RIT | Rochester Institute of Technology

CLEAR-ROAD: Extraction of Temporally Co-occurring yet Rare Critical Alerts

Gordon Werner

Ph.D. Candidate, College for Computing and Information Sciences

Rochester Institute of Technology

This talk includes efforts supported by NSF Awards # 1526383 and #1742789,
and RIT GCI Seed Fund

Cyber Defense From an Analyst's Perspective

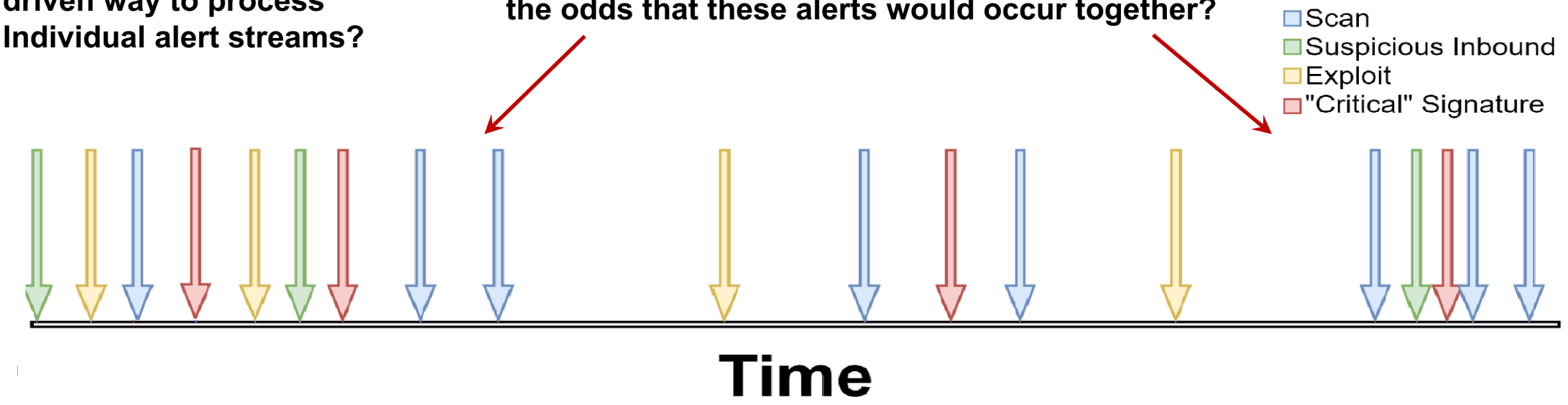
- Intrusion Detection Systems (IDSs) generate massive amounts of alerts
- Infeasible for analysts to process and find related alerts
  - Most existing approaches require training data built on expert knowledge
- Analysts are interested in specific signatures
  - Want to understand their occurrence patterns
  - What other signatures co-occur with them? In what timing profile?

Motivation

- Provide insight into alert occurrence patterns to analysts quickly
  - Using only recent network alerts
- More informative aggregation
- Data driven co-occurrence discovery

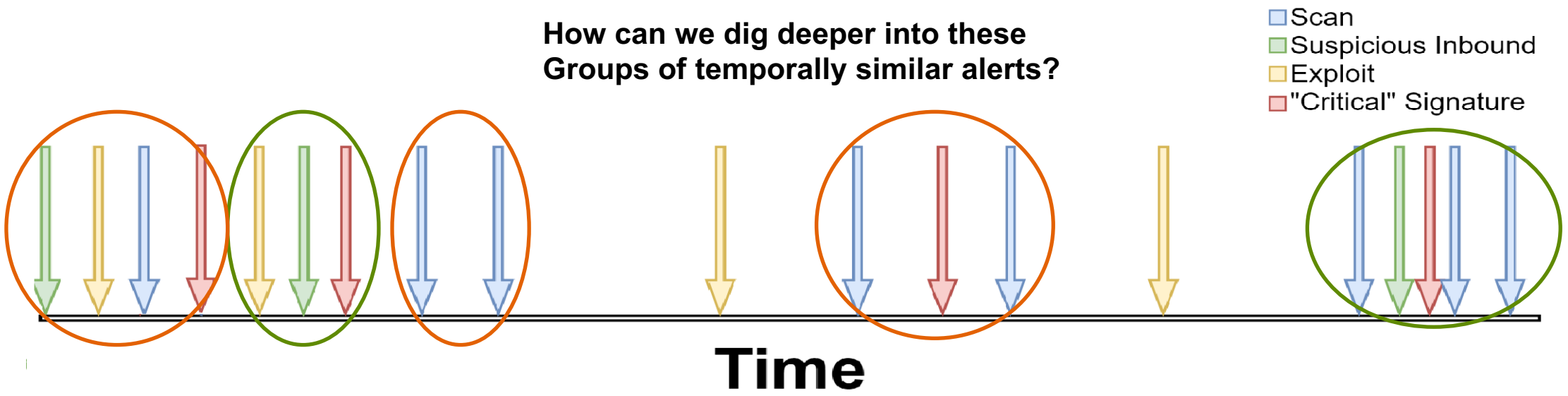
Is there an effective, data driven way to process Individual alert streams?

