INTERNATIONAL RESEARCH JOURNAL OF ENGINEERING & APPLIED SCIENCES
ISSN: 2322-0821(O) ISSN: 2394-9910(P)
VOLUME 10 ISSUE 3
Jul 2022 - Sep 2022
www.irjeas.org

Original Article

Artificial Intelligence Based Zero Trust Network

*Priya Parameswarappa [1]

[1]*Research Scholar, School of Information Technology, University of the Cumberland's, Kentucky, USA*
pparameswarappa69940@ucumberlands.edu https://orcid.org/0000-0003-2059-6043

**Corresponding Author – pparameswarappa69940@ucumberlands.edu*

DOI - https://doi.org/10.55083/irjeas.2022.v10i03013

Received: 15 July 2022; Received in revised form: 21 August 2022; Accepted: 01 September 2022

Abstract: Model-based security metrics are an emerging topic of cyber security research that focuses on assessing an information system's risk exposure. We propose an end-to-end solution with the deployment of a zero-trust network utilising Artificial Intelligence in this article to understand the security posture of a system before it is rolled out and as it matures. The major part contains a discussion about the key methods and techniques which was utilized in the development process and simplified operation principles of each developed process. Some developed processes were tested practically to evaluate the problems in the processes. Modules for automatic processing and data analysis were also developed. These modules can be connected in case it is needed. The most important data collection methods were benchmarked to detect problematic situations in the operation in different realistic situations. With the perception from the benchmark test, the problematic parts of the data collection were discovered and proposals for the solution were made which could be developed and tested in the next iterations of the development process. Working Artificial intelligence-based detection and data enrichment methods were created. The results of the article allow multiple continuous research and development projects related to data collection and data analysis with statistical and artificial intelligence-based methods.

Keywords: Artificial Intelligence, Anomaly detection, network monitoring, Zero Trust Network Access.

1. INTRODUCTION

The amount and versatility of different devices connected to the networks have increased significantly in ten years. Cisco VNI has estimated that devices and connections per capita will grow from 2.4 to 3.6 between the years 2017 and 2019 [1-4]. More than ten years ago, most devices connected to the network were mostly traditional workstations and servers. In the 2010s number of smart appliances such as smart phones & tablets, TV's and different IoT Devices has increased Significantly [15]. The evolution of the IT-industry has caused the trend in Cyber security to be changed. In the Solutions Review article that was written by Ben Canner in autumn 2019, the five most common attack vectors in endpoint security were employees, mobile devices, IoT, endpoint ports and applications [5].

Another growing trend in organizations is BYOD (Bring Your Own Device) culture. The culture gives benefits for both employers and employees. Employees feel comfortable when using their own devices at work and employers might be happy that there is no need for the acquisition of dedicated workstations which costs money. In most of the cases, the employer does not have

control of the security of the employee's BYOD devices which leads to devices can be unmonitored for long periods which allows data transfers in and out from the device without any regulations [5].

The third growing trend in attack vectors in the organization is the IoT, Internet of Things. IoT devices do not usually include any security protection and after the installation of the device, it is forgotten causing a blind spot in the network, which an attacker can penetrate.

The Enterprise Strategy Group has mentioned 5 top challenges in threat detection and response. Their recent research had 379 respondents mostly in the fields of cyber security and IT-professionals. 36% of the respondents said that their teams have spent most of the time to address high priority issues causing a stop for the evolving of strategy and process improvement. 30% said there are one or several blind spots. 26% said that their threat detection and response is anchored by manual processes that reduce their ability to keep up with the threats [6].

Meanwhile, an increasing number of different smart devices and other devices connected to the network have helped people in their lives, the number of attack vectors has increased and threats against, cyber-attack have arisen. While one or more of the employees might make the mistake someday and there might not be a legitimate possibility for employers to install required security software into employee's device, there aren't many choices as the solution [7].

An anomaly-based network traffic monitoring solution may help in the detection of unwanted data transfers inside the organizational networks. Because almost all malware or intruders cause network traffic at some point of their session, the network traffic monitoring and analysis is an ideal source for information about the health of the network or specific device [8].

