IST-652

Final Project: Security analytics and recommendations based on machine-learning in distributed-systems

(Proposal)

Professor: Debbie Landowski
Student: Sharat Sripada (vssripad)

Introduction

Securing enterprise/cloud data can be realized via what is commonly known as network firewall policies or rules. In the traditional world, perimeter firewalls can achieve this but with the advent of software defined networks (SDN) an administrator typically calls some APIs on an SDN controller which then pushes the intent down to a distributed data-plane where it is enforced on workloads like virtual-machines, PODs or containers.

Staying on top of constantly varying applications, traffic patterns and new attack vectors at scale can take administrators several repeated iterations to make efficient and this can be particularly daunting at large scale – commonly also known as Massively Scalable Datacenters (MSDC).

The project will attempt, through the knowledge gained in IST-652 to retrieve results of firewall actions written to log files from across several nodes in the distributed plane, analyze patterns and suggest recommendations (through machine learning algorithms) and present it via a security analytics dashboard.

Below is the high-level workflow and architecture:

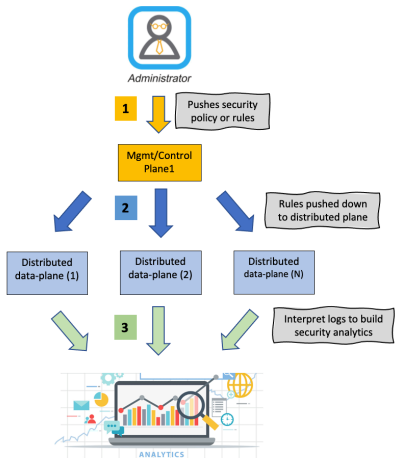

Fig-1: Workflow, process and Architecture

Data source and description

The source of the data will largely be logs written by a component enforcing network policies or rules. It captures traffic flows or tuples essentially hitting an action (allow/deny).

Here is an excerpt of the log:

```
2017-10-19T22:38:05.586Z 58734 INET match PASS domain-c8/1006 OUT 84 ICMP
172.18.8.121->172.18.8.119 RULE_TAG
2017-10-19T22:38:08.723Z 58734 INET match PASS domain-c8/1006 OUT 60 TCP
172.18.8.121/36485->172.18.8.119/22 S RULE_TAG
2017-10-19T22:38:18.785Z 58734 INET TERM domain-c8/1006 OUT ICMP 8 0
172.18.8.121->172.18.8.119 2/2 168/168 RULE_TAG
```

```
2017-10-19T22:38:20.789Z 58734 INET TERM domain-c8/1006 OUT TCP FIN
172.18.8.121/36484->172.18.8.119/22 44/33 4965/5009 RULE_TAG
```

Description

A traffic flow or tuple would comprise the following fields:

- src ip-address
- dst ip-address
- src port
- dst port
- TCP/IP Protocol
- Action – Deny/Permit

Viewing this through an example:

```
# tail -f /var/log/dfwpktlogs.log | grep 192.168.110.10

2015-03-10T03:20:31.274Z INET match DROP domain-c27/1002 IN 60 PROTO 1
192.168.110.10->172.16.10.12
2015-03-10T03:20:35.794Z INET match DROP domain-c27/1002 IN 60 PROTO 1
192.168.110.10->172.16.10.12
```

To generate a sufficiently large dataset, the idea is to simulate several traffic flows based on popular applications and therefore cause a corresponding number of actions to be logged.

Research Questions

Following are some of the questions we seek to answer:

1. What are the traffic patterns or flows across the datacenter?
   - Analyze and visualize traffic patterns – get counts of flows across TCP, UDP and ICMP protocols
2. Provide rule metrics across all network policies
3. Plot a time-series graphs:
   - For high hitting flows (allow/deny) identify spikes, saw-tooth behavior or sustained/repeated patterns etc.
4. Finally, based on traffic patterns make recommendations to add/remove/update network policies or rules. Explore machine-learning algorithms to achieve this.

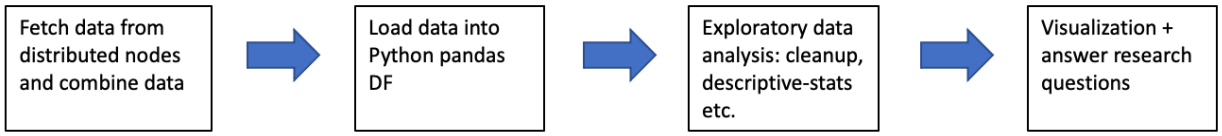

Fig-2: Workflow