Creating a Machine Learning Model for Detecting and Classifying Attacks on IoT Systems

Project for Summer Internship in Keystone Groupe

25/08/2023

Made by: Fedi Haddadi

Contents

I. Introduction:

This project was based on a research Article named: **CICIoT2023: A Real-Time Dataset and Benchmark for Large-Scale Attacks in IoT Environment.** The main goal of this research is to propose a novel and extensive IoT attack dataset to foster the development of security analytics applications in real IoT operations. To accomplish this, 33 attacks are executed in an IoT topology composed of 105 devices. These attacks are classified into seven categories, namely DDoS, DoS, Recon, Web-based, brute force, spoofing, and Mirai. In addition, all attacks are executed by malicious IoT devices targeting other IoT devices. This dataset includes multiple attacks not available in other IoT datasets and enables IoT professionals to develop new security analytics solutions. Furthermore, the data are available in different formats, allowing researchers to use features extracted in our evaluation or engineer new features.

The main contributions of this research are:

- We design a new realistic IoT attack dataset, CICIoT2023, using an extensive topology composed of several real IoT devices acting as either attackers or victims.
- We perform, document, and collect data from 33 attacks divided into 7 classes against IoT devices and demonstrated how they can be reproduced.
- We evaluate the performance of machine and deep learning algorithms using the CICIoT2023 dataset to classify and detect IoT network traffic as malicious or benign.

II. The Dataset:

Compared to the state-of-the-art publications, the CICIoT2023 dataset extends existing IoT security insights by using an extensive topology with a variety of IoT devices, executing several attacks never present in a single IoT security dataset. This section introduces the CICIot2023 dataset, the features collected and some insights about the attacks recorded in this dataset.

1. Recorded Attacks:

. For each attack, a different experiment is performed targeting all applicable devices. In all scenarios, the attacks are performed by malicious IoT devices targeting vulnerable IoT devices. For example, DDoS attacks are executed against all devices, whereas web-based attacks target devices that support web applications. Table 1 depicts all attacks alongside the number of rows generated. In addition, Figures 1 and 2 illustrate the instances count for each attack and category. The values are also presented in Table 1.

Table 1. Number of rows for each attack and category.

Attack	Rows	Attack	Rows	Category	Rows
DDoS-ICMP_Flood	7,200,504	DoS-TCP_Flood	2,671,445	DDoS	33,984,560
DDoS-UDP_Flood	5,412,287	DoS-SYN_Flood	2,028,834	DoS	8,090,738
DDoS-TCP_Flood	4,497,667	BenignTraffic	1,098,195	Mirai	2,634,124
DDoS-PSHACK_Flood	4,094,755	Mirai-greeth_flood	991,866	Benign	1,098,195
DDoS-SYN_Flood	4,059,190	Mirai-udpplain	890,576	Spoofing	486,504
DDoS-RSTFINFlood	4,045,285	Mirai-greip_flood	751,682	Recon	354,565
DDoS-SynonymousIP_Flood	3,598,138	DDoS-ICMP_Fragmentation	452,489	Web	24,829
DoS-UDP_Flood	3,318,595	MITM-ArpSpoofing	307,593	BruteForce	13,064
Recon-PingSweep	2262	Uploading_Attack	1252		
DDoS-UDP_Fragmentation	286,925	DDoS-HTTP_Flood	28,790		
DDoS-ACK_Fragmentation	285,104	DDoS-SlowLoris	23,426		
DNS_Spoofing	178,911	DictionaryBruteForce	13,064		
Recon-HostDiscovery	134,378	BrowserHijacking	5859		
Recon-OSScan	98,259	CommandInjection	5409		
Recon-PortScan	82,284	SqlInjection	5245		
DoS-HTTP_Flood	71,864	XSS	3846		
VulnerabilityScan	37,382	Backdoor_Malware	3218		

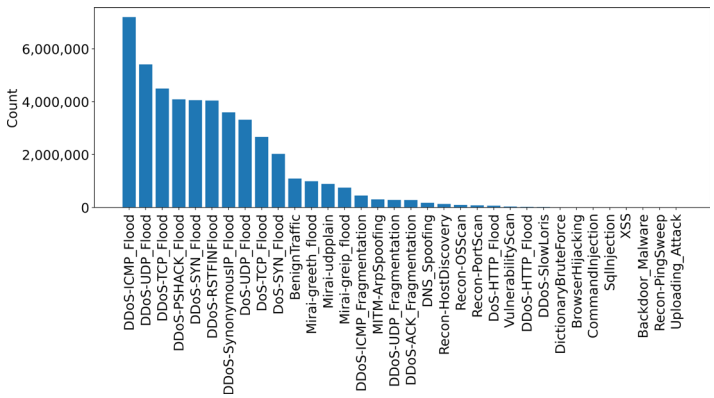

Figure 1. Number of rows for each scenario

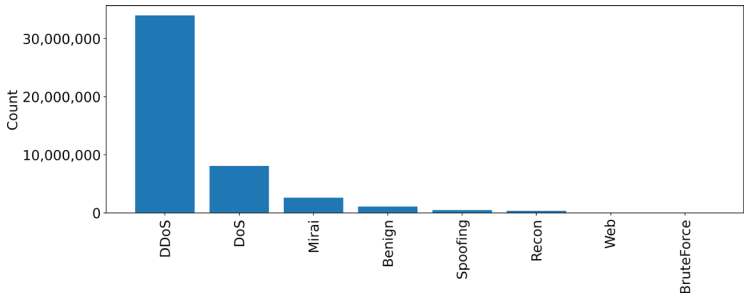

Figure 4. Number of rows for each category.

