A Seminar Report

on

IOT SECURITY: ANOMALY BASED INTRUSION DETECTION SYSTEM

*by*

Rhishabh Hattarki

Roll No: 305218

*Under the guidance of*

Prof. M. R. Dhage



Department of Computer Engineering

Sinhgad College of Engineering, Pune-41

*Accredited by NAAC*

**UNIVERSITY OF PUNE**

**2018-2019**

|  |  |
| --- | --- |
| Sinhgad Technical Education Society,  Department of Computer Engineering  Sinhgad College of Engineering , Pune-41 | New_Sinhgad_Logo_2013_300 |

Date:

**CERTIFICATE**

This is to certify that Mr. Rhishabh Hattarki has successfully completed his seminar work titled “IOT SECURITY: ANOMALY BASED INTRUSION DETECTION’’ at Department of Computer Engineering, SCOE, Pune for the partial fulfillment of the Bachelor Degree of Computer Engineering , Savitribai Phule Pune University, in semester-II, academic Year 2018-2019.

Prof. M. R. Dhage Prof. M. P. Wankhade

Internal Guide Head

Department of Computer Engineering

Dr. S. D. Lokhande

Principal

Sinhgad College of Engineering

**Abstract**

IoT devices are becoming ubiquitous. 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 anomaly-based intrusion detection system comes into effect when detecting newer attacks, that are not filtered by the firewall.

An anomaly-based intrusion detection system, is an intrusion detection system for detecting both network and computer intrusions and misuse by monitoring system activity and classifying it as either normal or anomalous. The normal system behavior is passively monitored and the normal activity profile is built. The suspicious activity that does not match this normal activity profile, the outlier is considered an anomaly. This system helps in automatically identifying suspicious IOT devices connected to the network.

It consist of the training phase where a profile of normal behaviors is built and testing phase where current traffic is classified as attack or normal with the profile created in the training phase. Machine learning ensemble model has been used, including several classifiers including J48, Meta Pagging, Random Forest, REPTree, AdaBoostM1, Decision Stump and Naïve Bayes. It is trained on the popular dataset of NSL-KDD.

(i)

**Acknowledgements**

I would like to express my deepest appreciation to my project guide, Prof. M. R. Dhage for ameliorating my seminar report by overseeing and guiding it throughout the semester.

I would like to show my gratitude to our Seminar Coordinator, Prof. Jayanthi E. for giving us valuable inputs to help us finish the report.

I would specially thank our head of department, Prof. M. P. Wankhade, for providing me with this opportunity to do the seminar report which helped me learn new concepts in a novel technology.

I am also thankful to our Principal, Dr. S. D. Lokhande, for granting me the possibility of doing this report.

Last, but not least, I’d thank my friends and family for helping me select the seminar topic in the stipulated time.

Rhishabh Hattarki

(ii)

**List of Tables**

|  |  |  |
| --- | --- | --- |
| **Table No.** | **Title** | **Page No.** |
| 3.1  3.2  4.1  4.2 | Confusion matrix  Overview of NSL-KDD data  Comparison of four classifiers  Accuracy in detection of normal and attack network flows by using the J48, SVM, Naïve Bayes and hybrid model classifier | 11  12  15  15 |
|  |  |  |

(iii)

**List of Figures**

|  |  |  |
| --- | --- | --- |
| **Figure No.** | **Title** | **Page No.** |
| 2.1  2.2  2.3  3.1  3.2 | Security requirements of IoT  Types of IDS based on attack type  Types of IDS based on approach  Position of IDS in IoT architecture  Functionality overview diagram | 5  7  7  10  11 |

(iv)

**Acronyms**

IoT Internet of Things

IDS Intrusion Detection System

ADS Anomaly Detection System

TCP Transmission Control Protocol

SQL Structured Query Language

IPS Intrusion Prevention System

NGIPS Next Generation Intrusion Prevention System

HIDS Host-based Intrusion Detection System

NIDS Network-based Intrusion Detection System

ML Machine Learning

WEKA Waikato Environment for Knowledge Analysis

SVM Support Vector Machine

(v)

