www.ierjournal.org

Type of manuscript

Survey on anomaly based intrusion detection systems in IoT networks

[#1]Rhishabh Hattarki, [#2]Shruti Houji, [#3]Sanika Patil, [#4]Sahil Dixit, [#5]Manisha Dhage

[1]rhish9h@gmail.com,
[2]shrutihouji98@gmail.com,
[3]patilsanika02@gmail.com,
[4]sahildxgeneration10@gmail.com,
[5]mrdhage.scoe@sinhgad.edu

[#12345]Computer Department, Sinhgad College of Engineering,
Pune, Maharashtra, 411041, India

ABSTRACT

IoT devices are becoming popular day-by-day. Several vulnerabilities in IoT present the need for IoT security. The number of attacks on these devices keep increasing and most of them are slight variations of the previously known attacks, which can bypass the conventional firewall systems.

The existing systems are not suitable for IOT devices as IOT devices have low computational power. Those that use signature-based intrusion detection work only on known patterns and attacks, hence they cannot recognize newer attacks with unknown pattern. Also, many systems use cloud computing, which has a downfall that it needs access to internet at all times and are most often paid.

The following survey discusses a few of the anomaly based intrusion detection systems built to tackle this situation. It shows the different methods used along with their drawbacks or future scope. Also, an anomaly-based detection system has been proposed. It comes into effect when detecting newer attacks, that are not filtered by the firewall. It is capable of handling newer/unknown attacks, which signature based cannot deal with. It would be set up on a local higher powered device rather than on cloud.

ARTICLE INFO

Keywords— Intrusion detection, Internet of Things, Network security, Machine learning

I. INTRODUCTION

The advancement in computer technologies such as mobile and pervasive computing, internet applications and services has led to the proliferation of IoT (Internet of Things) devices. [1] IoT consist of devices such as smart devices, vehicles, home appliances, industrial smart devices that contain electronics, software, sensors, actuators, and connectivity which allows these things to connect, interact and exchange data.

With myriads of devices generating, processing and exchanging a vast amount of data that consists of safety, security and privacy critical information, it is bound to attract the attention of cyber criminals. The security issues are often known to the device manufacturers. However, a large chunk of the development time and effort is dedicated into getting the product to the market prior to the deadline. Hence, the security in IoT devices is either neglected or left out. As a result IoT devices become an easier target for hackers. [2]. Using the vulnerabilities in IoT design, software and hardware, the adversaries can plant malicious code, install a virus, etc. and infiltrate the IoT network.

When an IoT network is compromised, the intrusion detection system works as a second line of defence after the firewall. If an attacker breaks into a system or system server by forwarding malicious packets to the user system, which is used to steal, modify or corrupt any private important information, it is said to be an Intrusion. The intrusion detection system monitors the network traffic for suspicious activity using either misuse or anomaly detection systems. In the misuse detection system, the suspicious activity is compared to an existing set of behavior patterns whereas in anomaly detection system, intrusions are detected by checking for irregularities from the normal network traffic. [3]

An anomaly, also referred to an outlier, is data that deviates from the normal system behavior, which is learned by the detection system by the provided set of accepted network behavior data. Anomalies indicate that the system has a fault or that a malicious event has compromised the system. [4] Although it is prone to high false alarm rates, where a particular activity is considered as a threat even when it isn't, Anomaly based detection system can be preferred over signature based methods as it allows the detection of unseen attacks, whose signatures do not exist with the system.

II. IOT SECURITY REQUIREMENTS

1. Data privacy, confidentiality and integrity:

Data in IoT networks travel through multiple nodes/devices before reaching the target destination. Hence, a good encryption mechanism is required so that the data is not seen by the anyone other than the rightful owner, that is it remains confidential. A compromised node can violate the privacy of the data stored on the devices in the network. The vulnerable devices can have their data tampered with in a malicious way by the attacker thereby affecting data integrity.

2. Authentication, authorization and accounting:

The authentication is needed between 2 parties communicating with one another to secure communication in IoT. The devices must prove that they're the designated party and not an unwarranted adversary. The main goal of authorization mechanisms is to confirm that the access to systems or information is provided to the licensed ones. Implementing authentication and authorization ensures trust and security while communicating. The accounting for resource usage, along with auditing and reporting gives a reliable mechanism for securing network management.

