IoT Architecture Security and Proposal for Semi-Markov Chain IDS

Joel Magee
The Department of Computer Science
Hampton University

Abstract

This research serves as a broad examination of the different threats and attacks against the IoT architecture. This research analyzes the different layers of the IoT architecture and the cyber attacks that threaten them each. Intrusion detection systems provide a means of protection against various attacks. Hence substantiating the proposal of a host-based signature type intrusion detection system utilizing the semi-markov process for IoT devices in a smart home environment. The semi-markov chain could potentially prove as an effective means to acutely identify behavioral anomalies associated with nodes within an IoT environment.

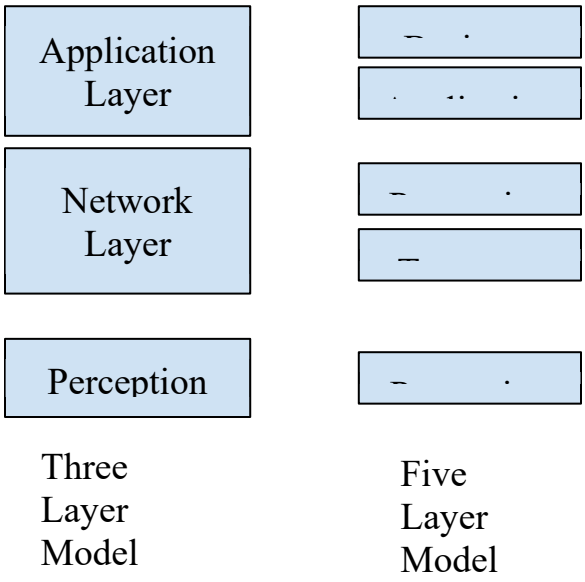
I. Introduction

The consistent development of technology has yielded innovative solutions to modern-day tasks. A sector of modern technology is the capability of physical devices to exchange and collect information across the internet. This technology is called the Internet of Things(IoT). IoT technology is widely used as 10 billion active device connections were reported in 2019 with applications in wearable devices, Healthcare environments, Energy grid environments, Smart Home environments, Smart City environments, Agriculture, and Industrial environments.[19] It is projected that the number of connected IoT devices will increase to 30.9 billion by 2025. [19] The concept of The Internet of Things was first officially presented in 1999 by Kevin Ashton of Auto-Labs at MIT [2]. Ashton proposed the idea of computers gaining knowledge about their surrounding environment, gathering data without human interaction with the intention of efficient accounting, reducing waste, and decreasing costs. In 1989, a group of students from Carnegie Mellon University were able to connect to a vending machine via ARPANET. By doing so they were able to identify the available beverages within the machine. [14]

As new technology emerges, new threats to that technology emerge as well. To combat against malicious actors technologies exist designed to detect and prevent cyber attacks. In the 1980's, Dorothy Denning and Peter Neumann developed the initial model of an intrusion detection system.[13] The model was based on the concept that an intruder's behavior substantially differentiates from legitimate users, therefore it can be detected by usage statistics analysis. The original model has evolved to give way to an intrusion prevention system. The modern day version of these security systems have been given various classifications such as Network specific or Host specific, knowledge-based or behavior based, and if it is an active or passive IDS. This paper will discuss a proposal classified as a behavioral, host-based intrusion

detection system. A behavioral intrusion detection system analyzes the behavior of traffic by comparing a baseline of standard system activity, identifying intrusion attempts. A host-based IDS primarily focuses on protecting a specific asset or device. Therefore the system is installed directly to the machine. [13]

The architecture of IoT is structured with three primary elements: the network layer, the perception layer, and the application layer. Several other architecture definitions exist expounding on the functions of each layer. The five-layer model introduces a processing layer and transport layer in place of the network layer and includes a business layer as the topmost layer above the business layer.



Cisco, IBM, and Microsoft each utilize a more complex model for referencing the structure of IoT systems. As impactful as this technology has become, it is still susceptible to cyber attacks. The variety of cyber attacks that threaten IoT environments can be associated with each of the different layers. This research will examine threats associated with the layers of the three-layer model.

