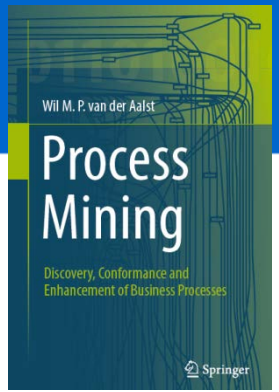
Process Mining: Data Science in Action

Aligning Observed and Modeled Behavior

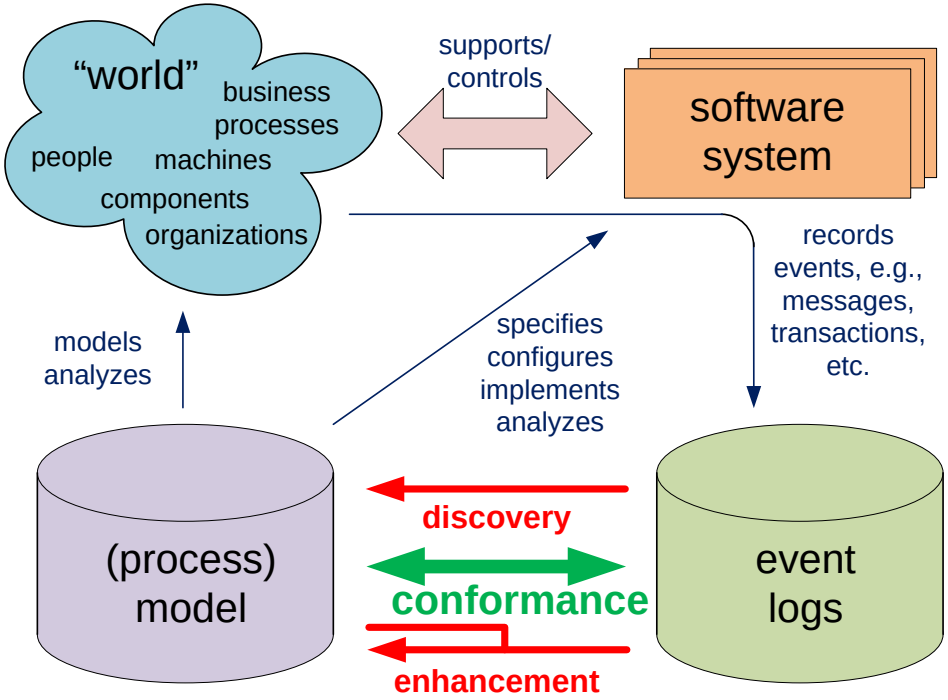
prof.dr.ir. Wil van der Aalst
www.processmining.org



Technische Universiteit Eindhoven University of Technology

Where innovation starts

Conformance checking



1. Conformance checking using causal footprints.
2. Conformance checking based on token-based replay.
3. **Alignment-based conformance checking.**

TU/e

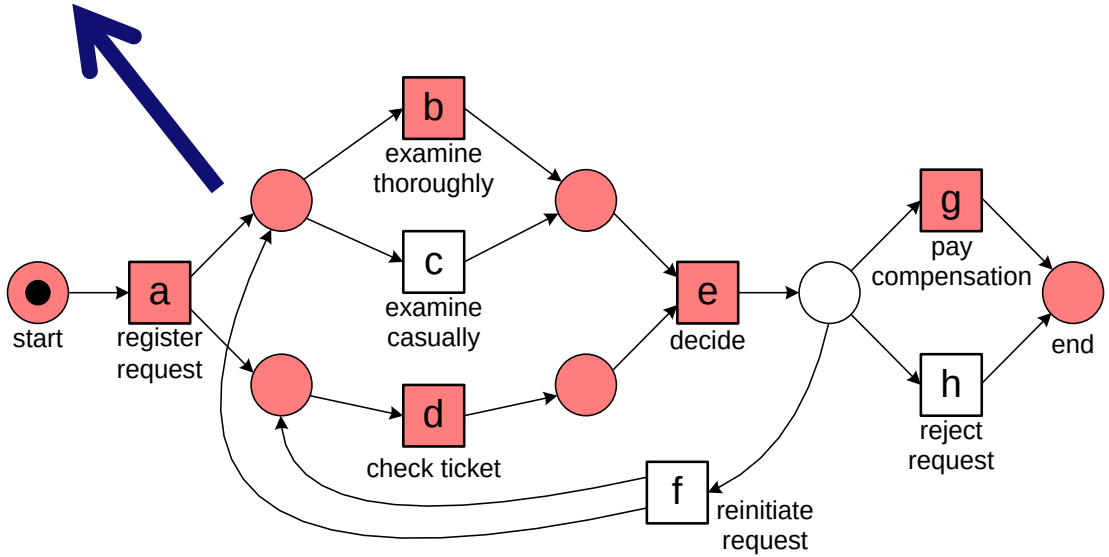
Requirements

- Conformance checking should not impose restrictions on the process notation (e.g., silent transitions and two transitions with the same label should be possible).
- Should provide a "closest matching path" through the process model for any trace in the event log.
  - Also required for performance analysis!
  - Beyond the analysis of replay fitness (advanced diagnostics, precision, generalization, etc.).

