Process Mining in Cybersecurity

Martin Macák

macak@mail.muni.cz

LAB OF SOFTWARE ARCHITECTURES
AND INFORMATION SYSTEMS

FACULTY OF INFORMATICS
MASARYK UNIVERSITY, BRNO

Summer School of
Applied Informatics,
Bedřichov
September 6, 2019

lasaris

Outline

1. Introduction to process mining
2. Process mining in cybersecurity

Process mining

- Process-centric data analysis

- What really happened in the past?
- Why did it happen?
- What is likely to happen in the future?
- When and why do people deviate?
- How to redesign a process to improve it?
- ...

Process mining

- Typically working with event logs which represent processes
- These logs have to contain cases (sequences of events)

```
Martin;order_start
Martin;select_hamburger
Martin;choose_card_payment
Martin;confirm_order
Martin;order_end
```

lasaris

Process mining

- Each event has:
  - caseId
  - activity
  - timestamp (optional)
  - resource (optional)
  - other data (optional)

```
1;order_accept;Dec 2, 2017 10:30:58 AM;Peter;21
1;order_cooked;Dec 2, 2017 10:39:24 AM;Victor;24
1;order_delivered;Dec 2, 2017 11:12:37 AM;Emma;19
```

lasaris

Process mining

- Sometimes, the mapping is not clear

```
1;order_accept;Dec 2, 2017 10:30:58 AM;Peter;21
1;order_cooked;Dec 2, 2017 10:39:24 AM;Victor;24
2;order_accept;Dec 2, 2017 10:40:21 AM;Peter;21
3;order_accept;Dec 2, 2017 10:42:19 AM;Greg;34
1;order_delivered;Dec 2, 2017 11:12:37 AM;Emma;19
2;order_cooked;Dec 2, 2017 11:17:04 AM;Victor;24
2;order_delivered;Dec 2, 2017 11:24:00 AM;Peter;21
```

- For example, the name of the worker can be:
  - resource
  - activity
  - caseId

lasaris

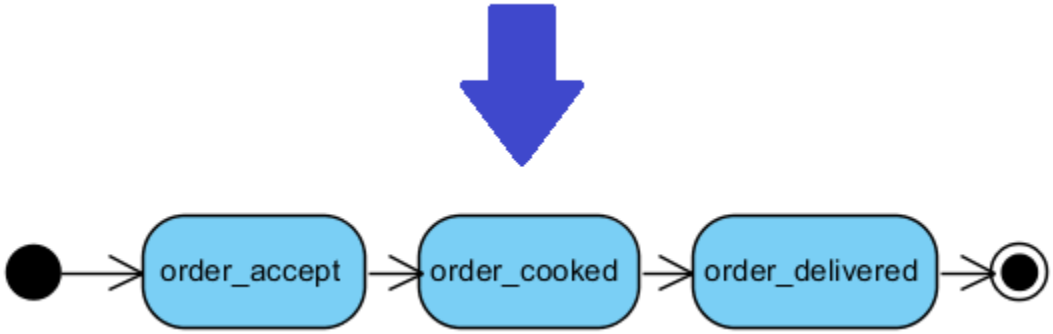
Analysis of the past

- Process discovery techniques
- From the event log, we create a model that represents how the process was executed in reality
- Model can be represented as a petri net, activity diagram, BPMN diagram, …
- Conformance checking techniques

lasaris

Process discovery

```
1;order_accept;Dec 2, 2017 10:30:58 AM;Peter;21
1;order_cooked;Dec 2, 2017 10:39:24 AM;Victor;24
2;order_accept;Dec 2, 2017 10:40:21 AM;Peter;21
3;order_accept;Dec 2, 2017 10:42:19 AM;Greg;34
1;order_delivered;Dec 2, 2017 11:12:37 AM;Emma;19
2;order_cooked;Dec 2, 2017 11:17:04 AM;Victor;24
2;order_delivered;Dec 2, 2017 11:24:00 AM;Peter;21
```