II. IoT Architecture Threats

Application Layer

As stated earlier, the architecture of IoT can be described in three essential pieces. The network layer, the perception layer, and the application layer. The top most layer is the application layer. The application layer is responsible for providing services to a user that are specific to the application. The specific application could be a smart home system or a smart healthcare environment. For instance, in a smart home system, the user would interact with the application layer to lock the doors or monitor the thermostat. Within the application layer, the IoT system is able to facilitate communication between devices on the local network, the internet, and servers storing system specific data. This layer is able to do so using various protocols such as Message Queueing Telemetry Transport, Constrained Application Protocol, Representational State Transfer, Advanced Message Queuing protocol, and Extensible Messaging and Presence Protocol.

As this layer utilizes messaging protocols it is susceptible to client-side and data attacks. There are numerous attacks that target client-side vulnerabilities, the most common are cross site scripting, Trojan Horse malware, and data overflow attacks.

- Cross-site scripting (XSS) is an attack that targets trusted websites. A malicious actor will utilize a web application to inject malicious code to a different end user.[17] This type of attack can result in loss of confidentiality for browser cookies or end user files. A XSS attack can be classified as either a stored, blind or reflected XSS attack. A stored attack occurs when the injected script is stored on the targeted server for when a victim

user requests the stored information. A blind XSS attack occurs when the payload of the attacker is stored on the server and reflected back to the victim from the application. A reflected XSS attack is when the injected script is reflected from the web server as a response that includes the input in the request. A system is vulnerable to a XSS attack when a web application generates output without validating the input from the user.

- Trojan Horse malware is the term given to malicious software that portrays itself as a benign program, deceiving the user into downloading the software. A trojan horse can be categorized as a remote access trojan, data sending trojan, destructive trojan, proxy trojan, FTP trojan, security software disable trojan, and a denial of service trojan. [21] Each category of a Trojan attack is to either destroy data, collect data or modify the settings of the victim system. The Cybersecurity & Infrastructure Security Agency (CISA) acknowledged the Trojan Horse Emotet as a "sophisticated trojan". The Emotet malware is designed to infiltrate a victims machine via phishing email attachments and links, which it then uses to proliferate with a network and writing to shared drives. [18]

- A Data Overflow attack is the term given to an attack that targets the memory component of a system. This attack is conducted when a program tries to write a value to a memory buffer that is too large, causing the buffer to overflow into different sections of memory.[12] This fault in memory systems grants attackers the ability to inject and execute malicious code. IoT devices possess a small amount of space of memory compared to more complex machines. This attribute allows for an easier compromise. IoT devices are also more highly susceptible to overflow attacks because they are written in the C++ or C programming language, which does not have a secure method for allocating space for excessive memory.

Network Layer

The network layer is the middle-ware layer. The network layer for the IoT architecture operates in a similar manner to the network layer in the TCP/IP and OSI model. This layer conducts the connectivity between other smart devices and servers on the network. The collected data from physical components in the system transition to this layer. It serves the purpose of transmitting and processing sensor data. In the case that a home automation system receives a request, such as an Alexa or Google Home, the network layer allows for the devices in the home to execute the received request. The User Datagram protocol is typically preferred in an environment with low power as it is connectionless with lower overhead, compared to TCP. This layer operates on several protocols such as Wi-Fi, Bluetooth, ZigBee, Z-Wave, and LoRaWan.

The network layer is typically threatened by attacks compromising integrity and bypassing authentication. The most common threats to the network layer are Man-In-The-Middle attacks, Sinkhole attacks, and Eavesdropping attacks:

- A Man-In-The-Middle attack occurs when internet communication is intercepted and modified during transmission between two nodes. IBM reported that in 2018, 35% of

exploitation activity involved attempted Man-in-the-Middle attacks. [24] In 2015, a group of hackers were caught facilitating a MiTM attack against major European companies. The attackers would infiltrate a company's network using social engineering and then established an illegitimate replica of a banks website, requested account credentials, and simultaneously set a transaction with the real website.[11] This attack is difficult to defend against, as attackers can spoof their addresses and other forms of identification.