Are these groups of alerts the same? What are the odds that these alerts would occur together?



CLEAR Temporal Actions

- Concept Learning for Intrusion Event Aggregation in Realtime (CLEAR) [1]
- CLEAR aggregates alerts in near-real time
  - Groups successive alerts with stationary Inter Arrival Times (IATs)
- Builds and leverages temporal “concepts” for aggregation



[1] “Near Real-time Intrusion Alert Aggregation Using Concept-based Learning,” CF ‘21,

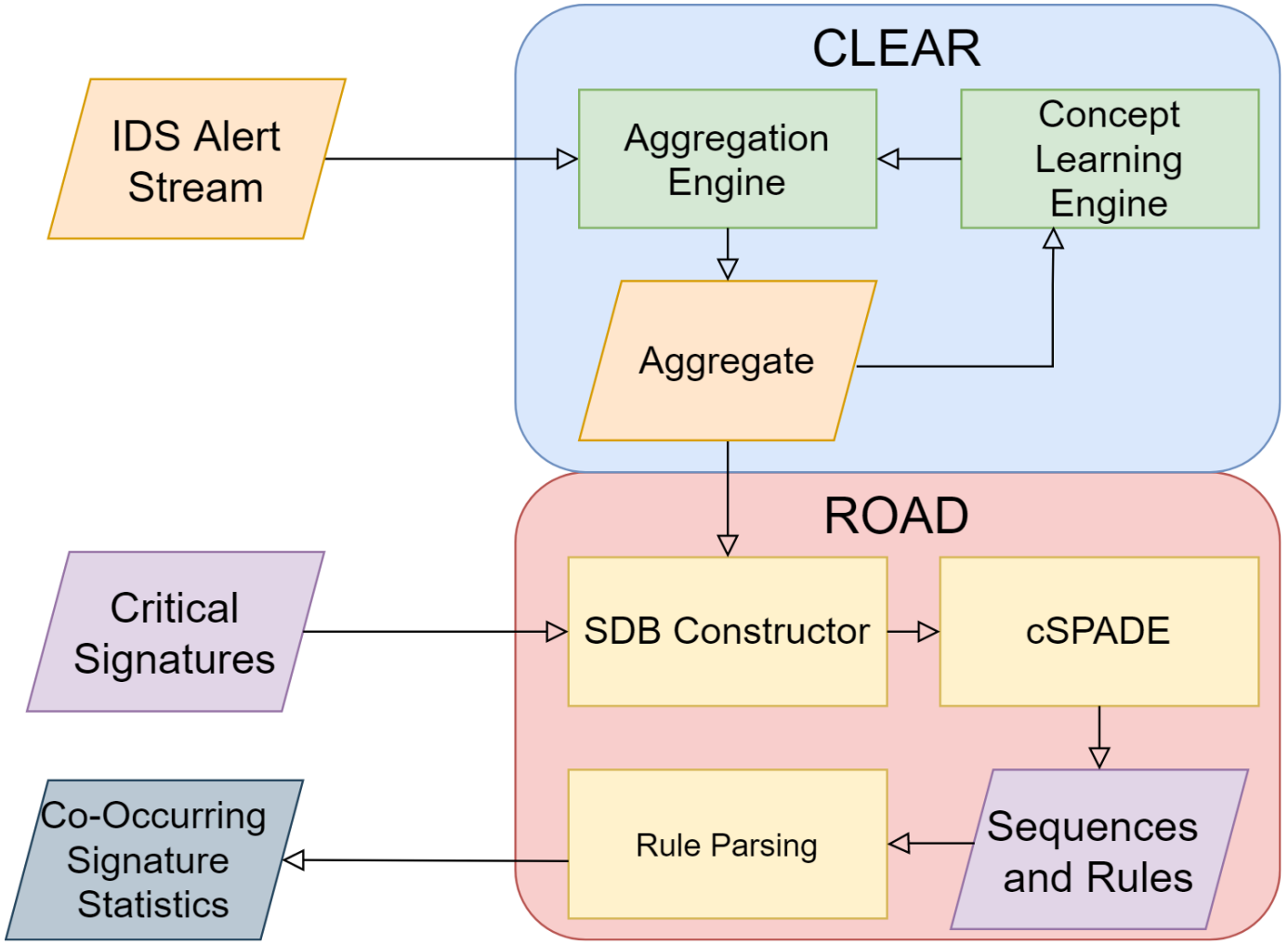
Pattern Mining Cyber Alerts

- Process database of sequences (SDB) of events to extract patterns and rules
- Techniques have been applied to cyber alerts in research [2]
  - Single adversarial IP used for individual sequences
  - Required offline processing, potentially high overheads [3]
- CLEAR aggregates are ideal candidates for sequencing
  - Each captures a stationary temporal “action”
- Focusing mining on specific and rare “critical” alerts can drastically reduce overhead
  - Constrained SDB

[2] “An incremental frequent structure mining framework for real-time alert correlation,” Computers and Security, vol. 28
[3] “On the sequential pattern and rule mining in the analysis of cyber security alerts,” ICARS ‘17

Concept Learning for Intrusion Event Aggregation in Realtime with Rare Co-Occurring Alert Signature Discovery (CLEAR-ROAD)

- Two-Step Approach:
  - CLEAR Learns & maintains unique, invariant temporal arrival patterns as "Concepts"
    - In NRT from incoming alerts
    - Captures individual ``actions" or "Aggregates" using concept statistics
  - ROAD Extracts "Co-Occurring Signatures" from Aggregates
    - Leverages pattern mining techniques
    - Sequence alert signatures using aggregates
    - Constrains sequence data base (SDB) to reduce computational overhead
    - Finds rules that exhibit statistical likelihood for signature co-occurrence



