Lasaris Process Mining Research

Martin Macák
macak@mail.muni.cz

Faculty of Informatics, Masaryk University

February 29, 2024

Process mining

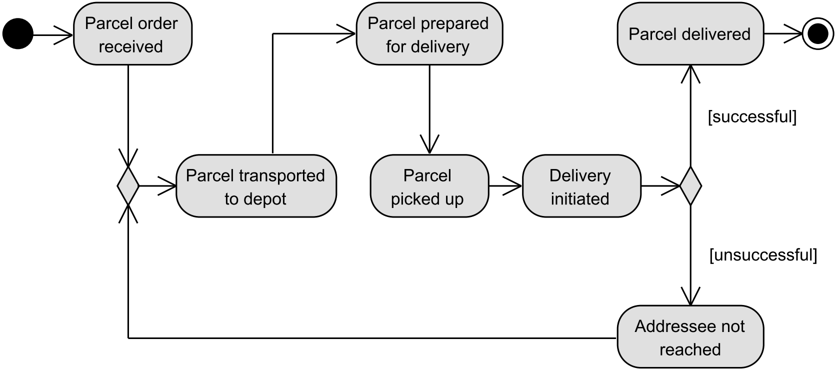
- Discipline proposed to give a better understanding of the processes by extracting knowledge from event logs.
- **Process:** a series of related activities that are performed in a specific sequence to achieve a particular goal within an organization.
- **Event log:** a structured collection of data that captures all the relevant events that occur during the execution of a process.

Why is process mining important?

- We assume how the process is performed.
- However, how the real process looks like?

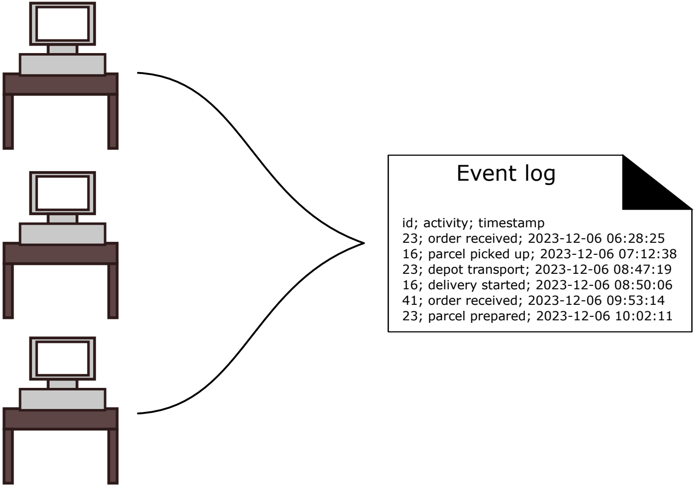
Process model

- A visual representation of a process.
- UML Activity diagram, BPMN, Petri net, ...



Event log

- **Case ID**: a unique identifier associated with a single process instance.
- **Activity**: a specific step in the process.
- **Timestamp**

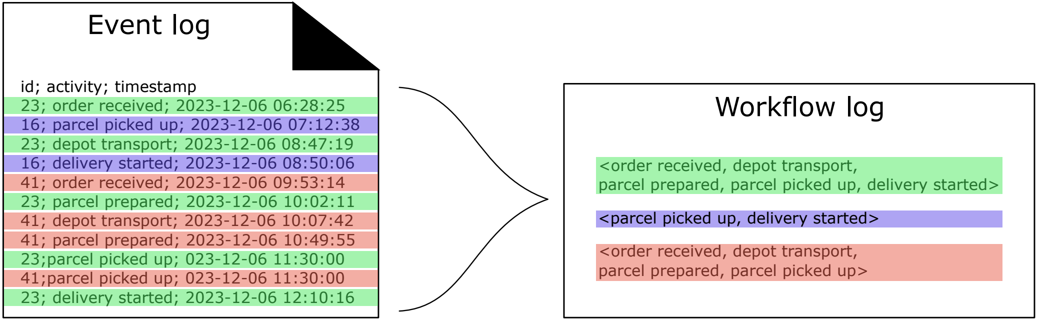


Event log

```
id; activity; timestamp
23; order received; 2023-12-06 06:28:25
16; parcel picked up; 2023-12-06 07:12:38
23; depot transport; 2023-12-06 08:47:19
16; delivery started; 2023-12-06 08:50:06
41; order received; 2023-12-06 09:53:14
23; parcel prepared; 2023-12-06 10:02:11
```

Trace

- **Trace** is a representation containing only ordered activities of a given case.
- A collection of traces is called **workflow log**.



