Machine learning-IoT 23

Jayant Singh Bains (jsbains@student.concordia.ab.ca)
Rahul Goyal (Rgoyal@student.concordia.ab.ca)
Hemanth Varma Kopanati (hkopanat@student.Concordia.ab.ca)
Bhargav Krishna Savaram (bsavaram@student.concordia.ab.ca)

ISSM 581: Research Project
Spring 2021

Research Project

Submitted to Faculty of Graduate Studies
Concordia University Of Edmonton

In Partial Fulfilment of the
Requirements of ISSM-581 Course

Concordia University Of Edmonton
FACULTY OF GRADUATE STUDIES
Edmonton , Alberta

Advisor: Dr. Sergey Butakov (sergey.butakov@concordia.ab.ca)
Department of Information Systems Security Management
Concordia University of Edmonton,
Edmonton T5B 4E4, Alberta, Canada

Machine learning-IoT 23

Jayant Singh Bains, Rahul Goyal, Hemanth Varma Kopanati, Bhargav Krishna Savaram

Approved:

Sergey Butakov [Original Approval on File]

Sergey Butakov Date: June 23, 2021

Primary Supervisor

Patrick Kamau [Original Approval on File]

Patrick Kamau, PhD, MCIC, PChem. Date: June 23, 2021

Dean, Faculty of Graduate Studies

Table of Contents

Lists of Figures

Lists of Tables

Machine learning-IoT 23

Jayant Singh Bains
Student ID:144103
jsbains@student.concordia.ab.ca

Hemanth Varma Kopanati
Student ID: 144357
hkopanat@student.Concordia.ab.ca

Rahul Goyal
Student ID: 143424
Rgoyal@student.concordia.ab.ca

Bhargav Krishna Savaram
Student ID: 142320
bsavaram@student.concordia.ab.ca

Advisor: Dr. Sergey Butakov
sergey.butakov@concordia.ab.ca

Department of Information Systems
Security Management
Concordia University of Edmonton,
Edmonton T5B 4E4, Alberta,
Canada

Abstract: IoT devices have become the future of many industries like household, security, etc. but these devices have a requirement that is that they should be connected to a network. This makes them vulnerable to outside attacks. If security is built into each device, it would cost more and would not reliable but if machine learning is used in the IDS and IPS to protect the system. This paper shows how machine learning can be used to detect malicious traffic in IoT datasets. The accuracy recorded ranges from 98.9%-100% depending on the dataset.

Keywords—Internet of Things (IoT), Intrusion detection systems, intrusion prevention system, Decision tree, Logistic regression, Artificial neural networks.

I. INTRODUCTION

Internet of Things (IoT) technology is penetrating the markets with unprecedented speed replacing traditional "dumb" equipment. Devices like sensors, intelligent controllers, smart TVs, smart wearables, etc. are all using IoT technology. The technology provides a great level of automating and information sharing to enhance usability and functionality [1]. Increasing the number of devices connected to the internet gives attackers more opportunities to tap on the data or remotely misuse IoT devices.

In 2018, there were 33,000 cases of cyber-attacks reported by the Canadian government, and the number keeps going up [2]. There is a constant arms race between attackers and defenders in cyberspace, but the attackers have a well-known advantage: while the defenders need to keep an eye on the entire security perimeter the attackers need to find only one gap to penetrate the protected system. In most scenarios where the internet is used, there is a remote interaction between the IoT device and the internet. This leads to more devices interacting and connecting to the network, having so many devices gives more ways for an attacker to attack a network if one of the devices is vulnerable the whole network becomes vulnerable.

MIRAI botnet was such an attack that compromised IoT devices and turned them into controlled bots. The main target for these botnets were cameras and home routers, after getting infected these devices scanned for IP addresses of IoT devices [3]. This attack left many IoT devices compromised, and it was one of the wakening calls

for the industry to pay closer attention to the security of IoT devices. More recent attacks happened in 2020 where Ubiquiti known for its IoT devices had a breach that compromised the personal information of their customers. The hacker(s?) got access by getting the credentials of an employee in the company and got access to all amazon web services (AWS) accounts.

Compromised IoT devices can be used to spy and violate people's privacy. The traffic that needs to be monitored is so huge that traditional security is not enough to protect IoT devices. The rise in attacks on IoT devices has been steadily growing, this leads to IoT devices vulnerable to many types of attacks that can threaten the user of IoT devices or the functionality of IoT devices themselves.

One potential solution to this continuous problem is to implement machine learning to automate the monitoring of security of IoT networks.

II. LITERATURE REVIEW

A. *Traditional Methods for Network Traffic Analysis*

The most common network traffic analysis methods involved the use of Intrusion Detection Systems (IDS) and the use of packet inspection.

1) Intrusion Detection Systems

The unauthorized access that affects the organization's information confidentiality, integrity, and availability is an intrusion [4]. The purpose of the IDS is not only to prevent the attack but also to identify and report. There are three learning techniques according to intrusion detection system. They are as follows:

i) Signature Based Detection:

Signature-based intrusion detection system uses known patterns or a signature of the malicious traffic to identify the attack traffic. The known patterns are stored in a database which includes the collection of suspicious activities and operations that can exploit the weaknesses of the information systems. In this technique, the pattern of the incoming traffic is compared with the pattern stored in the database to differentiate the attack traffic from the legitimate traffic [4]. The SNORT tool is the best example of a signature-based intrusion detection.

ii) Anomaly Based Detection:

The behavior of the network is an important parameter upon which the anomaly detection system relies. If the behavior of the network is within predefined behavior, the network will be allowed, or it triggers an alert in the Intrusion Detection System. The acceptable performance of the network can be predetermined, conditions set by the administrator or learnt from specifications. However, if the malicious traffic fall under the acceptable behavior, the traffic might get unnoticed [5].

