Time Series Data Mining for Context-Aware Event Analysis

Mona Lange

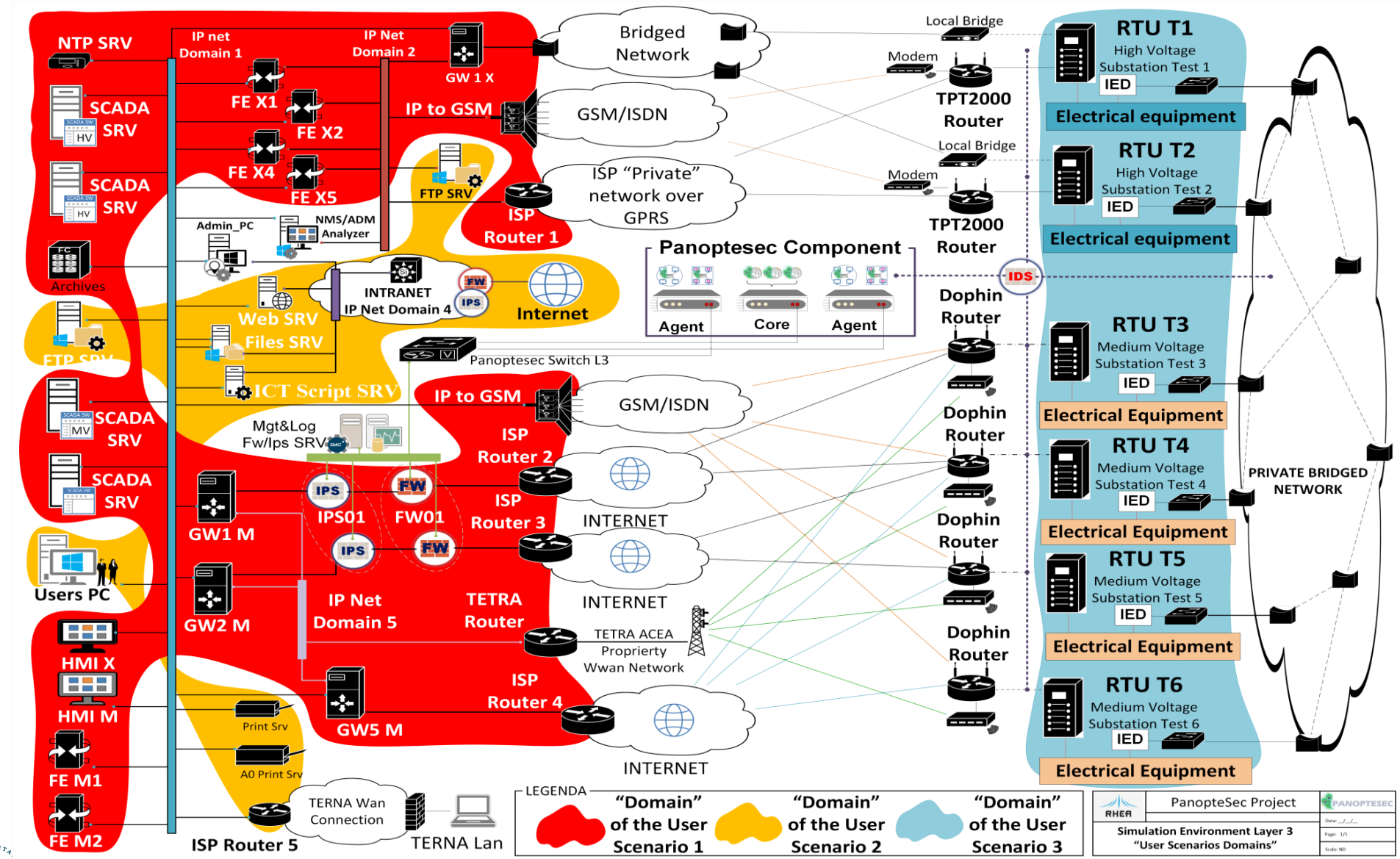
Characterization of the field of research

- IT security difficult to maintain / plethora of IDS/IPS/FW events
- Event fusion, filtering, prioritization / detecting important activities
  Mission-criticality tradeoff handled appropriately
- No human in the loop

What is the problem?
How do I address it.

