INF6953QE – Internet of Things Security

Fall 2024

Practical Work 2

Introduction to Machine Learning for IoT Security Datasets

Group 01

2117902 – Prince Madzou

Submit Date: Sunday 11th October 2024

1.Understanding of dataset and diagram

Network Configuration Table

Device	MAC Address	IP Address	Protocols	Ports
Computer	00:1A:2B:3C:4D:5E	192.168.0.100	HTTP, SSH, DHCP	80, 22
Switch (Netgear GS308)	NA	192.168.0.1	Layer 2	N/A
VeraPlus Advanced Home Controller	00:1A:22:34:56:78	192.168.0.150	Zigbee, Z-Wave, WiFi	8080
Amazon Alexa Echo Dot 1	1C:FE:2B:98:16:DD	192.168.0.101	HTTP, MQTT	80, 1883
Amazon Alexa Echo Dot 2	A0:D0:DC:C4:08:FF	192.168.0.102	HTTP, MQTT	80, 1883
Amazon Alexa Echo Spot	1C:12:B0:9B:0C:EC	192.168.0.103	HTTP, MQTT	80, 1883
Amazon Alexa Echo Studio	08:7C:39:CE:6E:2A	192.168.0.104	HTTP, MQTT	80, 1883
Google Nest Mini	CC:F4:11:9C:D0:00	192.168.0.105	HTTP, MQTT	80, 1883
Sonos One Speaker	48:A6:B8:F9:1B:88	192.168.0.106	HTTP, MQTT	80, 1883
AMCREST WiFi Camera	9C:8E:CD:1D:AB:9F	192.168.0.107	HTTP, RTSP	80, 554
Arlo Base Station	3C:37:86:6F:B9:51	192.168.0.108	HTTP, RTSP	80, 554
Arlo Q Camera	40:5D:82:35:14:C8	192.168.0.109	HTTP, RTSP	80, 554
Borun/Sichuan-AI Camera	C0:E7:BF:0A:79:D1	192.168.0.110	HTTP, RTSP	80, 554
DCS8000LHA1 D-Link Mini Camera	B0:C5:54:59:2E:99	192.168.0.111	HTTP, RTSP	80, 554
HeimVision Smart WiFi Camera	44:01:BB:EC:10:4A	192.168.0.112	HTTP, RTSP	80, 554
Home Eye Camera	34:75:63:73:F3:36	192.168.0.113	HTTP, RTSP	80, 554
Luohe Cam Dog	7C:A7:B0:CD:18:32	192.168.0.114	HTTP, RTSP	80, 554
Nest Indoor Camera	44:BB:3B:00:39:07	192.168.0.115	HTTP, RTSP	80, 554
Netatmo Camera	70:EE:50:68:0E:32	192.168.0.116	HTTP, RTSP	80, 554
Amazon Plug	B8:5F:98:D0:76:E6	192.168.0.117	MQTT, CoAP	1883, 5683
Atomi Coffee Maker	68:57:2D:56:AC:47	192.168.0.118	MQTT, CoAP	1883, 5683
Eufy HomeBase 2	8C:85:80:6C:B6:47	192.168.0.119	HTTP, MQTT	80, 1883
Globe Lamp ESP_B1680C	50:02:91:B1:68:0C	192.168.0.120	HTTP, MQTT	80, 1883
Gosund ESP_039AAF Socket	B8:F0:09:03:9A:AF	192.168.0.121	HTTP, MQTT	80, 1883
Gosund ESP_032979 Plug	B8:F0:09:03:29:79	192.168.0.122	HTTP, MQTT	80, 1883
Gosund ESP_10098F Socket	50:02:91:10:09:8F	192.168.0.123	HTTP, MQTT	80, 1883
Gosund ESP_0C3994 Plug	C4:DD:57:0C:39:94	192.168.0.124	HTTP, MQTT	80, 1883
HeimVision SmartLife Radio/Lamp	D4:A6:51:30:64:B7	192.168.0.125	HTTP, MQTT	80, 1883
Philips Hue Bridge	00:17:88:60:D6:4F	192.168.0.126	HTTP, CoAP	80, 5683
Ring Base Station AC:1236	B0:09:DA:3E:82:6C	192.168.0.127	HTTP, MQTT	80, 1883

iRobot Roomba	50:14:79:37:80:18	192.168.0.128	HTTP, MQTT	80, 1883
D-Link DCHS-161 Water Sensor	F0:B4:D2:F9:60:95	192.168.0.129	HTTP, MQTT	80, 1883
LG Smart TV	AC:F1:08:4E:00:82	192.168.0.130	HTTP, RTSP	80, 554
Netatmo Weather Station	70:EE:50:6B:A8:1A	192.168.0.131	HTTP, MQTT	80, 1883

Summary of IoT profiling and monitoring

The emergence of IoT devices can be traced back to 1991, and in 2000, LG introduced an innovative Smart Fridge, marking a pivotal moment in the industry's development. However, it wasn't until 2005 that the first formal reports on IoT began to surface. Surprisingly, despite the growing prominence of IoT devices, no standard or regulations have been established to address their security—unlike other widely adopted technologies. While IoT devices are inexpensive and easy to produce, they significantly lack memory and computational power, which has contributed to the current cybersecurity challenges. Unlike traditional devices, IoT systems are not equipped with standard security measures like Intrusion Detection Systems (IDS), proxy servers, firewalls, malware protection, or log management. This lack of visibility makes monitoring IoT devices particularly challenging, complicating the already complex cybersecurity landscape.