Finding Co-Occurring Signatures

- Frequent sequences are extracted from SDB based on support threshold

$$Sup(A) = \frac{A}{N}$$

- Association rules can be mined from frequent sequences

$$Conf(A \rightarrow B) = \frac{Sup(A \rightarrow B)}{Sup(A)}$$

- Lift measures the probability of occurrence between rule parts
  - Lift > 1 indicates statistical correlation in rule occurrence

$$Lift = \frac{Sup(A \rightarrow B)}{Sup(A) \cdot Sup(B)}$$

Constrained SDB Construction

- Critical signatures account for very small amount of alerts
- Processing all sequences would induce unnecessary overhead
  - Sequences with critical signature would never be “frequent”
- Limiting SDB to only include sequences with critical signatures drastically reduce size
  - Improves performance by 90+% in majority of cases

Experimental Datasets

- 2018 National Collegiate Penetration Testing Competition (CPTC)
  - Eight teams provided with identical but independent networks
- Real World SOC Operation
  - 1 Week of Operation (August 1-8 2020)
- Suricata used to generate alerts
  - Signatures mapped to various attack stages
- Five attack stages defined as critical
  - Based on discussions with real world SOC analysts
  - Arbitrary code execution, brute force creds, command & control, data exfiltration and privilege escalation

Quantitative Summary

- 62.8% of critical signatures found 1+ co-occurring signature (co-sig)
  - Most had 1+ co-sig that regularly was found co-occurring
- Many Command & Control signatures saw the same co-occurring signature in both datasets

TABLE I
SUMMARY RESULTS FOR CPTC CRITICAL SIGNATURES

Atk. Stage	Tot. Crit. Sig.	w/co-sig	w/reg. co-sig	In RSOC	Same Co-Sig
ARB. CODE EXE	55	35	34	8	1
BRUTE FORCE CREDS	4	3	2	3	1
COM. & CON.	19	9	9	10	8
DATA EXFIL	26	19	16	6	1
PRIV ESC	9	5	4	3	2

