The ELK Stack: an Open Source Solution to sFlow Storage

Nalin Suri

What is sFlow?

- Flow capture protocol maintained by sflow.org
- Stands for ‘sampled flow’
  - Sampling allows for flow capture to be scalable
  - Has sample rate, where 1 in N packets is captured
- Supports Layer 2 of OSI model

What is the ELK Stack?

- Combination of 3 applications created by Elastic
  - Elasticsearch
  - Logstash
  - Kibana
- Popular, open source solution to many different types of problems
- Stack can be utilized together, or with each application separately

Elasticsearch

- Distributed search and analytics engine
- For our use case: sFlow record database
  - Stores processed records from Logstash
  - Provides a RESTful API to query, maintain, and manipulate Elasticsearch cluster and records

Logstash

- A data processing pipeline
- For our use case: sFlow packet collector
  - Takes output from sflowtool as input using wrapper script
  - Parses and processes packet information
  - Sends processed packet to elasticsearch as output

Kibana

- Web interface to search and analyze elasticsearch records
- For our use case: sFlow record visualizations
  - Displays sFlow records and information
  - Allows us to create visualizations quickly

Setting up the server

- Update to Java version 8
- Install Elasticsearch, Logstash, and Kibana
  - `wget -qO - https://artifacts.elastic.co/GPG-KEY-elasticsearch | sudo apt-key add -`
  - `echo "deb https://artifacts.elastic.co/packages/5.x/apt stable main" | sudo tee -a /etc/apt/sources.list.d/elastic-5.x.list`
  - `sudo apt-get update`
  - `sudo apt-get install elasticsearch logstash kibana`
- Install sflowtool

Experiences

- Changed Elasticsearch, Kibana settings to allow for remote access
- Started with a Logstash config file from https://whiskeyalpharomeo.com/2015/06/13/logstash-and-sflow/
  - GeoIP lookup could not keep up with amount of sflow packets received
  - Removed many filters, and mutations to allow for us to keep up with the approximately 35 million records received per day totaling less than 10 GB/day
- Added ASN lookup into Logstash configuration to prevent from updating documents and increasing required storage space
  - Routers provide ASN, but switches do not
  - Update Logstash or plugin to support ASN lookup
  - https://discuss.elastic.co/t/create-a-field-based-on-ip-range/94139/2
- Created Elasticsearch index mapping to set each field to proper type
- Currently working on creating visualizations for reports in Kibana

Security

- Currently accessible through internet: http://kibana.ampath.net:5601/
  - Data can be accessed by anyone
  - Data can be manipulated by anyone
- Solutions:
  - X-Pack Plugin: After 30 day trial, licensing fees
  - X-Pack Open Source Alternative Plugins:
    - Search Guard
  - Use proxy server to only allow access based on IP addresses/authentication

Current Setup Files

- Logstash setup file: in /etc/logstash/logstash.yml
  https://github.com/NSuri1/Ciara_SDN_Automation/blob/master/Milestone3/logstash.yml
- Elasticsearch setup file: in /etc/elasticsearch/elasticsearch.yml
  https://github.com/NSuri1/Ciara_SDN_Automation/blob/master/Milestone3/elasticsearch.yml
- Kibana setup file: in /etc/kibana/kibana.yml
  https://github.com/NSuri1/Ciara_SDN_Automation/blob/master/Milestone3/kibana.yml
- Logstash config file: in /etc/logstash/conf.d/sflow-input.conf
  https://github.com/NSuri1/Ciara_SDN_Automation/blob/master/Milestone3/sflow-input.conf
- Elasticsearch index template: in /etc/logstash/sflow-elasticsearch-mapping.json
  https://github.com/NSuri1/Ciara_SDN_Automation/blob/master/Milestone3/sflow-elasticsearch-mapping.json
- Sflowtool wrapper script: in /usr/local/bin/sflowtool_wrapper.sh
  https://github.com/NSuri1/Ciara_SDN_Automation/blob/master/Milestone3/sflowtool_wrapper.sh

Server Specs: 4 core CPU, 8 GB RAM
Currently access Kibana through http://kibana.ampath.net:5601/
Access Elasticsearch from server through ‘curl -XGET “http://localhost:9200/”’