Research Article

Survey of Cyber security approaches for Attack Detection and Prevention

MalathiEswaran[1], S. Hamsanandhini[2], K IlakiyaLakshmi[3]

[1,2,3]Kongu Engineering College, Tamilnadu, India
malathieswaran@gmail.com,hamsanandhini@gmail.com, lucky.30091998@gmail.com

Article History: Received: 11 January 2021; Accepted: 27 February 2021; Published online: 5 April 2021

Abstract: In the world of modern technology many devices are frequently handled by the people via network. Since the network has been utilized in communication across the world and also in data sharing, there may be a chance of cyber-attacks and intruding into the personal data of the user. This survey provides a witness in large amount of cyber-attacks widespread in the recent times. The issue also deals with the system under use and with the storage devices concerned. Inorder to manage large amount of data, cloud computing plays a vital role in managing the data and also prevents data from intruders. Many intrusion detection systems help in detecting anomalies, that caused by various cyber-attacks. This proposed survey focuses on types of attacks and also the methodology involved in detecting such type of attacks.

Keywords: Cyber-attacks, Data sharing, Intruder, Networks.

1. Introduction

Cybersecurity should be given a great concern for managing large number of information that has been shared across the world. Aidin Ferdowsi and Walid Saadr [1] undergone a study in the intrusion detection in the area of Internet of Things (IoT). It provided a generative Adversarial Networks that detects anomalous nature without any particular centralized controller. Internet of Things is really communicable by connecting data from various resources and helps in reliable transmission of data. However an efficient detection system is required for managing and preventing the anomalous activities mainly during data transmission. The architecture helps not only in monitoring the own data but also in the neighbour IoT devices. Ibrahim Alrashdi and Ali Alqazzaz [2] suggested an anomaly detection method using Random Forest algorithm. Smart city in IoT incorporates all the smart systems along with communicable technology inorder to improve the various services that has been encountered within the city. There is necessity of ensuring high level storage for managing data. Hence an architecture governs the aggregation of various data.

- Cloud layer: It maintains a large amount of data and also contains servers in storing data
- Fog layer: It act as a bridge between cloud and IoT sensing layer that helps in managing the network edges.
- IoT sensing layer: It consists of sensors that are enabled within a city for data collection.

Distributed attack has to be detected in fog network rather than in sensing layer which helps in achieving reliable high accuracy of prediction.

Sushmitha R andDeepa N P [3] impost machine learning strategy which could lead to the development of various computer programs that could handle the data sufficiently. The datasets has been collected from Kaggle repository that includes various types of cyberattacks includes back doors, exploits, normal, shell code and warms. The anomaly detection could effectively provide security to various attacks specified and also reduce negative rate in the system. JadelAlsamiri and Khalid Alsubhi [4] provided a strategy by using machine learning algorithms in detecting cyber-attacks in IoT. It describes a method of implementing CICFlowMeter which helps in extracting flow-dependent features from network traffic. Sushmitha et al [5] used an AD-IoT system detection model based on the fog framework, Since the fog layer consist of enormous number of devices which results in cyber-attacks. Random Forest helps in choosing the parameters and is constructed to a forest using number of trees. The number of trees is nothing but the number of variables that are selected from dataset attributes as parameters. Here, the count of source to destination data packets are considered.

Rashid et al [6] suggested a methodology that helps in tracking the traffic traces that passes through each and every fog node. Since the fog architecture is very closer to IoT sensing devices it provides an effective way of identifying attacks in fog layer rather than in cloud layer. The dataset used here is CICIDF2017 datasetand it includes recent cyber-attackssuch as denial of services, SQL injection, Brute force and Infiltration. Hasan et al [7] suggested machine learning technique that serves as a skeleton for deep learning algorithms. The features are extracted and multiplied by random weights further it is added with a bias value. These values serve as an input for non-linear function. This study insist ROC curve it determining the performance of a specified classifier that has been imposed in feature extraction.

Shilpa Bahl and Suthir Kumar Sharma [8] provided a study in improving the rate of detection that uses each and every data features in detecting intrusions. Some of the important features are considered in U2R attack class that helps in increasing the overall accuracy in detection rate.

