Internet of Things (IoT) Device Investigative Analysis Using Machine-to-Machine (M2M) Framework

Phaneendra Kanakamedala¹, Yarra Harika¹, Mamilla Krishnaveni¹, Bhuvanesh Nallani¹
¹Lakireddy Balireddy College of Engineering, Mylavaram, India

Emails: phanikanakamedala@gmail.com, yarraharika2002@gmail.com, mamillakrishnaveni8@gmail.com, bhuvaneshnallani89@gmail.com

Abstract
As we know that now a days the possibility of the uninterrupted attacks on the IOT devices are increasing. The less memory and the minute processing power of these appliances make it tough for the security analyst to store the records of the different attacks. The forensic analysis is used to evaluate the damage done on the devices due to numerous attacks. In this mechanism the attacks on the IOT devices are detects undoubtedly by using machine-to-machine (M2M) framework. In addition to the using machine-to-machine framework the machine learning algorithms also been used to identify various attacks automatically. Here we use the third-party logging server in order to issue. The execution will be studied in the form of accuracy, precision and the Random Forest gives the most accuracy.

1. Introduction:
When connected to the internet, IoT devices may safely gather and share data (Vishwakarma and Jain). The wide range of applications in it and provide the evolution of a number of innovations, such as wearable technology, smart cities, smart metering, smart thermostats, and smart homes (Yang et al.). The Internet of Things has simplified human lives (Javaid and Khan).

In spite of the reality that the applications of IoT are continuously increasing, IoT tool reliability stays a limitation (Hossain et al.). The manufactures of these devices are often engrossed with getting original appealing capabilities and modifying to improve the devices’ intelligence and efficiency without increasing their price secured (Alladi et al.). In reality, There have been various cyberattacks on IoT devices in current as a result of years insufficient safety features (Almogren). The number of IoT devices and cyberattacks is each progressively growing (Sikder et al.).

Denial of service (DOS) is among the most familiar Infiltration of the IoT network. According to Cisco’s annual look at DDoS cyberattacks are expected to increase between 2018 and 2023 from the given diagram it compares the quantity of feasible DDoS assaults are every year (Hussain et al., “IoT DoS and DDoS Attack Detection using ResNet”). In addition, determined that IoT DoS attacks are constantly growing each day. It’s far suggested by using the Palo Alto Networks Unit 42 studies crew 98% of IoT devices’ traffic does not always encrypted, exposing the exclusive records network traffic and attacks on the network and systems at multiple levels (Stergiou et al.). It will increase the threat space for attackers, when these unprotected linked Internet of Things devices are on the network. In step with Kaspersky’s Research,
1.5 billion assaults on IoT hardware has been mentioned within the initial half of 2021 (Hussain et al., “Towards a Universal Features Set for IoT Botnet Attacks Detection”).

Moreover, IoT devices for smart homes which includes the various IoT devices, as smart cameras divide the 25% of the malware attacks in a botnet assault (Yousefnezhad, Malhi, and Främling). Through The Mirai attack takes use of default Credentials and obtained manipulate on hundreds million IoT devices and conducted a distributed denial of service (DDoS) assault against key systems (Tawalbeh et al.). HP additionally pronounced roughly 70% of machines are prone (Mariyanayagam, Shukla, and Virdee). Therefore, the security flaws for the tools need very secure (Gupta, R. Kumar, and A. Kumar). Security flaws is one of the weakness in devices which offers a prime target for hackers. Wrong component trying out, a scramble for price, as well as a scarcity of powerful rules are also the primary Reasons of IoT threats (Karabiyk and Akkaya). The structure is necessary to identify upon the assaults on various tools, store verification of those threats (Mazhar et al.). The vulnerabilities of devices can be mitigated by applying the forensic evaluation. Further, the assault and perpetrator may be clearly identified.

IoT devices are able to reduce a constrained procedure for variety a set of specified instructions (Haider et al.). Correspondingly, among the IoT devices, they are unable to acquire, interpret, or record, analyze connectivity. Because of the architecture of IoT device they are more complex for security analyst in case of store the data among the various attacks. Due to those constraints, the gathering of facts is a significant task in forensic analysis. To make IoT environment more secure and robust some of the special tools and approaches are required. In case of the research purpose the devices are Better appropriate analytical approaches should be developed and used.