- A sinkhole attack is an attack procedure in which an internal node is compromised and attracts nearby network traffic. It accomplishes this by dispersing fake routing information utilizing the routing metric associated with the routing protocol. Once compromised, the node can carry out any number of attacks such as a selective forwarding, denial of service or data fabrication.[7]

- An eavesdropping attack occurs when an attacker passively listens to communications on a network. The attacker would gain information such as node identification numbers, routing updates, or application specific data. Once the information is obtained, the attacker can carry out network disruption, compromise nodes or degrade application performance.

Perception Layer

The perception layer is the bottom layer. The sole purpose of this layer is to gather information from sensors within the environment. A sensor, in terms of IoT, is defined as a device or module that detects changes in its surroundings. [6] The physically quantified data is converted into digital signals to be transferred to the network layer for processing. The operation and data of the sensors are dependent on the application of the IoT system. The collected sensor information can range from motion, temperature, location, etc. A list of common sensors are humidity sensors, pressure sensors, proximity sensors and level sensors. The method for data collection also depends on the specific sensor. For instance, Robodo-Sen36 model pressure sensors utilize procedures developed by Honeywell.

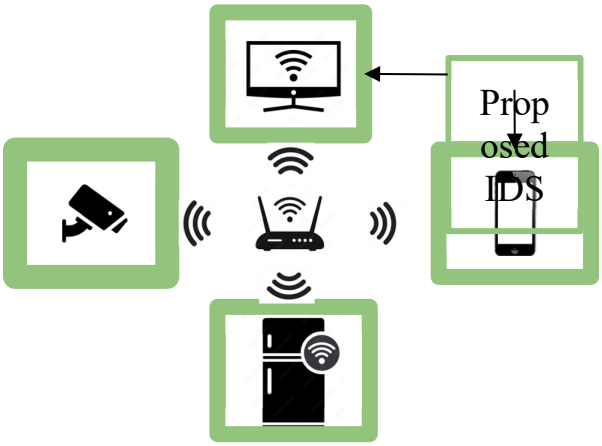
As the perception layer is composed of sensors and physical devices, it is especially susceptible to physical and signal interference attacks. The most notable attacks against the perception layer are replay attacks and jamming attacks.

- A replay attack is a form of man-in-the-middle attack but is primarily focused on the authentication aspect of communicating nodes. During a replay attack, the malicious actor will eavesdrop and collect authentication information between two nodes. The captured authentication information is then used to authenticate the hacker as the other node in the conversation. By sending the same authentication information, the victim is convinced that the hacker is authentic and has proven their identity. A form of defense against replay attacks is an IPSec protocol, VPN, in which messages are encrypted with a key and each packet is exchanged with a counter. In the case a node receives a packet with a number lower than what is to be expected, the packet is dropped. [22].

- A Jamming attack is an attack focused on the physical aspect of the IoT environment. A jamming attack is defined as the disruption of existing wireless communications by decreasing the signal-to-noise ratio on the receiver side during wireless signal transmission[4]. Jamming techniques have been classified into four different categories: proactive, reactive, function specific and smart-hybrid. A proactive jammer conducts interference techniques regardless of if data is transmitting throughout the network. A reactive jammer operates upon the observance of specific network activity condition. A function-specific jammer is implemented with a specific function. A smart-hybrid jammer is a reference to a general jammer that is energy efficient and can be implemented as both proactive and reactive.

III. IDS with Semi-Markov Model Proposal

I will now propose a host-based intrusion detection model utilizing the Semi-Markov Process (SMP) for baseline modeling and anomaly detection within a smart-home IoT environment. The Semi-Markov process is a state transition model. It utilizes a matrix to store the semi-markov chain values, whereas each entry is a time dependent probability that after a transition to a given state, the next state transition occurs in an amount of time less than or equal to the time associated with the previous state transition. The essence of this model is collecting data relating from the various events that occur in a non-compromised environment. The collected event data will be used to create a state matrix. The state matrix will demonstrate the typical time for conducting each event. The targeted events are outgoing internet connections from applications such as MQTT as well as inter-device communications.



