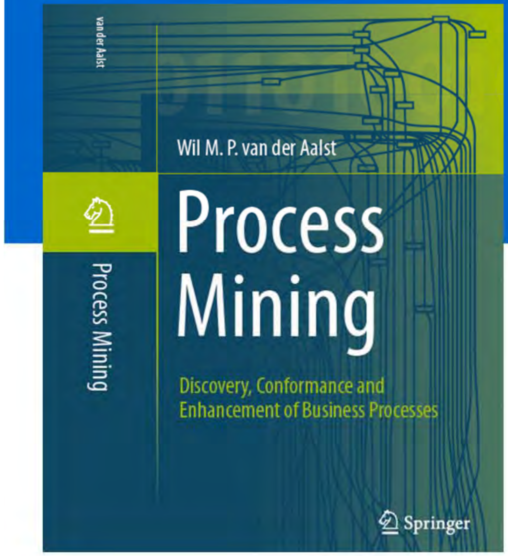
Discovering Concurrency

Learning (Business) Process Models from Examples

Invited Talk CONCUR 2011, 8-9-2011, Aachen.

prof.dr.ir. Wil van der Aalst

www.processmining.org



van der Aalst

Wil M. P. van der Aalst

Process Mining

Process Mining

Discovery, Conformance and Enhancement of Business Processes

Springer

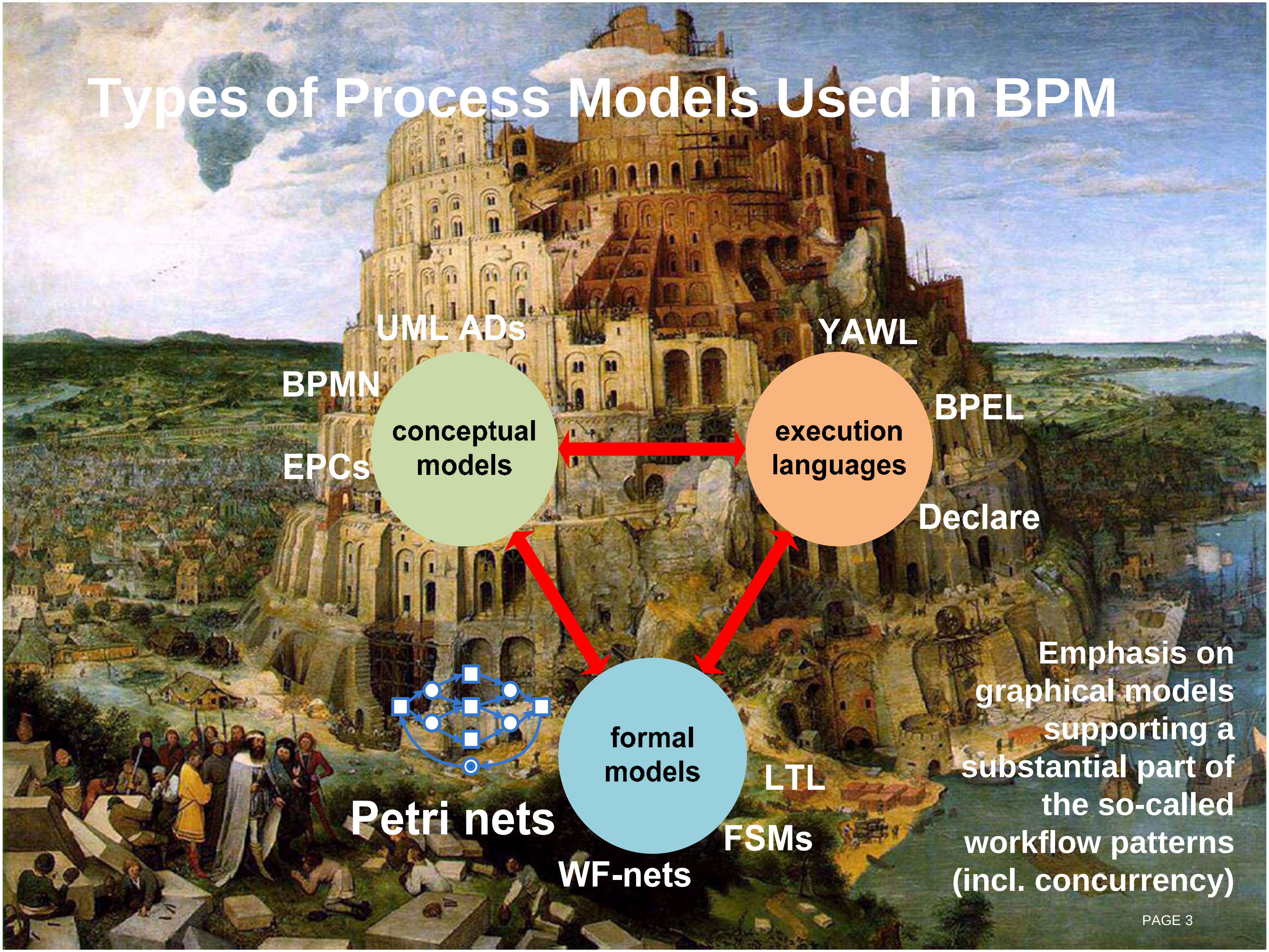
TU/e Technische Universiteit Eindhoven University of Technology

Where innovation starts

Business Process Management ?

Business
Process
Management !!

Types of Process Models Used in BPM



UML ADs

BPMN

EPCs

conceptual models

YAWL

BPEL

execution languages

Declare

formal models

Petri nets

WF-nets

LTL

FSMs

Emphasis on graphical models supporting a substantial part of the so-called workflow patterns (incl. concurrency)

Classical Challenges in BPM

- Verification (cf. soundness problem in WF-nets)
- Performance analysis (e.g., simulation)
- Converting models into running systems
- Providing flexibility without loosing control
- …

Many problems have been “solved”: the real problem is adoption in practice!

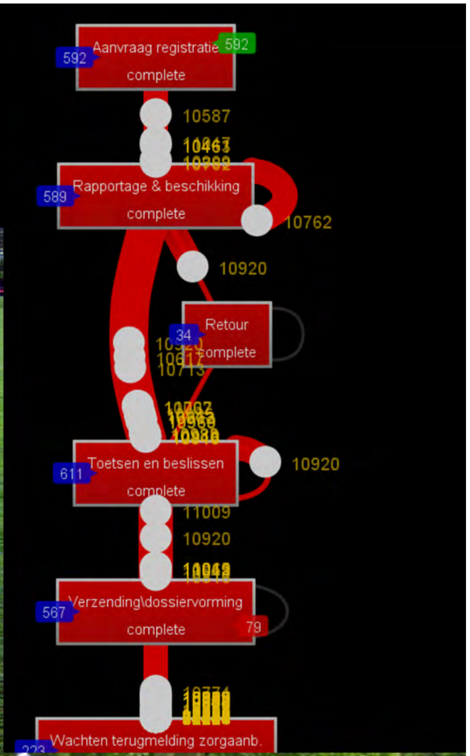
A more interesting challenge: Dealing with variability

- Variants of the same process exist in various domains, e.g., Dutch Municipalities, Hertz, Suncorp, Salesforce, Easychair, etc.
- Configurable process models to generate concrete processes, cf. C-YAWL.
- Merging process models is a challenging problem.
- Model similarity (rather than equivalence).

So we are interested in processes …

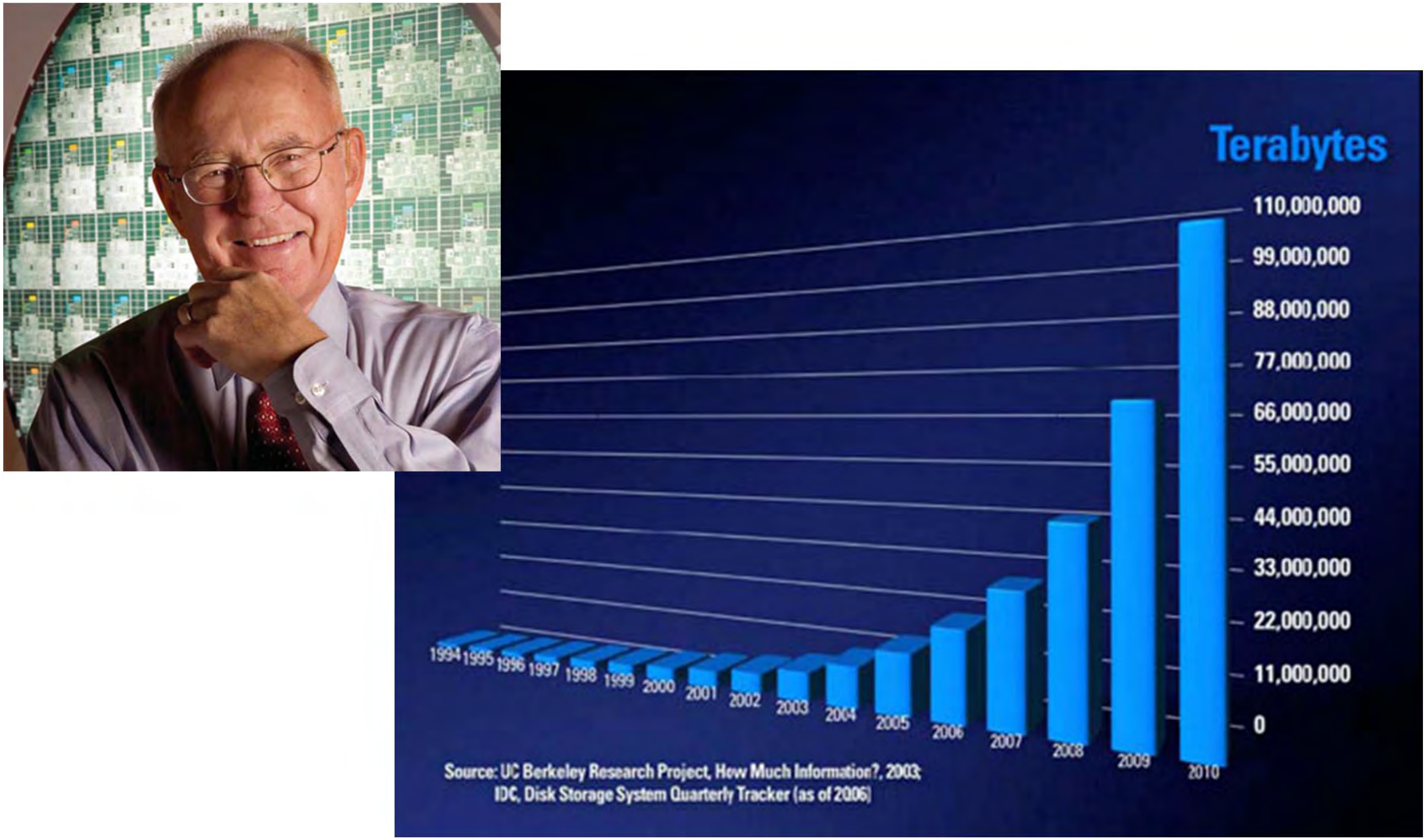
… but most of us do not study them!

Desire lines in process models



Aanvraag registratie complete
592
592
10587
Rapportage & beschikking complete
589
10762
10920
Retour complete
34
10920
10617
10713
Toetsen en beslissen complete
611
10920
11009
10920
Verzending\dossiervorming complete
567
79
Wachten terugmelding zorgaanb.

Data explosion



Source: UC Berkeley Research Project, How Much Information?, 2003;
IDC, Disk Storage System Quarterly Tracker (as of 2006)

Process Mining =

Event Data + Processes

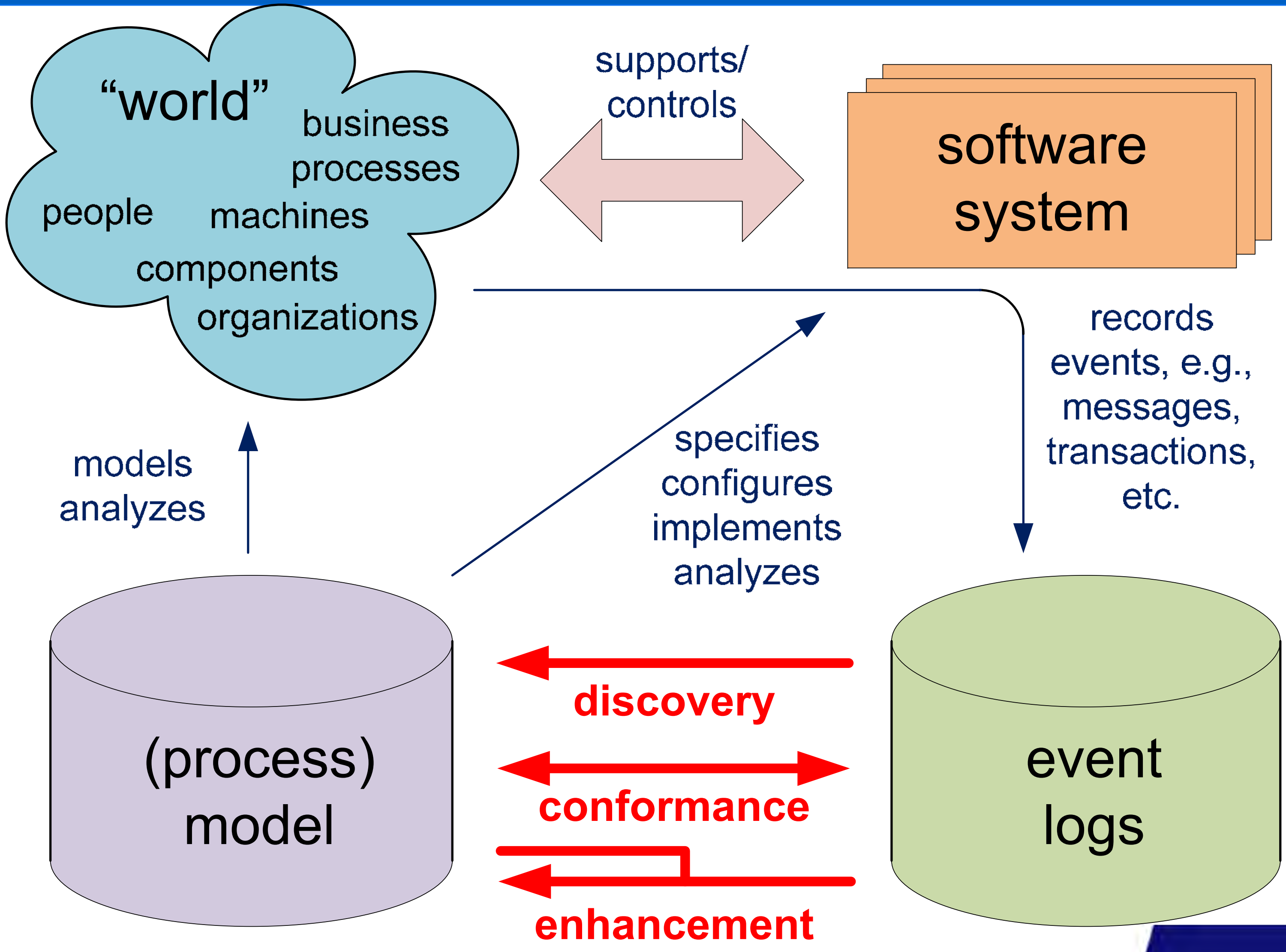
Data Mining + Process Analysis

Machine Learning + Formal Methods

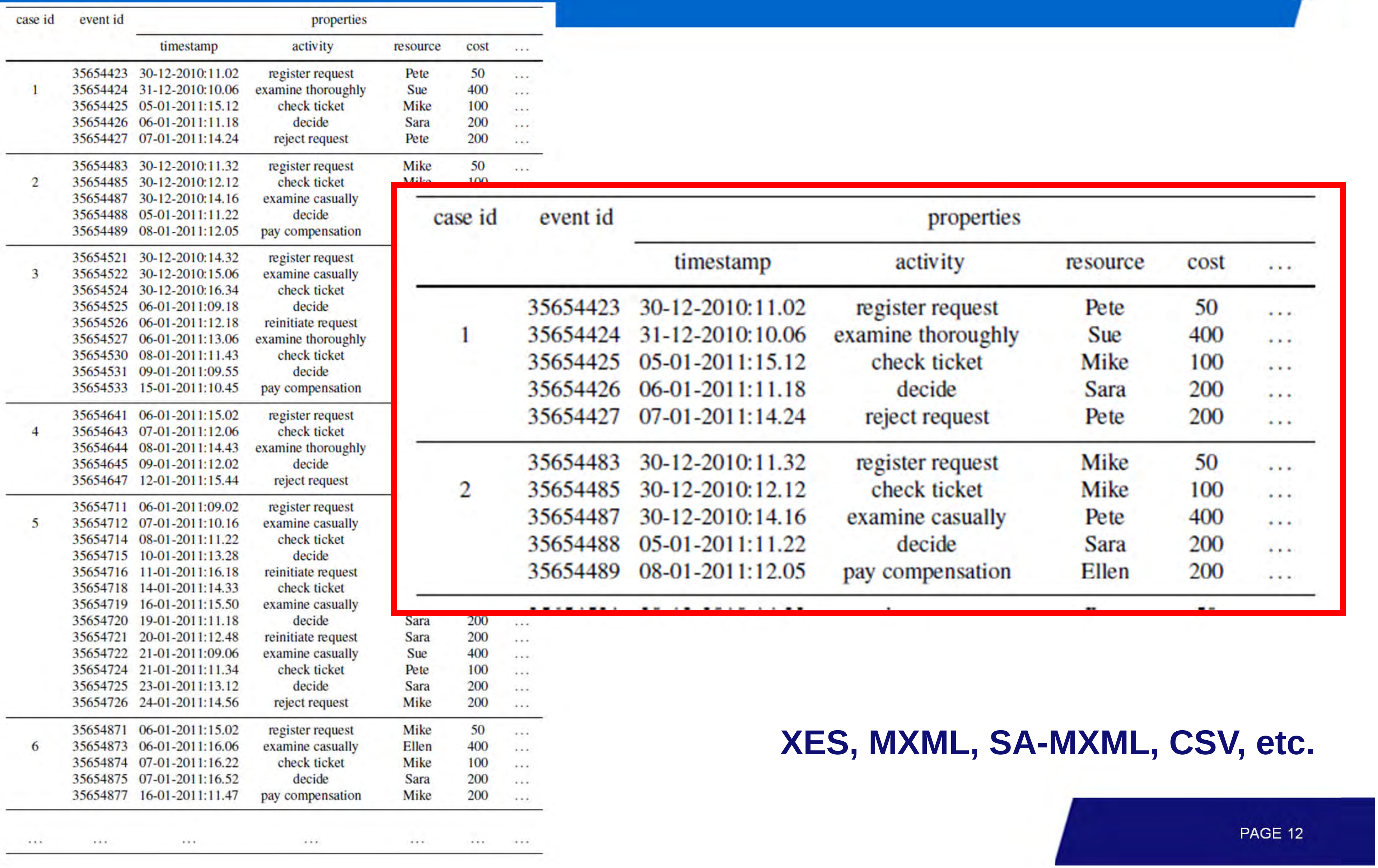
Process Mining

- Process discovery: "What is really happening?"
- Conformance checking: "Do we do what was agreed upon?"
- Performance analysis: "Where are the bottlenecks?"
- Process prediction: "Will this case be late?"
- Process improvement: "How to redesign this process?"
- Etc.

Process Mining



Starting point: event log

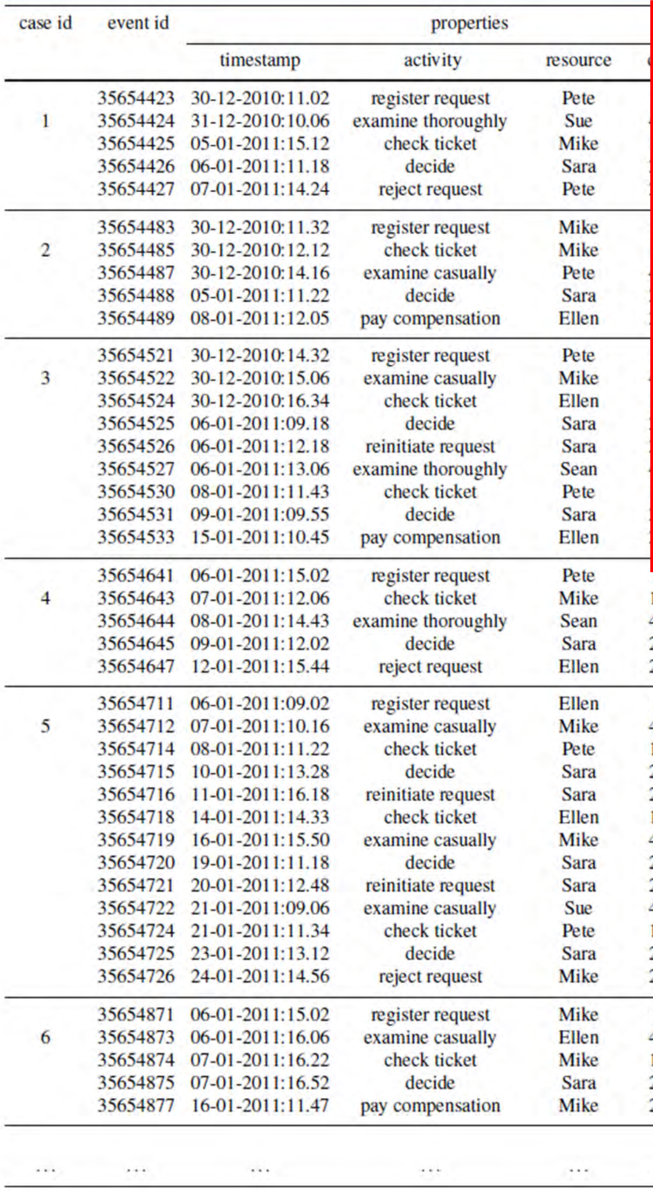


case id	event id	properties				
		timestamp	activity	resource	cost	…
1	35654423	30-12-2010:11.02	register request	Pete	50	…
	35654424	31-12-2010:10.06	examine thoroughly	Sue	400	…
	35654425	05-01-2011:15.12	check ticket	Mike	100	…
	35654426	06-01-2011:11.18	decide	Sara	200	…
	35654427	07-01-2011:14.24	reject request	Pete	200	…
2	35654483	30-12-2010:11.32	register request	Mike	50	…
	35654485	30-12-2010:12.12	check ticket	Mike	100	
	35654487	30-12-2010:14.16	examine casually			
	35654488	05-01-2011:11.22	decide			
	35654489	08-01-2011:12.05	pay compensation			
3	35654521	30-12-2010:14.32	register request			
	35654522	30-12-2010:15.06	examine casually			
	35654524	30-12-2010:16.34	check ticket			
	35654525	06-01-2011:09.18	decide			
	35654526	06-01-2011:12.18	reinitiate request			
	35654527	06-01-2011:13.06	examine thoroughly			
	35654530	08-01-2011:11.43	check ticket			
	35654531	09-01-2011:09.55	decide			
	35654533	15-01-2011:10.45	pay compensation			
4	35654641	06-01-2011:15.02	register request			
	35654643	07-01-2011:12.06	check ticket			
	35654644	08-01-2011:14.43	examine thoroughly			
	35654645	09-01-2011:12.02	decide			
	35654647	12-01-2011:15.44	reject request			
5	35654711	06-01-2011:09.02	register request			
	35654712	07-01-2011:10.16	examine casually			
	35654714	08-01-2011:11.22	check ticket			
	35654715	10-01-2011:13.28	decide			
	35654716	11-01-2011:16.18	reinitiate request			
	35654718	14-01-2011:14.33	check ticket			
	35654719	16-01-2011:15.50	examine casually			
	35654720	19-01-2011:11.18	decide	Sara	200	…
	35654721	20-01-2011:12.48	reinitiate request	Sara	200	…
	35654722	21-01-2011:09.06	examine casually	Sue	400	…
	35654724	21-01-2011:11.34	check ticket	Pete	100	…
	35654725	23-01-2011:13.12	decide	Sara	200	…
	35654726	24-01-2011:14.56	reject request	Mike	200	…
6	35654871	06-01-2011:15.02	register request	Mike	50	…
	35654873	06-01-2011:16.06	examine casually	Ellen	400	…
	35654874	07-01-2011:16.22	check ticket	Mike	100	…
	35654875	07-01-2011:16.52	decide	Sara	200	…
	35654877	16-01-2011:11.47	pay compensation	Mike	200	…
…	…	…	…	…	…	…

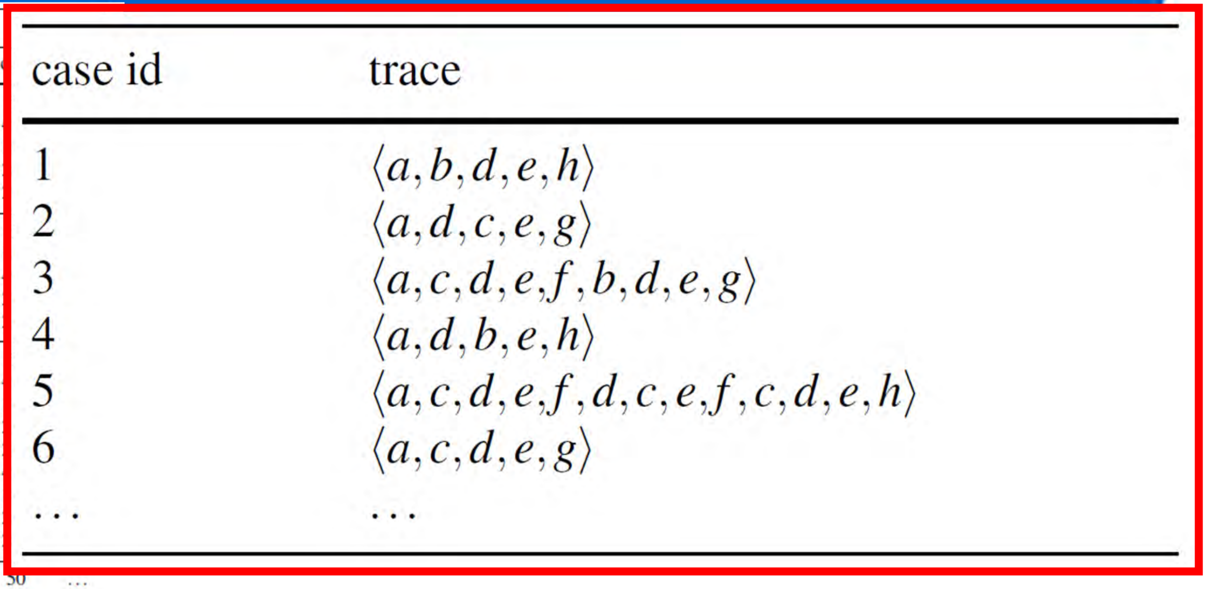
case id	event id	properties				
		timestamp	activity	resource	cost	…
1	35654423	30-12-2010:11.02	register request	Pete	50	…
	35654424	31-12-2010:10.06	examine thoroughly	Sue	400	…
	35654425	05-01-2011:15.12	check ticket	Mike	100	…
	35654426	06-01-2011:11.18	decide	Sara	200	…
	35654427	07-01-2011:14.24	reject request	Pete	200	…
2	35654483	30-12-2010:11.32	register request	Mike	50	…
	35654485	30-12-2010:12.12	check ticket	Mike	100	…
	35654487	30-12-2010:14.16	examine casually	Pete	400	…
	35654488	05-01-2011:11.22	decide	Sara	200	…
	35654489	08-01-2011:12.05	pay compensation	Ellen	200	…

XES, MXML, SA-MXML, CSV, etc.

