

Design of an anomaly-based intrusion detection system setup on a Raspberry Pi

Eddy Jesus Vargas

Knight Foundation School of Computing & Information Sciences

Florida International University

Miami Florida, U.S.A.

evarg062@fiu.edu

Abstract—In this paper we will be covering how we can implement an anomaly based intrusion detection system(IDS) in an IoT(Internet of Things) device so as to improve the security of these devices wherein they essentially forgo security in lieu of comfort and usability. This will involve the utilization of an open source network analysis program known as Zeek(previously named Bro) to help design an Intrusion Detection System that will be specific to the aspects and details of an Internet of Things device. In this case, we will be using the Raspberry Pi Model 4b as a test device for the IoT application. This will be the device having the IDS system installed and it will be run over a period of time to show the IDS what is anomalous behavior and what is the user's standard practice of use. At the end of this paper, it will hopefully provide an accurate basis for implementing an anomaly based Intrusion Detection System in an Internet of Things device so as to provide a better platform security-wise for these devices. This should then make the cyberspace more secure overall.

I. INTRODUCTION

Everyone nowadays has an Internet of Things(IoT) device even if they do not know it. These devices include various things from smartwatches that people wear everyday, to thermostats that individuals have configured to change the temperatures at certain times. There are even some individuals with smart refrigerators with is also an IoT device. While all of these devices can help us feel a lot more connected to the ever growing world around us as well as make some daily habits and chores a lot simpler, they also might put the users security at risk.

This is unfortunately the case in a world where these devices do not see many security features being implemented in them to make them as cheap and obtainable as possible. The dilemma is then, how to make these devices more secure in this day and age. This can be accomplished by many methods and it would take some time to verify and compare said methods to determine which one would be the best fit. These can include things such as network segmentation, through methods such as VLANs or even implementing an Intrusion Detection System on these devices.

Contributions: Any contributions that this paper will present is presented below

- Proposing a framework for configuring an Intrusion Detection System for an Internet of Things device in an efficient manner.

- Providing methods for seamless integration for the Intrusion Detection System to provide alerts and to remediate any issues.
- Showcasing how the technology works to better educate the public.

Organization: This paper will follow the following structure as a general guideline: Section I which we have just gone over will be the introduction, providing a good amount of background on the topic and any necessary knowledge of the paper itself. Section II will provide more information on IoT devices so the reader fully understands how they function, where their strong suits and weak points lie and how we can go about securing these devices. Section III will go over Intrusion Detection Systems in depth and also introduce our Intrusion Detection System Solution and how it functions. In section IV, we present the Security Information and Event Management(SIEM) that we will be using to supply alerts for the IDS and the system that we will be using to investigate any alerts and remediate said issues. Finally, in section V we will conclude this paper and summarize the information presented in an easily digestible manner.

In this paper, I will demonstrate how I was able to install an Intrusion Detection System on an Internet of Things device. This will allow us to know if these devices are running up to specifications and if they have not been compromised. If these devices are, in fact determined to be compromised, then we will be alerted about this intrusion and we will be able to begin taking action against said intrusion to contain it and then repair the issue so as to bring the unit back up and retain its functionality.

II. BACKGROUND

Network segmentation is a practice of creating smaller individual networks to help easily control different devices as well as the flow of network traffic.(Alto, 2023) While it may not seem like it has much to do with security, the topic itself actually lends itself heavily towards protecting devices from infiltration as well as from data extraction. This is due to the fact that while, one sub network might have been infiltrated by an attacker and they have gained access to every device in that network, they will not directly have access to devices in another network if they are segmented correctly. This is

then a decently cost effective method of preventing a threat actor from moving laterally across a companies network, in turn increasing security.

Intrusion Detection Systems, on the other hand may be more costly to implement, but you will be alerted if there is an intrusion on a device that is being monitored.(Alto, 2023) This is done by having the device actively monitor the network and can detect an intrusion based off of different systems such as rule-based system, an anomalistic-based system(which is the focus of this paper) and a signature-based system. There are also different types of models for the Intrusion Detection Systems.(Dig, 2023) These include models such as a Network-Based Intrusion Detection System(which monitors an entire network for infiltration), a Host-Based Intrusion Detection System(which is the basis for this paper and works by having the System installed on the host device so as to monitor that specific computer/device).(Dig, 2023) There are also other models which, while mentioned will not be explained and are merely for education's sake. These, in turn, are Protocol-Based Intrusion Detection Systems, Hybrid Intrusion Detection Systems, and Application Protocol-Based Intrusion Detection systems. These may also seem very similar to firewalls, but they are in fact different and work in a different manner. Furthermore, firewalls will not be the focus for this paper.