This type of system can be attached to both network-based intrusion detection systems (NIDS) and host-based intrusion detection systems (HIDS). Defining the rules is the main drawback of anomaly-based IDS.

iii) Rule Based Detection:

`Rule-based systems are used as the core of Intrusion Detection Systems. These systems are based on three components such as facts, rules, and inference. Whereas facts base contains facts on the states of the system. Rule base contains scenarios of rules.

The inference engines search the facts for that a rule is expected. The consequent action will be taken it finds any match. There is a number of ways to get the rules that describe the behavior of the user [6].

2) Packet inspection:

i) Stateful packet inspection

Stateful packet filtering only reads the header of the packet. It is done by using firewalls. By using deep packet inspection, we can overcome the drawbacks of stateful packet inspection [7].

ii) Deep packet inspection

Deep packet inspection is a data extraction or packet inspection method. Deep packet inspection carries out the information from both header and data part of the packet at a particular examination point. It checks for all 7 fields to allow or deny the packet from passing over the examination point. It makes

decisions about a specific packet should be dropped or forwarded to the destination. Continuously, it checks the contents of the network packets as per the guidelines or list of malicious signatures stored in a database assigned to the devices by the service providers or the network administrators [7].

B. *Approaches for detecting the IoT Malicious Traffic Using Machine Learning Techniques*

There are various approaches for malicious traffic detection which can be classified into three categories, they are supervised learning, unsupervised learning, and reinforcement learning.

1) *Supervised Learning*

In this type of learning, the algorithm is given a completely labeled dataset which can be used to test the accuracy of training data with respect to prior data. This model uses the training data to learn and defines its own patterns to determine whether the right outcome is achieved or not. The testing data can be used to test the model to see how accurately it learned during the training [8].
Supervised learning can be categorized into two:

a) *Classification*

Classification is used when predicting variables that falls within a group of values. Classification can either binary or multiclass. If there are only 2 options, then it is called binary classification else Multi-Class Classification when there are more than 2 options.

b) *Regression*

Regression is used when predicting the output variable as a number like height, weight, any prices, currency, and any other values. Few examples of regression are Linear Regression, Random Forest regressor, and Support vector machines (SVM) [9].

2) *Unsupervised Learning*

In this type of learning, the model is trained with data that is neither classified nor regressed. This technique then attempts to group the non-classified data by extracting useful patterns and features based on similarities and differences. They can be subdivided into Clustering and Dimensionality reduction problems [10].

3) *Reinforcement Learning*

In this type of learning, the aim is to monitor the environment, the Internet in this case continuously, and use the knowledge gained to improve the further performance of the model. The model works this way by trial and error through observation of the surrounding environment to achieve the final goal [11].

Classifier	Accuracy	Precision	Recall	F1-Score
NB	99.92	99.92	99.92	99.92
KNN	99.94	99.94	99.94	99.94
RF	100	100	100	100
LR	99.91	99.91	99.91	99.91

Table 1 Results of Mirai Attack detection on IoT-23 dataset [12]

Metrics	Classifiers RF	NB	ANN	SVM	ADA
Precision weighted	1.00	0.76	0.71	0.60	0.86
Macro	0.88	0.27	0.33	0.23	0.55
Recall weighted	1.00	0.23	0.66	0.67	0.87
Macro	0.85	0.38	0.14	0.14	0.35
F-1 score weighted	1.00	0.25	0.52	0.59	0.83
Macro	0.84	0.10	0.10	0.13	0.37
Accuracy	1.00	0.23	0.66	0.67	0.87

Table 2 The Results of the Classifier [13]

Tables 1 and 2 summarize the overall results acquired on testing the machine learning classifiers trained over the proposed universal features set for detecting different attacks across the dataset. It can be noticed that the RF classifier performed best for detecting the botnet attacks in the dataset.

All the learnings mentioned here used very large datasets. Many research groups attempted to create IoT datasets to serve as a common ground for researchers to improve the performance of ML models for malicious traffic detection in IoT networks. For example, the IoT-23 dataset, TLESS dataset, CGIAR dataset, etc. This research looked specifically at IoT-23 dataset in the attempt to build improved ML-based model that would allow reliable detection of traffic from infected devices.

III. METHODOLOGY

IoT-23 Dataset will be used to build a classification model which would classify the IoT network traffic as either malign or benign.

Modeling Unbalanced Classes

Classification algorithms are built to optimize accuracy, which makes it challenging to create a model when there is not a balance across the number of observations of different classes. Common methods that could be used to approach balancing in the classes are:

- Down sampling or removing observations from the most common class
- Upsampling or duplicating observations from the rarest class or classes
- A mix of downsampling and upsampling.

Exploratory Data Analysis and Data Cleaning.

Data Analysis and Cleaning is important because messy data will lead to unreliable outcomes. Some common issues that make data messy are duplicate or unnecessary data, inconsistent data and typos, missing data, outliers, and data source issues.

The following steps will be performed to analyze and clean the data.