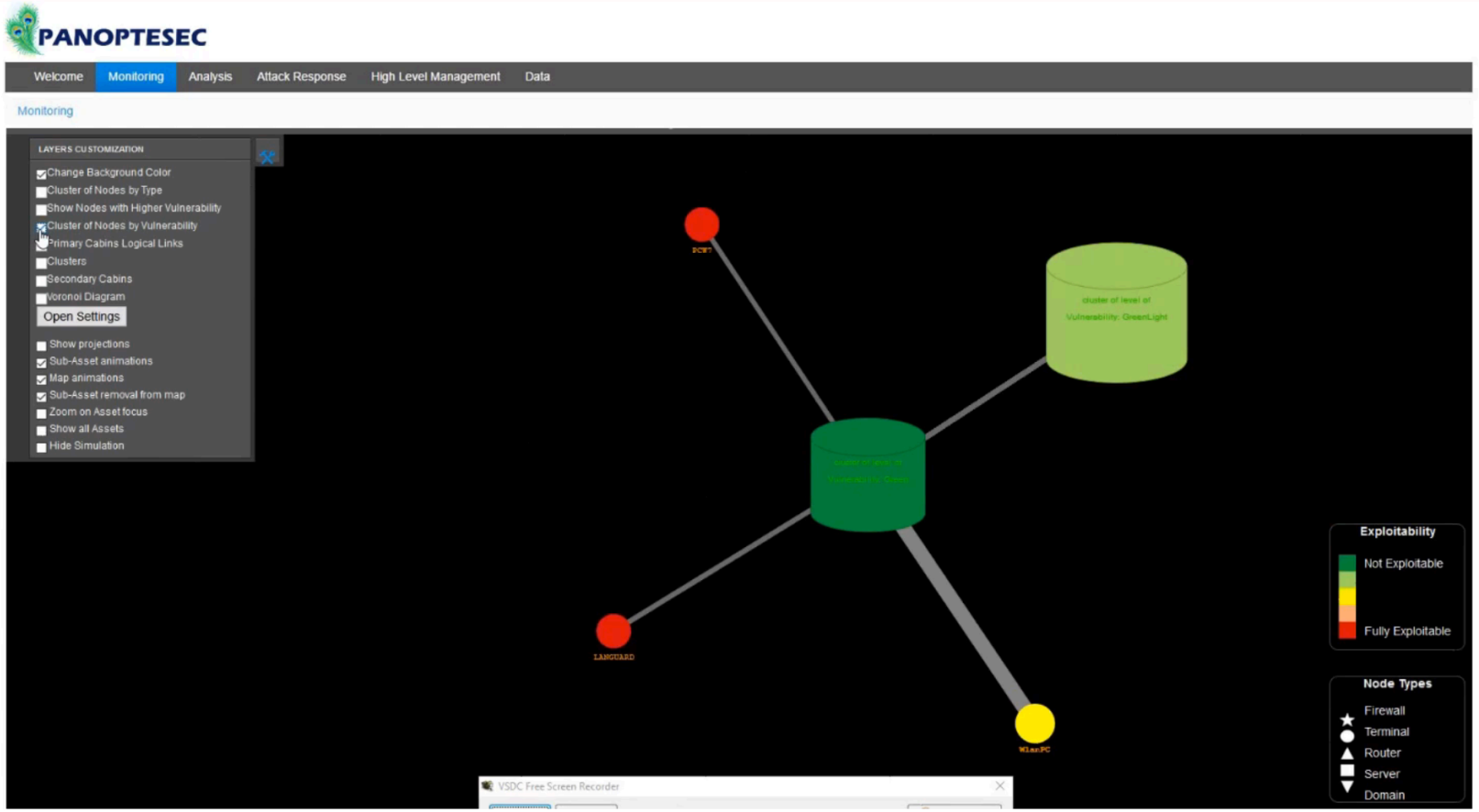
USP

Context: Critical Infrastructures – ACEA



Automatically Acquired: Vulnerabilities

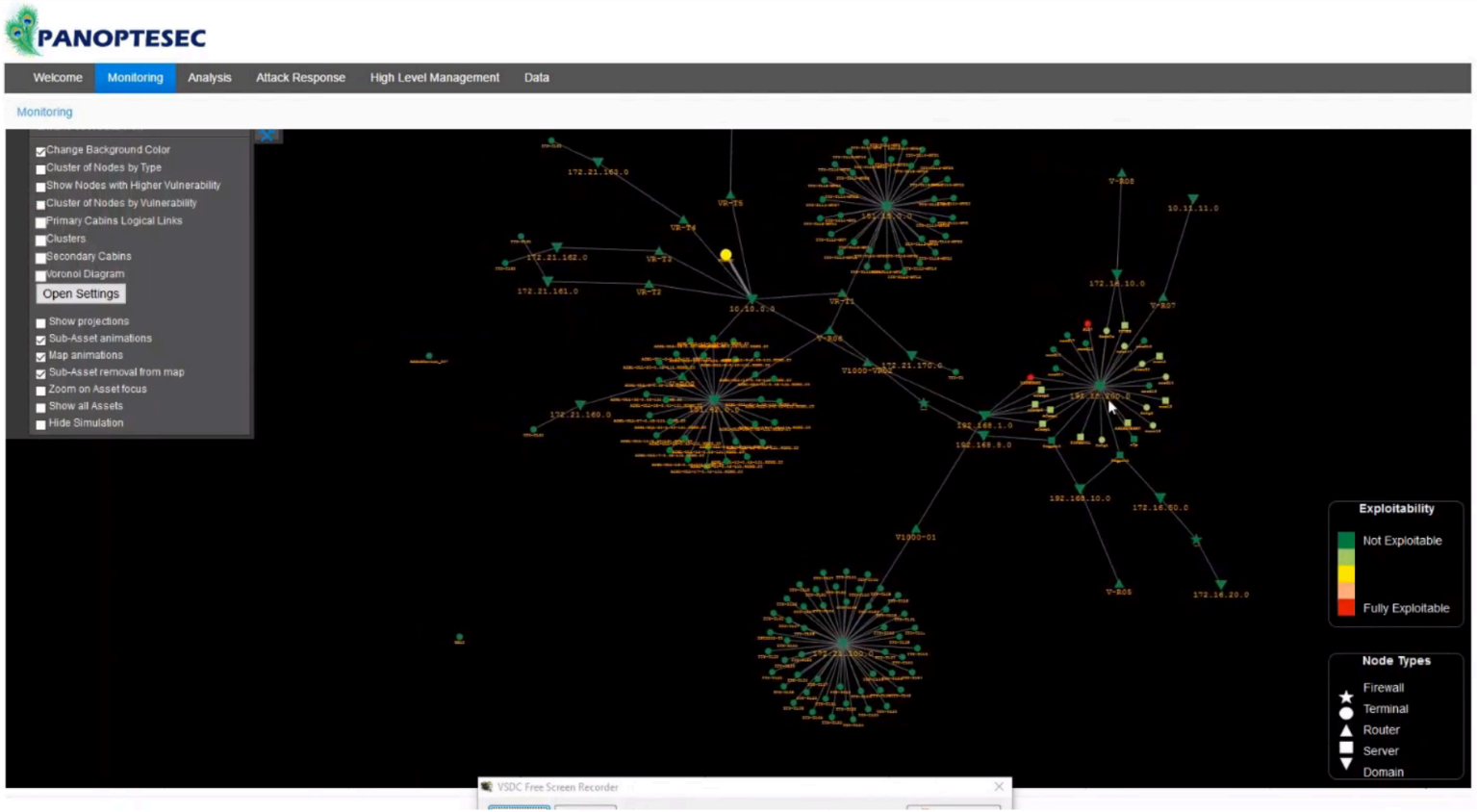
PANOPTESEC Network Topology overview



The Network Topology overview offers to the operator a view of the reachability of exploitable devices