Alignments

a	b	d	e	g
a	b	d	e	g

⟨a,b,d,e,g⟩

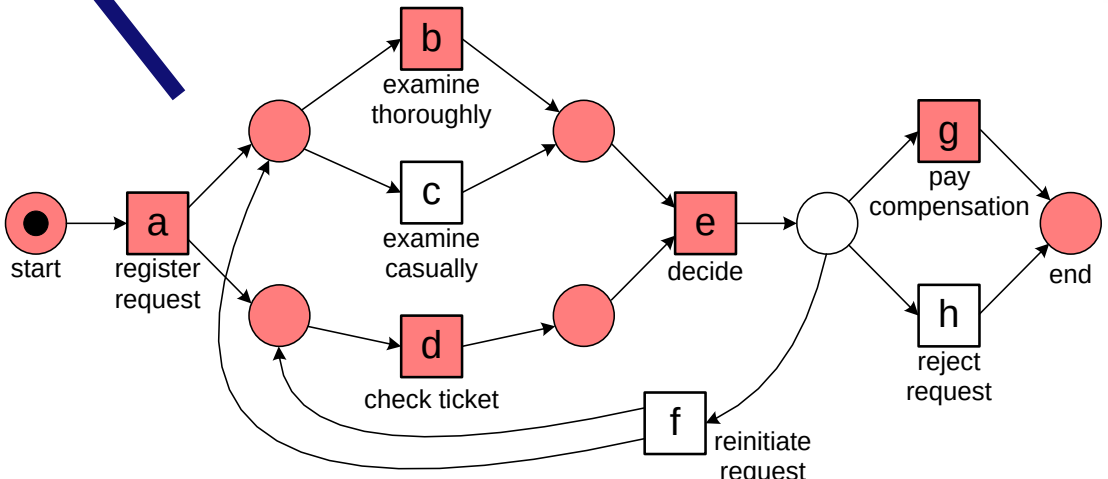


#	trace
455	acdeh
191	abdeg
177	adceh
144	abdeh
111	acdeg
82	adceg
56	adbeh
47	acdefdbeh
38	adbeg
33	acdefbdeh
14	acdefbdeg
11	acdefdbeg
9	adcefcdeh
8	adcefdbeh
5	adcefbdeg
3	acdefbdefdbeg
2	adcefdbeg
2	adcefbdefbdeg
1	adcefdbefbdeh
1	adbefbdefdbeg
1	adcefdbefcdefdbeg
1391	

Alignments

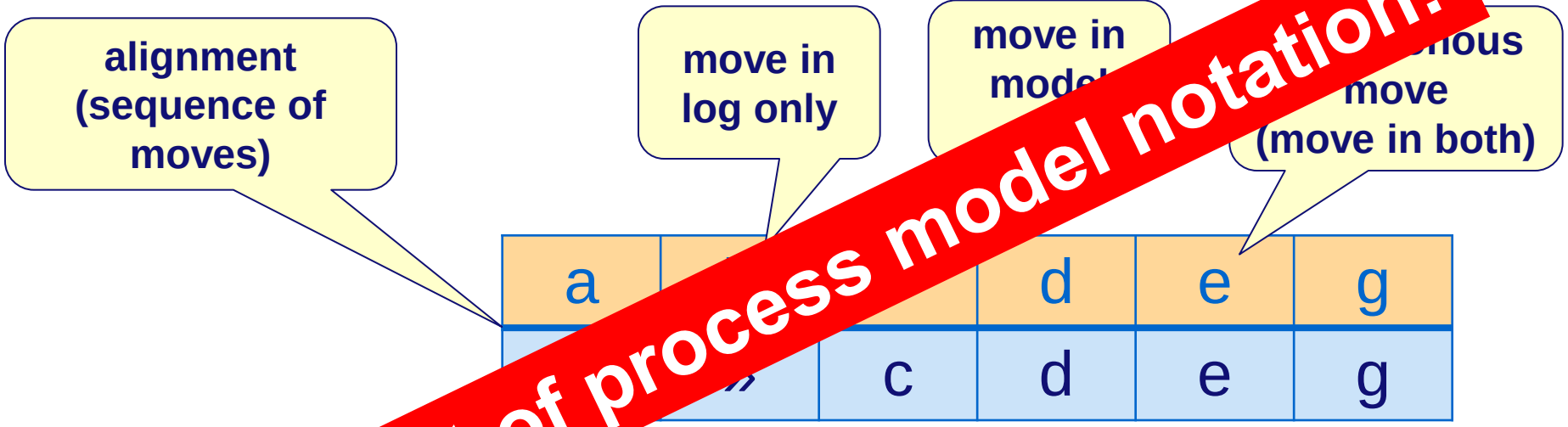
a	»	d	e	g	h
a	b	d	e	g	»

⟨a,d,e,g,h⟩



TU/e

Terminology



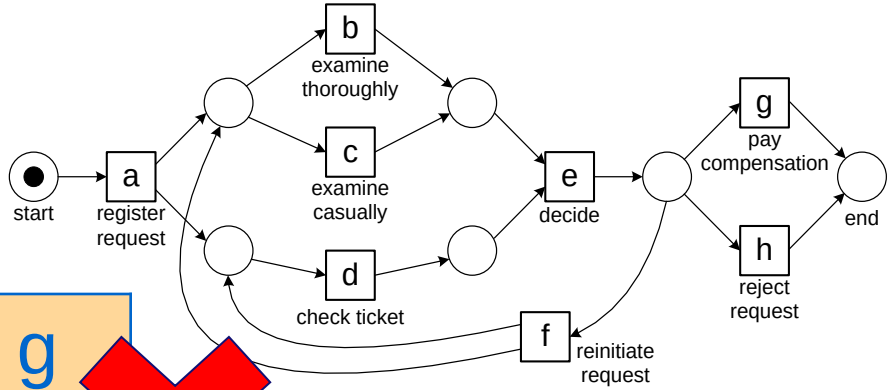
Independent of process model notation!