**Contents**

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | | **Page No.** |
|  | **Certificate** | |  |
|  | **Abstract** | | i |
|  | **Acknowledgement** | | ii |
|  | **List of Tables** | | iii |
|  | **List of Figures** | | iv |
|  | **Acronyms** | | v |
| **Chapter-1** | **Introduction** | | 1 |
|  | 1.1 | Motivation | 2 |
|  | 1.2 | Timeline /evolution | 2 |
|  | 1.3 | Organization of the report | 4 |
| **Chapter-2** | **Literature Review** | | 5 |
|  | 2.1 | Fundamentals  2.1.1 Security requirements for IoT  2.1.2 Types of Intrusion detection systems  2.1.3 Types of HIDS  2.1.4 Types of anomaly detectors | 5  5  7  7  8 |
|  | 2.2 | Related Work | 8 |
|  | 2.3 | Survey Conclusion | 9 |
| **Chapter-3** | 3.1  3.2  3.3  3.4  3.5 | Methodology  Functionality overview  Confusion Matrix  NSL-KDD dataset description  Classes based on Attack Types  Algorithm | 10  10  11  12  12  13 |
| **Chapter-4** |  | Discussion of Results | 15 |
| **Chapter-5** |  | Conclusion and Future work | 17 |
|  | 5.1 | Conclusion | 17 |
|  | 5.2 | Future work | 17 |
|  |  | **References** | 18 |
|  |  |  |  |

**Chapter 1**

**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. [7] 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. [16] According to statista, the number of IoT connected devices installed worldwide will exceed 75B by the year 2025. [14].

With myriads of devices generating, processing and exchanging a plethora 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. [6]. 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 defense 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. [12]

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. [5] 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.

* 1. **Motivation**

A single security layer in the IoT architecture is not sufficient to provide the necessary measures to prevent the system from being affected by attackers. The firewall that is widely used in all computer systems isn’t a comprehensive solution. An IDS, comes into play as a complementary layer to provide security. It does not need to know how the adversary penetrated the system in order to detect it. It helps improve security response, consequently reducing the chance of missing security threats which could compromise confidentiality, integrity, privacy or availability of mission critical assets and processes ergo providing business benefits. The IDS system can be automated reducing human interaction and making it more efficient to detect threats.

[10]

Signature-based IDS relies on spotting a duplication of events or types of attack that have happened before. In order to deal with threats, organizations would have to manually create thousands of signatures and even then run into situations where novel attacks with minor changes in signature would bypass the security layer. The new attacks for which a signature doesn’t exist can be detected with the Anomaly detection if it falls out of the normal traffic patterns. For example, when a new system is infected with a worm it usually starts scanning for other vulnerable systems at an accelerated or abnormal rate flooding the network with malicious traffic, thus triggering a TCP connection or bandwidth abnormality rule, which can be detected by the anomaly detection system. Anomaly testing methods guarantee to provide far more effective protection against hacker incidents.

* 1. **Timeline/Evolution**

1986: An Intrusion-Detection Model

[4] was the first academic paper written by Dorothy E. Denning about a real time IDS which led Stanford Research Institute (SRI) to develop the Intrusion Detection Expert System(IDES). That system used statistical anomaly detection, signatures and profiles of users and host systems to detect nefarious network behaviors. It used a rule-based Expert system to detect known types of intrusions plus a statistical anomaly detection component based on profiles of users, host systems, and target systems.

2000-2005: Intrusion Detection over Prevention

Firewalls were predominantly the only security feature available, but they lacked deep packet inspection, ie they had no visibility to the content or context of network traffic. In the early 2000s new threats like SQL injections and cross site scripting (XSS) attacks were becoming popular and these attacks would pass right by the firewall. This gave rise to the popularity of IDS. IDS mostly relied on signature based methods along with pattern matching, string matching, anomaly detection and heuristic-based detection.

2006-2010: Adoption of faster combined IDS and IPS

As IDS were getting better at detecting intrusions, it was also much better at not blocking harmless traffic. This gave rise to the use of combined IDS and IPS for more effective results.

In addition to pattern matching, string matching, anomaly detection and heuristic based detection, vendors added information to block malicious command & control IP addresses as well as websites that were known to host malware, reducing the time it takes to detect threats.

