force [3]. As a result, everything may be accessed at any time and from any location. Tiny sensors embedded in everyday items provide the backbone of an Internet of Things (IoT) network. There is no need for human involvement in the IoT devices’ interactions. IoT minimizes human effort, optimizes resource utilization, provides live tracking and monitoring, enhances data collection, and saves time. Wearable, smart farming, smart retail, and health monitoring are just a few examples of how the Internet of Things (IoT) might be used. New apps and services may be built using the IoT’s unique addressing techniques for interaction with other objects and things. Every day, new questions about the Internet of Things (IoT) arise. The Internet of Things (IoT) security concern is one that cannot be disregarded among many others. Untrusted networks, such as the Internet, are used to access IoT devices from anywhere. Consequently, IoT systems are vulnerable to a wide variety of malicious assaults, making them vulnerable to cyberattacks. The sensitive information might be hacked at any moment if security concerns are not addressed [4].

Using machine learning, it is possible to gain knowledge from previously accessible malicious records and to apply that knowledge effectively in order to behave more like a human brain when classifying future data pertaining to malware and intrusion. Malware data from the past are utilized to train machine learning algorithms in the present day and future. Machine learning algorithms become capable of accurately predicting future malware on the basis of what they have learned from prior data [5, 6].

Computer systems are monitored by IDS, which carries out postattack forensics. Check your network resources to see whether any intrusions or assaults have evaded any preventative measures you have put in place (firewall, router packet filtering, and proxy server). A system’s confidentiality, integrity, or availability has been compromised by an intrusion. True intruder detectors may be thought of as an approximate analogue to intrusion detection systems.

This article presents a paradigm for enhancing security in the Internet of Things that is based on feature selection and machine learning techniques. Because network data are inherently voluminous, it is necessary to reduce its dimensions before processing it. Dimension reduction is the process of creating a subset of an original data set in such a way that the extraneous material is removed from the fundamental data set. Dimension reduction is a technique used in data mining. Linear discriminant analysis (LDA) is used to reduce the number of dimensions in a dataset. After that, the data set with a decreased number of dimensions is fed into machine learning predictors as a training set. The effectiveness of machine learning approaches has been evaluated in terms of a variety of different metrics.

1.1. Literature Survey. Strazdins and Wang [7] cite security and privacy as the two most pressing concerns with cloud computing. The debate on these available positions brought up a number of key points. Authentication and authorization, key distribution and administration, data storage and secure processing, safe data transit, and protection against denial-of-service attacks were all taken into consideration while dealing with security concerns. Additionally, they addressed the privacy of passive users, privacy alternatives, identity management, and commercial needs while addressing privacy problems.

Weber and Boban [8] examined the security of IoT deployments. As the authors point out, there are a number of significant obstacles that must be overcome before widespread adoption of the Internet of Things can take place. Confidentiality and security are among the many challenges that must be addressed to preserve the integrity and confidentiality of data in a diverse environment.

Security threats and problems were examined at every stage of the IoT architecture [9]. It is the work of Farooq and others that introduces prioritizing data privacy and security in the design of their IoT infrastructure (IoT). The Internet of Things (IoT) security problems and limitations are addressed in a new categorization system. This category was created with the main goal of ensuring the security of IoT data. Classifying the assaults was done using a four-layer classification system: physical, network software, and encryption.

Sathish et al. [10] examined the current state of IoT security and the flaws they found. Some of the constraints have been alleviated by the implementation of a security architecture. The IoT’s susceptibility was assessed using the Threat Index (TI), which took a variety of factors from the IoT ecosystem into account. This TI and the index threshold may be used by the IoT provider to estimate the current level of security.

Sacira et al. [11] argue that IoT privacy and security issues are critical because of the wide range of technologies that make up the Internet of Things (IoT). Everyone who participates in the Internet of Things (IoT) must adhere to security and privacy rules and regulations.

Gupta and Shukla were particularly concerned about the security of IoT devices. [12]. The authors investigated the Internet of Things’ plethora of unresolved security and privacy vulnerabilities (IoT). It is imperative that all IoT infrastructure, applications, and backdoors are protected by these measures. In order to assure secrecy, integrity, and authentication, the authors explored every security approach available.

However, the proposed training approach by Ashok Kumar and Venugopalan [13] is capable of adapting/learning and detecting new assaults. Combining this technique with feature weights may still enhance the algorithm’s performance metrics. Despite its simplicity, the method has a lot of room for improvement in terms of parallelization.