Event log

```
id; activity; timestamp
23; order received; 2023-12-06 06:28:25
16; parcel picked up; 2023-12-06 07:12:38
23; depot transport; 2023-12-06 08:47:19
16; delivery started; 2023-12-06 08:50:06
41; order received; 2023-12-06 09:53:14
23; parcel prepared; 2023-12-06 10:02:11
41; depot transport; 2023-12-06 10:07:42
41; parcel prepared; 2023-12-06 10:49:55
23;parcel picked up; 023-12-06 11:30:00
41;parcel picked up; 023-12-06 11:30:00
23; delivery started; 2023-12-06 12:10:16
```

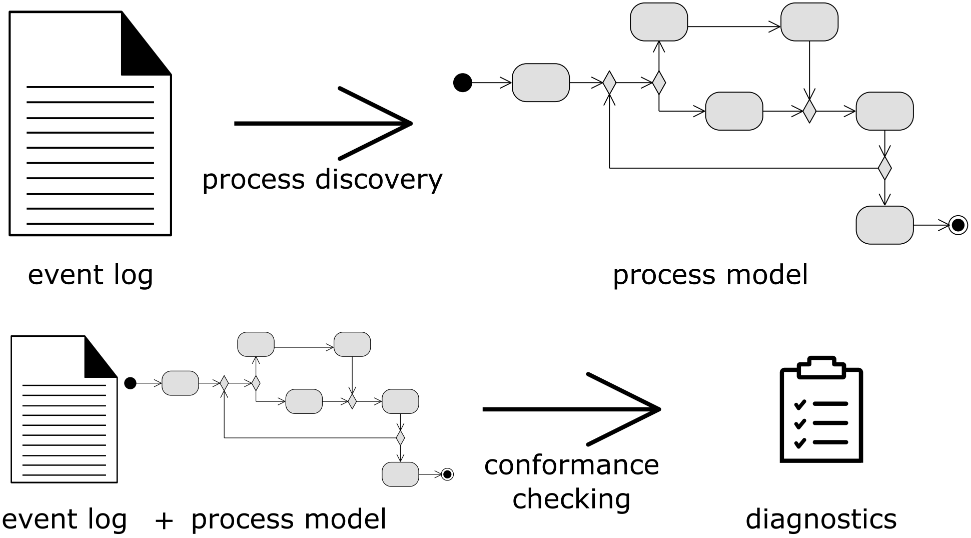
Workflow log

<order received, depot transport,
parcel prepared, parcel picked up, delivery started>