2011-2015: Next Gen IPS

The next gen IPS included features such as application and user control. A traditional IPS inspects network traffic looking for known attack signatures and either alerts on the traffic or stops it from proceeding into your network, depending on how it has been deployed. AN NGIPS does that and provides extensive coverage of network protocols to detect a wider range of attacks. It also provides Application Control to limit the portions of an application that users can and cannot use (e.g., users may be able to post on Facebook but unable to upload any photos), and provides User Control, which would allow only certain people access to the application.

2016 onwards: Machine Learning

The threat landscape is constantly changing, and now security vendors are focusing on high-fidelity machine learning, which uses algorithms to analyze files, and uses noise cancellation techniques like census and whitelist checking. [8]

* 1. **Organization of the Report**

The report comprises of 5 chapters – Chapter 1 consists of the Introduction along with the motivation to write the report and the timeline/evolution of the particular technology in consideration. Chapter 2 contains the Literature review where the fundamentals required to understand the IoT, IDS, ADS is given along with a survey of related researches. Chapter 3 explains the methodology used in the latest method (hybrid model of anomaly based intrusion detection) being reviewed. Chapter 4 discusses the results obtained from the experiments conducted by the researchers of the method considered. Chapter 5 provides the conclusion for the seminar report and discusses the scope and future works.

**Chapter 2**

**Literature Review**

Internet of things has certain security requirements which need to be fulfilled. Along with this, the types of IDS, types of HIDS and types of ADS are explained in the fundamentals which would be helpful in understanding the further report. Related papers to the hybrid efficient model (ensemble used for anomaly detection) have been reviewed in the related works and the pros and cons are discussed in the conclusion.

**2.1 Fundamentals**

2.1.1 Security requirements for IoT

For a secure IoT deployment, various mechanisms and parameters need to be established:

**Security requirements of IoT**

Data privacy, confidentiality and integrity

Authentication, authorization and accounting

Availability of services

Energy efficiency

Single points of failure

Figure 2.1. Security requirements of IoT.

1. Data privacy, confidentiality and integrity:

IoT data travels through multiple hops in a network. To ensure the confidentiality of data, a proper encryption mechanism is required. As the services, devices and network are integrated together, a compromised node can violate the privacy of the data stored on the devices in the network. The devices susceptible to attacks 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 required between two parties communicating with each other to secure communication in IoT. For privileged access to services, the devices must be authenticated. The authorization mechanisms ensure that the access to systems or information is provided to the authorized ones. A proper implementation of authorization and authentication results in a trustworthy environment which ensures a secure environment for communication. Moreover, the accounting for resource usage, along with auditing and reporting provide a reliable mechanism for securing network management.

3. Availability of services:

The attacks on IoT devices may hinder the provision of services through the conventional denial-of-service attacks. Various strategies including the sinkhole attacks, jamming adversaries or the replay attacks exploit IoT components at different layers to deteriorate the quality-of-service(QoS) being provided to IoT users.

4. Energy efficiency:

The IoT devices are typically resource-constrained and are characterized with low power and less storage. The attacks on IoT architectures may result in an increase in energy consumption by flooding the network and exhausting IoT resources through redundant or forged service requests.

5. Single points of failure:

A continuous growth of heterogeneous networks for the IoT based infrastructure may expose a large number of single-points of-failure which may in turn deteriorate the services envisioned through IoT. It necessitates the development of a tamper-proof environment for a large number of IoT devices as well as to provide alternative mechanisms for implementation of a fault-tolerant network. [11]

2.1.2 Types of IDS based on attack type:

IDS

Attack type

HIDS

NIDS

Figure 2.2. Types of IDS based on attack type.

There are two types of intrusion detection systems based on attack type commonly known as Host based Intrusion Detection systems (HIDS) and Network based Intrusion Detection systems (NIDS). Network based intrusion detection systems are used to monitor and analyze network traﬃc to protect a system from network-based threats. Network based IDS aims at collecting information from the packet itself and looks at the contents of individual packets with the aim to detect the malicious activity in network traﬃc. Host based intrusion detection systems are a network security technology originally built for detecting vulnerability exploits against a target application or computer system. A HIDS aims to collect information about events or system calls/logs on a particular system.