- Duplicate or unnecessary data will be deleted.
- Row with missing data will be either removed or the missing data will be replaced with the mean value of the particular feature.
- A column with variance less than 10% will be deleted.
- A column with skewness will be transformed using log function.
- Pair plot will be used to visualize the distribution of different attributes.
- Outliers will be either removed or its impact will be reduced by performing the transformation. Use of outlier resistant model will be the priority.
- If the required dimension of the data will also be reduced using Principal Component Analysis (PCA).
- Before feeding the data into the model, data will be scaled using **StandardScaler** of scikit- learn library

Model Building:

Training and Test Splits

Splitting the data into a training and a test set can help in choosing a model that has better chances at generalizing and is not overfitted. The training data is used to fit the model, while the test data is used to measure error and performance. stratified sampling method will be used for splitting the data into training and test data sets. stratified sampling method will help in getting a similar class balance in both train and test datasets.

Model Selection and Hyperparameter tuning

Machine learning algorithm:

Following algorithms will be used to build separate Machine learning models and the model which gives the best performance will be selected for final submission. Stratified Cross validation approach will be used to evaluate and select the best machine learning algorithm for the given IoT 23 Dataset,

- Logistic Regression
- K-Nearest Neighbors
- Decision Trees
- Support Vector Machines
- Random Forests
- Neural Networks

Hyperparameters of the models will be tuned using Randomised Grid Search Cross-Validation feature available in scikit- learn library.

Model Evaluation:

As efforts are being taken to make the dataset a balanced one, the accuracy will be a good choice to evaluate its performance. Along with accuracy, f1 score will also be used to verify the model performance. Once the model is trained and its performance meets minimum success criteria, the model could also be used for an Imbalanced dataset.

IV. VALIDATION OF THE KNOWN MODELS

IoT-23 dataset has 23 unbalanced datasets. The validation of the code was performed on a balanced dataset given in the cited project report [14]. The dataset has 13 labels. The code uses 2 datasets, one to train the model and the other one to test it.

A. Preprocessing

Preprocessing was done on the data sets to create 3 datasets. In this preprocessing the unimportant columns were removed, the sections with missing

values were changed with mean values. Also, the columns with categorical datatypes were transformed into numerical datatypes using label encoder. Feature scaling was performed using the standard scaler technique to standardize the range of the data to be compared on common grounds [14].

B. Models

4 models were used to validate the results, these models were decision tree classifier, logistic regression, random classifier, and artificial neural networks. All 3 datasets were used to train and test these models and validated the results given by previous models' execution [14].

V. Model Implementation and Results

A. Preprocessing

1) binarization of label with multi categories

The Datasets which are used are transformed such that the column containing the important categories such as benign or malicious is binarized and a new column is created where the value 0 represents the benign traffic and value 1 represents all the other malicious traffic. this is achieved using label encoder that gives values to all the categories in a label column. Once the value is given a function is run where the argument is passed which makes benign value as 0 and all the other value greater than 0 is given the value 1.

2) Cleaning and scaling the data

The dataset was cleaned by removing all the unnecessary columns, by replacing all the empty blocks with mean values, binarization was done on the label column but other columns were also converted to numerical datatypes. After the data was satisfactory the data were then scaled by using the standard scaler method. The data was then split into 2 datasets for training and testing.

B. Model implementation

1) Malware_1_1 Dataset

This dataset consists of 3 categories of malicious and benign data and they are Benign, C&C, PartOfAHorizontalPortScan.
Logistic regression is used on this dataset and an accuracy of 98.9% is achieved. The precision, recall, and F1 scores are shown in Figure 1.

```
Precision: 0.99022
Recall: 0.98858
F1-score: 0.98929
```

Figure 1 Classification report for Malware_1_1 dataset

The confusion matrix for this model is shown in Figure 2.

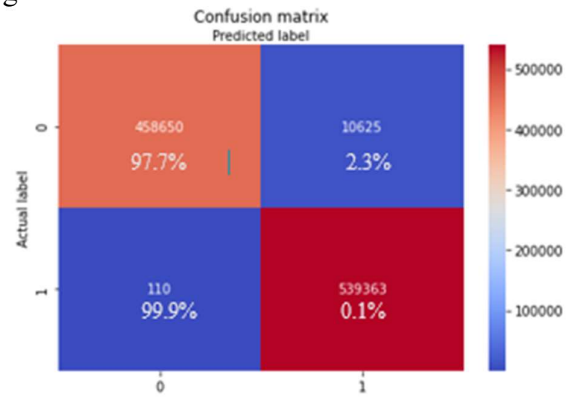

Figure 2 Confusion matrix for Malware_1_1 dataset

2) Malware_3_1 Dataset

This data consists of 4 categories of malicious and benign data. The label consists of Attack, Benign, C&C, PartOfAHorizontalPortScan.
Artificial neural networks are used on this dataset to achieve an accuracy of 99.9%. The classification report is shown in Figure 3.

	precision	recall	f1-score	support
0	1.00	0.98	0.99	4536
1	1.00	1.00	1.00	151567
accuracy			1.00	156103
macro avg	1.00	0.99	1.00	156103
weighted avg	1.00	1.00	1.00	156103

Figure 3 Classification report for Malware_3_1 dataset

The confusion matrix for artificial neural networks is shown in Figure 4.