3. Availability of services:

One of the most commonly implemented and tested attack, denial of service (DOS) is an attack where the attacker tries to deny service to the user by flooding the network with unnecessary traffic. This results in unavailability of services and eventually reduces quality of service (QOS).

4. Energy efficiency:

IoT devices are typically cursed with lack of resources like power and storage. They are lightweight and are susceptible to attacks that can leverage this vulnerability by flooding the devices' resources eventually leaving them unfit to use. Hence, security measures need to be energy efficient that can run on these low powered, low storage devices.

5. Single points of failure:

The explosive increase of IoT networks worldwide pose the issue of a huge number of single points of failure. This problem ruins the experience of IoT users that would have to deal with frequent device failures in their networks. This gives rise to the necessity of having implementations and measures that will provide robust networks. [5]

III. RELATED WORK

Abebe A. D. and Naveen C. have proposed a distributed rather than the centralized deep learning based IoT/Fog network attack detection system. [6] The distribution reduces the performance overhead from the individual IoT nodes which can be very helpful. The results of the distributed network show it to be better than the centralized network. However, the performance comparison of deep model vs shallow model is insufficient as only one shallow model is compared and no ensemble models are tested.

Ashima C. et. al. propose a CNN-GRU language model for the recently released ADFA-LD intrusion detection data set. [7] It uses the newer dataset; and Gated Recurrent Units rather than the normal LSTM networks to obtain a set of comparable results with reduced training times. However, it is an implementation of neural networks and it cannot match the performance of ensemble methods (stated in their conclusion).

Ebelechukwu N. et al. propose an anomaly detection algorithm for detecting anomalous instances of sensor based events in an IoT device using provenance graphs. [4] Provenance provides a comprehensive history of activity performed on a system's data. In this approach, an observed provenance graph is compared to an application's known provenance graph in order to detect the anomalies. This method has insufficient experimental data to prove it is better than other methods.

The hybrid efficient model proposed by Shadi A. et. al. [3], an ensemble consisting of J48, Meta Pagging, Random Forest, REPTree, AdaBoostM1, Decision Stump and Naïve Bayes is an anomaly based intrusion detection system. It is trained and tested on the NSL-KDD dataset and outperforms other simpler single machine learning models like the J48, SVM and Naïve Bayes in terms of accuracy in detecting the attacks.

Tagy Aldeen Mohamed ,Takanobu Otsuka and Takayuki Ito [8] proposed IDS based on machine learning techniques to be implemented into IOT platforms as a service. Random Forest is used as a classifier to detect intrusions. Also neural network classifier is used to detect the categorization of the detected intrusions. The experimental results showed the proposed model can efficiently use Cloud computing service to process data from different resources of IOT platforms.

GARUDA introduces a novel distance measure that can be used to perform feature clustering and feature representation for efficient intrusion detection. [9] GARUDA has better detection rates for U2R and R2L attack types, however it is more focused on feature reduction and dissimilarity measure used in the classifiers. Moreover, it did not improve classification accuracies on SVM.

Bayu A. T. and Kyung H. R. have presented an effective

anomaly detection by incorporating PSO-based feature selection and random forest model. [10] It uses a random forest model with particle swarm optimization-based feature selection. It focuses on feature selection and compares performance with only 2 other models.

Anomalous Payload-Based Network Intrusion Detection by Ke Wang and Salvatore J. Stolfo [11] have presented a detector called PAYL which establishes a base profile of the system using a byte frequency distribution, then detecting the anomaly by the standard deviation from the base using the Mahalanobis distance. The model improves on the false alarm rates and it is an incremental, unsupervised model. However, the dataset used, 1999 DARPA IDS dataset is fairly old and there is scope for more testing in live environments.

Belal S. K. et. al. [12] have implemented a fog computing based IDS for IoT using a Multilayer Perceptron (MLP) model. The fog layer sits between the cloud and the IoT network and improves on latency, mobility, user experience and geographical distribution. They have used the newer datasets ADFA-LD and ADFA-WD and raspberry pi as the fog node. This lightweight system addresses some of the issues from cloud based systems, although, they still have room for improvement in terms of accuracy and efficiency.

Aymen Yahyaoui et al. [13] have come up with a heirarchical solution for IoT Intrusion detection. They've used support vector machine for WSN intrusion detection and deep learning for detecting IP intrusions at the gateway. Historical data is used to determine the probability of getting an attack and SVM runs on the nodes with high probability of attack. This saves on energy which is critical in IoT networks. This study only covers few types of attacks and can be extended to detect more types of IoT attacks.