III. RELATED WORK

There are also other individuals who have done research on this topic as the security of Internet of Things devices has been questioned since their arrival into most networks. They also vary in specifics as one goes over smart homes, one utilizes Intrusion Prevention Systems, and one uses a Network Intrusion Detection System.

For example, the paper "Intrusion Detection in Internet of Things Systems: A Review on Design Approaches Leveraging Multi-Access Edge Computing, Machine Learning, and Datasets" by Eric Gyamfi and Anca Jurcut explains a lot of the issues that have arrived due to how rapidly IoT applications have grown in relation to increased network data and computational complexity with devices with low resource specifications. (Gyamfi Jurcut, 2022)Furthermore, it goes over how vulnerable they are to cyber attacks because of their limitations and how they cannot run bare security software and have identified it as a significant risk.(Gyamfi Jurcut, 2022) Lastly, it offers up multi-access edge computing, which works as an IoT gateway in a layer between mobile devices and the cloud, to fix these limitations and offload computational tasks to prevent resource exhaustion.

Furthermore, the paper "Intrusion detection and prevention system for an IoT environment" by Ajay Kumar et al goes into the importance of how we should bolster the defenses of IoT devices and networks as they are so vulnerable to unauthorized access. It also, goes over how one can go about setting up not only an Intrusion Detection System, but also a Network-based Intrusion Prevention System which would not only detect the unauthorized access attempts and intrusions,

but also do what it can to prevent it. (Lara et al., 2023) The NBIPS mentioned in this paper is typically situated behind the firewall to monitor network activity and prevent any malicious usage of devices.(Kumar et al., 2022)

Lastly, the paper "Smart home anomaly-based IDS: Architecture proposal and case study" by Antonio Estepa et al provides a structure for a Smart Home Intrusion Detection System that utilizes anomaly detection to detect any intrusions or bad actors. It also illustrates how it needs custom security designs due to the diverse and eclectic technologies that are involved in the IoT-space. It further shows the architecture that the paper provides so as to setup anomolay detection from application-layer information and general network traffic of an entire Smart Home to better monitor the security.

IV. IOT SYSTEMS IN DEPTH

Internet of Things devices are devices that can conduct simple functions, we have seen these such as things similar to that of Amazon Alexa and even Apple watches. These kinds of devices, while lacking various security features are connected to not only the internet, but also the network contained in your home. This means that when the device itself is hacked, it allows them to pivot to other devices in your network as they would already be in your network.

There are various threats when it comes to IoT devices that stem from different issues and these can all lead to your entire network being taken over depending on the level of attack. To begin with, they utilize a variety of technologies for any messages that they send which in turn makes it vary difficult to make a standard when it comes to the different types of devices that are housed under the IoT umbrella.(Net, 2023) Furthermore, there is a significant lack of security with these devices, as previously mentioned as not only is there rarely any security for the actual software that is implemented on the machine, but there is also a lack of security when it comes to the hardware components themselves, leaving them to be observed, examined, and reverse engineered to understand how they work before, ultimately disabling it or worse. In addition, not only is there an issue with the actual devices themselves, but there is also an issue with the user set that they typically market towards.(Gillis et al., 2023) This is due to the fact that many of these individuals have no security awareness and do not attempt to implement even the simplest of practices for security when it comes to these systems.(Gillis et al., 2023) Finally, one of the major dilemmas that these devices suffer from tend to be limited computational resources such as ram, storage and even CPU speeds. This can be seen in many different devices, including the Raspberry Pi Model 4b that was used for this project. (Gillis et al., 2023)

This can be seen in Figure 1 where we showcase the various RAM availability for the Raspberry Pi model 4 which was taken from the Raspberry Pi website where provide the specifications for their devices.(Pi, 2023) This is not the only instance either. In figure 2, we can see the specifications of various Amazon Alexa capable devices provided by Amazon which shows that the highest availability is 512Mb and the

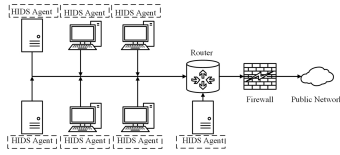


Fig. 4. In this figure, we see how an agent-based Intrusion Detection System functions and how it is able to get an accurate representation of the entire network without relying on network traffic due to each device on the network having an agent deployed, which would provide the data that they are receiving and sending so as to allow careful inspection of the devices and bolster the security of the machines.