2.1.3 Types of IDS based on approach:

IDS

Approach

Signature

Anomaly

Figure 2.3. Types of IDS based on approach

The two main types of IDS based on approach are signature-based and anomaly based. The signature based approach operates in much the same way as a virus scanner, by searching for identities or signatures of known intrusion events, while the anomaly based approach establishes a baseline of normal patterns. Anomaly based IDS allows the detection of unseen attacks, though resulting in higher false alarm rates but when paired with signature detection, can result in a powerful defense. [2]

2.1.4 Types of anomaly detectors:

There are two types of systems that are called anomaly detectors: those based upon a specification (or a set of rules) of what is regarded as “good/normal” behavior, and others that learn the behavior of a system under normal operation. The first type relies upon human expertise and may be regarded as a straightforward extension of typical misuse detection IDS systems. In the anomaly detection, the behavior of a system is automatically learned.

[9]

**2.2 Related Work**

Ebelechukwu N. et. al. propose an anomaly detection algorithm for detecting anomalous instances of sensor based events in an IoT device using provenance graphs. [5] 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.

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).

GARUDA introduces a novel distance measure that can be used to perform feature clustering and feature representation for efﬁcient intrusion detection. [13] 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.

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.

Bayu A. T. and Kyung H. R. have presented an effective anomaly detection by incorporating PSO-based feature selection and random forest model. [3] 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.

TagyAldeen M. et. al. have used random forest for detecting intrusions and then neural network classifier for the categorization of the detected intrusion. [15] The random forest classifier showed promising results, however the neural network showed consistently poor performance.

**2.3 Survey Conclusion**

There is a lot of research done on IDS, both signature based and anomaly based as well other novel methods. Most of the researches are done on the NSL-KDD dataset or the older KDDcup99 dataset. For anomaly based methods there is a trend of using data mining techniques and machine learning models to learn the normal behavior of the system and detecting the attacks based on outliers. Deep learning and artificial neural networks have been tested and give good results, however due to the higher overload, simpler methods like basic ML models are used. Research on provenance graphs has been done, however it has insufficient testing done and is at the moment theoretical. Cloud based systems have been tried, but has increase in time complexity as it will have to constantly send data to the cloud and time critical systems like health monitoring can suffer due to the given reason. The following method uses a hybrid system which is more accurate than the conventional ML models and more efficient.

**Chapter 3**

**Methodology**

The IDS acts as a second line of defense between the firewall and the IoT network. The traffic is filtered out of the firewall and passes through the IDS, where the Anomaly based detection system checks for anomalies and hence catches the intruders that made it through the firewall.

sensors

actuators

devices

IoT Network

IDS

Firewall

Gateway

Internet

Figure 3.1. Position of IDS in IoT architecture.

In [12], a hybrid machine learning model consisting of J48, Meta Pagging, Random Forest, REPTree, AdaBoostM1, Decision Stump and Naïve Bayes has been used. The NSL-KDD dataset has been used to train and test the model. The detection accuracy of the model has been tested on various attack type classifications.

**3.1 Functionality overview**

The diagram displays the overview of the process of detection of attacks with the help of the ensemble model. The diagram is followed by its description.

Pick dataset

Partition dataset

Preprocessing

Build hybrid model

Feature extraction

Optimization

Post processing

Figure 3.2. Functionality overview diagram

1. Picking the necessary dataset consisting of quality data such as the NSL-KDD.

2. Partitioning the dataset into 80% train and 20% test.

3. Preprocessing phase – involves the conversion of non-numerical values to numerical values, eliminating redundant data

4. Building a hybrid model consisting of the classifiers J48, Meta Pagging, Random Forest, REPTree, AdaBoostM1, Decision Stump, Naïve Bayes. The classification phase involves learning(supervised/unsupervised) and recognition/decision step.

5. Feature extraction – involves determining the right parameters that need to be used in the classification model, which can help improve the accuracy of the classification. VOTE scheme and Information Gain (IG) are used.