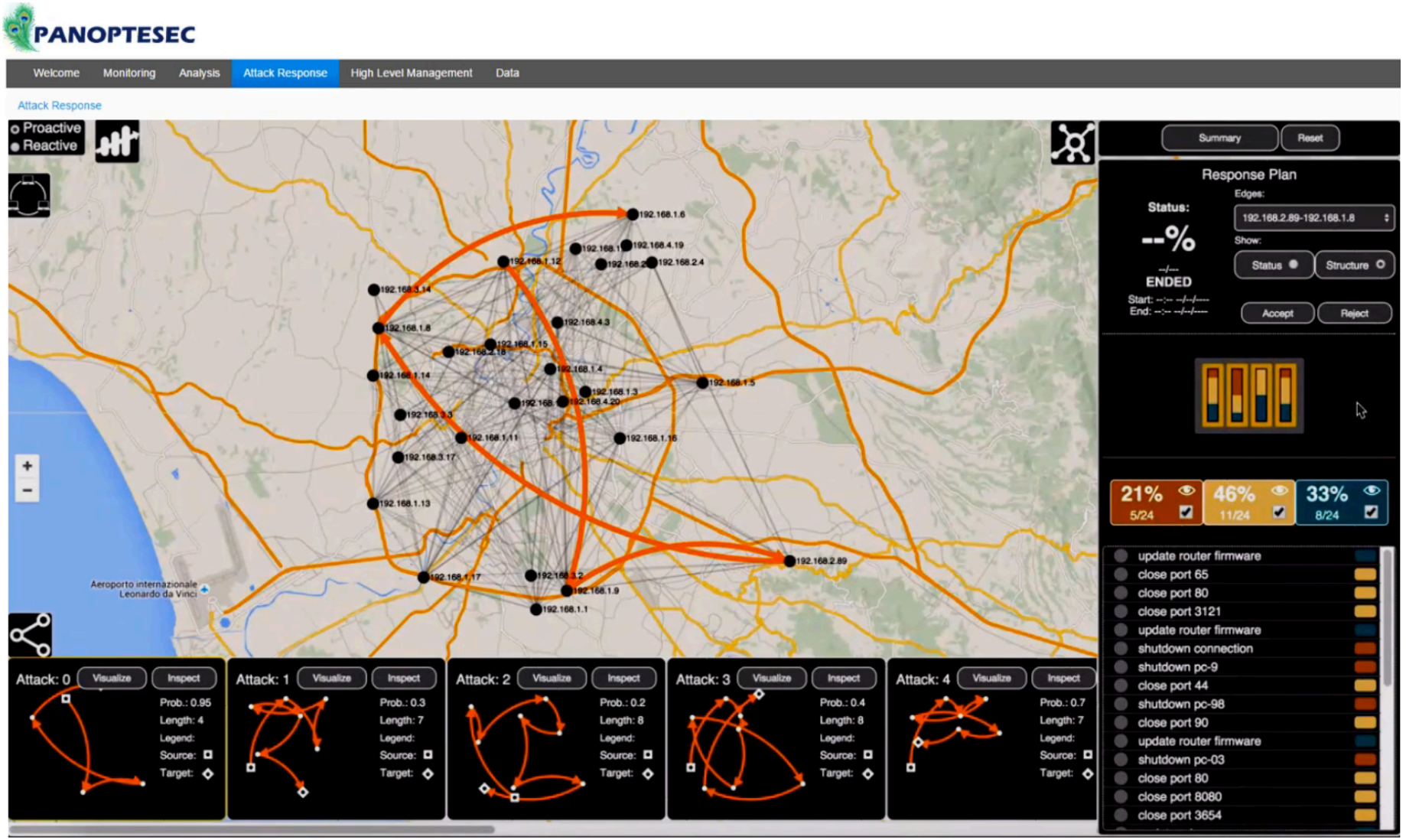
Automatically Acquired: Network Topology

PANOPTESEC Network Topology overview



The Network Topology overview offers to the operator a view of the reachability of exploitable devices. The operator can open the network view and observe vulnerabilities

