Aligning Event Logs and Process Models for Multi-perspective Conformance Checking: An Approach Based on ILP

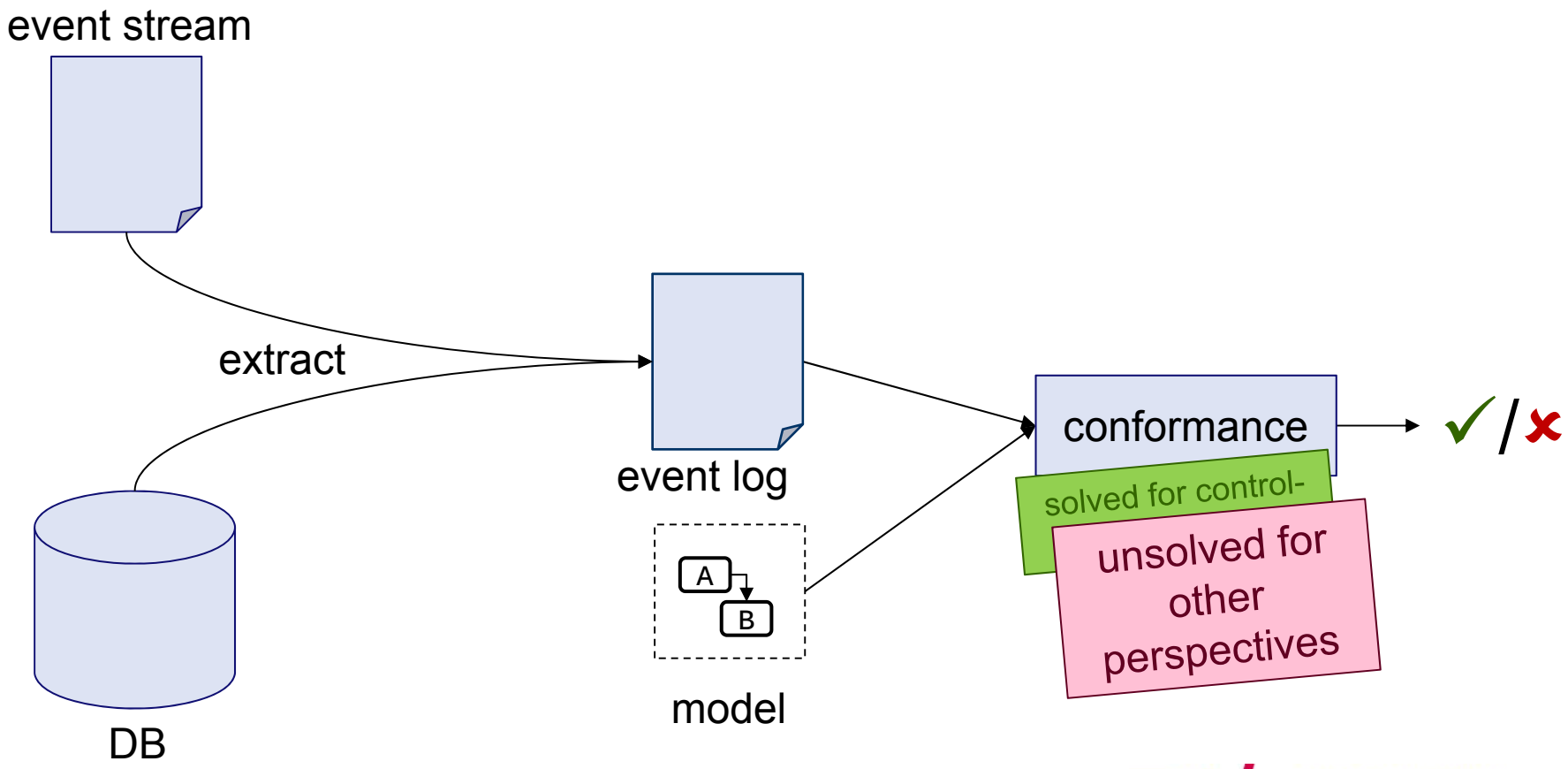
Massimiliano de Leoni

Wil M. P. van der Aalst

TU/e Technische Universiteit Eindhoven University of Technology

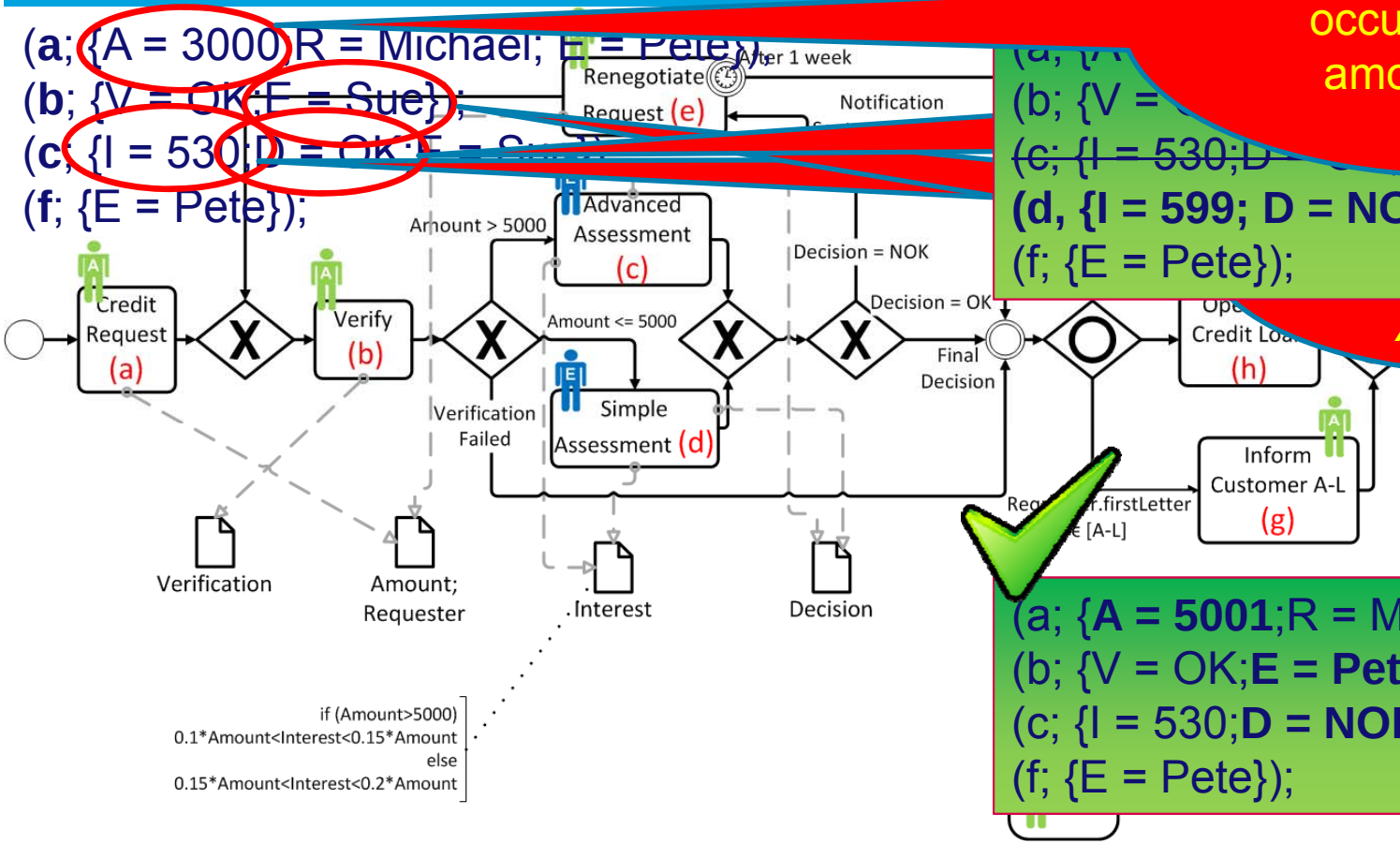
Where innovation starts

Conformance Checking



Example: A Credit Institute

(**a**; {A = 3000;R = Michael; E = Pete});
(**b**; {V = OK;E = Sue});
(**c**; {I = 530;D = OK;E = Sue});
(**f**; {E = Pete});