- Projecti… row (remove "no moves") corresponds to t… in the event log.
  …tion on bottom row (remove "no moves") …rresponds to a run of the model.

TU/e

Optimal alignment for ⟨a,b,d,e,g⟩

a	b	d	e	g
a	b	d	e	g

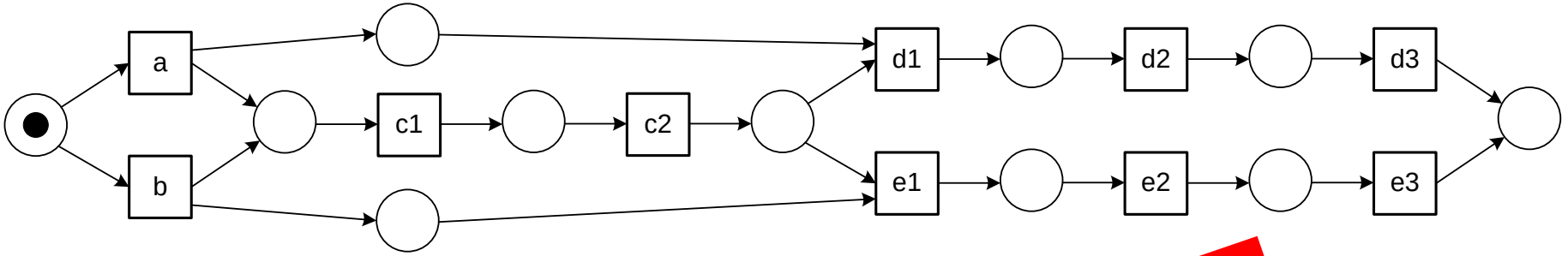


a	b	»	d	e	g
a	»	c	d	e	g

a	b	d	e	g	»	»	»	»	»
»	»	»	»	»	a	c	d	e	g

TU/e

Optimal alignment for ⟨a,c1,c2,e1,e2,e3⟩ ?



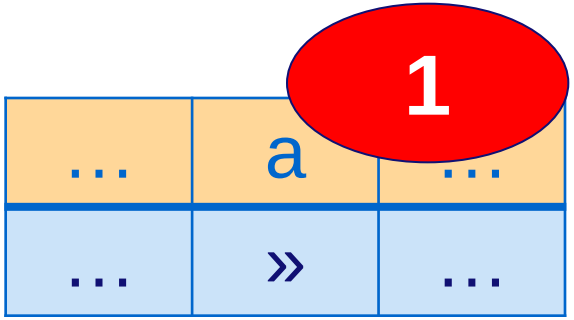
a	»	c1	c2	e1	e2	e3
»	b	c1	c2	e1		

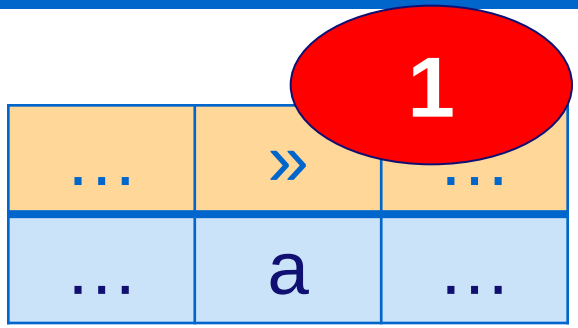
a	c1			»	e1	e2	e3
a	c	d1	d2	d3	»	»	»

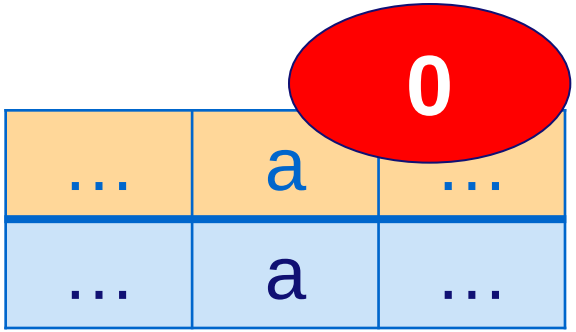
Depends on cost function!