Attacks: Reactive and Proactive View

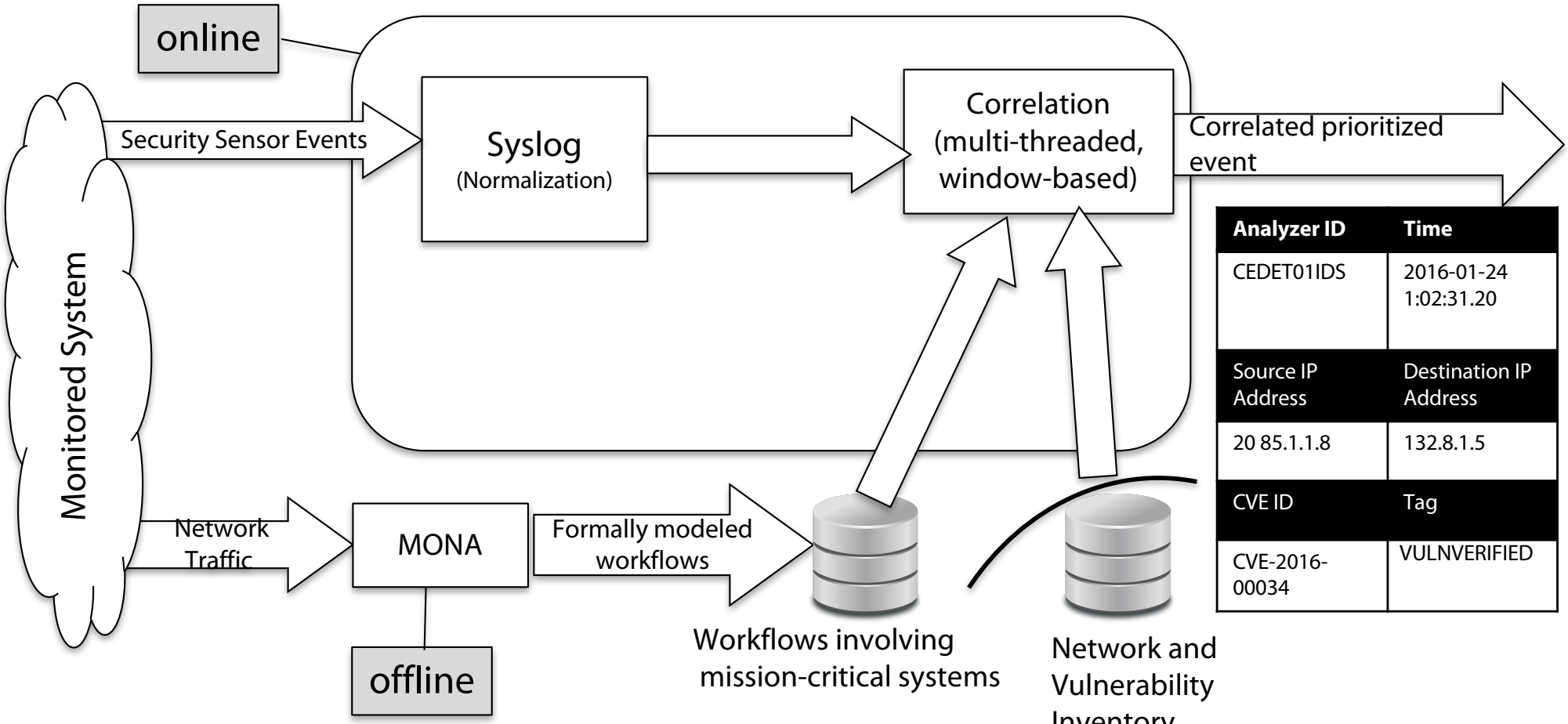


Objective of this Research

- Online: Context-Aware Event Analysis
  - Normalize heterogeneous events from multiple sources
  - Filter and fuse events
  - Prioritization by operational impact assessment based on important activities ("workflows")

- Offline: Time Series Data Mining
  - Learn to identify workflows based on mining network traffic
  - Formally represent workflows as stochastic processes
  - Mission Oriented Network Analysis (MONA)

Context-Aware Event Correlation



Analyzer ID	Time
CEDET01IDS	2016-01-24 1:02:31.20
Source IP Address	**Destination IP Address**
20 85.1.1.8	132.8.1.5
CVE ID	**Tag**
CVE-2016-00034	VULNVERIFIED

IP	Vulnerability identifier
132.8.1.5	CVE-2016-0034

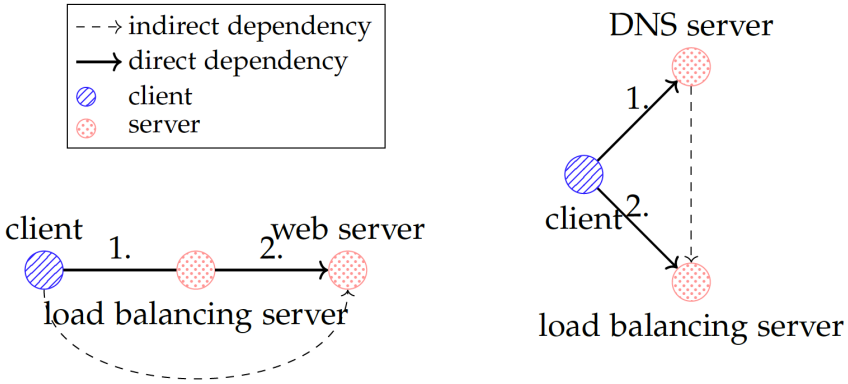
Support for Other Modules

- Enables other modules to work at all (normalization)
- Reduces load due to fusion and filtering
- Prioritization allows subsequent modules to focus on mission-critical events such that…
- ... attacks can be matched and ...
- ... relevant response plans can be generated ...
- ... in realtime

Network Service Dependency

Direct Dependency: A -> B, if A requires B to satisfy **certain** requests from its clients [Chen, Xu, al.]

Indirect Dependency: A -> B; A -> C, if request A -> B and A-> C are caused by the same activity



[1] L., Kuhr, Möller: **Using a Deep Understanding of Network Activities for Workflow Mining**, In: KI 2016, Springer

[2] L., Möller: **Time Series Data Mining for Network Service Dependency Analysis,** In: International Joint Conference SOCO 16-CISIS 16-ICEUTE, Springer

Detecting Dependencies

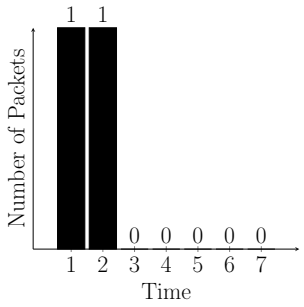
Normalized Cross-Correlation

$$\varrho_{r,s}(\tau) = \frac{\frac{1}{bins}\sum_{t=0}^{bins}(r_t - \mu_r)(s_{t+\tau} - \mu_s)}{\rho_r \rho_s},$$

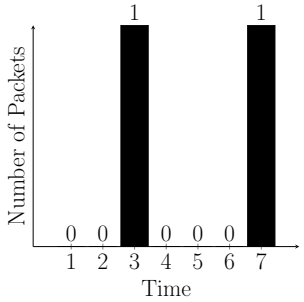
$$t_{delay} = \text{argmax}_{\tau=\{0,\ldots,(t_{max}-t_{min})\}\subseteq\mathbb{N}}\ \varrho_{r,s}(\tau).$$