Activity *d* should have occurred, since amount<5000

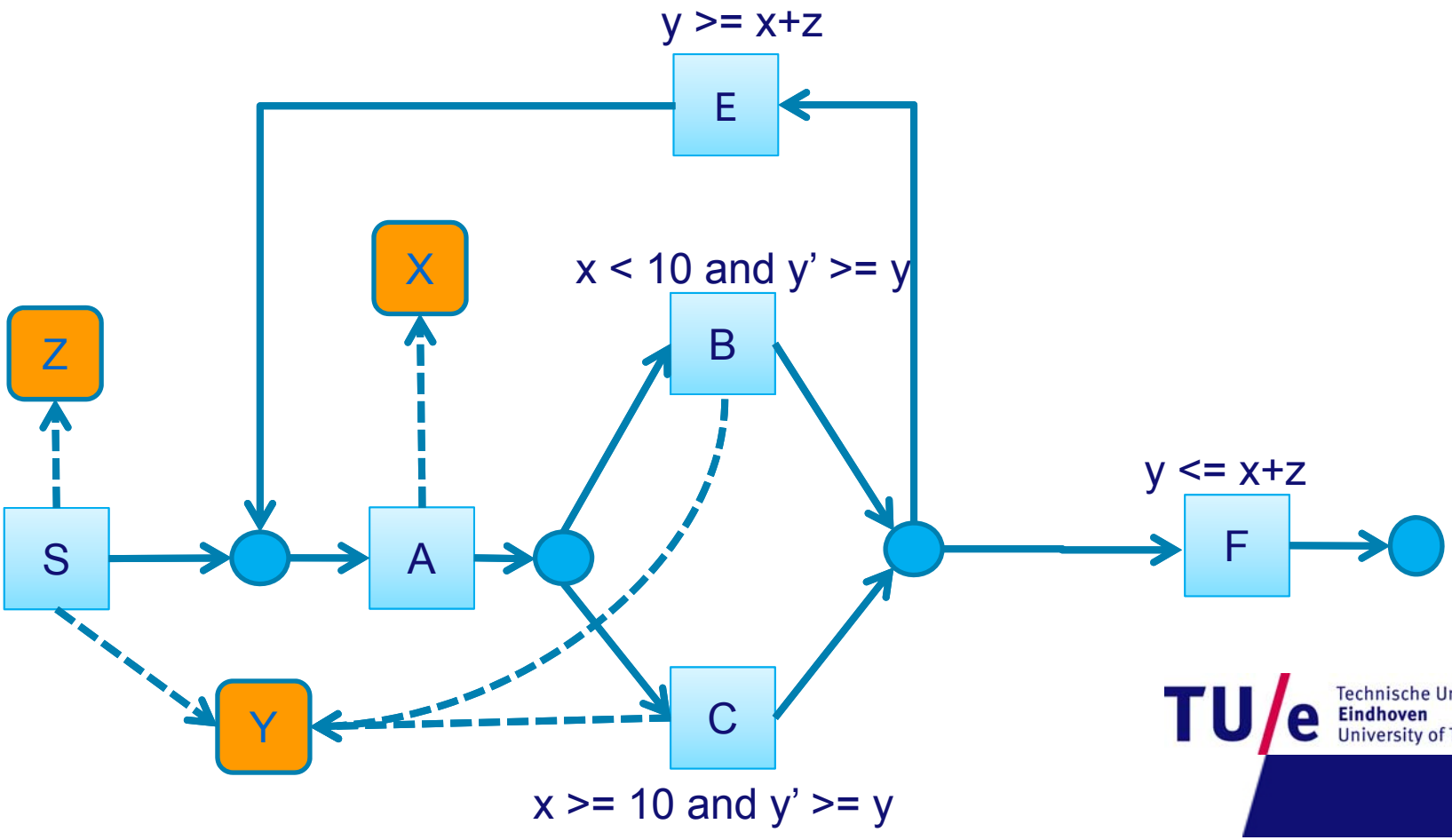
(a; {A = 3000;R = Michael; E = Pete});
(b; {V = OK;E = Sue});
(c; {I = 530;D = OK;E = Sue});
(d, {I = 599; D = NOK; E = Sue});
(f; {E = Pete});

Assistant

(a; {**A = 5001**;R = Michael; E = Pete});
(b; {V = OK;**E = Pete**});
(c; {I = 530;**D = NOK**;E = Sue});
(f; {E = Pete});

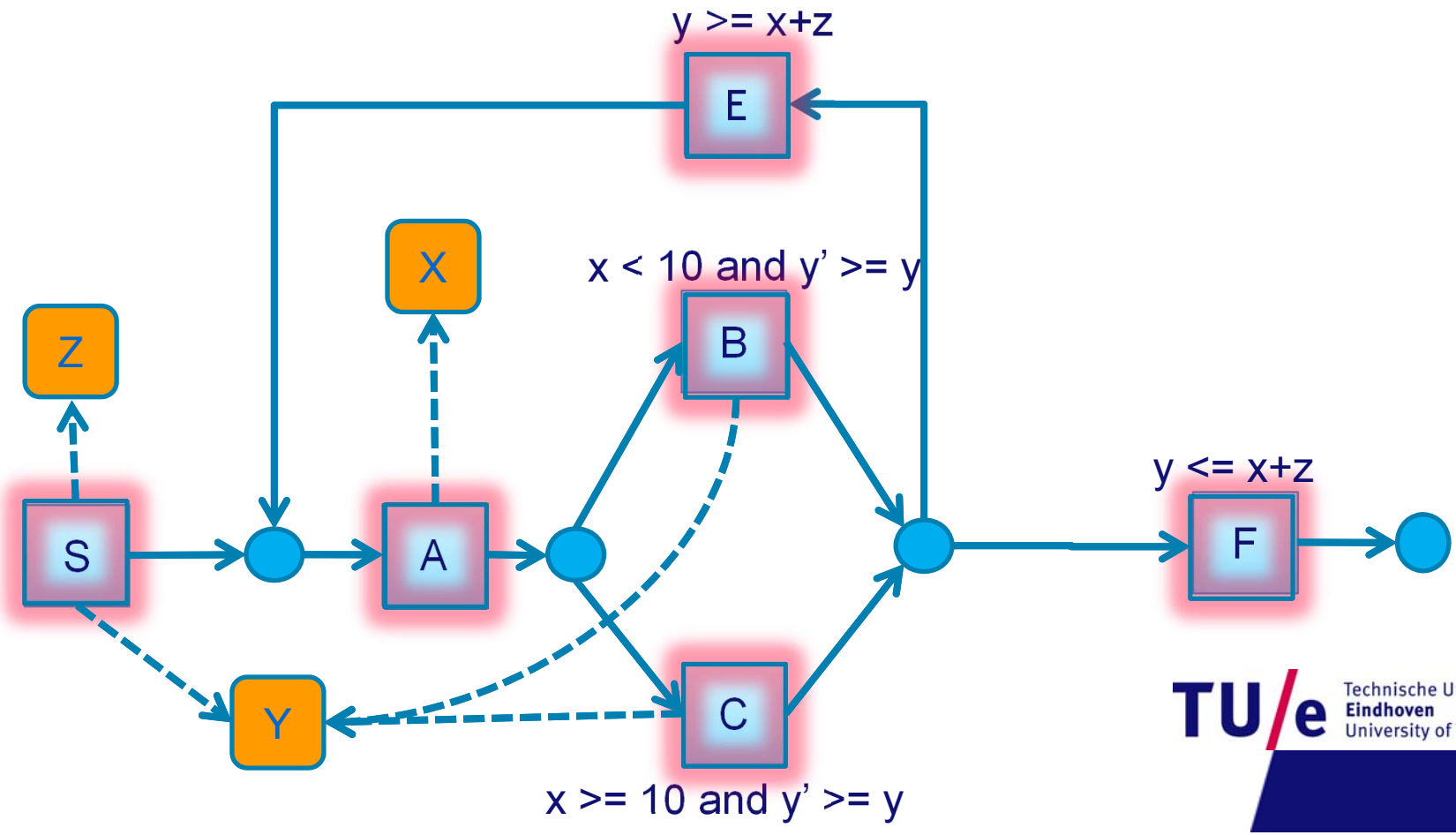
Technische Universiteit Eindhoven University of Technology