Veeramreddy J. and Koneti M. P. [14] proposed an Anomaly-Based IDS to develop generic meta-heuristic scale for both known and unknown attacks with a high detection rate and low false alarm rate by adopting efficient feature optimization techniques using the popular NSL-KDD dataset. Though existing intrusion detection techniques address the latest types of attacks like DoS, Probe, U2R, and R2L, reducing false alarm rate is a challenging issue with high classification accuracy.

Sébastien J.J. G., Alvin K.H. L., Craig O. F., et al. [15] have presented an anomaly based intrusion detection system using time based histogram comparison for the MIL-STD-1553 protocol (commonly used in military aircrafts). Baseline histograms are created for the time based features used in the considered protocol. They are compared with either the real time data or stored data histograms to determine the activity as anomalous or normal. It's a more basic approach compared to the modern techniques like machine learning, however it shows promising results for their requirements. There is room for improvement in terms of including multiple aircraft modes, using parallel processing for improved speed, adding additional features.

S. Venkatraman & B. Surendiran [16] have proposed a hybrid IDS for home IoT networks which includes the combination of signature based as well as anomaly based systems. The signature based IDS uses a crowd sourced framework for the pattern set. This hybrid proves to be more accurate than the systems running individually. However, this can be tested more extensively and compared with more similar IDSs, more number of IoT devices can be used.

Luis M.T, Eduardo M., Mikel I. et al. [17] proposed an anomaly-based IDS for IEEE 802.11 networks introducing a wireless IDS called S2WIDS. It is an anomaly based system that implements a multidisciplinary approach to detect the most common attacks in wireless environments and it may be able to fight some of the new threats that might arise in the future. However, it does not provide additional information about the AP (unauthorized wireless access point) to improve the corresponding part of the detection engine.

Ayyaz-ul-Haq Q. (B), Hadi L., Jawad A. et al. [18] have developed a heuristic intrusion detection system for IoT using a random neural network (RNN). The dataset used is NSL-KDD and the algorithm used is gradient descent. The dataset was reduced using feature selection and then tested using different learning rates. Highest accuracy was achieved at lower learning rates. The model faired well when it was compared to different models. This system can use more testing and implementation in an actual IoT network.

IV. PROPOSED METHODOLOGY

The system has been inspired by the already existing intrusion detection systems which attempt to protect the home-network against attacks and intrusions. We are taking these pre-existing models and combining them to get additional advantages and reduce the downsides as much as possible. Till date, there have been very few attempts at making an IDS for IOT devices. Most of the IDS present in the market cater to non-IDS networks. We are taking the principles of these IDS systems and modifying them as per requirement of an IOT network.

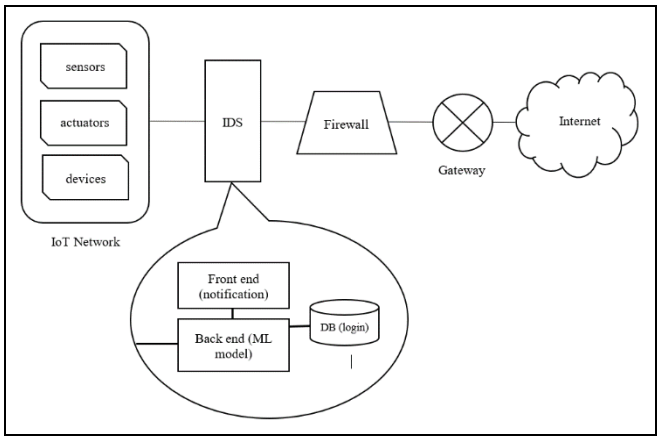

Fig 1. System Overview Diagram

The proposed system is a web app that constantly runs in the background. The IDS as shown in the Fig 1. Sits between the firewall and the IoT network.

A. Modules

1) User authentication

The intrusion detection system will get access only to the authorized personnel. Hence, the web app will have a user authentication system (login system).

2) Data connection establishment

Once the system has been started, it needs to check whether all the devices have been connected properly. If not, and error message will be sent to the front end.

3) Intrusion detection analysis

The machine learning model random forest will be used for the intrusion detection. The network data that is being monitored at run time will be preprocessed and sent to the model. This data will be analyzed here and classified as either normal or anomalous.