B. Solution

In this paper, our Intrusion Detection System solution is Zeek, which was previously named "Bro". (Ziobro, 2022) It was created in 1995 by Vern Paxson and is still used today by, not only individuals, but also government agencies and universities to help monitor networks to help accelerate threat detection and incident response times. (Ziobro, 2022) Furthermore, it is free to use as it is open source meaning that you can also improve it or modify it for your use case. It works by pulling more than 400 data entries from data traffic and also pulls data from more than 35 protocols all ranging from layer 3 to layer 7. (Kerravala, 2018) This is then compressed to 1 percent of the file size while retaining the integrity of the traffic capture and organizing said data by protocols in a PCAP file. (Kerravala, 2018)

In addition, Zeek includes its own programming language so we can create our own scripts to further make more custom detection rules for our networks and in this case the Raspberry Pi. (Edie, 2023) It would also allow for integration to a Security Information and Event Management (SIEM) system. This would allow for not only combining the information from Zeek to an endpoint solution, but would also allow for easier analysis. (Edie, 2023) It can also accept sensors that has throughput of more than 25Gb per second and can be installed quickly so as to be able to spend more time getting the configuration right for the system it is being installed on.

Essentially, the many benefits that Zeek can provide involve some of the following: It is an open source tool and integrates seamlessly into other third party tools to better analyze network traffic. (Bozdog, 2022) It can also assist in detecting advanced persist threats, intrusions and even minor anomalies. Furthermore, it can be setup to have encrypted traffic so as to protect the collected traffic that you are going to be analyzing and can be done through "specific raw indicators like JA3 fingerprints and TLS ciphers". Lastly, it is able to provide a usable security language so as to allow for better network monitoring and to assist in improving network security tools through custom scripts that, when customized will allow for a better overall Intrusion Detection System.

VI. ELASTICSEARCH

Elastic search is a Security Information and Event Management system that we have used in this paper in conjunction

with Zeek to provide anomaly intrusion detection possibilities. (Glossary, 2023) Elastic search is useful because not only does it allow you to serve as a base for all of the logs that Zeek provides, it can also integrate other programs to better help manage the security of any one device or even an entire network. (24x7, 2020) Furthermore, it searches any logs quickly due to the fact that it keeps this information in an index to structure JSON-based documents instead of looking through texts directly. These are then placed into indices, based on traits to establish keywords for a more efficient search. (Quach Hong, 2023)

There are various components that are involved in the utilization of elastic search. Some of these components include Clusters, Nodes of varying kinds to perform specific functions, Indices, Types, Documents, Mapping Types, Meta Fields, Properties, Simple Data Types, Special Data Types, Complex Data Types and Shards. (Quach Hong, 2023)

For example, Clusters are, in essence, different servers that work to provide indexing and search possibilities and can range from one node to thousands depending on how you are using them, which ends up resulting in the Elasticsearch cluster. (Singh, 2023) Furthermore, a node is any machine whether it be physical or virtual that contains that data and computational power to index and search said data. Their main jobs tend to be monitoring the health of the cluster, indexing, aggregation, and searching through data and also storing said data. The different kinds of nodes that can be implemented are those of the data node (for storage and computing), the master node (which is mainly an administrative node), and a coordinating node (which sends requests between master and data nodes to balance traffic).

Furthermore, the indices are essentially a container that stores data in manner similar to that of a relational database by sorting documents by whether they are logically or characteristically related, are named in a lowercase manner, and utilizes operations like add, delete and update. (Technologies, 2021) Type is a type of grouping belonging with indices and are based on the properties the document itself such as dates or strings. A document is what is indexed in the end by Elasticsearch which is simply a documented formatted in JSON so that it can be produced by the search. (Drudge, 2023) Mapping Types is, in essence, a method of configuring the index to determine how to use each JSON field in regards to indexing, and is formatted with Meta Fields and Properties. (Technologies, 2021) Meta fields involves supplemental information of the document, which can be edited to be align with how the index should register the document and there ten meta fields to configure. (Drudge, 2023) Properties contain data type information as well, but it is typically more focused such as the specifications of a device or even the web browser. (Team, 2023)

Lastly, we have simple data types, special data types, complex data types and shards. Simple data types are typically text based and store data such as product description, keyword based like an individuals age or even numeric data like short and float. (Team, 2023) Special data type involve geometric

data such as location or even the shape of the area a user may be in. Complex data types involve JSON oriented structures. This includes the standard object structure and nested object architecture. Finally, shards assist with horizontal scalability as they are a subset of indices and can help split up the incredible amount of documents that an indice is supposed to hold and can sometimes be seen in the same node of a cluster. Furthermore, they tend to influence the degree of parallelism involved in indexing and searching and the quantity of shards can be changed when an index is created.