AdaBoost and simple feed-forward neural network were two novel strategies introduced by Zhang et al. [14] to differentiate the strangeness of the system’s execution in terms of fractional characterization. For each approach,
Simulated data are first analyzed in order to determine how sensitive it is to the length and magnitude of the anomaly’s occurrence. There were no simulated anomalies missed by the accelerated decision tree technique, which proved to be excessively rapid (4 seconds per hour tested). After the anomaly has been identified, the order of the series list is restored. This is an extra benefit that the required amount of positive or false impact may be determined with an acceptable SSC limit. There was no hyperparameter optimization of the neural network model utilized, and the one network tested was less sensitive than the other to brief and low amplitude changes.

Sethi et al. [15] created a novel framework for intelligent malware analysis based on behavioral similarities for dynamic and static analysis of malware models. The proposed approaches for identifying and categorizing harmful files using Weka machine learning models have been tested and exhibit acceptable results. For accuracy and precision, decision tree J48 has been proven to be the most effective. There were only 220 examples of files to be studied, which may have been distorted since not all functionalities could have been included with this amount of files.

A variety of external and internal assaults are feasible on the network, according to Kaur et al. [16]. Internal assaults, on the other hand, are much more harmful. It is mostly an issue of internal assaults on the local wireless network and the local wired network, which are both vulnerable. IDS/IPS (intrusion detection or prevention system) signature-based technologies are currently available; however, they are insufficient because of the high incidence of false alerts. Then, using Wireshark, signature-based devices (Snort and Kismet), and machine learning methods is able to detect these assaults (WEKA) [17].

Burnap et al. [18] used machine action measures to identify toxic and trustworthy small executable programming tests. These attacks were inspired by the emergence of cyberattacks using technology previously used to transmit advanced persistent threats (APTs). APTs are becoming more complicated and ready to obscure many of their distinctive features via encryption, bespoke codebases, and in-memory execution. Using machine learning may achieve a high degree of accuracy in distinguishing between malicious and trustworthy samples, based on the specific imprint left on a PC framework during execution. With the help of these three components, you may monitor and manage data transfer at the byte and parcel level.

Hidden object recognition in PMMWIs was the focus of work by López-Tapia et al. [19]. SNR and nonstationary noise fill a picture, which makes this work challenging. This procedure can be done using a basic method; however, it is more efficient when working with high-quality photographs.

The performance of Snort and Suricata was tested at a network speed of 10 Gbps on two identical computer systems. Suricata has a lower packet rejection rate than Snort, but it uses more processing resources. It can manage greater network traffic speeds than Snort. Snort was chosen for additional testing because of its high level of detection accuracy. Snort was found to have generated a substantial number of erroneous alerts. An adaptive Snort plug-in was created to tackle this issue. In order to improve recognition accuracy, a mix of SVM and fuzzy logic was used. A firefly method with an RBF-optimized SVM yielded the greatest results [20].

2. Methodology

When it comes to the Internet of Things, security is a huge problem. IoT systems are vulnerable to a wide range of destructive actions, including cyberattacks, because they are interconnected. It is possible that sensitive information will be compromised at any time if security issues are not addressed. This section presents a methodology for strengthening security in the Internet of Things that is based on feature selection and machine learning techniques. Because network data are inherently voluminous, it is necessary to reduce its dimensions before processing it. Dimension reduction is the process of creating a subset of an original data set in such a way that the extraneous material is removed from the fundamental data set. Dimension reduction is a technique used in data mining. Linear discriminant analysis (LDA) is used to reduce the number of dimensions in a dataset. After that, the data set with a decreased number of dimensions is fed into machine learning predictors as a training set. Figure 1 shows the machine learning-based framework for enhancing security in the Internet of Things.

Using LDAs, dimensional reduction preprocessing is performed before the use of machine learning and model classification. Essentially, the purpose of this method is to condense a huge amount of data into a small area while keeping a high degree of class separation, in order to reduce the amount of time and effort required to calculate the results. With the use of LDA, we may examine the differences in performance between distinct members of a class whose frequencies are not exactly equal. The LDA technique maximises the relationship between a record’s class variance and the class variance of the records that follow it. As a result, it ensures the highest amount of isolation possible.

