[[1]](#footnote-1)

Anomaly based Intrusion Detection System for IOT devices using Random forest

Rhishabh H. Author, Shruti H. Author, Sanika P. Author, and Sahil D. Author

*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. It works 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, also the cloud services are most often paid.

In the proposed system, we are planning to use anomaly-based detection system. The anomaly-based intrusion detection system 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. Also we are setting up the IDS on a local higher powered device rather than on cloud. Machine learning ensemble model Random forest is used. The model will be trained on the DS2OS traﬃc traces dataset.

*Index Terms*—Intrusion detection, Internet of Things, Network security, Machine learning

# INTRODUCTION

Internet of things, or IoT, is a system of interrelated ”things ” that are provided with unique identiﬁers and the ability to transfer data over a network without requiring human-to-human or human-to-computer interaction. These ”things” could be computing devices, mechanical and digital machines such as build in sensors, monitors etc. A good example of a network of IoT devices is the security system implemented by various locations which include, CCTV cameras, motion sensors, automatic locks, smoke detectors, temperature sensors etc. IOT devices are becoming pervasive. They are used extensively in a lot of ﬁelds and their utility is just going to keep increasing. IOT devices help to automate things, reduce labor costs and facilitate smart living. Hence, it is important to build an optimal system that can provide proper safety and security measures for IOT devices. An intrusion detection system is one such system that can be a building block in making the IOT network as secure as possible. An intrusion detection system is a system that passively monitors the data exchange in the network and of the network with externals entities and looks for malicious activities that can be classiﬁed as an intrusion or attack. It then notiﬁes the user or sends a notiﬁcation to some other system which may or may not take action against the detected intruder. Simply put, if the network of devices is a home, an intrusion detection system is a CCTV camera. Also, a step up from just a intrusion detection system is an anomaly based intrusion detection system. This type of system creates a proﬁle of normal behavior, and any activity that falls outside of the normal category is marked as anomalous. Anomaly based system is better suited to defend the system against zero-day attacks. There are certain challenges while building this system which are speciﬁc to IOT devices. For example, the low computational powers and limited amount of resources (such as space/memory). The IDS system should be build keeping in mind all these challenges and they should be overcome in the most eﬃcient manner.

# 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. [1] 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. [2] 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. [3] 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. [4], 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.

# system overview

Product Perspective

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.

Product Functions

1) User authentication

2) Data connection establishment

3) Intrusion detection analysis

4) Notiﬁcation

# 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.

• 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 nonparametric 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

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.

Acknowledgment

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.

References

1. 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.
2. Ashima C., Brian L., Sheila F., Paul J., ”Host based Intrusion Detection System with Combined CNN/RNN Model”, IWAISe 2018- ECML PKDD Conference, 2018.
3. 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.
4. 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.
5. 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.

**ALL BIOS ARE REQUIRED**

**First A. Author** (M’76–SM’81–F’87) and the other authors may include biographies at the end of regular papers. Biographies are often not included in conference-related papers. This author became a Member (M) of IEEE in 1976, a Senior Member (SM) in 1981, and a Fellow (F) in 1987. The first paragraph may contain a place and/or date of birth (list place, then date). Next, the author’s educational background is listed. The degrees should be listed with type of degree in what field, which institution, city, state, and country, and year degree was earned. The author’s major field of study should be lower-cased.

The second paragraph uses the pronoun of the person (he or she) and not the author’s last name. It lists military and work experience, including summer and fellowship jobs. Job titles are capitalized. The current job must have a location; previous positions may be listed without one. Information concerning previous publications may be included. Try not to list more than three books or published articles. The format for listing publishers of a book within the biography is: title of book (city, state: publisher name, year) similar to a reference. Current and previous research interests end the paragraph.

The third paragraph begins with the author’s title and last name (e.g., Dr. Smith, Prof. Jones, Mr. Kajor, Ms. Hunter). List any memberships in professional societies other than the IEEE. Finally, list any awards and work for IEEE committees and publications. If a photograph is provided, the biography will be indented around it. The photograph is placed at the top left of the biography. Personal hobbies will be deleted from the biography.

1. Manuscript received October 9, 2001. (Write the date on which you submitted your paper for review.) This work was supported in part by the U.S. Depart­ment of Com­merce under Grant BS123456 (sponsor and financial support acknowledgment goes here). Paper titles should be written in uppercase and lowercase letters, not all uppercase. Avoid writing long formulas with subscripts in the title; short formulas that identify the elements are fine (e.g., "Nd–Fe–B"). Do not write “(Invited)” in the title. Full names of authors are preferred in the author field, but are not required. Put a space between authors’ initials.

   F. A. Author is with the National Institute of Standards and Technology, Boulder, CO 80305 USA (corresponding author to provide phone: 303-555-5555; fax: 303-555-5555; e-mail: author@ boulder.nist.gov).

   S. B. Author, Jr., was with Rice University, Houston, TX 77005 USA. He is now with the Department of Physics, Colorado State University, Fort Collins, CO 80523 USA (e-mail: author@lamar.colostate.edu).

   T. C. Author is with the Electrical Engineering Department, University of Colorado, Boulder, CO 80309 USA, on leave from the National Research Institute for Metals, Tsukuba, Japan (e-mail: author@nrim.go.jp). [↑](#footnote-ref-1)