

Extending Process Logs with Events from Supplementary Sources

Felix Mannhardt (@fmannhardt),
Massimiliano de Leoni, Hajo A. Reijers

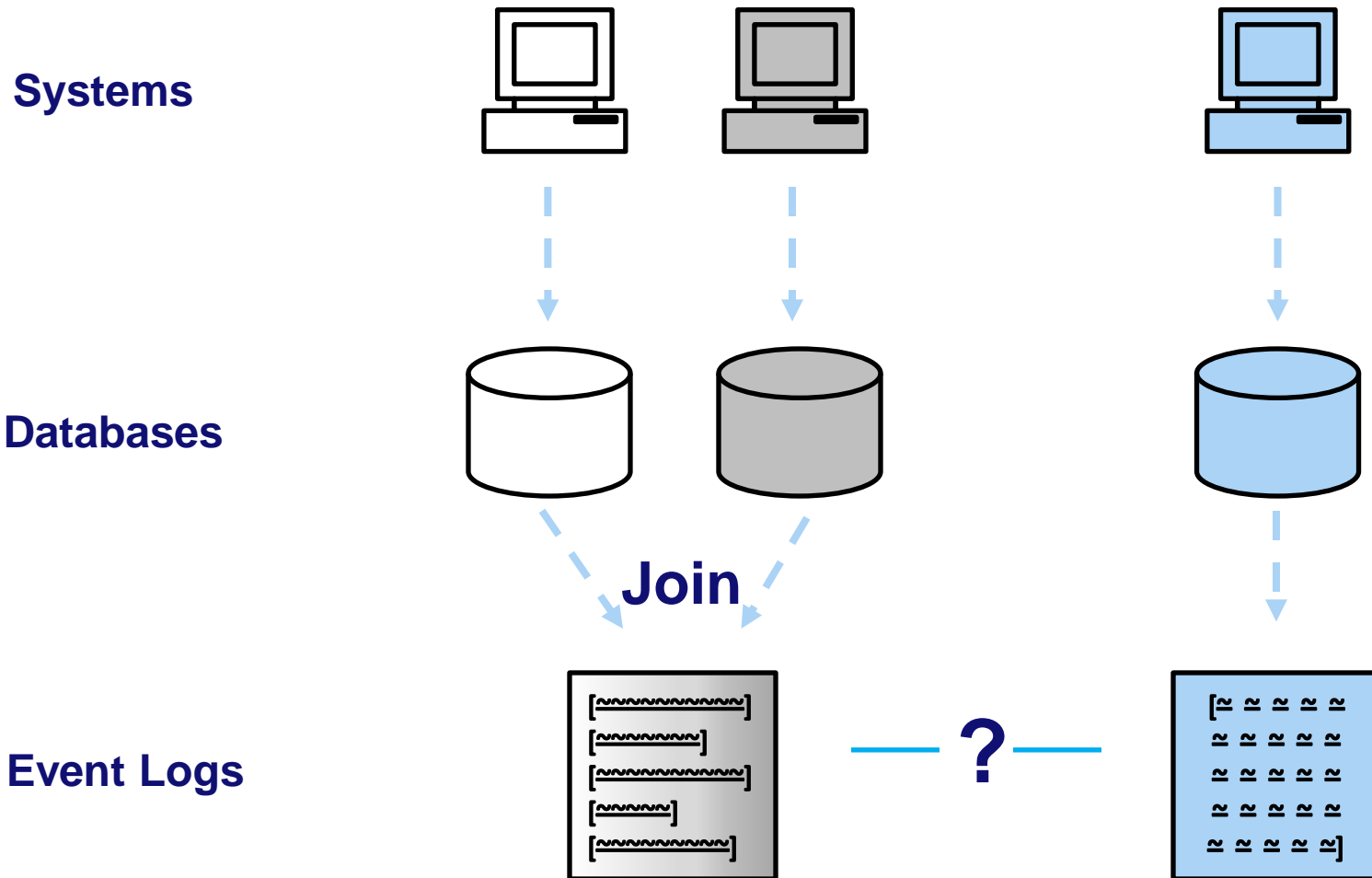


TU / **e**

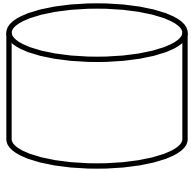
Technische Universiteit
Eindhoven
University of Technology

Where innovation starts

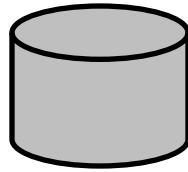
Heterogeneous Information Systems



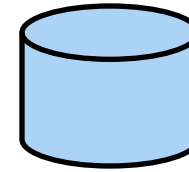
Relating Events to Cases



PK	Event
1	Create Fine
1	Payment
2	Create Fine



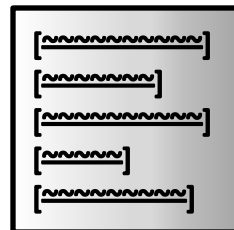
PK	FK	Event
22	1	Send Fine
23	2	Send Fine



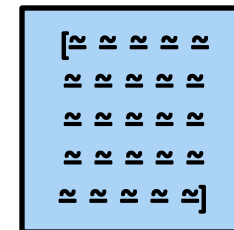
Event	Amount
Collection	120
Collection	60