Simplified event log



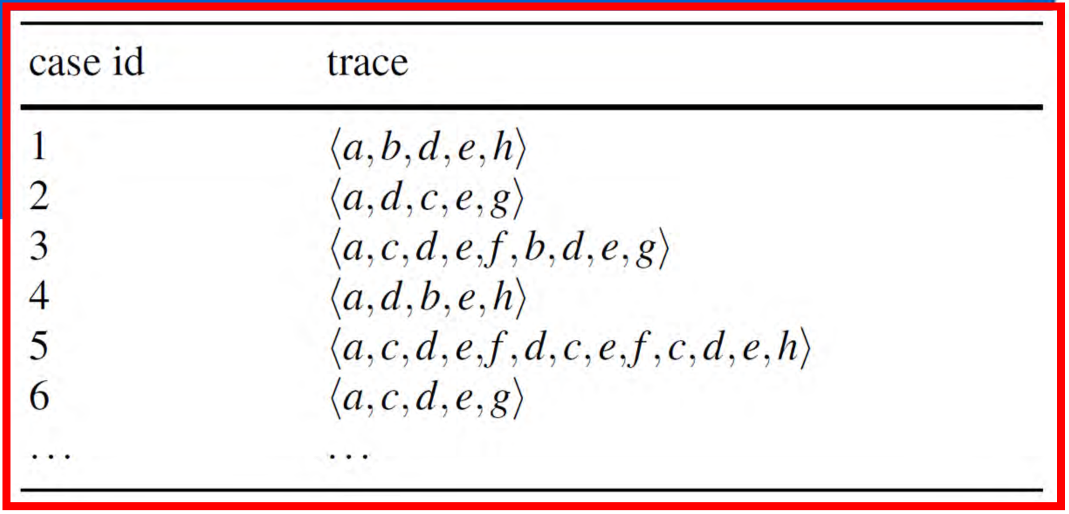
case id	event id	properties				
		timestamp	activity	resource		
1	35654423	30-12-2010:11.02	register request	Pete		
	35654424	31-12-2010:10.06	examine thoroughly	Sue		
	35654425	05-01-2011:15.12	check ticket	Mike		
	35654426	06-01-2011:11.18	decide	Sara		
	35654427	07-01-2011:14.24	reject request	Pete		
2	35654483	30-12-2010:11.32	register request	Mike		
	35654485	30-12-2010:12.12	check ticket	Mike		
	35654487	30-12-2010:14.16	examine casually	Pete		
	35654488	05-01-2011:11.22	decide	Sara		
	35654489	08-01-2011:12.05	pay compensation	Ellen		
3	35654521	30-12-2010:14.32	register request	Pete		
	35654522	30-12-2010:15.06	examine casually	Mike		
	35654524	30-12-2010:16.34	check ticket	Ellen		
	35654525	06-01-2011:09.18	decide	Sara		
	35654526	06-01-2011:12.18	reinitiate request	Sara		
	35654527	06-01-2011:13.06	examine thoroughly	Sean		
	35654530	08-01-2011:11.43	check ticket	Pete		
	35654531	09-01-2011:09.55	decide	Sara		
	35654533	15-01-2011:10.45	pay compensation	Ellen		
4	35654641	06-01-2011:15.02	register request	Pete	50	...
	35654643	07-01-2011:12.06	check ticket	Mike	100	...
	35654644	08-01-2011:14.43	examine thoroughly	Sean	400	...
	35654645	09-01-2011:12.02	decide	Sara	200	...
	35654647	12-01-2011:15.44	reject request	Ellen	200	...
5	35654711	06-01-2011:09.02	register request	Ellen	50	...
	35654712	07-01-2011:10.16	examine casually	Mike	400	...
	35654714	08-01-2011:11.22	check ticket	Pete	100	...
	35654715	10-01-2011:13.28	decide	Sara	200	...
	35654716	11-01-2011:16.18	reinitiate request	Sara	200	...
	35654718	14-01-2011:14.33	check ticket	Ellen	100	...
	35654719	16-01-2011:15.50	examine casually	Mike	400	...
	35654720	19-01-2011:11.18	decide	Sara	200	...
	35654721	20-01-2011:12.48	reinitiate request	Sara	200	...
	35654722	21-01-2011:09.06	examine casually	Sue	400	...
	35654724	21-01-2011:11.34	check ticket	Pete	100	...
	35654725	23-01-2011:13.12	decide	Sara	200	...
	35654726	24-01-2011:14.56	reject request	Mike	200	...
6	35654871	06-01-2011:15.02	register request	Mike	50	...
	35654873	06-01-2011:16.06	examine casually	Ellen	400	...
	35654874	07-01-2011:16.22	check ticket	Mike	100	...
	35654875	07-01-2011:16.52	decide	Sara	200	...
	35654877	16-01-2011:11.47	pay compensation	Mike	200	...
...	...	...	...	...	...	...



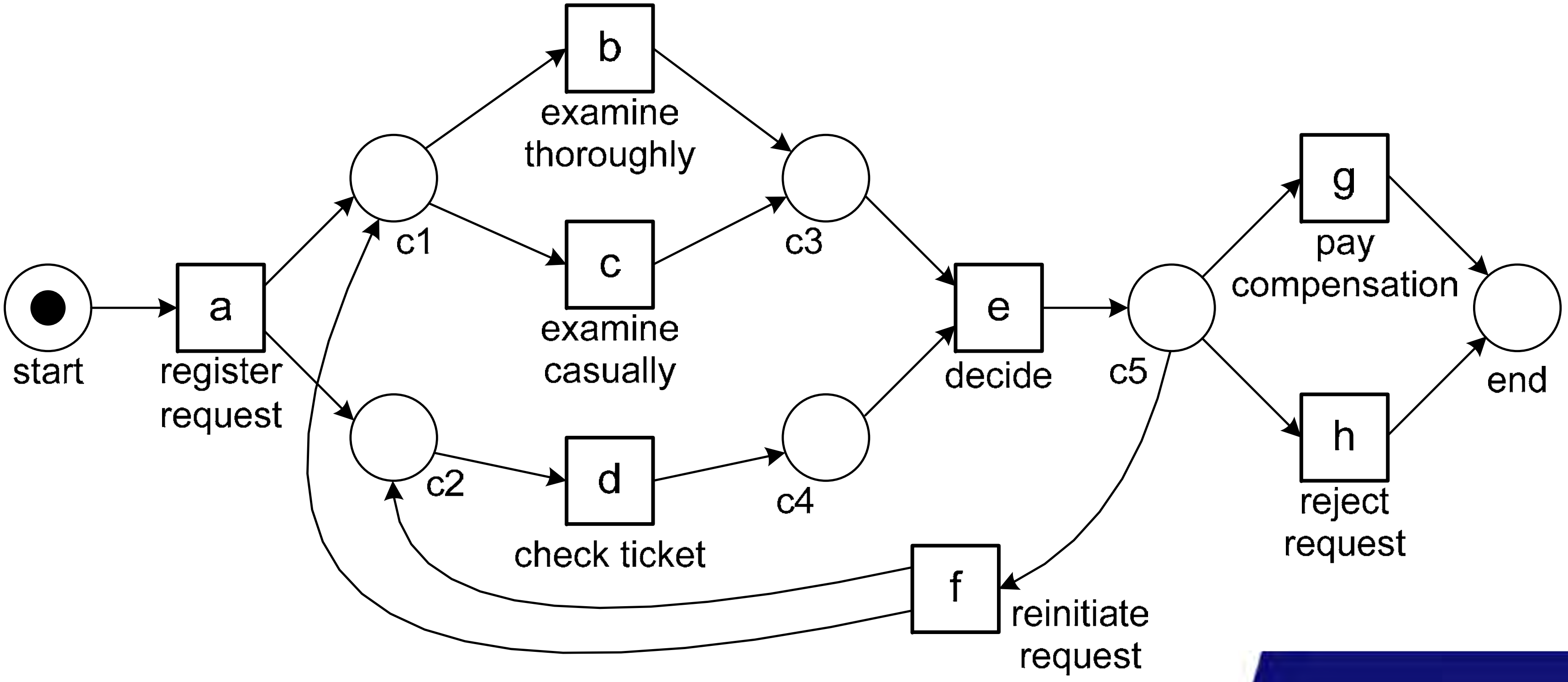
case id	trace
1	$\langle a,b,d,e,h \rangle$
2	$\langle a,d,c,e,g \rangle$
3	$\langle a,c,d,e,f,b,d,e,g \rangle$
4	$\langle a,d,b,e,h \rangle$
5	$\langle a,c,d,e,f,d,c,e,f,c,d,e,h \rangle$
6	$\langle a,c,d,e,g \rangle$
...	...

a = register request,
b = examine thoroughly,
c = examine casually,
d = check ticket,
e = decide,
f = reinitiate request,
g = pay compensation,
and h = reject request

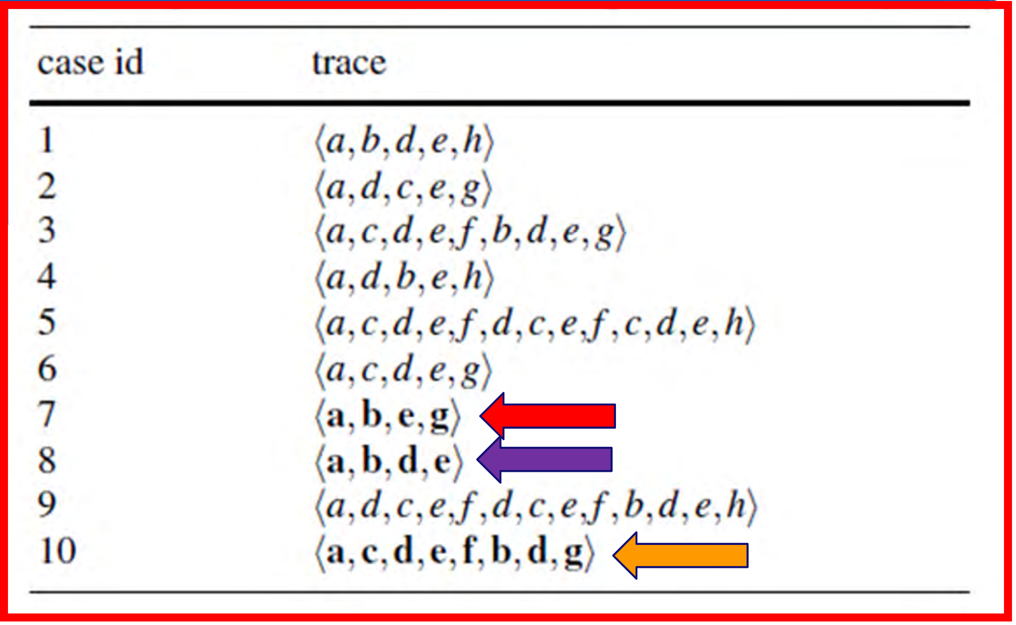
Process discovery



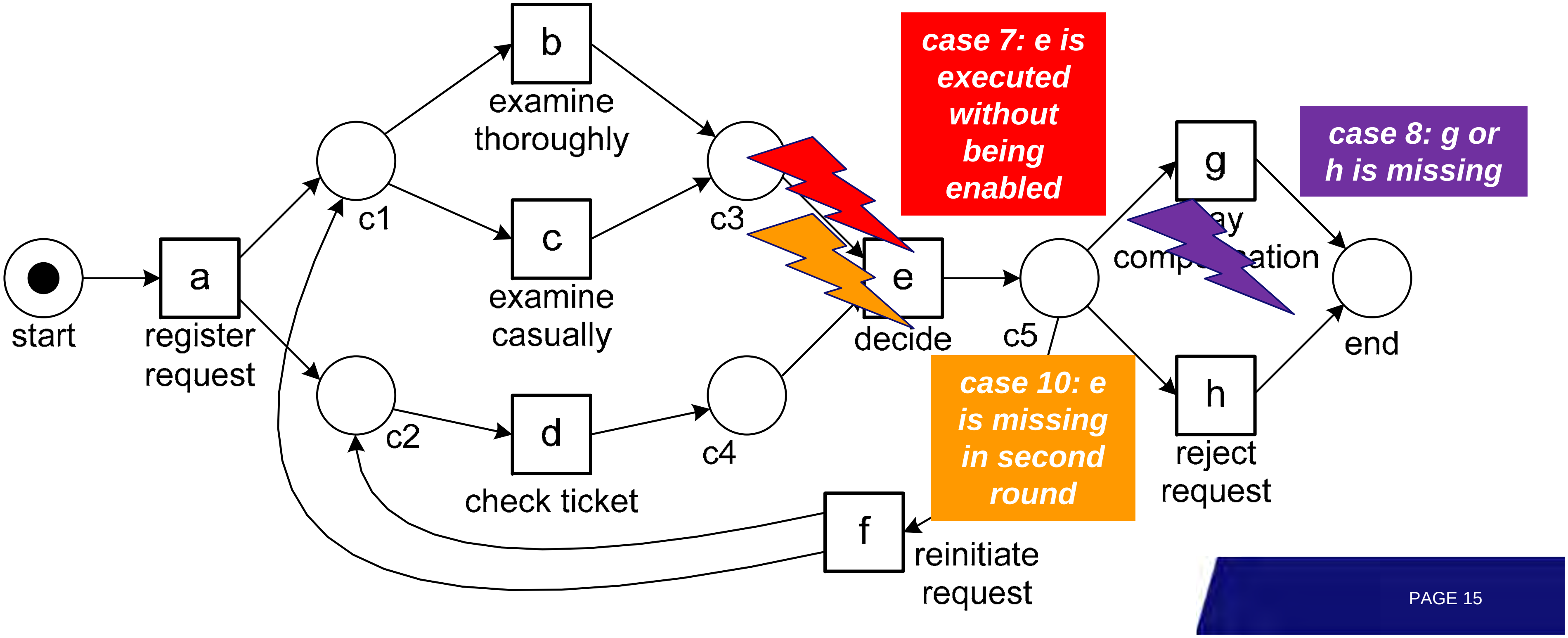
case id	trace
1	$\langle a,b,d,e,h \rangle$
2	$\langle a,d,c,e,g \rangle$
3	$\langle a,c,d,e,f,b,d,e,g \rangle$
4	$\langle a,d,b,e,h \rangle$
5	$\langle a,c,d,e,f,d,c,e,f,c,d,e,h \rangle$
6	$\langle a,c,d,e,g \rangle$
...	...



Conformance checking



case id	trace
1	$\langle a,b,d,e,h \rangle$
2	$\langle a,d,c,e,g \rangle$
3	$\langle a,c,d,e,f,b,d,e,g \rangle$
4	$\langle a,d,b,e,h \rangle$
5	$\langle a,c,d,e,f,d,c,e,f,c,d,e,h \rangle$
6	$\langle a,c,d,e,g \rangle$
7	$\langle \mathbf{a,b,e,g} \rangle$
8	$\langle \mathbf{a,b,d,e} \rangle$
9	$\langle a,d,c,e,f,d,c,e,f,b,d,e,h \rangle$
10	$\langle \mathbf{a,c,d,e,f,b,d,g} \rangle$



case 7: e is executed without being enabled

case 8: g or h is missing

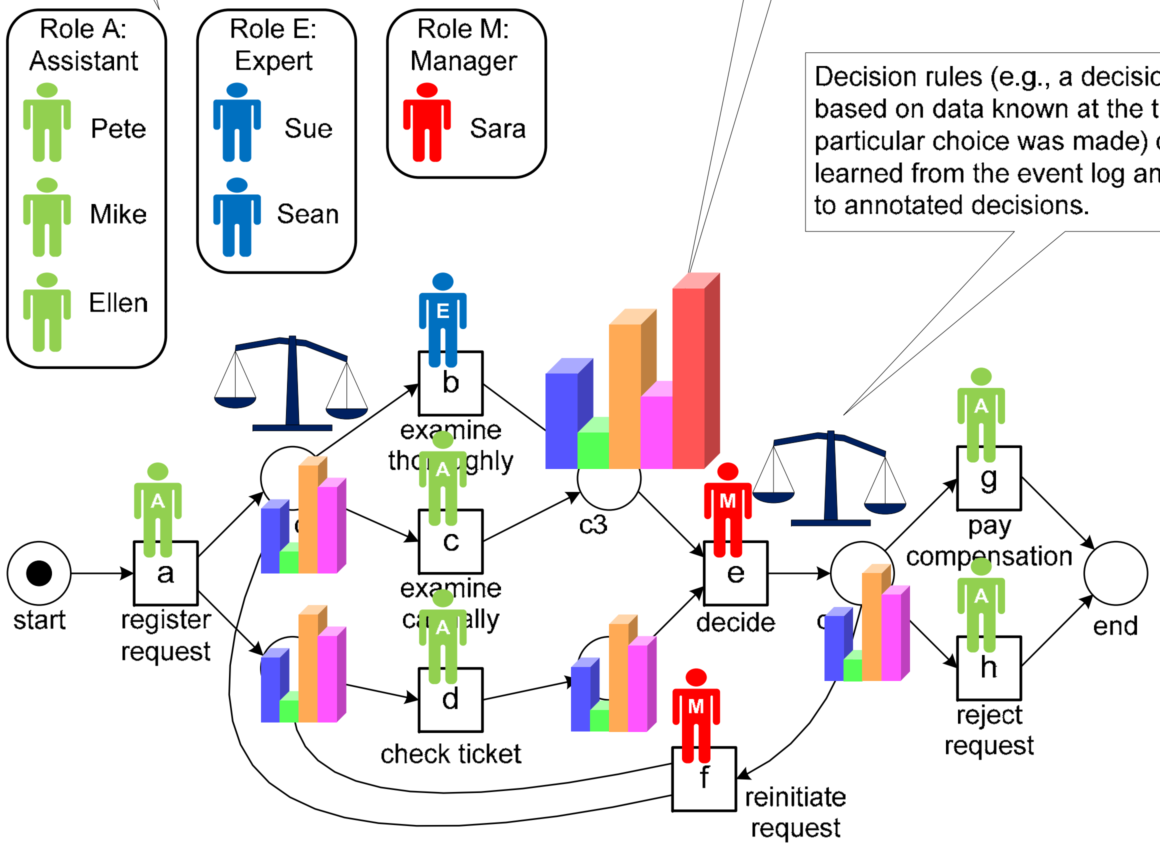
case 10: e is missing in second round

Extension: Adding perspectives to model based on event log

The event log can be used to discover roles in the organization (e.g., groups of people with similar work patterns). These roles can be used to relate individuals and activities.

Performance information (e.g., the average time between two subsequent activities) can be extracted from the event log and visualized on top of the model.

Decision rules (e.g., a decision tree based on data known at the time a particular choice was made) can be learned from the event log and used to annotated decisions.



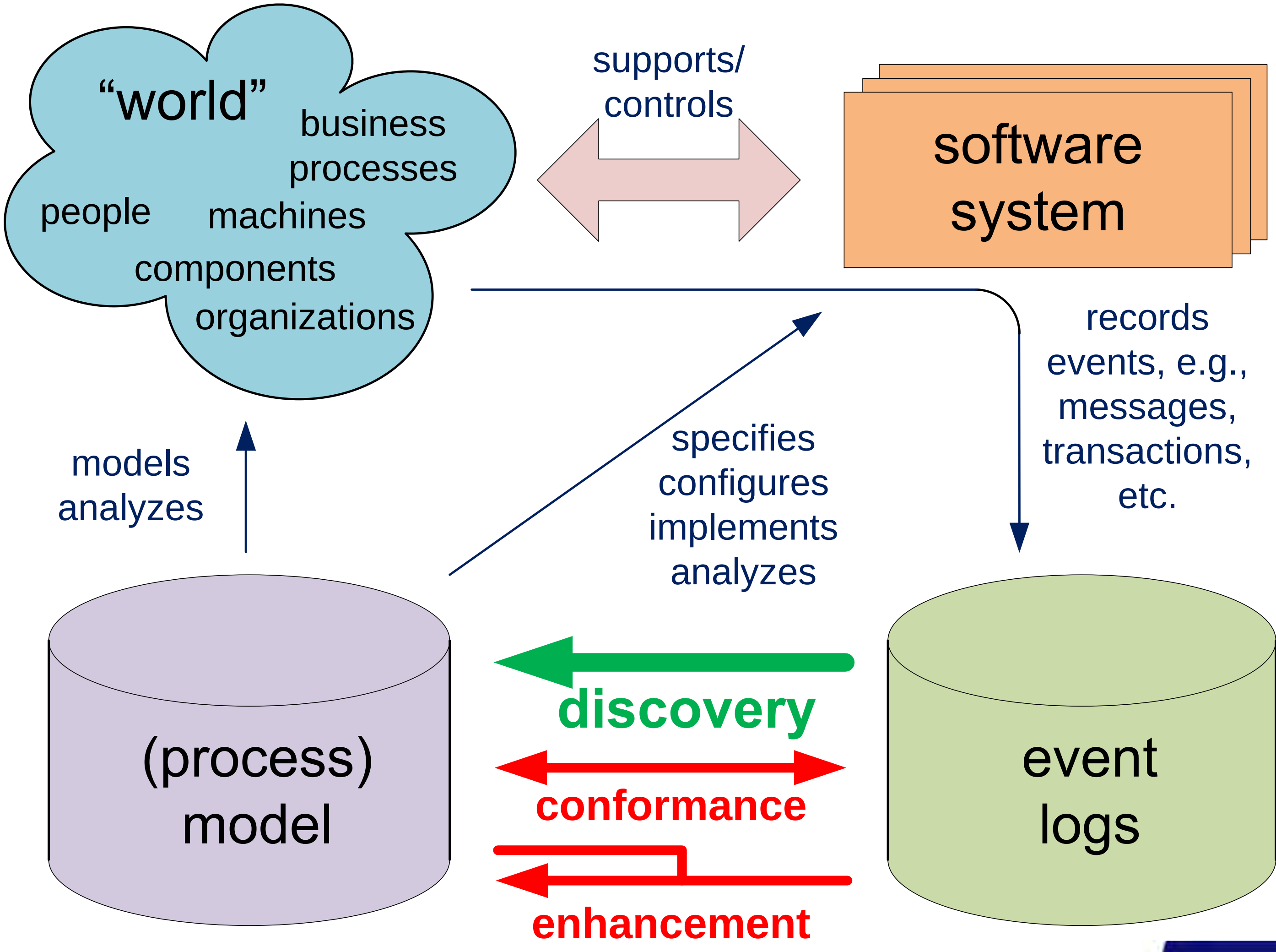
We applied ProM in >100 organizations

- **Municipalities** (e.g., Alkmaar, Heusden, Harderwijk, etc.)
- **Government agencies** (e.g., Rijkswaterstaat, Centraal Justitieel Incasso Bureau, Justice department)
- **Insurance related agencies** (e.g., UWV)
- **Banks** (e.g., ING Bank)
- **Hospitals** (e.g., AMC hospital, Catharina hospital)
- **Multinationals** (e.g., DSM, Deloitte)
- **High-tech system manufacturers and their customers** (e.g., Philips Healthcare, ASML, Ricoh, Thales)
- **Media companies** (e.g. Winkwaves)
- **...**

All supported by …

- Open-source (L-GPL), cf. www.processmining.org
- Plug-in architecture
- Plug-ins cover the whole process mining spectrum and also support classical forms of process analysis

Process discovery



Process Discovery Techniques (small selection)

distributed genetic mining

automata-based learning

language-based regions

heuristic mining

state-based regions

partial-order based mining

genetic mining

LTL mining

pattern-based mining

neural networks

stochastic task graphs

fuzzy mining

mining block structures

hidden Markov models

multi-phase mining

α algorithm

conformal process graph

α# algorithm

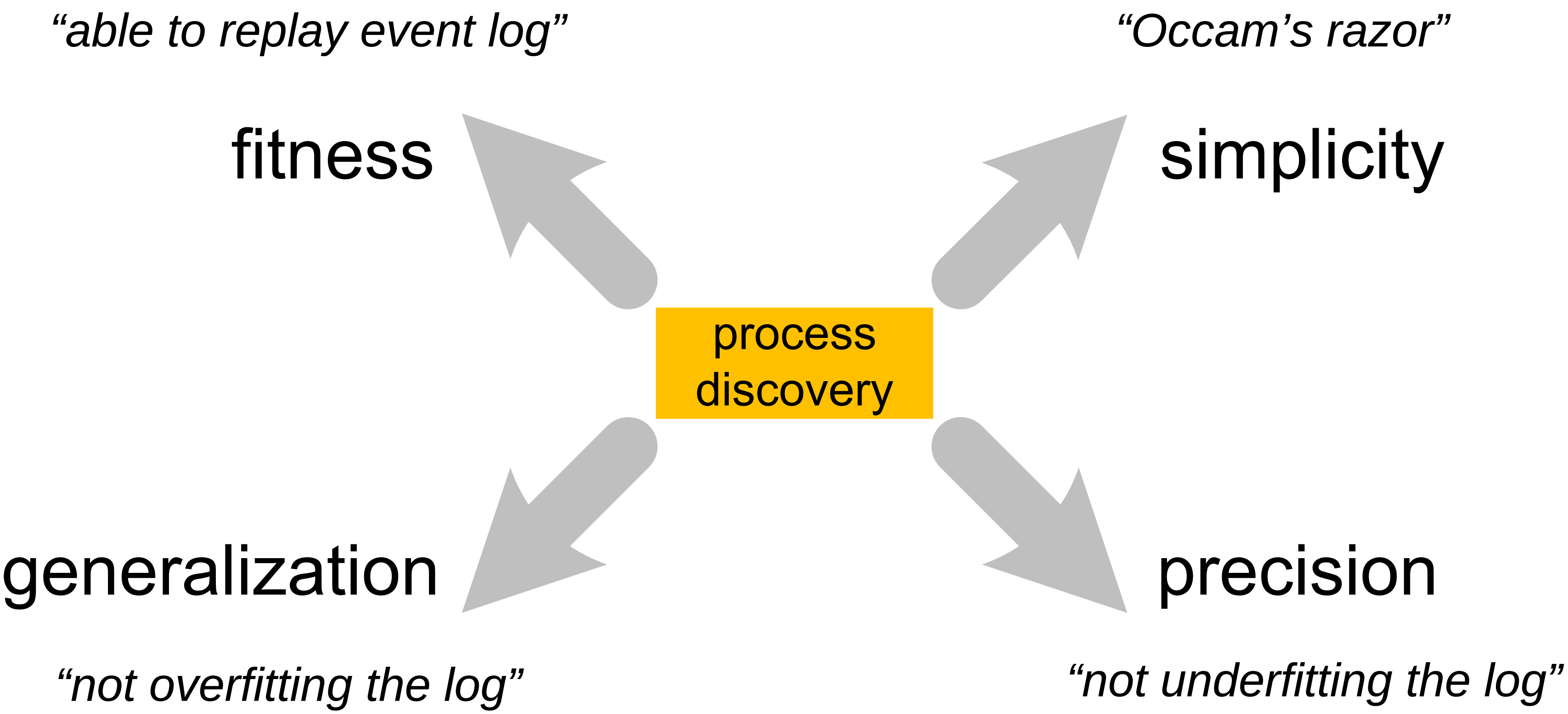
ILP mining

α++ algorithm

Why is process discovery such a difficult problem?

- There are no negative examples (i.e., a log shows what has happened but does not show what could not happen).
- Due to concurrency, loops, and choices the search space has a complex structure and the log typically contains only a fraction of all possible behaviors.
- There is no clear relation between the size of a model and its behavior (i.e., a smaller model may generate more or less behavior although classical analysis and evaluation methods typically assume some monotonicity property).