2. Generative Adversarial Networks

The main goal of GAN is to discover a discriminator without sharing the data such that each discriminator could discriminate like a new point proceeds the distribution phase. Since it understands the distribution of entire data it could easily to detect intrusion on other discriminator. This approach also helps in receiving the loss value from each and every discriminator. Since the GAN is distributed there is no need of central units in detecting the intrusion to the system. The output points are compared to ½. If it is closer to half the device is in normal state and if it is closer to 0 or 1, then the device is under attack. This preserve privacy because, sharing of data is not allowed instead weights are shared among the discriminators. The analytical study proves that IDS could perform as standalone and could fit to the real world. This achieves the lower negative rate and higher accuracy when compared to traditional intrusion detecting systems.

3. AD-IOT Using Machine Learning

It mainly deals with handling attacks in the layers that could lead to intrusion. HIDS (Host-based-IDS) makes the software to be installed on the computer inorder to monitor any malware activities or any intrusion behaviours in connection to local networks. Adversary model has been considered that the attacker provides a method in scanning the internet that results in vulnerability of IoT devices which are connected to different routers. The design mode consists of components such as IoT devices that are connected to fog networks, gateways and IDS system. First the dataset replaced by using Pandas framework that helps in grouping the features. Pre-processing is done to enhance the large amount of network traffic by saving it to HDF5. The features chosen are various traffics that includes DoS, generic, analysis, exploits, attacks, back doors and normal. The vulnerability in IoT devices is considered in public networks. NIDS uses various machine learning algorithm such as DT, KNN, RF, etc. which helps in classifying and finding the malicious nature especially in fog networks. This model could identify normal and abnormal attacks with reduced negative rate. Evaluation metrics such as precision and recall are calculated in ensuring the performance of the proposed model.

4. Machine Learning for Anomaly Detection of Cyber-Attacks

Machine learning helps in exploring the datasets from various locations and make accurate predictions by continuous training phase. The four main categories implied in this study includes – supervised, unsupervised, semi-supervised and reinforcement. In supervised learning, data has been labelled. In unsupervised category, the model is used to identify the features based on the training enforced on them. There will be no data labelling for unsupervised learning. Some functions are done in reinforcement learning inorder to make decisions based on the result obtained. The UNSW-NB15 dataset has been used in this machine learning model to train the implementations done. Random Forest algorithm is done by taking the features into consideration from the dataset provided. A forest is constructed using enormous amount of decision tree framework. After that, data are needed to be categorized based on the feature extraction for classification purpose. The two parameters taken into consideration for Random Forest are:-

- In a dataset, the number of features chosen.
- Number of trees used in building the forest.

Anomaly detection helps in analysing the possibilities of attacks that could happen via data sharing and could easily spot the cyber-attacks. Also, when considering accuracy, it make much better results in analysing the cause of attacks based on the feature chosen in constructing the forest.1

5. Machine Learning Approach in Detecting Cyber-Attacks

The approaches uses different pre-processing methods inorder to detect anomalies by using various machine learning techniques. The data were extracted from CICFlowMeter and flow based factors are drawn from the raw dataset. The data are classified as testing and training set. After that, feature extraction is done by the suitable machine learning algorithm. Here, the chosen dataset is Bot-IoT. K-Nearest Neighbour (KNN) helps in associating recent data points with existing one. Quadratic Discriminant Analysis is used in assigning data to the groups. Here number of groups is lower when compared to the number of samples. ID3 (Iterative Dichotomiser 3 helps in creating the decision tree using the dataset extracted. AdaBoost mainly focuses on the issues that is faced by the classification technique and converts the classifiers which are weaker to the efficient one. Multi-layer Perceptron uses 3 layers namely – input, output and hidden. Naïve Bayes classifier helps in handling the features of traffic network that has been classified independently. The Random regressor is the effective method in reducing the dataset dimensions. More than 80 features of network traffic has been used to train the machine learning model. The evaluation metrics such as Precision, Recall, F-measure and Accuracy were used in the

estimation of performance of the trained models. It has been suggested that if it is multi-layered architecture, the accuracy in detecting the anomaly can be increased.