Join

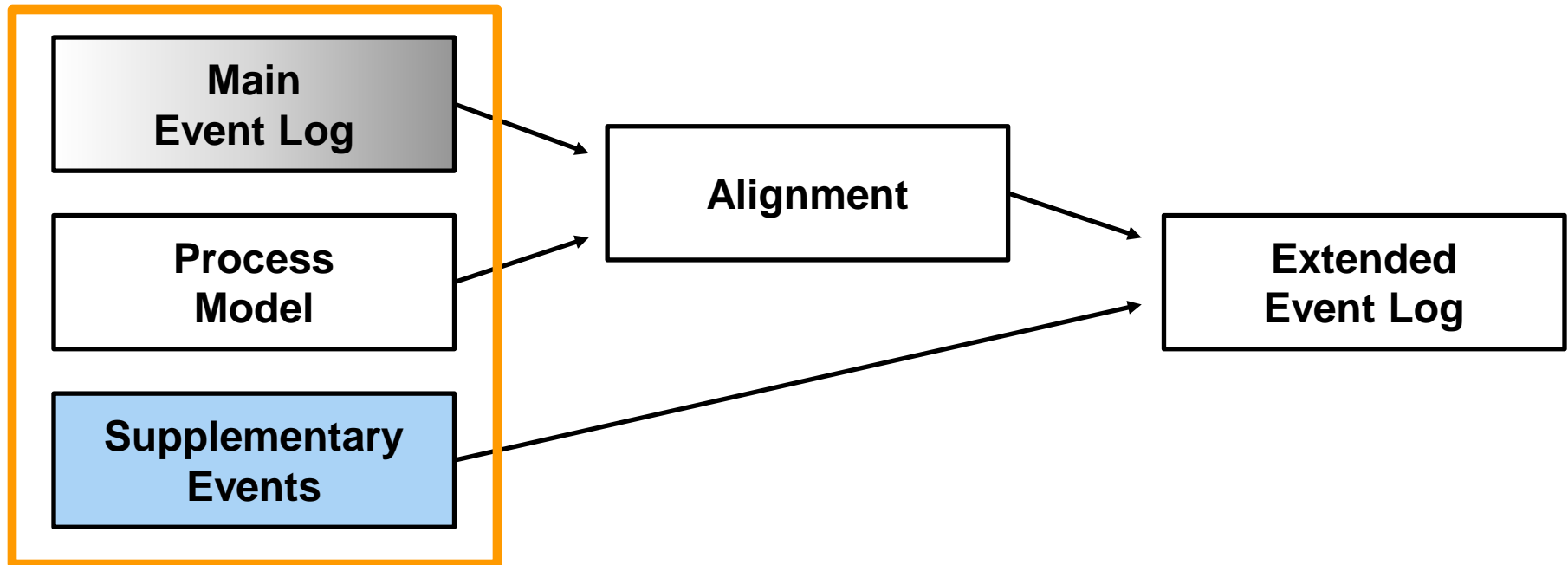


Main Event Log



Supplementary Events

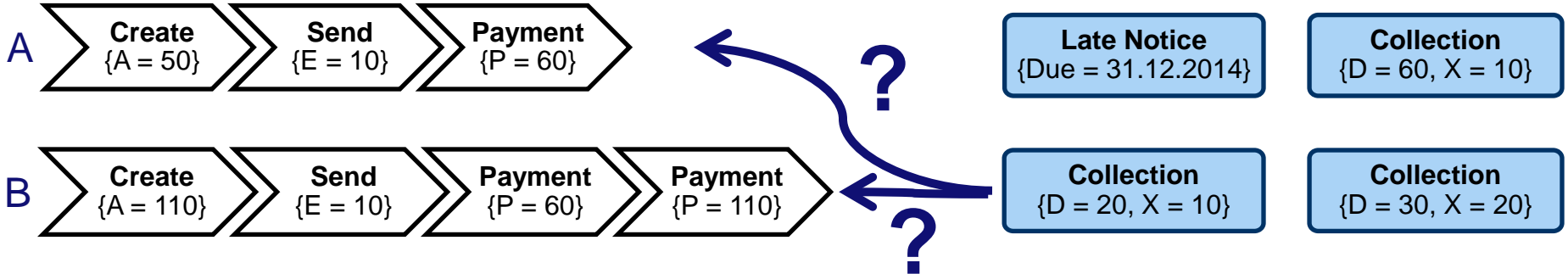
Overview



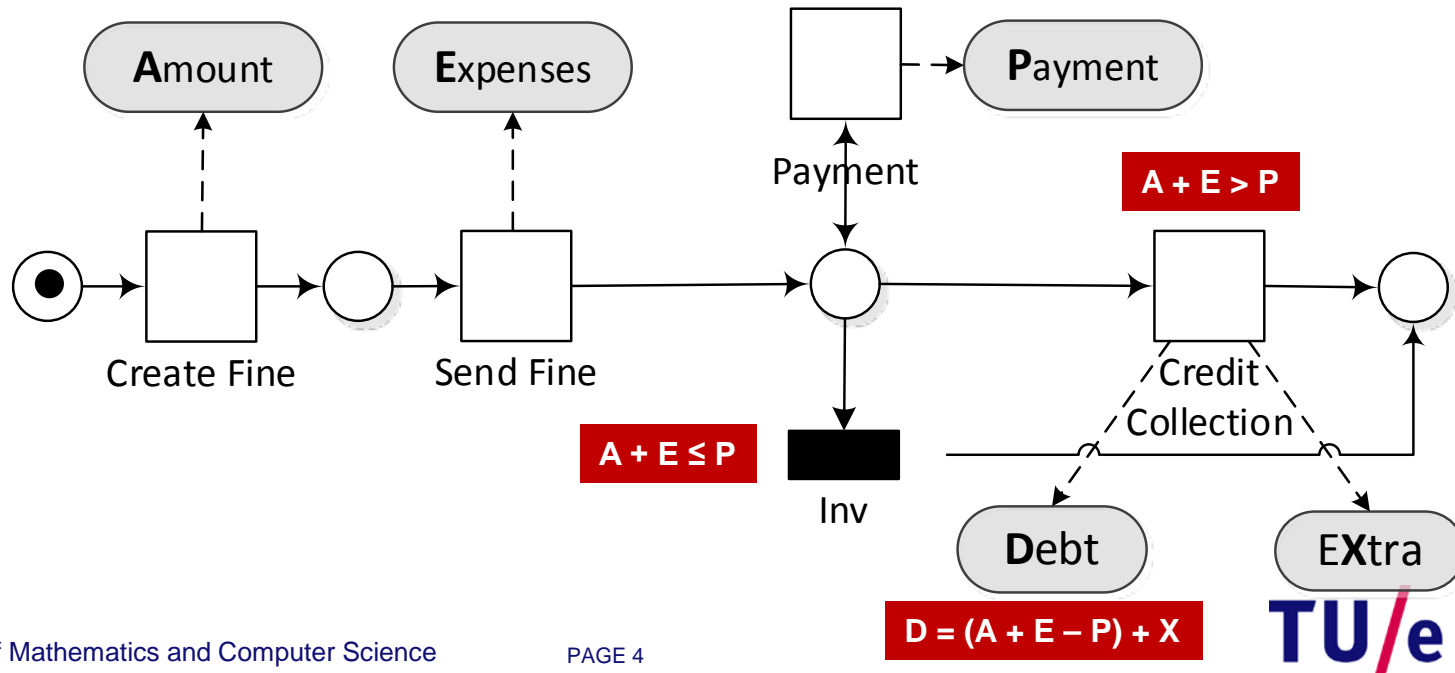
Ingredients

Main Event Log

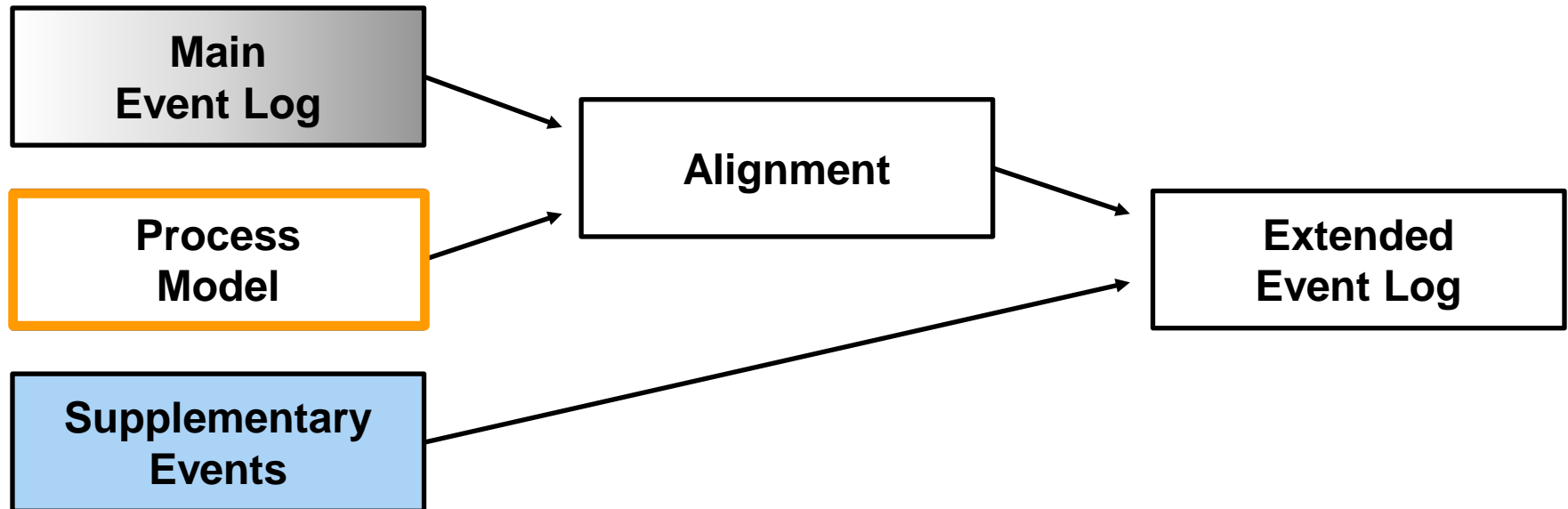
Supplementary Events



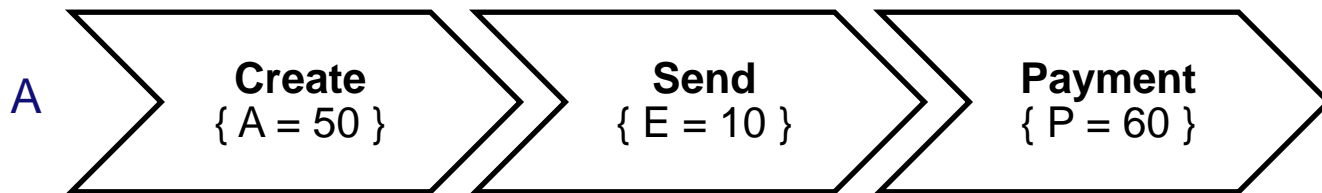
Process Model



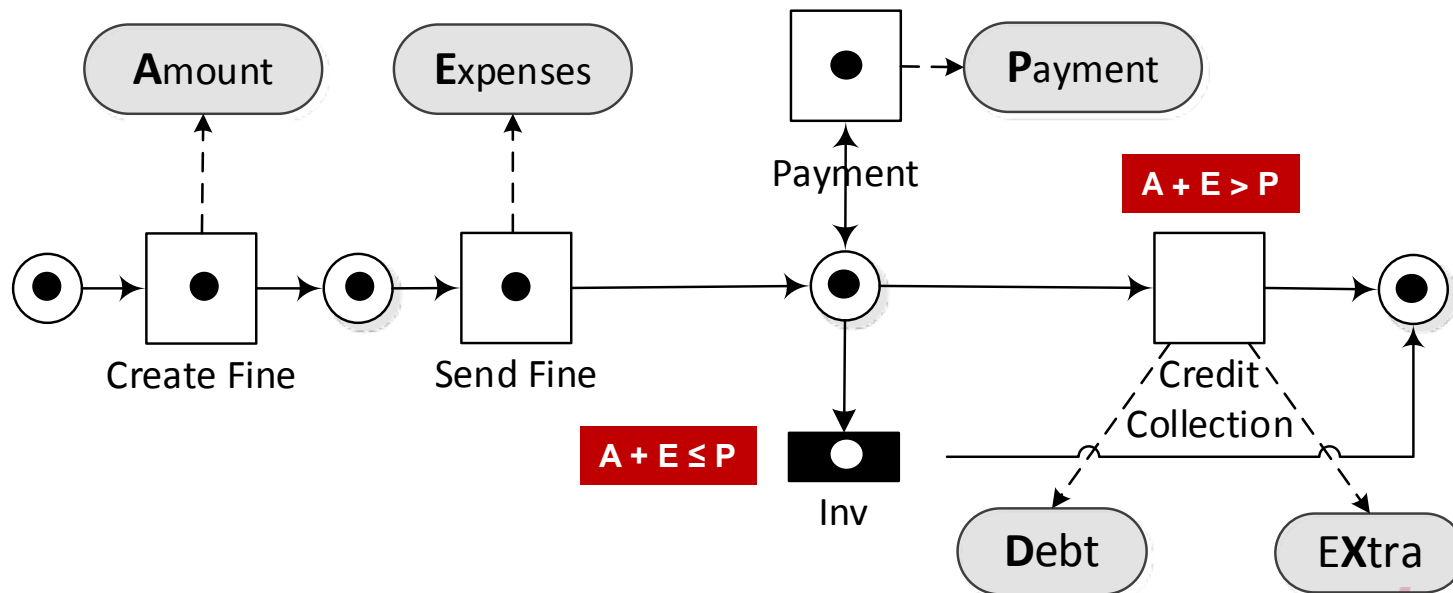