$$\varrho_{r,s}(t_{delay}) \geq \theta$$

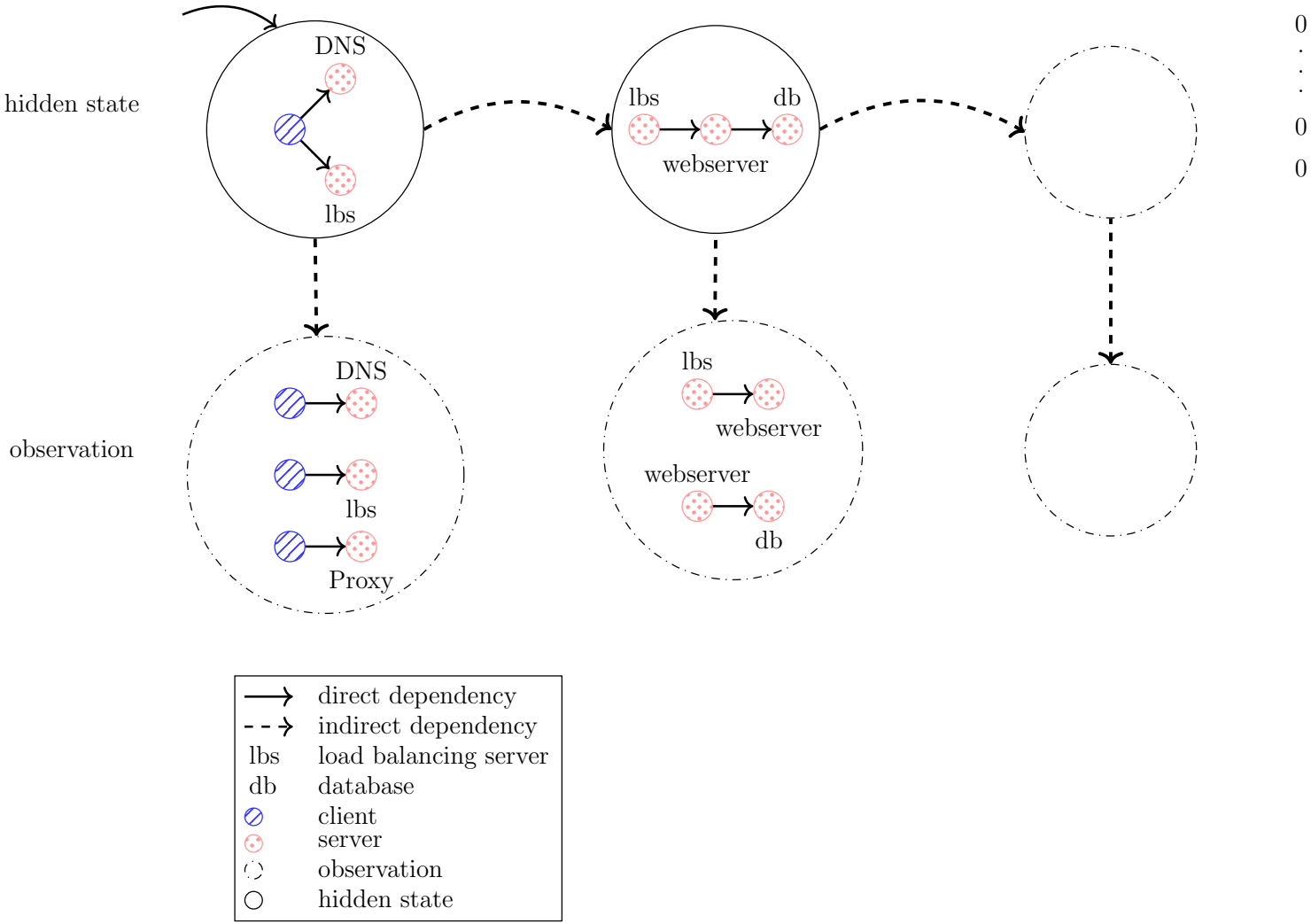
HMM for Workflow Modeling



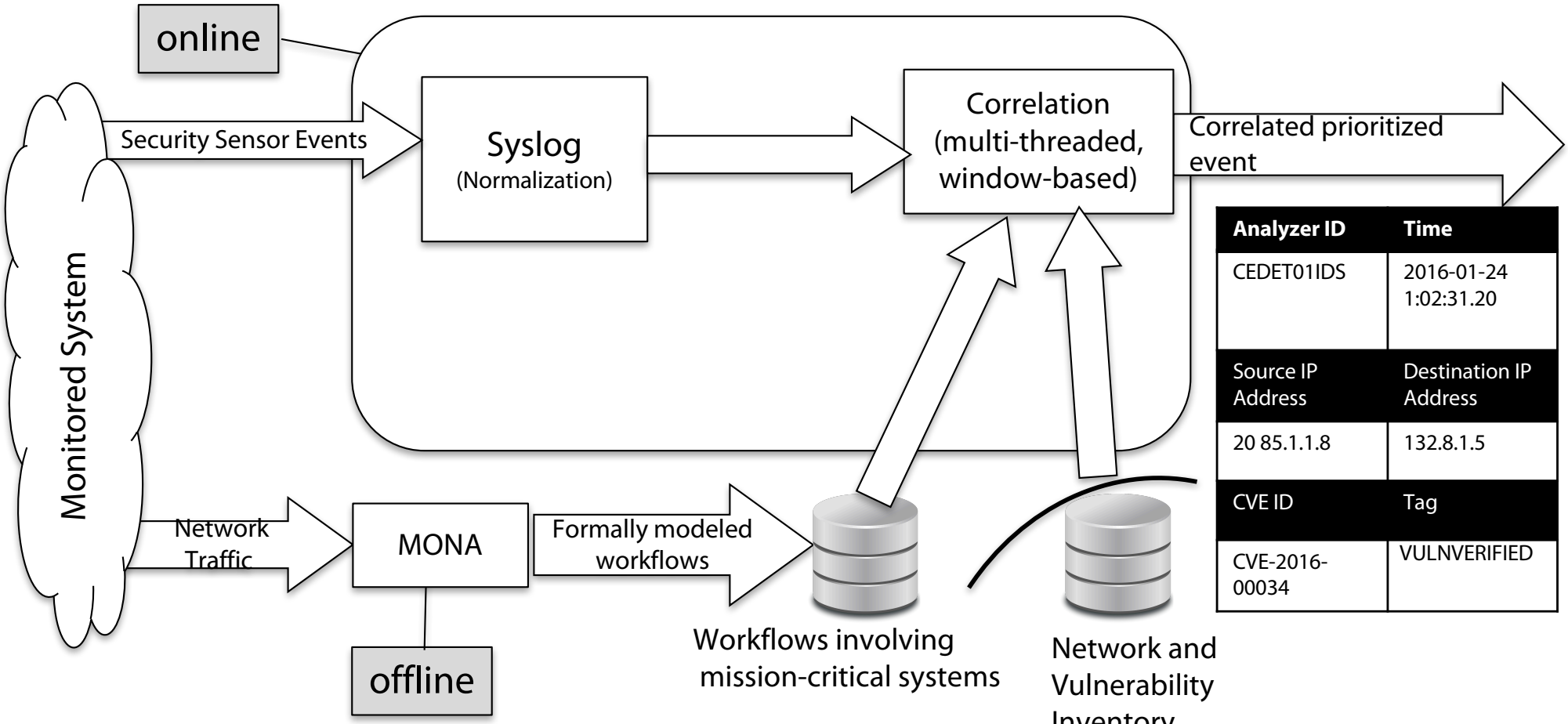
(a) Client→DNS Server



(b) Client→Load balancing server



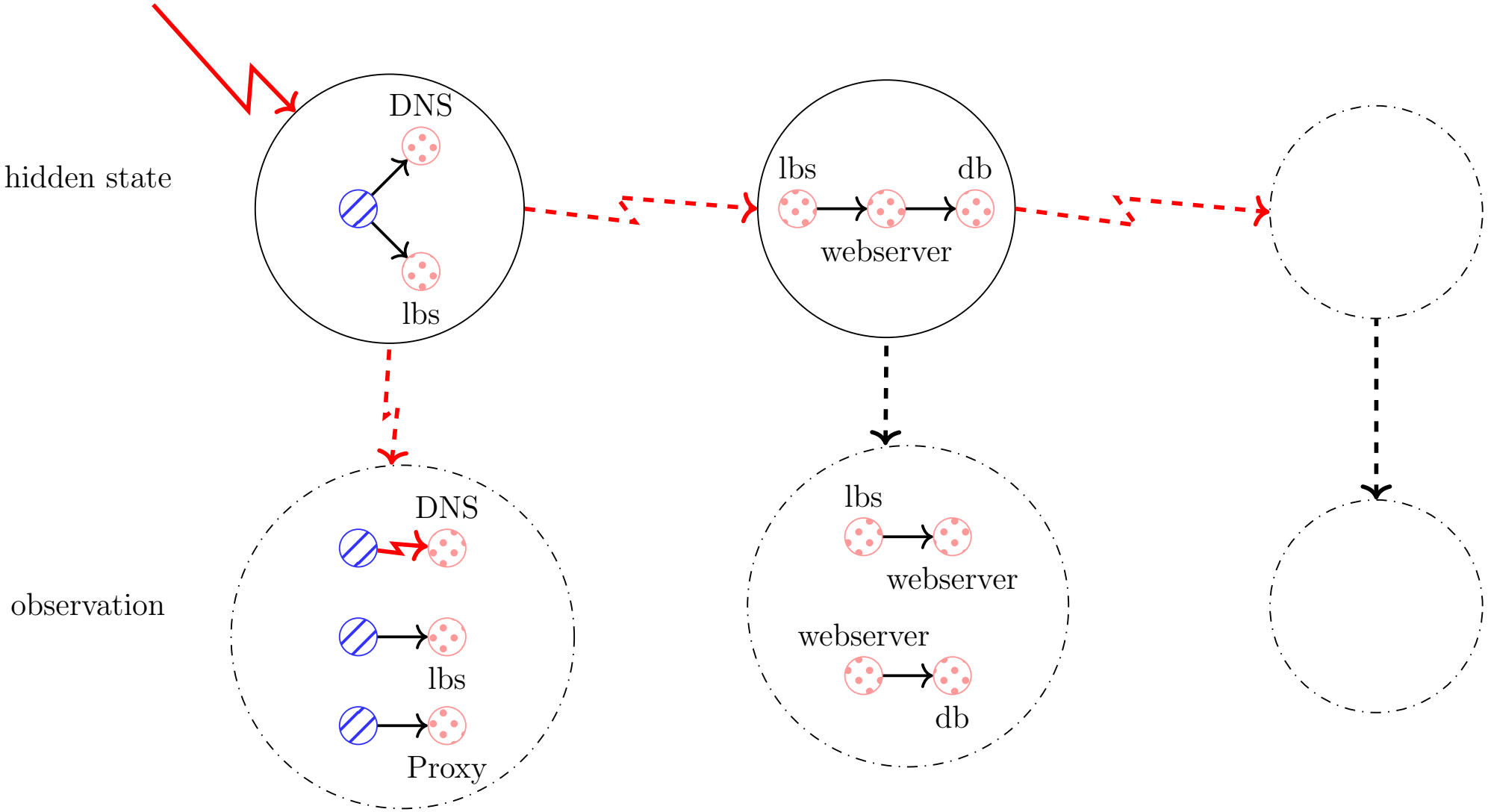
Context-Aware Event Correlation



Analyzer ID	Time
CEDET01IDS	2016-01-24 1:02:31.20
Source IP Address	**Destination IP Address**
20 85.1.1.8	132.8.1.5
CVE ID	**Tag**
CVE-2016-00034	VULNVERIFIED

IP	Vulnerability identifier
