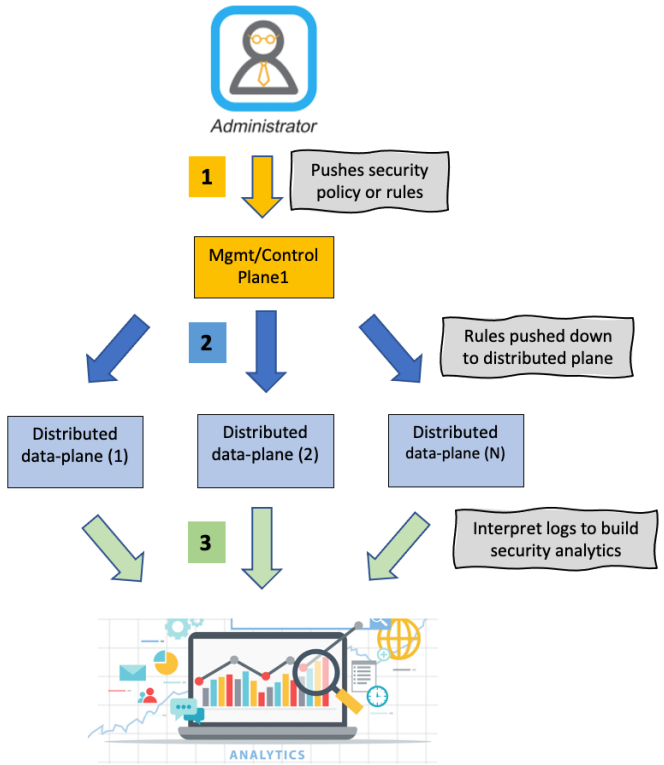
IST-652: FINAL PROJECT PRESENTATION

SECURITY ANALYTICS AND RECOMMENDATIONS IN DISTRIBUTED SYSTEMS

SHARAT SRIPADA

INTRODUCTION



De-centralized/distributed Firewall

- Securing enterprise/cloud data can be realized via what is commonly known as network firewall policies or rules
- In traditional network architecture, perimeter firewalls can achieve this but with the advent of software defined networks (SDN), firewalls are de-centralized and distributed

GOAL:

- Collect logs (un-structured data) across the distributed plane and build an analytics/recommendation system based on machine-learning

RESEARCH QUESTIONS

Workflow

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

SOURCE OF DATA

```
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
```

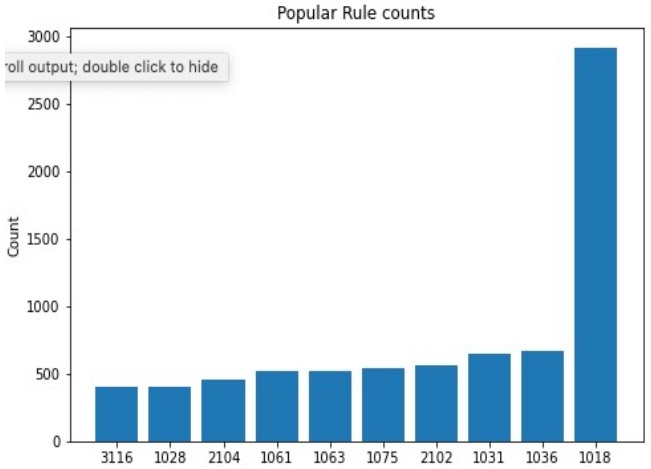
Log excerpt on distributed-plane
(raw un-structured data)

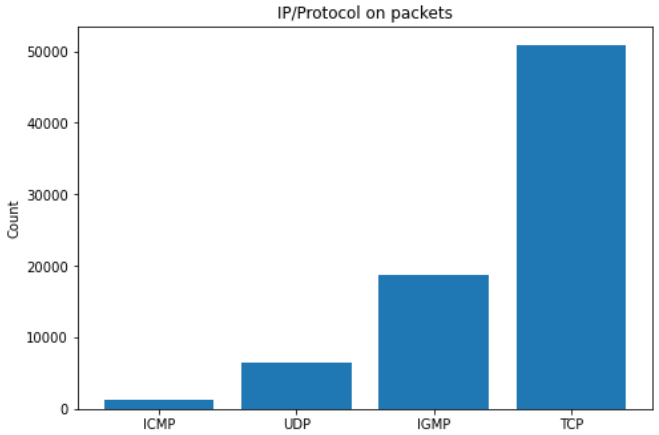
	time	reason	source	action	rule	dir	pktlen	proto	sip	dip	sport	dport
0	2020-12-05T20:51:02.220Z	match	20.20.177.78	DROP	2026	IN	36	IGMP	0.0.0.0	224.0.0.1	0	0
1	2020-12-05T20:51:02.220Z	match	20.20.177.78	DROP	2026	IN	36	IGMP	0.0.0.0	224.0.0.1	0	0
2	2020-12-05T20:51:02.220Z	match	20.20.177.78	DROP	2026	IN	76	ICMP	fe80::ffff:ffff:ffff:ffff	ff02::1	0	0
3	2020-12-05T20:51:02.220Z	match	20.20.177.78	DROP	2026	IN	36	IGMP	0.0.0.0	224.0.0.1	0	0
4	2020-12-05T20:51:02.220Z	match	20.20.177.78	DROP	2026	IN	76	ICMP	fe80::ffff:ffff:ffff:ffff	ff02::1	0	0