<parcel picked up, delivery started>

<order received, depot transport,
parcel prepared, parcel picked up>

Process mining types



event log

process discovery

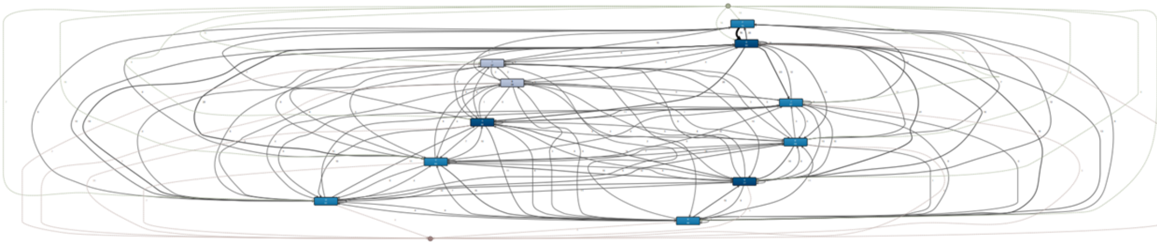
process model

event log + process model

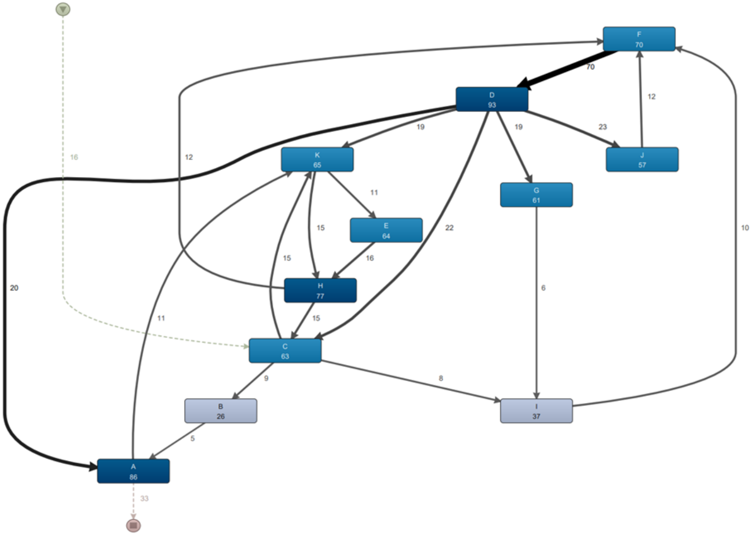
conformance checking

diagnostics

Declarative paradigm

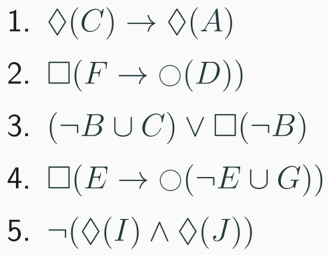


Declarative paradigm

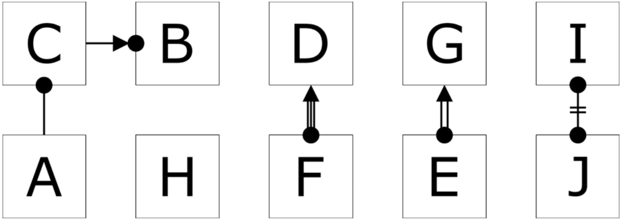


Declarative paradigm - Rules example

1. If C occurs, then A occurs.
2. Each time F occurs, then D occurs immediately after F.
3. B occurs only if preceded by C.
4. Each time E occurs, then G occurs afterward before E recurs.
5. I and J never occur together.



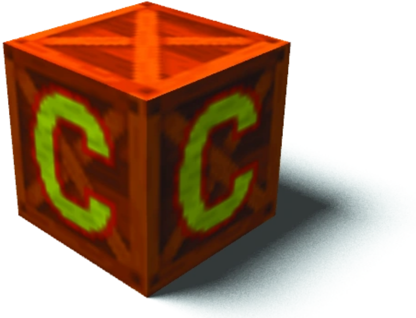
1. $\Diamond(C) \rightarrow \Diamond(A)$
2. $\Box(F \rightarrow \bigcirc(D))$
3. $(\neg B \cup C) \vee \Box(\neg B)$
4. $\Box(E \rightarrow \bigcirc(\neg E \cup G))$
5. $\neg(\Diamond(I) \wedge \Diamond(J))$



ProcessM.NET

- For implementation of process mining algorithms, you can check our .NET process mining library: `https://github.com/lasaris/ProcessM.NET`.

Checkpoint



Where we use process mining in Lasaris?

1. Cybersecurity
2. Software engineering

Cybersecurity

1. Insider Attack Detection
2. Cybersecurity Training Session Analysis
3. Internet of Secure Things

Insider Attack Detection

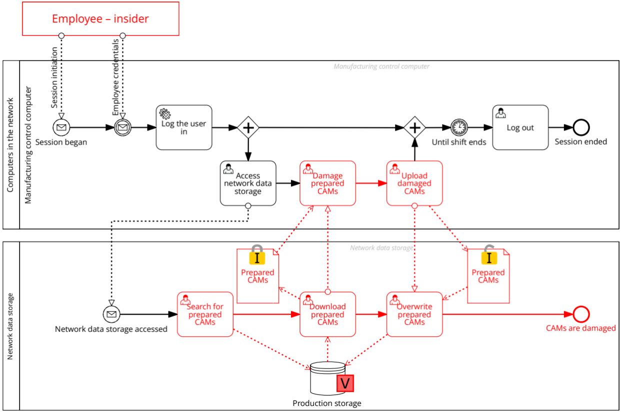
- Audit logs
- Manufacturing
- Windows Processes
- Information Systems

Audit logs

- The goal was to explore the potential of process mining analysis types.
  - User application activity.
  - Data flow.
  - File log.
  - Collaboration with Safetica.