132.8.1.5	CVE-2016-0034

Workflows for Event Prioritization

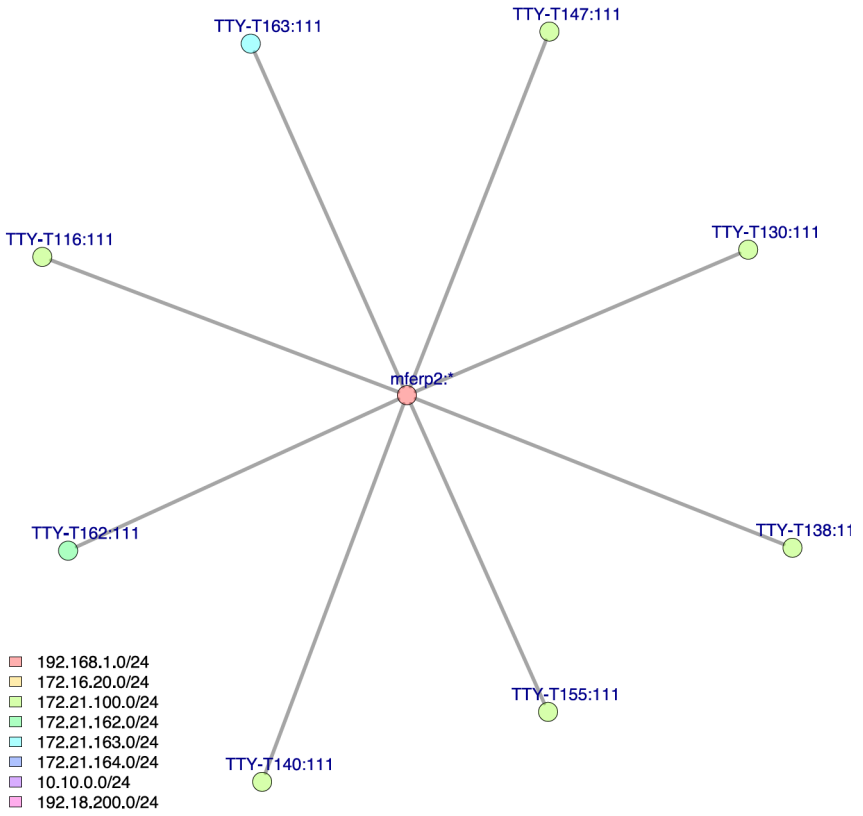


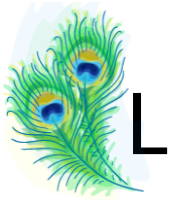
[3] Kott, L., Ludwig: **Assessing Mission Impact of Cyber Attacks: Towards a Model-Driven Paradigm**, In: IEEE Security Privacy, 2016

Using Workflows for Event Prioritization

Using a list of **mission-critical network devices**, workflows can be used to identify whether mission-critical network devices are affected.