Summary Results for Select Signatures

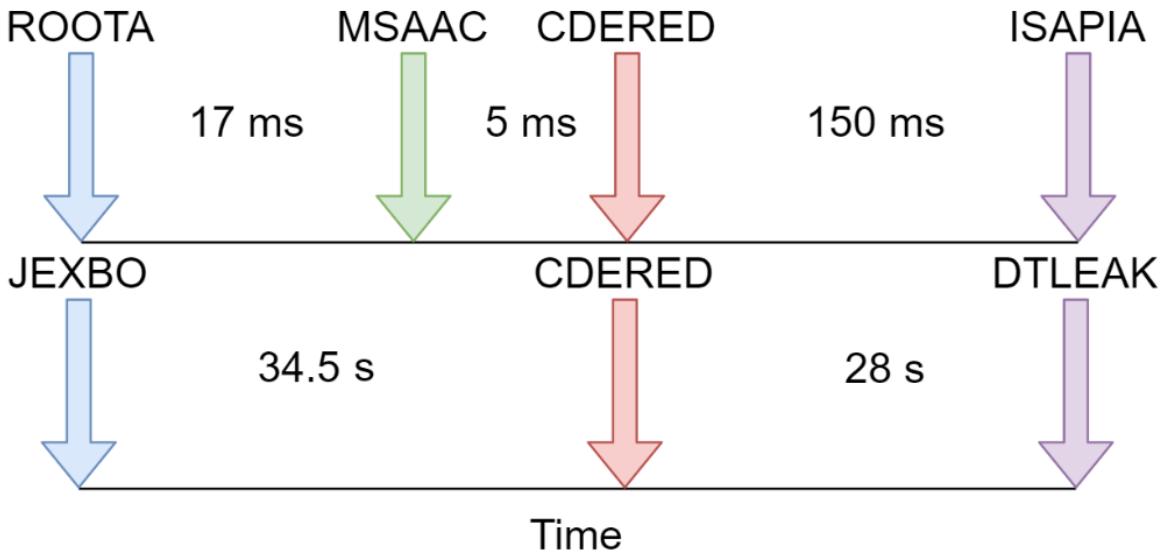
- Critical signatures are very rare in both datasets
- CPTC aggregates saw higher number of unique signatures
- RSOC timing is much higher, live network v. closed environment of CPTC

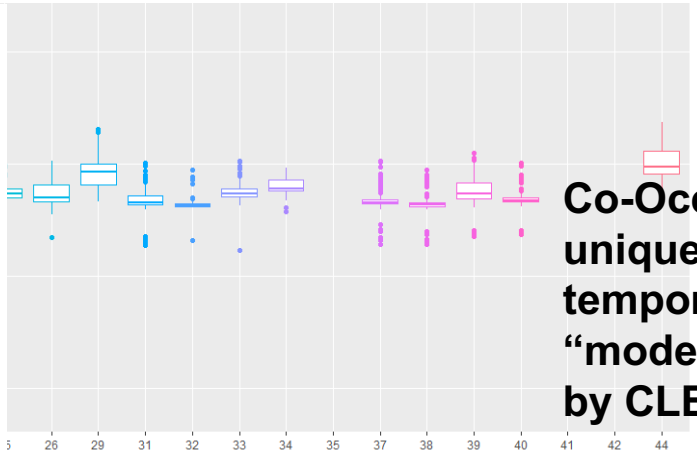
Cri.Sig.Abr.	Dataset	Rarity(%)	Agg.Sigs.	μ Lift	μ IAT
CFADMN	CPTC	1.39	43	4.67	6.1 ms
CFADMN	RSOC	1.39	19	1.45	1.2 s
CFAPIA	CPTC	0.16	27	4.52	1.5 ms
CFAPIA	RSOC	0.07	14	1.75	170 ms
CFUTIL	CPTC	0.16	27	4.52	1.5 ms
CFUTIL	RSOC	0.04	11	1.75	322 ms
DRUPAL	CPTC	0.13	20	5.78	10.8 ms
DRUPAL	RSOC	0.02	15	1.33	1.6 s
CDERED	CPTC	0.39	15	7.8	15.8 s
SMPURI	CPTC	0.04	9	196	2.6 ms
SSPURI	CPTC	0.04	9	196	2.6 ms
DIFURI	CPTC	0.04	9	196	2.6 ms
OBDURI	CPTC	0.04	9	196	2.6 ms