Overview: Process Model



Process Model: Data Petri Nets

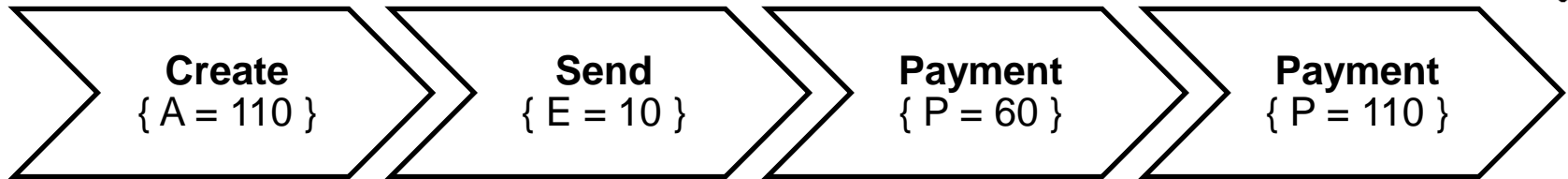


Process Model

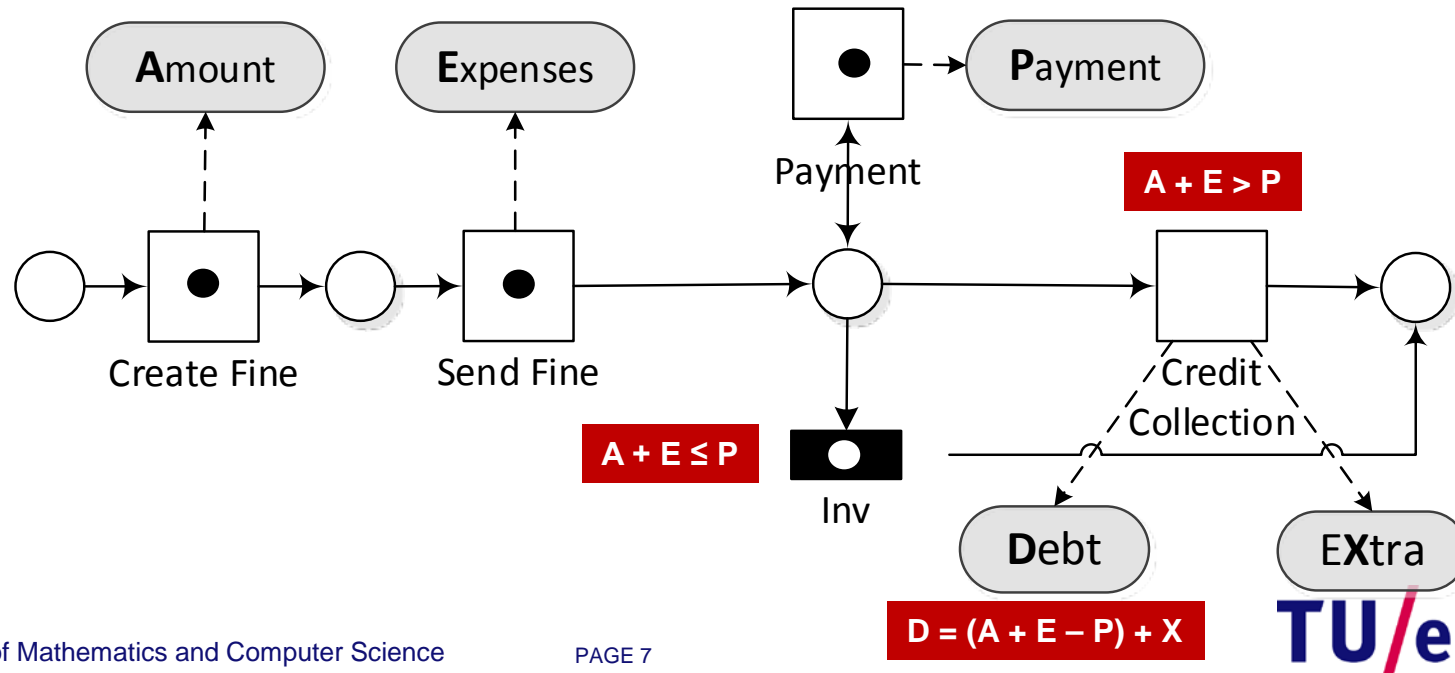


Data Petri Nets: Non-Conforming Trace

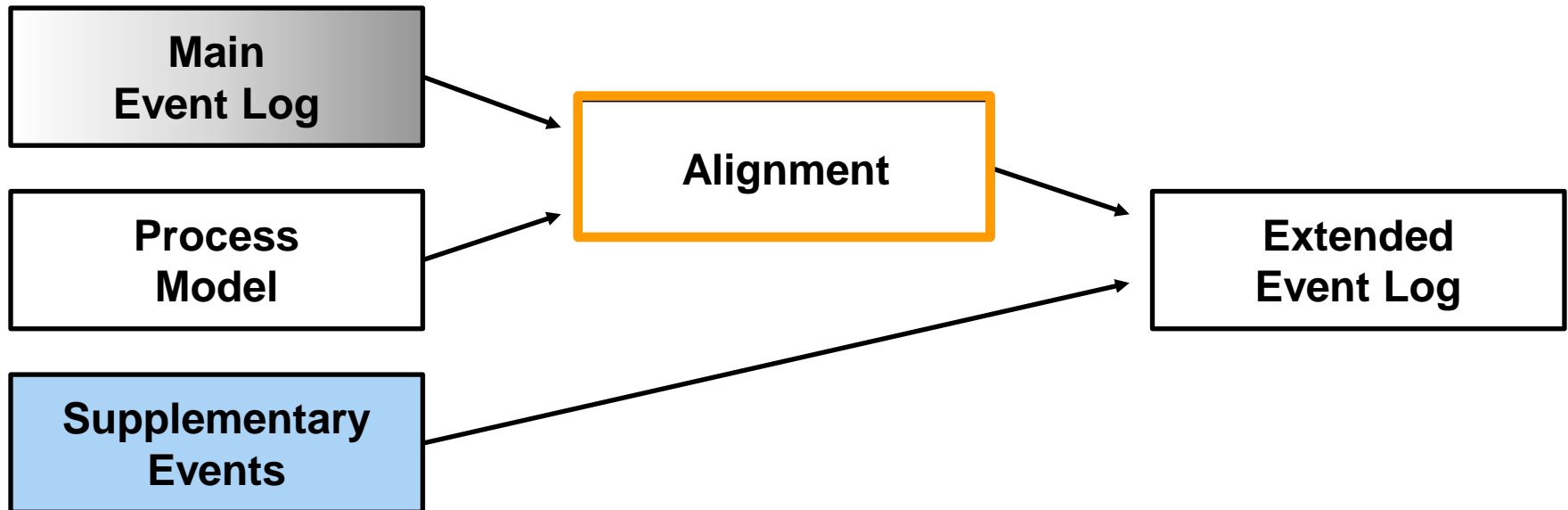
B



Process Model

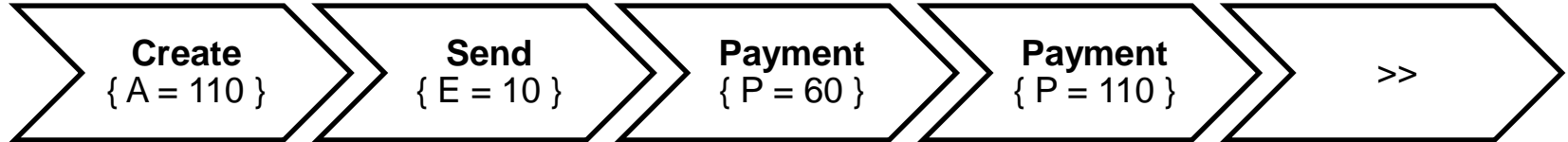


Overview: Alignment

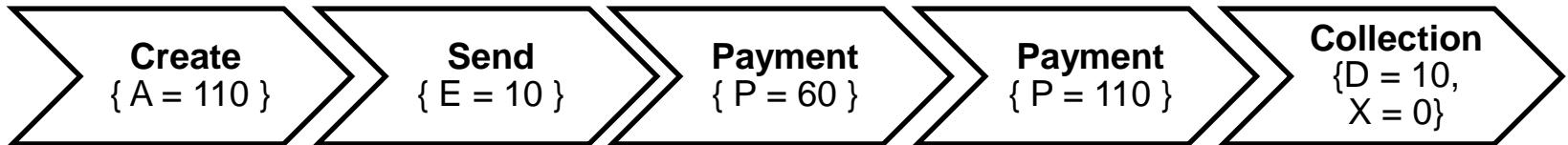


Alignment: Identify Missing Events