Challenge: four competing quality criteria



“able to replay event log”

fitness

“Occam’s razor”

simplicity

process discovery

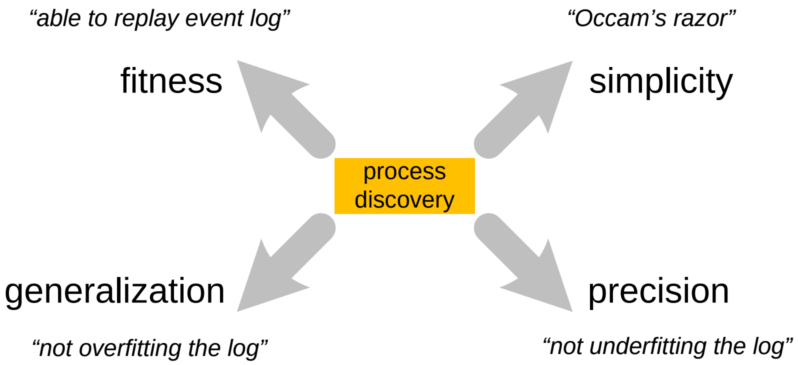
generalization

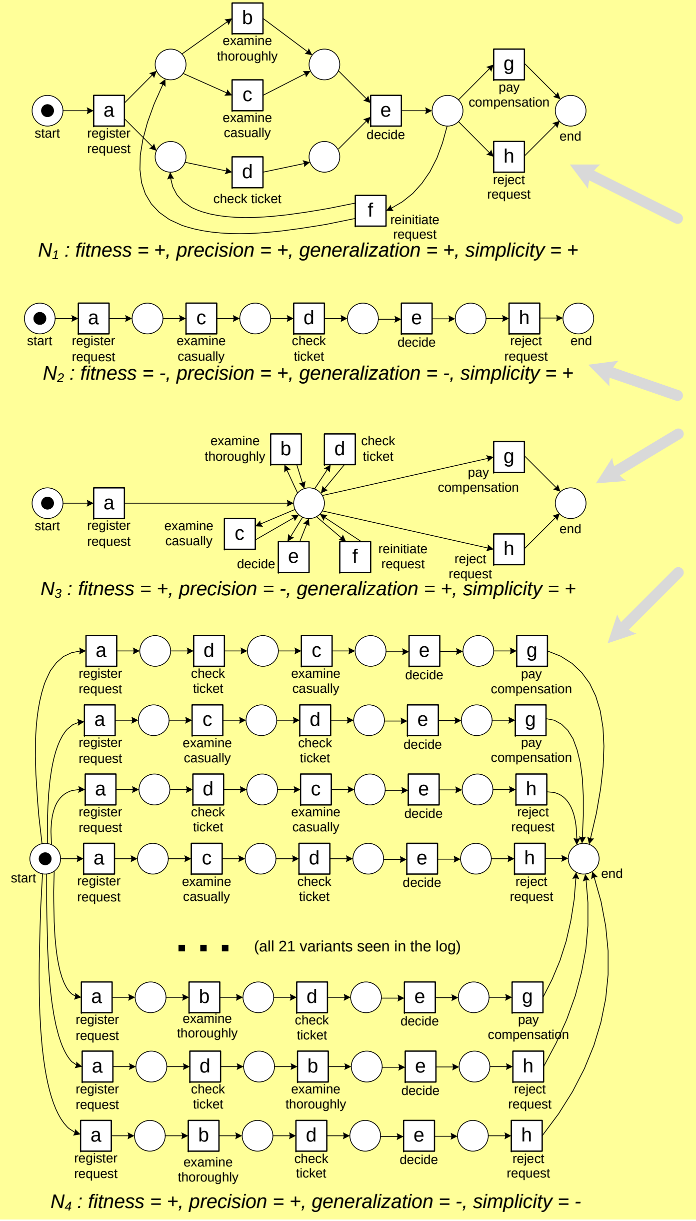
“not overfitting the log”

precision

“not underfitting the log”

Example: one log four models


“able to replay event log”
fitness
“Occam’s razor”
simplicity
process discovery
generalization
“not overfitting the log”
precision
“not underfitting the log”



N_1 : *fitness = +, precision = +, generalization = +, simplicity = +*

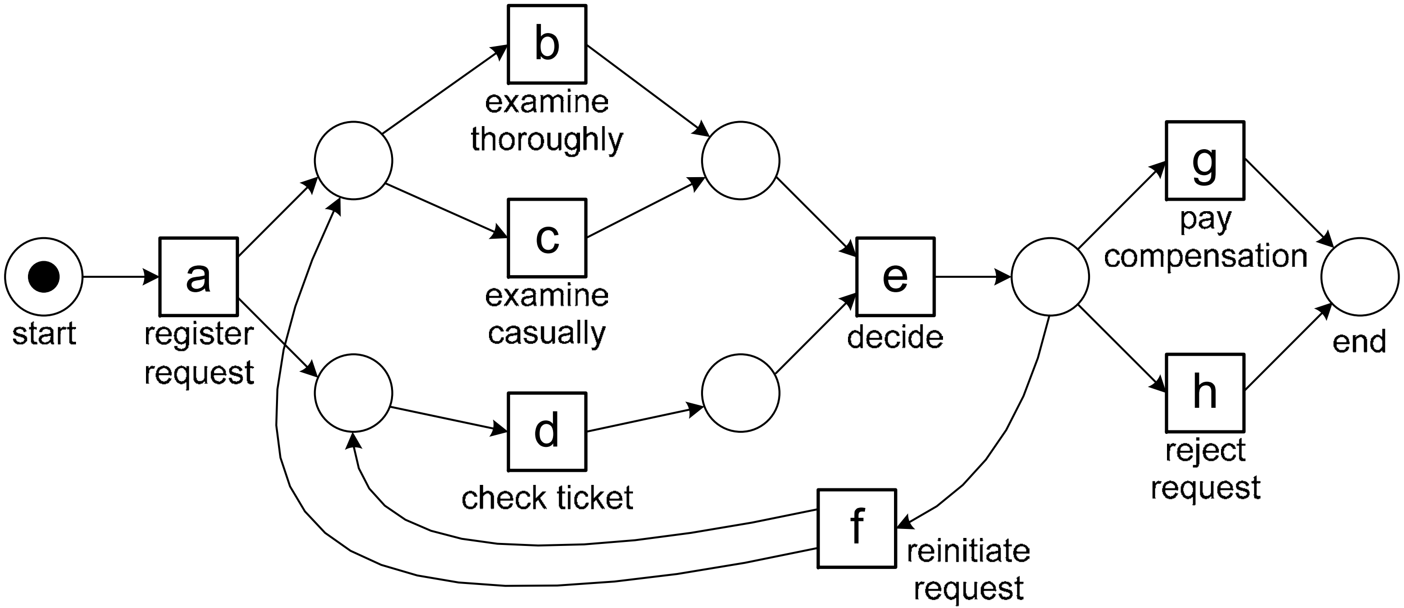
N_2 : *fitness = -, precision = +, generalization = -, simplicity = +*

N_3 : *fitness = +, precision = -, generalization = +, simplicity = +*

N_4 : *fitness = +, precision = +, generalization = -, simplicity = -*

#	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	

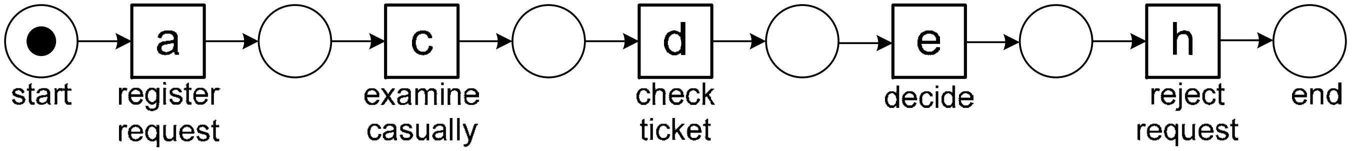
Model N_1



N_1 : fitness = +, precision = +, generalization = +, simplicity = +

#	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	

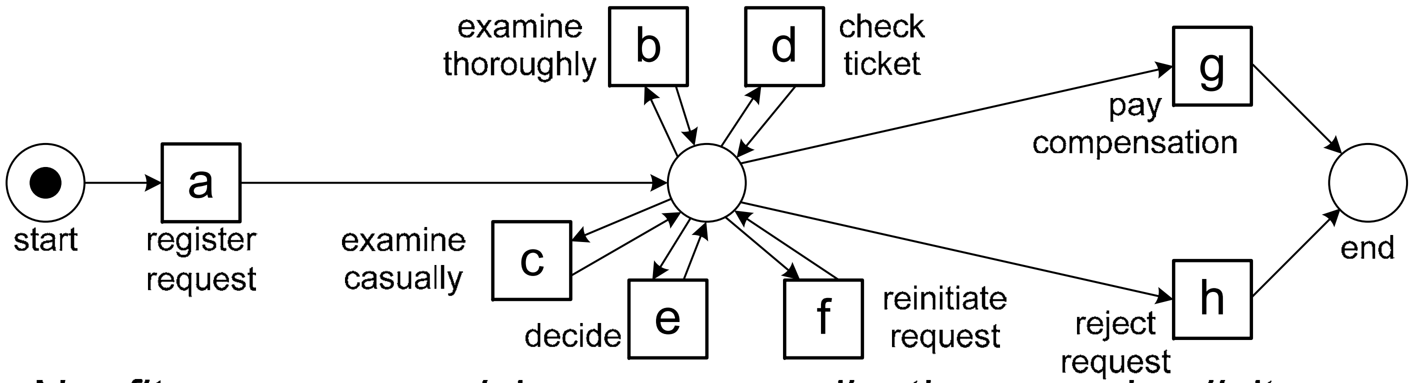
Model N_2



N_2 : *fitness* = -, *precision* = +, *generalization* = -, *simplicity* = +

#	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	

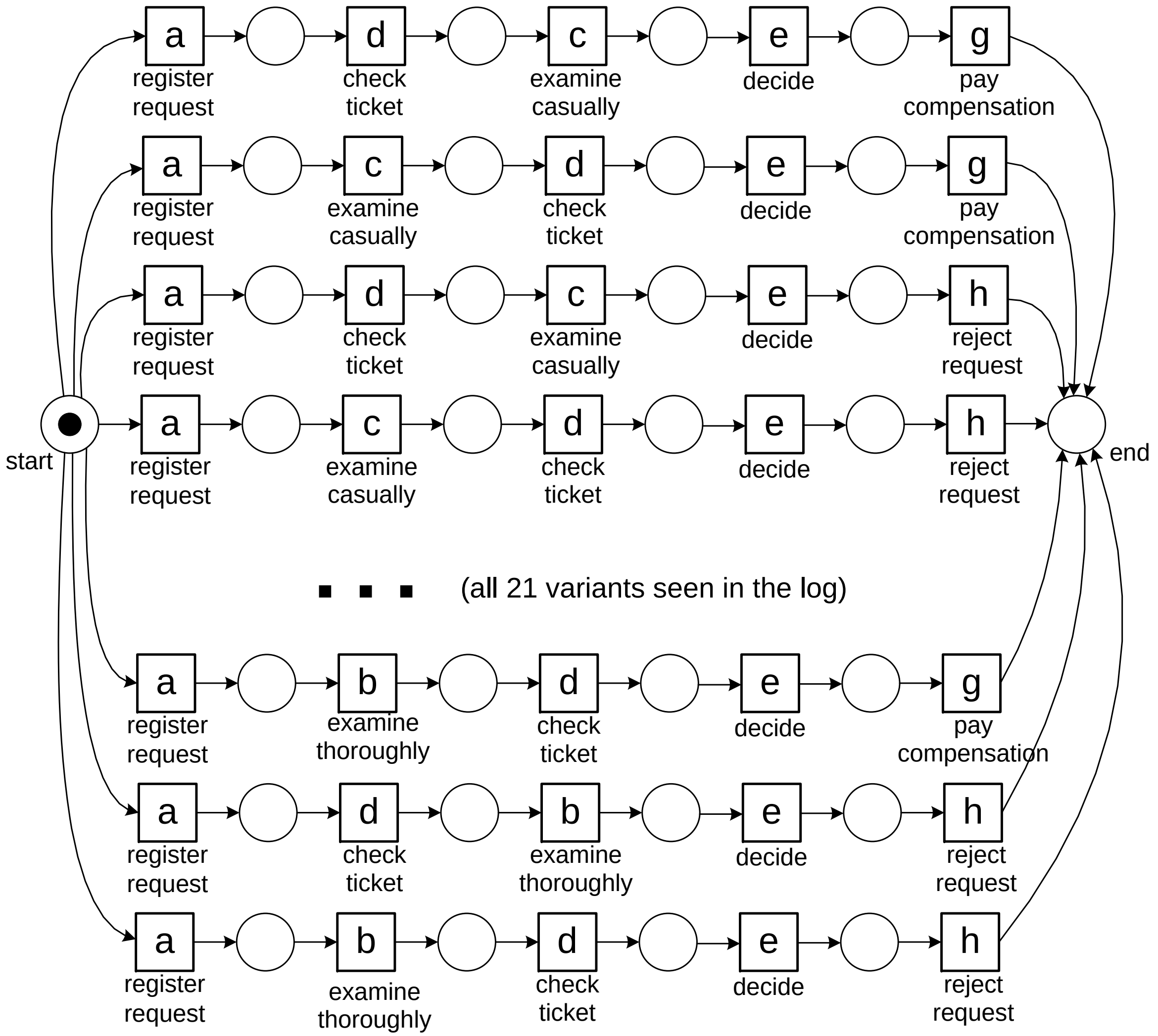
Model N_3



N_3 : fitness = +, precision = -, generalization = +, simplicity = +

#	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	

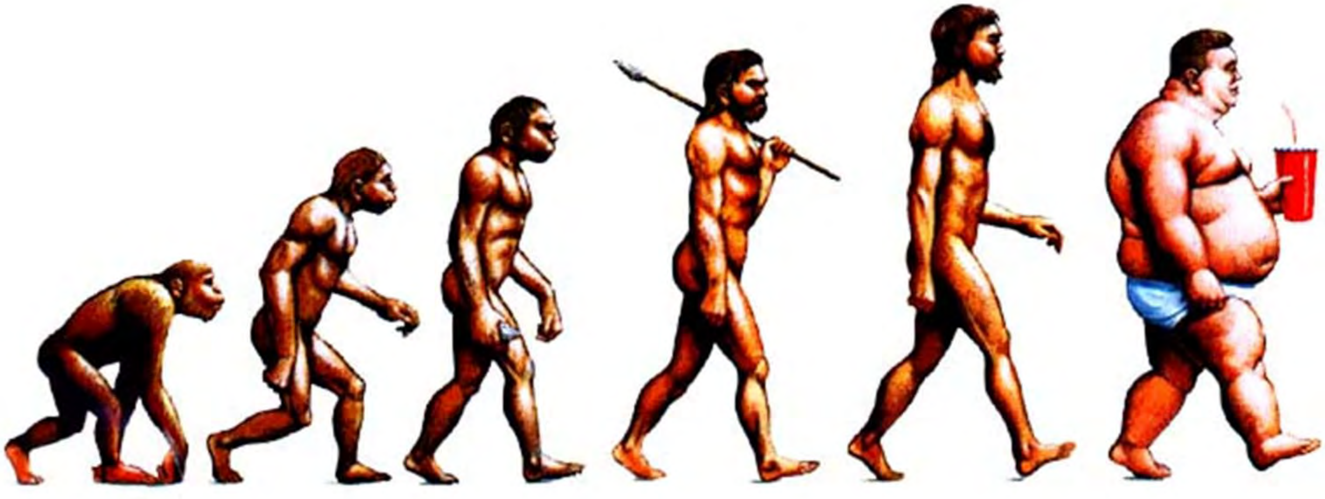
Model N_4



N_4 : fitness = +, precision = +, generalization = -, simplicity = -

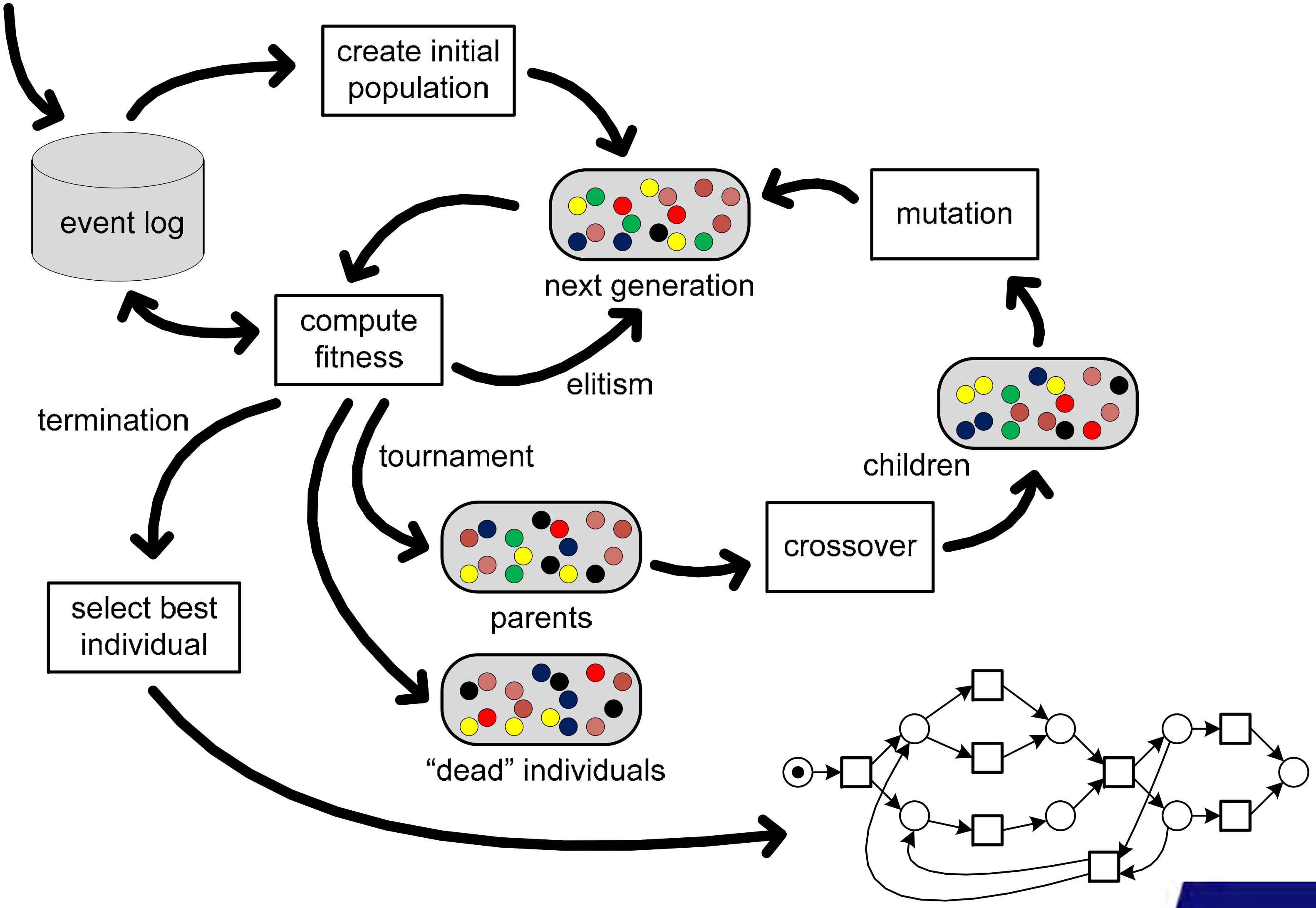
#	trace
455	acdeh
191	abdeg
177	adceh
144	abdeh
111	acdeg
82	adceg
56	adbeh
47	acdefdbeh
38	adbeg
33	acdefbdeh
14	acdefbdeg
11	acdefdbeg
9	adcefcdeh
8	adcefdbeh
5	adcefbdeg
3	acdefbdefdbeg
2	adcefdbeg
2	adcefbdefbdeg
1	adcefdbefbdeh
1	adbefbdefdbeg
1	adcefdbefcdefdbeg
1391	

Example of a process discovery technique: Genetic Mining

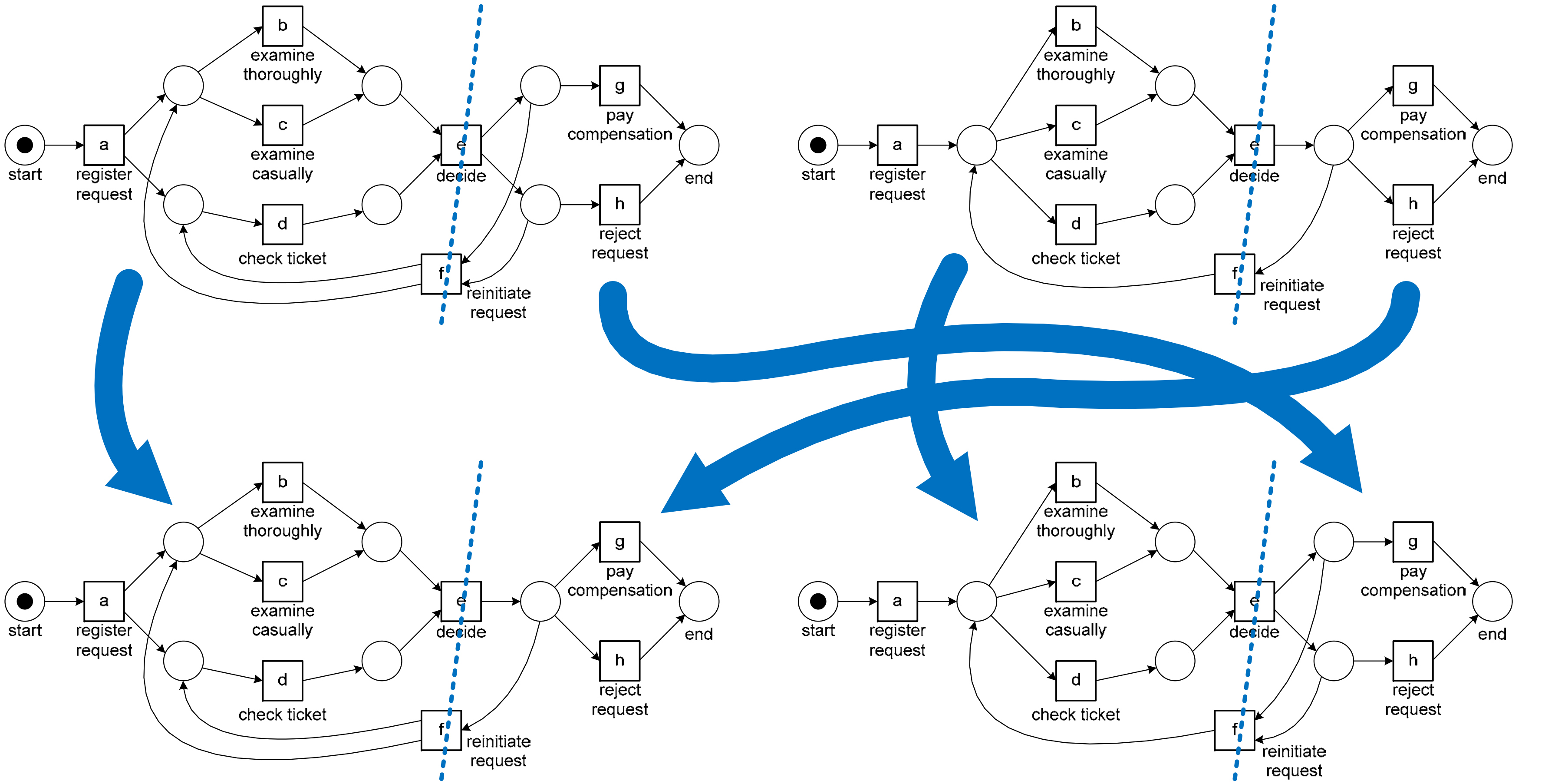


- **Characteristics**
  - **requires a lot of computing power, but can be distributed easily,**
  - **can deal with noise, infrequent behavior, duplicate tasks, invisible tasks,**
  - **allows for incremental improvement and combinations with other approaches (heuristics post-optimization, etc.).**