A powerful method for handling machine learning challenges, APSO-SVM (adaptive particle swarm optimization-support vector machine), has a strong mathematical foundation and may be used to solve many problems. In order to build its response from the aspect of division, SVM is trained with training input. SVM applications include classification, regression, novelty detection, and feature reduction, all of which fall under the SVM umbrella. One can learn a great deal about SVM by looking at it from a geometric approach. For what it is worth, the generative technique, which is more commonly employed when attempting to identify outliers, shows more potential. The discriminant technique, on the other hand, necessitates the use of fewer training data and computing resources. It is necessary to obtain the equation for a multidimensional plane that best separates the distinct groups in the highlighted area in order to train a classifier in this manner. SVM is recognised as a discriminant method because of the logical approach it takes to tackle the twisted optimization problem.
it was designed to solve. A further advantage of these SVMs is that their hyperplane factor is optimal [21].

Numerous methods are used in pattern recognition to categorize and sort things. K-NN is a classification method that classifies objects based on their proximity to training samples. K-NN is an example of an event-based learning algorithm. Calculations are deferred until after classification when a locally estimated function is used. When there is little previous information about the distribution of the data, the KNN classification technique is the simplest. KNN is a well-known classification approach for pattern recognition. Numerous research on diverse data sets has shown exceptional results when the KNN algorithm is used [22].

The naive Bayes strategy is a straightforward technique for selecting class labels for issue occurrences using feature value vectors while developing classifier models. Rather than depending on a single technique, these classifiers are trained using a range of methodologies based on the same fundamental principle. With just the class variable as a guide, there is no way to determine whether one characteristic is more significant than another [23].

3. Result and Discussion

NSL KDD [24] was the data set utilized in the experiment. The NSL-KDD dataset contains 24 different forms of assaults, and the data are either normal or one of them. Probe, DoS, R2L, and U2R are all attacks that fit into one of these four categories. In this experimental work, the NSL KDD data set is used as an input data set. This contains 125973 instances. Twenty percent of the NSLKDD dataset is testing data (25192 records), and the remaining eighty percent is training data (100781 records).

The result comparison of these three classifiers is shown below in Figure 2.

Accuracy = (TP + TN)/(TP + TN + FP + FN)
Sensitivity = TP/(TP + FN)
Specificity = TN/(TN + FP)

4. Conclusion

In order to function, the Internet of Things (IoT) network requires a backbone of small sensors implanted in everyday objects. In the interactions of IoT devices, there is no requirement for human intervention. No one can overlook the security threat posed by the Internet of Things (IoT). When it comes to remote access to IoT devices, untrusted networks, such as the Internet, are used. A wide range of destructive actions, including cyberattacks, can take advantage of IoT systems as a result of this. It is possible that sensitive information will be compromised at any time if security issues are not addressed. With the help of feature selection and machine learning, this article describes an approach for improving security in the Internet of Things. Given the large amount of data generated by networks, it is necessary to reduce their dimensions. Using dimension reduction, a subset of the original data collection is created that is devoid of any superfluous data. It is necessary to minimize the dimensions using linear discriminant analysis (LDA). A machine learning predictor is then used to make predictions from the data set with fewer dimensions. APSO-SVM is the most accurate algorithm for the classification of malware data related to the Internet of Things.

Data Availability

The data shall be made available on request to the corresponding author.

Conflicts of Interest

The authors declare that they have no conflicts of interest.
References

[1] A. Raghuvanshi and U. K. Singh, "Internet of Things for smart cities: security issues and challenges," Elsevier Materials Today, 2020.

[2] M. P. Lokhande, D. D. Patil, L. V. Patil, and M. Shabaz, "Machine-to-Machine communication for device identification and classification in secure telerobotics surgery," in Security and Communication Networks, C. Chakraborty, Ed., vol. 2021, pp. 1–16, 2021.

[3] T. Gera, J. Singh, A. Mehbodniya, J. L. Webber, M. Shabaz, and D. Thakur, "Dominant feature selection and machine learning-based hybrid approach to analyze android ransomware," in Security and Communication Networks, J. Cui, Ed., vol. 2021, pp. 1–22, 2021.

[4] V. Hemamalini, S. Rajarajeswari, S. Nachiyappan et al., "Food quality inspection and grading using efficient image segmentation and machine learning-based system," Journal of Food Quality, vol. 2022, pp. 1–6, 2022.

[5] A. Raghuvanshi, U. K. Singh, and C. Joshi, "A review of various security and privacy innovations for IoT applications in healthcare," Advanced Healthcare Systems, pp. 43–58, 2022.