6. Anomaly Detection of IOT Cyber-Attacks

The Ad-IoT model is implied in detecting the various cyber-attacks. It is mainly based on the fog layer since many of the IoT devices has been connected which resulted in attacks of the IoT layer. The UNSW-NB15 dataset is used as an input. The data are classified under binary classification. Inorder to test the data and classify them as "Normal" and "Attack", random forest is implied. The parameters chosen are reduced to limited threshold values. Here, one of the parameter named spks encountered that is the count of packets from source to destination that has been ranged from 20 to 28 which represents the attack defined as reconnaissance. The features taken into account are protocol, service, state, packet count, source bytes, destination bytes time to live for both source and destination. If the category falls to be Normal, the label has been assigned as 0, else it is for attacks, and then label is assigned as 1. The confusion matrix helps in determining the accuracy of the detection.

7. IOT Based on Ensemble Techniques

The ensemble based Intrusion Detection System has been used in enhancing the accuracy over various classifiers. The features has been extracted using BFS, Genetic and Rank search techniques. The dataset used here is NSL-KDD. The optimal feature selection paves the way for the high degree of accuracy when detecting the anomalies over smart cities. The features such as Normal, Generic, Exploits, Fuzzers, Denial of Service and Backdoor. The ensemble models has devised the fog layer without increasing the system latency. The fog layer has been devised without increasing the system latency. The stacking of the implemented classifiers could easily detect the various cyber-attacks from samples of benign samples.

8. Attacks and Anomaly Detection in IOT Sensors Sites

The data has been obtained from the open -source data repository Kaggle. The features that servers for anomaly detection in IoT sites includes:-

- Denial of Service
- Scan
- Data Type Probing
- Spying
- Wrong setup
- Malicious control

Random Forest is used in creating the forest using various decision trees. The decision trees altogether forms the random forest, it is predicted by taking average for each tree component. Artificial Neural Network is used in training the data. It takes longer time period in optimising the error value. The features extracted are multiplied by using random weights and the bias values are added with them. The evaluation metrics such as True positive, True negative, False positive and False negative are done in the construction of Confusion matrix. Random Forest is suggested to be the best classifier in the feature extraction.

9. Detecting User-To-Root (U2R) Attacks

Intrusion detection is process of analysing data. The mapping of data items into several predefined classifications are done. It will be helpful in gathering sufficient data from the user programs such as "Normal" and "Abnormal". Then classification algorithms is applied that will determine what data are to be belonging to the normal and abnormal class. The main problem that
results in higher false rates is that difficulty in finding the features, and also performance requests in higher degree exits. The machine learning models implied here helps in detecting the process of anomalies that exists in the attacks especially root type cyber-attacks. The classification machine learning algorithms helps in classifying datasets are Random forest, Multi-layer perceptron, Naïve Bayes and JRIP.

Table 1. Analysis of Cyber-attack detection

Approaches	Dataset	Pre-Processing	Feature Extraction	Classification Techniques	Merits	Demerits	Ref
Generative Adversarial Network	HTTP dataset CSIC	Extracts security data from	Host, type of network, malware	t-SNE algorithm	Generation of data similar to	Training the data is harder	[1]