Execution

The IDS was selected to be a host-based system to account for device-to-device communication. This communication would bypass an intermediary medium such as a router. Also, in the case the IDS was network based, the end nodes would still be susceptible to perception layer attacks such as jamming attacks. The SMP IDS model will create a baseline by developing a SMP state matrix composed of probabilistic values that represent the time for events to complete. Consider k and n represent two states referencing an IoT environment specific event such as a data processing task.

$$P_{k,n} = E_{k,n}$$

The equation represents that the probabilistic value is saved as an entry into the SMP state matrix. The process will operate during peak time and downtimes of the environment to create a sufficient baseline. This will create the average expected time for state transitions associated with environment events.

To ensure detection, post baseline creation, the IDS will use the matrix to compare ongoing

events with stored values. In the case a new event, X, is not within the expected range, $E_{k,n} > X_{k,n} > E_{k,n}$, the system will notify the user of an attempted intrusion. This proposed solution should be evaluated for time effective notification as well as the potential for false positives.

Testing Methodology

The tests for the proposed model were focused on identifying CPU exhaustion attacks. The Semi-Markov chain was implemented by evaluating the difference in time of transitions from CPU utilization states. The CPU utilization states are four groups representing the levels of CPU utilization.

States of CPU Utilization

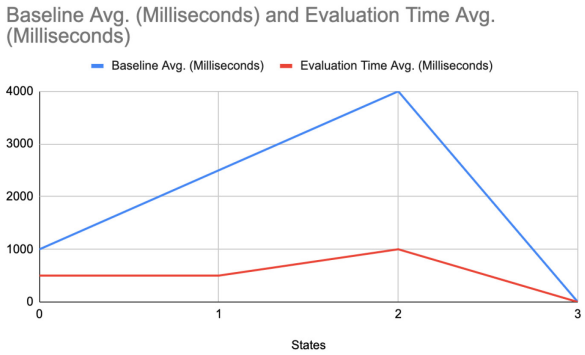
As the overall system CPU utilization increases and decreases, if it transitions to another state the time between entering the previous state and entering the current state will be recorded and evaluated.

The tests utilized several baselines, varying in length of runtime. One baseline was created within 5 minutes, the second baseline was created within ten minutes and the last baseline was created within fifteen minutes. The testing concept was to evaluate the effectiveness of identification depending on how long the baseline was established.

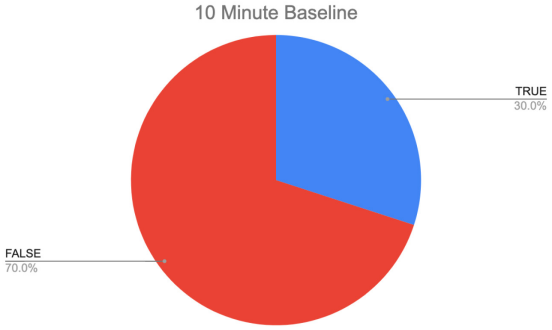
The baselines were created using a python script that monitored the CPU utilization and identified if the value would qualify for a state transition. The evaluation was conducted every five milliseconds. In the case that a state transition occurred, the time was recorded as a single state change for the previous state and added to an average of time for that specific state.

The model was tested on a virtual machine operating with the Windows 7 operating system. The virtual machine was infected with a sample of the Trojan Horse “Trojan.Defi.Gen.1”. The range for anomaly identification was 50 milliseconds. The IDS system was allowed to run for 5 minutes for each test.