Petri Nets with Data



A trace without problems

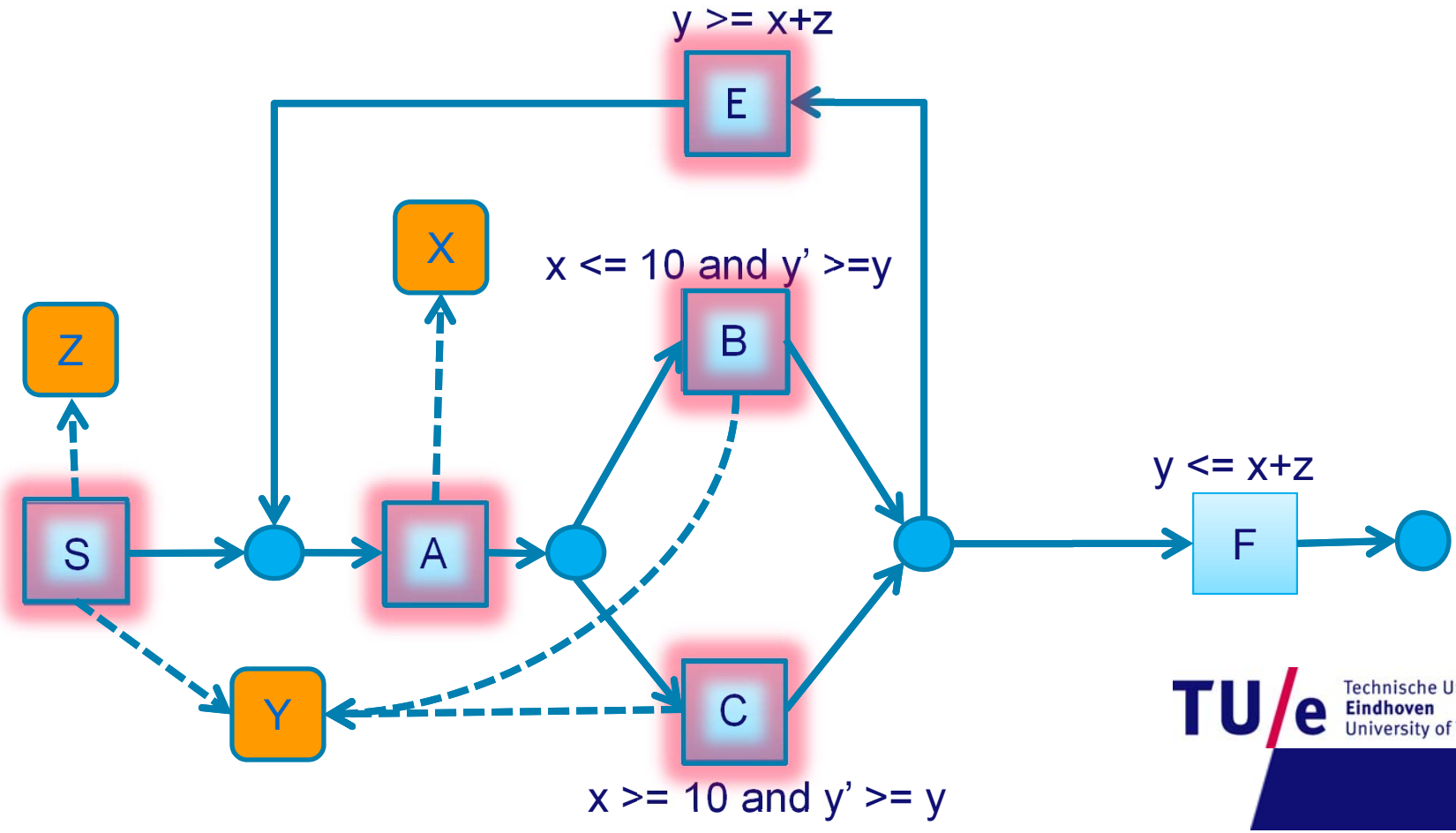
S {z=1,y=0} – A{x=10} – C{y=11} – E – A{x=10} – B{y=11} - F