Note: Due to limitation of computing power and time. We will use just a portion of the original dataset. The CICIot2023 dataset have around 47 million recorded attacks. We will be using around 2.12% (around 1 million attacks) throughout the training and evaluation phases.

2. Extracted Features

These features are extracted based on proposals present in the literature regarding IoT security [8,46]. In fact, although these features have been used and validated in other efforts, our main goal is to present a flexible approach to training ML models with multiple features. Thus, several other features can be extracted or engineered based on the scripts used in this research as well as the raw network traffic (i.e., pcap files).

Table 4. Features extracted from the network traffic.

#	Feature	Description
1	ts	Timestamp
2	flow duration	Duration of the packet's flow
3	Header Length	Header Length
4	Protocol Type	IP, UDP, TCP, IGMP, ICMP, Unknown (Integers)
5	Duration	Time-to-Live (ttl)
6	Rate	Rate of packet transmission in a flow
7	Srate	Rate of outbound packets transmission in a flow
8	Drate,	Rate of inbound packets transmission in a flow
9	fin flag number	Fin flag value
10	syn flag number	Syn flag value
11	rst flag number	Rst flag value
12	psh flag numbe	Psh flag value
13	ack flag number	Ack flag value
14	ece flag numbe	Ece flag value
15	cwr flag number	Cwr flag value
16	ack count	Number of packets with ack flag set in the same flow
17	syn count	Number of packets with syn flag set in the same flow
18	fin count	Number of packets with fin flag set in the same flow
19	urg coun	Number of packets with urg flag set in the same flow
20	rst count	Number of packets with rst flag set in the same flow
21	HTTP	Indicates if the application layer protocol is HTTP
22	HTTPS	Indicates if the application layer protocol is HTTPS
23	DNS	Indicates if the application layer protocol is DNS
24	Telnet	Indicates if the application layer protocol is Telnet
25	SMTP	Indicates if the application layer protocol is SMTP
26	SSH	Indicates if the application layer protocol is SSH
27	IRC	Indicates if the application layer protocol is IRC
28	TCP	Indicates if the transport layer protocol is TCP
29	UDP	Indicates if the transport layer protocol is UDP
30	DHCP	Indicates if the application layer protocol is DHCP
31	ARP	Indicates if the link layer protocol is ARP
32	ICMP	Indicates if the network layer protocol is ICMP
33	IPv	Indicates if the network layer protocol is IP
34	LLC	Indicates if the link layer protocol is LLC
35	Tot sum	Summation of packets lengths in flow
36	Min	Minimum packet length in the flow
37	Max	Maximum packet length in the flow
38	AVG	Average packet length in the flow
39	Std	Standard deviation of packet length in the flow
40	Tot size	Packet's length
41	IAT	The time difference with the previous packet
42	Number	The number of packets in the flow
43	Magnitude	(Average of the lengths of incoming packets in the flow + average of the lengths of outgoing packets in theflow)$^{0.5}$
44	Radius	(Variance of the lengths of incoming packets in the flow + variance of the lengths of outgoing packets in the flow)$^{0.5}$
45	Covariance	Covariance of the lengths of incoming and outgoing packets
46	Variance	Variance of the lengths of incoming packets in the flow/ variance of the lengths of outgoing packets in the flow
47	Weight	Number of incoming packets - Number of outgoing packets

III. Machine Learning (ML) Evaluation

In order to demonstrate how the CICIoT2023 dataset can be used to train machine learning (ML)-based attack detection and classification methods. Firstly, we combine all datasets produced. Once the data are integrated, we evaluate ML performance from three different perspectives:

- multiclass classification, focusing on classifying 33 individual attacks.
- grouped classification, considering 7 attack groups (e.g., DDoS and DoS).
- binary classification (i.e., malicious and benign traffic classification).

In each case, the dataset is divided into the train (80%) and test (20%) sets, which are normalized using the **StandardScaler** method before the actual training process. Finally, the results obtained are summarized as integrated results.