Log

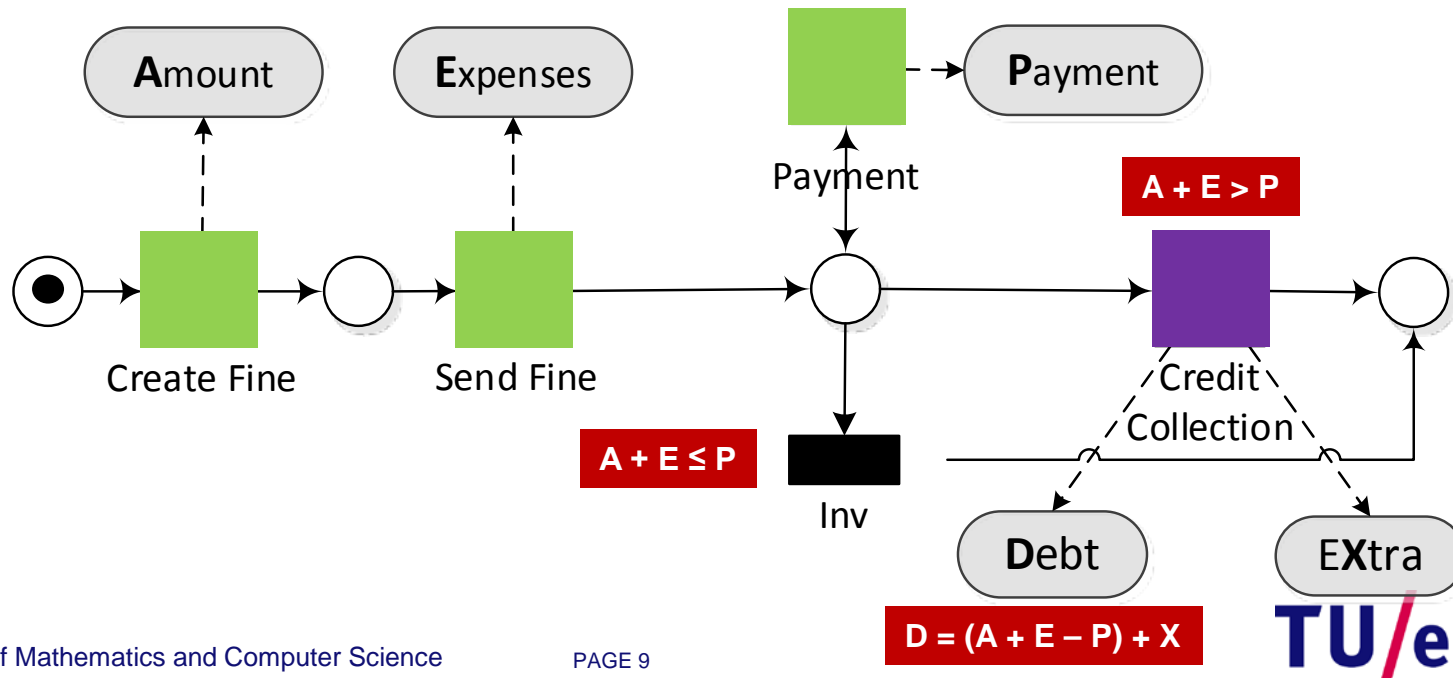


Process



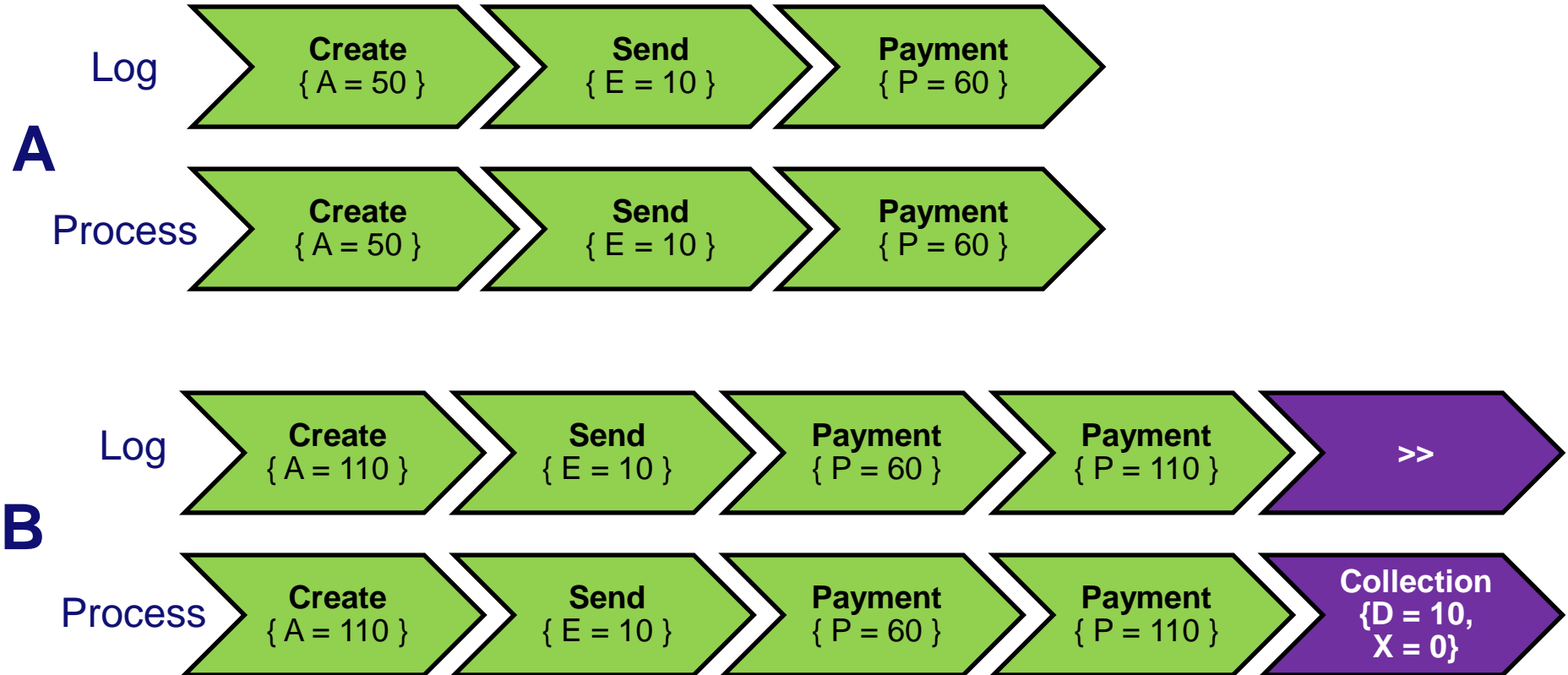
Missing Event

Process Model

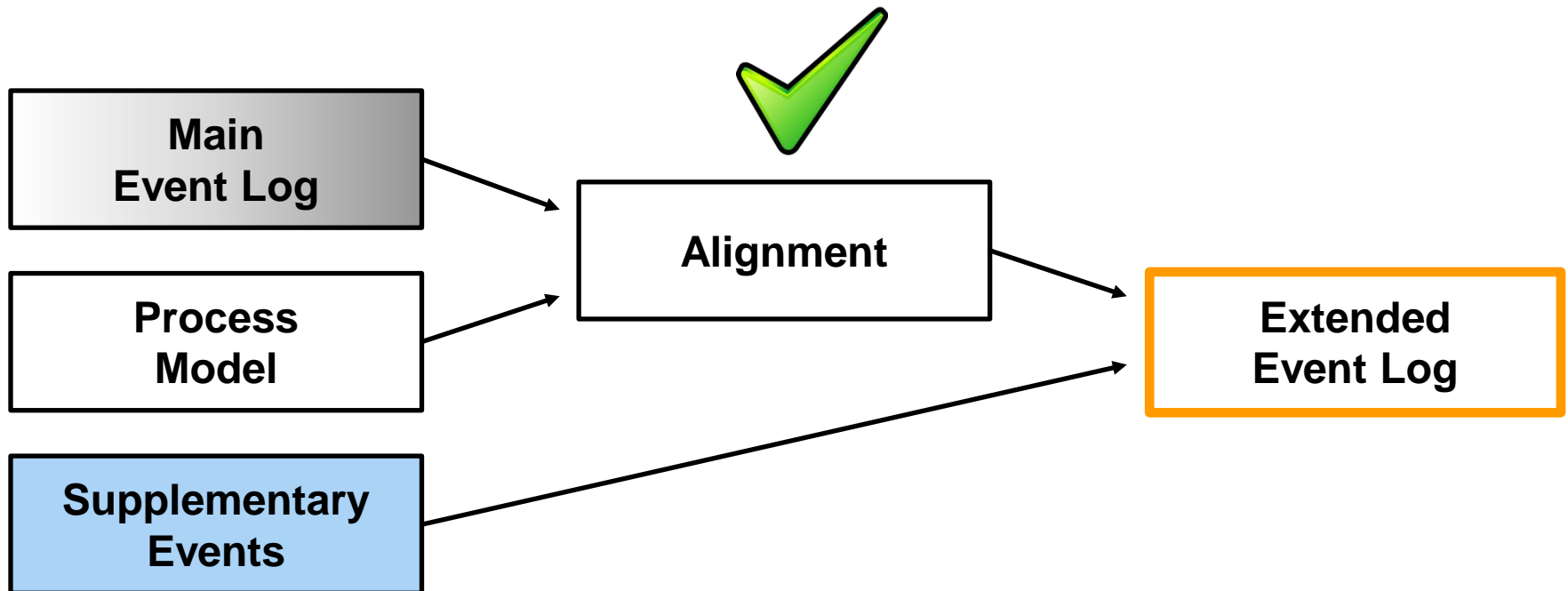


Alignment with Missing Events

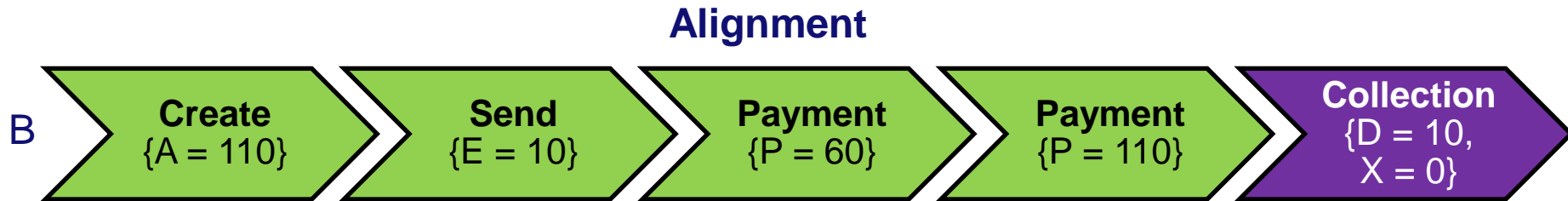
Alignment of Main Event Log



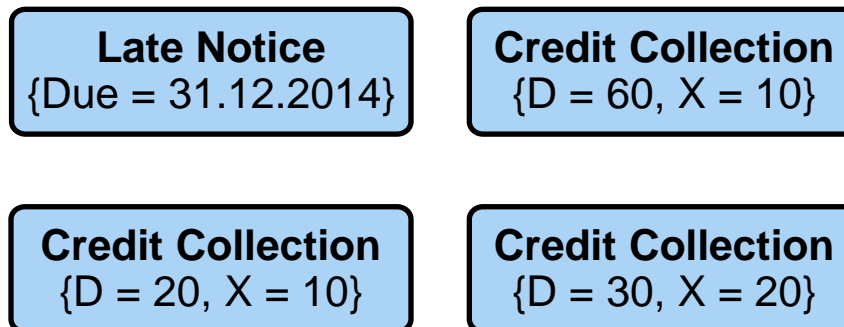
Overview: Extending the Event Log



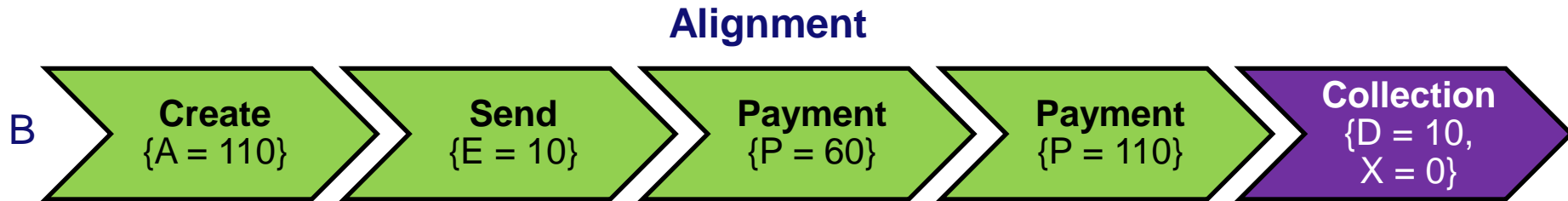