6. Develop the model that is optimized to provided the best accuracy and performance.

7. Correction methods are implemented in post processing phase to give better recognition rate.

**3.2 Confusion Matrix**

The confusion matrix is used to represent the information related to the actual and predicted classifications performed by the classification system.

Table 3.1. Confusion matrix.

|  |  |  |
| --- | --- | --- |
|  | Predicted Normal | Predicted Attack |
| Actual Normal | TN | FP |
| Actual Attack | FN | TP |

Where TN = True negatives (normal behavior predicted as normal),

FP = False positives (normal behavior predicted as attack),

FN = False negatives (attack predicted as normal behavior),

TP = True positives (attack predicted as attack).

Accuracy = (total number of correct predictions)

True positive rate = (correctly identified attacks)

False positive rate = (normal cases that have been incorrectly classified as attacks)

True negative rate = (normal cases that were correctly classified)

False negative rate = (attacks that have been classified incorrectly as normal)

**3.3 NSL-KDD dataset description**

NSL-KDD is a popular dataset used by researchers, commonly used to come up with effective IDSs. It is a refined version of KDDcup99 dataset. It is made up of a large amount of data. For the purpose of experiments, 10% of the main dataset was under consideration for training. It consists of 41 attributes representing different flow features. Each sample is labeled either normal or attack type. It takes into account the following protocols: UDP, TCP, ICMP.

Table 3.2. Overview of NSL-KDD data.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Data set type | No of data samples | | | | | |
| Records | Normal | DoS | Probe | U2R | R2L |
| NSL-KDD Train | 125973 | 67343 | 45927 | 11656 | 52 | 995 |
| % | 53.46 | 36.45 | 9.25 | 0.04 | 0.79 |
| NSL-KDD Test | 22543 | 9711 | 7458 | 2421 | 200 | 2754 |
| % | 43.08 | 33.08 | 10.74 | 0.89 | 12.22 |

**3.4 Classes based on Attack Types**

Aside from the normal data, records corresponding to 39 different attack types are found in the dataset. They can be categorized in 4 main attack types.

1. Denial of service attack(DoS) - when a legitimate user is denied access to the network by consuming memory or computer’s resources. Types – neptune, teardrop, ping of death, back, mail bomb, smurf, land
2. User-to-root attack(U2R) – when an adversary gets root access from a node within the network itself by exploiting the existing system weaknesses. Types – buffer overflow, load-module, rootkit, perl
3. Remote-to-local attack(R2L) – when an attacker who does not own an account remotely accesses a local machine account using machine vulnerabilities. Types – phf, warezclient, warezmaster, spy, ftp write, imap, multihop, guess passwd
4. Probing attack(PROBE) – when attacker dodges security and obtains data from the computers in the network. Types – nmap, ipsweep, satan, portsweep

Therefore classification is done based on 5 classes – Normal, DoS, Probe, R2L, U2R. Furthermore, the DoS class has 10 attack types, Probe has 6 types, R2L has 16 types, U2R has 7 types.

**3.5 Algorithm**

1. procedure model()

2. InputFn = NSL-KDD data set possessing 41 features f1, f2, f3… f42

3. Reduce 41 features to 8 features based on a number of the proposed filters

4. Use Vote scheme

5. Develop a robust model M

6. Propose the model

7. for every feature Fn

8. Provide Fn to J48, Meta pagging, RandomForest, REPTree, AdaBoostM1, DecisionStump and Naïve Bayes using NSL-KDDTrain+20%

9. Calculate

10. A1=J48 model accuracy

11. A2=Meta Pagging model accuracy

12. A3=RandomForest model accuracy

13. A4=REPTree model accuracy

14. A5=AdaBoostM1 model accuracy

15. A6=DecisionStump model accuracy

16. A7=NaiveBayes model accuracy

17. E=Ensemble representing J48, Meta Pagging, Random Forest, REPTree, AdaBoostM1, DecisionStump and NaiveBayes with NSL-KDDTrain+20%

18. Compare of the accuracy of A1, A2, A3, A4, A5, A6, A7, E

19. Select the best model M=E.

**Chapter 4**

**Discussion of Results**

Classification of 80% of the training data was done, using Naïve Bayes, J48, Random Forest, Hybrid model. The following table compares the true positives, false positives and accuracy rates. The classifier that shows the hybrid model has the highest percentage with 99.81% in terms of successfully classifying the instances. Thus, the comparison shows the hybrid model surpasses the other models in terms of accuracy.