The following mentioned issue is avoided along with assistance of forensic analysis approach. We provided a framework which executes the threats with detection, recognition and generate the records and warnings for these threats. This is focused with security Incident control (SIM) for recognizing security incidents are done at the computer network, after which proper precautions are taken and finished some constraints of security ideas which are harmed. Forensics analysis is distinct to network auditing as it is the pre-examination of a network’s flaws while forensic analysis is the post-examination of security misfeasance that What happened to the document and when it happened.

Acquiring the data is one of the problem that is addressed by employing a third-party logging server. The traffic generated towards the devices are routed towards the server, in this forensic analysis is used to generate and store the logs and alerts of malicious attacks. to acquire data about the assaults and the perpetrators the previously saved records are recreated, and analyzed in a server. The detection of these machine learning is used to perform assaults using a dataset it built all these data.

The generalized forensic analysis procedure consists of four steps: data collection, evaluation, processing, verification and the report. [23]. In first step, the data according to a particular attack are gathered. Data acquisition is a major issue due to the limitation of IoT devices have limited computing power. In this case attacks with the evidences are not found. This issue can be handled by our intended system the log server is introduced to identify the threats and also the logs of malicious traffic are maintained, and warnings are generated.

And the next step is to gathered information is analyzed to the different information which is pertinent. The traffic should be redirected in IoT devices are configured by using an IP table. The log server writes the logs related to the particular threats as well as embedding them into the detecting engine. In the process of traffic redirection related data is
taken and other packets are removed. The obtained information is analyzed for the purpose of useful data. Server security onion gives the information related to when records are collected and regenerated. Security onion gives the information to identify the kind of assault and the assailants and the source and destination ports. Snort gives the alerts for these kinds of attacks. These kinds of attacks are inspected and gives the data help to identify the attacker and the extent of the harm caused by the threats. In the final step, the analytical findings are obtained.

2. Literature Survey

Various forensic tools and frameworks are designed for detecting the attacks in IoT devices.

Fagbola et al. proposed a framework for smart digital forensic readiness (SIoTDFR). SIoTDFR includes six distinct stages, device connection, device identification, device monitoring, digital evidence gathering, digital evidence preservation and secure storage. It shows the tiniest PDE and when an incident occur it monitors the criminal activity easily.

Aslan et al. gives the idea on various malware detection techniques along with their advantages and disadvantages. To detect the both signature-based and heuristic-based detection methods have proven to be effective in detecting malware but to identify the detection method relying on known malware signatures has proven to be unsuccessful. Different approaches like behaviour based, Cloud-based methodologies exhibit strong performance in terms of efficiency complicated malware, some parts of the known and unknown malware are detected by using some approaches

Schedit et al. introduced a system for recognizing IoT devices through the utilization of DNA has been devised. With the help the buyer’s details and the unique identification number of the device were kept confidential DNA of IoT devices are created. By using the DNA, The signs of assaults on these devices can be quickly recognized through their distinct fingerprints. The Hybrid Forensic IoT server was introduced in order to help the present IoT forensic investigation process.

Shrivastava et al., Examined the threats and the utilization of machine learning algorithms on IoT devices was executed to improve their performance. To detect a malicious network traffic some of the classification-based are used, among all algorithms SVM gives the more accuracy. They examine some commands and identifies how the malicious activity is performed by the attackers.

Hegarty the authors tackled the intricate nature of digital forensics in the Internet of Things and suggested a cloud computing solution for conducting digital investigations.

An overall examination of suggested remedies and a blueprint of the system are included in the effort. Nonetheless, they lack a plan for implementing their idea.

Oriwoh et al. These hypothetical situations were developed through a comprehensive examination of individuals who employed a novel approach in committing their digital offenses Upon evaluating and deducing insights from the research findings, a framework was established which utilizes regions as the basis for exploring the IoT ecosystem, and centers on three key elements.

Nisais Nimalasingam An effective approach to detecting IoT malware through forensic analysis is to focus on the most distinct network traffic features and combine them with the binary characteristics of various malware families. A massive collection of network traffic data was utilized, featuring various network traffic characteristics. As a result of the feature extraction process for each malware type, the proposed model demonstrated an impressive detection accuracy of nearly 96% during the experimentation stage of the research.