A trace with problems

S {z=10,y=0} – A{x=1} – C{y=11} – E – A{x=3} – B{y=13} >>

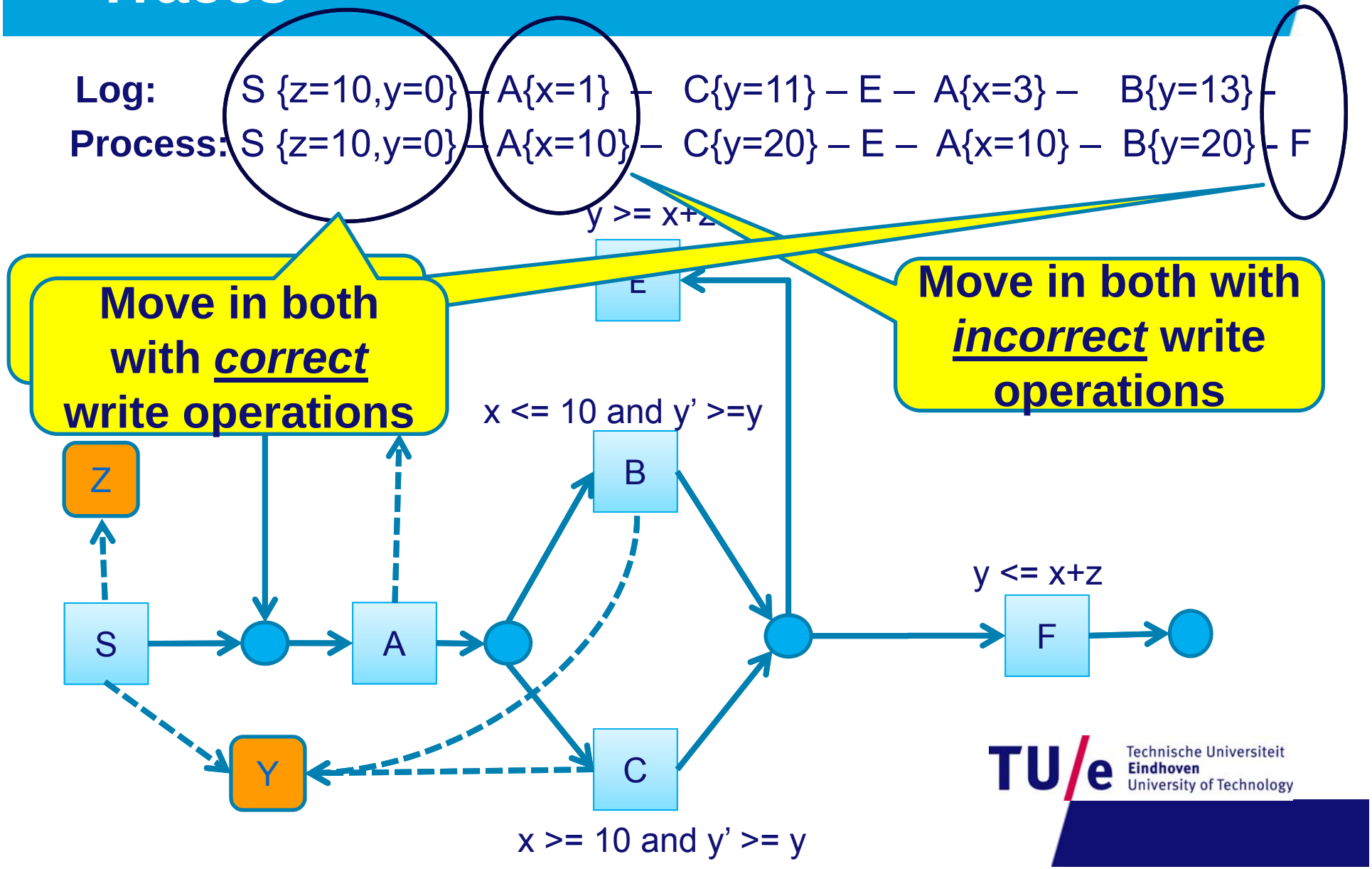
A{x=10} C{y=20} A{x=10} B{y=20} F



Alignments between Log and Process Traces

Log: S {z=10,y=0} – A{x=1} – C{y=11} – E – A{x=3} – B{y=13} –

Process: S {z=10,y=0} – A{x=10} – C{y=20} – E – A{x=10} – B{y=20} – F



Many Alignment Exists!

Log: S {z=10,y=0} – A{x=1} – C{y=11} – E – A{x=3} – B{y=13} -

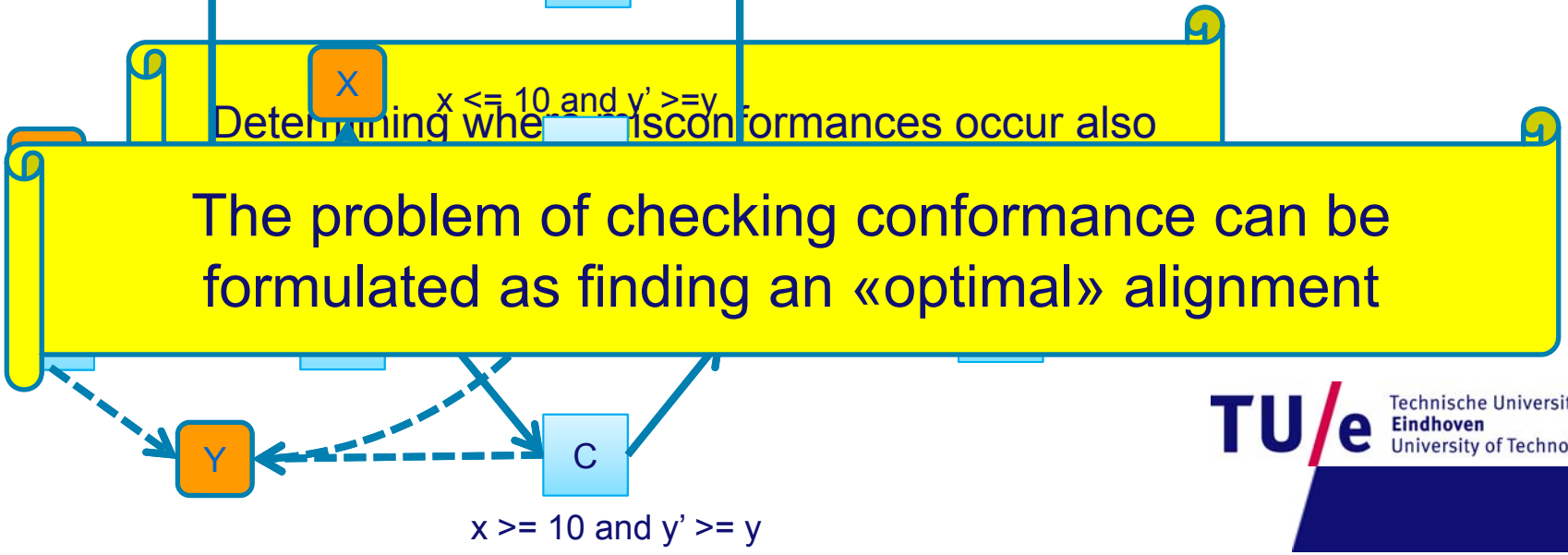
Process: S {z=10,y=0} – A{x=10} – C{y=20} – E – A{x=10} – B{y=20} - F

