RATING LOG EVENTS USING REPUTATION AND ANOMALY SCORES

BC. JAN ZBOŘIL

ABSTRACT: COMPUTE ANOMALY AND REPUTATION SCORES FROM NETWORK TRAFFIC LOGS. RATE LOG EVENTS WITH THE COMBINATION OF BOTH SCORES. REPORT SUSPICIOUS EVENTS. **GOAL: REDUCE AMOUNT OF LOG RECORDS**

ANOMALIES

- TRAFFIC NOT CONFORMING TO BASELINE

DETECTION
- PRINCIPAL COMPONENT ANALYSIS
- PER IP ADDRESS, FOR EACH LOG RECORD
- TRAIN/TEST SPLIT BY TIME

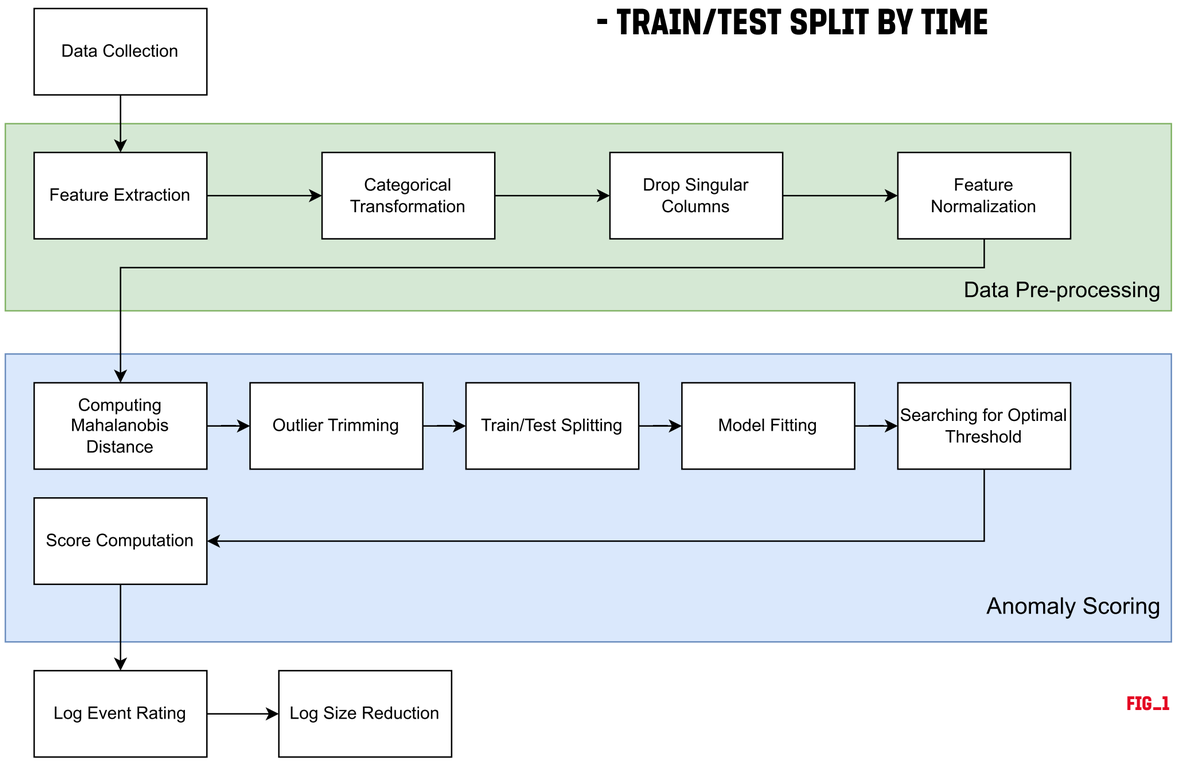


FIG_1

Algorithm PCA Anomaly Detection - Time based method

Require: Preprocessed numeric normalized dataset D, PCA number of components nc, Set of outliers O in D

