PV226: The usage of Process Mining

Martin Macák

State of the Art – Process Mining domains

- Healthcare
- Manufacturing
- Finance
- Public sector
- Usability
- Robotics, industry 4.0
- Utility
- Advisory, audits
- Biology
- Agriculture
- ICT
- Education
- Logistics
- Security
- Call center
- Entertainment
- Garment
- Retail
- Hotel

More details: [1]

State of the Art – Cybersecurity domains

- Used in domains:
  - Network (IS, DNS, IDS, websites)
  - Smart grids (anomalous behaviour of energy usage)
  - Smartphones (social engineering attacks, malwares)
  - Banking (frauds, security deviations)
  - Industrial Control Systems (cyberattacks)
  - Business processes (anomalies, deviations)

State of the Art – Techniques

1. Target period of the analysis
   - past
   - present
2. Domain awareness
   - with domain knowledge
   - without any domain knowledge
3. Analysis of a discovered process model
   - visually
   - programmatically
4. Detection technique
   - outlier behavior detection
   - abnormal behavior detection (only in supervised analysis)
   - conformance checking

Process discovery: DNS traces

- Event log built from DNS traces (caseID, activity, timestamp)
- caseID= client, DNS Server
- activity = query/response, type
- Detection of spambots

Process discovery: DNS traces

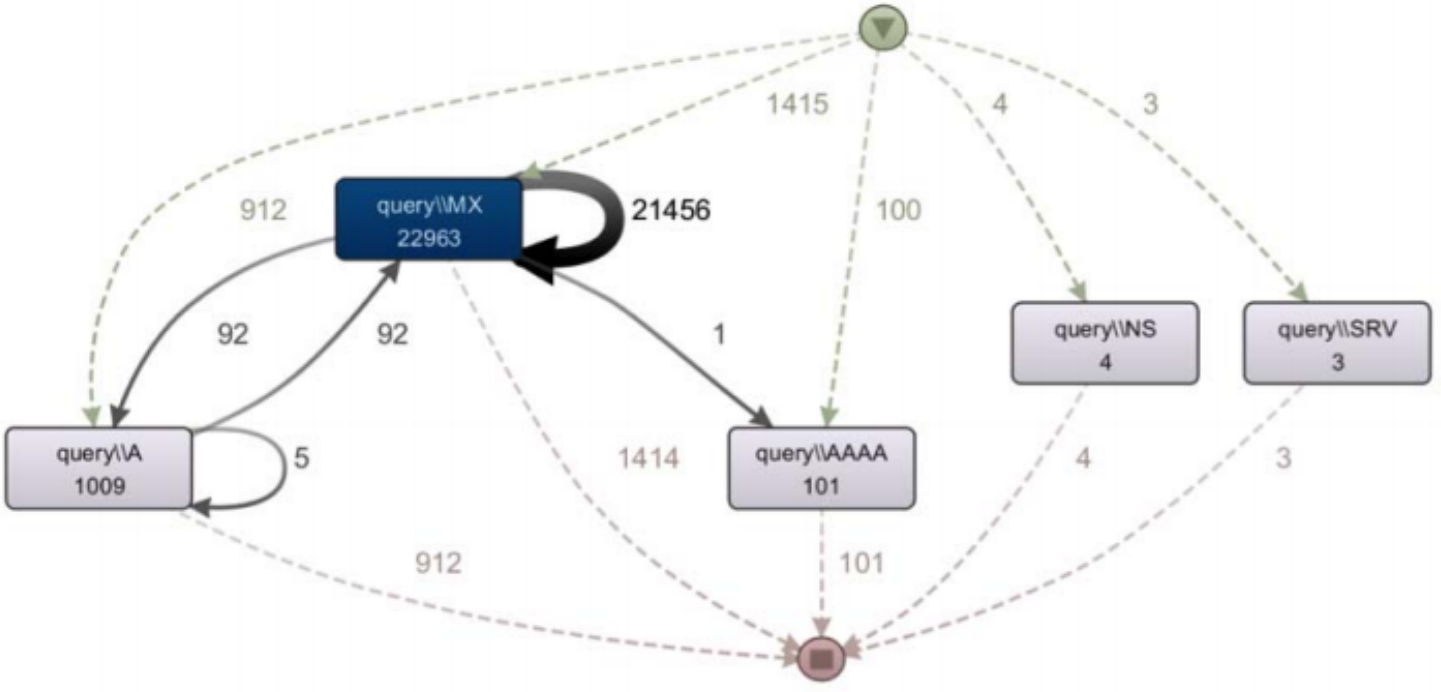


Fig. 6. Simplified graph of the attack shown in Figure 5. We show the model after filtering the 10% of most active IPs.