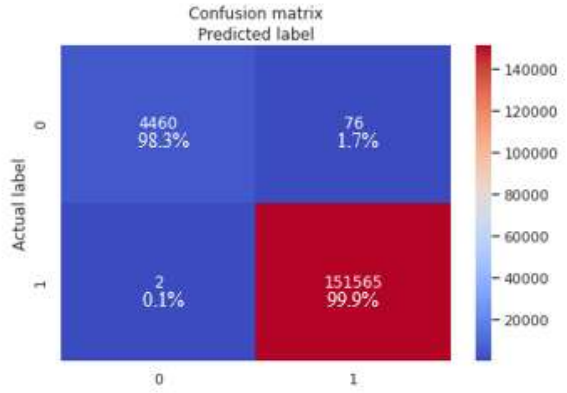

Figure 4 Confusion matrix for Malware_3_1 dataset

3) Malware_7_1 Dataset

Malware_7_1 dataset has 4 labels and they are Benign, C&C Hearbeat, Okiru, DDoS.
Decision tree classifier is used on this dataset to achieve an accuracy of 100%. The classification report is shown in Figure 5.

	precision	recall	f1-score	support
0	1.00	1.00	1.00	1344
1	1.00	1.00	1.00	1047224
micro avg	1.00	1.00	1.00	1048568
macro avg	1.00	1.00	1.00	1048568
weighted avg	1.00	1.00	1.00	1048568
samples avg	1.00	1.00	1.00	1048568

Figure 5 Classification report for Malware_7_1 dataset

4) Malware_8_1 Dataset

Malware_8_1 dataset has 2 labels and they are benign and C&C.
Random forest classifier is used on this dataset to achieve an accuracy of 100%. The classification report is shown in Figure 6.

	precision	recall	f1-score	support
0	1.00	1.00	1.00	2181
1	1.00	1.00	1.00	8222
micro avg	1.00	1.00	1.00	10403
macro avg	1.00	1.00	1.00	10403
weighted avg	1.00	1.00	1.00	10403
samples avg	1.00	1.00	1.00	10403

Figure 6 Classification report for Malware_8_1 dataset

5) Malware_9_1 Dataset

Malware_9_1 dataset has 2 labels and they are benign and PartOfAHorizontalPortScan .
Logistic regression is used on this dataset to achieve an accuracy of 99.9%. The classification report is shown in Figure 7.

```
Precision: 1.00000
Recall: 0.99922
F1-score: 0.99961
```

Figure 7 Classification report for Malware_9_1 dataset

The confusion matrix for Logistic regression is shown in Figure 8.

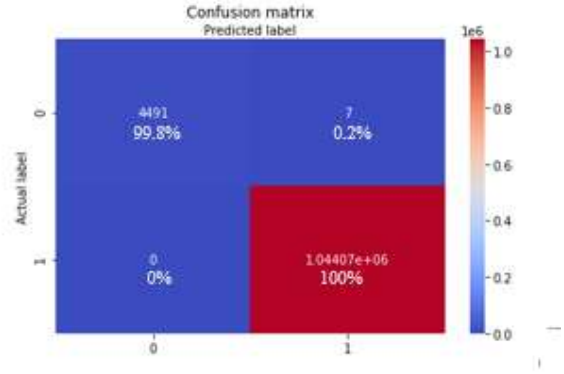

Figure 8 Confusion matrix for Malware_9_1 dataset

6) Malware_20_1 Dataset

This dataset consists of 2 categories of malicious and benign data and they are Benign and C&C-Torii.
Logistic regression is used on this dataset and an accuracy of 100% is achieved. The precision, recall, and F1 scores are shown in Figure 9.

	precision	recall	f1-score	support
0	1.00	1.00	1.00	3193
1	1.00	1.00	1.00	16
micro avg	1.00	1.00	1.00	3209
macro avg	1.00	1.00	1.00	3209
weighted avg	1.00	1.00	1.00	3209
samples avg	1.00	1.00	1.00	3209

Figure 9 Classification report for Malware_20_1 dataset

7) Malware_21_1 Dataset

This dataset consists of 2 categories of malicious and benign data and they are Benign and C&C-Torii.

Decision tree classifier is used on this dataset and an accuracy of 100% is achieved. The precision, recall, and F1 scores are shown in Figure 10.

	precision	recall	f1-score	support
0	1.00	1.00	1.00	3272
1	1.00	1.00	1.00	14
micro avg	1.00	1.00	1.00	3286
macro avg	1.00	1.00	1.00	3286
weighted avg	1.00	1.00	1.00	3286
samples avg	1.00	1.00	1.00	3286

Figure 10 Classification report for Malware_21_1 dataset

8) Malware_60_1 Dataset

Malware_60_1 dataset has 3 labels and they are benign, C&C Heartbeat, and DDoS.
Decision tree classifier is used on this dataset to achieve an accuracy of 100%. The classification report is shown in Figure 11.

	precision	recall	f1-score	support
0	1.00	1.00	1.00	700
1	1.00	1.00	1.00	1047868
micro avg	1.00	1.00	1.00	1048568
macro avg	1.00	1.00	1.00	1048568
weighted avg	1.00	1.00	1.00	1048568
samples avg	1.00	1.00	1.00	1048568

Figure 11 Classification report for Malware_60_1 dataset

9) Malware_49_1 Dataset

This dataset consists of 4 categories of malicious and benign data and they are Benign, C&C, C&C-FileDownload, and PartOfAHorizontalPortScan.
Random forest classifier is used on this dataset and an accuracy of 100% is achieved. The precision, recall, and F1 scores are shown in Figure 12.

	precision	recall	f1-score	support
0	1.00	1.00	1.00	1371
1	1.00	1.00	1.00	1047197
micro avg	1.00	1.00	1.00	1048568
macro avg	1.00	1.00	1.00	1048568
weighted avg	1.00	1.00	1.00	1048568
samples avg	1.00	1.00	1.00	1048568

Figure 12 Classification report for Malware_49_1 dataset

10) Malware_48_1 Dataset

Malware_48_1 dataset has 6 labels and they are benign, Attack, C&C Heartbeat-attack, C&C Heartbeat -FileDownload, C&C PartOfAHorizontalPortScan ,and PartOfAHorizontalPortScan .