$split \leftarrow \frac{2}{3}$ of all days in dataset
$train \leftarrow \{x : x \in D, x.date < split, x \notin O\}$
$test \leftarrow \{x : x \in dataset, x.date \geq split\}$
$pca \leftarrow PCA(nc).fit(train)$ ▷ Fits PCA model to training data
for each $x \in D$ **do** ▷ $x = (x_1, x_2, \ldots, x_p)$; p = number of dimensions in D
  $x' \leftarrow pca.transform(x)$ ▷ Apply dimesion reduction for sample x
  $x'' \leftarrow pca.reverseTransform(x')$ ▷ Approximate original x
  $l_x = \sum_{i=1}^{p}(x_i - x''_i)^2$ ▷ Sum accross all dimensions
end for
$\boldsymbol{l_n} \leftarrow \frac{\boldsymbol{l} - min(\boldsymbol{l})}{max(\boldsymbol{l}) - min(\boldsymbol{l})}$ ▷ Loss min-max normalization
$t \leftarrow max(\boldsymbol{l_n}) \cdot \frac{1}{2}$ ▷ Threshold is one half of max value in $\boldsymbol{l_n}$
$bt, bm \leftarrow IterThreshold(\boldsymbol{l_n}, t, O, D)$ ▷ Get best threshold, confusion matrix
return bt, bm

ALG_1

REPUTATION

- MEASURE OF BAD BEHAVIOUR
- PER IP ADDRESS, DAY

DNS, HTTP
ANOMALY, ALERT
-> DAILY SUBSCORES

DAILY SUBSCORES -> DAILY SCORE

$$S_{daily} = 1 - \frac{1}{1.05^{W_{AN}S_{AN}+W_{AL}S_{AL}+W_{DNS}S_{DNS}+W_{HTTP}S_{HTTP}}}$$

EQ_1

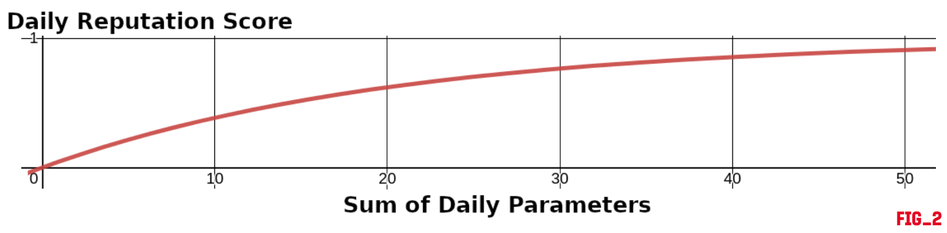


FIG_2

DAILY SCORE -> OVERALL SCORE

$$S_f = \frac{\sum_{i=1}^{n}(i)S_i}{\sum_{i=1}^{n} i}$$

EQ_2

DATASET

- EVE JSON LOGS OF THE SURICATA IDS
- UNLABELLED
- ENHANCED NETFLOW RECORDS
- 3 MONTHS OF DATA (11-12 2022, 01 2023)


```
{
  "timestamp": "2022-10-26T17:10:22.821458+0200",
  "flow_id": 229369739830515,
  "event_type": "http",
  "src_ip": "X.X.X.X",
  "src_port": 49528,
  "dest_ip": "Y.Y.Y.Y",
  "dest_port": 80,
  "proto": "TCP",
  "http": {
    "hostname": "registry.npmjs.org",
    "http_user_agent": "JetBrains IDE",
    "http_method": "GET",
    "protocol": "HTTP/1.1",
    "status": 301,
  }
}
```