Extending the Event Log



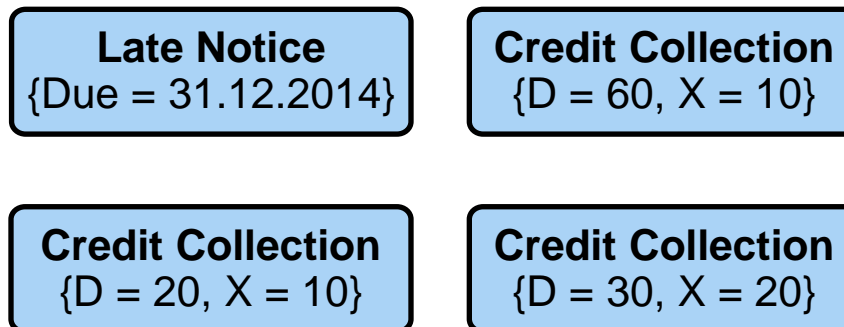
Supplementary Events



Events that relate to the Missing Activity

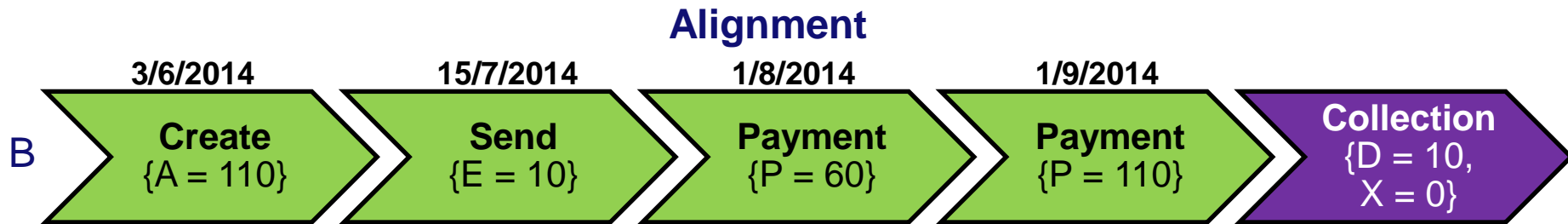


Event Candidates

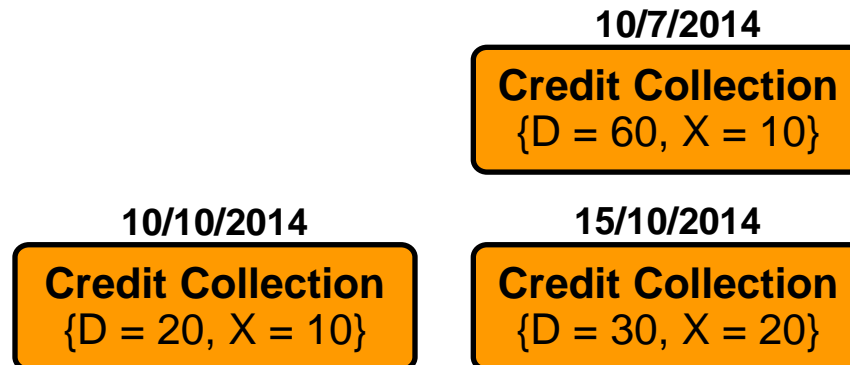


Events relating the
correct missing activity

Events with Compatible Timestamps



Supplementary Events

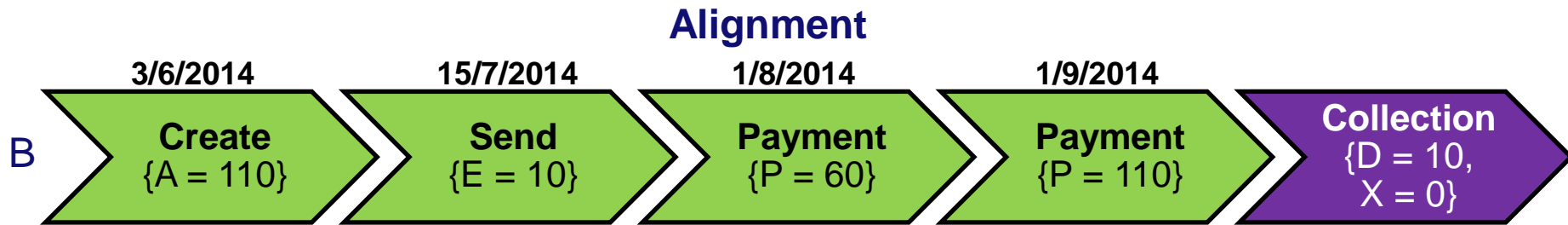