Artificial neural networks are used on this dataset to achieve an accuracy of 100%. The classification report is shown in Figure 13.

	precision	recall	f1-score	support
0	1.00	1.00	1.00	1371
1	1.00	1.00	1.00	1047197
accuracy			1.00	1048568
macro avg	1.00	1.00	1.00	1048568
weighted avg	1.00	1.00	1.00	1048568

Figure 13 Classification report for Malware_48_1 dataset

11) Malware_44_1 Dataset

This dataset consists of 4 categories of malicious and benign data and they are Benign, C&C, C&C-FileDownload, and DDoS.
Logistic regression is used on this dataset and an accuracy of 100% is achieved. The precision, recall, and F1 scores are shown in Figure 14.

```
Precision: 1.00000
Recall: 1.00000
F1-score: 1.00000
```

Figure 14 Classification report for Malware_44_1 dataset

The confusion matrix for Logistic regression is shown in Figure 15.

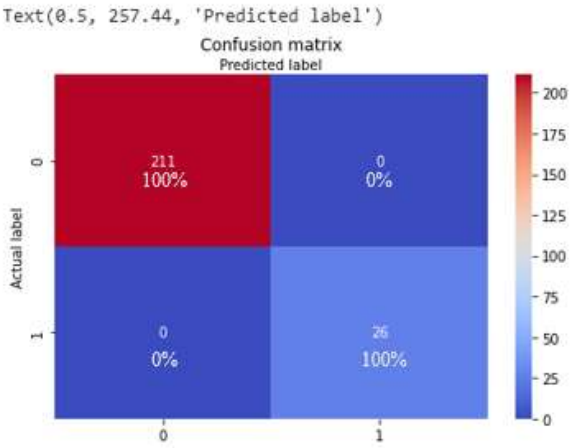

Figure 15 Confusion matrix for Malware_44_1 dataset

12) Malware_42_1 Dataset

This dataset consists of 4 categories of malicious and benign data and they are Benign, C&C, C&C-FileDownload, and FileDownload.
Decision tree classifier is used on this dataset and an accuracy of 100% is achieved. The precision, recall, and F1 scores are shown in Figure 16.

	precision	recall	f1-score	support
0	1.00	1.00	1.00	4420
1	1.00	1.00	1.00	6
micro avg	1.00	1.00	1.00	4426
macro avg	1.00	1.00	1.00	4426
weighted avg	1.00	1.00	1.00	4426
samples avg	1.00	1.00	1.00	4426

Figure 16 Classification report for Malware_42_1 dataset

13) Malware_36_1 Dataset

This dataset consists of 4 categories of malicious and benign data, and they are Benign, C&C-HeartBeat, Okiru, and Okiru-Attack. Random forest classifier is used on this dataset and an accuracy of 100% is achieved. The precision, recall, and F1 scores are shown in Figure 17.

	precision	recall	f1-score	support
0	1.00	1.00	1.00	188
1	1.00	1.00	1.00	1048380
micro avg	1.00	1.00	1.00	1048568
macro avg	1.00	1.00	1.00	1048568
weighted avg	1.00	1.00	1.00	1048568
samples avg	1.00	1.00	1.00	1048568

Figure 17 Classification report for Malware_36_1 dataset

14) Malware_35_1 Dataset

This dataset consists of 5 categories of malicious and benign data, and they are Benign, C&C-FileDownload, C&C, Attack, and DDoS. Artificial neural networks are used on this dataset and an accuracy of 99.9% is achieved. The precision, recall, and F1 scores are shown in Figure 18.

	precision	recall	f1-score	support
0	1.00	1.00	1.00	1048476
1	1.00	0.86	0.92	92
accuracy			1.00	1048568
macro avg	1.00	0.93	0.96	1048568
weighted avg	1.00	1.00	1.00	1048568

Figure 18 Classification report for Malware_35_1 dataset

The confusion matrix for artificial neural networks is shown in Figure 19.

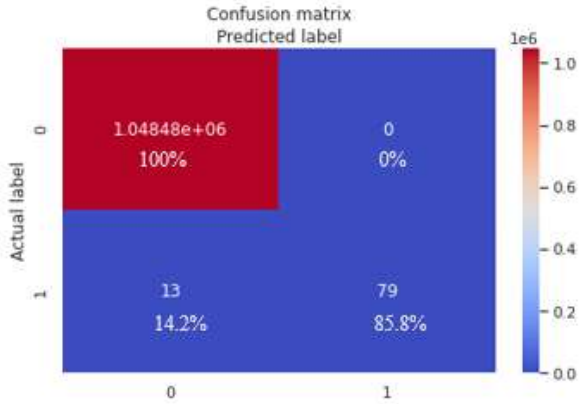

Figure 19 Confusion matrix for Malware_35_1 dataset

15) Malware_34_1 Dataset

This dataset consists of 4 categories of malicious and benign data and they are Benign, PartOfAHorizontalPortScan, C&C, and DDoS. Logistic regression is used on this dataset and an accuracy of 99.8% is achieved. The precision, recall, and F1 scores are shown in Figure 20.

```
Precision: 0.99778
Recall: 0.99000
F1-score: 0.99386
```

Figure 20 Classification report for Malware_34_1 dataset

The confusion matrix for this model is shown in Figure 21.