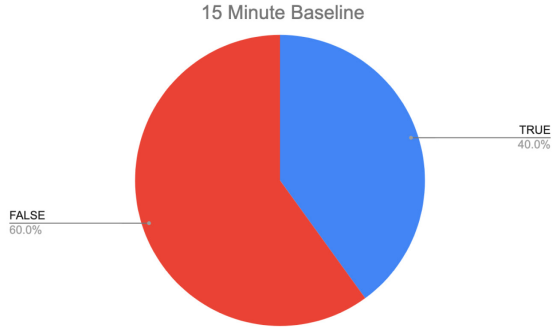
Results



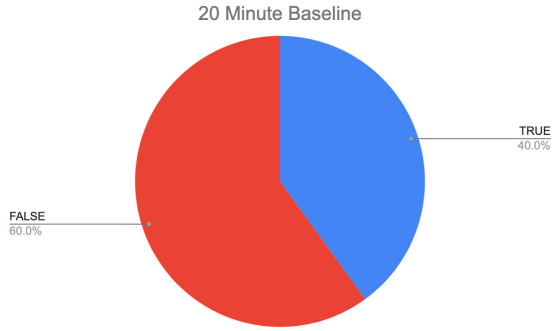
The chart above represents a sample of a test in which the times were recorded between each transition. As reflected in the graph for this particular test, the evaluation time transition was below the baseline time. This chart would reflect anomaly detection.
However, after conducting the test 10 times for each baseline, the results did not prove the theorized model effective.



In the 10 minute baseline, the model only identified an anomaly for 30% of the tests.



In the 15 minute baseline, the model only identified an anomaly for 40% of the tests.



In the 20 minute baseline, the model only identified an anomaly for 40% of the tests. As reflected through the data, this model is not effective in detecting anomalies with CPU utilization.

VI. Summary

In conclusion, the Internet of Things is a rapidly evolving arena of technology. It provides a wide range of innovative solutions. It is composed of three essential layers, the network layer, the perception layer and the application layer. Although additional models exist, such as the 5 layer model, which includes a business layer, processing layer and transport layer, they were not discussed in this research. Each of these layers are each respectively susceptible to cyber attacks. However, intrusion detection systems and intrusion prevention systems are designed to protect systems from cyber attacks by using various methods such as analyzing traffic behavior on a host system. An IDS was proposed using the semi-markov chain process. The IDS would use an averaged value utilizing SMC to develop a baseline of regular activity to compare against anomaly behavior. Upon testing the proposed model, it can be concluded that the implementation of Semi-Markov chains with CPU utilization state sectors was not effective.

ACKNOWLEDGEMENTS

This work is partly supported by the National Science Foundation CyberCorps: Scholarship for Service program under grant award# 1754054.

V. References

[1] arvindpdmn devbot5S, "IoT Security Model," *Devopedia*, Jun. 25, 2021. https://devopedia.org/iot-security-model (accessed Feb. 7, 2022).

[2] K. D. Foote, "A Brief History of the Internet of Things," *DATAVERSITY*, Aug. 16, 2016. https://dev.dataversity.net/brief-history-internet-things/ (accessed Feb. 12, 2022).

[3] C. Gregersen, "A Complete Guide to IoT Protocols & Standards In 2021," *Nabto*, Dec. 18, 2020. https://www.nabto.com/guide-iot-protocols-standards/ (accessed Feb. 13, 2022).

[4] K. Grover, A. Lim, and Q. Yang, "Jamming and anti-jamming techniques in wireless networks: a survey," *IJAHUC*, vol. 17, no. 4, p. 197, 2014, doi: 10.1504/IJAHUC.2014.066419.

[5] "Grover et al. - 2014 - Jamming and anti-jamming techniques in wireless ne.pdf." Accessed: Feb. 4, 2022. [Online]. Available: https://scholarworks.montana.edu/xmlui/bitstream/handle/1/9021/jamming.pdf?sequence=1

[6] A. Patel, "3 Layer Architecture of IoT | BelkIoT | Perception, Network, Application," *BelkIoT*, Jan. 02, 2021. https://belkiot.in/3-layered-architecture-of-iot/ (accessed Feb. 12, 2022).