		the different sensor	functionality		the original set of data		
Ad-IoT using machine learning	UNSW-NB15 dataset	Malware and bengin separation	Traffic such as DoS, Worms, exploits, Attacks, Backdoor and normal	Decision Tree, Random Forest and K – Nearest Neighbour	Analysing the attacks that could high affect the security	Using malicious data into the machine learning algorithm	[2]
Machine learning for anomaly detection of cyber-attacks	UNSW-NB15 dataset	Network traffic extraction	Prototype, service, state, packet count from source to destination	K – Nearest Neighbour	Large amount of data features are taken into consideratio n	More time in labelling process	[3]
Machine learning in detecting cyber-attacks	Bot-IoT	CICFlow Meter used in extraction of flow based features	Attacks such as analysis, normal, shell code, reconnaissa nce and worms	Naïve Bayes and Decision Tree classification	Ability to make the model to be self-trained in analysing attacks	Environme ntal factors is not given a concern	[4]
Anomaly detection of IoT attacks	UNSW-NB15 dataset	Retrieval of Fog node features	Flow based features from the traces of raw network traffic	Support vector machine and Logistic regressions	Consideratio n of the node features in the fog layer	High reliance on the data presentatio n	[5]
IoT based on ensemble techniques	NSL-KDD dataset	Feature extraction of correlation based features	Correlation combination features	Logistic Regression and Multi class classification model	Provides better prediction than single contribution model	Not given a concern on unknown differences between test and train dataset	[6]
Attacks and anomaly detection in IoT sensor sites	KDD dataset	Count of packets from source to destination	Type of layer concerned in fog architecture	Logistic Regression, Random Forest, Support vector regression and Naïve Bayes	More samples are given training in differentiati on of Normal and DoS (Denial of Service)	Drooped some essential features when pre-processing is done	[7]
Detection of U2R attacks	KDD99 dataset	Mapping of symbolic attributes to numerical attributes	Duration, type of the protocol, src_bytes, dst_bytes, flag, wrong_frag ment and urgent	Two tier classification	The classifier is retrieved predict the connection level attack	Well define patterns of the extracted features exploits the system resources	[8]
Detection of DDOS attacks	DARP A 2000, CAIDA	Identifying clusters with non-	Entropy values, Various	EM-CURE cluster analysis	An efficient model in analysing	Takes much time in referring	[9]

using cluster analysis	2007 and CAIDA 2008 datasets	spherical shapes	phases of DDOS attacks		attack and construction has been done using clustering analysis	the DDoS attacks that contributes to the exploitation of the sub system	
Network intrusion detection system based on Recursive feature addition and bigram technique	ISCX 2012 dataset	Removing the features that deviates and results in over-fitting	Recursive Feature Addition in extracting the correlation based features	SVM classifier	Performance of the classification in featured extraction is increased	Feature extract do not help independently during classification	[10]
High Performance Attack Estimation	KDD99 and CIC-IDS datasets	Remove the known attack signatures	Correlation based feature selection, Gain ratio and Information gain	Decision tree classifier, SVM and KNN	High level of accuracy is concerned	Only the minimum set of supported features is concerned	[11]
Analysis of Network Intrusion Detection	KDD99 and NSL-KDD dataset	Reduction of network traffic that deviates from plotting	Denial of service, Buffer overflow, Denial of service	ANN, SVM and Random Forest	Labelling ensures efficient classification	There may be probability of wrong training	[12]
Anomaly Detection of network traffic based on packet bytes	DARPA datasets	Traffic is filtered	Probe, Denial of service, data. U2R and R2L	NETAD models	Many of the events that are hostile are located	Does not monitor the incoming packets and outsourcing traffic	[13]
Detection of light weight intrusion using wrapper approach	KDD CUP99 dataset	Redundant instances that are unbiased are removed	Patterns that categorize normal and anomaly	Naïve Bayes, Random Forest and Decision stump	Constructs neuro-tree that helps in achieving better degree of accuracy detection	Contributes higher training time and security	[14]
Network intrusion detection using Firefly algorithm	KDD CUP99 dataset	Removal of redundant features of network traffic	Various attacks that results in network traffic (Bengin / Normal)	Bayesian network and Firefly algorithm	Maximum efficiency and lower false rate in detection of anomaly	Extracting the features of High dimensional network traffic is tedious	[15]

10. Discussion

The data has been extracted from various resources such as Open Source Kaggle, UNSW-NB15 and CICFlowMeter. The basic features mainly that affects or influences the attacks are taken into consideration. The fog architecture is closely examined since many of the IoT devices has been connected in the IoT layer. The classification method chosen must be a multi-layered classifier so the degree of accuracy is relatively high. Also, many performance measure is considered in the estimation of machine learning algorithms. Among all, Random

Forest is better suited for the classification mechanism. Moreover, this study proposed proves to be helpful in data extraction and pre-processing. Also, the feature extraction and classification algorithm specified in this survey helps in choosing a better one in anomaly detection of cyber-attacks in IoT.