lasaris

Process discovery challenges

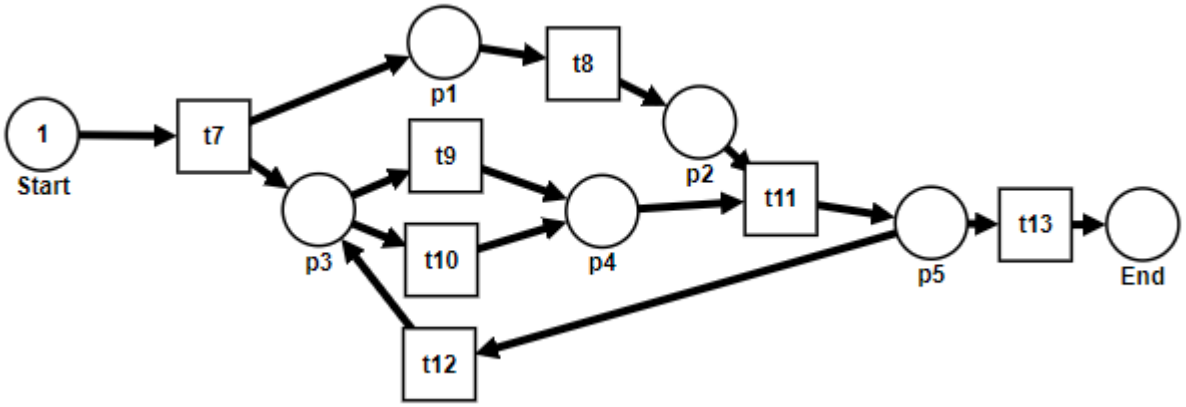
- Concurrency
- Loops
- Noisy behavior
- No negative examples in the log
- Too many allowed behaviors

Process discovery activities

- Explore processes at run-time
- Discover process models
- Compare the model of desired behavior with the model of reality
- Check the deviations in historic data
- Promote the model that shows the desired behavior