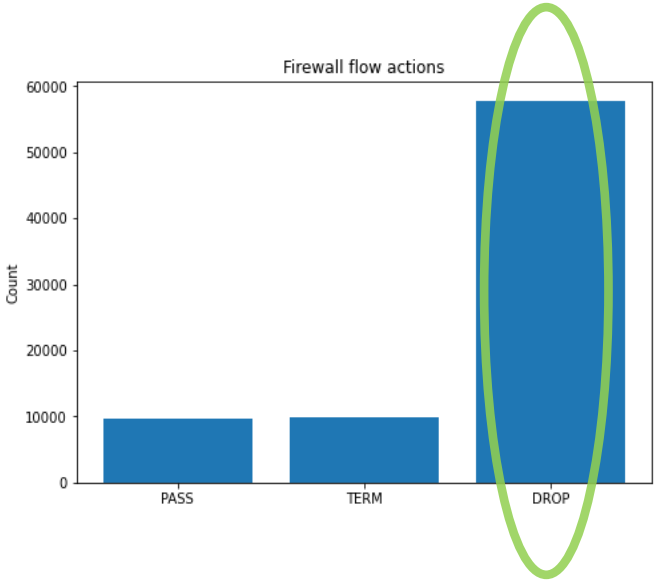
Data loaded from Mongo-DB into
data-frame

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

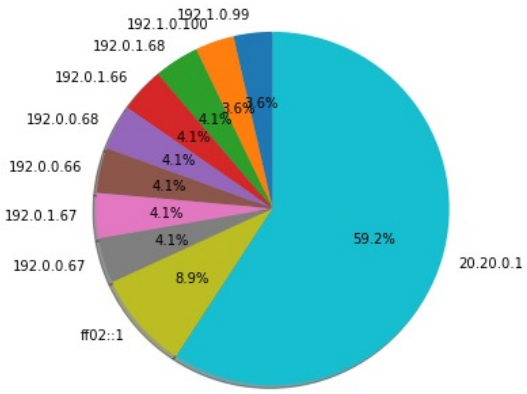
ANALYTICS-DASHBOARD



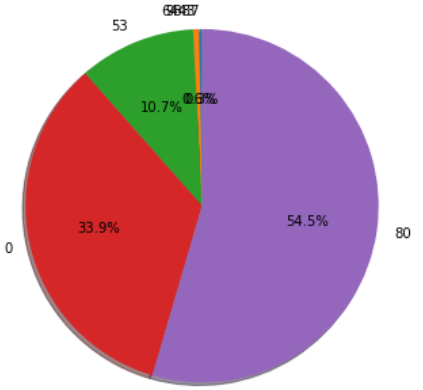




Zoom in on DROP flows



Drop flows categorized by source



Drop flows categorized by L4 ports

RECOMMENDATIONS AND CONCLUSION

```
# Step-5:
# Check the accuracy
print("Accuracy:", metrics.accuracy_score(y_test, y_pred))

# Conclusion:
# Using Decision-Trees (GINI method), given certain packet/flow attributes we
# predict with upto 78% accuracy if Firewall action is DROP, PASS or TERM
```

```
Accuracy: 0.7804326767839845
```

Fig. Result of prediction using Decision-Trees

- Decision-Tree (used sklearn libraries) to train, test and predict outcomes given a flow tuple:
  - Data-split: 80% train, 20% test
  - Result: Accuracy of 78%

CONCLUSION AND FUTURE-WORK:

- Improve accuracy of prediction with usage of Random Forests (ensemble methods) or other kernel SVM techniques
- Explore application of deep-learning methods
  - Research if database signatures of commonly known threat/attack vectors can be maintained in an MNIST-like datastore
- Build a simple utility – given a packet tuple (s.ip, d.ip, ports, protocol) predict Firewall action based on flows extracted