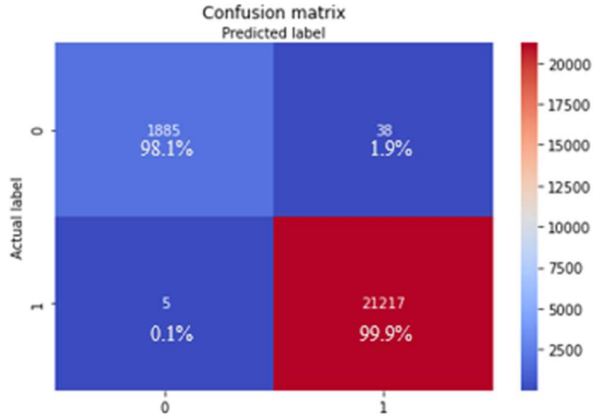

Figure 21 Confusion matrix for Malware_34_1 dataset

16) Malware_52_1 Dataset

This dataset consists of 4 categories of malicious and benign data and they are Benign, C&C-FileDownload, C&C, and PartOfAHorizontalPortScan.

Artificial neural networks are used on this dataset and an accuracy of 100% is achieved. The precision, recall, and F1 scores are shown in Figure 22.

	precision	recall	f1-score	support
0	1.00	1.00	1.00	1794
1	1.00	1.00	1.00	19779584
micro avg	1.00	1.00	1.00	19781378
macro avg	1.00	1.00	1.00	19781378
weighted avg	1.00	1.00	1.00	19781378
samples avg	1.00	1.00	1.00	19781378

Figure 22 Classification report for Malware_52_1 dataset

17) Malware_17_1 Dataset

This dataset consists of 7 categories of malicious and benign data and they are Benign, Attack, DDoS, C&C-Hearbeat, PartOfAHorizontalPortScan, PartOfAHorizontalPortScan-Attack and Okiru. Random forest classifier is used on this dataset. The dataset being very large is split into 3 datasets and the accuracy recorded ranges from 99.9% - 100%. The precision, recall, and F1 scores are shown in Figures 23, 24 and 25.

	precision	recall	f1-score	support
0	1.00	1.00	1.00	9969
1	1.00	1.00	1.00	18210042
micro avg	1.00	1.00	1.00	18220011
macro avg	1.00	1.00	1.00	18220011
weighted avg	1.00	1.00	1.00	18220011
samples avg	1.00	1.00	1.00	18220011

Figure 23 Classification report for Malware_17_1 dataset (part 1 of the split dataset)

	precision	recall	f1-score	support
0	1.00	1.00	1.00	9015
1	1.00	1.00	1.00	15545420
micro avg	1.00	1.00	1.00	15554435
macro avg	1.00	1.00	1.00	15554435
weighted avg	1.00	1.00	1.00	15554435
samples avg	1.00	1.00	1.00	15554435

Figure 24 Classification report for Malware_17_1 dataset (part 2 of the split dataset)

	precision	recall	f1-score	support
0	1.00	1.00	1.00	9015
1	1.00	1.00	1.00	15545420
micro avg	1.00	1.00	1.00	15554435
macro avg	1.00	1.00	1.00	15554435
weighted avg	1.00	1.00	1.00	15554435
samples avg	1.00	1.00	1.00	15554435

Figure 25 Classification report for Malware_17_1 dataset (part 3 of the split dataset)

18) Malware_33_1 Dataset

This dataset consists of 4 categories of malicious and benign data and they are Benign, C&C-Hearbeat, PartOfAHorizontalPortScan, and Okiru-Attack.
Random forest classifier is used on this dataset. The dataset being very large is split into 4 datasets and the accuracy recorded ranges from 99.9% - 100%. The precision, recall, and F1 scores are shown in Figures 26, 27, 28 and 29.

	precision	recall	f1-score	support
0	1.00	1.00	1.00	346278
1	1.00	1.00	1.00	13267282
micro avg	1.00	1.00	1.00	13613560
macro avg	1.00	1.00	1.00	13613560
weighted avg	1.00	1.00	1.00	13613560
samples avg	1.00	1.00	1.00	13613560

Figure 26 Classification report for Malware_33_1 dataset (part 1 of the split dataset)

	precision	recall	f1-score	support
0	1.00	1.00	1.00	346278
1	1.00	1.00	1.00	13267282
micro avg	1.00	1.00	1.00	13613560
macro avg	1.00	1.00	1.00	13613560
weighted avg	1.00	1.00	1.00	13613560
samples avg	1.00	1.00	1.00	13613560

Figure 27 Classification report for Malware_33_1 dataset (part 2 of the split dataset)

	precision	recall	f1-score	support
0	1.00	1.00	1.00	346297
1	1.00	1.00	1.00	13267289
micro avg	1.00	1.00	1.00	13613586
macro avg	1.00	1.00	1.00	13613586
weighted avg	1.00	1.00	1.00	13613586
samples avg	1.00	1.00	1.00	13613586