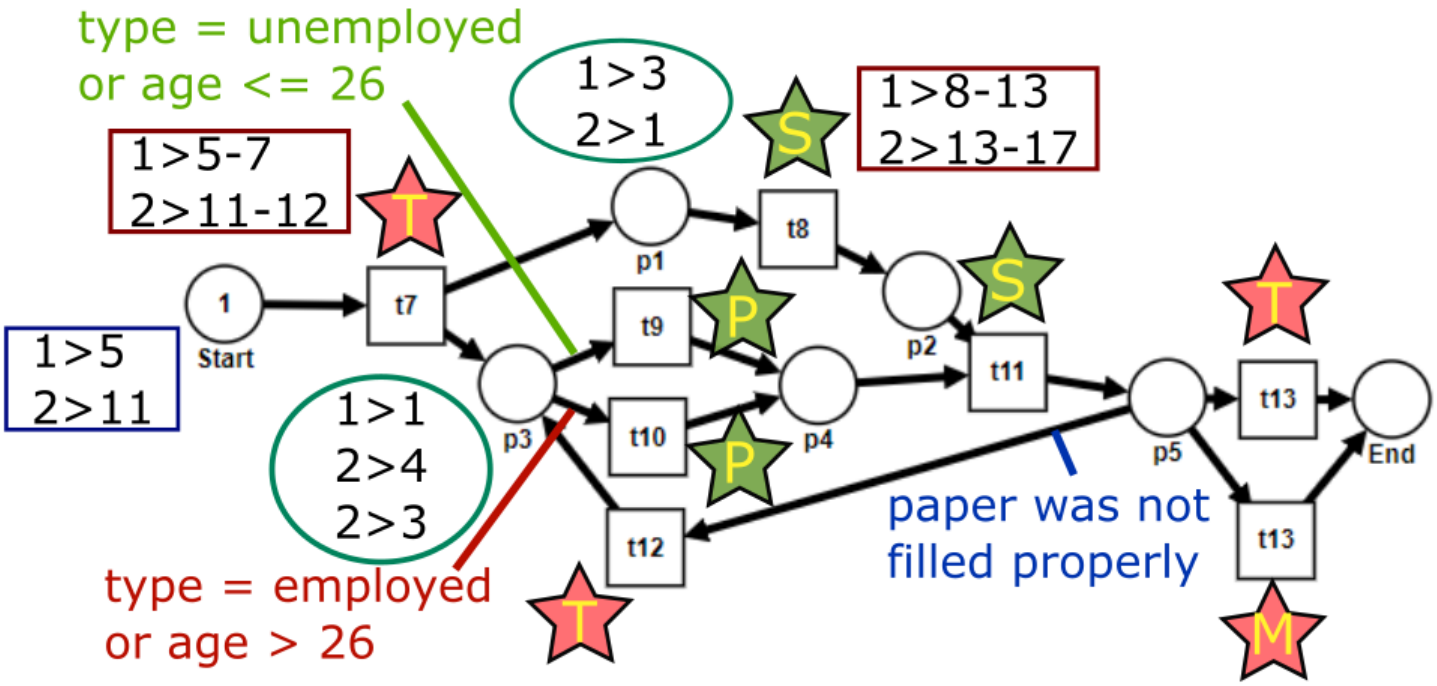
Adding additional perspectives

- Control flow is not the only perspective
- We can enhance the existing process models with:
  - Social network analysis
  - Organizational structures
  - Resource behavior analysis
  - Time perspective
  - Decision points mining
  - …



Additional perspectives

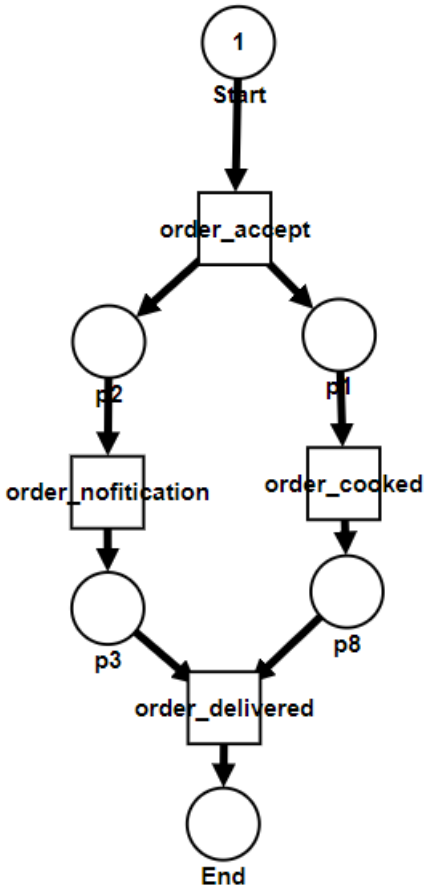
- We can add many others
- We can combine them to the integrated model
- Our model is enhanced, we might get better results



Conformance checking

- We can use the existing model to identify deviations in the behavior from logs

```
1;order_accept;
1;order_nofitication;
1;order_delivered;
1;order_cooked;
//NOK
```



Analysis of the present

- Also called operational support
- We use our model to analyze running cases
- We can:
  - Detect deviations in real-time data using the model of the desired behavior
  - Do real-time predictions (prob. of success, remaining time,…)
  - Make recommendations

Operational support: Detect deviations

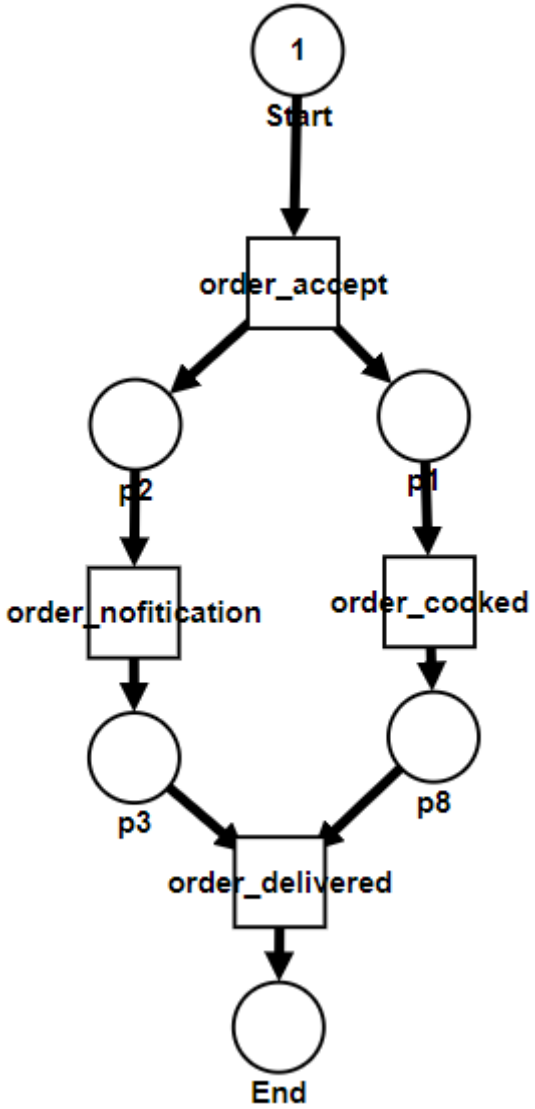
- We consider only the partial trace of a particular case
- We want immediate response when the deviation occurs