Log: S {z=10,y=0} – A{x=1} – C{y=11} – E – A{x=3} – B{y=13} -

Process: S {z=1, y=0} – A{x=10} – C{y=11} – E – A{x=3} – B{y=13} - F

Log: S {z=10,y=0} – A{x=1} C{y=11} – E – A{x=3} – B{y=13} -

Process: S {z=1, y=0} – A{x=1} B{y=11} – E – A{x=3} – B{y=13} - F

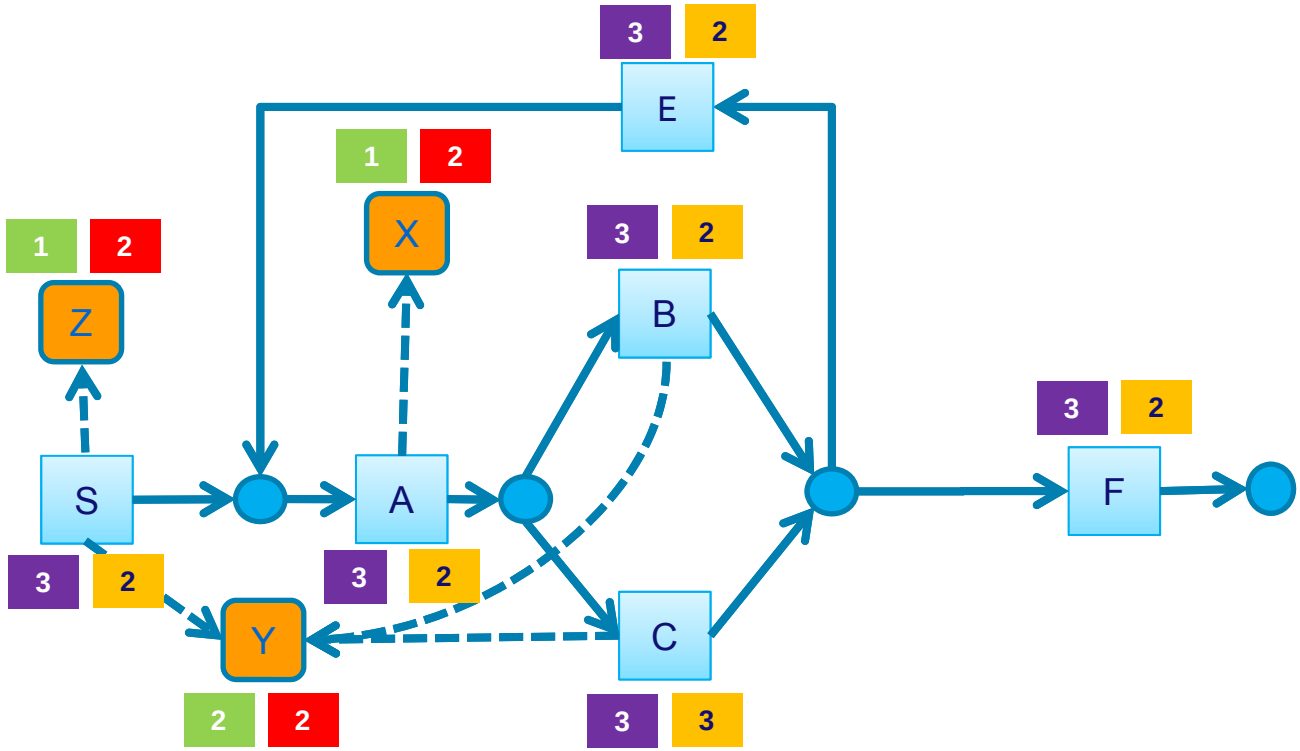


Determining where misconformances occur also

The problem of checking conformance can be formulated as finding an «optimal» alignment

Cost of alignments

- Each move is associated with a cost
- Cost of alignment is the sum of the costs of its moves



<x> : Cost of "move on model"

<y> : Cost of "move on log"

<w> : Cost of writing a wrong value

<z> : Cost of non-writing a variable

Cost of alignments: some examples

Log: S {z=10,y=0} – A{x=1} – C{y=11} – E – A{x=3} – B{y=13} -

Process: S {z=10,y=0} – A{x=10} – C{y=20} – E – A{x=10} – B{y=20} - F

8

Log: S {z=10,y=0} – A{x=1} – C{y=11} – E – A{x=3} – B{y=13} -

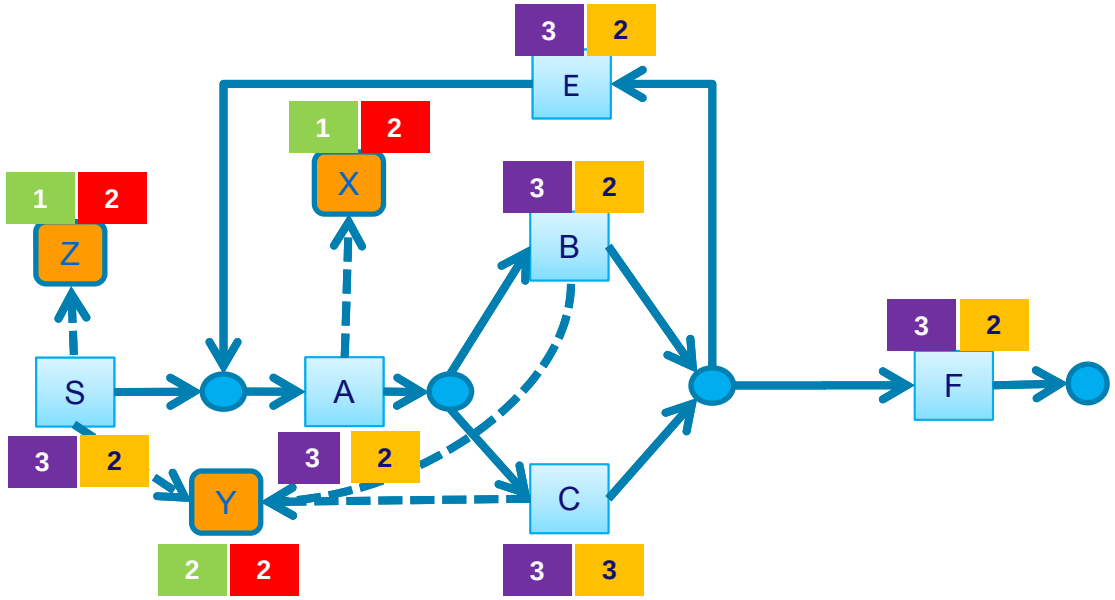
Process: S {z=1, y=0} – A{x=10} – C{y=11} – E – A{x=3} – B{y=13} - F

4

Log: S {z=10,y=0} – A{x=1} – C{y=11} – E – A{x=3} – B{y=13} -

Process: S {z=1, y=0} – A{x=1} – B{y= – E – A{x=3} – B{y=13} - F

5



An *optimal alignment*: an alignment with the lowest cost

Finding an optimal alignment

1. Computing the control-flow alignment using existing techniques [1]
2. Enriching the alignment with the data operations.

- The alignment is enriched, thus minimizing the cost of the alignment
- Naturally formulated as an Mixed Integer Linear Program

Log:	S {z=10,y=0}	–	A{x=1}	–	C{y=11}	–	E	–	A{x=3}	–	B{y=13}	-
Process:	S {z=1, y=0}	–	A{x=10}	–	C{y=11}	–	E	–	A{x=3}	–	B{y=13}	- F

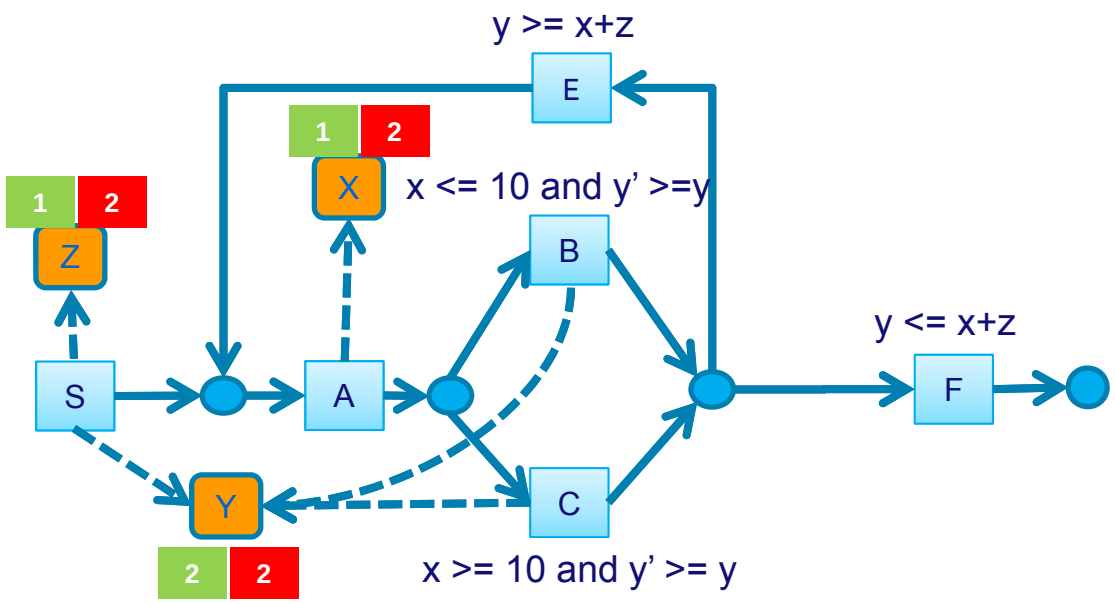
[1] **A. Adriansyah, A., B. F. van Dongen, B.F., W.M.P. van der Aalst,W.M.P. :** Conformance Checking Using Cost-Based Fitness Analysis. *IEEE International Enterprise Distributed Object Computing Conference (EDOC 2011)*

