Security Information and Event Management tool & ML

Riddhesh Markandeya & Malasree Rallapalli
Wayne State University
Department of Computer Science
6th December 2022

Components

Elasticsearch

Kibana

Filebeat

Suricata

Suricata

- Suricata is a high performance, open-source network analysis and threat detection software used by most private and public organizations and embedded by major vendors to protect their assets.
- Suricata can generate log events, trigger alerts, and drop traffic when it detects suspicious packets
- By default, Suricata works as a passive Intrusion Detection System (IDS)
- It can also be configured as an active Intrusion Prevention System (IPS)
- By default, "ET Open Ruleset" included in suricata signature

Suricata signatures:

Action

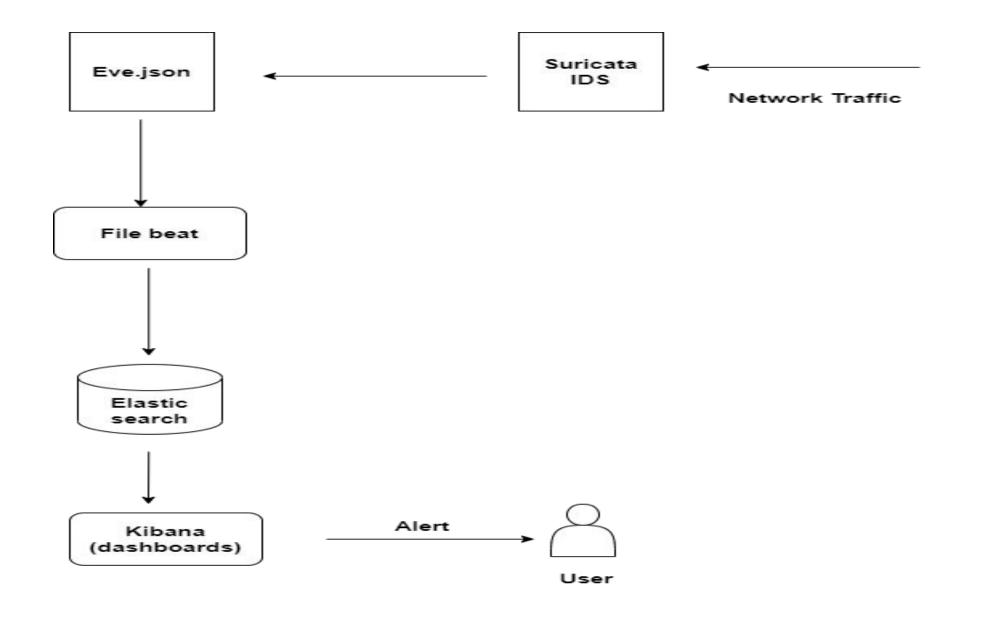
Header

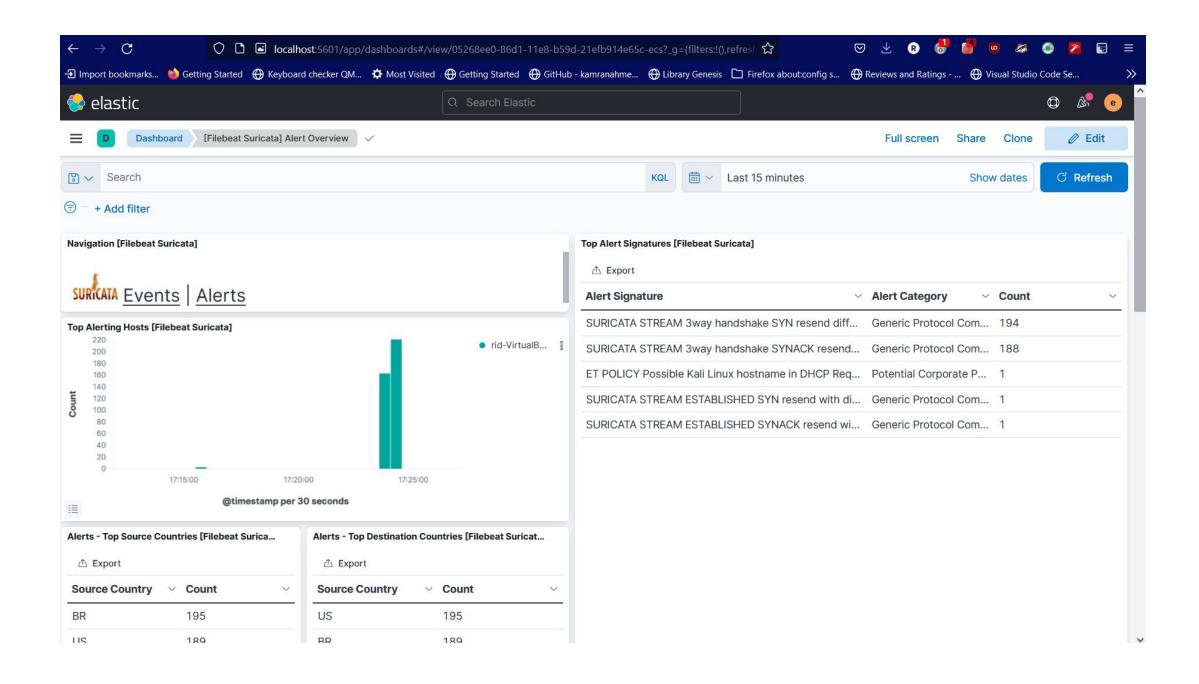
Options

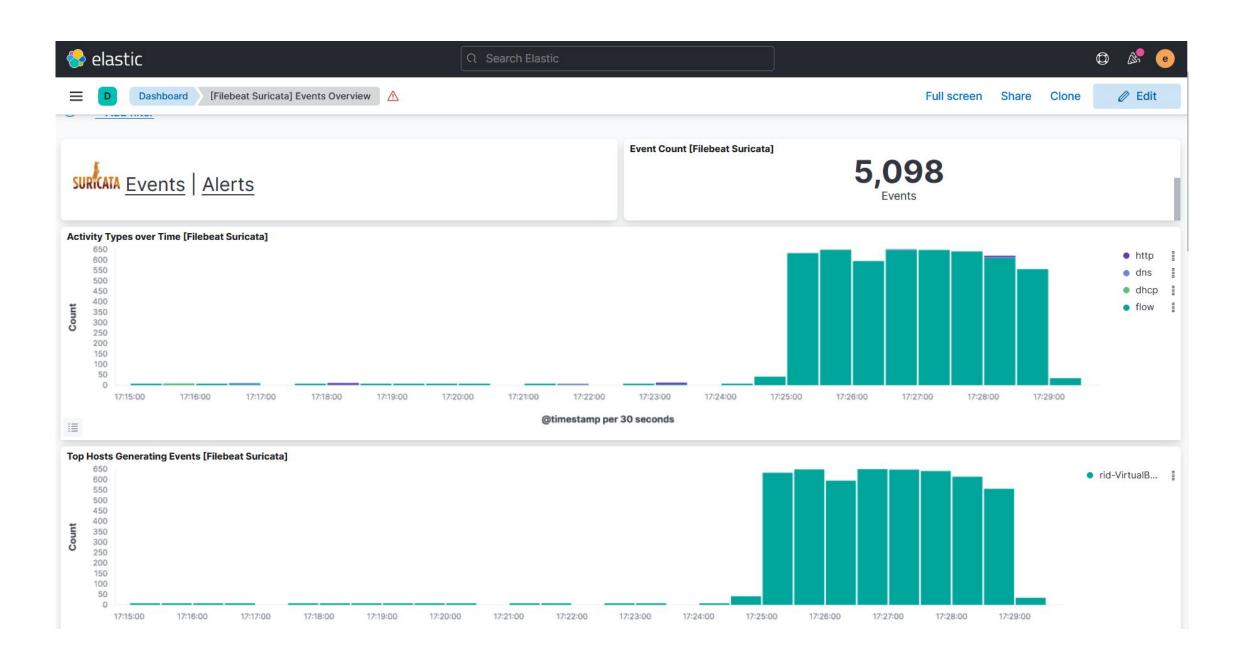
Example:

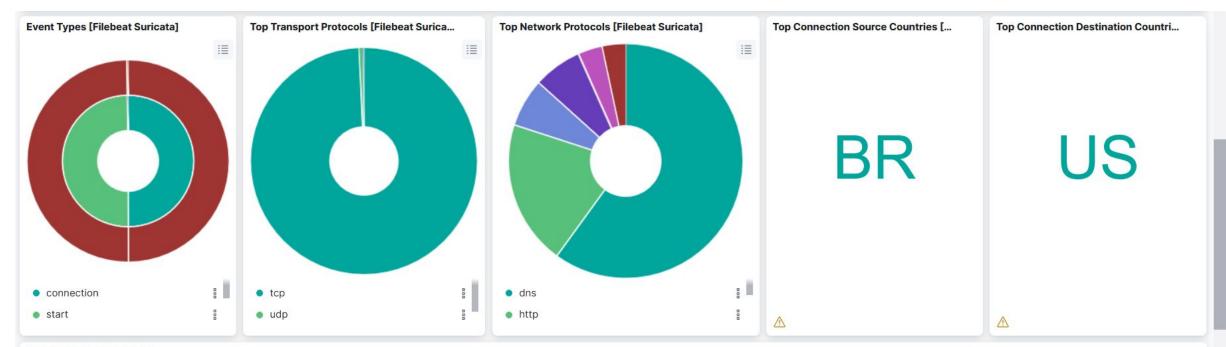
Generic Rule structure: Action Header Options

```
alert ip any any -> any any (msg:"GPL ATTACK_RESPONSE id check returned root"; content:"uid=0|28|root|29|"; classtype:bad-unknown; sid:2100498; rev:7; metadata:created_at 2010_09_23, updated_at 2010_09_23;)
```





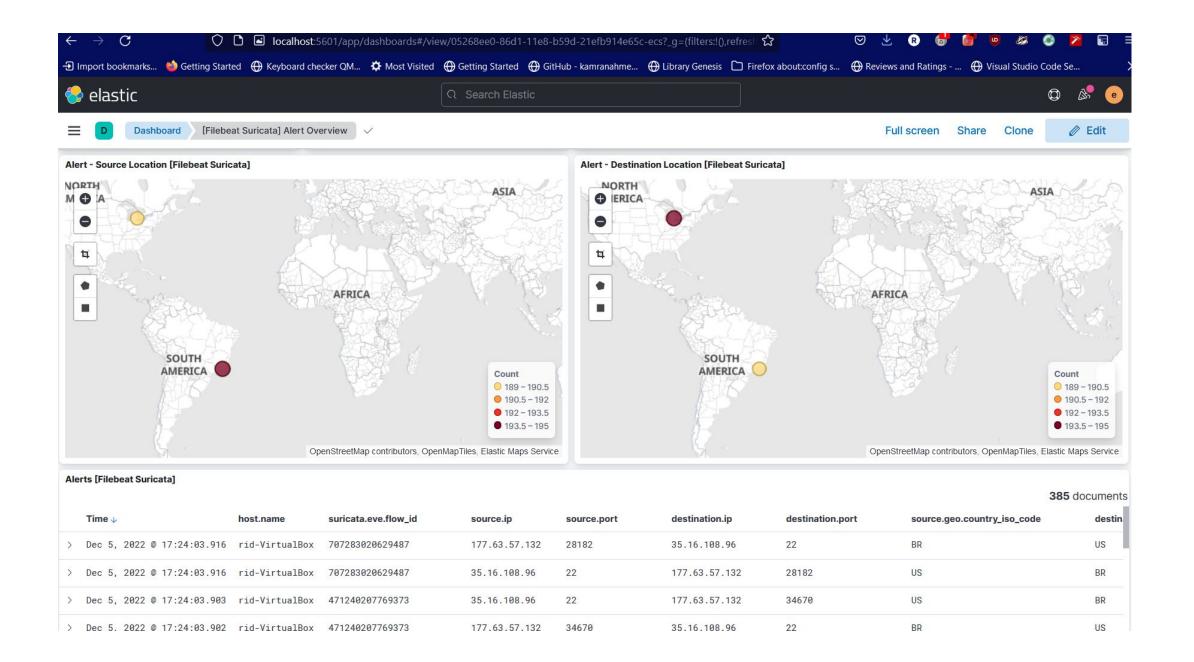




Events [Filebeat Suricata]

5096 documents

	Time ↓	host.name	suricata.eve.flow_id	network.transport	source.ip	source.port	destination.ip	destination.port	destination.geo.re
>	Dec 5, 2022 @ 17:29:10.273	rid-VirtualBox	1119245545488258	tcp	177.63.57.132	18857	35.16.108.96	22	Michigan
>	Dec 5, 2022 @ 17:29:10.273	rid-VirtualBox	134280695506333	tep	177.63.57.132	14780	35.16.108.96	22	Michigan
>	Dec 5, 2022 @ 17:29:09.592	rid-VirtualBox	978439337660877	tcp	177.63.57.132	28397	35.16.108.96	22	Michigan
>	Dec 5, 2022 @ 17:29:03.401	rid-VirtualBox	130490387634876	tcp	177.63.57.132	42432	35.16.108.96	22	Michigan