2. LITERATURE REVIEW

There are available several other products and services that are related to the topic and which use artificial intelligence for solving related cyber security problems or for finding anomalies in collected data. In the field of AI and cyber security, there are solutions such Pattern EX AI2 developed by MIT and CSAIL, Amazon Macie, Cyberlytinc WWW-threat detection tool, Cylance Protect, Drak trace, Deep Instinct, Spark Cognition Deep Armor and Vectra Network Cognito. Cisco DNA Center, DNAC is a management system for the intent-based enterprise networks. It uses artificial intelligence and machine learning for monitoring the network proactively. It also helps in network optimization and troubleshooting. DNAC leans to group-based policies, micro-segmentation and AI to improve network security [10-12].

IBM Qradar is a security information and event management, SIEM powered with artificial intelligence and machine learning. Like most of the SIEM solutions, it allows centralized insight for the traffic and logs that collect around the devices. It correlates and aggregates gathered data into single events to accelerate incident analysis and remediation [9].

For software development, there are several libraries available that are useful for artificial intelligence and machine learning implementation. Few useful ones would be Tensor Flow which is an end-to-end open-source platform for machine learning developed by Google. NumPy is the fundamental package for scientific computing for Python programming language. In the field of data collection and management, there are available solutions like ELK Stack and Splunk [11-13]

Network architectures like Zero Trust Networks keep extended visibility of its users and devices mandatory. It helps security teams to gauge security risk related to network users and device. Passwords and other authentication methods may not be enough to ensure that the correct person is using the credentials. A user behaviour analytics is in close relationship with SIEM and it focuses to determine user activity in the IT-networks and systems to indicate anomalies in user behaviour. This might be the key thing in the future to detect an intruder in the network. It cannot prevent an intruder from getting into the system, but it can be useful for quick detection of intrusion [14-18].

3. METHODOLOGY

In the original plan, it was supposed to receive the Net Flow records from the core switches of the network directly without the dedicated packet capture solution. VSS, Virtual Switching System feature that is in use in the core switches caused problems for the integrity of the Net Flow data. About 50% of the traffic flow information was newer exported by the Cisco 4500-X switches. The reason for that was that the switch which was in the standby state was not able to export the flow records of the traffic while another switch was active.

Centralized data collection is one of the core parts of this article. It consists of data collectors and processors. The collected data is used to create a baseline model of the data and detect anomalies

from the collected data against the baseline model and create events from the detected anomaly.