Case Study 1 - CodeRed

- CodeRed was found in multiple temporal patterns by CLEAR
- Each temporal action corresponded to unique co-occurring signatures

Cri.Sig.Abr.	Co.Sig.Abr.	Apps.	Dir	μ Lift	μ IAT
CDERED	ISAPIA	28	cri→co	18.9	150 ms
CDERED	ROOTA	22	co→cri	24.5	17 ms
CDERED	MSAAC	4	co→cri	21	5 ms
CDERED	JEXBO	6	co→cri	3	34.5 s
CDERED	DTLEAK	4	cri→co	6	28 s



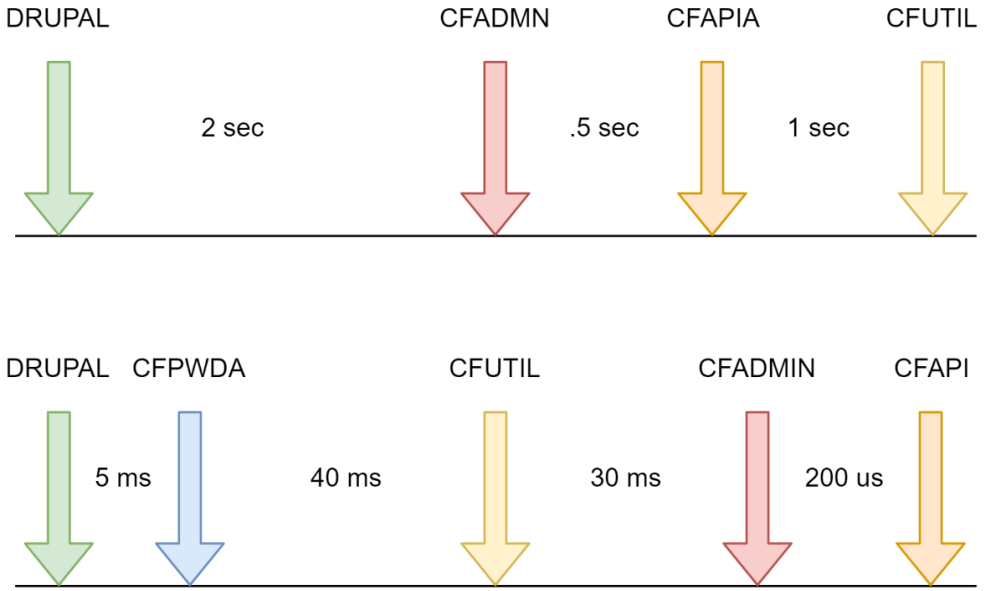


Co-Occurrence is unique to the 2 temporal "modes" found by CLEAR

Case Study 2 – ColdFusion

- Found in both datasets with the same co-occurring signatures
- # of appearances highlights the discrepancy in dataset size
  - CFADMN is equally rare in both datasets (%)

Cri.Sig.	Co.Sig.	Dataset	Apps.	Dir	μ L	μ IAT
CFADMN	CFAPIA	RSOC	1630	cri→co	1.81	0.5 s
CFADMN	CFAPIA	CPTC	2	cri→co	27.6	184 us
CFADMN	CFUTIL	RSOC	903	cri→co	1.82	1.52 s
CFAPIA	CFUTIL	CPTC	1	co↔cri	13	2 8ms
CFADMN	DRUPAL	RSOC	158	co→cri	1.61	1.97 s
CFADMN	DRUPAL	CPTC	1	co→cri	3	70.6 ms
CFADMN	CFPWDA	CPTC	4	co→cri	23	71 ms
CFADMN	NMAPSC	CPTC	82	co↔cri	1.93	5.5 ms
CFADMN	PHPINA	CPTC	31	cri↔co	11.5	20.1 ms
DRUPAL	STREX	RSOC	370	co→cri	1.2	1.84 s



Conclusion

- CLEAR-ROAD can quickly process massive numbers of alerts and provide beneficial insight to analysts
- Temporal occurrence relationships across cyber alert signatures can be extracted
  - With no external training, in near-real time
  - Unique temporal patterns can reflect unique relationships
- In some cases, these relationships persist across networks
  - Timing and frequency may be affected due to network differences

Comments, Ideas, & Questions