Events relating the correct missing activity



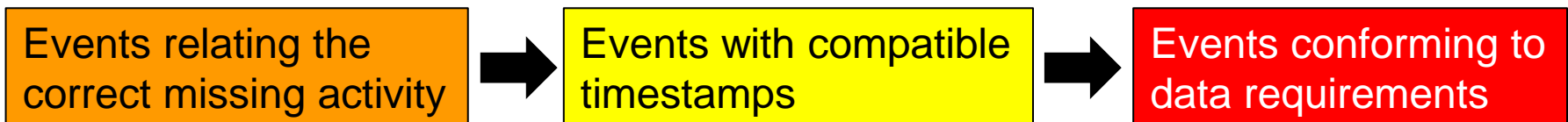
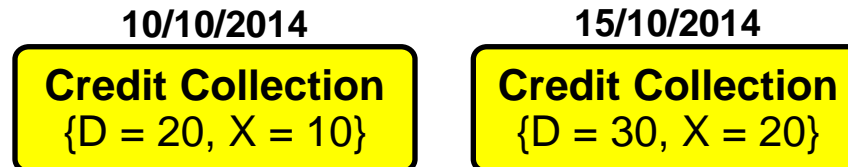
Events with compatible timestamps

Events that Conform to Data Requirements

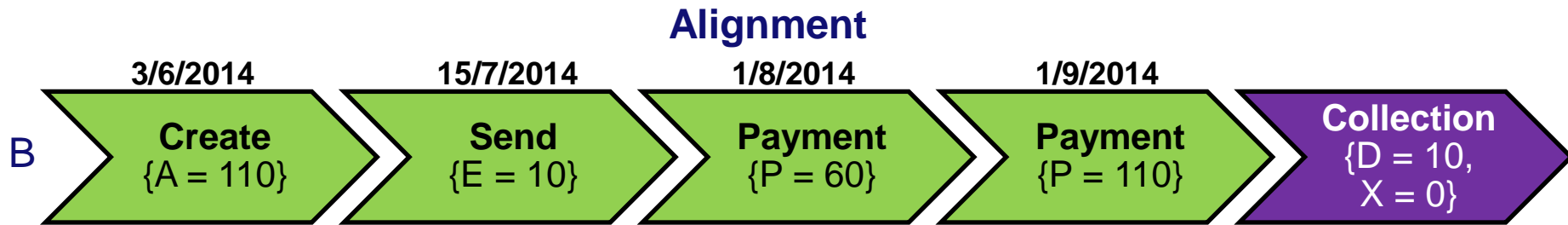


Event Candidates

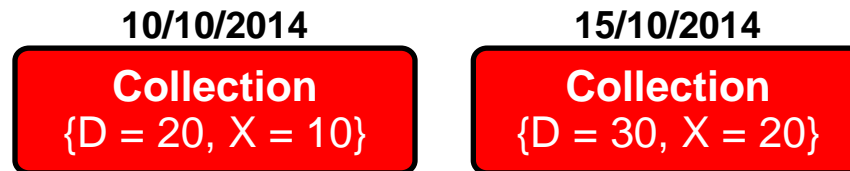
$$D \stackrel{\text{def}}{=} A + E - P + X = 10 + X$$