    a) Token-based replay
    b) Business rules

Detect deviations: Token-based replay

- Check the conformance with the model

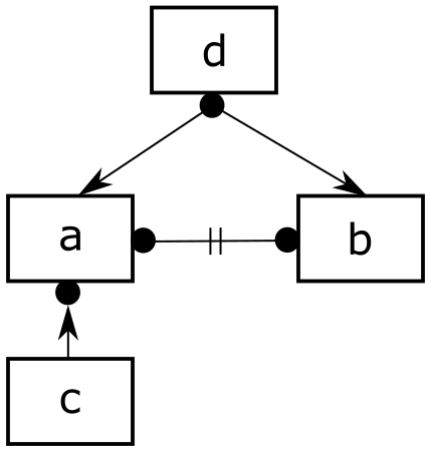
```
1;order_accept;          //OK
1;order_nofitication;    //OK
1;order_cooked;          //OK
1;order_delivered;       //OK

5;order_accept;          //OK
5;order_nofitication;    //OK
5;order_delivered;       //NOK
```



Detect deviations: Business rules

- Specific rules we want to follow
- To define them, we can use *Declare*
  - Constraint-based workflow language that uses graphical notations and semantics based on Linear Temporal Logic
- Example:
  - **a** and **b** cannot happen in the same case
  - **a** cannot happen before **c** has happened
  - every **d** have to be eventually followed by **a** or **b**

$$!((\blacklozenge a) \wedge (\blacklozenge b))$$
$$(!a)\ W\ c$$
$$\square(d \Rightarrow (\blacklozenge(a \vee b)))$$



lasaris

Operational support: Predict & Recommend

- We can apply data mining techniques (supervised learning, …)
- Examples of predictions:
  - Total cost of the current case
  - Total service time for the current case
  - Probability of meeting the deadline
  - Remaining flow time
- Examples of recommendations:
  - Minimize the total costs
  - Maximize the number of accepted cases
  - Minimize resource usage
  - Minimize the remaining flow time