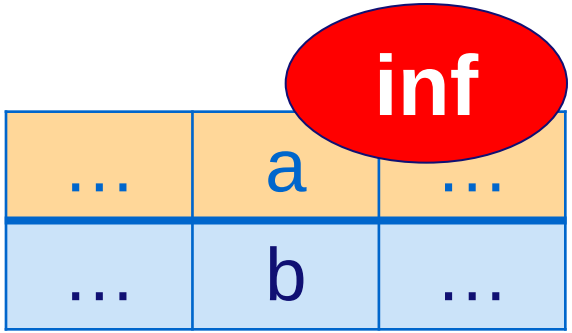
TU/e

Standard cost function
count »'s in alignment









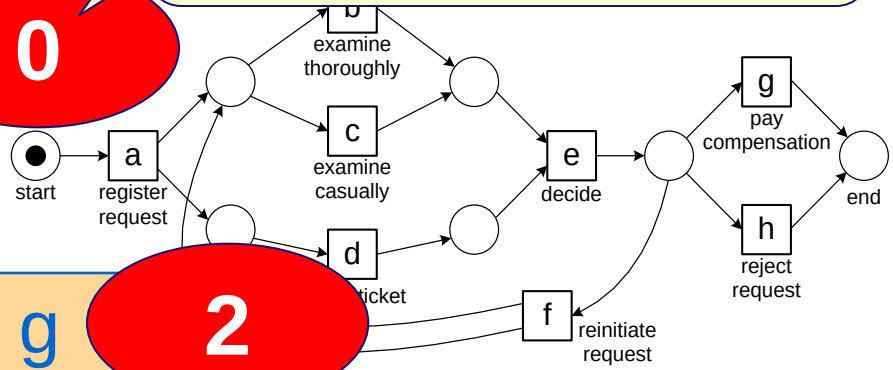
TU/e

Using the standard cost f

optimal
there is no other alignment that has lower costs

a	b	d	e	g
a	b	d	e	g

0



a	b	»	d	e	g
a	»	c	d	e	g

2

a	b	d	e	g	»	»	»	»	»
»	»	»	»	»	a	c	d	e	g

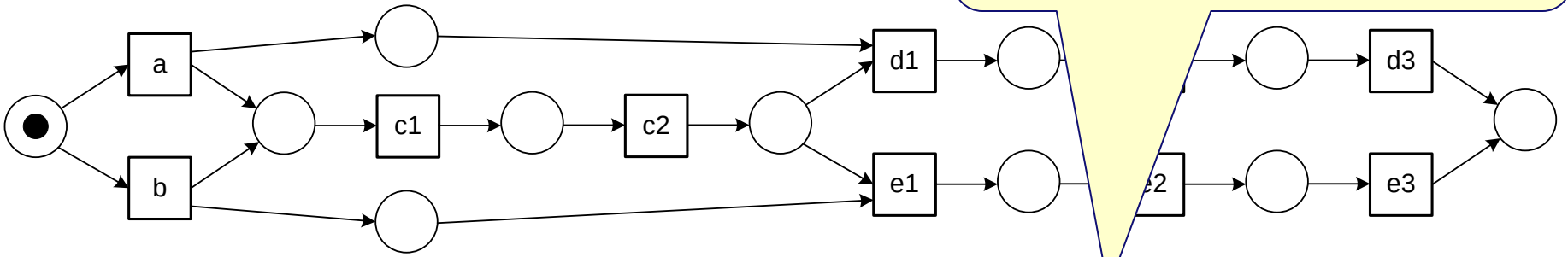
10

TU/e

Using the standard cost fu

trace in log: ⟨a,c1,c2,e1,e2,e3⟩

optimal
there is no other alignment that has lower costs



a	»	c1	c2	e1	e2	e3
»	b	c1	c2	e1	e2	e3

2

a	c1	c2	»	»	»	e1	e2	e3
a	c1	c2	d1	d2	d3	»	»	»

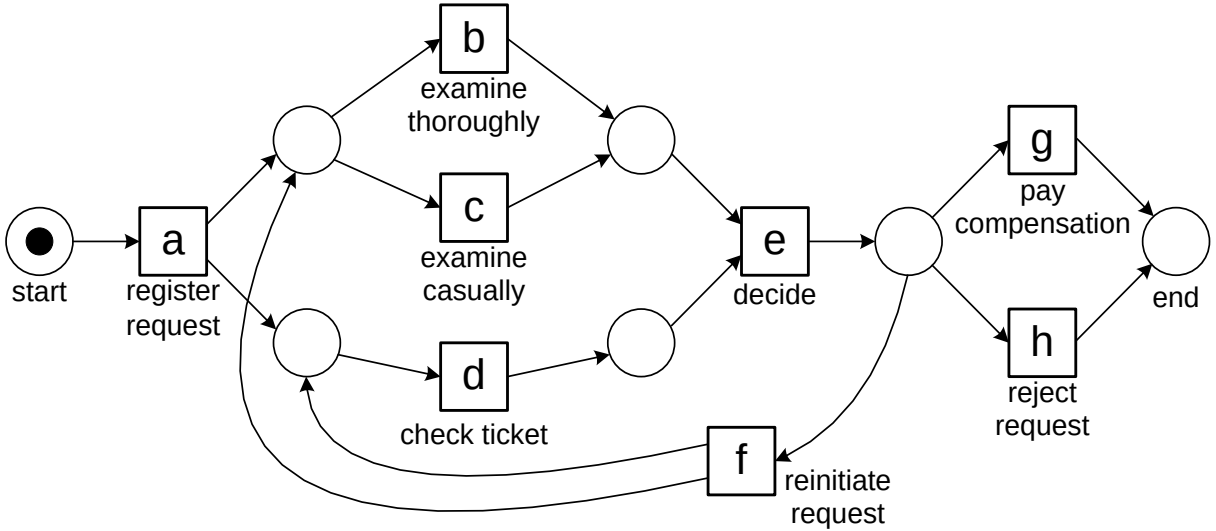
6

TU/e

Optimal alignment for ⟨a,b,e,f,d,e,g⟩ (1/2)