VII. PERFORMANCE EVALUATION

Results for this paper will be mostly based on observations as the majority of the analysis involving the system that has been designed cannot be quantified. Overall, it would be better if the system could be one tool instead of two tools as issues with one can cause an issue with the entire Intrusion Detection System. As far as the performance overall, the system is able to detect anomalistic behavior utilizing the provided Zeek logs. The system can be improved upon, however as the system takes about 30 minutes to register said anomaly as well as the other rules that are deployed for the system. While this may work in a business setting with no issue, for the purpose of this project a quicker detection time would be more effective. It also brings up the question of will a quicker detection time also increase the false positive rate.

VIII. CONCLUSION

Overall, in this paper we have gone over various sections including introduction, background, the related works, an in depth look at IoT systems as well as broad discussion over them, an overlook of Intrusion Detection Systems and what our Intrusion Detection System is comprised of, how it pairs with Elasticsearch so as to provide anomalistic intrusion detection and an evaluation of the performance of the system as a whole.

The introduction elaborates on what the project will consist of as well as how it is organized and what it will contribute. The background section allowed for an explanation of the topics to know to better understand the paper overall and introduces the basics of some of the tools that we utilized in this paper so we can explain them more in depth later in the paper. The related works section explained how investigation into IoT systems is not a new topic and how there have been many attempts to design an Intrusion Detection System to help quell the issues that these devices have with security. The in-depth look at IoT systems informed the reader on the general downfalls of Internet of Things devices as a whole, as well as the downfalls of the Raspberry Pi system that we utilize in this project.

We then had the IDS Introduction and Solution section which explained the various types of Intrusion Detection Systems and how one may use them as well as our solution "Zeek" as well as how it works and how we utilized it. We then had the Elasticsearch section after the IDS solution section to show they went in hand and went about providing the anomalistic intrusion detection with the logs provided by Zeek

as well as other capabilities. Lastly, we evaluated how the system performed as a whole based off of observation as well as some mild metrics to showcase the architecture that has been presented in this project.

There are areas for improvement, such as seeing how this system works in a live network as well as how it may be able to work with a full network instead of just one system that it is monitoring. Would the false positive and false negative rates change or stay the same and how would the anomaly detection manage an entire network? Furthermore, there are plenty of future studies that can be conducted off not only this paper, but also the related works previously mentioned and even others that were not discovered at the time of this writing. There could be a possible study done on how this system would work implemented on switches or even another possible study done where the basis of the project uses the IDS/IPS tool Snort instead of Zeek to supply a SIEM with the data and how those two systems would be able to interact.