[2]

Model comparison: Smart Grids

- Anomaly detection of power consumption
- Classification of consumption to levels
- Then they discover graphs of consumption per short period
- Time-evolving graph approach: comparing consecutive graphs
- They chose Hamming distance and cosine similarity measure

Model comparison: Smart Grids

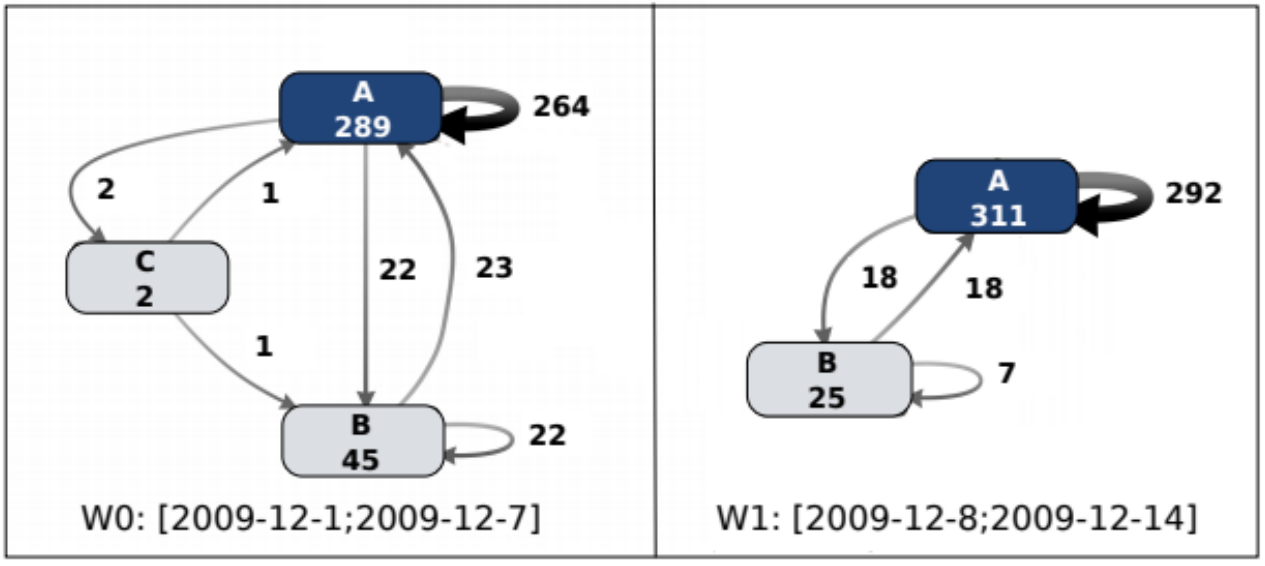


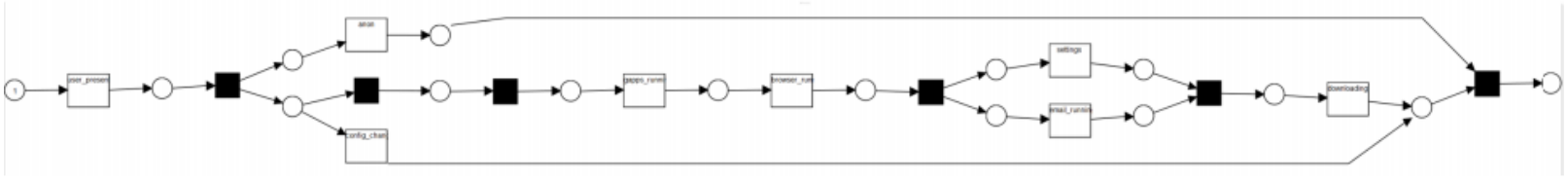
Figure 3. Consumption graphs of customer #1565 of two consecutive weeks.