a	b	e	f	d	e	g
a	b	»	»	d	e	g

2



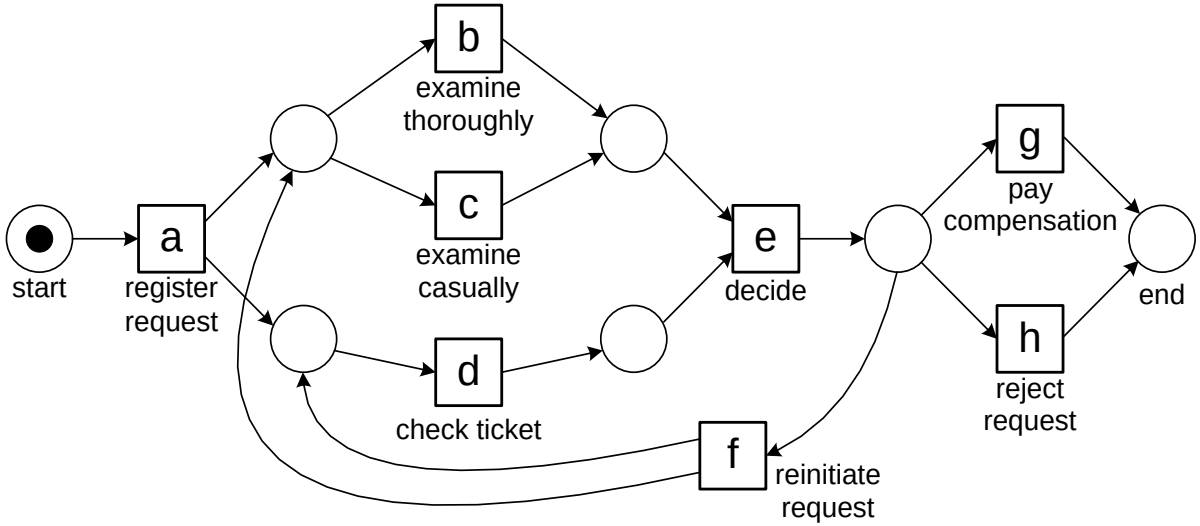
loop is not taken: e and f in event log are discarded

TU/e

Optimal alignment for ⟨a,b,e,f,d,e,g⟩ (2/2)

a	b	»	e	f	d	»	e	g
a	b	d	e	f	d	b	e	g

2



loop is taken: d and b are missing in event log

TU/e

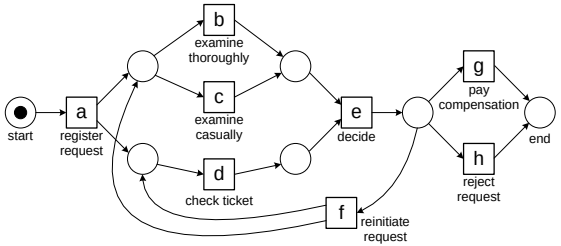
Not one unique optimal alignment

a	b	»	e	f	d	»	e	g
a	b	d	e	f	d	b	e	g

2

a	b	e	f	d	e	g
a	b	»	»	d	e	g

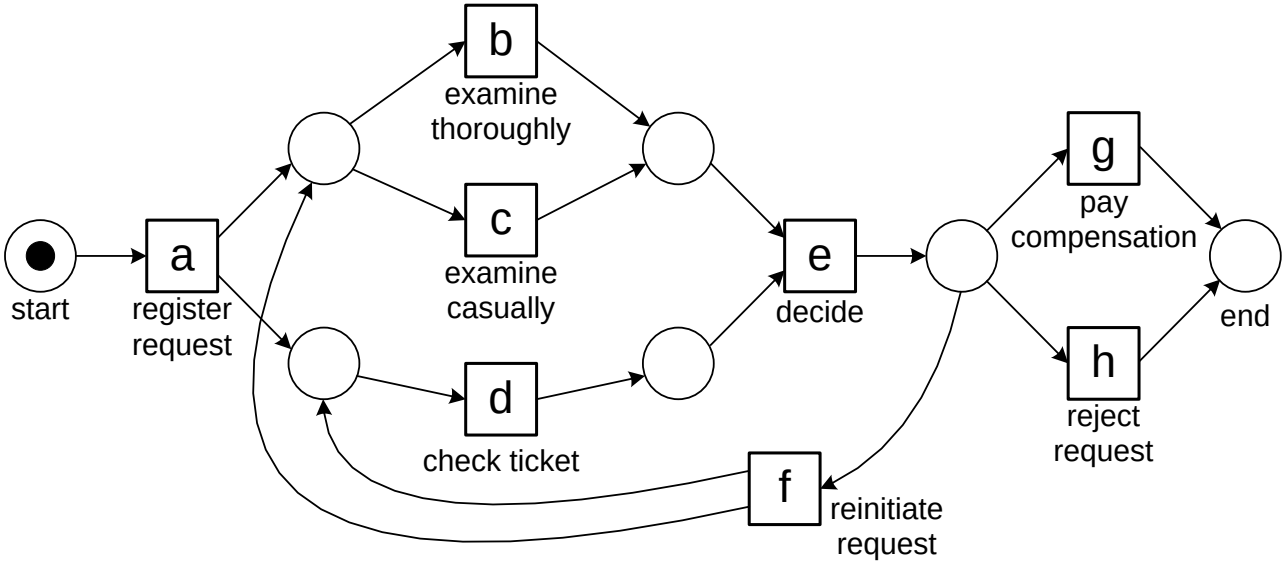
2



■ ■ ■

TU/e

Question: How many optimal alignments are there for ⟨a,b,e,f,d,e,g⟩ ?



⟨a,b,e,f,d,e,g⟩

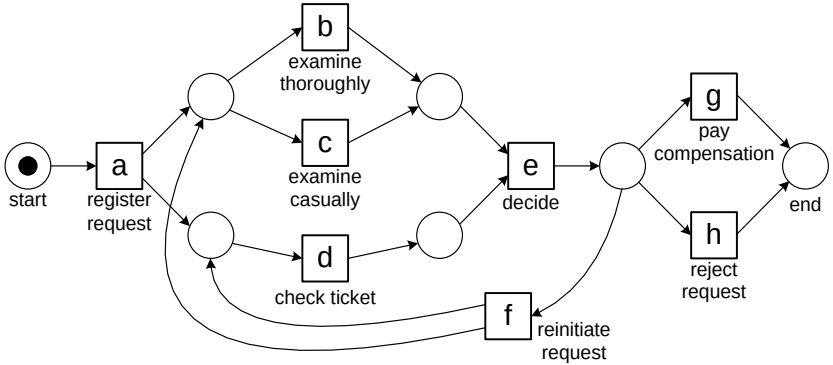
TU/e

Answer: 9

1+(2x2)+(2x2)=9 optimal alignments having cost 2

a	b	e	f	d	e	g
a	b	»	»	d	e	g

1x



move in model can be reordered in concurrent part

4x

a	b	»	e	f	d	»	e	g
a	b	d	e	f	d	b	e	g

4x

a	b	»	e	f	d	»	e	g
a	b	d	e	f	d	c	e	g

TU/e

Any cost structure is possible

...	send-letter(John,4 weeks, $400)	...
...	send-email(Sue,3 weeks,$500)	...