Construction of the ILP problem : converting guards to ILP constraints

Log: S {z=10,y=0} – A{x=1} – C{y=11} – E – A{x=3} – B{y=13} -

Process: S {z= z_1,y=y_1} – A{x=x_1} – C{y= y_2} – E – A{x= x_2} – B{y= y_3} - F

Variables: v_i = the i-th write operation for variable v



$$
\begin{cases}
\dots \\
y_2 \leq x_2 + z_1 \\
y_3 \geq y_2 \\
x_1 \geq 10 \\
x_2 \leq 10 \\
y_2 \geq x_1 + z_1 \\
y_2 \geq y1 \\
x_1, x_2, y_1, y_2, y_3, z_1 \in N
\end{cases}
$$

Technische Universiteit Eindhoven University of Technology

Construction of the ILP problem : the objective function

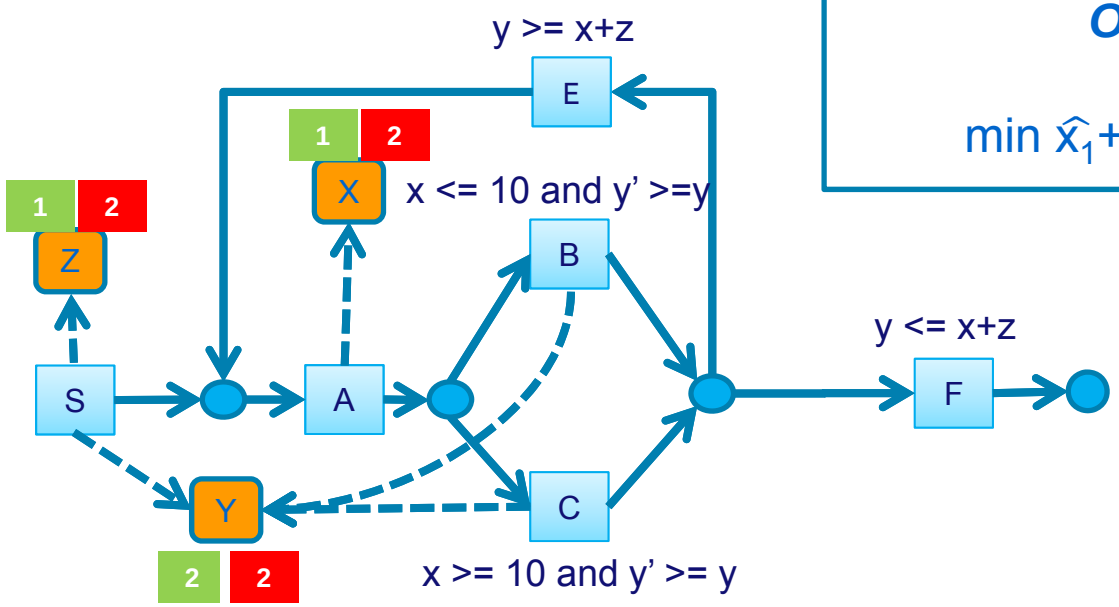
Log: S {z=10,y=0} – A{x=1} – C{y=11} – E – A{x=3} – B{y=13} -

Process: S {z= z_1,y=y_1} – A{x=x_1} – C{y= y_2} – E – A{x= x_2} – B{y= y_3} - F

Additional variables: $\hat{v}_i =$

$$\begin{cases} 0 & \text{if value assigned to } v_i \text{ is the same as in the corresponding log event} \\ 1 & otherwise \end{cases}$$

E.g., for the first write operation for x: $\hat{x}_1 = 0 \Leftrightarrow x_1 = 1$



Objective Function:

$$\min \hat{x}_1 + \hat{x}_2 + \hat{x}_3 + 2\hat{y}_1 + 2\hat{y}_2 + 2\hat{y}_3 + \hat{z}_1$$

Construction of the ILP problem : converting guards to constraints

Log: S {z=10,y=0} – A{x=1} – C{y=11} – E – A{x=3} – B{y=13} -

Process: S {z= z_1,y=y_1} – A{x=x_1} – C{y= y_2} – E – A{x= x_2} – B{y= y_3} - F

$\min \hat{x}_1 + \hat{x}_2 + \hat{x}_3 + 2\hat{y}_1 + 2\hat{y}_2$

$y_2 \leq x_2 + z_1$

$y_3 \geq y_2$

$x_2 \leq 10$

$y_2 \geq x_1 + z_1$

$y_2 \geq y1$

$x_1 \geq 10$

$\hat{x}_1 = 0 \Leftrightarrow x_1 = 1$

$\hat{x}_2 = 0 \Leftrightarrow x_2 = 3$

$\hat{y}_1 = 0 \Leftrightarrow y_1 = 0$

$\hat{y}_2 = 0 \Leftrightarrow y_2 = 11$

$\hat{y}_3 = 0 \Leftrightarrow y_3 = 13$

$\hat{z}_1 = 0 \Leftrightarrow z_1 = 10$

$\hat{x}_1, \hat{x}_2, \hat{y}_1, \hat{y}_2, \hat{y}_3, \hat{z}_1 \in \{0,1\}$

$x_1, x_2, y_1, y_2, y_3, z_1 \in N$

$\hat{v}_i = 0 \Leftrightarrow v_i = k$

can be rewritten as two linear constraints:

$v_i - M\hat{v}_i \leq k$

$-v_i - M\hat{v}_i \leq -k$

$= 13$

- **…ment cost=**

Value Objective Function at minimum

+

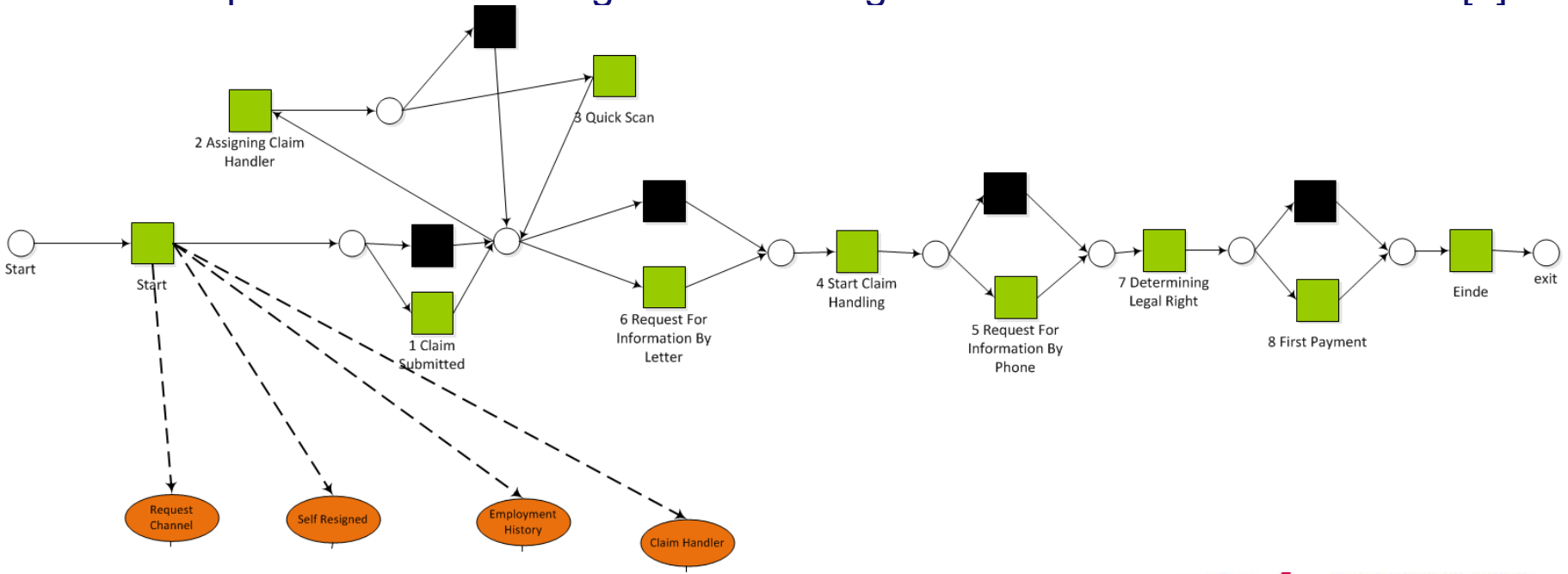
Cost of Log/Process Move=2+2

Additional features

- Support for the OR operator in the transition guards
- Support for any time of primitive types (through a bidirectional conversion to numerical value):
  - Boolean
  - Timestamp
  - String

Implementation and Experiments

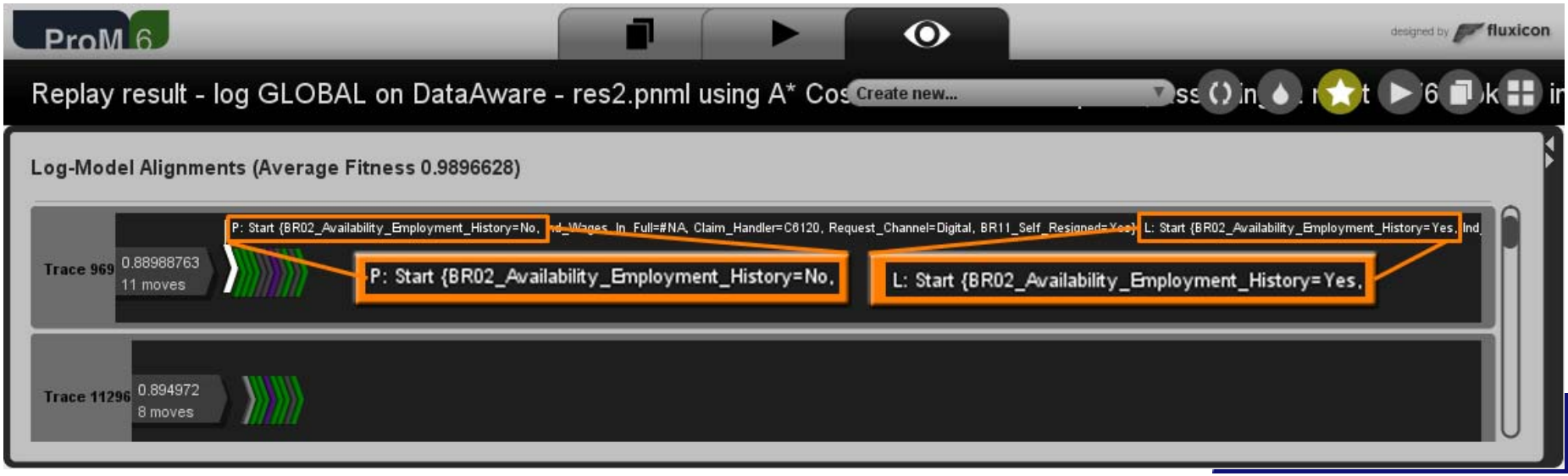
- Implemented as ProM plugins
- Tested with a real-life event log
  - The event log recorded the execution of 12319 process cases with a Dutch insurance institute
  - The process model was generated using the new ProM *Decision Miner* [2]



[2] **Massimiliano de Leoni, Wil M. P. van der Aalst**: Data-aware process mining: discovering decisions in processes using alignments. *28th Annual ACM Symposium on Applied Computing (SAC '13)*

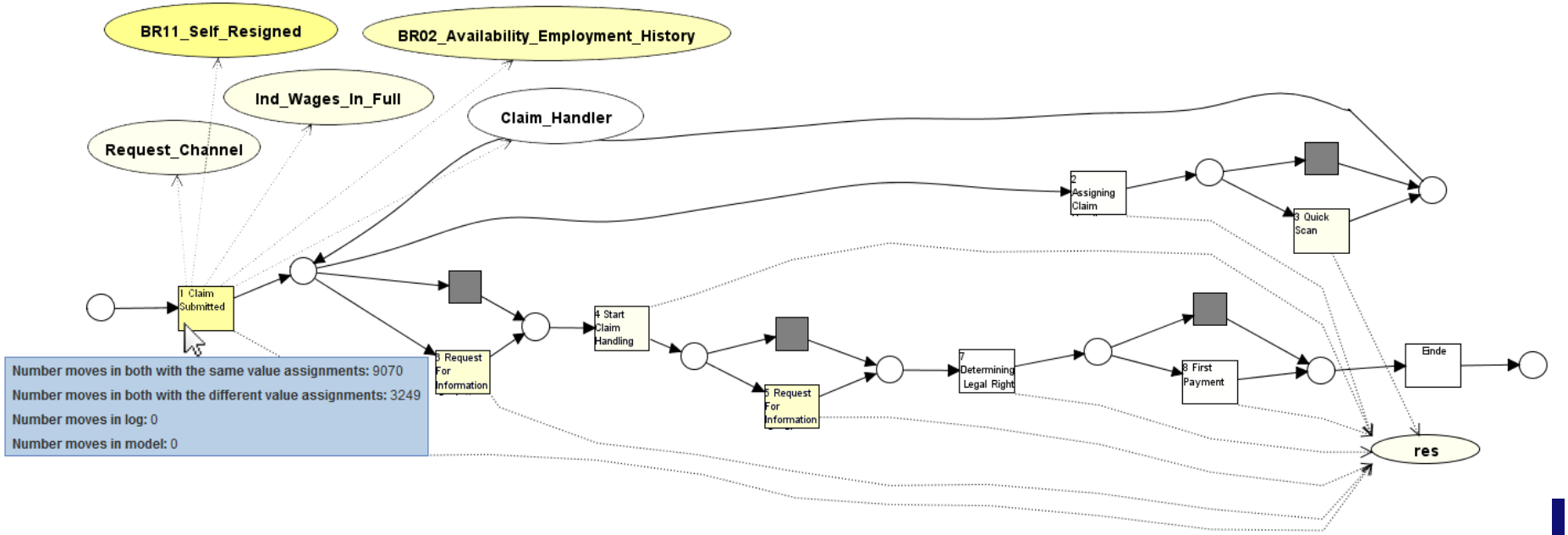
Visualization of the optimal alignments in ProM

- Each row is the optimal alignment of a different log trace
- Alignments are visualized a sequence of triangles
  - Each triangle corresponds to a move
  - The type of move is identified by the colour
    - Green: move in both with correct write operations
    - White: move in both with incorrect write operations
    - Purple: move in model; Yellow: move in log