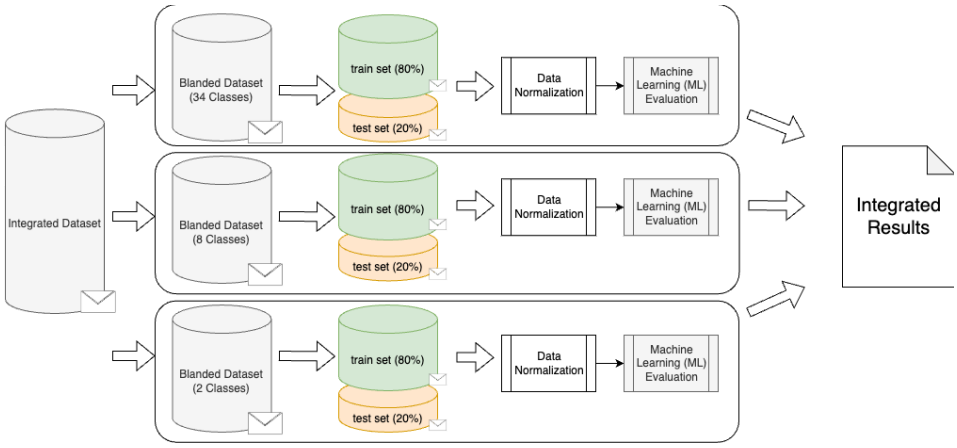

Figure 3. Machine learning (ML) evaluation pipeline adopted in this research.

1. Metrics :

The evaluation of different ML models and configurations is conducted based on evaluation metrics. Given that TP represents the True Positives, TN the True Negatives, FP the False Positive, and FN the False Negatives, the metrics used in this research are :

• Accuracy: responsible for evaluating the classification models by depicting the propor- tion of correct predictions in a given dataset and is based on the following expression:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN}$$

- Recall: the ratio of correctly identified labels to the total number of occurrences of that particular label:

$$Rec = \frac{TP}{TP + TN}$$

- Precision: the ratio of correctly identified labels to the total number of positive classi- fications:

$$Pre = \frac{TP}{TP + FP}$$

- F1-Score: geometric average of precision and recall:

$$F1 = 2 \times \frac{Pre \times Rec}{Pre + Rec}$$

2. Evaluation:

In the evaluation process, we adopted five ML methods that have been successfully used in different applications, including cybersecurity: Logistic Regression, Adaboost, Random Forest. The following illustrates the performance of all methods when framing the classification problem as binary (i.e., malicious and benign), multiclass with 8 classes (i.e., benign and attack categories), and multiclass with 34 classes (i.e., benign and all individual attacks).

Table3. Results obtained in the classification process conducted using different machine learn- ing models.

	34 classes	8 classes	2 classes
Logistic Regression	accuracy_score: 0.801519 recall_score: 0.602292 precision_score: 0.485841 f1_score: 0.49386	accuracy_score = 0.8312326 recall_score = 0.7133377 precision_score = 0.5195019045 f1_score = 0.549203529	accuracy_score: 0.98831536 recall_score: 0.886749 precision_score: 0.8530086030 f1_score: 0.869065
Adaboost	accuracy_score: 0.6922489 recall_score: 0.58150 precision_score: 0.471957 f1_score: 0.476322	accuracy_score = 0.7291914864492008 recall_score = 0.5118663624984297 precision_score = 0.46339704892496364 f1_score = 0.32863948843199586	accuracy_score: 0.996186 recall_score: 0.95162716 precision_score: 0.9684483 f1_score: 0.959875564827
Random Forest	accuracy_score: 0.9915322 recall_score: 0.78615981 precision_score: 0.70363 f1_score: 0.713013	accuracy_score = 0.99444 recall_score = 0.942271 precision_score = 0.7172001 f1_score = 0.740204	accuracy_score: 0.99672539 recall_score: 0.96369226 precision_score: 0.96632340 f1_score: 0.9650039122

For the binary classification, the results show that all methods present high performance, whereas accuracy is a metric that all methods reach over 98%, and the F1-score highlights the difference among these approaches. For example, Logistic Regression achieves 86%, showing that it suffers since the minority class (i.e., benign) is misclassified more often. In the classification of attack groups (i.e., eight classes), the overall performance is degraded since the classification task becomes more challenging. The Logistic Regression, and Adaboost methods show a significant decrease in accuracy.

This impact is even more perceptible for F1-score. However, Random Forest is able to maintain high accuracy and F-1 score. These methods also present a decrease in performance but are capable of achieving F1 scores of 70%.

Finally, the most challenging classification task is represented by a multiclass classification of individual attacks (i.e., 34 classes). In this scenario, Random Forest could maintain high accuracy with very similar results. The same applies to F1-score since a slight reduction was perceived (around 1%) compared to the eight-class challenge. Furthermore, this case study shows that the Logistic Regression,, and Adaboost methods are not able to categorize attacks as efficiently, given that the average accuracy is below 81% and F1-score is less than 50% in both cases.

IV. Conclusion

These results show how ML methods can be used to classify attacks against IoT operations. In fact, this is a starting point that can be considered in any ML-based cybersecurity solutions for IoT operations. This effort not only highlights that the use of other ML methods is possible (e.g., optimized methods), but also enables the adoption of similar strategies to solve IoT-specific problems. Finally, although we are focusing on 33 different attacks, future directions could also be tailored to address issues related to individual attacks or categories.

Made by: Fedi Haddadi

+216 26864820

Haddadifedi33@gmail.com