[3]

Conformance checking: Smartphones

- Attack: user activated a malicious URL, which resulted in downloading personal user data via known vulnerability
- They designed a model of this attack from OS-generated information about performed actions, browser history, and network connection log
- Token-based replay with this model

Conformance checking: Smartphones



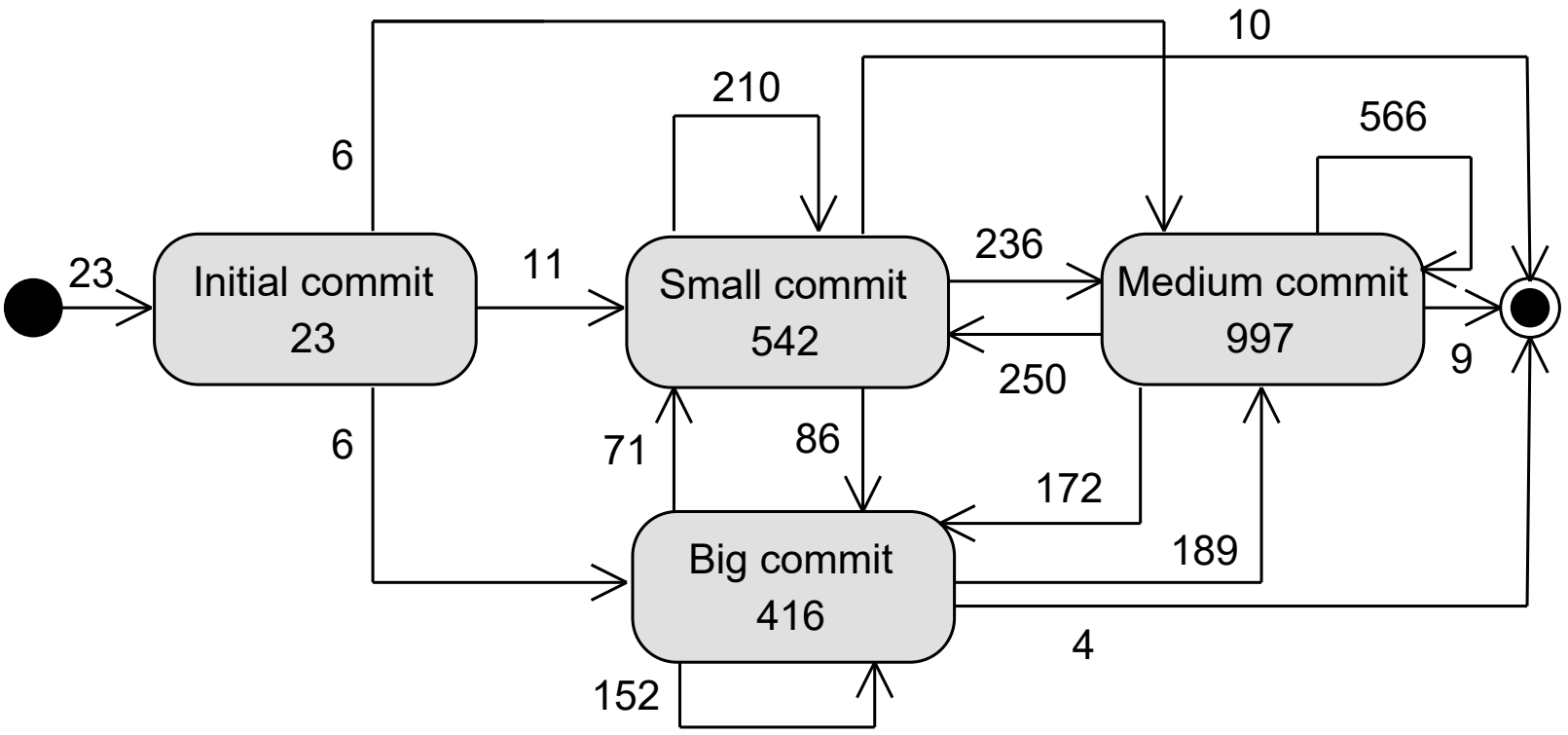
[4]