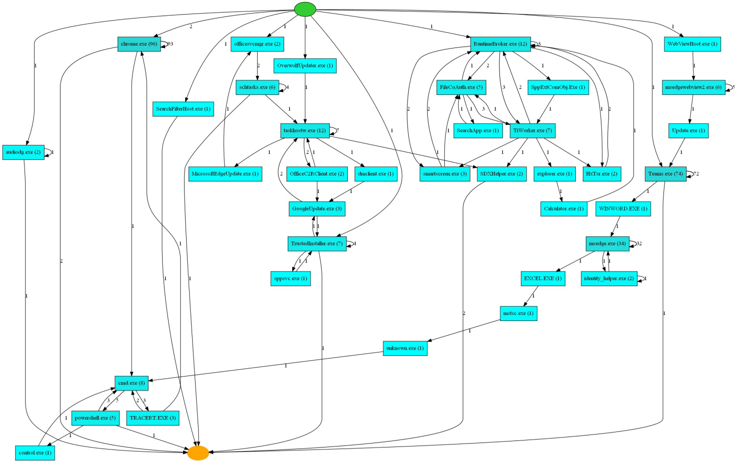
Manufacturing

- The goal was to identify the scenarios where process mining can help with the detection of insider attacks.
- Collaboration with a manufacturing company.



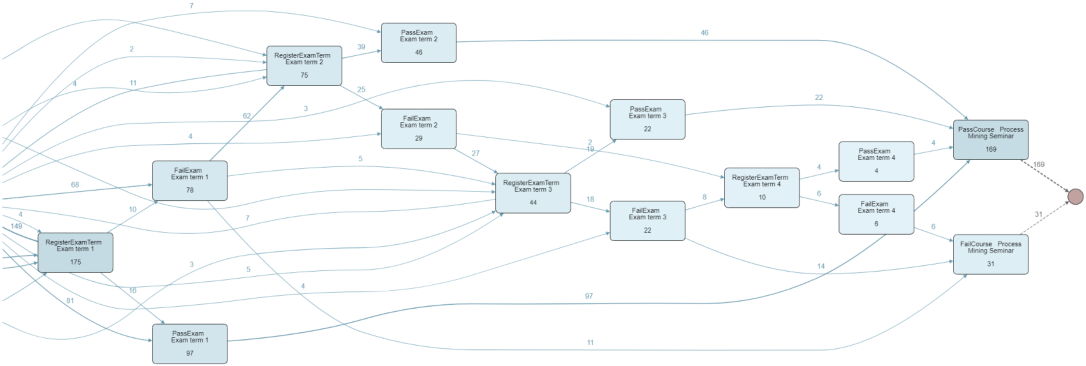
Windows Processes

- We proposed the technique for masquerader traitor detection.
- We performed a case study where we evaluated the process models.

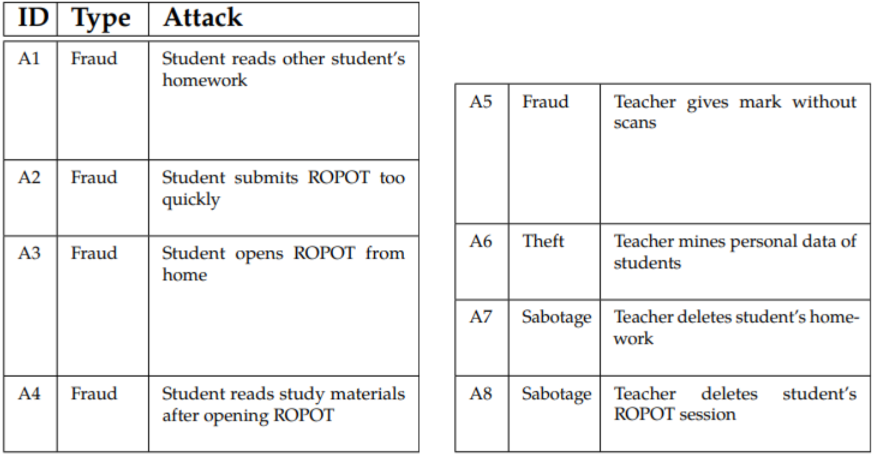


Information Systems

- We proposed the framework for insider attack detection using process mining.
- We performed a case study where we evaluated the framework.