TU/e

Any cost structure is possible

...	send-letter(John,4 weeks, $400)	...
...	send-email(Sue,3 weeks,$500)	...

similar activities (lower costs for related activities)

TU/e

Any cost structure is possible

...	send-letter(John,4 weeks, $400)	...
...	send-email(Sue,3 weeks,$50	...

resource-related conformance costs (done by someone that does or does not have the specified role)

TU/e

Any cost structure is possible

...	send-letter(John,4 weeks, $400)	...
...	send-email(Sue,3 weeks,$500)	...

time-related conformance costs (activity should happen within a preset deadline)

TU/e

Any cost structure is possible

...	send-letter(John,4 weeks, $400)	...
...	send-email(Sue,3 weeks,$500)	...

data-related conformance costs (routing condition is violated, e.g., path only for more valuable orders)

TU/e

Any cost structure is possible

...	send-letter(John,4 weeks, $400)	...
...	send-email(Sue,3 weeks,$500)	...

risk-related conformance costs, context-dependent conformance costs, …

TU/e

Computing fitness

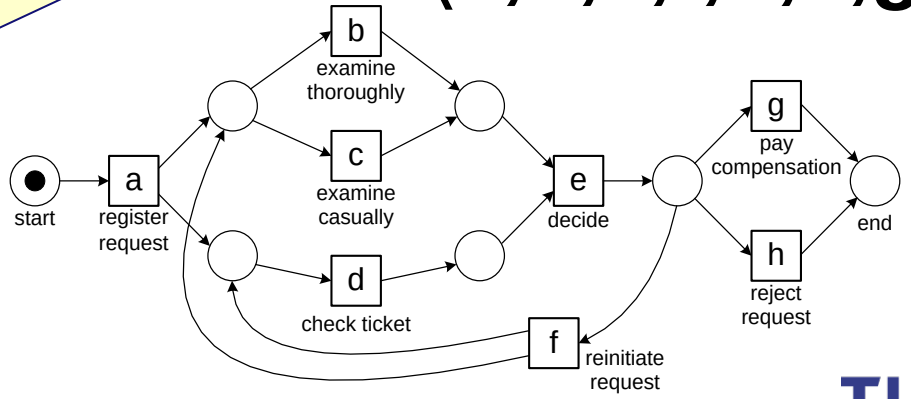
C… …n optimal alignment with a mo… …model.