Lasaris Process Mining research

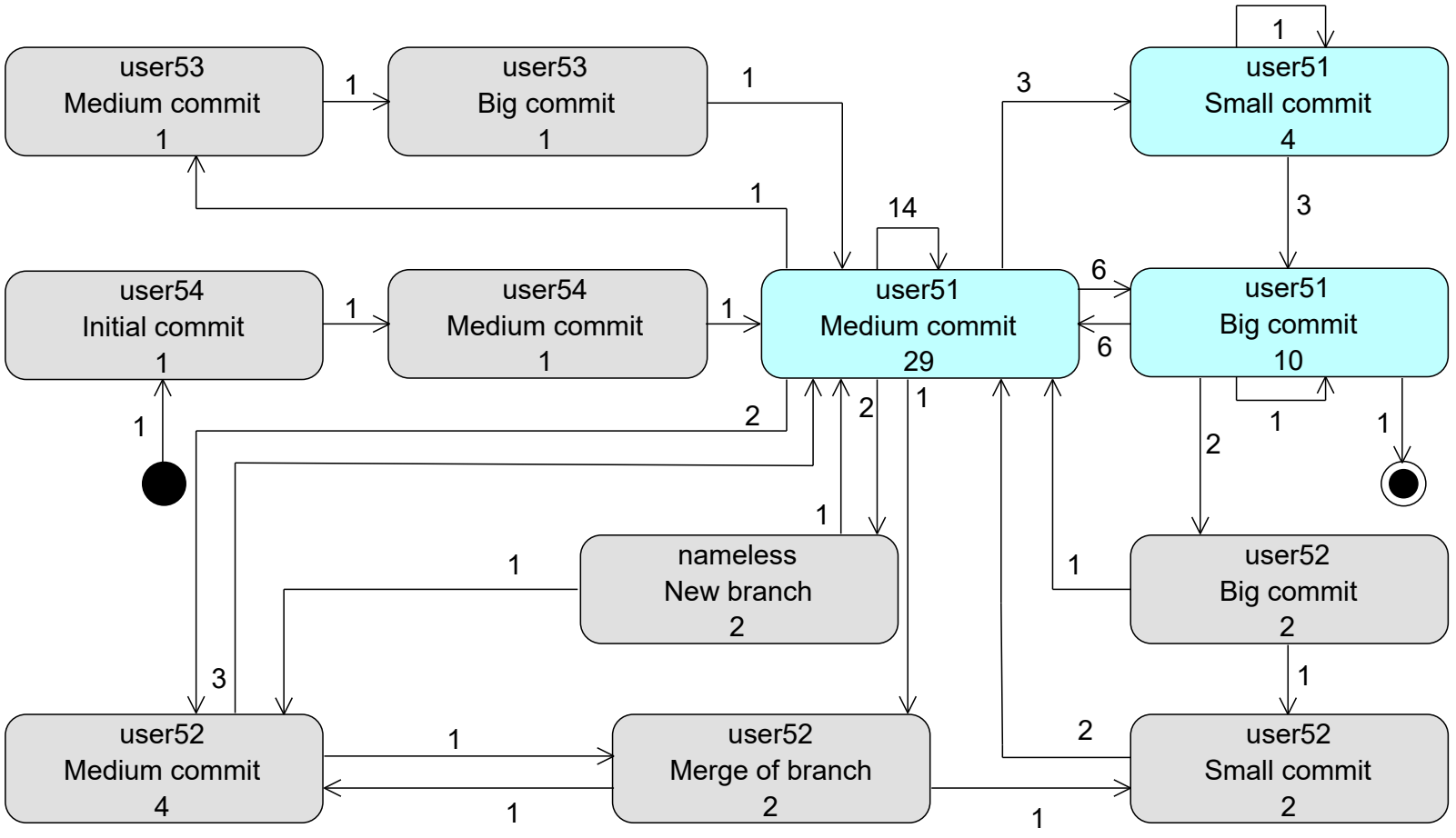
- Verification of Forensic Readiness in Software Development
- **Git Log Analysis of Projects in PV179**
- Detection of masquerading traitors from the process vizualization
- **Cybersecurity KYPO Training Analysis**
- Simulation Games Platform for Unintentional Perpetrator Attack Vector Identification
- **Insider Threat Detection from Audit Logs**

  + development for further research
- Process mining library ProcessM.NET
- Declarative Process mining tool