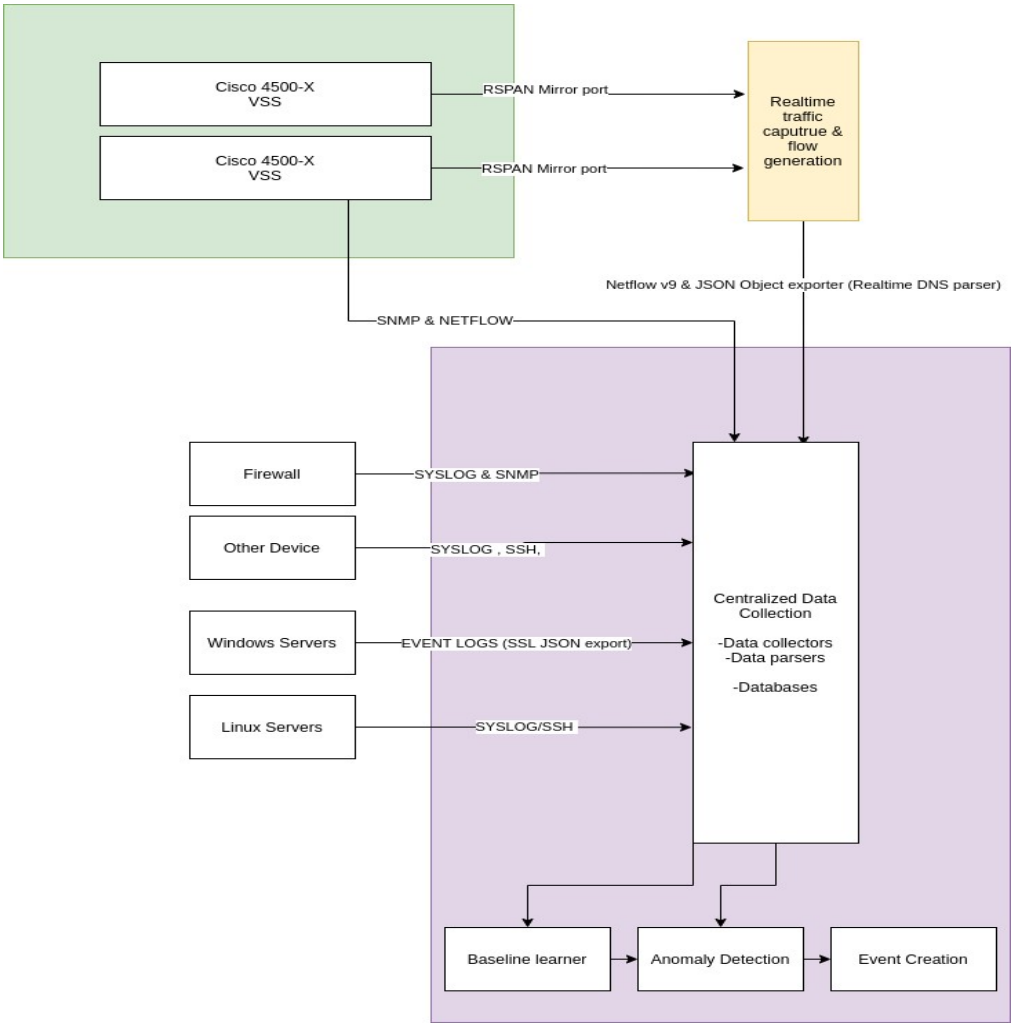

Figure 1- Working Methodology

4. IMPLEMENTATION

The implementation of the data gathering process is the core part of this article. Data gathering techniques were selected to be universal or freely available to allow implementation without obstacles caused by financial resources. The data gathering processes were programmed by using Python3.10.5. Each of the processes was programmed as modules, which allows easy reuse of code in other projects.

For this article project, four data collection techniques were implemented. The methods or techniques are the following:

- Real-time network traffic capture
- Microsoft Windows Event Log forwarder
- Syslog server
- SSH stream reader

Design each of them tries to follow similar design with each other and confidently, integrity and availability in the case that is reasonable, either possible to implement with the corresponding technology. Each of the datasets that are output from the module is in JSON format, JavaScript Object Notation format.

In the collector initialization phase, the required variables are set for the module. After that, the start call initializes the worker thread for the module and it becomes active. Each of the collectors has collector statistics, which stores the numerical statistics of the collector. This statistic can be used as data for EPS (Events Per Second) metric or calculating the latency of the data collector.

The operation of the worked thread is relatively simple: Receive data and fire call-back functions on different events and update collector statistics. Parsing the collected data in the case that is required to do and other data handling is also done in the worker thread. In the case the ingress volume of the data collector process is high and the data handling process is relatively heavy, it is recommended to separate the data handler to another thread to prevent the saturation of the collector.

This design prevents other collectors from blocking other collectors and the system operates in non-blocking mode. The modular design also allows easy integration to the other implementations of AI processes or other data science projects relating to the topic.

5. RESULTS AND DISCUSSION

Traffic flow classification with artificial neural networks was tested for the further implementation of the traffic behaviour classification process. Due to the time taken in model compilations, the complete implementation of this process was moved to the next development cycle. The method uses supervised learning with predefined datasets. The process automatically classifies the flow behaviour to categories such as:

- HTTP Large activity
- HTTP Small activity
- SSH small activity
- SSH large activity
- SNMP poll
- ICMP echo/reply

The test implementation of flow classification by using an artificial neural network was created with Keras library, which is a versatile and simple deep learning module for Python 3.10.5 programming language. Keras library is a user-friendly interface for the Tensor Flow library.

src port	dst port	octets	packets	app_id	output
49141	22	120	2	(ssh)	1 (SSH small activity, input)
49141	22	140	2	(ssh)	1 (SSH small activity, input)
49141	22	400	3	(ssh)	1 (SSH small activity, input)
22	49141	500	3	(ssh)	2 (SSH small activity, output)
22	49141	620	2	(ssh)	2 (SSH small activity, output)
22	49141	700	3	(ssh)	2 (SSH small activity, output)
48134	443	6000	5	(tls)	3 (TLS small activity, input)
48134	443	12498	8	(tls)	3 (TLS small activity, input)
48134	443	13311	10	(tls)	3 (TLS small activity, input)
443	48134	64234	80	(tls)	4 (TLS large activity, output)
443	48134	341241	50	(tls)	4 (TLS large activity, output)
443	48134	123141	30	(tls)	4 (TLS large activity, output)

Malware beacon interval 60 seconds, jitter 0%

In this test, a beacon call home interval value was set to one minute and jitter was set to 0%. The test simulated the case where malware was active and it was communicating with its host.