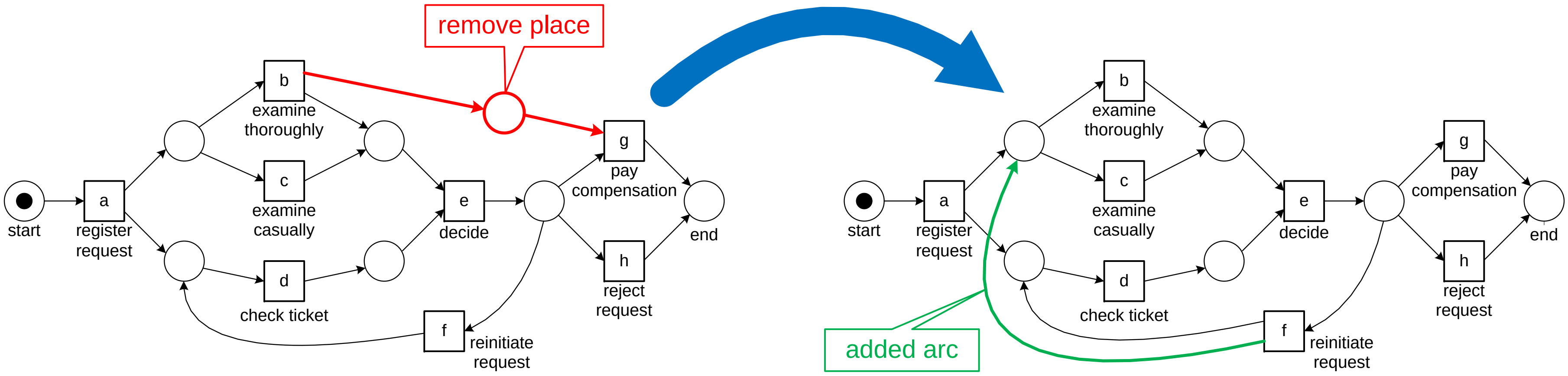
Genetic process mining: Overview



Example: crossover



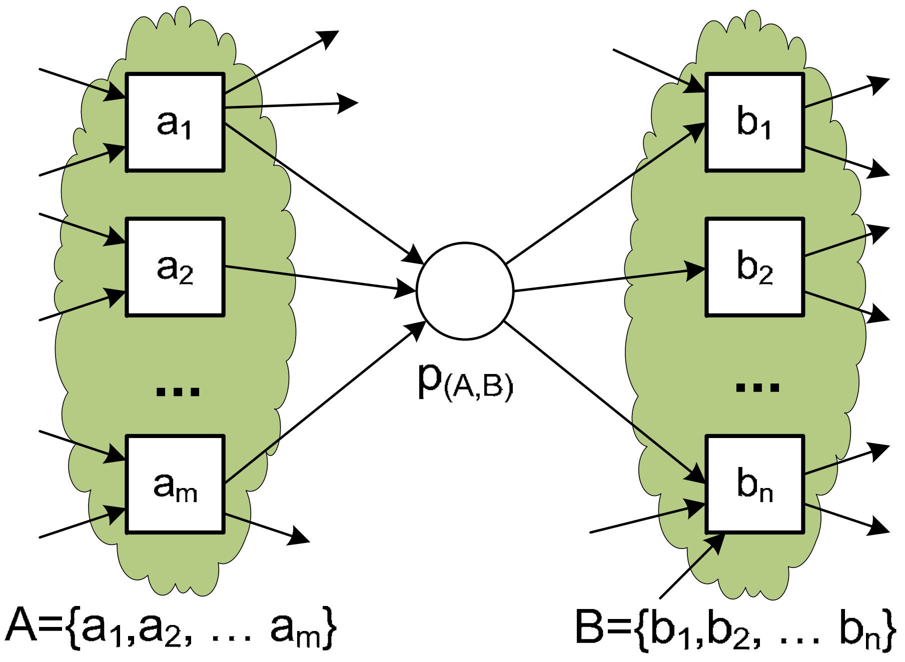
Example: mutation


remove place

added arc

Example of a process discovery technique: Theory of Regions

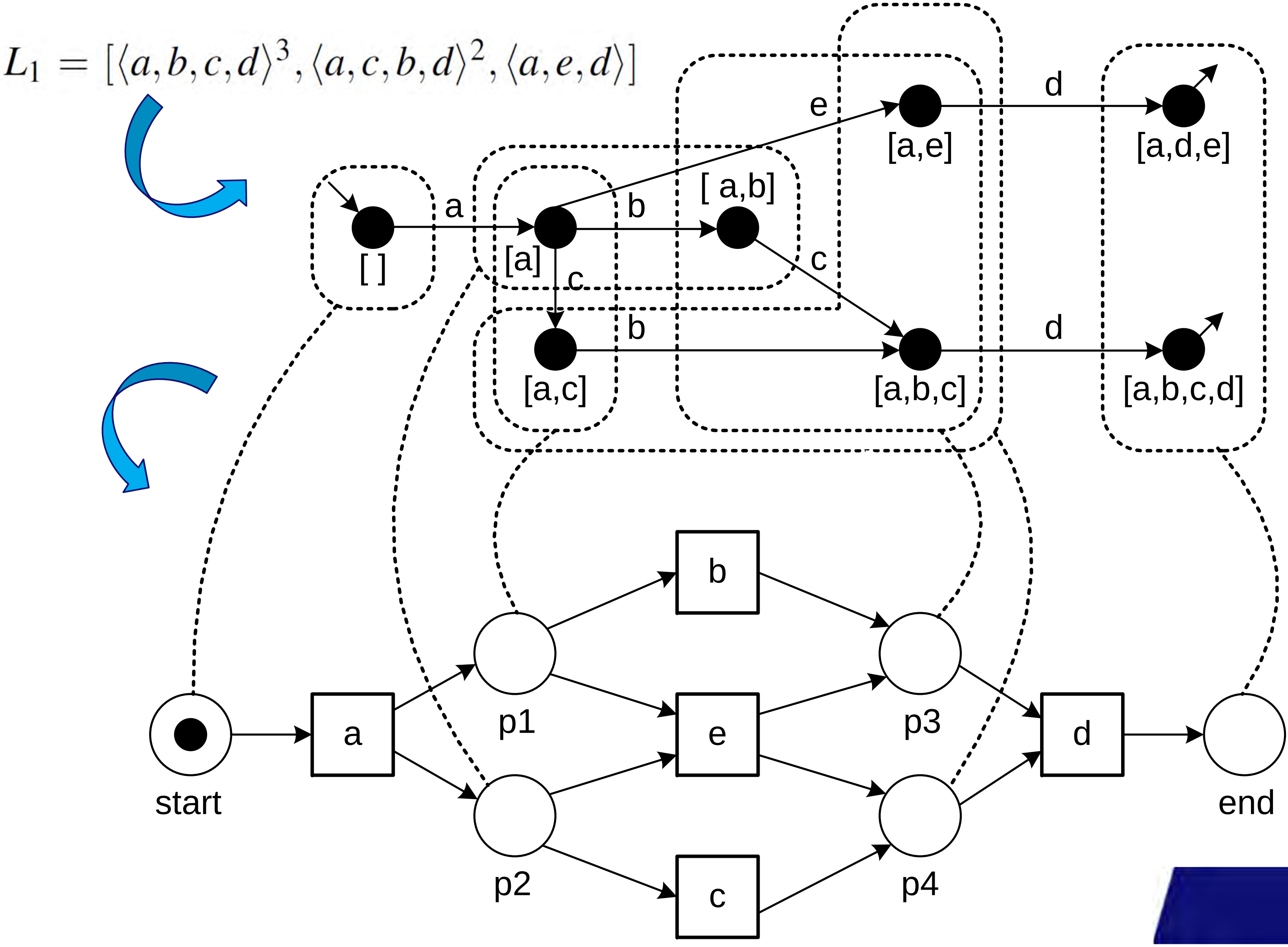
- Two types of regions theory:
  - State-based regions
  - Language-based regions



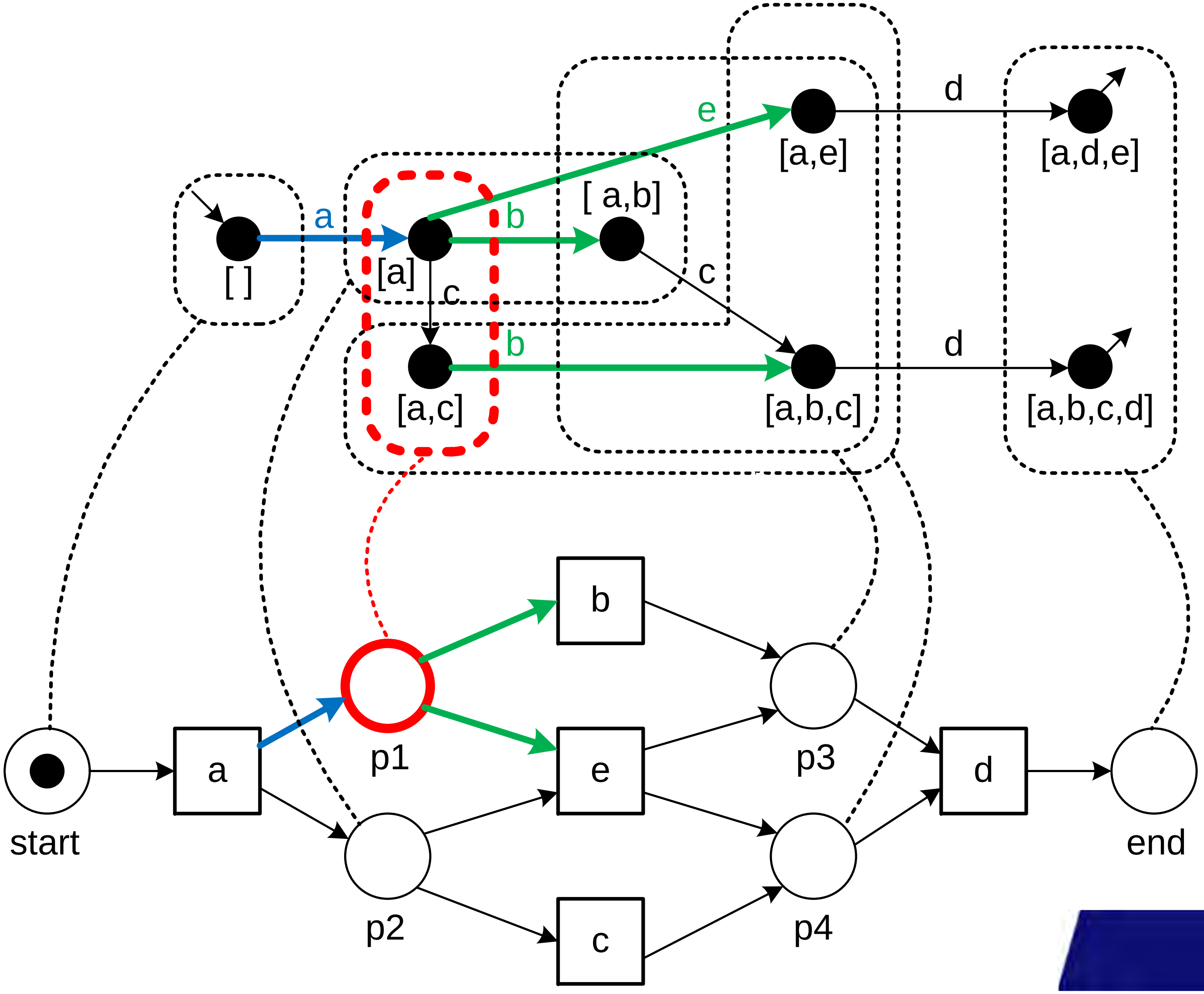
All about discovering places!

State-based regions: Two step approach

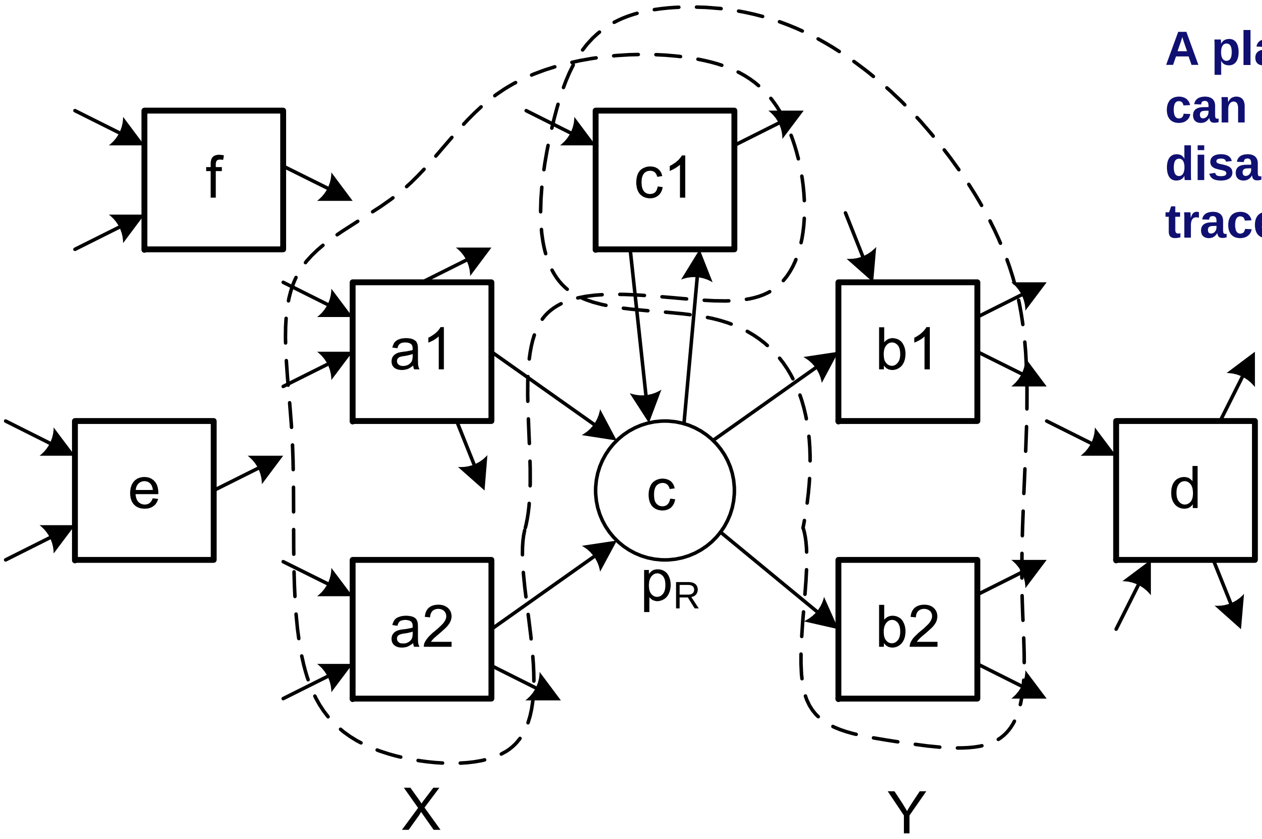
$L_1 = [\langle a,b,c,d\rangle^3, \langle a,c,b,d\rangle^2, \langle a,e,d\rangle]$



Example region
a enters, b and e exit, c and d do not cross



Language-based regions

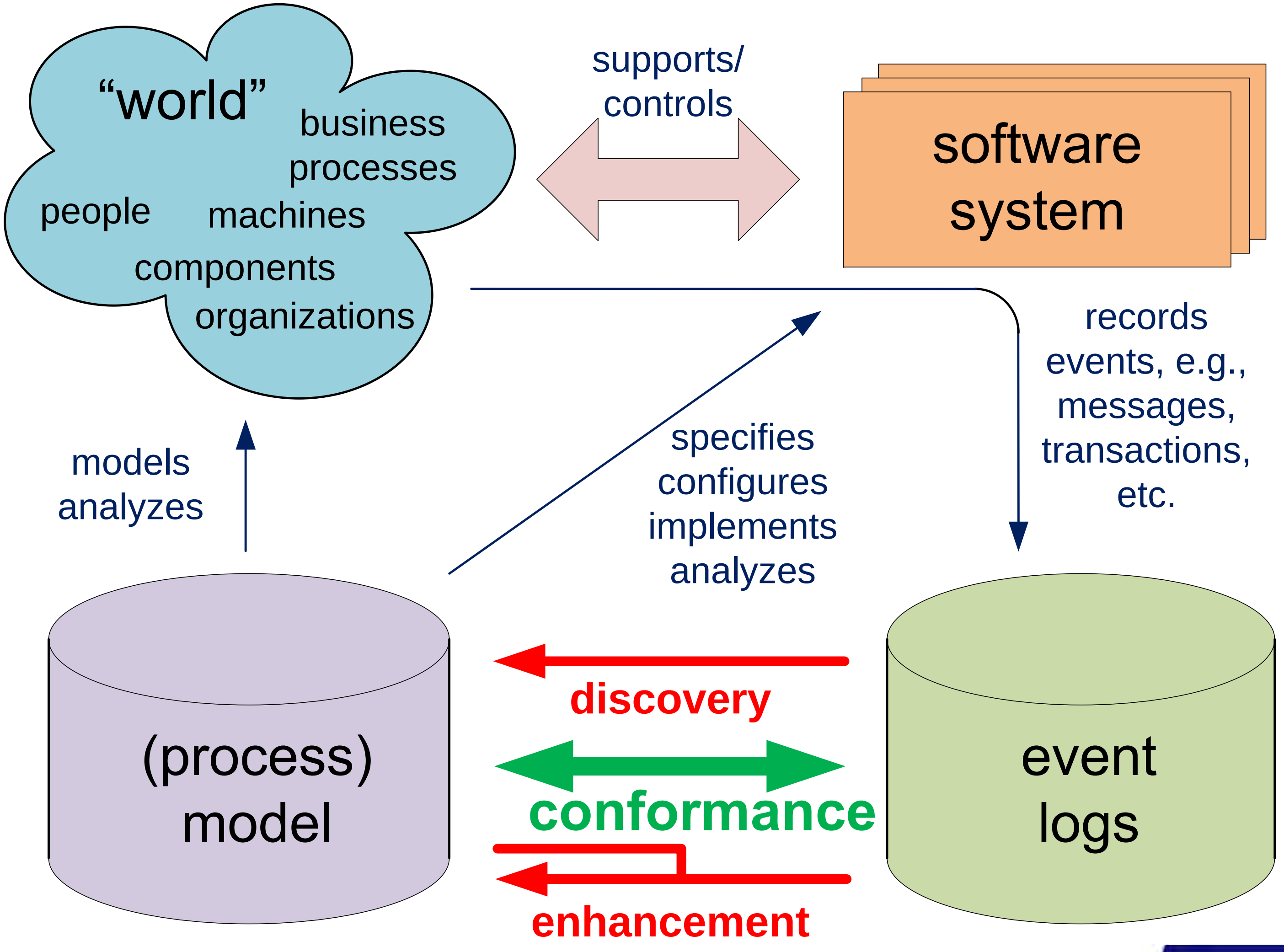


A place is feasible if it can be added without disabling any of the traces in the event log.

for any $\sigma \in L, k \in \{1, \ldots, |\sigma|\}, \sigma_1 = hd^{k-1}(\sigma), a = \sigma(k), \sigma_2 = hd^k(\sigma) = \sigma_1 \oplus a$:

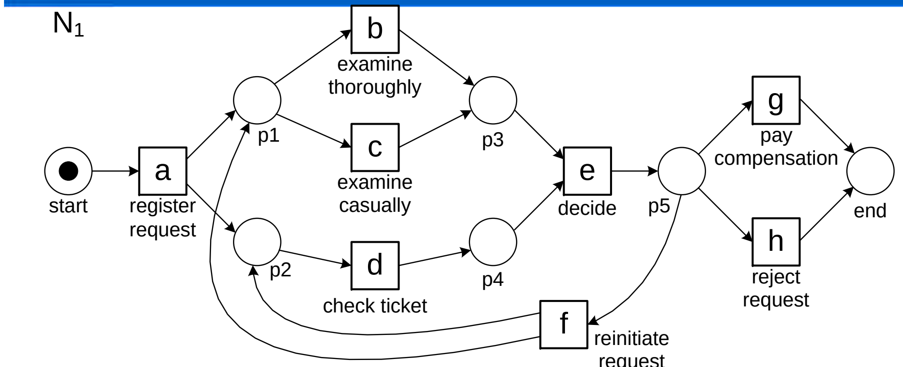
$$c + \sum_{t \in X} \partial_{multiset}(\sigma_1)(t) - \sum_{t \in Y} \partial_{multiset}(\sigma_2)(t) \geq 0.$$

Conformance checking

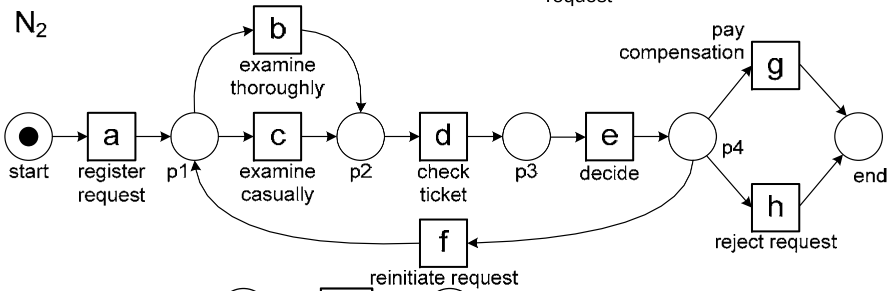


Four models, one log

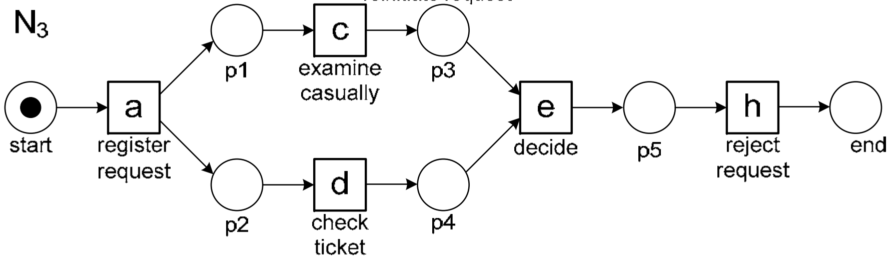
N_1



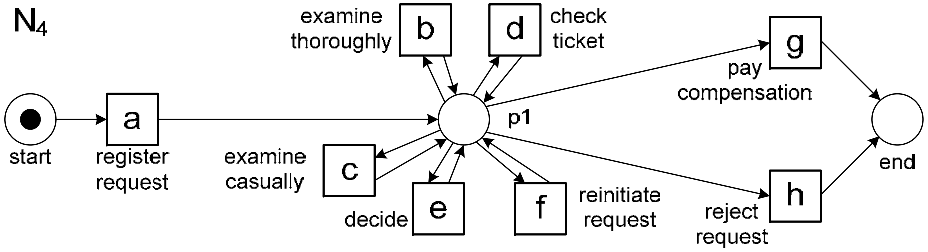
N_2

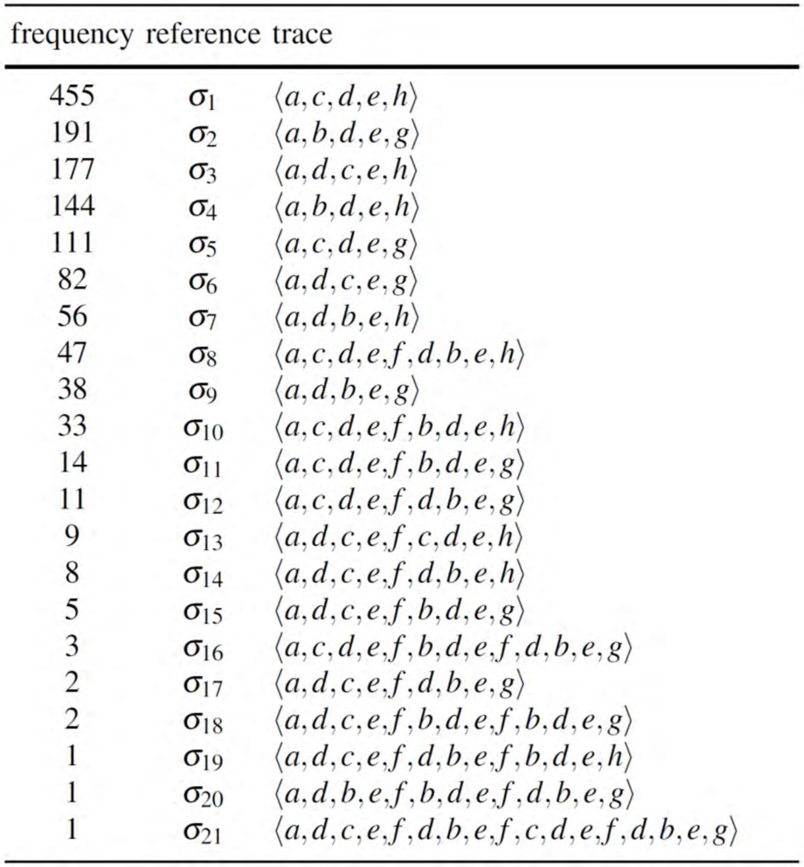


N_3



N_4



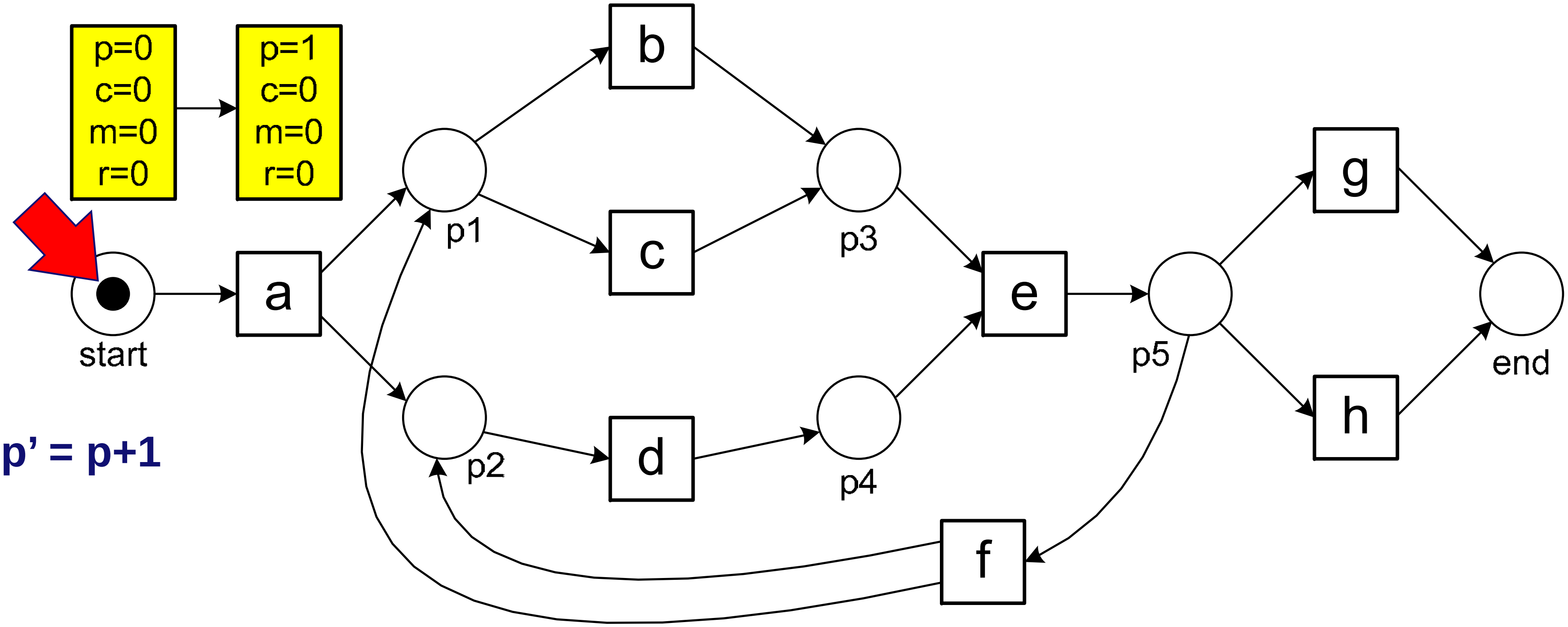


frequency	reference	trace
455	σ_1	$\langle a,c,d,e,h\rangle$
191	σ_2	$\langle a,b,d,e,g\rangle$
177	σ_3	$\langle a,d,c,e,h\rangle$
144	σ_4	$\langle a,b,d,e,h\rangle$
111	σ_5	$\langle a,c,d,e,g\rangle$
82	σ_6	$\langle a,d,c,e,g\rangle$
56	σ_7	$\langle a,d,b,e,h\rangle$
47	σ_8	$\langle a,c,d,e,f,d,b,e,h\rangle$
38	σ_9	$\langle a,d,b,e,g\rangle$
33	σ_{10}	$\langle a,c,d,e,f,b,d,e,h\rangle$
14	σ_{11}	$\langle a,c,d,e,f,b,d,e,g\rangle$
11	σ_{12}	$\langle a,c,d,e,f,d,b,e,g\rangle$
9	σ_{13}	$\langle a,d,c,e,f,c,d,e,h\rangle$
8	σ_{14}	$\langle a,d,c,e,f,d,b,e,h\rangle$
5	σ_{15}	$\langle a,d,c,e,f,b,d,e,g\rangle$
3	σ_{16}	$\langle a,c,d,e,f,b,d,e,f,d,b,e,g\rangle$
2	σ_{17}	$\langle a,d,c,e,f,d,b,e,g\rangle$
2	σ_{18}	$\langle a,d,c,e,f,b,d,e,f,b,d,e,g\rangle$
1	σ_{19}	$\langle a,d,c,e,f,d,b,e,f,b,d,e,h\rangle$
1	σ_{20}	$\langle a,d,b,e,f,b,d,e,f,d,b,e,g\rangle$
1	σ_{21}	$\langle a,d,c,e,f,d,b,e,f,c,d,e,f,d,b,e,g\rangle$

frequency	reference	trace
455	σ_1	$\langle a,c,d,e,h\rangle$
191	σ_2	$\langle a,b,d,e,g\rangle$
177	σ_3	$\langle a,d,c,e,h\rangle$
144	σ_4	$\langle a,b,d,e,h\rangle$
111	σ_5	$\langle a,c,d,e,g\rangle$
82	σ_6	$\langle a,d,c,e,g\rangle$
56	σ_7	$\langle a,d,b,e,h\rangle$
47	σ_8	$\langle a,c,d,e,f,d,b,e,h\rangle$
38	σ_9	$\langle a,d,b,e,g\rangle$
33	σ_{10}	$\langle a,c,d,e,f,b,d,e,h\rangle$
14	σ_{11}	$\langle a,c,d,e,f,b,d,e,g\rangle$
11	σ_{12}	$\langle a,c,d,e,f,d,b,e,g\rangle$
9	σ_{13}	$\langle a,d,c,e,f,c,d,e,h\rangle$
8	σ_{14}	$\langle a,d,c,e,f,d,b,e,h\rangle$