ML Classifiers on CICIDS-2017

CICIDS-2017

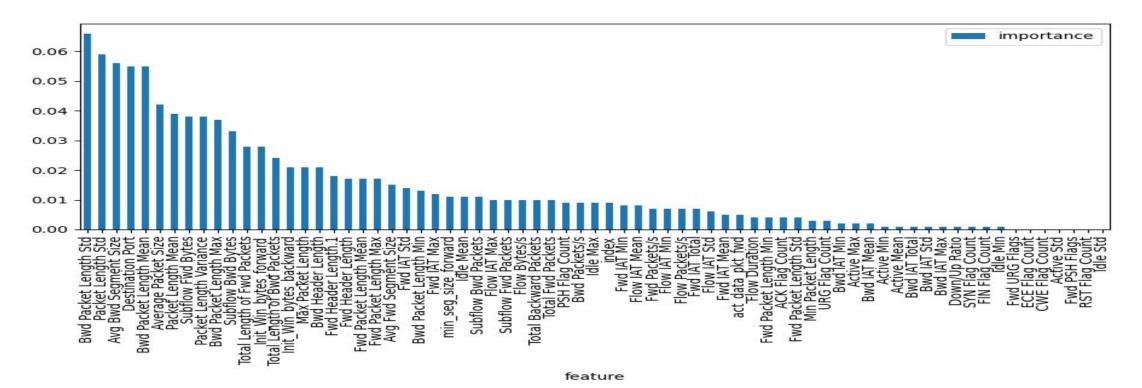
- PCAP data that resembles the true real-world traffic data spanning over 5 days.
- Multi class dataset: Benign, DoS Hulk, PortScan, DDoS, DoS GoldenEye, FTP-Patator, DoS slowloris etc.
- 15 labels: 1 Normal + 14 Attack labels
- 3,119,345 instances and 83 features
- Features: Flow Duration, Flow Bytes/s, SYN Flag Count, Fwd Header Length, etc.

CICIDS-2017

Class Labels	Number of instances
BENIGN	2359087
DoS Hulk	231072
PortScan	158930
DDoS	41835
DoS GoldenEye	10293
FTP-Patator	7938
SSH-Patator	5897
DoS slowloris	5796
DoS Slowhttptest	5499
Botnet	1966
Web Attack – Brute Force	1507
Web Attack – XSS	652
Infiltration	36
Web Attack – Sql Injection	21
Heartbleed	11

Data Preprocessing and Feature Reduction

- Remove instances with missing class labels, missing values, NaNs, Infinity, scale numerical values to have 0 mean, 1 variance.
- Remove less important features



Feature Reduction Contd.

- Least important features: Bwd PSH Flags, Bwd URG Flags, Fwd Avg Bulk Rate, Fwd Avg Bytes/Bulk etc.
- Important features: Avg Packet Size, Flow_duration, Flow Packets/sec etc.
- Feature reduction is done by using RandomForest Classifier.

Classifiers

Decision Trees(entropy, maxdepth=6)

Naive Bayes

Random Forest

Model Evaluation

	Predicted Positive	Predicted Negative
Actual Positive Class	TP	FN
Actual Negative Class	FP	TN

- Accuracy = (TP +TN)/ (TP +TN +FN +FP)
- Precision(True Positive Rate) = TP/(TP +FP)
- Recall(Sensitivity) = TP/(TP +FN)
- F1_Score = 2TP / (2TP + FP + FN)
- Data was split 70-30% for training-testing.

Model Evaluation Contd.

	Accuracy	Precision	Recall	F1_Score
Decision Tree	0.97	0.87	0.78	0.82
Naive Bayes	0.72	0.90	0.71	0.80
Random Forest	0.98	0.93	0.79	0.85

Random Forest is performing best.

Conclusion & Improvements

- Using Suricata and Elasticsearch/Kibana one can create a robust, scalable network monitoring and alerting system.
- ML classifiers can help detect anomalies in the network
- One should run both rule-based IDS and ML based IDS simultaneously and improve the ML models over the time.
- Integrating Suricata with ML model
- Reducing classes labels in CICIDS-2017 dataset. Classes with high records influences the model more. Merge classes with similar effects.

Q&A