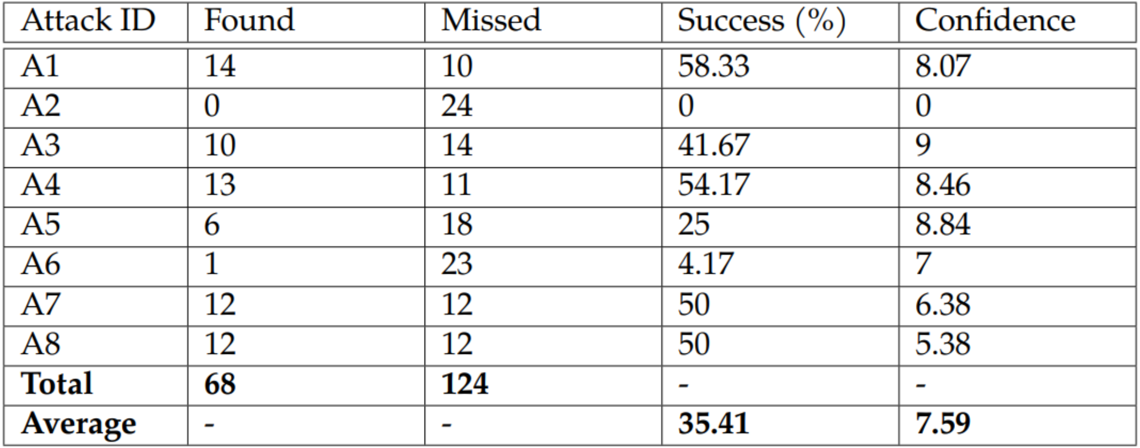


Information Systems



ID	Type	Attack
A1	Fraud	Student reads other student's homework
A2	Fraud	Student submits ROPOT too quickly
A3	Fraud	Student opens ROPOT from home
A4	Fraud	Student reads study materials after opening ROPOT
A5	Fraud	Teacher gives mark without scans
A6	Theft	Teacher mines personal data of students
A7	Sabotage	Teacher deletes student's homework
A8	Sabotage	Teacher deletes student's ROPOT session

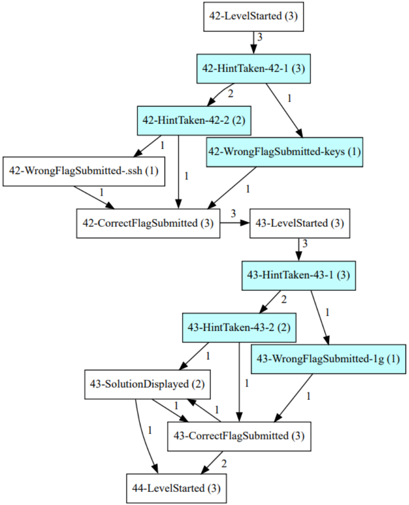
Information Systems



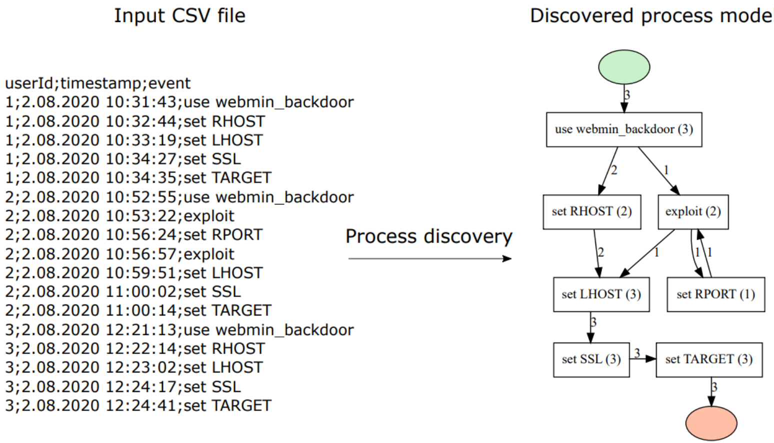
Attack ID	Found	Missed	Success (%)	Confidence
A1	14	10	58.33	8.07
A2	0	24	0	0
A3	10	14	41.67	9
A4	13	11	54.17	8.46
A5	6	18	25	8.84
A6	1	23	4.17	7
A7	12	12	50	6.38
A8	12	12	50	5.38
Total	**68**	**124**	-	-
Average	-	-	**35.41**	**7.59**

Cybersecurity Training Session Analysis

- Goal is to identify the issues with the training design.
- Collaboration with KYPO.