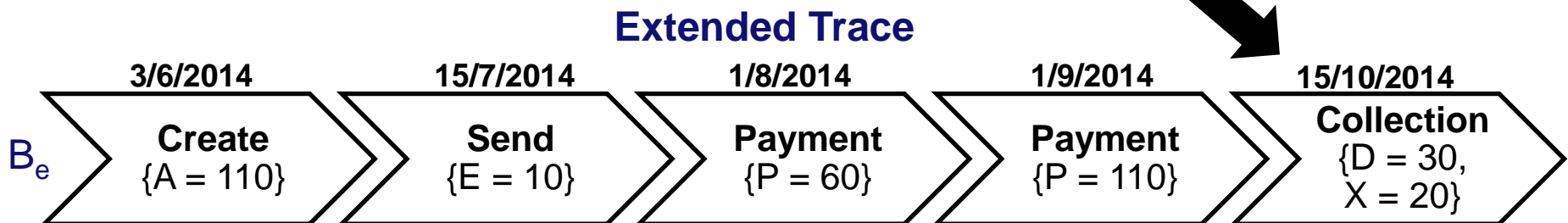
Choosing the “correct” Event



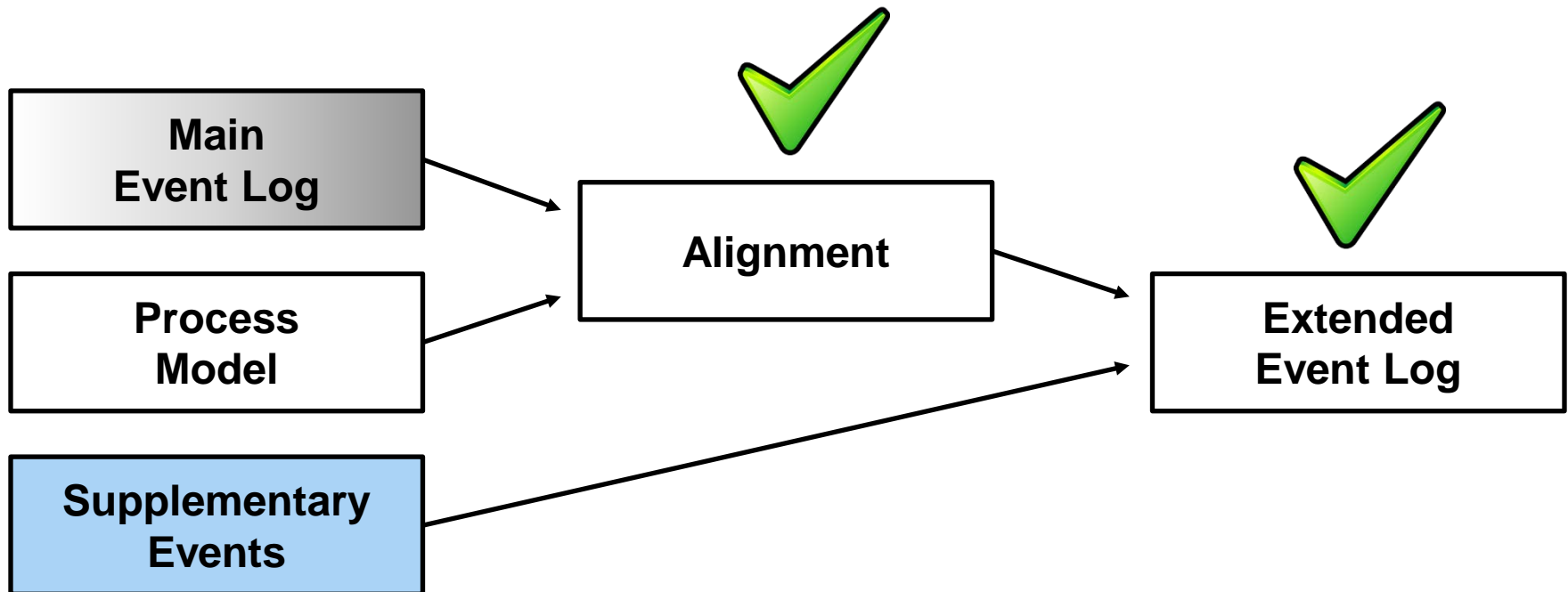
Matching Events



Random / Time-based
etc..

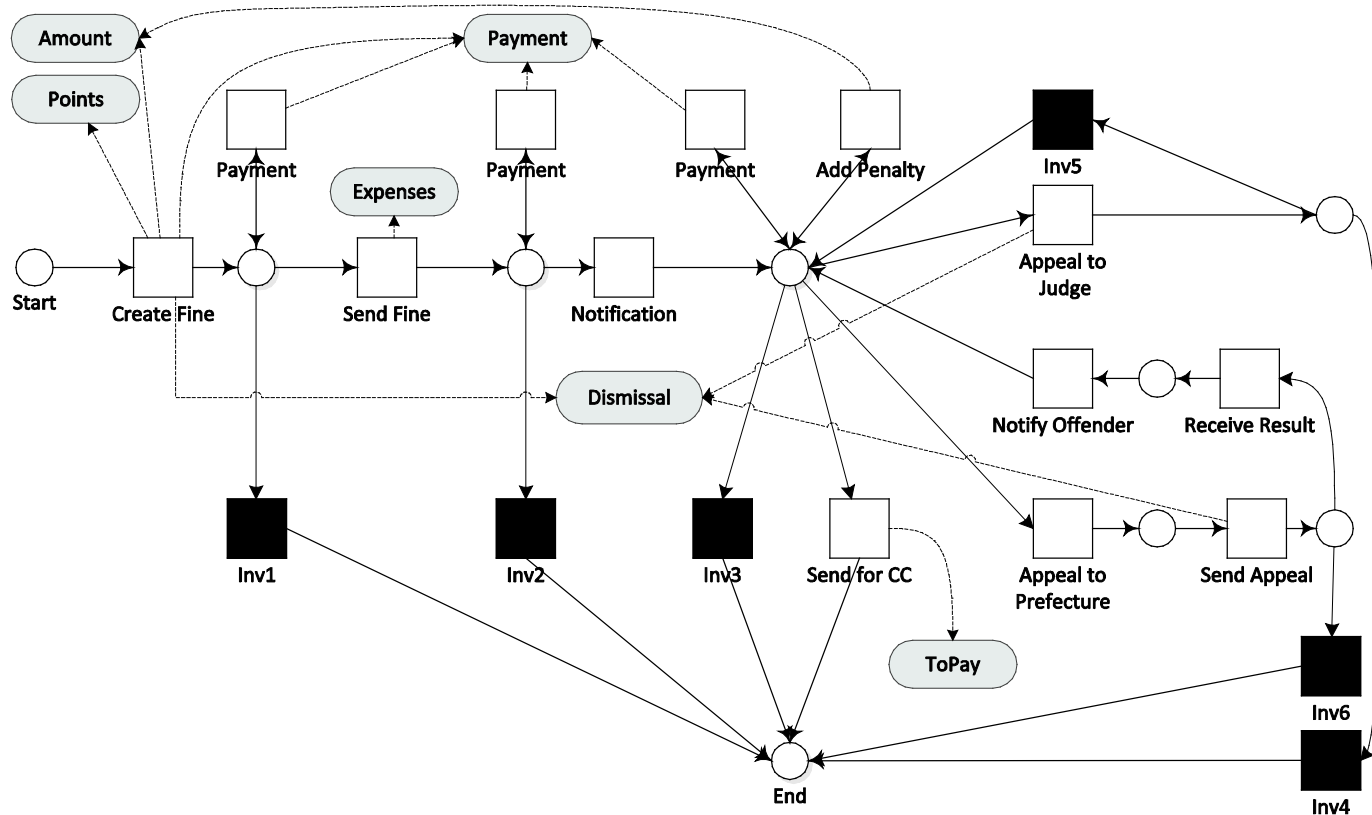


Overview



Evaluation

- Real-life dataset of a Road Traffic Fines process
- ~150.000 traces with ~ 540.000 events