Replaying (1/7)
σ_1 on N_1

$$\sigma_1 = \langle a, c, d, e, h \rangle$$



p=0
c=0
m=0
r=0

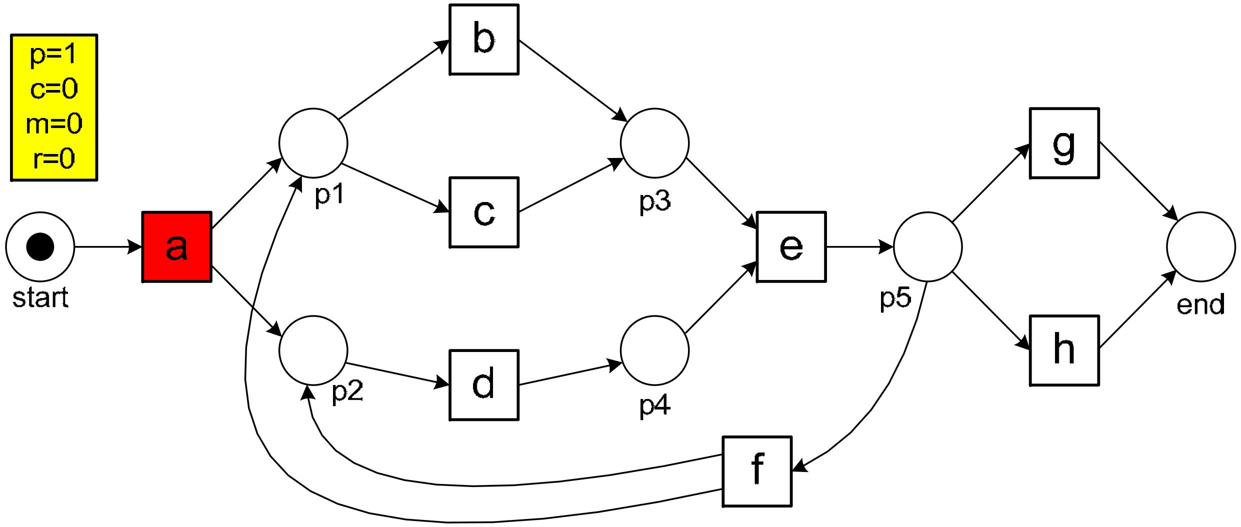
p=1
c=0
m=0
r=0

p’ = p+1

p = produced
c = consumed
m = missing ≤ c
r = remaining ≤ p

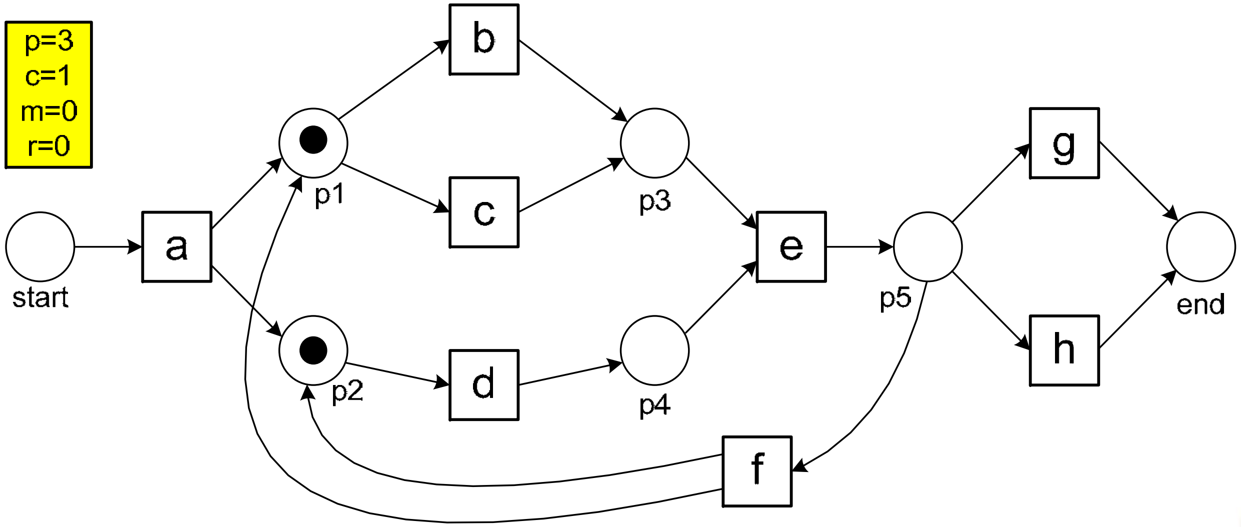
Replaying (2/7)
σ_1 on N_1

$$\sigma_1 = \langle a, c, d, e, h \rangle$$



p=1
c=0
m=0
r=0

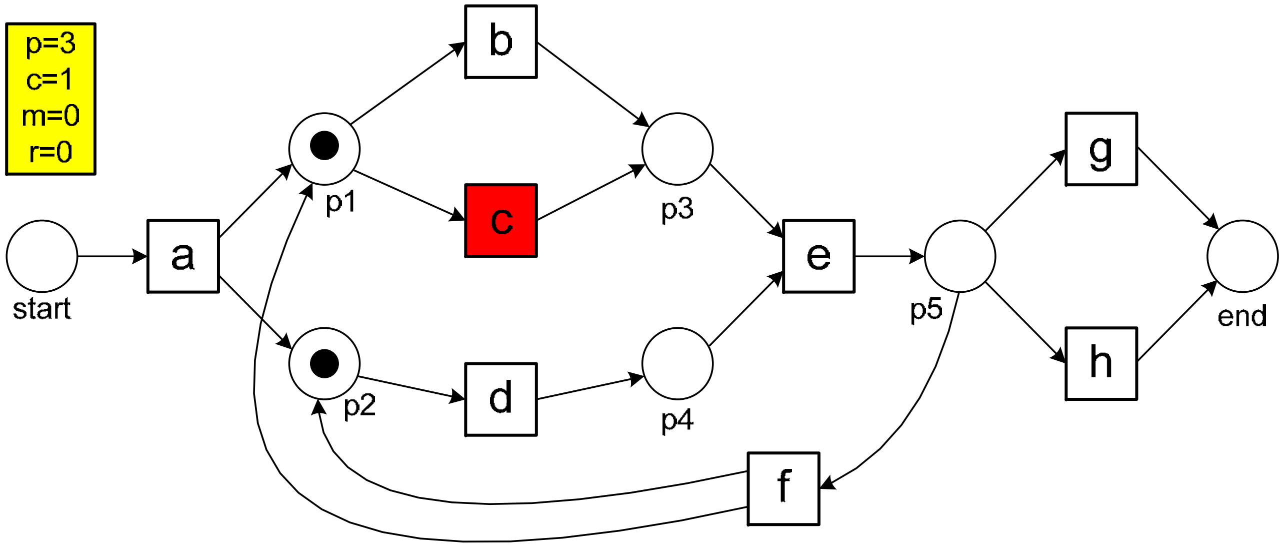
$p' = p+2$
$c' = c+1$



p=3
c=1
m=0
r=0

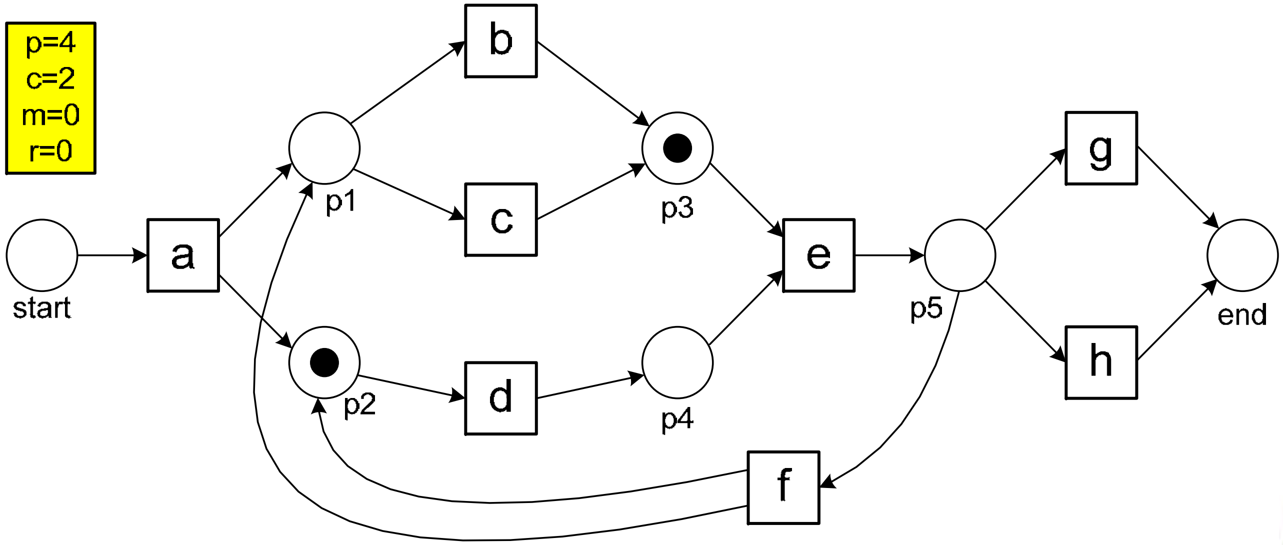
Replaying (3/7)
σ_1 on N_1

$$\sigma_1 = \langle a, c, d, e, h \rangle$$



p=3
c=1
m=0
r=0

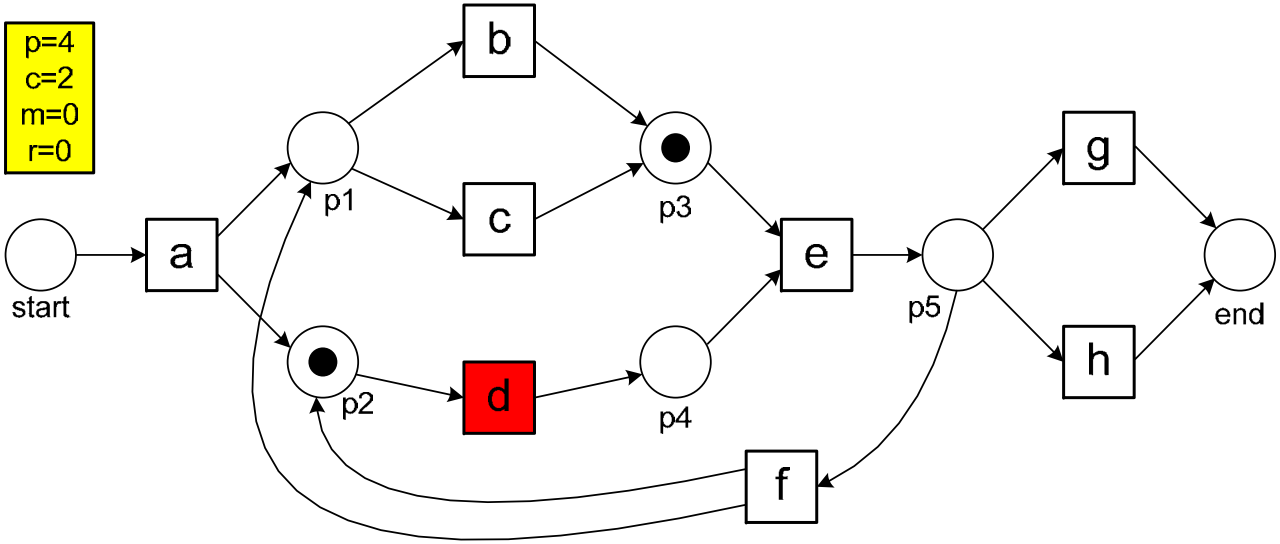
$p' = p+1$
$c' = c+1$



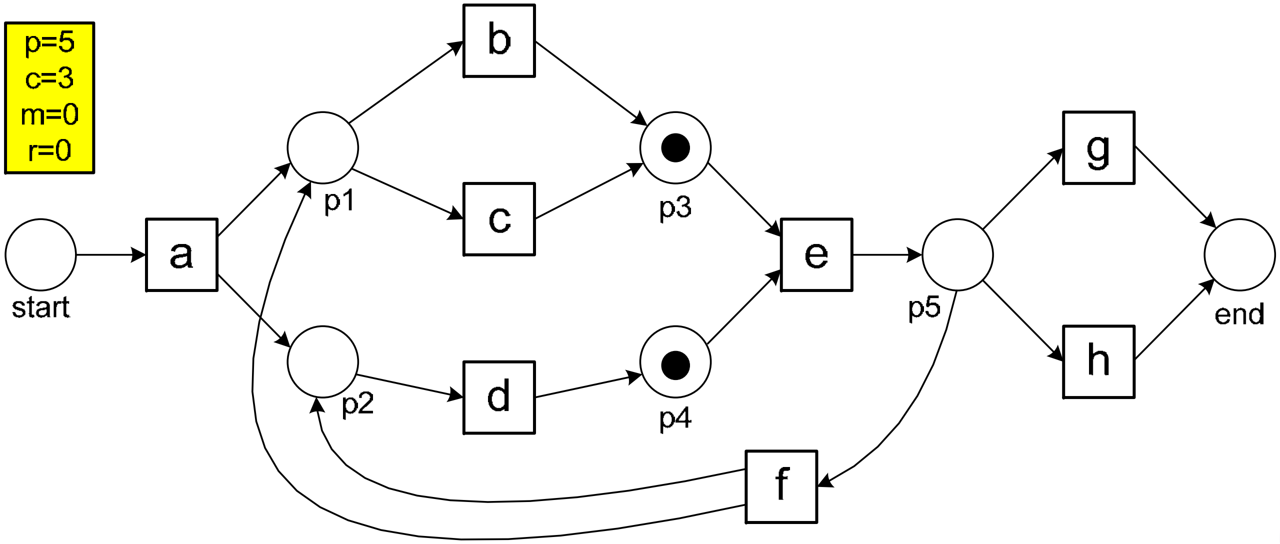
p=4
c=2
m=0
r=0

Replaying (4/7)
σ_1 on N_1

$\sigma_1 = \langle a, c, d, e, h \rangle$



p=4
c=2
m=0
r=0

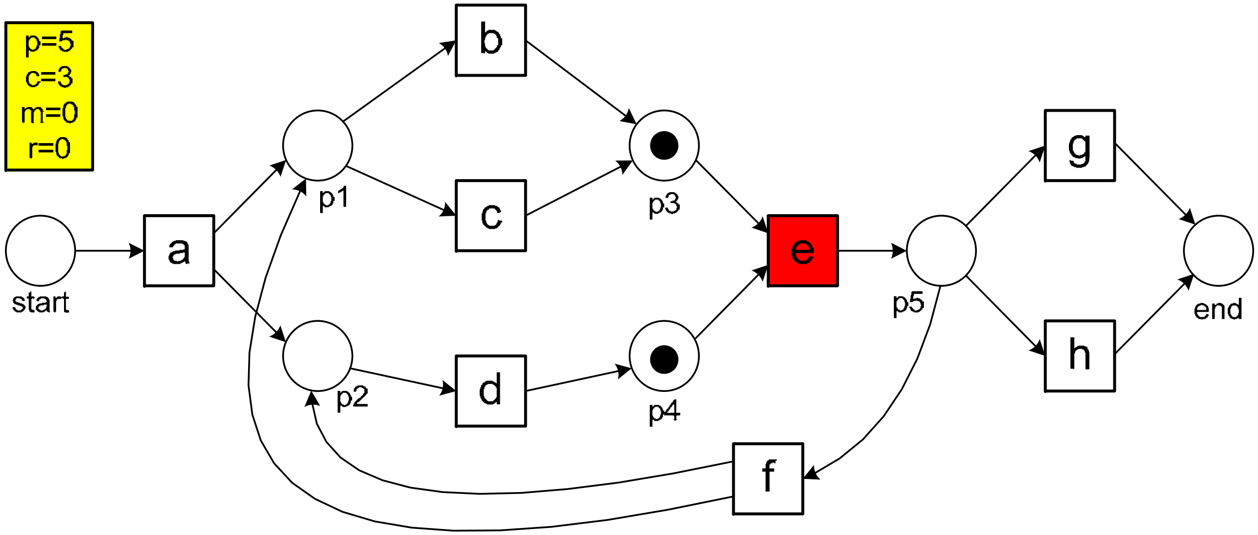


p=5
c=3
m=0
r=0

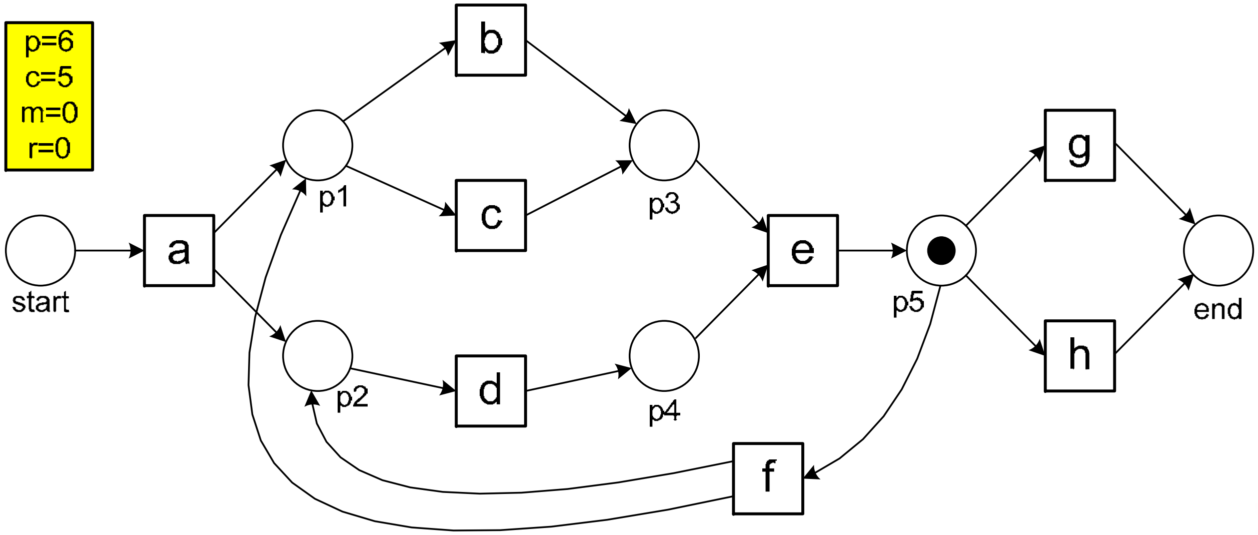
p’ = p+1
c’ = c+1

Replaying (5/7)
σ_1 on N_1

$$\sigma_1 = \langle a, c, d, e, h \rangle$$



p=5
c=3
m=0
r=0

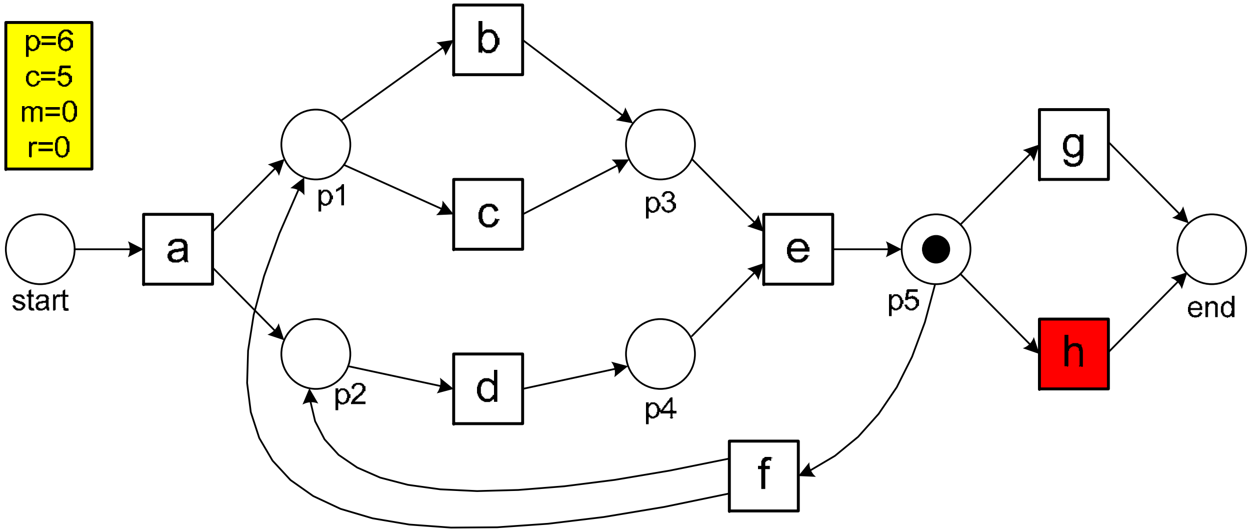


p=6
c=5
m=0
r=0

p’ = p+1
c’ = c+2

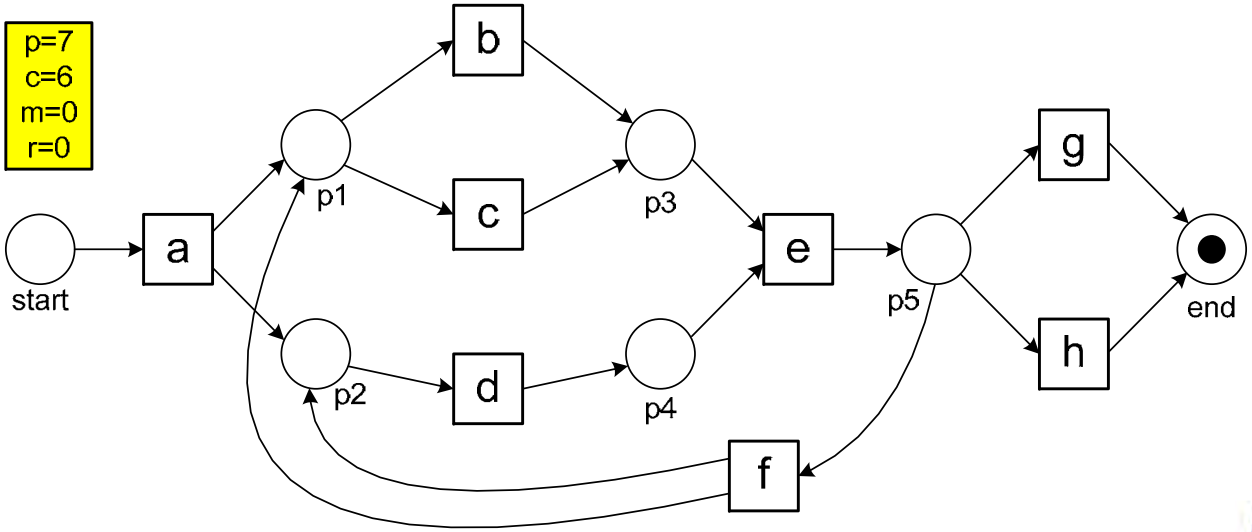
Replaying (6/7)
σ_1 on N_1

$$\sigma_1 = \langle a, c, d, e, h \rangle$$



p=6
c=5
m=0
r=0

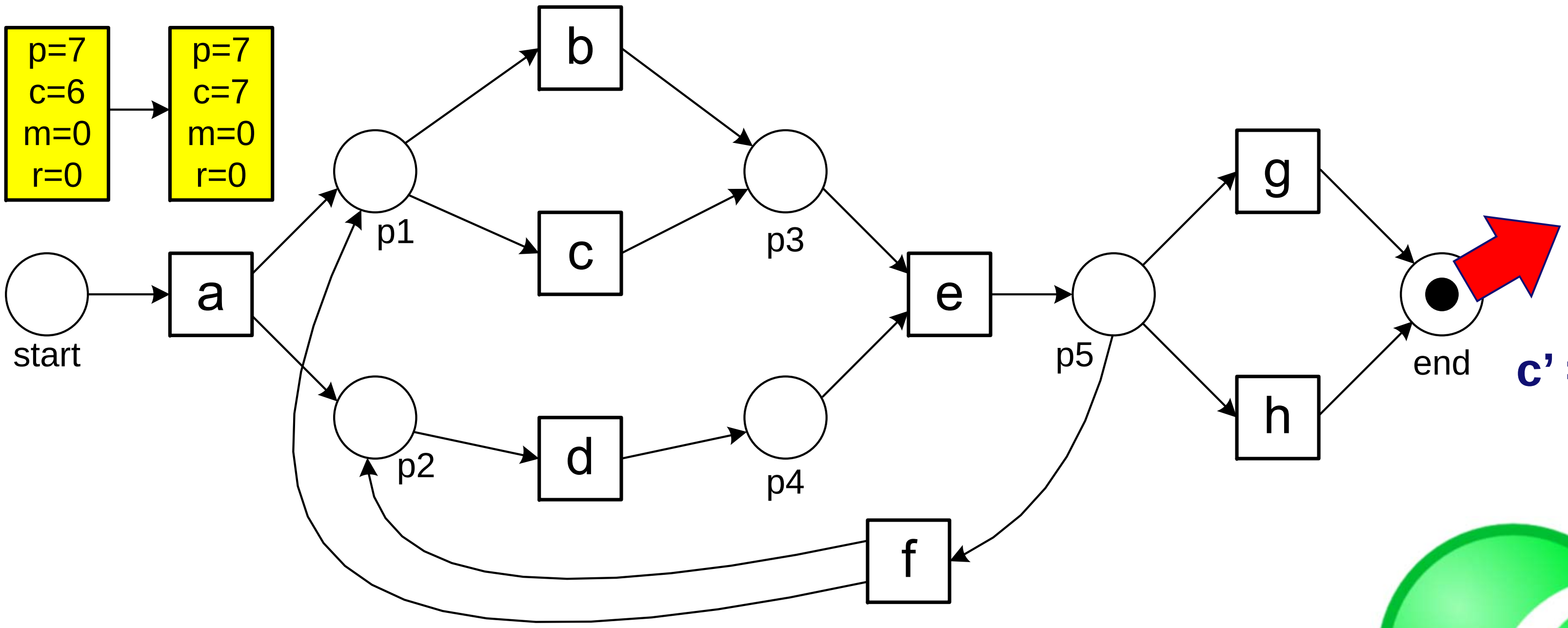
p' = p+1
c' = c+1



p=7
c=6
m=0
r=0

Replaying (7/7)
σ_1 on N_1

$$\sigma_1 = \langle a, c, d, e, h \rangle$$



c’ = c+1

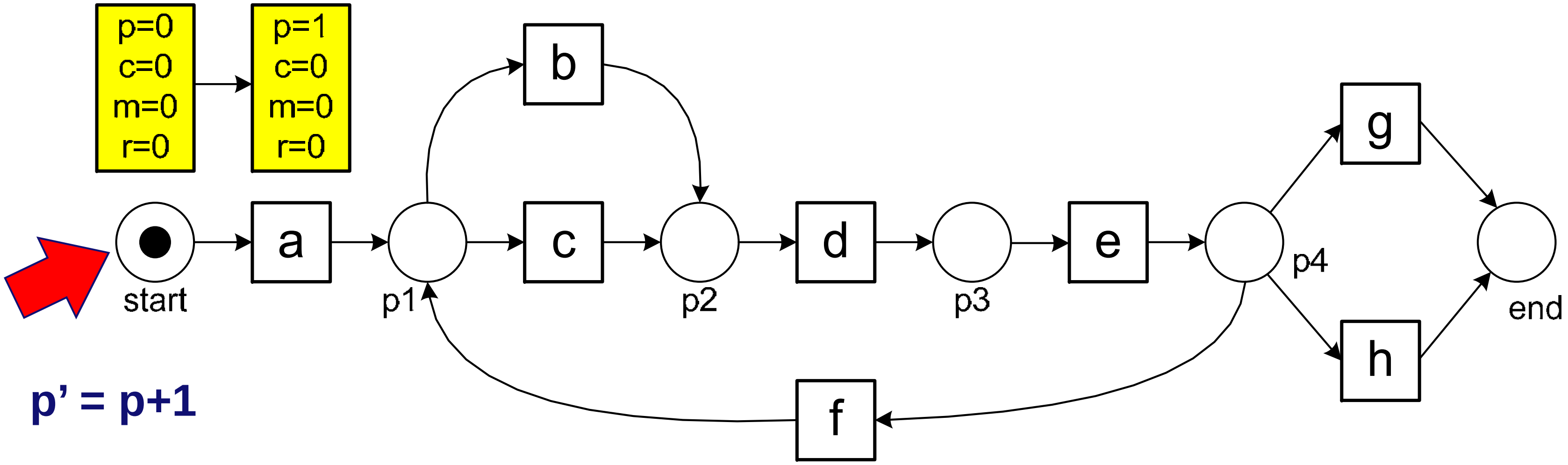
m = 0

r = 0