The next table shows the comparison of the hybrid model with other classifiers in terms of multiclass classification as compared to the binary classification(normal/attack) in previous table. There as well, the hybrid model exhibited the highest percentages in all classes (99.7, 99.9, 96.2, 99.1, 97.9), thus proving it is superior to the other classifiers.

Table 4.1 Comparison of four classifiers

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Classifier | TP | FP | Correctly classified instance | Incorrectly classified instance |
| Naïve Bayes | 0.903 | 0.102 | 90.2876 | 9.7124 |
| J48 | 0.997 | 0.003 | 99.74 | 0.26 |
| Random Forest | 0.997 | 0.003 | 99.747 | 0.253 |
| Proposed Model | 0.997 | 0.003 | 99.81 | 0.25 |

Table 4.2 Accuracy in detection of normal and attack network flows by using the J48, SVM, Naïve Bayes and hybrid model classifier.

|  |  |  |
| --- | --- | --- |
| Classification Algorithm | Class Name | Test Accuracy |
| J48 | |  | | --- | | Normal | | DoS | | Probe | | U2R | | R2L | | |  | | --- | | 99.8 | | 99.1 | | 98.9 | | 98.7 | | 97.9 | |
| SVM | |  | | --- | | Normal | | DoS | | Probe | | U2R | | R2L | | |  | | --- | | 98.8 | | 98.7 | | 91.4 | | 94.6 | | 92.5 | |
| Naïve Bayes | |  | | --- | | Normal | | DoS | | Probe | | U2R | | R2L | | |  | | --- | | 74.9 | | 75.2 | | 74.1 | | 72.3 | | 70.1 | |
| Proposed Hybrid Model | |  | | --- | | Normal | | DoS | | Probe | | U2R | | R2L | | |  | | --- | | 99.7 | | 99.9 | | 96.2 | | 99.1 | | 97.9 | |

**Chapter 5**

**Conclusion and Future Work**

**5.1 Conclusion**

Security is integral in the development of IoT. With the heterogeneous nature of IoT consisting of various types of devices, different protocols used and massive amounts of information generated and the architectural constraints that exist, a second line of defense against adversaries is beneficial, which comes in the form of intrusion detection system. The hybrid efficient model proposed by Shadi A. et. al. [12], 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.

**5.2 Future Work**

The hybrid model is already very accurate and efficient, however, research and experiments on including this model into fog computing/ cloud computing networks can help reduce the resource and computational load on the IoT devices, where major processing for detection of attacks would be done away from the sensor devices that lack the power. The NSL-KDD dataset is very apt for research on IDS, however, more experiments on the newer, updated datasets such as the ADFA should be done in order to test and improve on the model performance.

**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] 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

[4] Dorothy E. D., "An Intrusion-Detection Model", IEEE Transactions on software engineering, vol SE-13, No. 2, 1987, 222-232

[5] 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

[6] 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

[7] 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

[8] John P., "The Evolution of Intrusion Detection/Prevention: Then, Now and the Future", https://www.secureworks.com/blog/the-evolution-of-intrusion-detection-prevention, 2017

[9] Ke W., Salvatore J. S., "Anomalous Payload-Based Network Intrusion Detection", Recent advances in intrusion detection: 7th international symposium, vol. 3224, 2004, pp 203-222

[10] Margaret R., "intrusion detection system (IDS)", https://searchsecurity.techtarget.com/definition/intrusion-detection-system, 2018

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

[12] 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

[13] 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

[14] Statista, "Internet of Things (IoT) connected devices installed base worldwide from 2015 to 2025 (in billions)", https://www.statista.com/statistics/471264/iot-number-of-connected-devices-worldwide/

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

[16] Wikipedia, "Internet of things", https://en.wikipedia.org/wiki/Internet\_of\_things#cite\_note-Linux\_Things-1