Given the absence of tools and the inherent difficulties in monitoring IoT devices, we are forced to approach this problem from scratch. A foundational step involves identifying IoT devices based on their network traffic and communication protocols. Once identified, profiles can be built for each device, allowing us to model their behavior. This approach is a solid starting point, but several challenges quickly emerge. Monitoring individual IoT devices is crucial due to their heterogeneous nature—varying protocols, behaviors, and operational stages. These dimensions significantly increase the complexity of generalizing across IoT environments. Although we can capture different behavioral stages, a general classifier may not be effective due to the sheer diversity of devices. In fact, devices of the same category, and sometimes even the same brand, can behave differently, further complicating matters.

In-depth IoT profiling involves gathering and analyzing information about individual devices, including packet characteristics, behavior patterns, and other metrics. By building detailed profiles, we can gain a deeper understanding of how these devices interact with the network. Multiple profiles can be created for the same device across different topologies or environments, helping to generalize device behavior in a wide range of scenarios. This approach is especially valuable in identifying vulnerabilities that might otherwise go unnoticed. Having pre-built profiles allows us to detect anomalies in device behavior, which could signal malicious activity or potential attacks. Device

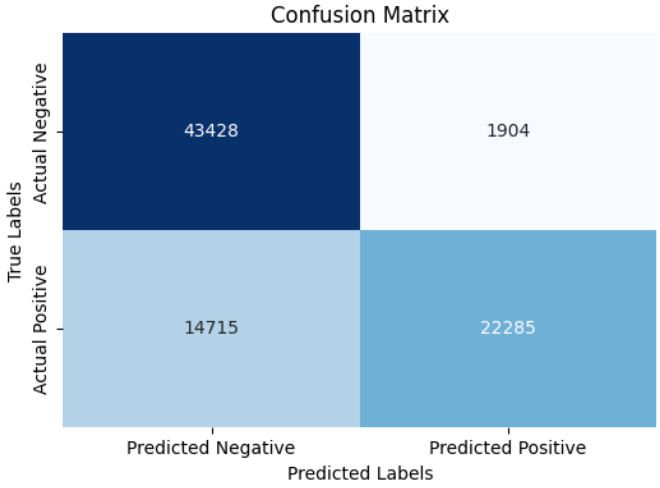
impersonation is another significant concern, but robust profiling can help uncover the identity of a device and prevent such attacks.

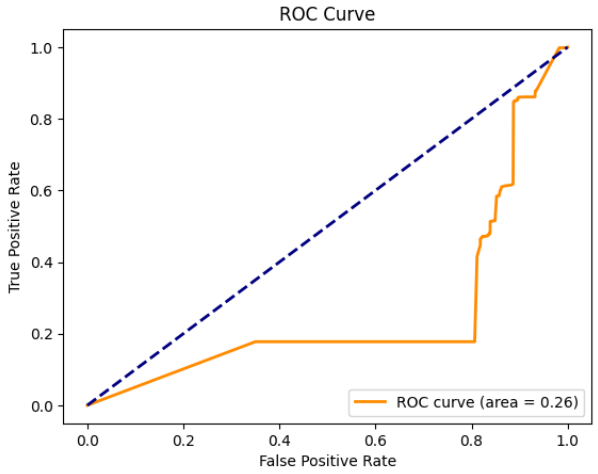
However, the process of data profiling is not without its risks. One major concern is the potential for data poisoning, where the data used to build profiles is intentionally corrupted, undermining the integrity of the model. Techniques like label flipping, where the class labels of profiled devices are manipulated, pose serious threats. Fortunately, several mitigation strategies can address these issues, but the challenge remains substantial.