[6] V. D. P. Jasti, A. S. Zamani, K. Arumugam et al., "Computational technique based on machine learning and image processing for medical image analysis of breast cancer diagnosis," Security and Communication Networks, vol. 2022, pp. 1–7, 2022.

[7] G. Strazdins and H. Wang, "Open security and privacy challenges for the Internet of Things," 2015 10th International Conference on Information, Communications and Signal Processing (ICICS), pp. 1–4, 2015.

[8] M. Weber and M. Boban, "Security challenges of the internet of things," in 39th International Convention on Information, Electronics and Microelectronics (MIPRO), pp. 638–643, IEEE, 2016.

[9] M. U. Farooq Farooq, M. Waseem, A. Khairi, and S. Mazhar, "A critical analysis on the security concerns of Internet of Things (IoT)," International Journal of Computer Application, vol. 111, no. 7, pp. 1–6, 2015.

[10] A. K. Sathish, V. Tyler, and S. Harshit, "Security in internet of things: challenges, solutions and future directions," in 2016 49th Hawaii International Conference on System Sciences (HICSS), pp. 5772–5781, IEEE, 2016.

[11] S. Sicari, A. Rizzardi, L. A. Grieco, and A. Coen-Porisini, "Security, privacy and trust in internet of things: the road ahead," Computer Networks, vol. 76, pp. 146–164, 2015.

[12] K. Gupta and S. Shukla, "Internet of Things: security challenges for next generation networks," in 2016 International Conference on Innovation and Challenges in Cyber Security (ICICCS-INBUSH), pp. 315–318, IEEE, 2016.

[13] D. Ashok Kumar and S. R. Venugopalan, "A novel algorithm for network anomaly detection using adaptive machine learning," Advances in Intelligent Systems and Computing, vol. 564, pp. 59–69, 2018.

[14] J. Zhang, R. Gardner, and I. Vukotic, "Anomaly detection in wide area network meshes using two machine learning algorithms," Future Generation Computer Systems, vol. 93, pp. 418–426, 2019.

[15] K. Sethi, S. K. Chaudhary, B. K. Tripathy, and P. Bera, "A novel malware analysis framework for malware detection and classification using machine learning approach," ACM International Conference on Distributed Computing and Networking, 2018.

[16] J. Kaur, "Wired LAN and wireless LAN attack detection using signature based and machine learning tools," Networking Communication and Data Knowledge Engineering, vol. 3, pp. 15–24, 2018.

[17] L. Xiao, X. Wan, X. Lu, Y. Zhang, and D. Wu, "IoT security techniques based on machine learning: how do IoT devices use AI to enhance security?" IEEE Signal Processing Magazine, vol. 35, no. 5, pp. 41–49, 2018.

[18] P. Burnap, R. French, F. Turner, and K. Jones, "Malware classification using self organising feature maps and machine activity data," Computers & Security, vol. 73, pp. 399–410, 2018.

[19] S. López-Tapia, R. Molina, and N. Pérez de la Blanca, "Using machine learning to detect and localize concealed objects in passive millimeter-wave images," Engineering Applications of Artificial Intelligence, vol. 67, pp. 81–90, 2018.

[20] B. Gulzar and A. Gupta, "DM: a theoretical framework for SensorSecurity in IoT applications," INTERNATIONAL JOURNAL OF NEXT-GENERATION COMPUTING, vol. 12, no. 3, 2021.

[21] Y. Li, C. Wang, J. Xu, Z. Zhou, J. Xu, and J. Cheng, "Rockburst prediction based on the KPCA-APSO-SVM model and its engineering application," Shock and Vibration, vol. 2021, pp. 1–12, 2021.

[22] A. Gupta and L. K. Awasthi, "Security issues in cross-organizational peer-to-peer applications and some solutions," in Communications in Computer and Information Science, pp. 422–433, Springer Berlin Heidelberg, 2009.

[23] Z. Shen, Y. Zhang, and W. Chen, "A bayesian classification intrusion detection method based on the fusion of PCA and LDA," Security and Communication Networks, vol. 2019, pp. 1–11, 2019.

[24] S. Revathi and A. Malathi, "A detailed analysis on NSL-KDD dataset using various machine learning techniques for intrusion detection," International Journal of Engineering Research and Technology, vol. 2, no. 12, pp. 1848–1853, 2013.