Cybersecurity Training Session Analysis



```
userId;timestamp;event
1;2.08.2020 10:31:43;use webmin_backdoor
1;2.08.2020 10:32:44;set RHOST
1;2.08.2020 10:33:19;set LHOST
1;2.08.2020 10:34:27;set SSL
1;2.08.2020 10:34:35;set TARGET
2;2.08.2020 10:52:55;use webmin_backdoor
2;2.08.2020 10:53:22;exploit
2;2.08.2020 10:56:24;set RPORT
2;2.08.2020 10:56:57;exploit
2;2.08.2020 10:59:51;set LHOST
2;2.08.2020 11:00:02;set SSL
2;2.08.2020 11:00:14;set TARGET
3;2.08.2020 12:21:13;use webmin_backdoor
3;2.08.2020 12:22:14;set RHOST
3;2.08.2020 12:23:02;set LHOST
3;2.08.2020 12:24:17;set SSL
3;2.08.2020 12:24:41;set TARGET
```

Figure 1. Process discovery example of the hacking process using Metasploit command line history

Internet of Secure Things

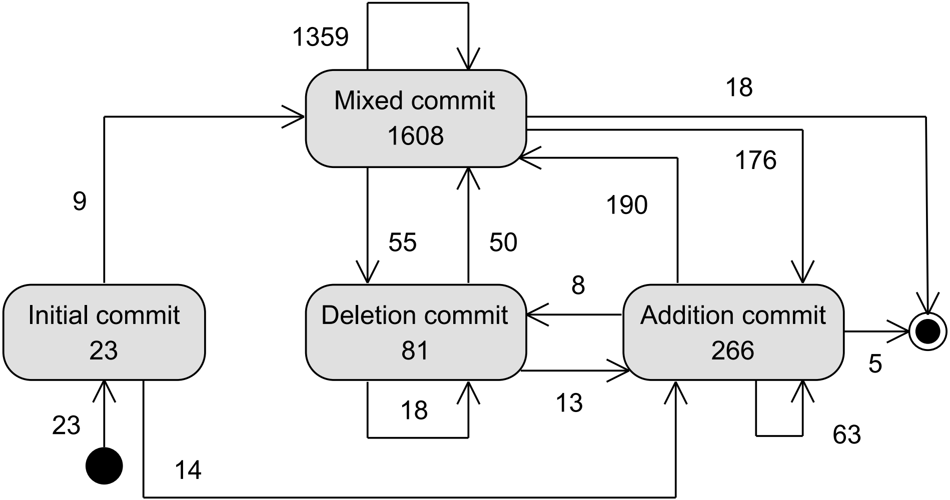
- Topics connected to Smart Cities.
  - Smart parking.
  - Vehicle sharing.
- Collaboration with University of Tartu.

Software engineering

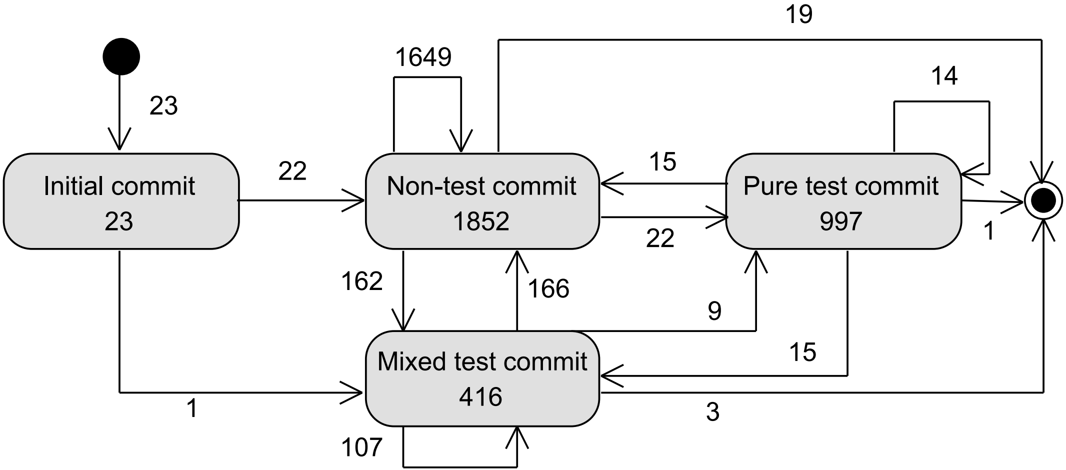
1. Git Logs
2. Application Usage: Steam Achievements
3. Software Development and Maintainability
4. Gen AI Usage in Programming
5. Software Testing

Git Logs

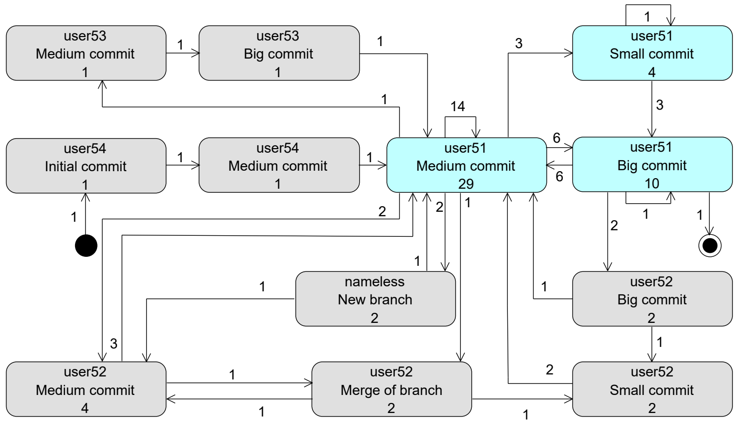
- Goal is to explore the usage of Git in team projects.