[7] M. Rassam, A. Zainal, M. Maarof, and M. Al-Shaboti, "A sinkhole attack detection scheme in Mintroute wireless Sensor

Networks," Nov. 2012, pp. 71–75. doi: 10.1109/ISTT.2012.6481568.
[8] S. Richardson, "Table 38 Buffer Overflow - Network Security," *Cisco Certified Expert*, Jan. 06, 2021. https://www.ccexpert.us/network-security-2/table-38-buffer-overflow.html (accessed Feb. 12, 2022).
[9] R. Sahner, K. S. Trivedi, and A. Puliafito, "Semi-Markov Chains," in *Performance and Reliability Analysis of Computer Systems: An Example-Based Approach Using the SHARPE Software Package*, R. Sahner, K. S. Trivedi, and A. Puliafito, Eds. Boston, MA: Springer US, 1996, pp. 143–149. doi: 10.1007/978-1-4615-2367-3_8.
[10] P. Sethi and S. R. Sarangi, "Internet of Things: Architectures, Protocols, and Applications," *Journal of Electrical and Computer Engineering*, vol. 2017, p. e9324035, Jan. 2017, doi: 10.1155/2017/9324035.
[11] "49 busted in Europe for Man-in-the-Middle bank attacks," *Naked Security*, Jun. 11, 2015. https://nakedsecurity.sophos.com/2015/06/11/49-busted-in-europe-for-man-in-the-middle-bank-attacks/ (accessed Feb. 12, 2022).
[12] "Internet of Things Security Vulnerabilities: It's All About Buffer Overflow," *Altium*, Dec. 15, 2017. https://resources.altium.com/p/internet-of-things-security-vulnerabilities-all-about-buffer-overflow (accessed Feb. 15, 2022).
[13] "IDS: History, Concept and Terminology," *OSTEC | Segurança digital de resultados*, Mar. 01, 2018. https://ostec.blog/en/perimeter/ids-history-concept-terminology/ (accessed Feb. 14, 2022).
[14] "The History of the Internet of Things |," *Eleven Fifty Academy*, Dec. 28, 2020. https://elevenfifty.org/blog/the-history-of-the-internet-of-things/ (accessed Feb. 13, 2022).
[15] "Application Layer Protocols for IOT : IOT Part 11," *Engineers Garage*, Mar. 06, 2021. https://www.engineersgarage.com/application-layer-protocols-for-iot-iot-part-11/ (accessed Feb. 14, 2022).
[16] "State of IoT 2021: Number of connected IoT devices growing 9% to 12.3 billion globally, cellular IoT now surpassing 2 billion," *IoT Analytics*, Sep. 22, 2021. https://iot-analytics.com/number-connected-iot-devices/ (accessed Feb. 10, 2022).
[17] "Cross Site Scripting (XSS) Software Attack | OWASP Foundation." https://owasp.org/www-community/attacks/xss/ (accessed Feb. 10, 2022).
[18] "Emotet Malware | CISA." https://www.cisa.gov/uscert/ncas/alerts/aa20-280a (accessed Feb. 11, 2022).
[19] "Global IoT and non-IoT connections 2010-2025," *Statista*. https://www.statista.com/statistics/1101442/iot-number-of-connected-devices-worldwide/ (accessed Feb. 4, 2022).
[20] "Markov Chain - an overview | ScienceDirect Topics." https://www.sciencedirect.com/topics/computer-science/markov-chain (accessed Feb. 14, 2022).
[21] "Trojan Horse Software Attack | OWASP Foundation." https://owasp.org/www-community/attacks/Trojan_Horse (accessed Feb. 12, 2022).
[22] "What is anti-replay protocol and how does it work?," *SearchNetworking*. https://www.techtarget.com/searchnetworking/definition/anti-replay-protocol (accessed Feb. 12, 2022).
[23] "What is the Internet of Things (IoT)?" https://www.oracle.com/internet-of-things/what-is-iot/ (accessed Feb. 10, 2022).
[24] "What is a man-in-the-middle attack? How MitM attacks work and how to prevent them | CSO Online." https://www.csoonline.com/article/3340117/what-is-a-man-in-the-middle-attack-how-mitm-attacks-work-and-how-to-prevent-them.html (accessed Feb. 9, 2022).