Figure 28 Classification report for Malware_33_1 dataset (part 3 of the split dataset)

	precision	recall	f1-score	support
0	1.00	1.00	1.00	346297
1	1.00	1.00	1.00	13267289
micro avg	1.00	1.00	1.00	13613586
macro avg	1.00	1.00	1.00	13613586
weighted avg	1.00	1.00	1.00	13613586
samples avg	1.00	1.00	1.00	13613586

Figure 29 Classification report for Malware_33_1 dataset (part 4 of the split dataset)

19) Malware_39_1 Dataset

This dataset consists of 4 categories of malicious and benign data and they are Benign, C&C, PartOfAHorizontalPortScan, and Attack. Random forest classifier is used on this dataset. The dataset being very large is split into 4 datasets and the accuracy recorded ranges from 99.9% - 100%. The precision, recall, and F1 scores are shown in Figures 30, 31, 32, 33.

	precision	recall	f1-score	support
0	1.00	1.00	1.00	346297
1	1.00	1.00	1.00	13267289
micro avg	1.00	1.00	1.00	13613586
macro avg	1.00	1.00	1.00	13613586
weighted avg	1.00	1.00	1.00	13613586
samples avg	1.00	1.00	1.00	13613586

Figure 30 Classification report for Malware_39_1 dataset (part 1 of the split dataset)

	precision	recall	f1-score	support
0	1.00	1.00	1.00	965
1	1.00	1.00	1.00	8257631
micro avg	1.00	1.00	1.00	8258596
macro avg	1.00	1.00	1.00	8258596
weighted avg	1.00	1.00	1.00	8258596
samples avg	1.00	1.00	1.00	8258596

Figure 31 Classification report for Malware_39_1 dataset (part 2 of the split dataset)

	precision	recall	f1-score	support
0	1.00	1.00	1.00	836
1	1.00	1.00	1.00	9409540
micro avg	1.00	1.00	1.00	9410376
macro avg	1.00	1.00	1.00	9410376
weighted avg	1.00	1.00	1.00	9410376
samples avg	1.00	1.00	1.00	9410376

Figure 32 Classification report for Malware_39_1 dataset (part 3 of the split dataset)

	precision	recall	f1-score	support
0	1.00	1.00	1.00	1034
1	1.00	1.00	1.00	8331085
micro avg	1.00	1.00	1.00	8332119
macro avg	1.00	1.00	1.00	8332119
weighted avg	1.00	1.00	1.00	8332119
samples avg	1.00	1.00	1.00	8332119

Figure 33 Classification report for Malware_39_1 dataset (part 4 of the split dataset)

20) Malware_43_1 Dataset

This dataset consists of 7 categories of malicious and benign data and they are Benign, C&C, C&C FileDownload, Okiru, DDoS, and PartOfAHorizontalPortScan. Random forest classifier is used on this dataset. The dataset being very large is split into 7 datasets and the accuracy recorded ranges from 99.9% - 100%. The precision, recall, and F1 scores are shown in Figures 34, 35, 36, 37, 38, 39, 40.

	precision	recall	f1-score	support
0	1.00	1.00	1.00	2577514
1	1.00	1.00	1.00	5840121
micro avg	1.00	1.00	1.00	8417635
macro avg	1.00	1.00	1.00	8417635
weighted avg	1.00	1.00	1.00	8417635
samples avg	1.00	1.00	1.00	8417635

Figure 34 Classification report for Malware_43_1 dataset (part 1 of the split dataset)

	precision	recall	f1-score	support
0	1.00	1.00	1.00	2579467
1	1.00	1.00	1.00	5838565
micro avg	1.00	1.00	1.00	8418032
macro avg	1.00	1.00	1.00	8418032
weighted avg	1.00	1.00	1.00	8418032
samples avg	1.00	1.00	1.00	8418032

Figure 35 Classification report for Malware_43_1 dataset (part 2 of the split dataset)

	precision	recall	f1-score	support
0	1.00	1.00	1.00	2568343
1	1.00	1.00	1.00	5849425
micro avg	1.00	1.00	1.00	8417768
macro avg	1.00	1.00	1.00	8417768
weighted avg	1.00	1.00	1.00	8417768
samples avg	1.00	1.00	1.00	8417768

Figure 36 Classification report for Malware_43_1 dataset (part 3 of the split dataset)

	precision	recall	f1-score	support
0	1.00	1.00	1.00	2580286
1	1.00	1.00	1.00	5837681
micro avg	1.00	1.00	1.00	8417967
macro avg	1.00	1.00	1.00	8417967
weighted avg	1.00	1.00	1.00	8417967
samples avg	1.00	1.00	1.00	8417967

Figure 37 Classification report for Malware_43_1 dataset (part 4 of the split dataset)

	precision	recall	f1-score	support
0	1.00	1.00	1.00	2544970
1	1.00	1.00	1.00	5851746
micro avg	1.00	1.00	1.00	8396716
macro avg	1.00	1.00	1.00	8396716
weighted avg	1.00	1.00	1.00	8396716
samples avg	1.00	1.00	1.00	8396716

Figure 38 Classification report for Malware_43_1 dataset (part 5 of the split dataset)

	precision	recall	f1-score	support
0	1.00	1.00	1.00	2574722
1	1.00	1.00	1.00	5843135
micro avg	1.00	1.00	1.00	8417857
macro avg	1.00	1.00	1.00	8417857
weighted avg	1.00	1.00	1.00	8417857
samples avg	1.00	1.00	1.00	8417857