Git Logs


19
1649
14
23
Initial commit
23
22
Non-test commit
1852
15
Pure test commit
997
22
1
162
166
9
15
Mixed test commit
416
1
3
107

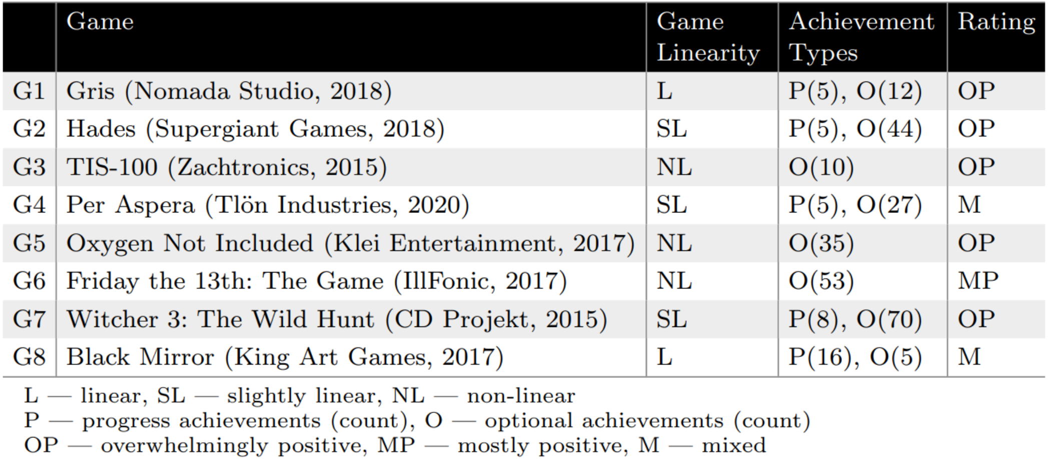
Git Logs



user53
Medium commit
1
1
user53
Big commit
1
1
3
user51
Small commit
4
1
1
14
3
user54
Initial commit
1
1
user54
Medium commit
1
1
user51
Medium commit
29
6
6
user51
Big commit
10
1
2
2
1
1
2
1
1
1
nameless
New branch
2
1
user52
Big commit
2
3
1
2
user52
Medium commit
4
1
1
user52
Merge of branch
2
1
user52
Small commit
2

Application Usage: Steam Achievements

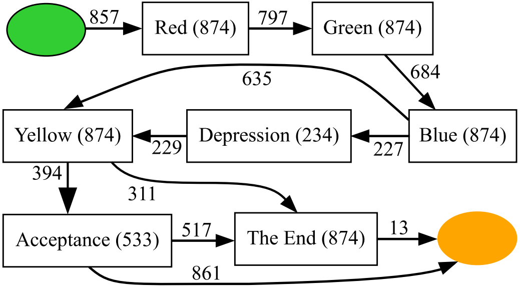
- Goal was to explore the usability of Steam data for a game playthrough analysis.

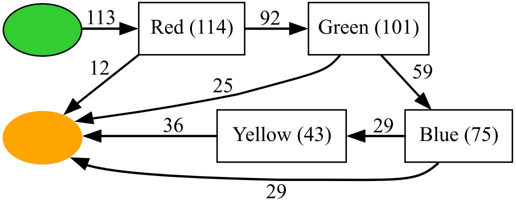


	Game	Game Linearity	Achievement Types	Rating
G1	Gris (Nomada Studio, 2018)	L	P(5), O(12)	OP
G2	Hades (Supergiant Games, 2018)	SL	P(5), O(44)	OP
G3	TIS-100 (Zachtronics, 2015)	NL	O(10)	OP
G4	Per Aspera (Tlön Industries, 2020)	SL	P(5), O(27)	M
G5	Oxygen Not Included (Klei Entertainment, 2017)	NL	O(35)	OP
G6	Friday the 13th: The Game (IllFonic, 2017)	NL	O(53)	MP
G7	Witcher 3: The Wild Hunt (CD Projekt, 2015)	SL	P(8), O(70)	OP
G8	Black Mirror (King Art Games, 2017)	L	P(16), O(5)	M