no problems encountered

Replaying (1/7)
σ_3 on N_2

$$\sigma_3 = \langle a, d, c, e, h \rangle$$

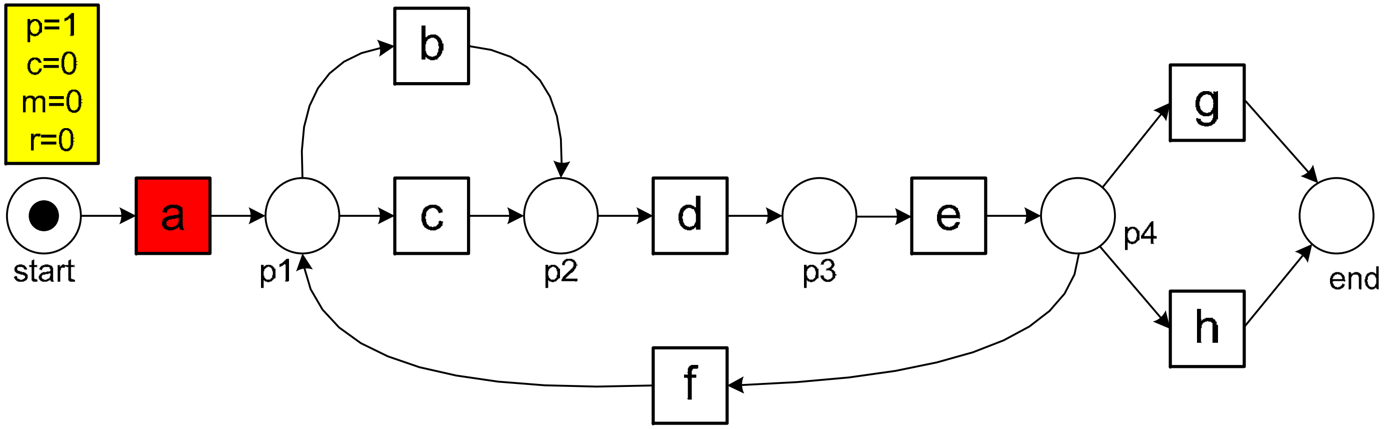


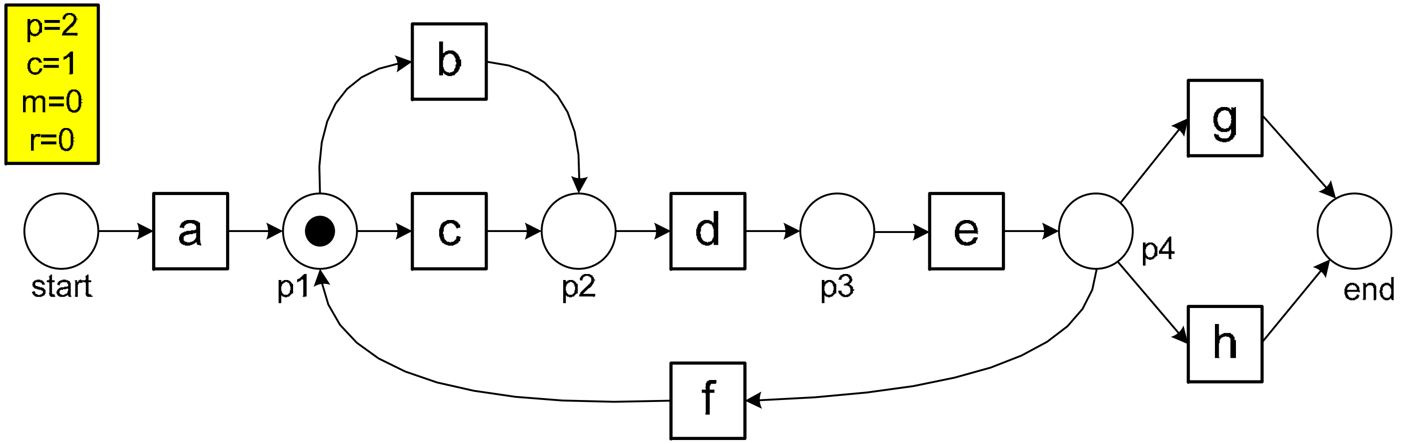
$p' = p+1$

p = produced
c = consumed
m = missing ≤ c
r = remaining ≤ p

Replaying (2/7)
σ_3 on N_2

$$\sigma_3 = \langle a, d, c, e, h \rangle$$

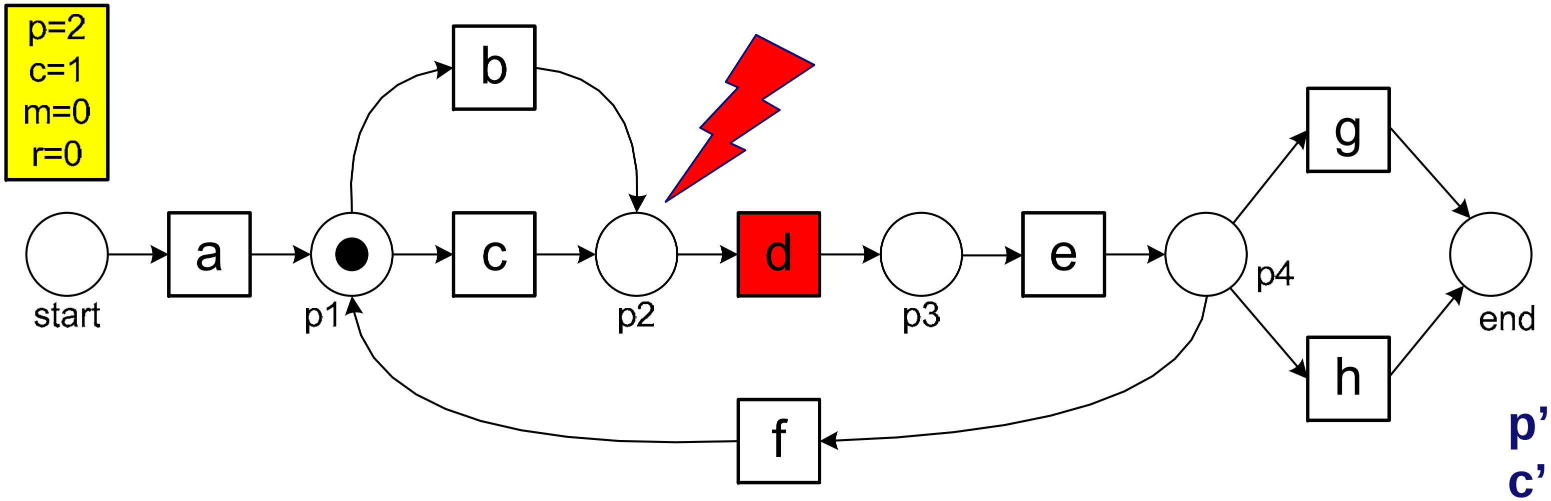

p=1
c=0
m=0
r=0


p=2
c=1
m=0
r=0

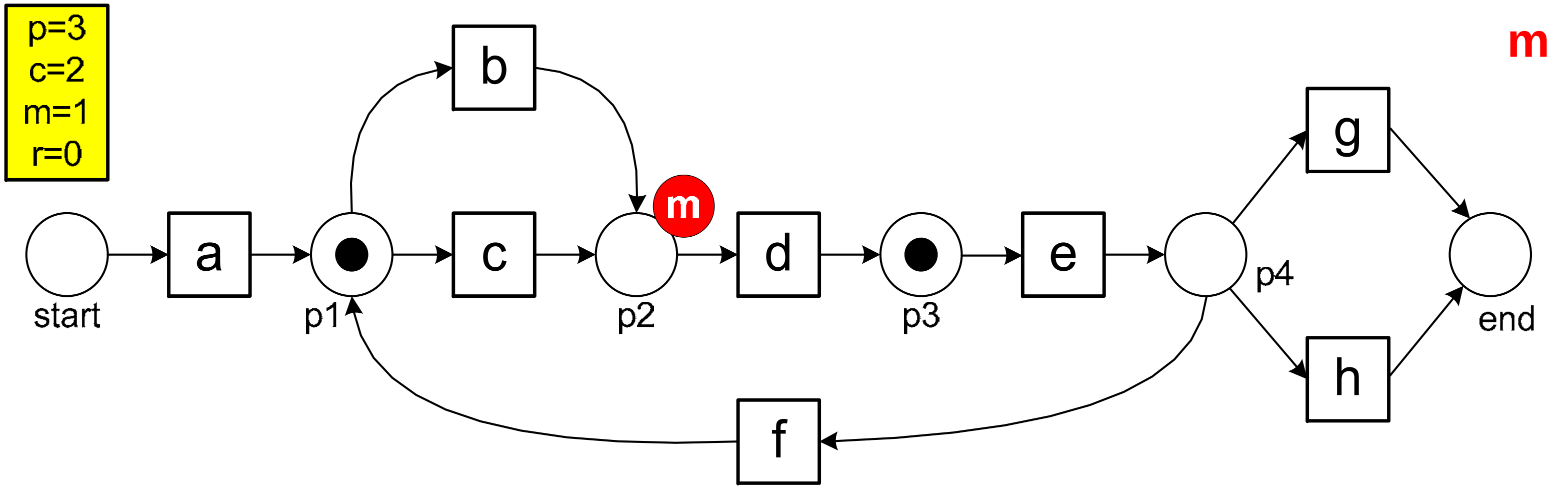
p’ = p+1
c’ = c+1

Replaying (3/7)
σ_3 on N_2

$$\sigma_3 = \langle a, d, c, e, h \rangle$$



p=2
c=1
m=0
r=0

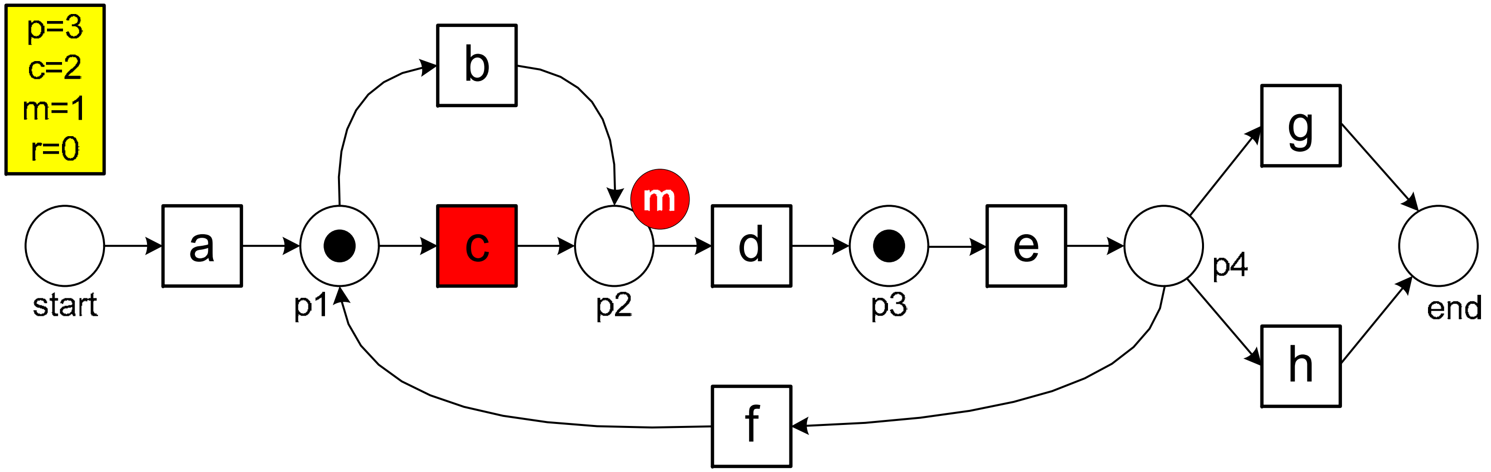


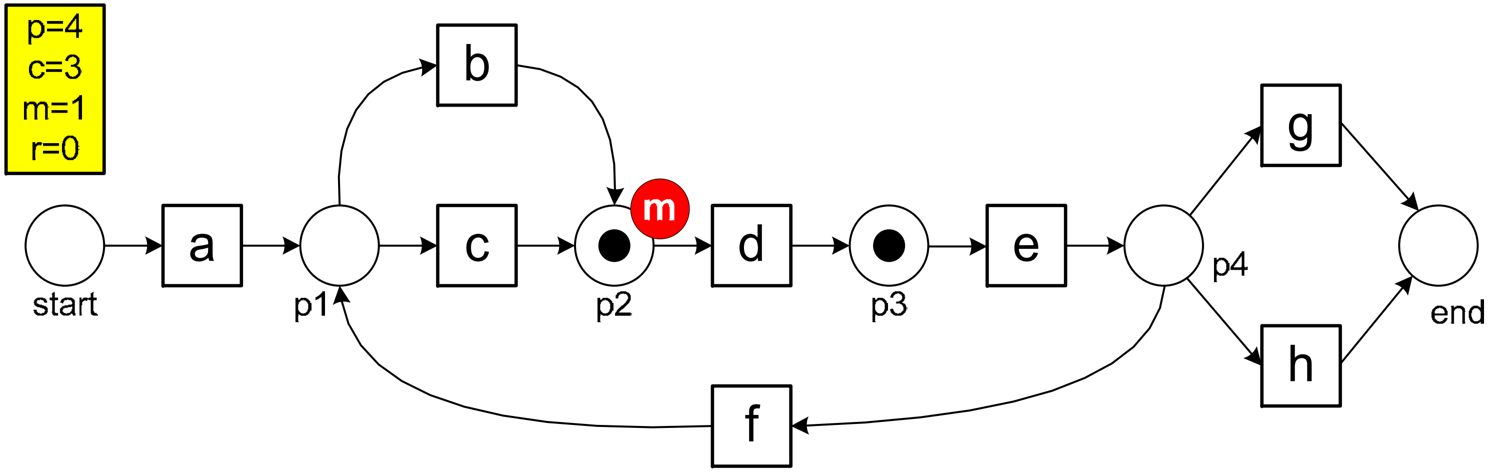
p=3
c=2
m=1
r=0

p’ = p+1
c’ = c+1
m’ = m+1

Replaying (4/7)
σ_3 on N_2

$$\sigma_3 = \langle a, d, \underset{\uparrow}{c}, e, h \rangle$$

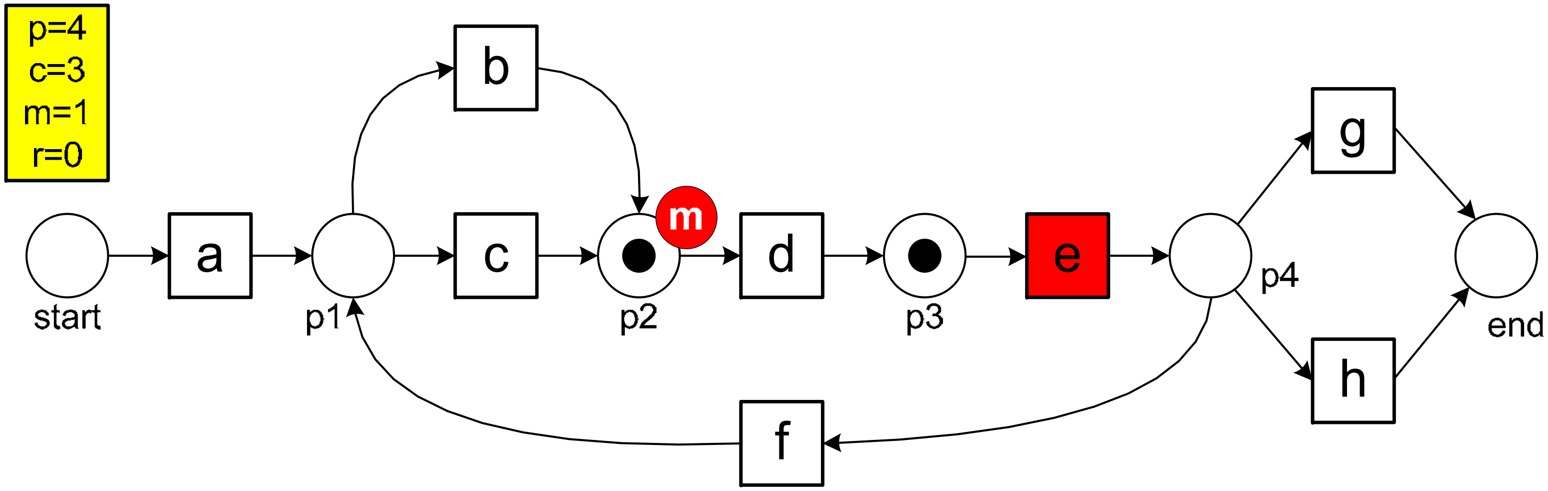

p=3
c=2
m=1
r=0


p=4
c=3
m=1
r=0

p’ = p+1
c’ = c+1

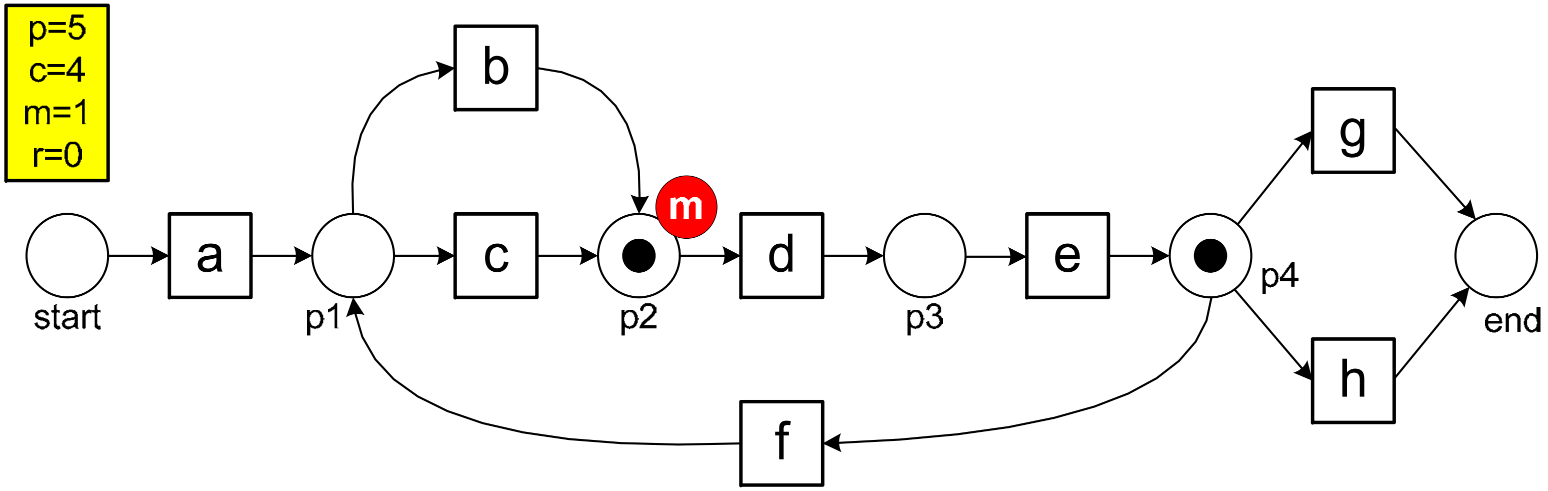
Replaying (5/7)
σ_3 on N_2

$$\sigma_3 = \langle a, d, c, e, h \rangle$$



p=4
c=3
m=1
r=0

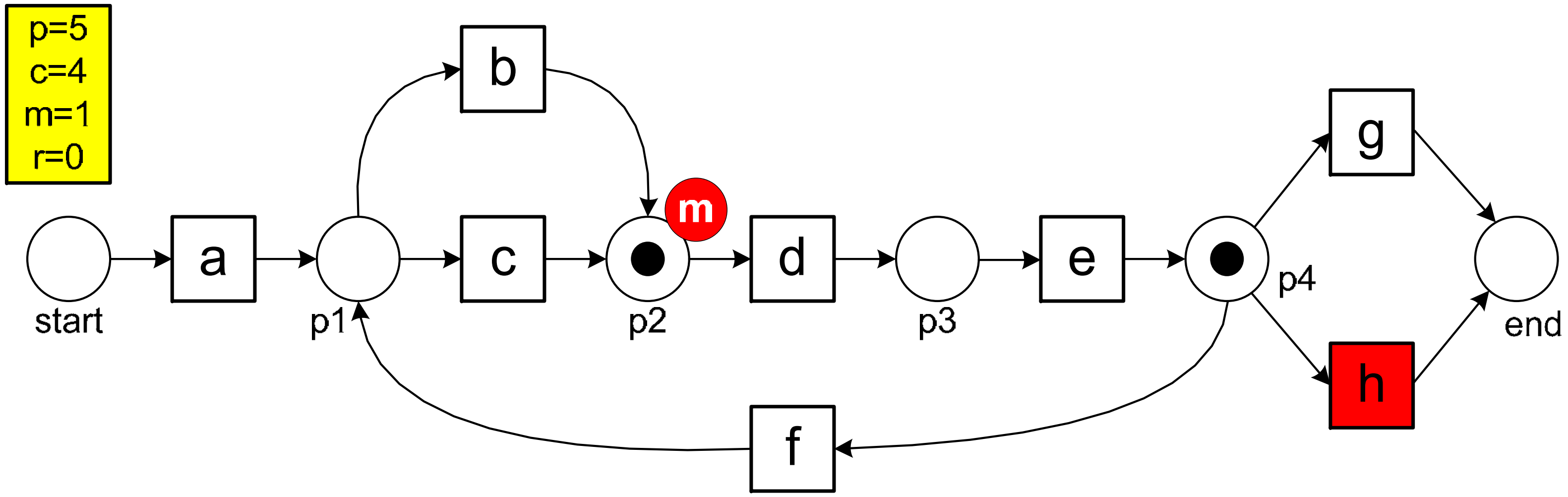
$p' = p+1$
$c' = c+1$



p=5
c=4
m=1
r=0

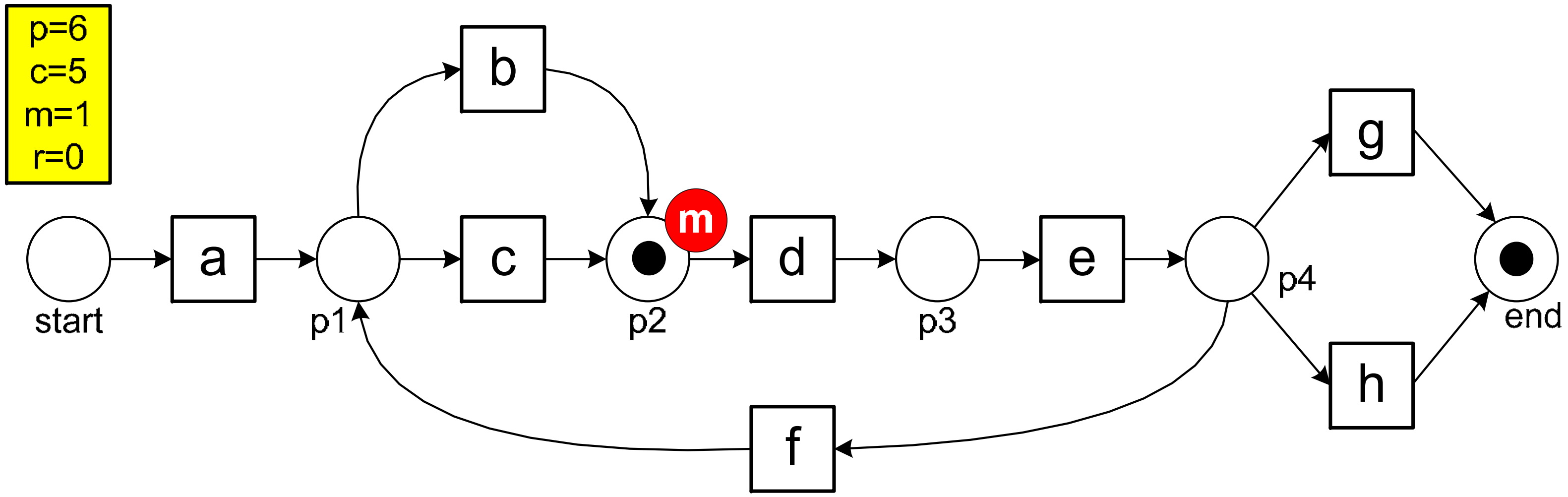
Replaying (6/7)
σ_3 on N_2

$$\sigma_3 = \langle a, d, c, e, h \rangle$$



p=5
c=4
m=1
r=0

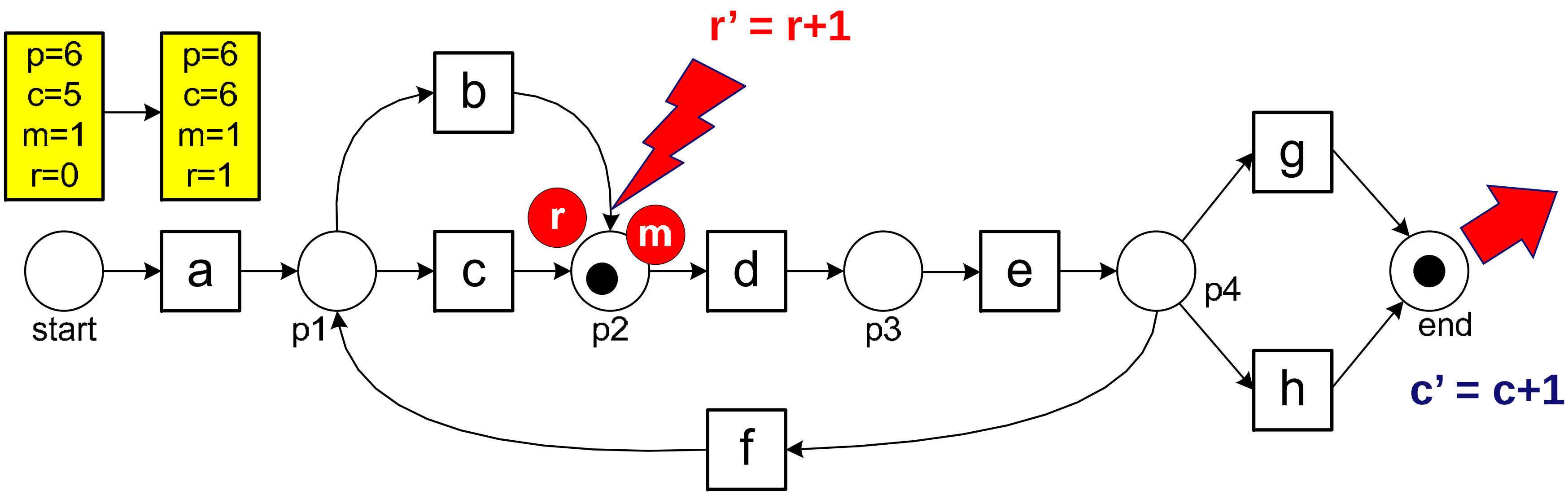
p' = p+1
c' = c+1



p=6
c=5
m=1
r=0

Replaying (7/7)
σ_3 on N_2

$$\sigma_3 = \langle a, d, c, e, h \rangle$$

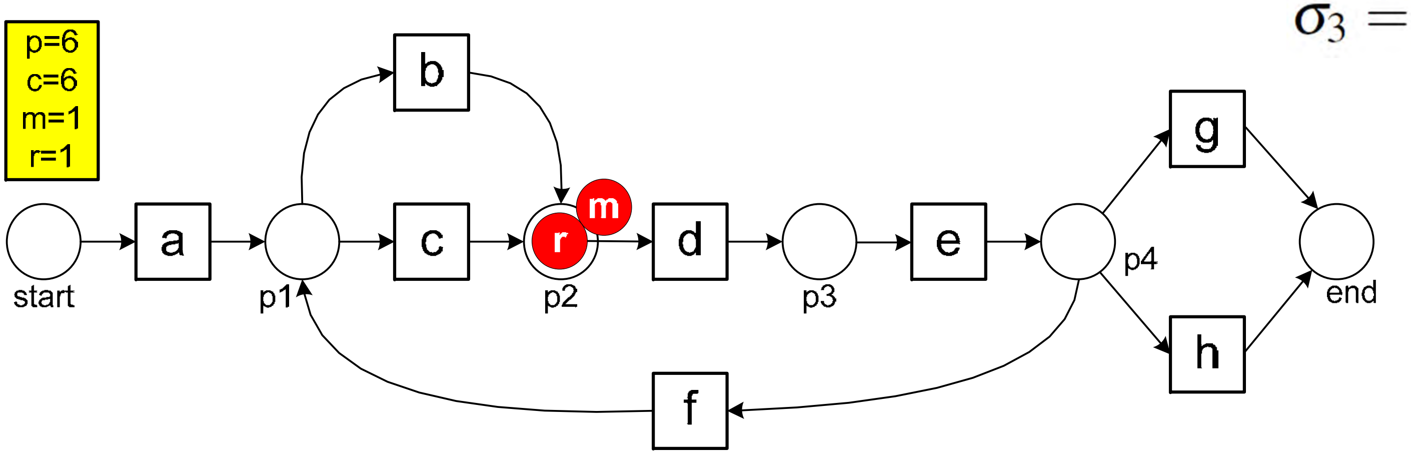


Problems:

- **m = 1 : d was forced to occur without being enabled**
- **r = 1 : output of c was not used by d**

Computing fitness at trace level

$$\sigma_3 = \langle a, d, c, e, h \rangle$$

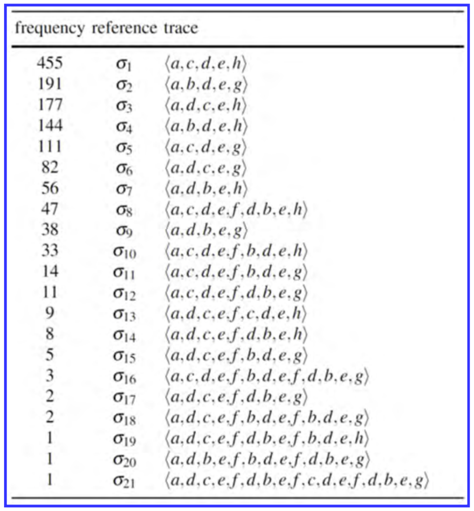


$$fitness(\sigma, N) = \frac{1}{2}\left(1 - \frac{m}{c}\right) + \frac{1}{2}\left(1 - \frac{r}{p}\right)$$

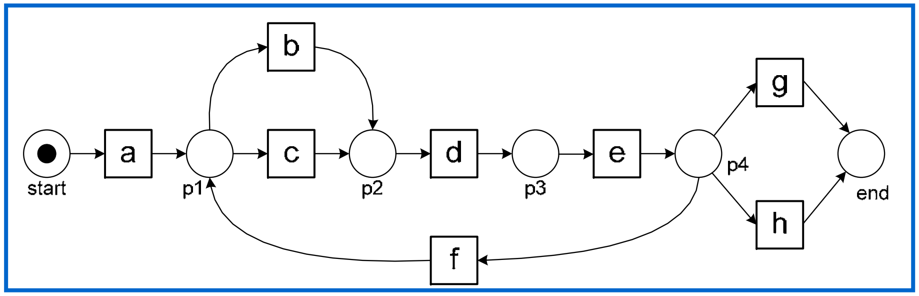
$$fitness(\sigma_3, N_2) = \frac{1}{2}\left(1 - \frac{1}{6}\right) + \frac{1}{2}\left(1 - \frac{1}{6}\right) = 0.8333$$

Computing fitness at the log level

$$fitness(L,N) = \frac{1}{2}\left(1 - \frac{\sum_{\sigma \in L} L(\sigma) \times m_{N,\sigma}}{\sum_{\sigma \in L} L(\sigma) \times c_{N,\sigma}}\right) + \frac{1}{2}\left(1 - \frac{\sum_{\sigma \in L} L(\sigma) \times r_{N,\sigma}}{\sum_{\sigma \in L} L(\sigma) \times p_{N,\sigma}}\right)$$

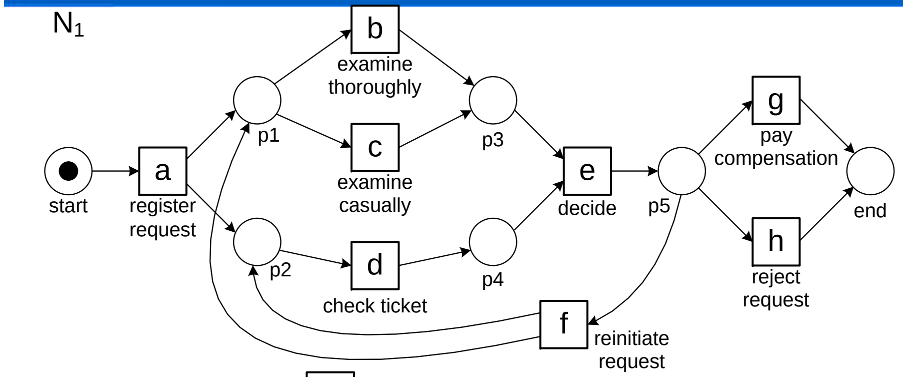


frequency	reference	trace
455	σ_1	$\langle a,c,d,e,h \rangle$
191	σ_2	$\langle a,b,d,e,g \rangle$
177	σ_3	$\langle a,d,c,e,h \rangle$
144	σ_4	$\langle a,b,d,e,h \rangle$
111	σ_5	$\langle a,c,d,e,g \rangle$
82	σ_6	$\langle a,d,c,e,g \rangle$
56	σ_7	$\langle a,d,b,e,h \rangle$
47	σ_8	$\langle a,c,d,e,f,d,b,e,h \rangle$
38	σ_9	$\langle a,d,b,e,g \rangle$
33	σ_{10}	$\langle a,c,d,e,f,b,d,e,h \rangle$
14	σ_{11}	$\langle a,c,d,e,f,b,d,e,g \rangle$
11	σ_{12}	$\langle a,c,d,e,f,d,b,e,g \rangle$
9	σ_{13}	$\langle a,d,c,e,f,c,d,e,h \rangle$
8	σ_{14}	$\langle a,d,c,e,f,d,b,e,h \rangle$
5	σ_{15}	$\langle a,d,c,e,f,b,d,e,g \rangle$
3	σ_{16}	$\langle a,c,d,e,f,b,d,e,f,d,b,e,g \rangle$
2	σ_{17}	$\langle a,d,c,e,f,d,b,e,g \rangle$
2	σ_{18}	$\langle a,d,c,e,f,b,d,e,f,b,d,e,g \rangle$
1	σ_{19}	$\langle a,d,c,e,f,d,b,e,f,b,d,e,h \rangle$
1	σ_{20}	$\langle a,d,b,e,f,b,d,e,f,d,b,e,g \rangle$
1	σ_{21}	$\langle a,d,c,e,f,d,b,e,f,c,d,e,f,d,b,e,g \rangle$

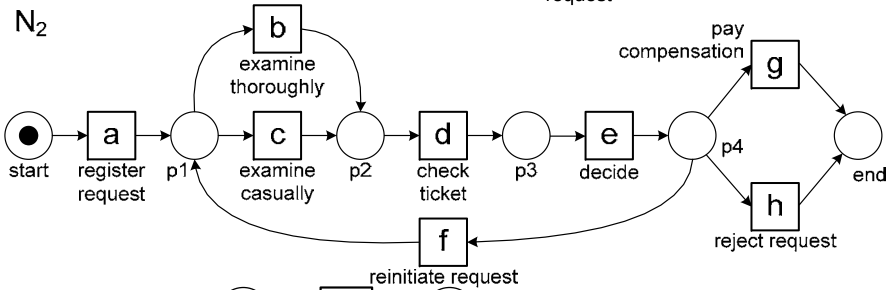


Example values

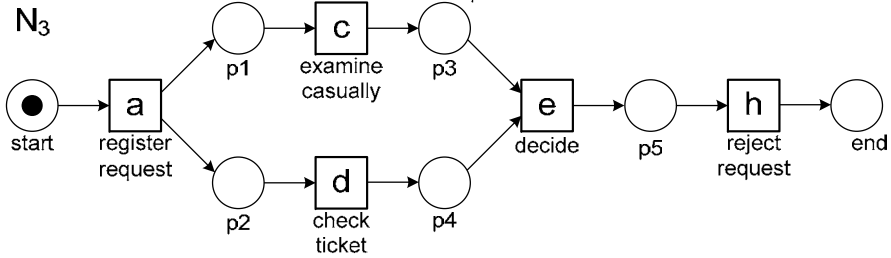
N1



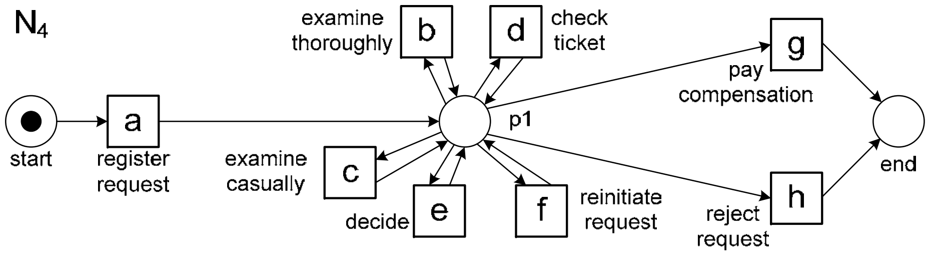
N2

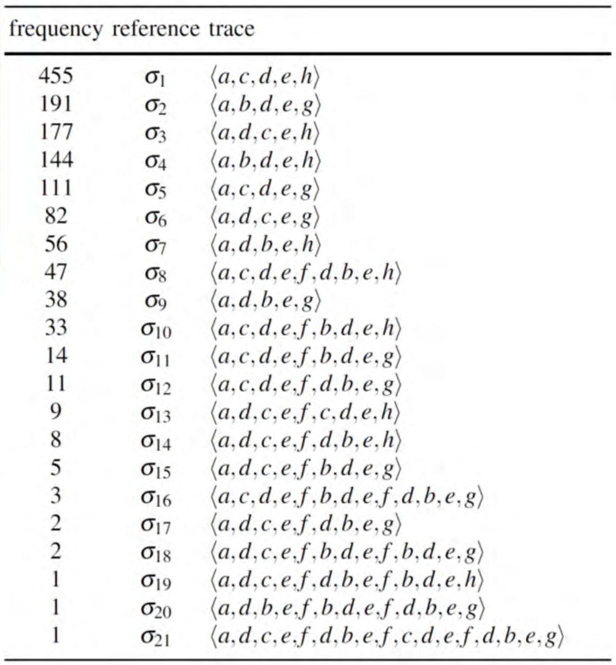


N3



N4





frequency	reference	trace
455	σ_1	$\langle a,c,d,e,h\rangle$
191	σ_2	$\langle a,b,d,e,g\rangle$
177	σ_3	$\langle a,d,c,e,h\rangle$
144	σ_4	$\langle a,b,d,e,h\rangle$
111	σ_5	$\langle a,c,d,e,g\rangle$
82	σ_6	$\langle a,d,c,e,g\rangle$
56	σ_7	$\langle a,d,b,e,h\rangle$
47	σ_8	$\langle a,c,d,e,f,d,b,e,h\rangle$
38	σ_9	$\langle a,d,b,e,g\rangle$
33	σ_{10}	$\langle a,c,d,e,f,b,d,e,h\rangle$
14	σ_{11}	$\langle a,c,d,e,f,b,d,e,g\rangle$
11	σ_{12}	$\langle a,c,d,e,f,d,b,e,g\rangle$
9	σ_{13}	$\langle a,d,c,e,f,c,d,e,h\rangle$
8	σ_{14}	$\langle a,d,c,e,f,d,b,e,h\rangle$
5	σ_{15}	$\langle a,d,c,e,f,b,d,e,g\rangle$
3	σ_{16}	$\langle a,c,d,e,f,b,d,e,f,d,b,e,g\rangle$
2	σ_{17}	$\langle a,d,c,e,f,d,b,e,g\rangle$
2	σ_{18}	$\langle a,d,c,e,f,b,d,e,f,b,d,e,g\rangle$
1	σ_{19}	$\langle a,d,c,e,f,d,b,e,f,b,d,e,h\rangle$
1	σ_{20}	$\langle a,d,b,e,f,b,d,e,f,d,b,e,g\rangle$
1	σ_{21}	$\langle a,d,c,e,f,d,b,e,f,c,d,e,f,d,b,e,g\rangle$

$$fitness(L_{full}, N_1) = 1$$

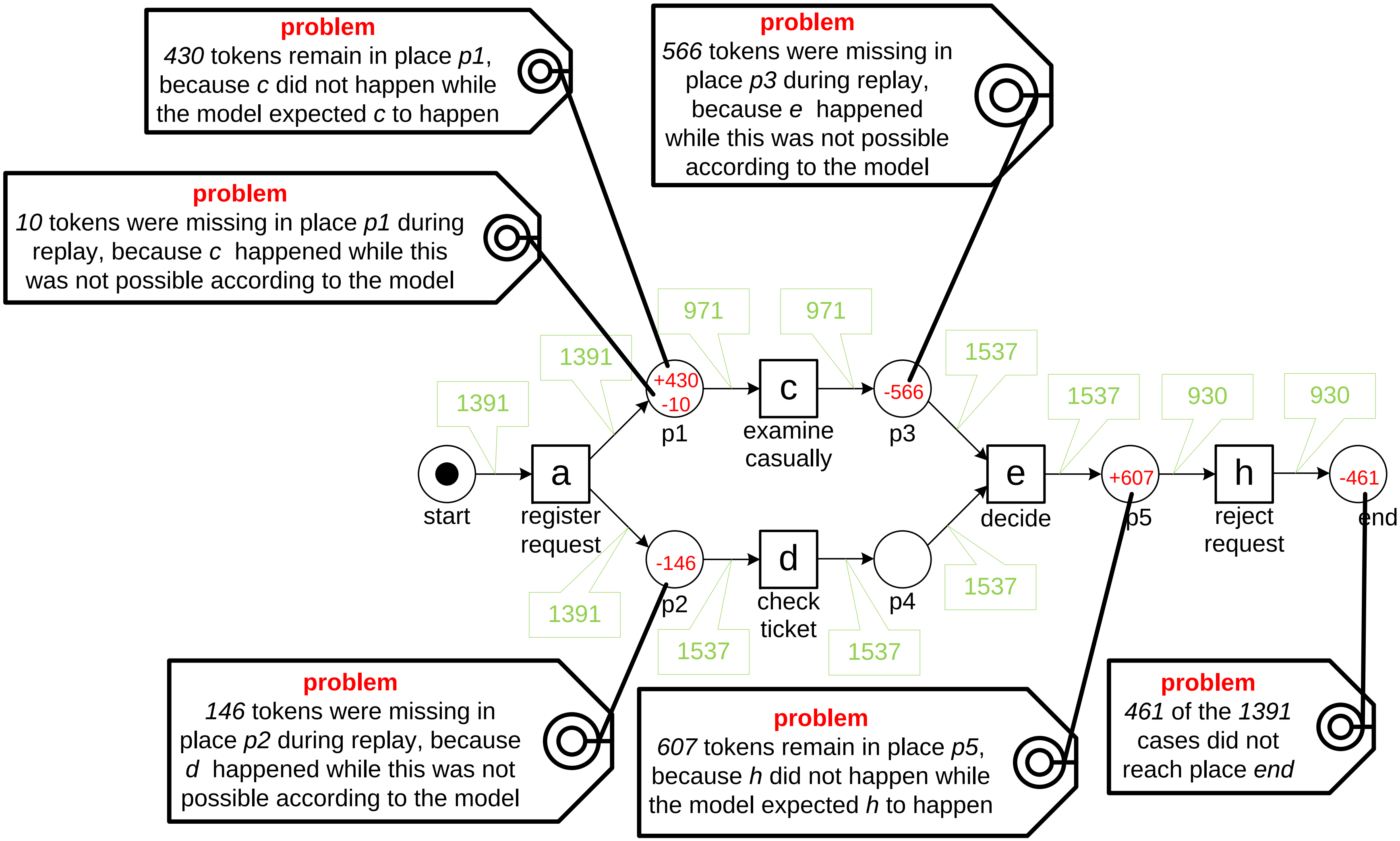
$$fitness(L_{full}, N_2) = 0.9504$$

$$fitness(L_{full}, N_3) = 0.8797$$

$$fitness(L_{full}, N_4) = 1$$

Diagnostics

$$(fitness(L_{full}, N_3) = 0.8797)$$



problem
430 tokens remain in place *p1*, because *c* did not happen while the model expected *c* to happen

problem
566 tokens were missing in place *p3* during replay, because *e* happened while this was not possible according to the model

problem
10 tokens were missing in place *p1* during replay, because *c* happened while this was not possible according to the model

problem
146 tokens were missing in place *p2* during replay, because *d* happened while this was not possible according to the model

problem
607 tokens remain in place *p5*, because *h* did not happen while the model expected *h* to happen

problem
461 of the *1391* cases did not reach place *end*

Challenges related to conformance checking

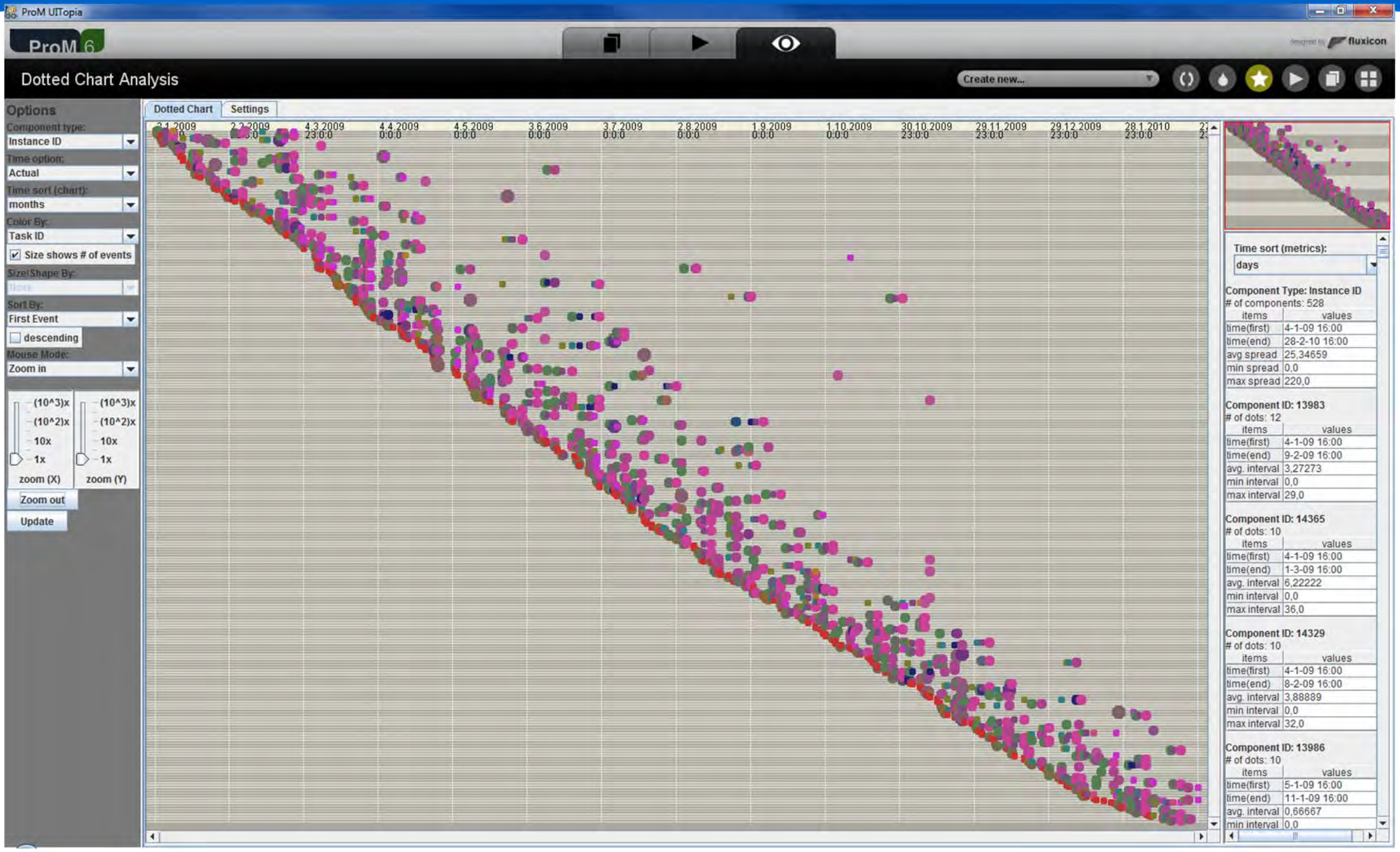
- Not as simple as it seems!
- In case of duplicate tasks (two transition with the same label) or silent tasks (τ labeled transitions), multiple paths need to be considered (state space analysis, heuristics, or optimization).
- More general formulation of the problem with costs associated to skipping/inserting particular tasks, see ProM latest conformance checker (A* algorithm).
- Computing the most likely alignment is needed for other types of process mining (time analysis, measuring precision, social network analysis, etc.).

How can process mining help?

- Detect bottlenecks
- Detect deviations
- Performance measurement
- Suggest improvements
- Decision support (e.g., recommendation and prediction)

- Provide mirror
- Highlight important problems
- Avoid ICT failures
- Avoid management by PowerPoint
- From “politics” to “analytics”

Example of a Lasagna process: WMO process of a Dutch municipality



Dotted Chart Analysis

Options

Component type: Instance ID

Time option: Actual

Time sort (chart): months

Color By: Task ID

☑ Size shows # of events

Size/Shape By: None

Sort By: First Event

☐ descending

Mouse Mode: Zoom in

Time sort (metrics): days

Component Type: Instance ID
# of components: 528

items	values
time(first)	4-1-09 16:00
time(end)	28-2-10 16:00
avg spread	25,34659
min spread	0,0
max spread	220,0

Component ID: 13983
# of dots: 12

items	values
time(first)	4-1-09 16:00
time(end)	9-2-09 16:00
avg. interval	3,27273
min interval	0,0
max interval	29,0

Component ID: 14365
# of dots: 10

items	values
time(first)	4-1-09 16:00
time(end)	1-3-09 16:00
avg. interval	6,22222
min interval	0,0
max interval	36,0

Component ID: 14329
# of dots: 10

items	values
time(first)	4-1-09 16:00
time(end)	8-2-09 16:00
avg. interval	3,88889
min interval	0,0
max interval	32,0

Component ID: 13986
# of dots: 10

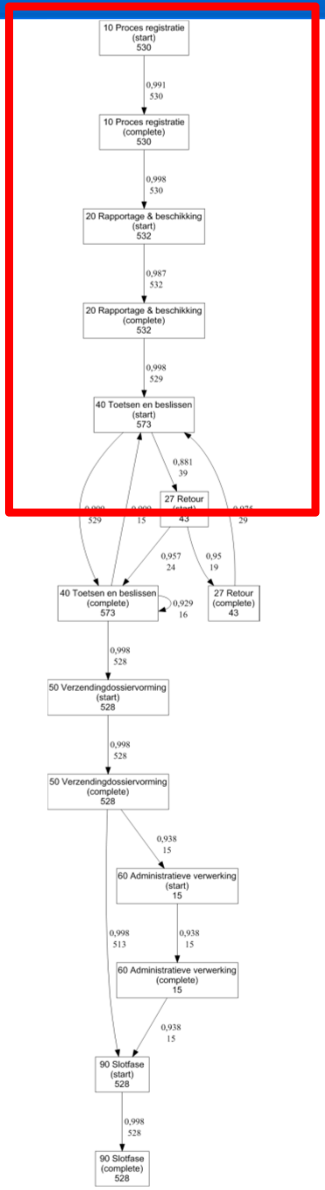
items	values
time(first)	5-1-09 16:00
time(end)	11-1-09 16:00
avg. interval	0,66667
min interval	0,0

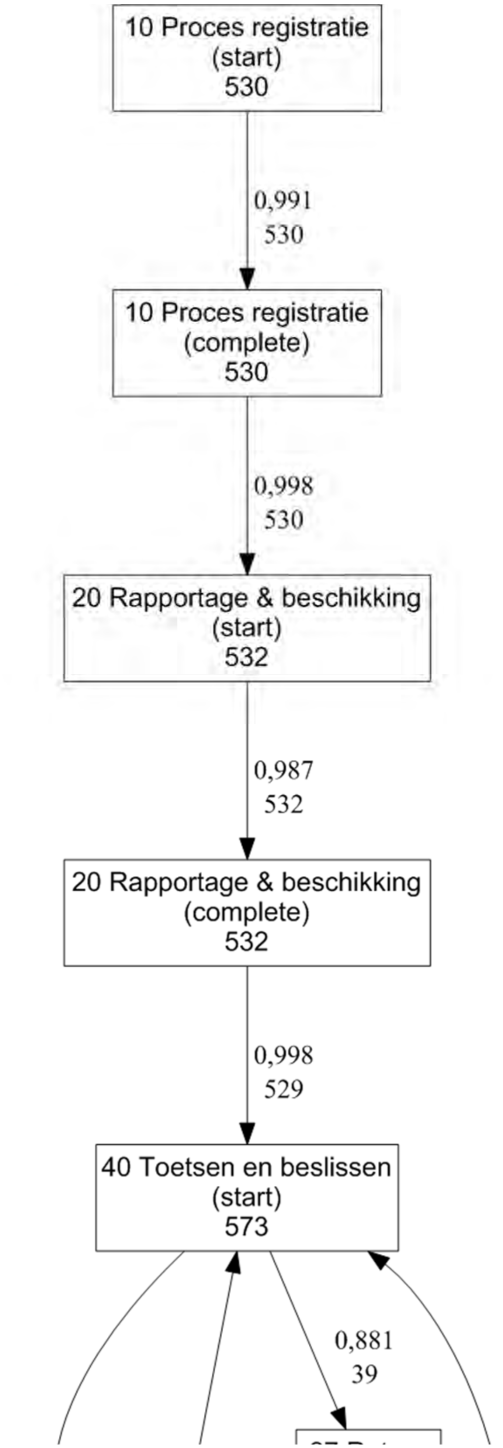
Each line corresponds to one of the 528 requests that were handled in the period from 4-1-2009 until 28-2-2010. In total there are 5498 events represented as dots. The mean time needed to handled a case is approximately 25 days.