4) Notification

If an intrusion is detected, a notification will be sent to the user. The user will then be able to see more details if required.

B. Mathematical model

Let S be the solution set for the given problem statement. S={Input,Function,Output, Terminate,Success,Failure}. Where,Input =Input to the System. Function =Functions of the system. Output =Output of the system. Terminate= Terminating Condition of the System. Success =Success cases for the System. Failure =Failure cases for the system.

1. Input ={UserName,Password,Data Packets}
   a. UserName :-use rid
   b. Password : user passsword
   c. Data Packets:- A data packet is a unit of data made into a single package that travels along a given network path.

2. Function={Login auth,Network connection,train test,intrusion detection,Notification}

a. Login auth: This function will take Username and Password as the input and gives the authorization rights to the user accordingly.

IF Y=F(X) is the login auth function then

X:- Username and Password taken as an input.

Y:- Authorization rights of a user.

b. Network connection =This function will take port number, ip address as an input to give the status of the connection network.

IF Y=F(X) is the Network Connection Function, then

X:- Inputs to setup a network, connecting to a device ,routing the incoming data and device configuration.

Y:- Status of the network connection.

c. train test: This function will take features as an input to give the accuracy of the model and further it will be turned to improve the accuracy of the model. Random Forest has been used as a Machine learning algorithm for building the model.

IF Y=F(X) is the function to test and train the model then,

X:- Features of the DS2OS dataset as an input.

Y:- Accuracy of the model

d. intrusion detection : This function will take features like sourceID, sourceAddress, sourceType, sourceLocation, destinationServiceAddress, destinationServiceType, destinationLocation, accessedNodeAddress, accessedNodeType, operation, value, timestamp as an input and notifies the user the system for the intrusion.

C. Algorithm

Random Forest Classifier has been used for the building and implementing the model.

Basically, a random forest is an average of tree estimators. As with non-parametric regression, simple and interpretable classifiers can be derived by partitioning the range of X.

Let n = A1, . . . , AN be a partition of X. Let Aj be the partition element that contains x. Then h(x) = 1 if XiAj Yi XiAj (1 Yi) and h(x) = 0 otherwise.

We conclude that the corresponding classification risk satisfies R(h) R(h) = O(n1/(d+2)).

Trees are useful for their simplicity and interpretability. But the prediction error can be reduced by combining many trees.

These are bagged trees except that we also choose random subsets of features for each tree. The estimator can be written as

$$m(x) = \frac{1}{M}\sum_{j} m_j(x)$$

where mj is a tree estimator based on a subsample (or bootstrap) of size a using p randomly selected features. The trees are usually required to have some number k of observations in the leaves. There are three tuning parameters: a, p and k. You could also think of M as a tuning parameter but generally we can think of M as tending to . For each tree, we can estimate the prediction error on the un-used data. (The tree is built on a subsample.) Averaging these prediction errors gives an estimate called the out-of-bag error estimate. IF Y=F(X) is the function then,

X: sourceID, sourceAddress, sourceType, sourceLocation, destinationServiceAddress, destinationServiceType, destinationLocation, accessedNodeAddress, accessedNodeType, operation, value, timestamp as an input

Y: Notifies the system.

e. Notification: This function takes the input which is provided by the function (intrusion detection) and notifies the user.

IF Y=F(X) is the function then,

X: Output of the function(intrusion detection). Y: Intrusion message to the user.

3. Output = {display intrusionmsg }
   a. display intrusionmsg : display error message if any intrusion occurs.

4. Intermediate Results
   a. Successful working of module.
   b. Successful Working of Network.
   c. Successful User authentication.

5. Terminate= {Invalid details, Network failure, Timeout
   a. Invalid User Authentication.
   b. Network failure
   c. timeout

6. Success
   a. Successful user login.
   b. Successful connection establishment of nodes and ids.
   c. Successful detection of intrusion.
   d. Displaying the results.
   e. Appropriate error messages in case of invalid input.

7. Failure
   a. Web app Failure.
   b. Hardware faults.
   c. Network establishment failure.
   d. Not displaying required results.

V. CONCLUSION

Intrusion detection systems act as a second line of defense after the firewall and are beneficial for the security of IoT networks. The paper provides a survey of few of the different types of intrusion detection systems, mostly anomaly based systems. A system has been proposed using the pointers from the survey to detect intrusions in IoT networks. This can be used in a real time IoT networks to improve its security.