Ayush Kumar and Teng Joon Lim ML classification algorithms were used to offer EDIMA, a modular solution for the identification of network activity coming from IoT threats. Obtained features are collected from network traffic samples at the accessing gateway level and given target class. Several common machine learning (ML) techniques were trained using some of the selected features that have been retrieved, and the resulting ML models were again deployed to analyse data collected with their classification scores provided.

Meffert et al. The challenges posed by IoT devices, such as the absence of a uniform standard, are emphasized as being numerous and complex. There are many different communication protocols used by different electronic devices, while some of these devices often use Real-Time Operating Systems
Internet of Things (IoT) Device Investigative Analysis Using Machine-to-Machine (M2M) 2023, Vol. 05, Issue 05S May

(RTOS), which often have very little storage space or none.

3. Methodology

As shown in Figure 2, the suggested architecture for analysis of devices during assault is composed of four components. First, traffic generation attack is in charge of generating attacks from the Kali Linux system to the tools used in experimentation. Second, there is a network redirection logging server where the warnings are provided in charge for routing traffic to the devices and the servers, analyses communication and provides records only the network matches with server rules. Third, analysis with the server is in charge of regenerating records obtained from the network. The records are rebuilt, the useful data about assault and the assailants is taken. Finally, the analysis in charge of detecting threats with the help various machine learning models.

FIGURE 2. Approach for forensic analysis of network.

Different devices are used in our designed experimentation, like The Raspberry Pi, utilized as an Internet of Things device, can be equipped with a Pi camera. snort as a logging server, security onion as a forensic server, Kali Linux to produce assault possibilities. These devices are all linked to a common system. the given diagram depicts about device configuration.

3.1. Assault Network Generation

In this architecture the initial step involves using Kali Linux, to perform several assaults on the Raspberry Pi, which has an IP address of 192.168.56.101. On the board, IoT devices are built with connectivity options Sensors, cameras, and other devices of various types are used. The Internet of Things board is built on an open-source platform. In this experiment, a Raspberry Pi is used as an IoT device, Ettercap, HPING3, NMAP, Metasploit, and Wireshark were among the Kali Linux tools we used. These are Raspberry Pi-based attacks.

1. NMAP port scanning;
2. attack through brute force with Metasploit;
3. Synthetic flooding with DoS utilizing HPING3;
4. Via Ettercap, perform MITM ARP spoofing

3.2. Diversion of communication & creation of intrusion records and signals

Under this architecture the traffic from the devices is diverted to a server, additional records were produced. To prevail the limitations of IoT device traffic is forwarded to a server at 192.168.56.2 as its IP address. Every device has an M2M connection and can communicate with one another directly. Regarding server-side record archiving, It was done using a tracking interface (WAZUH). A third-party server is used. However, the Raspberry Pi’s ARM architecture does not support it. The network traffic is then routed through IP tables towards a snort gateway, which has the address of 192.168.56.2. Snort is installed on the logging server. Snort guidelines are created and added to the setup file for certain attacks. Snort examines a network arriving at the device, compares
towards a snort guideline. When the server’s identification finds it is match, then threats were detected by a snort alert, & records are kept in the server. Otherwise, packets are outmoded. Snorts generates attack logs in pcap format. The open-source CIC flow metre these pcap files are transformed to CSV files using program. These machine learning models are utilized using a CSV file of records because ML can’t be performed to pcap files.

3.3. Analytical Approaches Using Security Onion
Logs are saved for analysis after snort detects an attack. These network files included details on the kind of threats, source & destination address, and additional information. Security onion has two network cards and an IP address of 192.168.56.4. The first is used for management, and the second sniffs network packets to find illegal activity on the network. The logging server record solve a problem of evidence acquisition. Security onion includes a number of built-in tools for analysing logs, including squil and squert. Sguil is the graphical user interface for snort, a command-line tool. As we have captured the log, to gather details about assaults and assailants, these logs are periodically created.