all events cause a move in log only

shortest path from initial state to final state

$$1 - \frac{2}{7+5} = 0.833$$

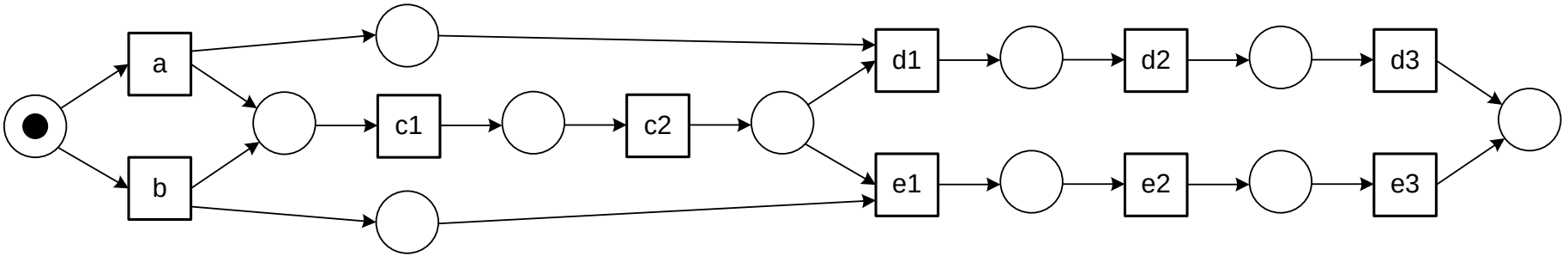
⟨a,b,e,f,d,e,g⟩



TU/e

Question: Compute alignment-based fitness

$$\langle a, c1, c2, e1, e2, e3 \rangle$$



TU/e

Answer

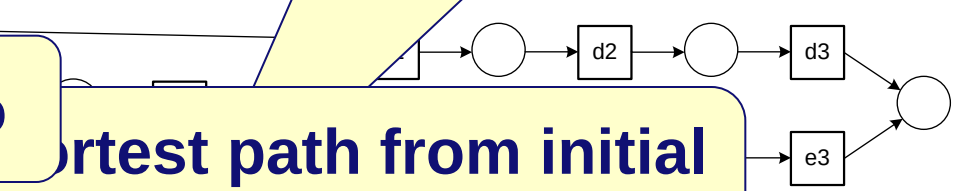
cost of optimal alignment = 2

⟨a,c1,c2,e1,e2,e3⟩

all ev… move in log only

worst-case scenario

…ortest path from initial state to final state



a		c1	c2		e2	
»		c1	c2	e	e2	e

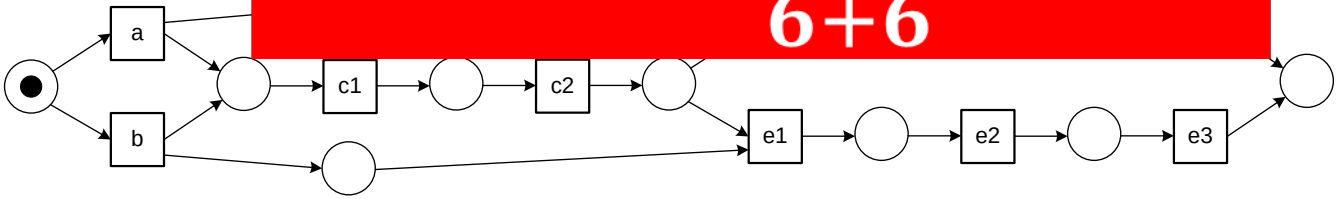
a	c1	c2	e1	e2	e3	»	»	»	»	»	»
»	»	»	»	»	»	a	c1	c2	d1	d2	d3