Conclusion

Cyber-attacks is the major threat in data sharing and also the main area to be noted while the data packets are transferred from source to destination. Machine learning algorithms suggested helps in identifying the related features that correlate with the cyber-attacks. The data analysis and pre-processing methods proposed in this survey helps in getting a better accuracy range of prediction. From all the above mentioned, multi-layered perceptron proved to a better one in analysing the cyber-attacks other than pre-existing machine learning algorithms.

References

11. Ferdowsi, A., & Saad, W. (2019). Generative Adversarial Networks for Distributed Intrusion Detection in the Internet of Things. *2019 IEEE Global Communications Conference (GLOBECOM)*.
12. Alrashdi, I., Alqazzaz, A., Aloufi, E., Alharthi, R., Zohdy, M., & Ming, H. (2019). AD-IoT: Anomaly Detection of IoT Cyberattacks in Smart City Using Machine Learning. *2019 IEEE 9th Annual Computing And Communication Workshop And Conference (CCWC)*.
13. Sushmitha R & Deepa N P. (2020).Machine Learning Approach for Anomaly Detection of IoT Cyber-attacks in smart city, International Research Journal of Engineering and Technology, 7(6), 6945-6947.
14. Alsamiri, J., & Alsubhi, K. (2019). Internet of Things Cyber Attacks Detection using MachineLearning. *International Journal Of Advanced Computer Science And Applications*, *10*(12), 627-634.
15. Sushmitha, R., Deepa, N., and Sudha, K L . (2020). Anomaly Detection of IoT Cyber-attacks in Smart City build on Machine Learning Algorithms,*Journal of Seybold Report*, 15(9), 71-84.
16. Rashid, M., Kamruzzaman, J., Hassan, M., Imam, T., & Gordon, S. (2020). Cyberattacks Detection in IoT-Based Smart City Applications Using Machine Learning Techniques. *International Journal Of Environmental Research And Public Health*, *17*(24), 9347.
17. Hasan, M., Islam, M., Zarif, M., & Hashem, M. (2019). Attack and anomaly detection in IoT sensors in IoT sites using machine learning approaches. *Internet Of Things*, *7*, 100059.
18. Bahl, S., & Sharma, S. (2015). Detection rate analysis for user to root attack class using correlation feature selection. *International Conference On Computing, Communication & Automation*,66-71.
19. Bhaya, W., & EbadyManaa, M. (2017). DDoS attack detection approach using an efficient cluster analysis in large data scale. *2017 Annual Conference On New Trends In Information & Communications Technology Applications (NTICT)*, 168-173.
20. Hamed, T., Dara, R., & Kremer, S. (2018). Network intrusion detection system based on recursive feature addition and bigram technique. *Computers & Security*, *73*, 137-155.
21. Freas, C., Harrison, R., & Long, Y. (2018). High Performance Attack Estimation in Large-Scale Network Flows. *2018 IEEE International Conference On Big Data (Big Data)*, 1-8.
22. Saddam Hossen & Anirudh Janagam 2018). Analysis of Network Intrusion Detection System with Machine Learning Algorithms (Deep Reinforcement Learning Algorithm), Master of Science in Electrical Engineering with emphasis on Telecommunication Systems, 1-80.
23. Mahoney, M. (2003). Network traffic anomaly detection based on packet bytes. *Proceedings of The 2003 ACM Symposium On Applied Computing - SAC '03*, 346-350.
24. Sivatha Sindhu, S., Geetha, S., & Kannan, A. (2012). Decision tree based light weight intrusion detection using a wrapper approach. *Expert Systems With Applications*, *39*(1), 129-141.
25. Selvakumar. B., & Muneeswaran. K. (2019). Firefly algorithm based feature selection for network intrusion detection. *Computers & Security*, *81*, 148-155.