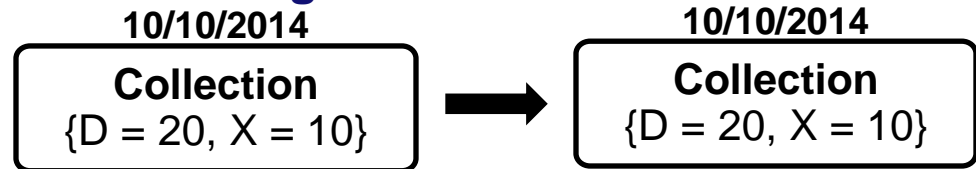
Evaluation

- Real-life dataset of a Road Traffic Fines process
- ~150.000 traces with ~ 540.000 events
- Two scenarios: For each, assume that an activity is done with a different system:
 - Send Fine
 - Send for Credit Collection
- Split the Log into a *Main Event Log* and *Supplementary Events*
- Apply our approach
- Check how events are correlated

Measures

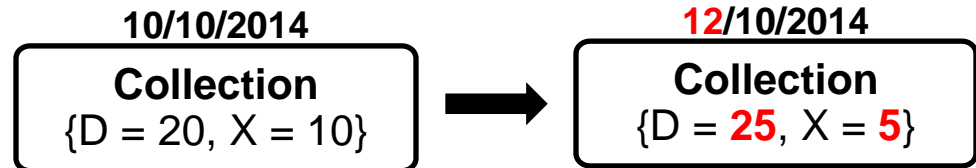
% Perfectly correlated →

Same as original trace



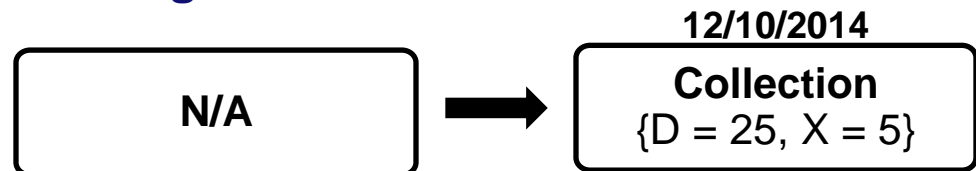
% Approx. correlated →

Mismatch in data or time



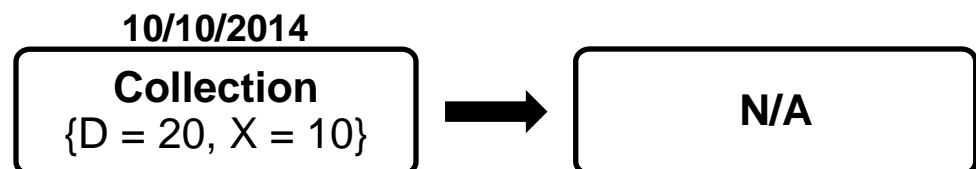
% Wrongly correlated →

Missing / added events



% Ignored →

Unmapped events



Evaluation

Scenario	C	l_M	% Perfectly	% Approx.	% Wrongly	% Ignored
A: Send Fine (Compliant)	33178	512	22.0	13.9	1.5	62.6
		1024	34.0	30.3	3.1	32.2
		2048	38.6	44.8	5.2	11.3
B: Send Fine	101950	512	5.1	5.1	0.4	89.5
		1024	11.6	15.5	0.8	72.1
		2048	14.1	31.5	1.7	52.8
C: Send for CC (Compliant)	28049	512	29.1	4.3	0	66.6
		1024	51.8	6.1	0	42.1
		2048	69.6	12.4	0	18.1
D: Send for CC	54232	512	20.8	3.6	3.8	71.7
		1024	29.9	4.7	5.6	59.8
		2048	40.3	6.9	7.3	45.5

Parameter to steer the reliability

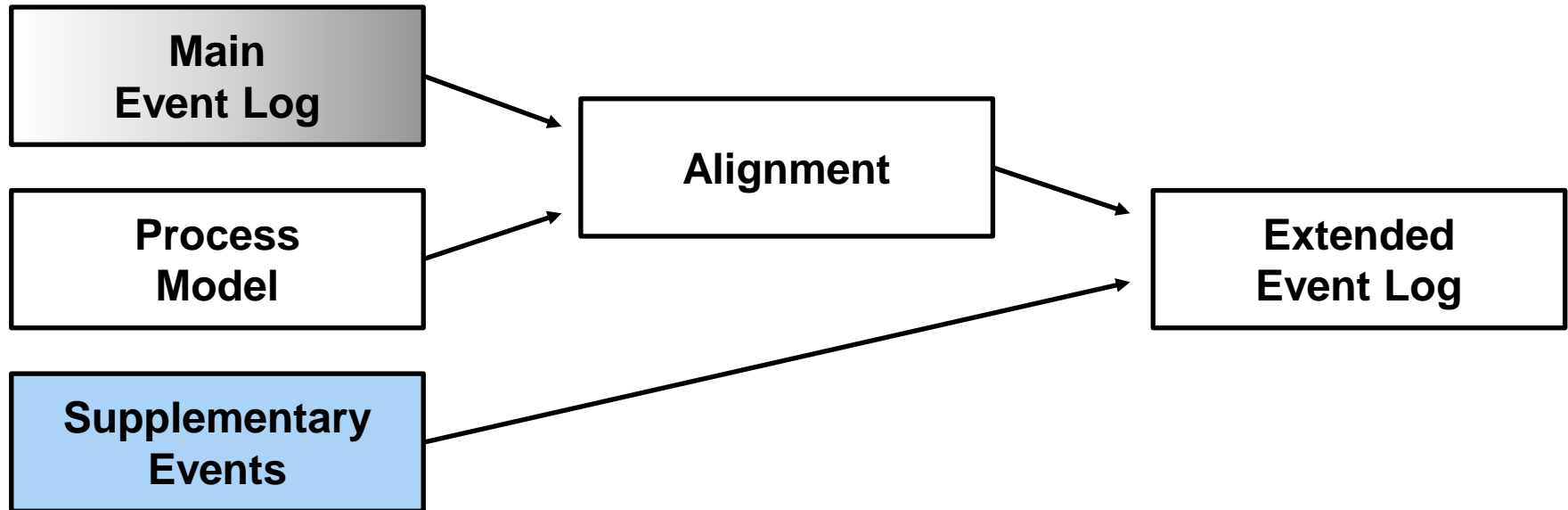
% Approx.: up to 80%

% Ignored too high

Future Work

- **Improve the evaluation**
 - ➔ **Multiple events missing per trace**
 - ➔ **Measures to compare wrt. existing work**
- **Find strategies for “leftover” non-mapped events**
- **How to use (estimated) activity durations:**
 - ➔ **Better time constraints**
- **Improve the prototype:**
 - ➔ ***LogEnhancement* package in ProM**

Extending Process Logs with Events from Supplementary Sources



- **Technique to extend an Event Log with Events**
- **Using an Alignment (with Data) to identify *Gaps***
- **Finding the most likely Event to fill these *Gaps***

Remarks? Questions?



Contact: f.mannhardt@tue.nl | [@fmannhardt](https://twitter.com/fmannhardt)

Image source: http://commons.wikimedia.org/wiki/File:Pictofigo_-_Idea.png