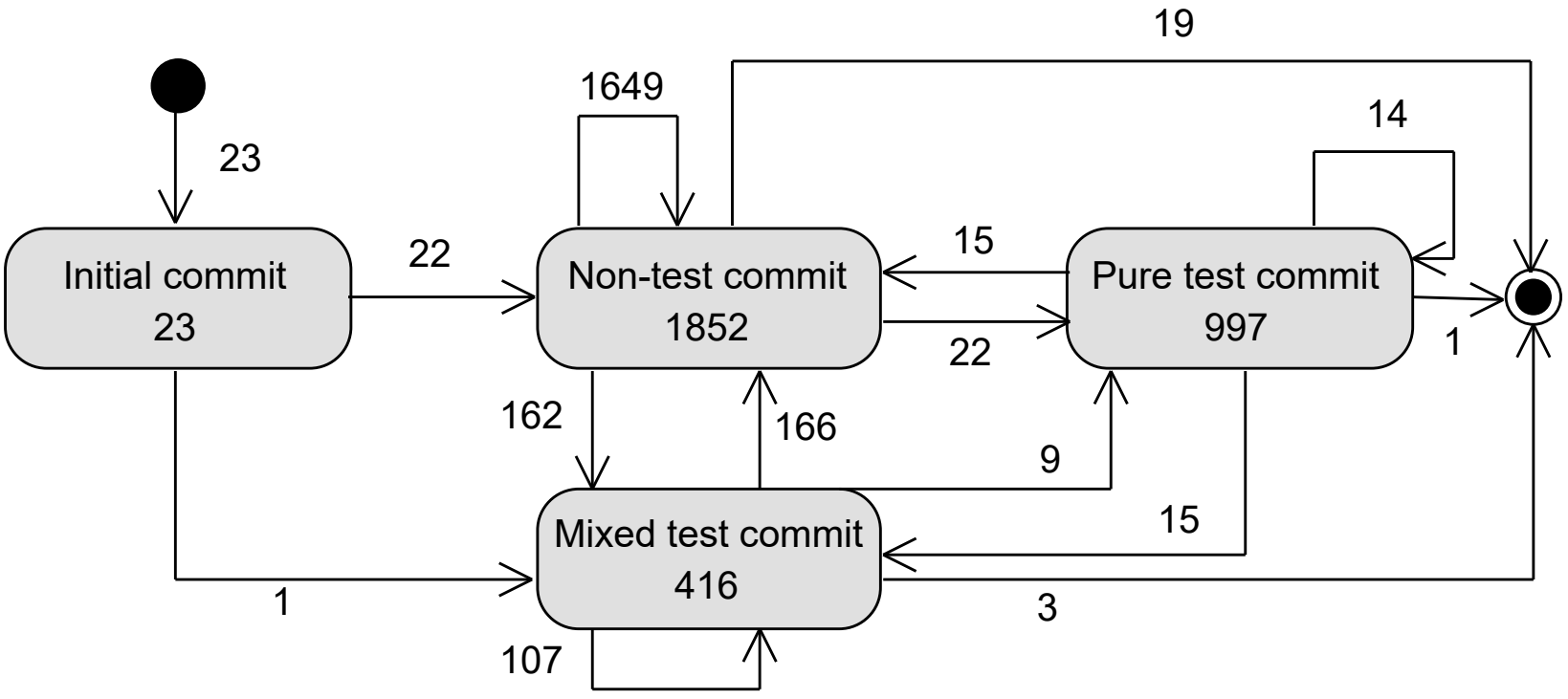
Git Log Analysis of Projects in PV179



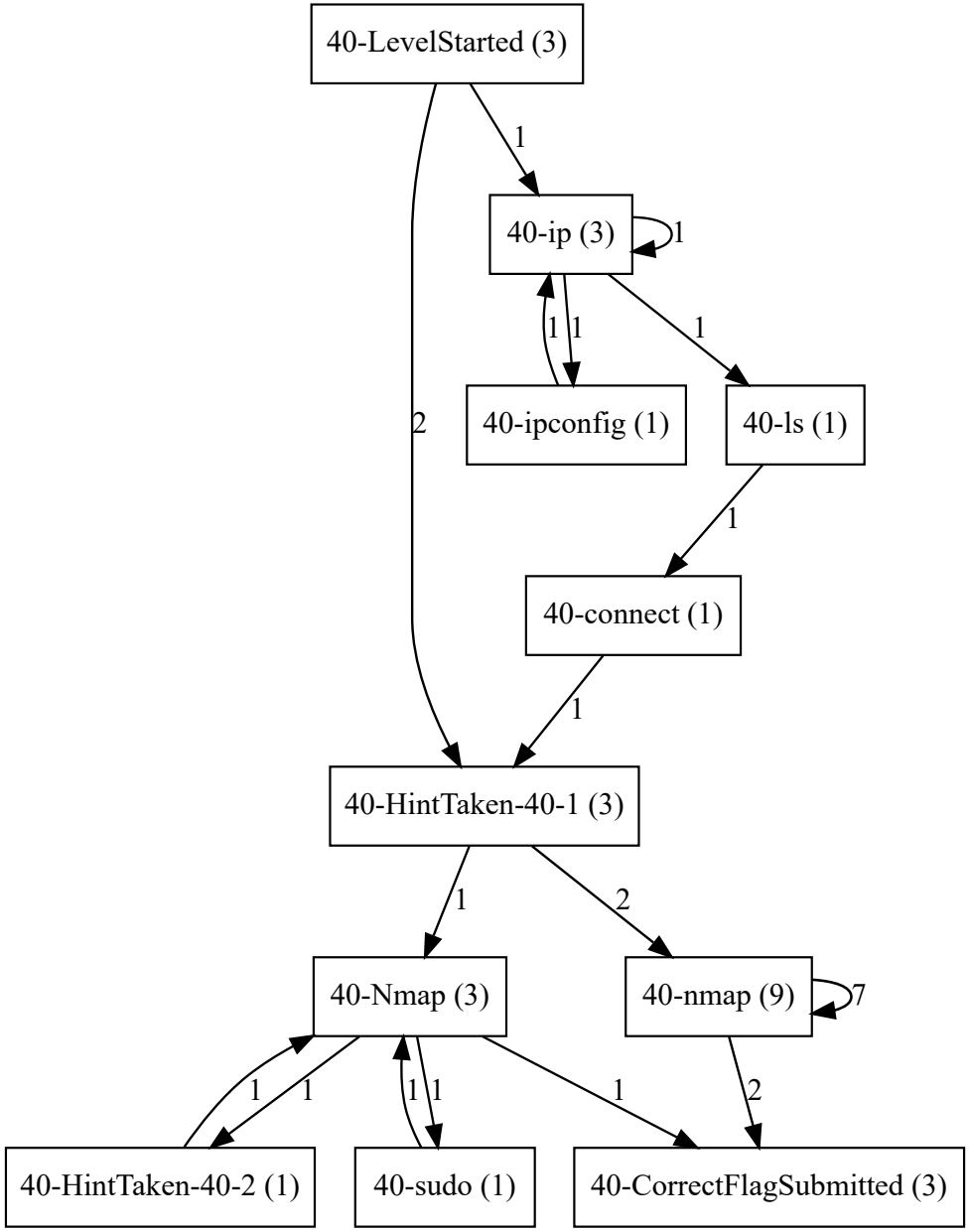
Git Log Analysis of Projects in PV179



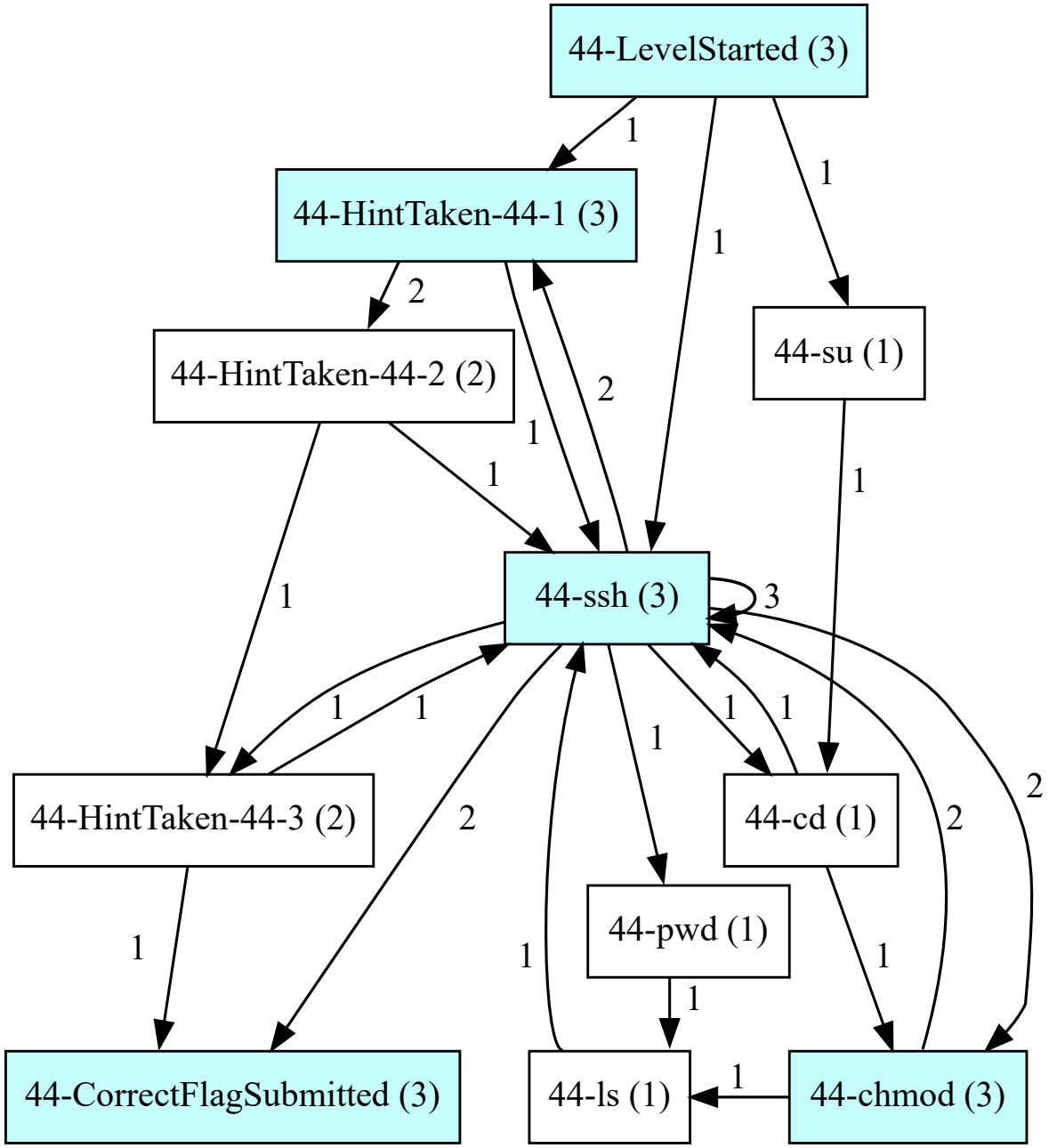
Git Log Analysis of Projects in PV179



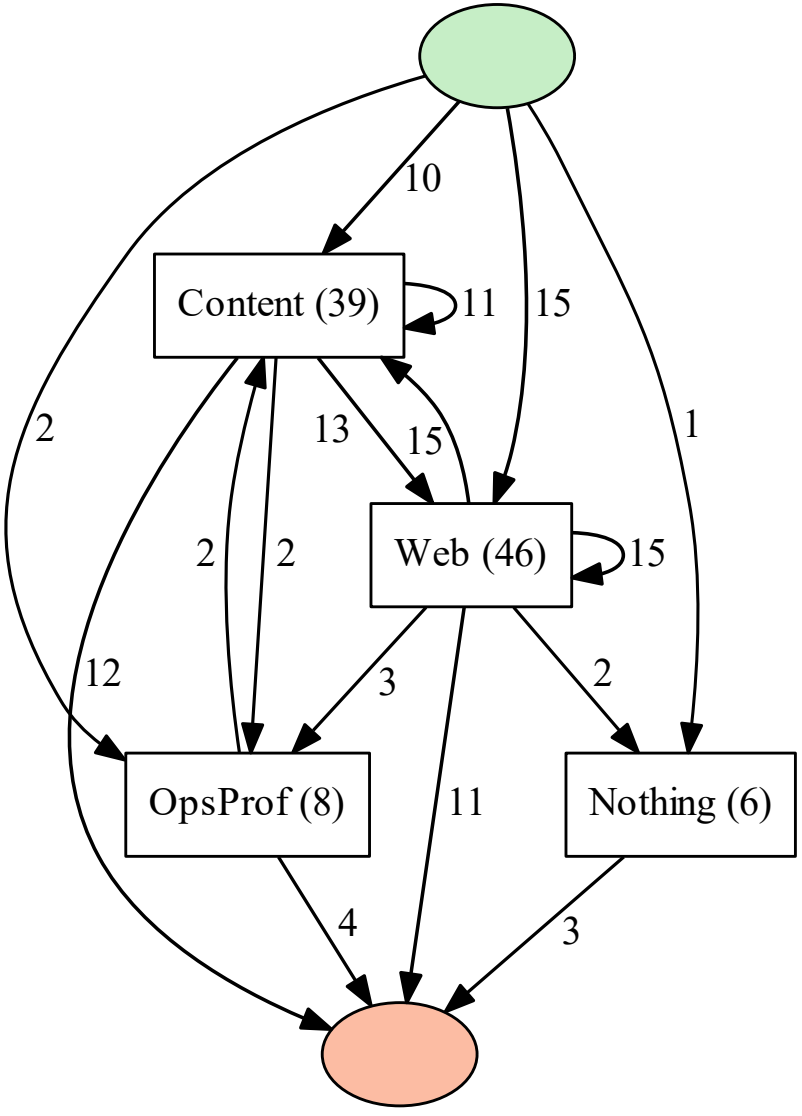
Cybersecurity KYPO Training Analysis



Cybersecurity KYPO Training Analysis



Insider Threat Detection from Audit Logs



Traces	Fitness
s1, s2, s3, s4, s5	0.85
s1, s2, s3, s4, m5	0.76
s1, s2, s3, m4, m5	0.56
s1, s2, m3, m4, m5	0.48
s1, m2, m3, m4, m5	0.36
m1, m2, m3, m4, m5	0.29

What can WE do?

- Generally:
  - Discover the process from event logs
  - Replay the log on top of that model and visually detect deviations
  - Filter the event log and manually analyse the desired cases