3.4. Machine learning for forensic analysis
Machine learning algorithms are used to detect attacks on IoT devices. Automatic detection using Snort is not possible in which We utilize IDS to various threats each time. Through an artificial assault prediction system, ML by using various types of classifiers and labelling generates CSV-formatted logs. This information was split into training and testing groups after pre-processing. After extracting the features, we developed ML models and evaluated them using real-time traffic and the testing dataset.

3.4.1. Data Labelling and Flow Aggregation
Because PCAP files cannot be used by machine learning models, To transform into CSV format a CIC flow meter is utilized . Traffic behaviour,some analytical traits are taken out. after that ,Those attributes catered to machine learning model, which detects threats to devices. Information is labelled to identify standard and anomalous behaviour, shown in Table 3.

3.4.2. Data Pre-processing
To guarantee data reliability, integrity, and stability we eliminate irrelevant fields and any attributes which encrypt qualitative features, can never aid to categorization, & scale properties of numerals between 0 and 1. In order to prevent model efficiency deterioration and source bias, previously labelled fields such as category of threat and IP & port address must be removed. Various strategies are applied to eliminate anomalies and incomplete data.

3.4.3. Dividing the Dataset for Training & Testing
Furthermore, the dataset is divided as two subsets: training and testing. With the use of trained data, the model is developed and tested. Thirty percent of the dataset is used for testing, and 70% for training.

3.4.4. Identification and Analysis of Attributes
A machine learning algorithm’s recognise effectiveness is decreased by related attributes. To select attributes, Backward elimination , k-best , and attribute value were utilized. We chose k-best for feature extraction. As shown in Table 4, K = 10 yields the best accurate results.

3.4.5. Model development and evaluation
After the feature extractor extracts the features from the inputs, within the learning phase, inputs and labels are supplied to machine learning algorithms. The optimal model is built using a combination of machine learning methods. Each model operates

### Table 1. Analysis and Configuring possibilities

<table>
<thead>
<tr>
<th>Category</th>
<th>Type</th>
<th>Symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>Normal</td>
<td>0</td>
</tr>
<tr>
<td>Anomaly</td>
<td>Dos</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Brute force</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Attack</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Shell Code</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Backdoors</td>
<td></td>
</tr>
</tbody>
</table>

### Table 2. Test config settings

<table>
<thead>
<tr>
<th>Selected</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>Flow_Byts/s,Pkt_Len_Var,Flow_Pkts/s, Fwd_Pkts/s,Bwd_Pkts/s, Bwd_IAT_Max,Src_Port, Bwd_IAT_Mean,Bwd_IAT_Tot, Flow_Duration</td>
</tr>
</tbody>
</table>

International Research Journal on Advanced Science Hub (IRJASH)
differently due to the multiple domains on which input is trained. During the testing step, input usually passed to the pattern generator to acquire the features. They are provided to extract the labels into predefined classifier models predictions. We labelled our data as before. As a result, the threat is deduced & predicted from the model. Figure 4 depicts a typical representation of the training and testing stage. They made utilize of certain assessment criteria to determine effectiveness of predictions, including the using confusion matrix’s F1 score, recall, accuracy, and precision. The trained models are evaluated using the testing dataset. Efficiency and other metrics won’t accurately depict the actual fault when we only utilise the training dataset. Additionally, during training, cross-validation is utilised to fine-tune the models and enhance the performance measures. On the basis of their accuracy, precision, F1 score, and recall, our models were assessed.

The analytical system security onion retrieves the records from the server, where network packets are created & examined. To automate this proposed model, ML models are applied with the dataset. When Cyberattack detection seems to be more accurate when forensic tool analysis and machine learning study are combined.

4. Existing state-of-the-art methods for forensic analysis of IoT devices include:

Digital Forensic Investigation of Internet of Things Devices: This method involves analyzing the digital artifacts on IoT devices to identify evidence of cybercrime incidents. The method uses traditional forensic techniques, such as data carving and analysis of file system metadata.

Network Forensics Analysis for Internet of Things: This method involves analyzing the network traffic between IoT devices and other devices or services to identify evidence of cybercrime incidents. The method uses network traffic analysis tools to capture and analyze the traffic.

Accuracy:

Accuracy is a statistical metric that is commonly used to evaluate the performance of a model, classifier, or algorithm. It is defined as the ratio of the correctly predicted instances to the total number of instances in each dataset.

\[
\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \tag{1}
\]

TP-True Positive, TN-True Negative, FP-False Positive, FN-False Negative.