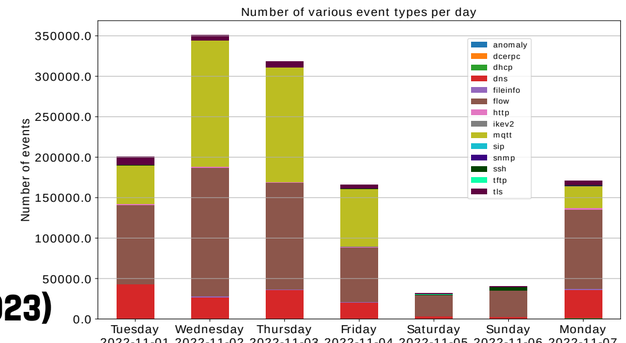
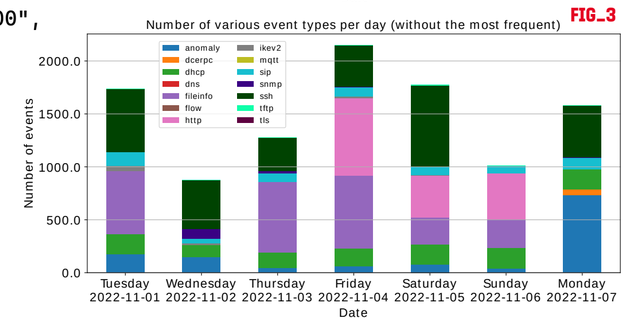



FIG_3



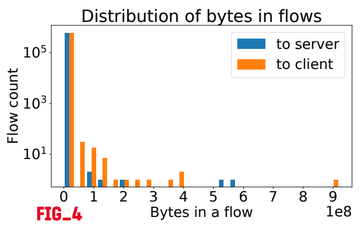


FIG_4

```
http,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,
registry.npmjs.org,JetBrains IDE,
text/plain,GET,301,0,-1,-1,-1,-1,-1,-1,X.X.X.X
```