In step one, 1 hour of NetFlow data was fetched from the database and regular traffic detection was performed in a regular traffic detection module. Fetching and processing one hour of NetFlow data took 19 seconds which contained 465667 flow records.

The parameters for the module were set to the following:

- Minimum interval 10 sec
- Maximum variance between intervals 2 seconds

With these settings, 1283 regular traffic flows were detected. A relatively small minimum interval caused a lot of false-positive results because of a regular web browsing activity, WWW-trackers and protocols such as SNMP, ICMP and NTP for example.

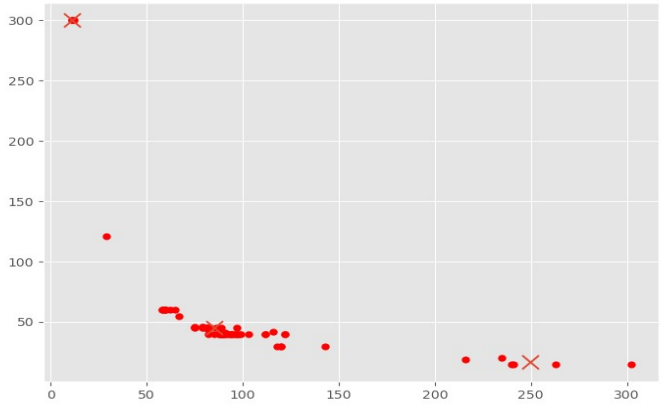

Figure 2- Scatter of all detected regular flows. X = Amount of flows Y = Interval

1283 Data points

In figure 2, the scatter of regular flows in a given time window, a cluster of regular flows can be seen at the point or near where also the command and control channel of the malware is located. It is possible to perform some filtration for the data. Cross marks in the graph present the centre of the cluster of the multiple points. The mark does not have any actual meaning in these tests.

In the next step, all traffic flows that were not present in the whole time window were removed. These flows were caused by transient WWW-browsing for example. Also, the port filter was enabled to filter all ports, except 53(DNS), 80(HTTP) and 443(TLS/HTTPS) away from the dataset.

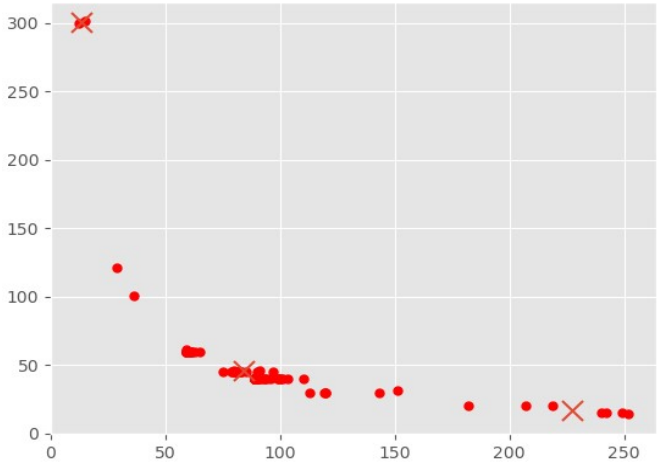

Figure 3- Scatter of regular flow after filtration. X = Amount of flows Y = Interval

111 Datapoints

As shown in figure 3, the filtration did not remove the noise around the command and control channel significantly. After a more detailed look at the data, most of the noise was caused by Microsoft 365, Microsoft Teams, Windows Telemetry Services and Skype which all are used in the workstations located in the environment. Due to the lack of an automatic filtration of IP-addresses with a good reputation, those cannot be automatically filtered away in this stage. By filtering them out manually the noise disappears around the command-and-control channel.

6. CONCLUSION

In this article, the prototype of a modular framework for network security monitoring was successfully implemented. The implementation requires a lot of improvements in areas because some features were also implemented partially just to the needs of the initial research. Also, the codebase would require more standardization to make this for wider use.