Equation (1) represents the proportion of correct predictions among all predictions made.

Recall:

Recall is a statistical metric used to evaluate the performance of a classification model or algorithm in correctly identifying the positive instances. It is also known as sensitivity or true positive rate (TPR). The formula to calculate Recall is:

\[
\text{Recall} = \frac{TP}{TP + FN} \tag{2}
\]

Equation (2) represents the proportion of true positives among all actual positive cases.

Precision:

Precision is a commonly used performance metric in machine learning and information retrieval that measures the proportion of true positives among the
instances predicted as positive. It is calculated using the following formula:

\[
\text{Precision} = \frac{\text{True positives}}{\text{True positives} + \text{False Positives}}
\]  

(3)

Equation (3) represents the proportion of true positives among all positive predictions made.

**F-Score**:

The F-score (also called F1-score) is a statistical measure that combines precision and recall into a single value. It is used to evaluate the performance of a classifier or model in binary classification problems.

The formula to calculate the F-score is:

\[
F1 \text{ - score} = \frac{2 \times (\text{precision} \times \text{recall})}{(\text{precision} + \text{recall})}
\]  

(4)

Equation (4) represents the harmonic mean of precision and recall, which gives a balanced score that takes both precision and recall into account.

**FIGURE 5. Random Forest Confusion Matrix**

**Random Forest Confusion Matrix**: It shows the performance of a Random Forest classification model on a set of test data. It summarizes the number of correct and incorrect predictions made by the model for each class, organized by true and predicted labels.

**Decision tree confusion matrix**:  
It summarizes the performance of a Decision Tree classification model on a set of test data. It shows the number of correct and incorrect predictions made by the model for each class, organized by true and predicted labels.

**FIGURE 6. Decision tree confusion matrix**

**FIGURE 7. Extra trees confusion matrix**

**Extra trees confusion matrix**:  
It summarizes the performance of an Extra Trees classification model on a set of test data. It shows the number of correct and incorrect predictions made by the model for each class, organized by true and predicted labels.

**MLP Confusion Matrix**:  
It summarizes the performance of an MLP classification model on a set of test data. It shows the number of correct and incorrect predictions made by the model for each class, organized by true and predicted labels.

**Gradient Boosting Classifier**:  
It summarizes the performance of a Gradient Boosting Classifier on a set of test data. It shows the number of correct and incorrect predictions made by the model for each class, organized by true and predicted labels.
It used for classification tasks. It is an ensemble learning method that combines multiple weak prediction models to create a strong classifier. The algorithm builds the model in a step-by-step manner by minimizing the error in each iteration.

Various machine learning models are applied to the dataset obtained from records gathered by the logging server. We employed strategies like component significance, k-best, and backward elimination to identify & choose features. We utilised the k-best feature extraction and selection approach to choose the most optimal characteristic in the database while executing several tests. This information is separated as 70% for training and 30% for testing. We assessed the performance of the models by taking into account various parameters, including accuracy, recall, precision, and F1-score. Table 3 compares the efficiency of ML algorithms. Among all the algorithms random forest gives the more accuracy.

5. Future scope:
As the Internet of Things (IoT) continues to grow and become more prevalent in our daily lives, the need for forensic analysis on IoT devices using machine-to-machine (M2M) frameworks is becoming increasingly important. Here are some potential future developments and applications in this field like Increased complexity of IoT devices, Improved security, Increased demand for forensic analysis and Cross-disciplinary collaboration

6. Conclusion:
The proposed system is aimed at detecting attacks on IoT devices through the implementation of machine learning techniques like Random Forest, Decision Tree, Extra Trees, Gradient Boosting Classifier, MLP. A confusion matrix is a useful tool for evaluating the performance of different classifiers used on a test dataset where the true values are already known. It presents the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) for each classifier. The true positives are the number of correctly classified positive instances, while true negatives are the number of correctly classified negative instances. False positives are the number of negative instances that are mistakenly classified as positive, and false negatives are the number of positive instances that are mistakenly classified as negative. A confusion matrix allows us to compare the classification results of different models and assess their overall accuracy.
References


Embargo period: The article has no embargo period.