TU/e

Answer

$$\text{fitness} = 1 - \frac{2}{6+6} = 0.833$$

⟨a,c1,c2,e1,e2,e3⟩



2

a	»	c1	c2	e1	e2	e3
»	b	c1	c2	e1	e2	e3

6

a	c1	c2	e1	e2	e3	»	»	»	»	»	»
»	»	»	»	»	»	a	c1	c2	d1	d2	d3

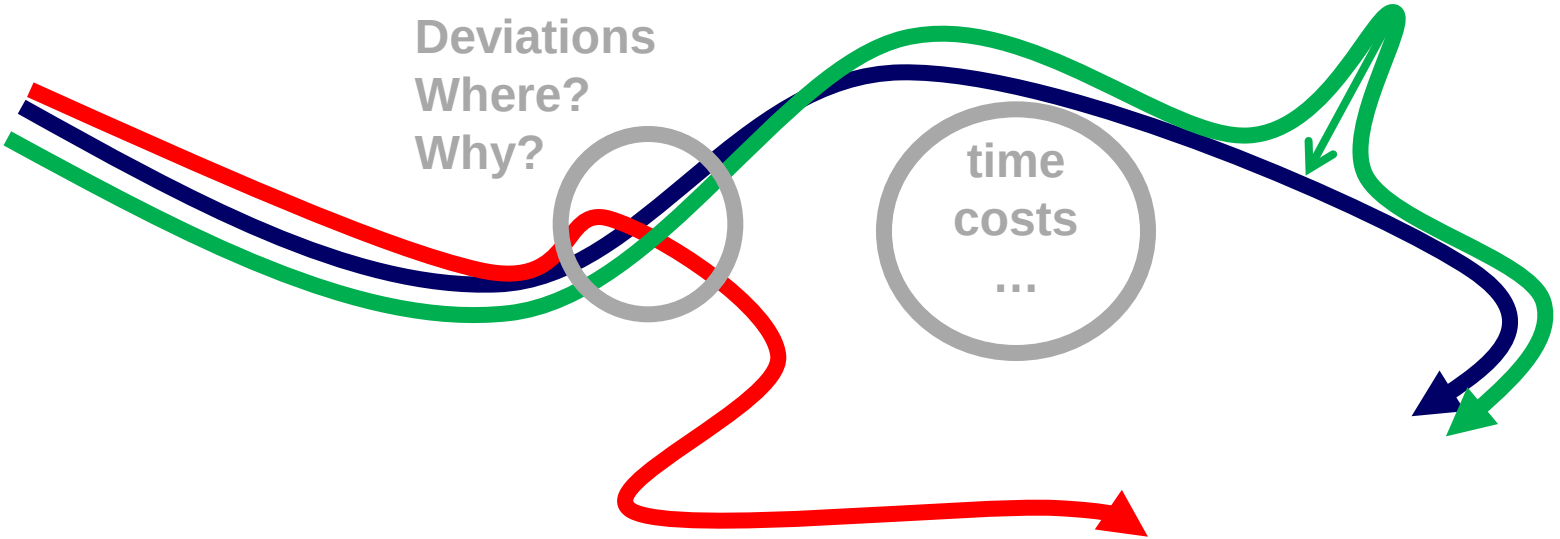
6

TU/e

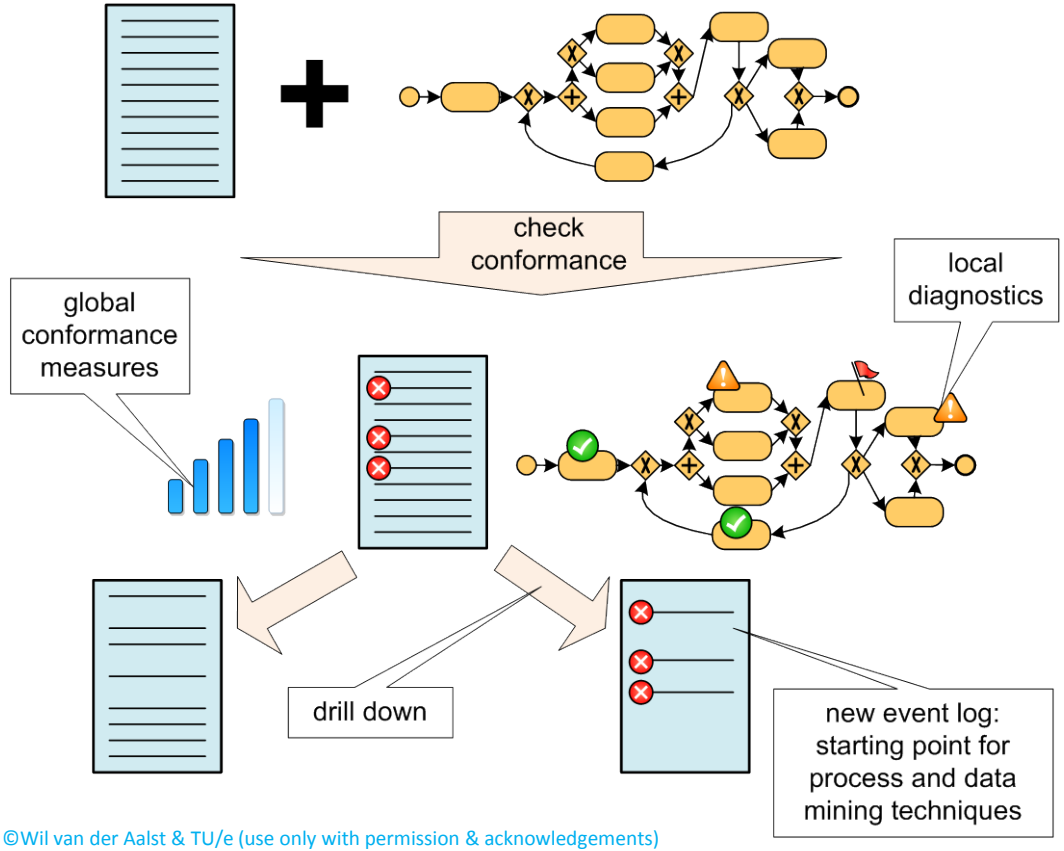
Advantages of aligning log and model

- Observed behavior is directly related to modeled behavior.
- Very flexible (any cost structure).
- Detailed diagnostics.
- After aligning log and model, other quality dimensions can be investigated (separation of concerns).

TU/e


Deviations
Where?
Why?
time
costs
...

Drilling down

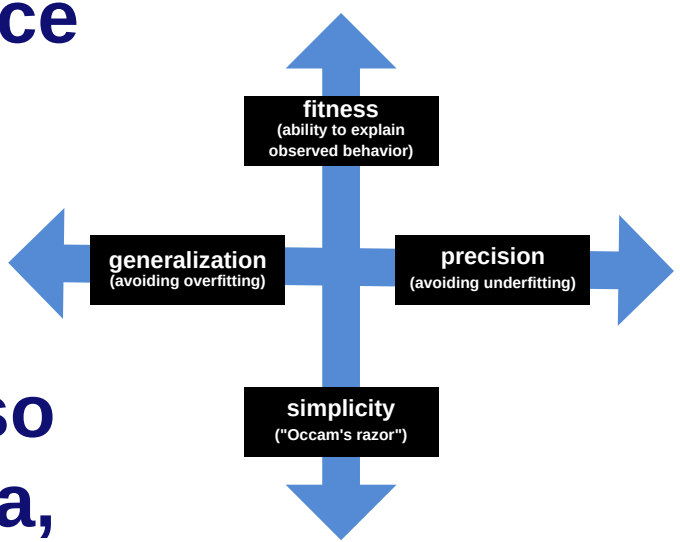


- Create event log containing deviating (or non-deviating) cases.
- Apply process mining to new log.
- Comparative process mining.

TU/e

Beyond fitness and control-flow

- There are also solid conformance measures for precision, generalization, and simplicity.
- Multiple definitions possible.
- Conformance checking may also include other perspectives (data, resources, time, cost, etc.).
- Example: data-aware alignments.



TU/e

Part I: Preliminaries

- Chapter 1: Introduction
- Chapter 2: Process Modeling and Analysis
- Chapter 3: Data Mining

Part II: From Event Logs to Process Models

- Chapter 4: Getting the Data
- Chapter 5: Process Discovery: An Introduction
- Chapter 6: Advanced Process Discovery Techniques

Part III: Beyond Process Discovery

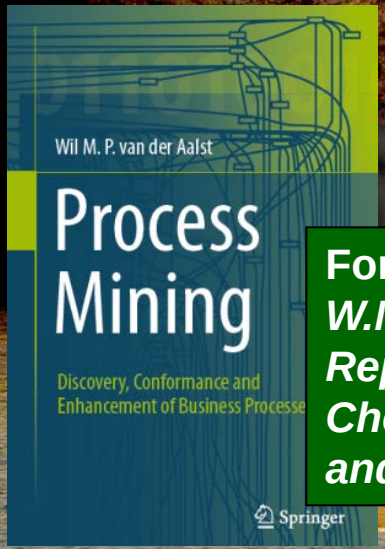
- Chapter 7: Conformance Checking
- Chapter 8: Mining Additional Perspectives
- Chapter 9: Operational Support

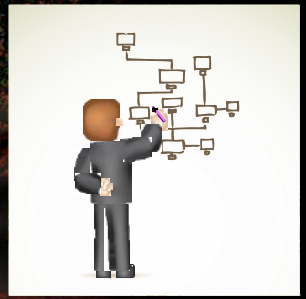
Part IV: Putting Process Mining to Work

- Chapter 10: Tool Support
- Chapter 11: Analyzing "Lasagna Processes"
- Chapter 12: Analyzing "Spaghetti Processes"

Part V: Reflection

- Chapter 13: Cartography and Navigation
- Chapter 14: Epilogue





For more details see:
W.M.P. van der Aalst, A. Adriansyah, and B. van Dongen. Replaying History on Process Models for Conformance Checking and Performance Analysis. WIREs Data Mining and Knowledge Discovery, 2(2):182-192, 2012.

TU/e