Figure 39 Classification report for Malware_43_1 dataset (part 6 of the split dataset)

	precision	recall	f1-score	support
0	1.00	1.00	1.00	2576794
1	1.00	1.00	1.00	5841169
micro avg	1.00	1.00	1.00	8417963
macro avg	1.00	1.00	1.00	8417963
weighted avg	1.00	1.00	1.00	8417963
samples avg	1.00	1.00	1.00	8417963

Figure 40 Classification report for Malware_43_1 dataset (part 7 of the split dataset)

C. *Summary table with results*

As it can be seen from the evaluation models were able to achieve acceptable results on full datasets. 99.9% accuracy on malware recognition would allow effectively suppress traffic from infected devices. 98% accuracy of recognition on the benign traffic would lead to rare cases when benign traffic would not reach the target. This may be tolerable for consumer-level devices and would require special attention in critical services such as remote sensors in industrial applications.

IoT datasets	Precision	recall	F1-Score	Accuracy
Malware_1_1	99%	98.8%	98.9%	98.9%
Malware_3_1	100%	98%	99%	99.9%
Malware_7_1	100%	100%	100%	100%
Malware_8_1	100%	100%	100%	100%
Malware_9_1	100%	99.9%	99.9%	99.9%
Malware_20_1	100%	100%	100%	100%
Malware_21_1	100%	100%	100%	100%
Malware_60_1	100%	100%	100%	100%
Malware_49_1	100%	100%	100%	100%
Malware_48_1	100%	100%	100%	100%
Malware_44_1	100%	100%	100%	100%
Malware_42_1	100%	100%	100%	100%
Malware_36_1	100%	100%	100%	100%
Malware_35_1	99.9%	100%	86%	92%
Malware_34_1	99.8%	99.7%	99%	99.3%
Malware_52_1	100%	100%	100%	100%
Malware_17_1	100%	100%	100%	100%
	100%	100%	100%	100%
	100%	100%	100%	100%
Malware_33_1	100%	100%	100%	100%
	100%	100%	100%	100%
	100%	100%	100%	100%
	100%	100%	100%	100%
Malware_39_1	100%	100%	100%	100%
	100%	100%	100%	100%
	100%	100%	100%	100%
	100%	100%	100%	100%
Malware_43_1	100%	100%	100%	100%
	100%	100%	100%	100%
	100%	100%	100%	100%
	100%	100%	100%	100%
	100%	100%	100%	100%
	100%	100%	100%	100%
	100%	100%	100%	100%

Table 3 results of all 20 datasets

CONCLUSION

More devices on the network mean more vulnerable access points for attackers and due to increasing numbers of IoT devices, the attacks targeting IoT devices have also increased. Due to continuously expanding zoo of IoT devices and growing application areas for IoT traditional rule-based methods may lag behind in terms of malware traffic detection. Experiments with ML-based detection models indicate the feasibility to use them to protect IoT devices. Accuracy in 98%-99.9% range allows effective suppression of the attacks with some occasional false positives that would be typically taken care of by the upper-level protocol. Using machine learning in IDS/IPS will

increase the speed of detection of malicious traffic, will lessen the time of response to an attack, and increase the security overall. IDS/IPS using machine learning will be able to handle large influxes of traffic and detect malicious traffic with respectable accuracy.

VI. BIBLIOGRAPHY

[1] J.Pike, "internet of things - standard of things," 2014.

[2] Statistics Canada, "Cybercrime in Canada," 2019. [Online]. Available: https://www150.statcan.gc.ca/n1/pub/89-28-0001/2018001/article/00015-eng.htm.

[3] Zeifman, Igal, Bekerman, Dima and Herzberg, Ben, "Breaking Down Mirai: An IoT DDoS Botnet Analysis," 2016.

[4] M. K. Asif, Umar Naeem, Sufyan Yakoob and Talha A. Khan, "Network Intrusion Detection and its strategic," 2013.

[5] . K. Munivara, J. Veeramreddy and P. , "Anomaly-Based Intrusion Detection System," 2019.

[6] G. a. Mikhail, "Intrusion Detection: Techniques and Approaches".

[7] C. Brook, "What is Deep Packet Inspection? How It Works, Use Cases for DPI, and More," 5 December 2018.

[8] Data Robot, "supervised machine learning," [Online]. Available: https://www.datarobot.com/wiki/supervised-machine-learning/.

[9] J. Brownlee, "Supervised and Unsupervised machine learning algorithm," [Online]. Available: https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/.

[10] S. Shai and B. Shai, "Understanding Machine Learning: From Theory to Algorithm," *Cambridge University press,* 2014.

[11] K. Ansam and A. Ammar, "A critical review of intrusion detection," 2021.

[12] F. Hussain, G. A. Syed, U. Ubaid, A. Ghlib, T. Abdullah and A. Ahmad, "Towards a Universal Features Set for IoT Botnet Attacks Detection," 1 Dec 2020.

[13] N.-A. Stotian, "Machine Learning for Anomaly Detection in IoT networks:Malware analysis on the IoT-23 Data set," p. Available: http://essay.utwente.nl/81979/1/Stoian_BA_EEMCS.pdf Online, 2020.

[14] C.V. Oha, F.S. Farouk, P.P. Patel, P. Meka, S. Nekkanti, B. Nayini, S.X. Carvalho, N. Desai, M. Patel and S. Butakov, "Machine Learning Models for Malicious Traffic Detection in IoT networks".