  - Find the deviations in an event log from the existing model using conformance checking
  - Real-time conformance checking
  - ...
- The biggest challenge might be to **find** suitable data to analyze and to **clean** them
- However, you can generate or create your own prototype event logs

Previous PV226 projects

- Analysis of:
  - Git activity
  - Game achievements
  - Software development from Jira logs
  - Tickets from ticket system
  - Sentiments of news articles
  - E-learning course about Python
  - COVID vaccinations
  - Traffic accidents
  - Counter Strike rounds

What can WE do?

- Specifically:
  - Your project should be interesting for you
  - You need to achieve something that can be presented
- Real examples of a project:
  - Process discovery from real datasets, for example: https://data.gov.cz/datové-sady, https://data.brno.cz/search?collection=Dataset
  - Process discovery of some groups (e.g., Big Data repositories of open source tools)
  - Process analysis of the behavior of people in some context
  - Extension or application of PM libraries
- Tools:
  - Disco
  - ProM
  - RapidMiner
  - Python (PM4Py) or C# (ProcessM.NET)

Resources

[1] C. dos Santos Garcia, A. Meincheim, E. R. F. Junior, M. R. Dallagassa,D. M. V. Sato, D. R. Carvalho, E. A. P. Santos, and E. E. Scalabrin,"Process mining techniques and applications - a systematic mappingstudy,"Expert Systems with Applications, vol. 133, pp. 260 – 295, 2019.doi: https://doi.org/10.1016/j.eswa.2019.05.003. [Online].

[2] J. Bustos-Jiménez, C. Saint-Pierre, and A. Graves, "Applying process mining techniques to dns traces analysis," in 2014 33rd International Conference of the Chilean Computer Science Society (SCCC) , Nov 2014. doi: 10.1109/SCCC.2014.9. ISSN 1522-4902 pp. 12–16

[3] S. Bernardi, R. Trillo-Lado, and J. Merseguer, "Detection of integrity attacks to smart grids using process mining and time-evolving graphs," in 2018 14th European Dependable Computing Conference (EDCC), Sep. 2018. doi: 10.1109/EDCC.2018.00032 pp. 136–139.

[4] L. Hluchý and O. Habala, "Enhancing mobile device security with process mining," in 2016 IEEE 14th International Symposium on Intelligent Systems and Informatics (SISY) , Aug 2016. doi: 10.1109/SISY.2016.7601493. ISSN 1949-0488 pp. 181–184.