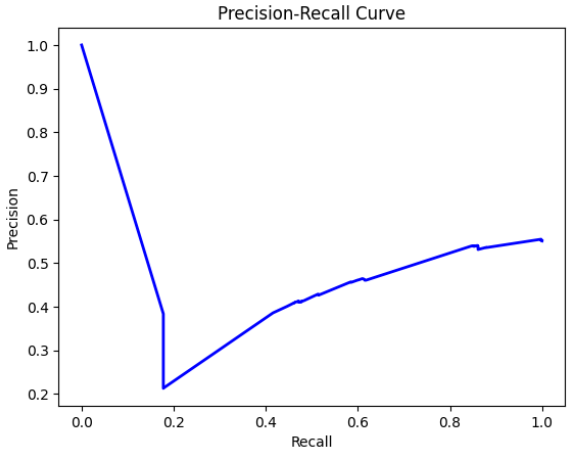
In summary, IoT profiling plays a critical role in anomaly detection and vulnerability identification by creating detailed profiles of devices and monitoring their behavior across diverse environments. While the complexity of IoT ecosystems and the sheer volume of potential scenarios make this task challenging, it remains an essential tool for securing IoT networks. Nonetheless, the profiling process itself is vulnerable to attacks like data poisoning, which could compromise the reliability of these models at their core

2.Application of ML-based algorithm on the UNSW_NB15

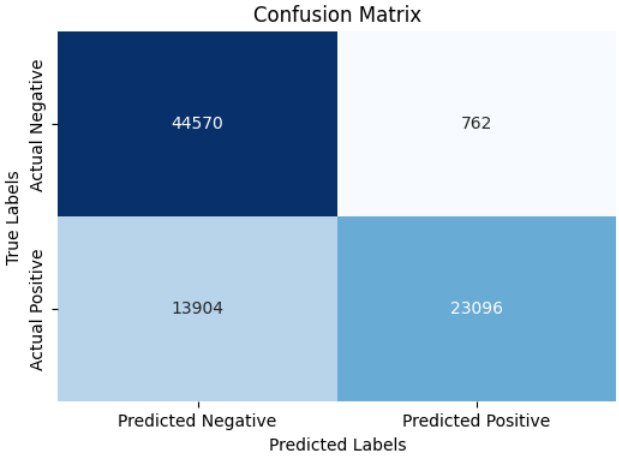
Decision Tree

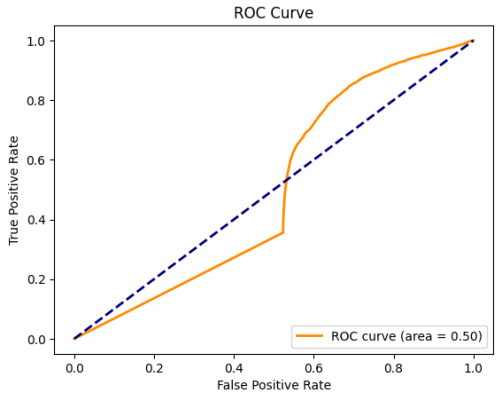


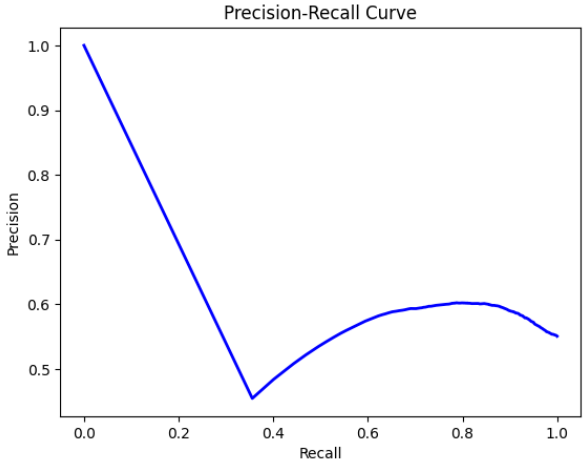




K-Nearest Neighbors







3.Comparison of obtained results

	Decision Tree Classifier	K-Nearest Neighbors
Accuracy	79.81465 %	82.18676 %
Precision	92.12865 %	96.80610 %
Recall	60.22973 %	62.42162 %
F-Score	72.83989 %	75.90128 %

Based on recent observations, we can clearly determine that K-Nearest Neighbors (KNN) outperforms the Decision Tree Classifier (DTC) when comparing model accuracy. However, accuracy alone is not sufficient to declare one model superior. Even though the training dataset contains a roughly equal distribution of classes, precision provides a more nuanced assessment, particularly for the prediction of positive outcomes. In this case, KNN achieves a true positive rate of approximately 96%, while DTC follows with around 92%.

Moreover, KNN demonstrates a higher recall, meaning it more effectively identifies positive cases, leading to fewer false negatives compared to DTC. This is particularly important because false negatives—failing to detect an attack when one is present—are critical errors that can have severe consequences. Reducing these errors is vital for improving model reliability.

Finally, the F1-score, which combines precision and recall into a single metric, also favors KNN over the Decision Tree Classifier. This further reinforces KNN's overall performance in distinguishing between normal communication and potential attacks.

In summary, K-Nearest Neighbors consistently outperforms the Decision Tree Classifier across multiple evaluation metrics, making it the more reliable model for predicting whether a communication is an attack or not.