IPFIX Processor for Apache NiFi

Mahesh Addanki, Michael Spainhour, Nakiyah Wright
George Mason University - Cyber Security Engineering Department

Introduction

When analyzing network traffic, it's typically not as important to look at the contents of the packet; rather the information about them, where they are going and how it got there. This metadata, or as explained as data about data, can reveal interesting information about your network such as: policy abuses, security incidents and possibly uncover misconfigurations. While Internet traffic has become increasingly encrypted, the metadata associated with this traffic is becoming convoluted. It has become difficult to parse through the data to collect information that could be useful to improve the design and analysis of network systems. Our goal is to streamline raw packet information into usable data.

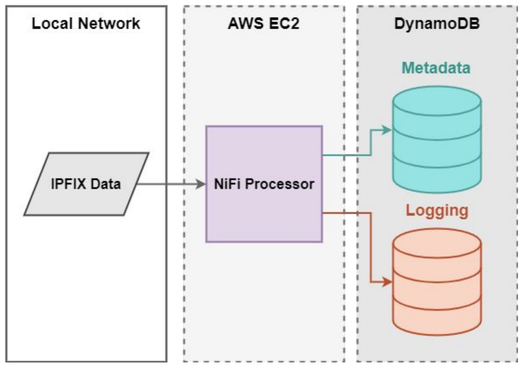
Requirements

Processor must output a 5-tuple containing
- Source IP
- Destination IP
- Source Port
- Destination Port
- Protocol Identifier

Database as storage for output (DynamoDB)
- Output as JSON, a popular human-readable data format

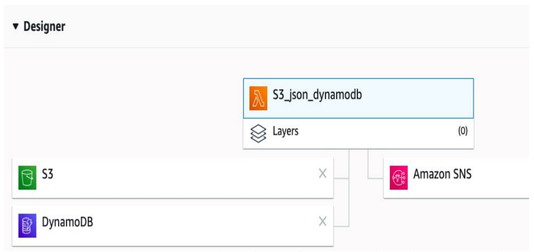
Must be scalable to meet demands for larger organizations
- Support additional protocols and data points
- Work in real-time

Conops



1. IPFIX Collector sends input data through the NiFi processor.
2. The 5-tuple data is exported as a JSON file to a database (DynamoDB) for storage. Run-time information is also stored in a log file for detailed information about how the data was processed.
3. The user can access the stored data to quickly identify traffic patterns or anomalies.

DynamoDB Conops

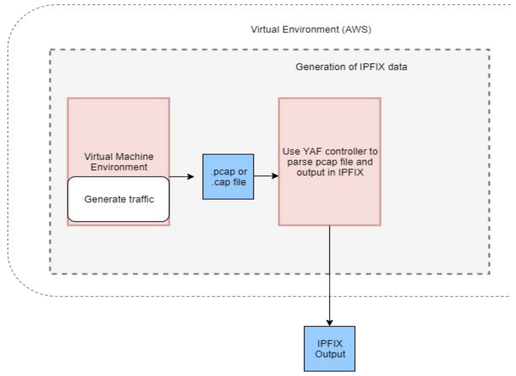


DynamoDB is comprised us four main elements in order to perform proper functionality.

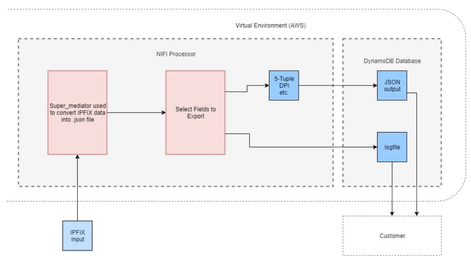
1. AWS S3 bucket is utilized to store the JSON files.
2. The DynamoDB table operates as a GUI for visual representation of the IPFIX information.
3. The Lambda function shown above automates the transfer of data from the S3 bucket to the databse.
4. 4. The SNS alaram send alerts for failed and successful data transfers.

Architecture

Add your information, graphs and images to this section.



1. Network traffic recorded in pcap or cap file format.
2. YAF converts pcap info to IPFIX. After the YAF controller converts the pcap file to IPFIX data, super_mediator will convert it to a JSON file in next steps



1. IPFIX info read by super_mediator
2. Selected fields (source/destination IP, source/destination port and protocol) exported to JSON stored in DynamoDB.
3. Log file with information on program operation (failures, errors and other metadata) stored as well.

Future Improvements

- As the amount of data in the original .pcap or .cap files increases, it is important to keep **costs** as low as possible; this is associated with offloading the database information
- Implementing a more robust and detailed **logging system,** seperate from the Flow Information Database
- Create CloudWatch Logs for **DynamoDB** (Query, Archive Log Data, Monitor)
- Create **Ansible scripts** to build infrastructure for replicated machines

Acknowledgements

Jerry Dowdy: Raytheon
Phil Harvey: Raytheon
David Vogel: Raytheon
Thomas Winston: George Mason University

References

[1] "Ansible Documentation¶," *Ansible Documentation - Ansible Documentation*, 03-Apr-2020. [Online]. Available: https://docs.ansible.com/ansible/latest/index.html. [Accessed: 13-Apr-2020].

[2] "Apache NiFi Overview," *Apache NiFi Documentation*. [Online]. Available: https://nifi.apache.org/docs.html. [Accessed: 13-Apr-2020].
[3] AWS Data Pipeline. (2020). Retrieved from https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/dp-importexport-ddb-console-start2.html

[4] B. Claise, "Specification of the IP Flow Information Export (IPFIX) Protocol for the Exchange of Flow Information," *IETF Tools*, Sep-2013. [Online]. Available: https://tools.ietf.org/html/rfc7011. [Accessed: 13-Apr-2020].

[5] "CERT YAF." CERT NetSA Security Suite, Carnegie Melon University, tools.netsa.cert.org/yaf/.

[6]Hendrix, R. W. (1983). Lambda. Retrieved from https://docs.aws.amazon.com/lambda/latest/dg/welcome.html

[7]Jenkins, G. (2000). Clouds Project CloudWatch. Retrieved from https://aws.amazon.com/cloudwatch/features/

[8]North, F. (1998). Getting started. Retrieved from https://aws.amazon.com/getting-started/hands-on/create-nosql-table/

[9] "super_mediator," *New Releases*. [Online]. Available: https://tools.netsa.cert.org/super_mediator/docs.html. [Accessed: 13-Apr-2020].