The data collection methods would require more security improvements. However, some of them are insecure by default and it might not be possible to improve the security of them without using third party methods to provide a secure communication channel between the collector and the device.

REFERENCES

[1] Abrahamsson, P., Salo, O., Ronkainen, J., Warsta, J., 2017. Agile Software Development Methods: Review and Analysis. arXiv:1709.08439 [cs].

[2] Akbas, E., 2019. SureLog SIEM Data Enrichment [WWW Document]. Medium. URL https://drertugrulakbas.medium.com/surelog-siem-data-enrichment7125a5ed27b1 (accessed 11.12.20).

[3] Anderson, B., Paul, S., McGrew, D., 2016. Deciphering Malware's use of TLS (without Decryption). arXiv:1607.01639 [cs].

[4] Cisco, n.d, Five Steps to Perimeter-Less Security Conklin, K., 2018. What is Network Flow Monitoring? [WWW Document]. URL https://www.whatsupgold.com/blog/network-monitoring/why-you-need-networkflow-monitoring (accessed 05.27.22).

[5] Canner, B., 2019. The 5 Most Common Attack Vectors in Endpoint Security [WWW Document]. Best Endpoint Security Protection Software and Vendors. URL https://solutionsreview.com/endpoint-security/the-5-most-common-attackvectors-in-endpoint-security/ (accessed 05.23.22).

[6] Constantin, L., 2012. Malware uses Google Docs as proxy to command-and-control server [WWW Document]. Computerworld. URL https://www.computerworld.com/article/2493242/malware-uses-google-docs-asproxy-to-command-and-control-server.html (accessed 05.10.22).

[7] Cullen, C., 2016. Password security fail? Add multifactor with user behavior analytics [WWW Document]. TechBeacon. URL https://techbeacon.com/security/password-security-fail-add-multifactorauthentication-behavior-analytics (accessed 05.10.22).

[8] Davide, C., n.d. Figure 5.3: NetFlow collector architecture [WWW Document]. ResearchGate. URL https://www.researchgate.net/figure/NetFlow-collectorarchitecture_fig8_237812998 (accessed 06.12.22).

[9] IBM, 2019b. IBM QRadar SIEM - Overview - Finland [WWW Document]. URL https://www.ibm.com/fi-en/marketplace/ibm-qradar-siem (accessed 05.24.22).

[10] Davidoff, S., Ham, J., 2012. Network forensics: tracking hackers through cyberspace. Prentice Hall, Upper Saddle River, NJ.

[11] Douligeris, C., Serpanos, D.N., 2007. Network Security: Current Status and Future Directions.

[12] Edn, 2009. Improvement of libpcap for lossless packet capturing in Linux using PF_RING kernel patch - EDN [WWW Document]. URL https://www.edn.com/improvement-of-libpcap-for-lossless-packet-capturing-inlinux-using-pf_ring-kernel-patch/ (accessed 05.12.22).

[13] Endace, n.d. NetFlow Versus Full Packet Capture: understand the difference - Endace [WWW Document]. URL https://www.endace.com/articles/NetFlow-v-fullpacket-capture (accessed 05.21.22).

[14] Gardiner, J., Cova, M., Nagaraja, S., 2014. Command & Control - Understanding, Denying and Detecting.

[15] Hetting, C., 2019. Smart home Wi-Fi devices to grow to 17 billion units by 2030. Wi-Fi NOW Events. URL https://wifinowevents.com/news-and-blog/researchsmart-home-wi-fi-devices-to-grow-to-17-billion-units-by-2030/ (accessed 05.23.22).
[16] Hintze, A., 2016. Understanding the Four Types of Artificial Intelligence [WWW Document]. URL https://www.govtech.com/computing/Understanding-the-FourTypes-of-Artificial-Intelligence.html (accessed 05.30.22).
[17] Hutchins, E.M., Cloppert, M.J., Amin, R.M., n.d. Intelligence-Driven Computer Network Defense Informed by Analysis of Adversary Campaigns and Intrusion Kill Chains 14.
[18] URL https://www.ibm.com/security/artificial-intelligence (accessed 06.4.22).

Conflict of Interest Statement: *The author declares that there is no conflict of interest regarding the publication of this paper.*