Event Prioritization

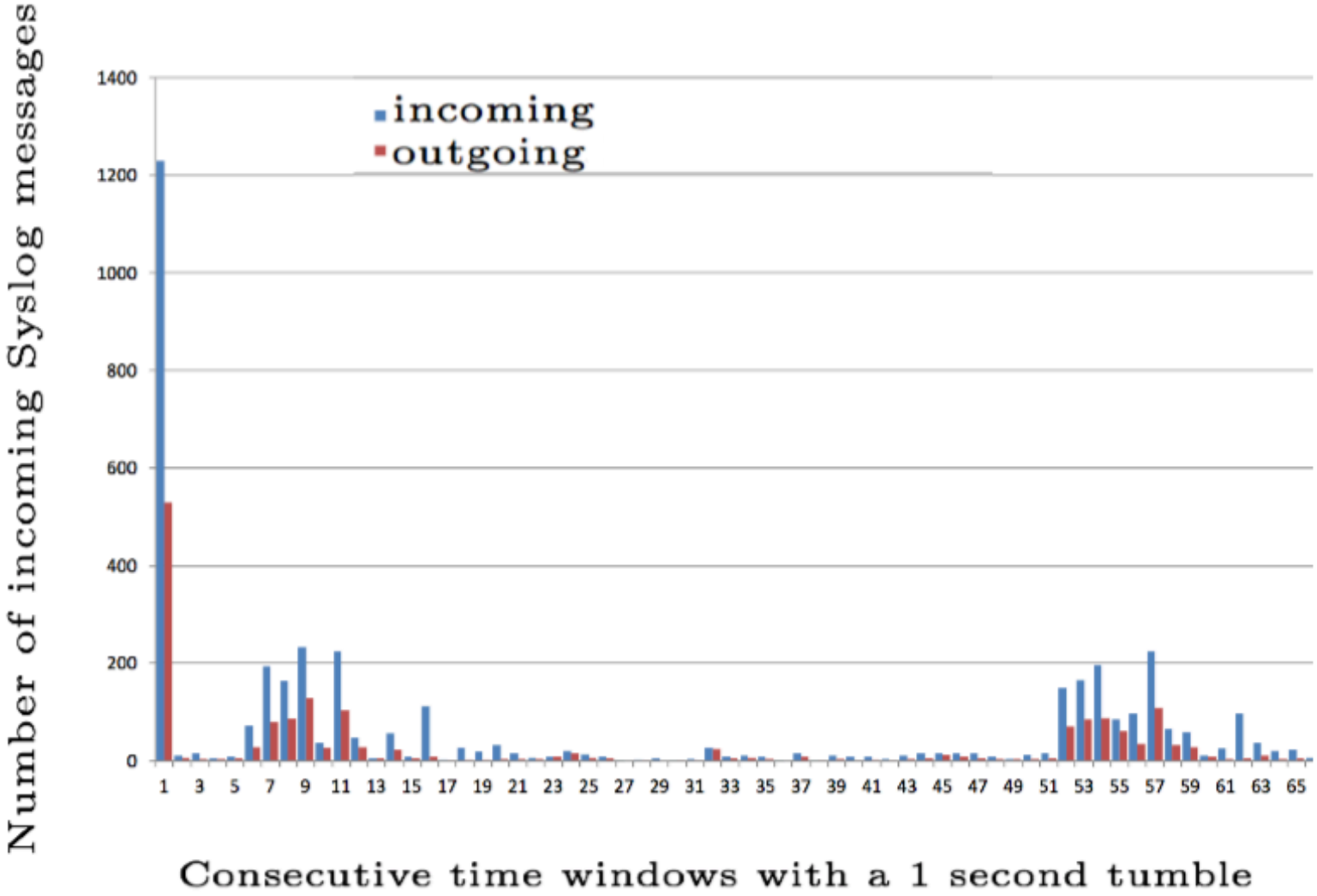




LLC – Scalability Tests

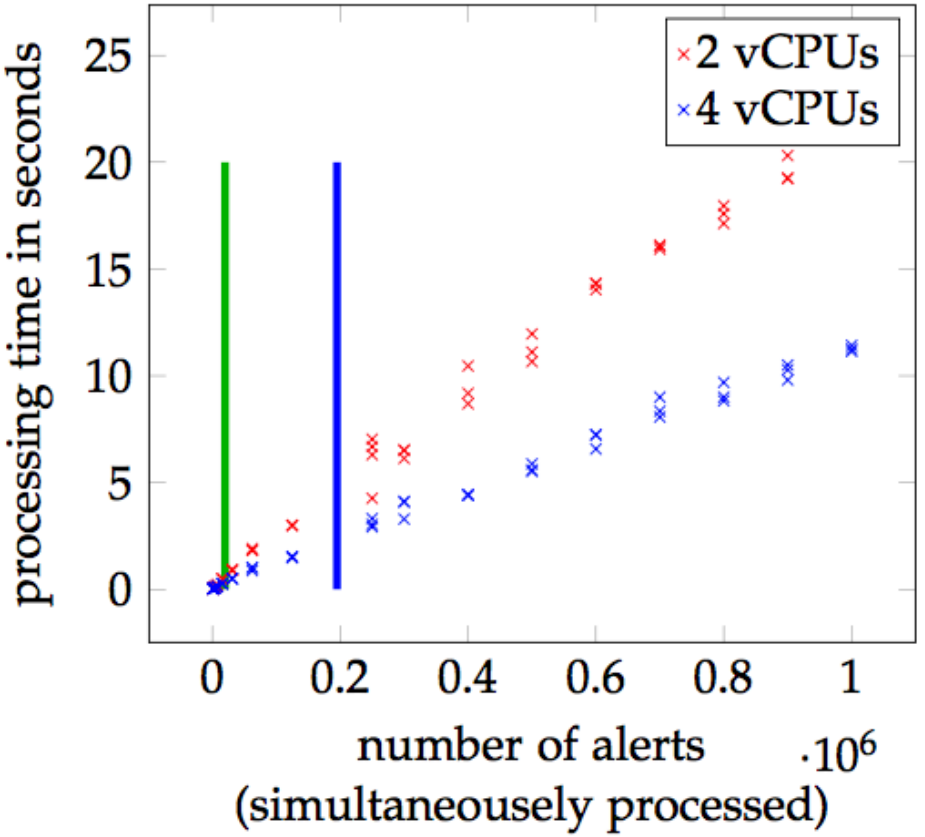
- Production environment 19.7.16 for about 7 hours
- LLC successfully deployed
- Overall >6M Syslog messages were received
- Due to the criticality of the production environment, IPS sensors and FWs block unexpected attempts of communication (white listing).
  - Therefore, as was expected, no LLC alerts were produced
  - Only events were processed
- LLC is able to perform within an operational environment
- Reduce the overall number of reported events by at least a factor of 2