ACKNOWLEDGEMENT

We thank Prof. M.R Dhage for her expert guidance and continuous encouragement throughout to see that adequate research has been conducted to approve the project. Collectively, we would also like to thank our project committee members Prof. E. Jayanthi and Prof. A.S. Kalaskar for their time, suggestions, and for graciously agreeing to be on our committee, and always making themselves available. We express deepest appreciation towards Prof. M. P. Wankhade, Head of Department of Computer Engineering, Dr. S. D. Lokhande, Principal, Sinhgad College Of Engineering.

REFRENCES

[1] Jesus P., Salim H., "IoT Security Framework for Smart Cyber Infrastructures", IEEE 1st International Workshops on Foundations and Applications of Self* Systems, 2016, pp. 242-247

[2] Jacob W., Khoa H., Orlando A., Ahmad-Reza S. et. al., "Security analysis on consumer and industrial IoT devices", 21st Asia and South Pacific Design Automation Conference (ASP-DAC), 2016

[3] Shadi A., Monther A., Muneer B. Y., ”Anomaly-based intrusion detection system through feature selection analysis and building hybrid efficient model”, Journal of Computational Science, 25, 2018, 152-160.

[4] Ebelechukwu N., Andre C., Gedare B., ”Anomaly-based Intrusion Detection of IoT Device Sensor Data using Provenance Graphs”, 1st International Workshop on Security and Privacy for the Internet-of-Things (IoTSec), 2018.

[5] Minhaj A. K., Khaled S., "IoT security: Review, blockchain solutions, and open challenges", Future Generation Computer Systems, 82, 2018, 395-411

[6] Abebe A. Diro., Naveen C., ”Distributed attack detection scheme using deep learning approach for Internet of Things”, Future Generation Computer Systems, 82, 2017, 761-768.

[7] Ashima C., Brian L., Sheila F., Paul J., ”Host based Intrusion Detection System with Combined CNN/RNN Model”, IWAISe 2018- ECML PKDD Conference, 2018.

[8] TagyAldeen M., Takanobu O., Takayuki I., ”Towards Machine Learning Based IoT Intrusion Detection Service”, Recent Trends and Future technology in Applied Intellignece, 2018, 580-585.

[9] Shadi A. A., Radhakrishna V., "GARUDA: Gaussian dissimilarity measure for feature representation and anomaly detection in Internet of things", The Journal of Supercomputing, 2018, pp 1-38

[10] Bayu A. T., Kyung-Hyune R., "An Integration of PSO-based Feature Selection and Random Forest for Anomaly Detection in IoT Network", MATEC Web of Conferences, vol 159, 2018

[11] Ke W., Salvatore J. S., “Anomalous Payload-Based Network Intrusion Detection”, RAID 2004, LNCS 3224, pp. 203–222, 2004

[12] Belal S. K., Ainuddin W. B. A. W., Mohd Y. I. B. I., et. al., “A Lightweight Perceptron-Based Intrusion Detection System for Fog Computing”, Appl. Sci. 2019, 9, 178; doi:10.3390/app9010178

[13] Aymen Y., Takoua A., Rabah A., “Hierarchical anomaly based intrusion detection and localization in IoT”, 2019, 15th International Wireless Communications & Mobile Computing Conference (IWCMC), pp 108-113

[14] Veeramreddy J., Koneti M. P. , “Anomaly-Based Intrusion Detection System”, IntechOpen, 2019.

[15] Sébastien J.J. G., Alvin K.H. L., Craig O. F., et. al., “MAIDENS: MIL-STD-1553 Anomaly-Based Intrusion Detection System Using Time-Based Histogram Comparison”, IEEE Transactions on Aerospace and Electronic Systems, 2019

[16] S. Venkatraman & B. Surendiran, “Adaptive hybrid intrusion detection system for crowd sourced multimedia internet of things systems”, Multimed Tools Appl, 2019

[17] Luis M.T, Eduardo M., Mikel I. et. al. , “An anomaly-based intrusion detection system for IEEE 802.11 networks”, 2010 IFIP Wireless Days, 2010

[18] Ayyaz-ul-Haq Q. (B), Hadi L., Jawad A. et. al., “A Heuristic Intrusion Detection System for Internet-of-Things (IoT)”, CompCom 2019, AISC 997, 2019, pp. 86–98