REFERENCES

- [1] F. Net, "Top IOT device vulnerabilities: How to secure IOT devices," IoT Device Vulnerabilities, <https://www.fortinet.com/resources/cyberglossary/iot-device-vulnerabilities> (accessed Nov. 27, 2023).
- [2] T. 24x7, "How to install Elk or elastic stack (Elasticsearch, Kibana, Logstash, and Winlogbeat) on windows," How To Install ELK Or Elastic Stack (Elasticsearch, Kibana, Logstash, and Winlogbeat) on Windows, <https://elasticsearch.tutorials24x7.com/blog/how-to-install-elasticsearch-kibana-and-logstash-elk-elastic-stack-on-windows> (accessed Nov. 27, 2023).
- [3] E. Bozdog, "Anomaly-based intrusion detection system using unsupervised ML approach," Medium, <https://medium.com/hootsuite-engineering/anomaly-based-intrusion-detection-system-using-machine-learning-a18e88694ce0> (accessed Nov. 27, 2023).
- [4] A. Lara, V. Mayor, R. Estepa, A. Estepa, and J. E. Díaz-Verdejo, "Smart home anomaly-based ids: Architecture proposal and case study," Internet of Things, <https://www.sciencedirect.com/science/article/pii/S2542660523000963> (accessed Nov. 27, 2023).
- [5] R. Pi, "Raspberry pi 4 model B specifications," Raspberry Pi, <https://www.raspberrypi.com/products/raspberry-pi-4-model-b/specifications/> (accessed Nov. 27, 2023).
- [6] B. Mars, "Sizing Up CPU, Memory, and Storage for Your Alexa Built-in Device," Amazon, <https://developer.amazon.com/en-US/blogs/alexa/post/2a32d792-d471-4136-8262-79962a2b4d72/cpu-memory-and-storage-for-alexa-built-in-device.html> (accessed Nov. 27, 2023).
- [7] P. Alto, "What is network segmentation?," Palo Alto Networks, <https://www.paloaltonetworks.com/cyberpedia/what-is-network-segmentation> (accessed Nov. 27, 2023).
- [8] P. Alto, "What is an intrusion detection system?," Palo Alto Networks, <https://www.paloaltonetworks.com/cyberpedia/what-is-an-intrusion-detection-system-ids> (accessed Nov. 27, 2023).
- [9] S. Dig, "What is HIDS (host-based Intrusion Detection System)?," Sysdig, <https://sysdig.com/learn-cloud-native/detection-and-response/what-is-hids/> (accessed Nov. 28, 2023).
- [10] M. F. Elrawy, A. I. Awad, and H. F. A. Hamed, "Intrusion Detection Systems for IOT-based Smart Environments: A Survey - Journal of Cloud Computing," SpringerOpen, <https://journalofcloudcomputing.springeropen.com/articles/10.1186/s13677-018-0123-6/figures/6> (accessed Nov. 29, 2023).
- [11] Z. Ziobro, "What is Zeek and why is it important?," NetQuest, <https://netquestcorp.com/3-minute-crash-course-on-zeek/> (accessed Dec. 3, 2023).
- [12] Z. Kerravala, "Zeek: A free, powerful way to monitor networks, detect threats," CSO Online, <https://www.csoonline.com/article/566421/zeek-a-free-powerful-way-to-monitor-networks-detect-threats.html> (accessed Dec. 3, 2023).

- [13] G. Glossary, "Definition of security information and Event Management (SIEM) - gartner information technology glossary," Gartner, <https://www.gartner.com/en/information-technology/glossary/security-information-and-event-management-siem> (accessed Dec. 4, 2023).
- [14] S. Quach and C. Hong, "Elasticsearch: What it is, how it works, and what it's used for," Knowi, <https://www.knowi.com/blog/what-is-elastic-search/> (accessed Dec. 5, 2023).
- [15] T. Singh, "Everything You Need to Know About Elasticsearch," Everything you need to know about Elasticsearch, <https://www.nitorinfotech.com/blog/everything-you-need-to-know-about-elasticsearch/> (accessed Dec. 5, 2023).
- [16] A. Kumar, A. Kumar, M. R. Ghalib, and X. Cheng, "Intrusion detection and prevention system for an IOT environment," Digital Communications and Networks, <https://www.sciencedirect.com/science/article/pii/S2352864822001201> (accessed Dec. 5, 2023).
- [17] I. Center, "Structure your paper," IEEE Author Center Conferences, <https://conferences.ieeeauthorcenter.ieee.org/write-your-paper/structure-your-paper/> (accessed Dec. 6, 2023).
- [18] E. Gyamfi and A. Jurcut, "Intrusion detection in internet of things systems: A review on Design Approaches Leveraging Multi-Access Edge Computing, machine learning, and datasets," Sensors (Basel, Switzerland), <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9143513/> (accessed Dec. 5, 2023).
- [19] V. Technologies, "Elasticsearch 101: Fundamentals & Core Components," Medium, <https://medium.com/velotio-perspectives/elasticsearch-101-fundamentals-core-components-a1fdc6090a5e> (accessed Dec. 6, 2023).
- [20] C. Drudge, "Elastic search 101: Getting started with Elasticsearch," Medium, <https://levelup.gitconnected.com/elastic-search-101-de7ef8aa2469> (accessed Dec. 6, 2023).
- [21] D. Team, "Elasticsearch: Everything you need to know about this software," DataScientest, <https://datascientest.com/en/elasticsearch-everything-you-need-to-know> (accessed Dec. 6, 2023).
- [22] A. S. Gillis, B. Posey, and S. Shea, "What are IOT devices?: Definition from TechTarget," IoT Agenda, <https://www.techtarget.com/iotagenda/definition/IoT-device> (accessed Dec. 7, 2023).
- [23] J. A. Edie, "Network security monitoring with zeek," Pluralsight, <https://www.pluralsight.com/paths/network-security-monitoring-with-zeek> (accessed Dec. 7, 2023).