lasaris

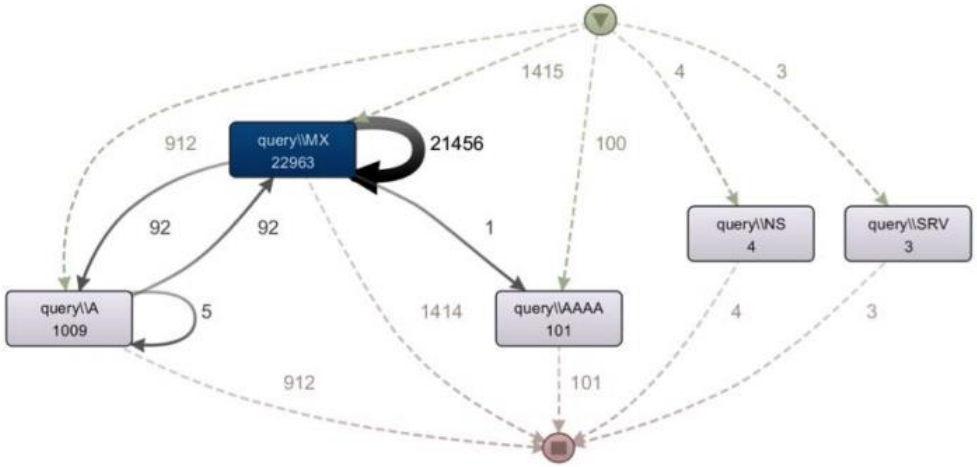
Process Mining in Cybersecurity

- Visual analysis of model
- Model comparison
- Conformance checking

lasaris

Visual analysis of model: DNS traces

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

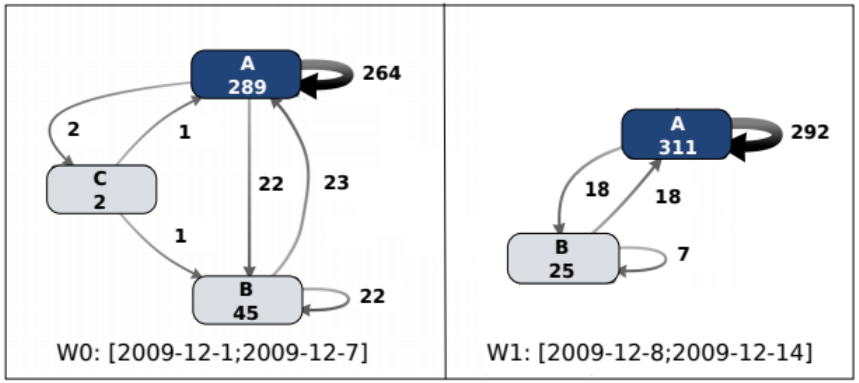


[1]

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

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 using a distance or similarity function
- They chose Hamming distance and cosine similarity measure

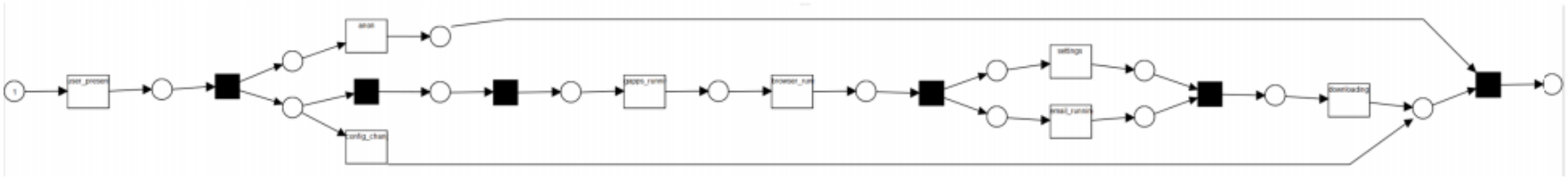


[2]

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

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



[3]

lasaris

Process mining cybersecurity domains (fields)

- 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 in the event data)

lasaris

Conclusion

Process mining: discover, enhance, operational support

Usage of process mining in cybersecurity:

- Visual analysis of model
- Model comparison
- Conformance checking

Current domains:

- Smart grids, Banking, Smartphones, Business processes, Network, Industrial Control Systems

Martin Macák, FI MU Brno
macak@mail.muni.cz

lasaris

Questions & comments?

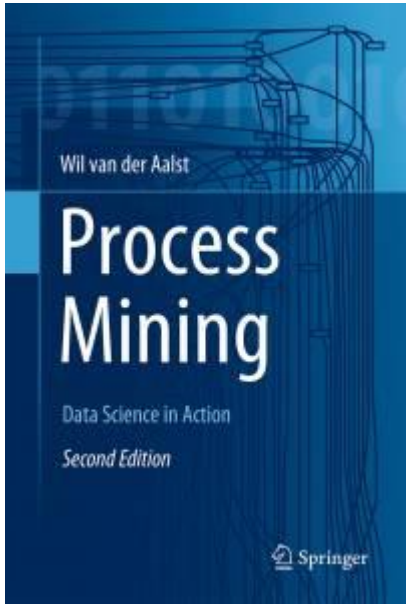
lasaris

Sources

- [1] 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
- [2] 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.
- [3] 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.

Sources

- Presentation based on the book **Process Mining: Data Science in Action**
- https://www.springer.com/gp/book/9783662498507



lasaris