L — linear, SL — slightly linear, NL — non-linear
P — progress achievements (count), O — optional achievements (count)
OP — overwhelmingly positive, MP — mostly positive, M — mixed

Application Usage: Steam Achievements

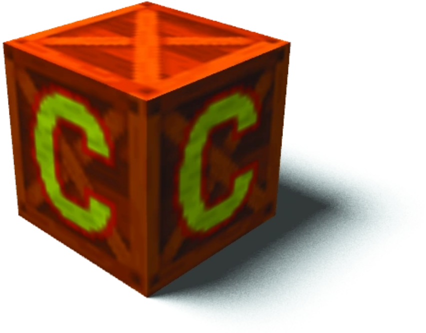




Other research

- Software Development and Maintainability
  - Process mining of Jira issues.
  - Collaboration with SAP Signavio.
- Gen AI Usage in Programming
  - Process mining of text prompts.
  - Collaboration with NLP lab.
- Software Testing
  - Process mining of application testing.

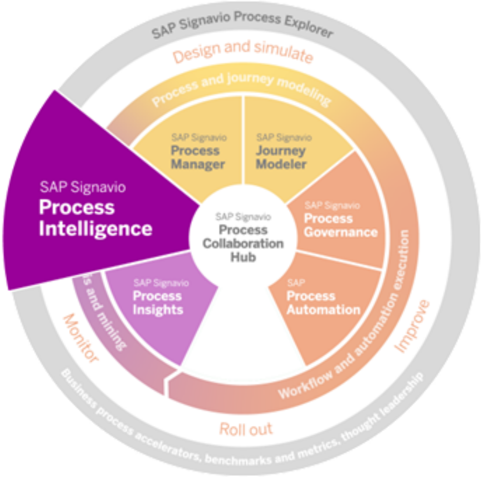
Checkpoint



SAP Signavio

- Core in Berlin.
- Acquired by SAP 3 years ago.
- 2 years in Brno (30 people).

SAP Signavio – Process Mining


SAP Signavio Process Explorer
Design and simulate
Process and journey modeling
SAP Signavio Process Manager
SAP Signavio Journey Modeler
SAP Signavio Process Governance
SAP Process Automation
SAP Signavio Process Collaboration Hub
SAP Signavio Process Intelligence
SAP Signavio Process Insights
s and mining
Monitor
Workflow and automation execution
Improve
Roll out
Business process accelerators, benchmarks and metrics, thought leadership

Blue Sky Team

- The initiative started a year ago.
- In 2023 they had:
  - 15+ Scientific papers,
  - 10+ BA/MA students,
  - 5 PhD students, 8 visiting researchers,
  - 5 university partners,
  - 25+ academic co-authors.
- The team consists of:
  - Thesis/Working students,
  - PhD students,
  - Visiting researchers.

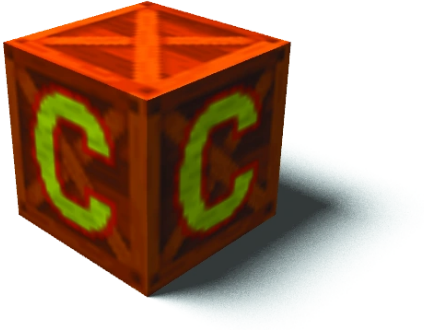
Focus topics

- BPI & Generative AI
- Advanced Conformance Checking
- Foundations of Process Querying
- Recommender Systems

Research in Brno

- I was the first researcher located in Czech republic.
- My topic was the Concept drift detection using process metrics.
  - Applied for software development and maintainability process using Jira dataset.
- Located with the local Brno team focused on the development.
- Very enjoyable cooperation and support for 6 months.
- Open for further collaboration in teaching and research with our university.

Checkpoint



Conclusion

1. Process Mining
2. Process Mining in Lasaris
   - Audit logs
   - Manufacturing
   - Windows Processes
   - Information Systems
   - Cybersecurity Training Sessions
   - Internet of Secure Things
   - Git Logs
   - Steam Achievements
   - Jira
   - GenAI
   - Testing
3. Process Mining in SAP Signavio

Call to action!

- Contact me: macak@mail.muni.cz
- Register PV226
- Find your consultant
- Let’s go to lunch