FIG_5

RESULTS

- ANOMALY SCORE
-> 0.62 % POSITIVES
- BOTH SCORES
-> 0.1 % POSITIVES

NO CORRELATION BETWEEN SCORES

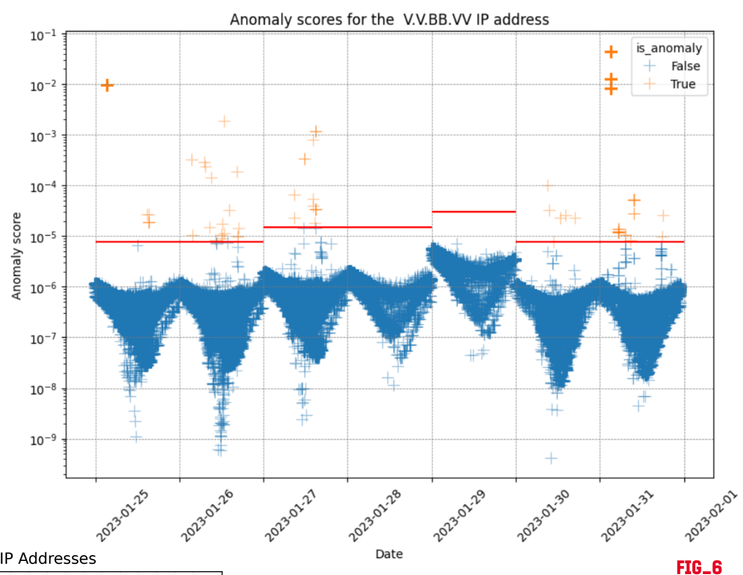


FIG_6

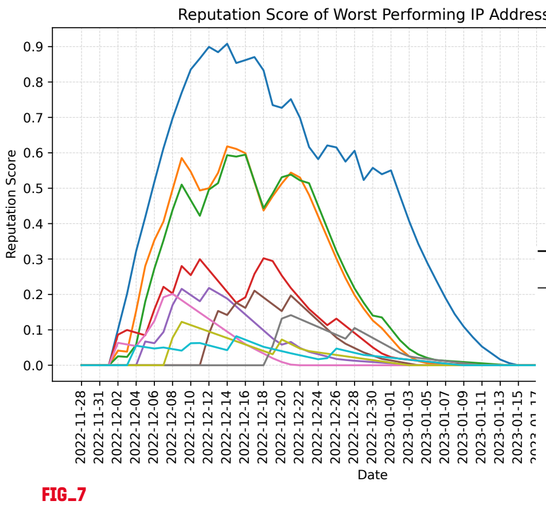


FIG_7

TAB_1

	Count	% of All	% of Anomalous
All records	727 602	100.00	-
Non Anomaly	723 060	99.38	-
Anomaly	4 542	0.62	-
Reputation = 0	222 328	30.56	-
Reputation > 0	505 274	69.44	-
Reputation >= 0.1	503 925	69.26	-
Reputation >= 0.2	386 337	53.10	-
Reputation >= 0.3	100 714	13.84	-
Anomaly & Reputation = 0	2 616	0.36	57.60
Anomaly & Reputation > 0	1 926	0.26	42.40
Anomaly & Reputation >= 0.1	1 904	0.26	41.92
Anomaly & Reputation >= 0.2	1 444	0.20	31.79
Anomaly & Reputation >= 0.3	743	0.10	16.36

IP	Thesis		NERD - 12. 3. 2024				Talos - 12. 3. 2024		
	Computed Score	Date	Score	Added	Last Activity	Other	Web Rep.	Block List	Email Rep.
V.V.BB.QQ	0.908057	14.12.2022	-	-	-	-	-	-	-
76.223.92.165	0.618265	14.12.2022	-	2024-03-09	-	1 list, 443	-	-	-
13.248.212.111	0.594867	16.12.2022	-	2024-02-16	-	1 list, 443	-	-	-
V.V.BB.LL	0.302143	18.12.2022	-	-	-	-	-	-	-
V.V.BB.VV	0.218269	12.12.2022	-	-	-	-	-	-	-
139.59.152.202	0.210607	17.12.2022	0.000	2024-01-01	2024-02-26	3 lists, Scan, 22, 80	-	Expired	-
141.94.110.90	0.201445	08.12.2022	0.000	2024-01-01	2024-02-26	4 lists, Scan, 22, 8069	-	Expired	-
165.232.69.156	0.141505	21.12.2022	0.000	2024-01-01	2024-02-26	4 lists, Scan, 22, 8069	-	Expired	-
178.60.204.50	0.117610	13.12.2022	-	-	-	-	-	Expired	Poor
186.122.177.117	0.105949	06.12.2022	-	-	-	-	-	Expired	-
149.202.74.37	0.096894	01.01.2023	-	2024-03-11	-	3 lists	-	Expired	-
57.128.11.39	0.072859	22.12.2022	0.582	2023-11-19	2024-03-12	8 lists, Scan, 22, 80, 111, 8081	Untrusted	Yes	-
43.138.17.151	0.063309	05.12.2022	-	-	-	-	-	-	Poor
101.43.110.129	0.063309	18.12.2022	-	2024-03-11	-	1 blacklist	-	Expired	-
192.145.127.42	0.006803	19.12.2022	0.764	2023-08-15	2024-03-12	5 lists, Scan	Untrusted	Yes	-
139.28.218.34	0.006803	27.12.2022	0.500	2023-11-02	2024-03-12	7 lists, Scan	-	Expired	Poor
91.207.175.154	0.006803	03.12.2022	0.652	2021-11-23	2024-03-12	7 lists, Scan	-	Expired	-
185.245.86.226	0.006803	22.12.2022	0.750	2021-09-09	2024-03-12	10 lists, Scan, 123	-	Expired	Poor
190.171.189.85	0.006803	14.12.2022	0.0445	2023-10-27	2024-03-12	9 lists, Scan, Login attempts, 22, 25, 53, 80, 443, self-sign., iot, eol, database, starttls	-	Expired	-

TAB_2