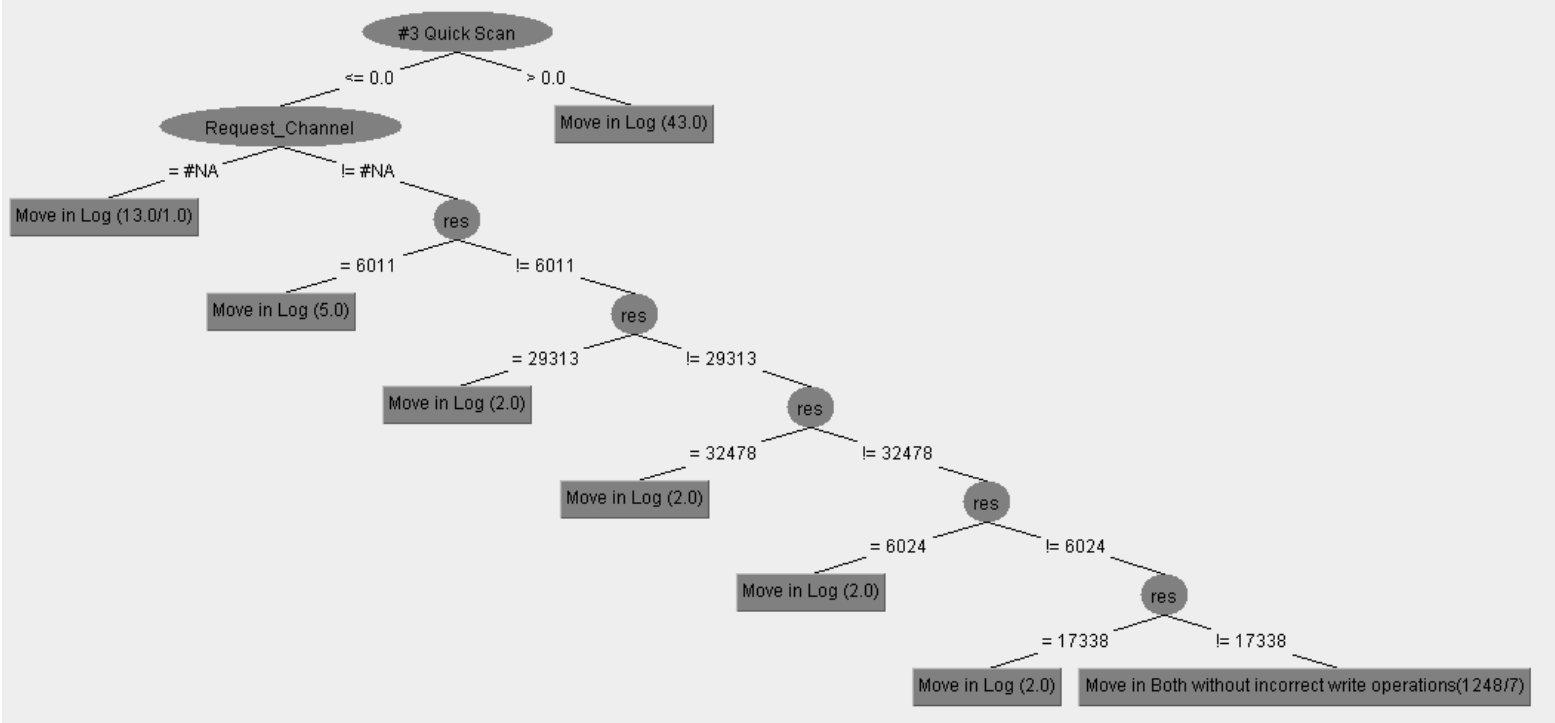
A Helicopter view on the optimal alignments : projection onto models

- When there are so many alignments (12319), difficult to gain a general insight into the most common problems
- We also provide a visualization where alignments are projected onto the model
  - Darker colour for activities/variables more involved into deviations

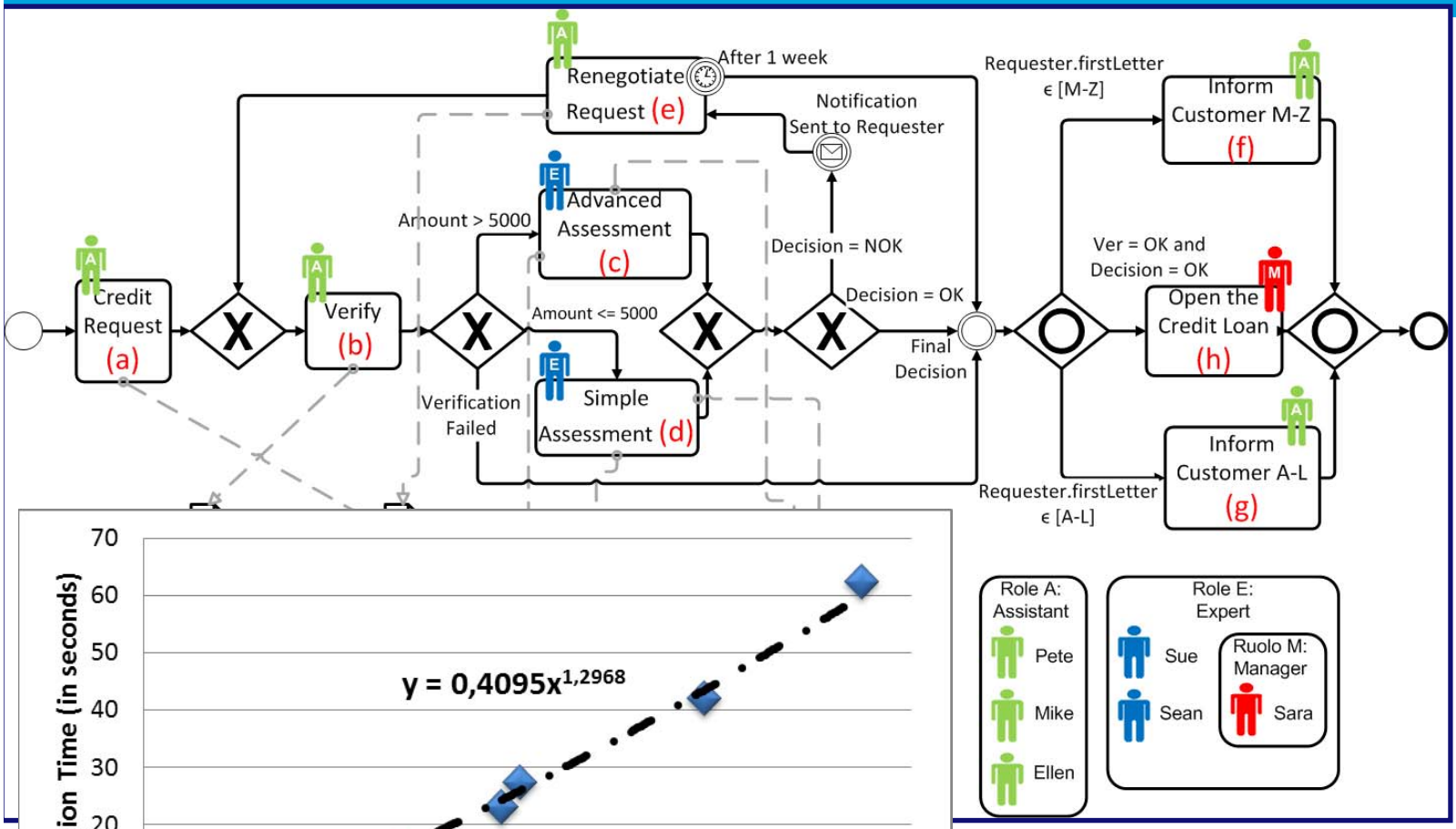


A Helicopter view on the optimal alignments : root causes of deviations

- Transitions are associated with decision trees:
  - Features: Process Variables + #execution of transitions in the prefix
  - Classification Attribute: The move types (in log, model, etc.)



Execution-Time Analysis



Technische Universiteit
Eindhoven
University of Technology