LLC – Scalability Tests



LLC – Functionality and Performance Tests

- Emulation environment
- Functionality
  - Provides input for both HOC implementations
  - Used in operational workshop w/o any problems
- Performance
  - 10,000 events/sec 2CPUs
  - 100,000 events/sec 4CPUs
  - 1000,000 events/10sec 4CPUs



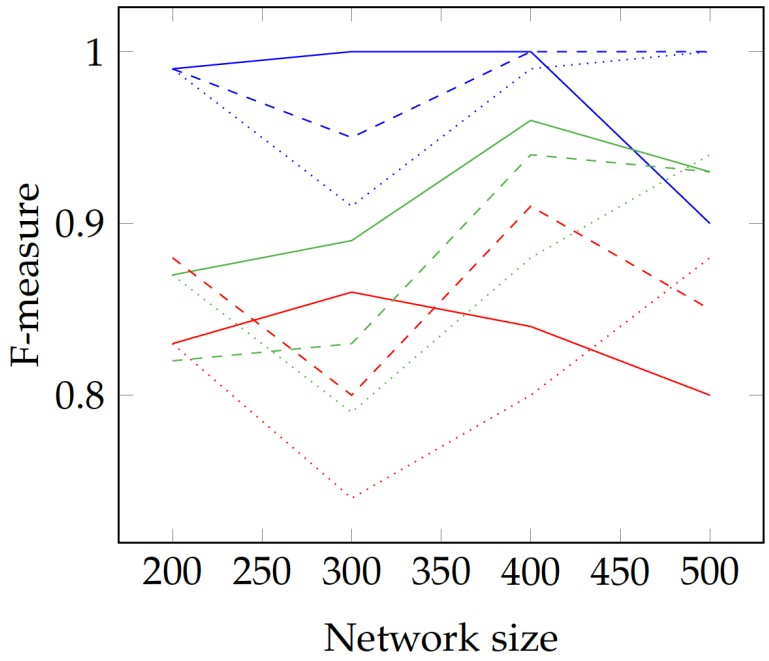
[4] L., Kuhr, Möller: **Using a Deep Understanding of Network Activities for Network Vulnerability Assessment**, PrAISe@ECAI 2016

[5] L., Kuhr, Möller: **Using a Deep Understanding of Network Activities for Network Vulnerability Assessment**, In: ECAI 2016

[6] L., Kuhr, Möller: **Using a Deeper Understanding of Network Activities for Security Event Management**, In: International Journal of Network Security & Its Applications (IJNSA), 2016

MONA: Performance Analysis

ACEA-Network + Synthetic networks



$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{FP}{TP + FN}$$

$$F - measure = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

True Positives (TP),

False Positives (FP),

False Negatives (FN)

Flows per communication between indirectly dependent network services:					
MONA	5 – 10 flows		5 – 50 flows		5 – 90 flows
Sherlock	5 – 10 flows		5 – 50 flows		5 – 90 flows
Orion	5 – 10 flows		5 – 50 flows		5 – 90 flows

Summary

- Online: Context-Aware Event Analysis
  - ✓ Normalize heterogeneous events from multiple sources
  - ✓ Filter and fuse events
  - ✓ Prioritization by operational impact assessment based on important activities ("workflows")

- Offline: Time Series Data Mining
  - ✓ Learn to identify workflows based on mining network traffic
  - ✓ Formally represent workflows as stochastic processes
  - ✓ Mission Oriented Network Analysis (MONA)

Bibliography

[1] Mona Lange, Ralf Möller: **Time Series Data Mining for Network Service Dependency Analysis,** In: International Joint Conference SOCO 16-CISIS 16-ICEUTE 16, San Sebastián, Spain, October 19-21, 2016, Manuel Graña, López-Guede, José Manuel, Oier Etxaniz, Álvaro Herrero, Héctor Quintián, Emilio Corchado (Ed.), Springer International Publishing, p.584-594

[2] Mona Lange, Felix Kuhr, Ralf Möller: **Using a Deep Understanding of Network Activities for Workflow Mining**, In: KI 2016: Advances in Artificial Intelligence - 39th Annual German Conference on AI, Klagenfurt, Austria, September 26-30, 2016, Springer, Lecture Notes in Computer Science, Vol.9904, p.177-184

[3] Mona Lange, Felix Kuhr, Ralf Möller: **Using a Deep Understanding of Network Activities for Network Vulnerability Assessment**, In: Proceedings of the 1st International Workshop on AI for Privacy and Security, PrAISe@ECAI 2016, The Hague, Netherlands,29.08.-02.09., 2016, ACM, p.6:1-6:8

[4] Mona Lange, Felix Kuhr, Ralf Möller: **Using a Deep Understanding of Network Activities for Network Vulnerability Assessment**, In: ECAI 2016 - 22nd European Conference on Artificial Intelligence, 29 August-2 September 2016, The Hague, The Netherlands - Including Prestigious Applications of Artificial Intelligence (PAIS 2016), 2016, Gal A. Kaminka, Fox, Bouquet, Hüllermeier, Dignum, Dignum, Frank van Harmelen (Ed.), IOS Press, Frontiers in Artificial Intelligence and Applications, Vol.285, p.1583-1585

[5] Mona Lange, Felix Kuhr, Ralf Möller: **Using a Deeper Understanding of Network Activities for Security Event Management**, In: International Journal of Network Security & Its Applications (IJNSA), 2016