WMO process
(Wet Maatschappelijke Ondersteuning)

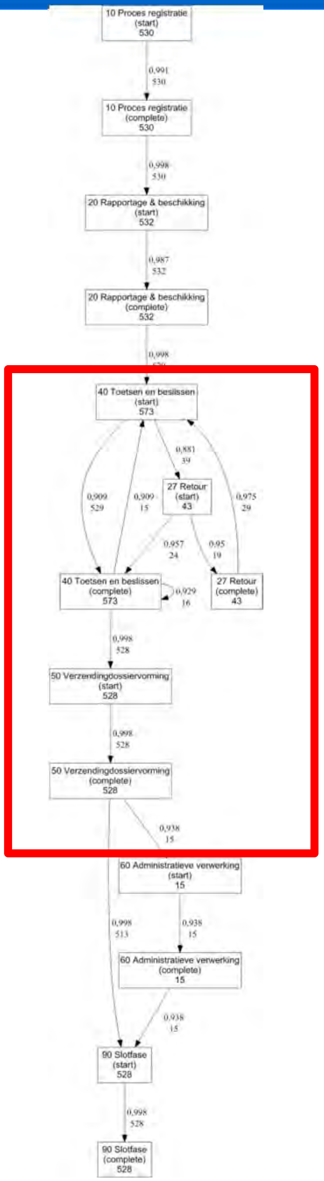
- WMO refers to the social support act that came into force in The Netherlands on January 1st, 2007.
- The aim of this act is to assist people with disabilities and impairments. Under the act, local authorities are required to give support to those who need it, e.g., household help, providing wheelchairs and scootmobiles, and adaptations to homes.
- There are different processes for the different kinds of help. We focus on the process for handling requests for household help.
- In a period of about one year, 528 requests for household WMO support were received.
- These 528 requests generated 5498 events.

C-net discovered using heuristic miner (1/3)

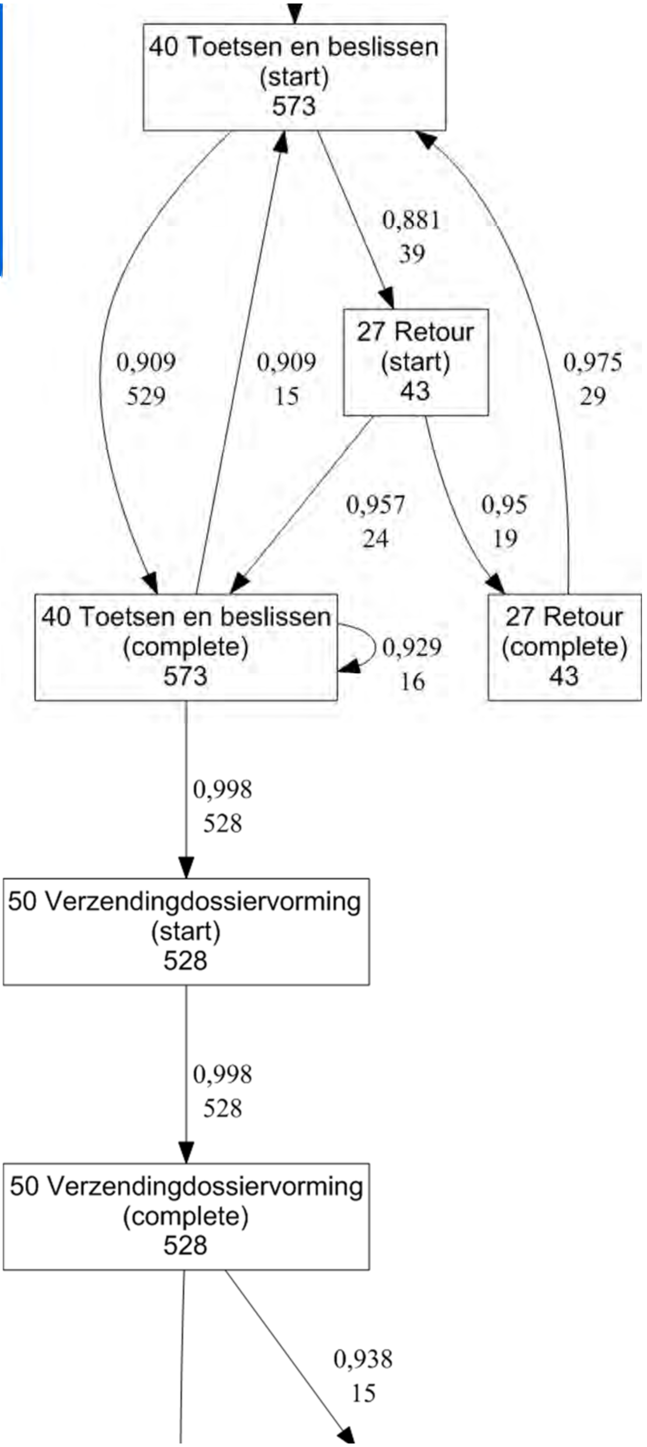

10 Proces registratie (start) 530
0,991 530
10 Proces registratie (complete) 530
0,998 530
20 Rapportage & beschikking (start) 532
0,987 532
20 Rapportage & beschikking (complete) 532
0,998 529
40 Toetsen en beslissen (start) 573
0,881 39
27 Retour (start) 43
529
15
29
0,957 24
0,95 19
40 Toetsen en beslissen (complete) 573
0,929 16
27 Retour (complete) 43
0,998 528
50 Verzendingdossiervorming (start) 528
0,998 528
50 Verzendingdossiervorming (complete) 528
0,938 15
60 Administratieve verwerking (start) 15
0,998 513
0,938 15
60 Administratieve verwerking (complete) 15
0,938 15
90 Slotfase (start) 528
0,998 528
90 Slotfase (complete) 528


10 Proces registratie (start) 530
0,991 530
10 Proces registratie (complete) 530
0,998 530
20 Rapportage & beschikking (start) 532
0,987 532
20 Rapportage & beschikking (complete) 532
0,998 529
40 Toetsen en beslissen (start) 573
0,881 39

C-net discovered using heuristic miner (2/3)

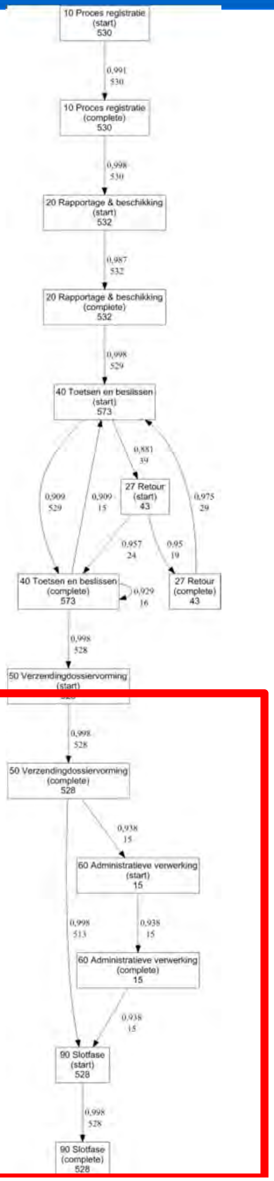


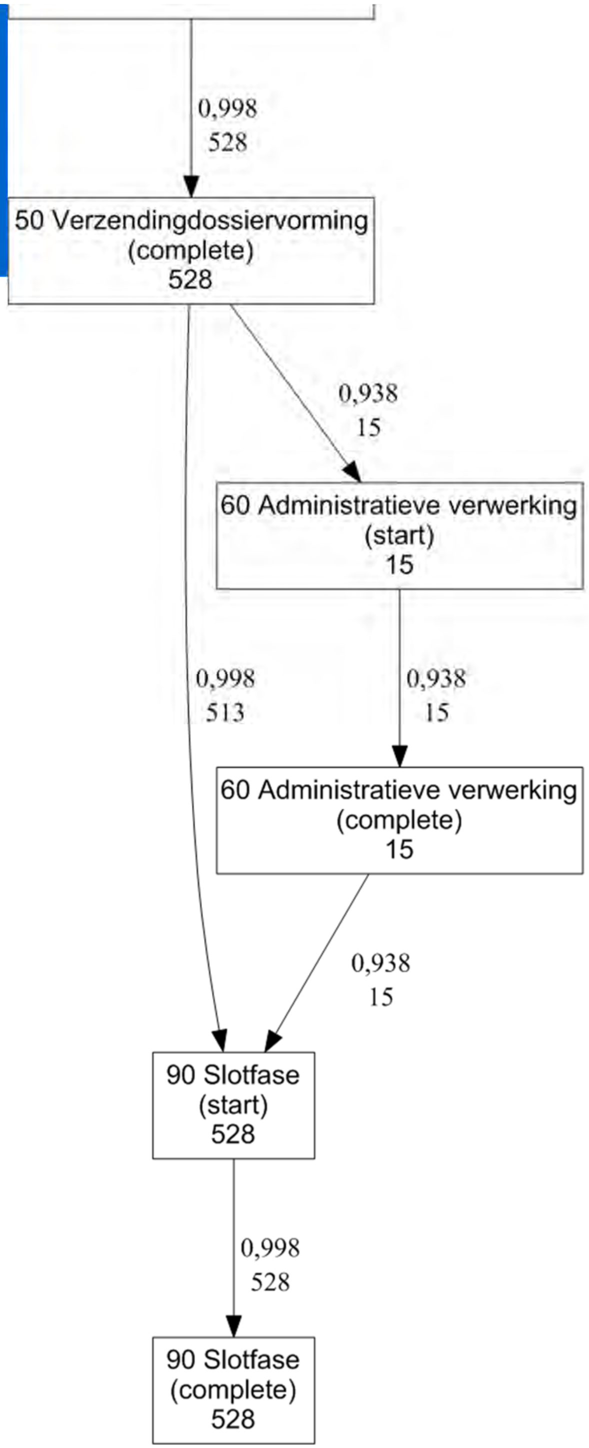
10 Proces registratie (start) 530
0,991 530
10 Proces registratie (complete) 530
0,998 530
20 Rapportage & beschikking (start) 532
0,987 532
20 Rapportage & beschikking (complete) 532
0,998
40 Toetsen en beslissen (start) 573
0,881 39
27 Retour (start) 43
0,909 529
0,909 15
0,975 29
0,957 24
0,95 19
40 Toetsen en beslissen (complete) 573
0,929 16
27 Retour (complete) 43
0,998 528
50 Verzendingdossiervorming (start) 528
0,998 528
50 Verzendingdossiervorming (complete) 528
0,938 15
60 Administratieve verwerking (start) 15
0,998 513
0,938 15
60 Administratieve verwerking (complete) 15
0,938 15
90 Slotfase (start) 528
0,998 528
90 Slotfase (complete) 528



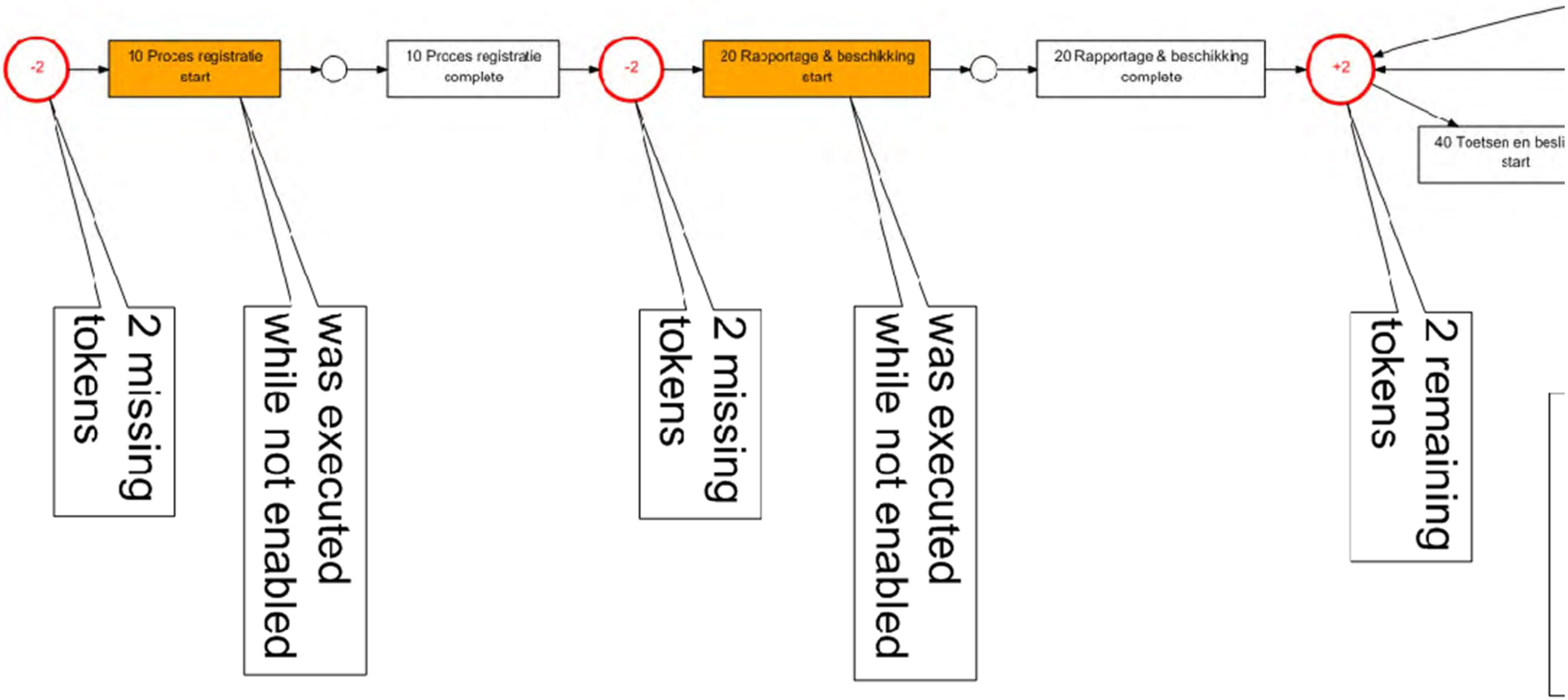
40 Toetsen en beslissen (start) 573
0,881 39
27 Retour (start) 43
0,909 529
0,909 15
0,975 29
0,957 24
0,95 19
40 Toetsen en beslissen (complete) 573
0,929 16
27 Retour (complete) 43
0,998 528
50 Verzendingdossiervorming (start) 528
0,998 528
50 Verzendingdossiervorming (complete) 528
0,938 15

C-net discovered using heuristic miner (3/3)

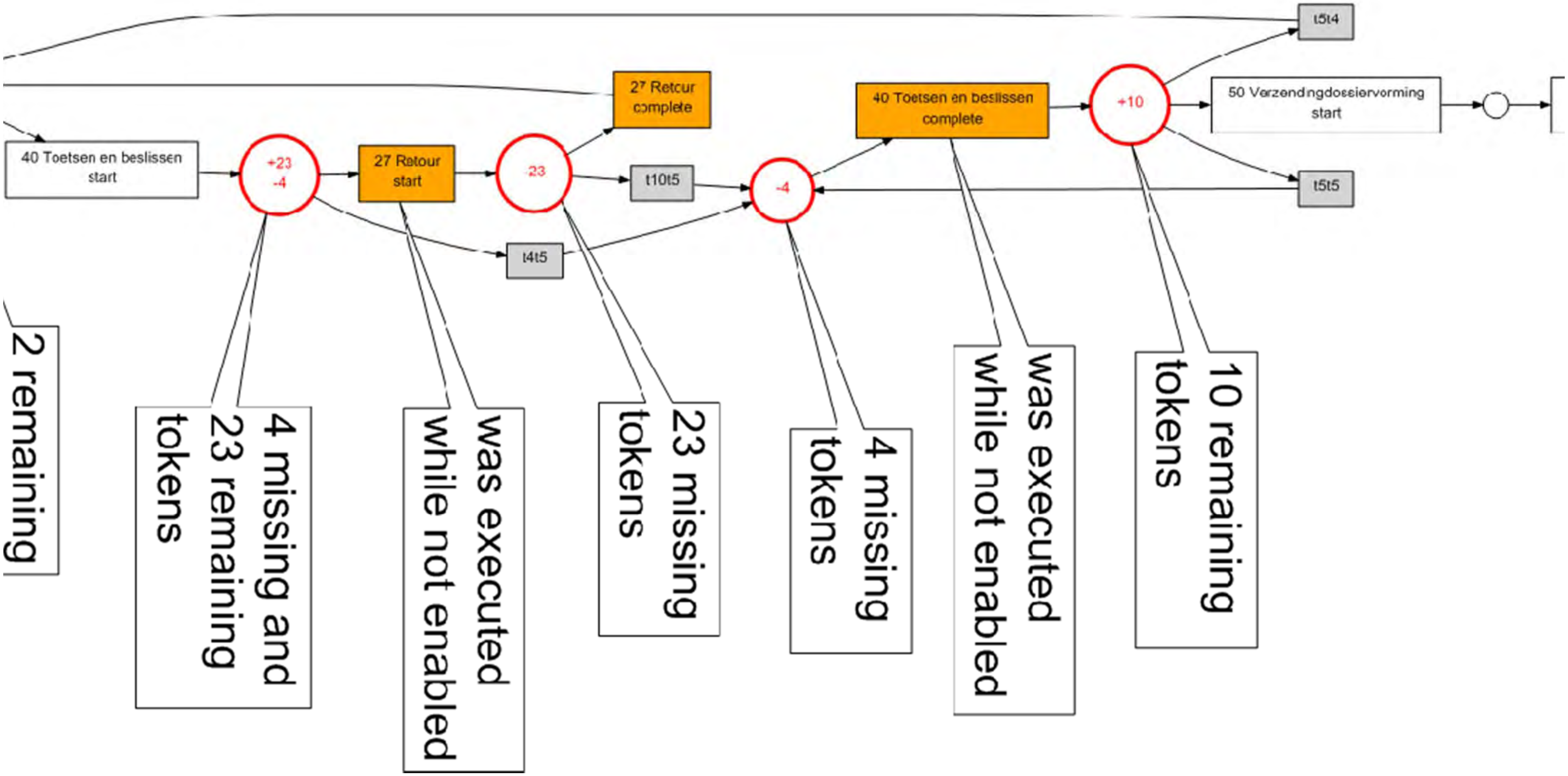

10 Proces registratie (start) 530
0,991 530
10 Proces registratie (complete) 530
0,998 530
20 Rapportage & beschikking (start) 532
0,987 532
20 Rapportage & beschikking (complete) 532
0,998 529
40 Toetsen en beslissen (start) 573
0,881 39
27 Retour (start) 43
0,909 529
0,909 15
0,975 29
0,957 24
0,95 19
40 Toetsen en beslissen (complete) 573
0,929 16
27 Retour (complete) 43
0,998 528
50 Verzendingdossiervorming (start) 528
0,998 528
50 Verzendingdossiervorming (complete) 528
0,938 15
60 Administratieve verwerking (start) 15
0,998 513
0,938 15
60 Administratieve verwerking (complete) 15
0,938 15
90 Slotfase (start) 528
0,998 528
90 Slotfase (complete) 528


0,998
528
50 Verzendingdossiervorming (complete) 528
0,938
15
60 Administratieve verwerking (start) 15
0,998
513
0,938
15
60 Administratieve verwerking (complete) 15
0,938
15
90 Slotfase (start) 528
0,998
528
90 Slotfase (complete) 528

Conformance check WMO process (1/3)


-2
10 Proces registratie start
10 Proces registratie complete
-2
20 Rapportage & beschikking start
20 Rapportage & beschikking complete
+2
40 Toetsen en besli start
2 missing tokens
was executed while not enabled
2 missing tokens
was executed while not enabled
2 remaining tokens

Conformance check WMO process (2/3)


t5t4
27 Retour complete
40 Toetsen en beslissen complete
+10
50 Verzendingdossiervorming start
40 Toetsen en beslissen start
+23
-4
27 Retour start
-23
t10t5
-4
t5t5
t4t5

2 remaining

4 missing and 23 remaining tokens

was executed while not enabled

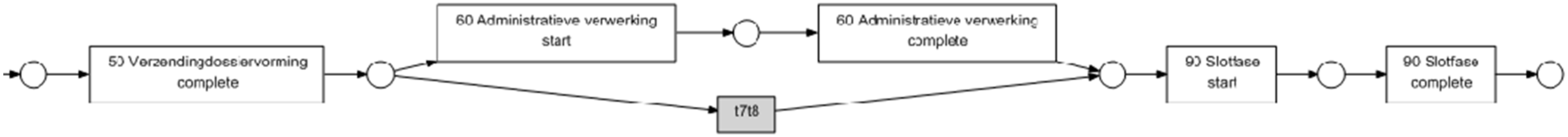
23 missing tokens

4 missing tokens

was executed while not enabled

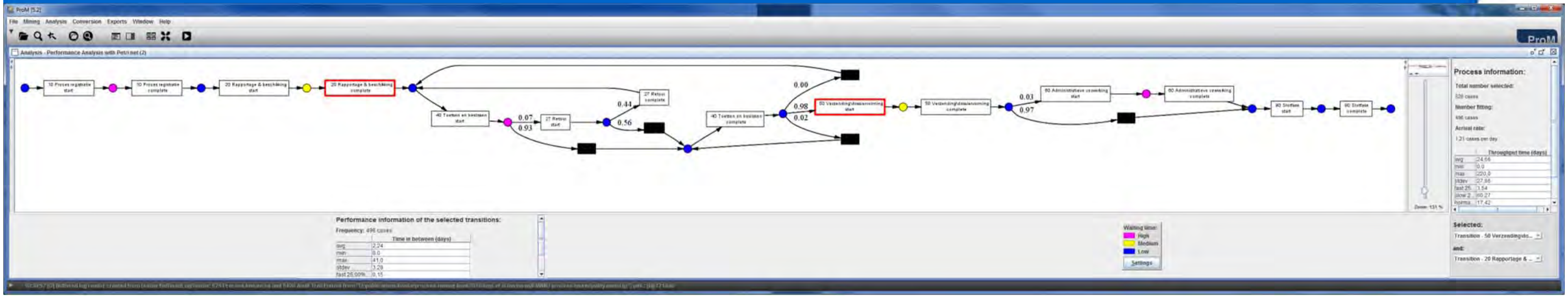
10 remaining tokens

Conformance check WMO process (3/3)


50 Verzendingdossiervorming complete
60 Administratieve verwerking start
60 Administratieve verwerking complete
t7t8
90 Slotfase start
90 Slotfase complete

The fitness of the discovered process is 0.99521667. Of the 528 cases, 496 cases fit perfectly whereas for 32 cases there are missing or remaining tokens.

Bottleneck analysis WMO process (1/3)


ProM [5.2]
File Mining Analysis Conversion Exports Window Help
Analysis - Performance Analysis with Petri net (2)
10 Proces registratie start
10 Proces registratie complete
20 Rapportage & beschikking start
20 Rapportage & beschikking complete
40 Toetsen en beslissen start
0.07
0.93
27 Retour start
0.44
0.56
27 Retour complete
40 Toetsen en beslissen complete
0.00
0.98
0.02
50 Verzending/dossiervorming start
50 Verzending/dossiervorming complete
0.03
0.97
60 Administratieve verwerking start
60 Administratieve verwerking complete
90 Slotfase start
90 Slotfase complete
Zoom: 131 %
Waiting time:
High
Medium
Low
Settings
ProM

Process information:

Total number selected:

528 cases

Number fitting:

496 cases

Arrival rate:

1.21 cases per day

	Throughput time (days)
avg	24.66
min	0.0
max	220.0
stdev	27.86
fast 25...	3.54
slow 2...	60.27
norma...	17.42

Selected:

Transition - 50 Verzending/do...

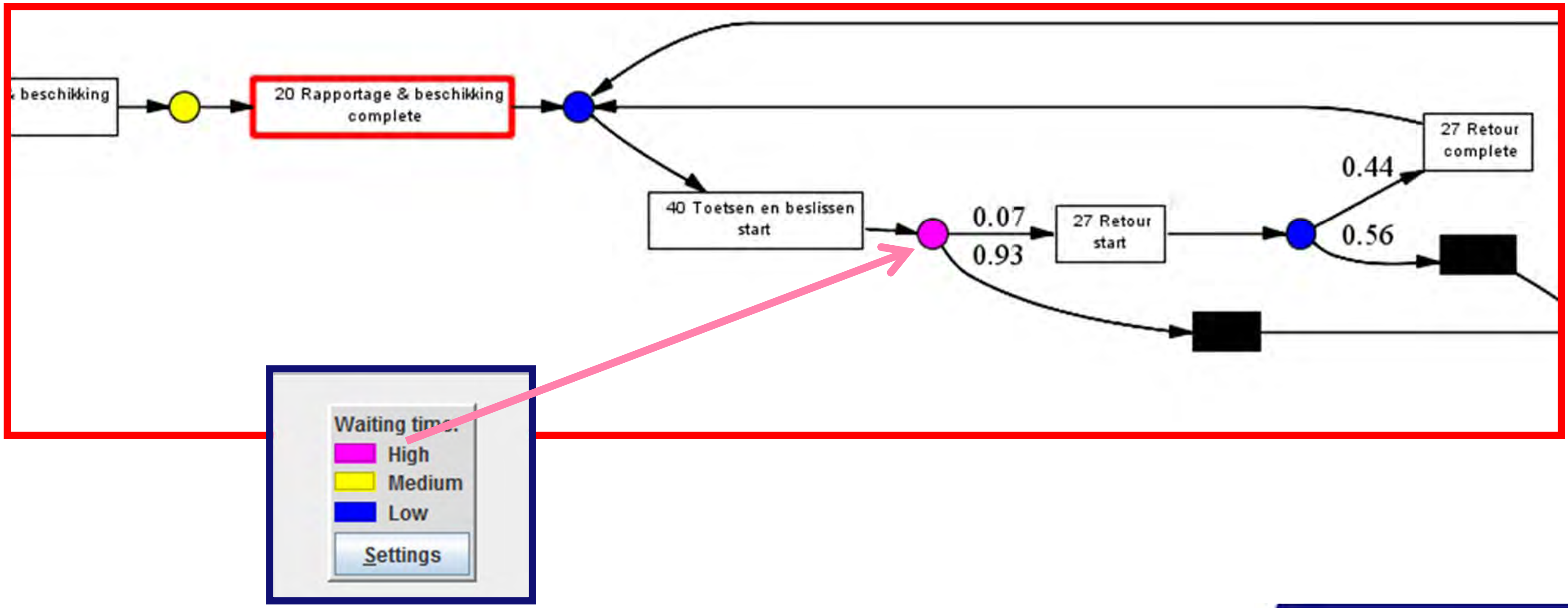
and:

Transition - 20 Rapportage & ...

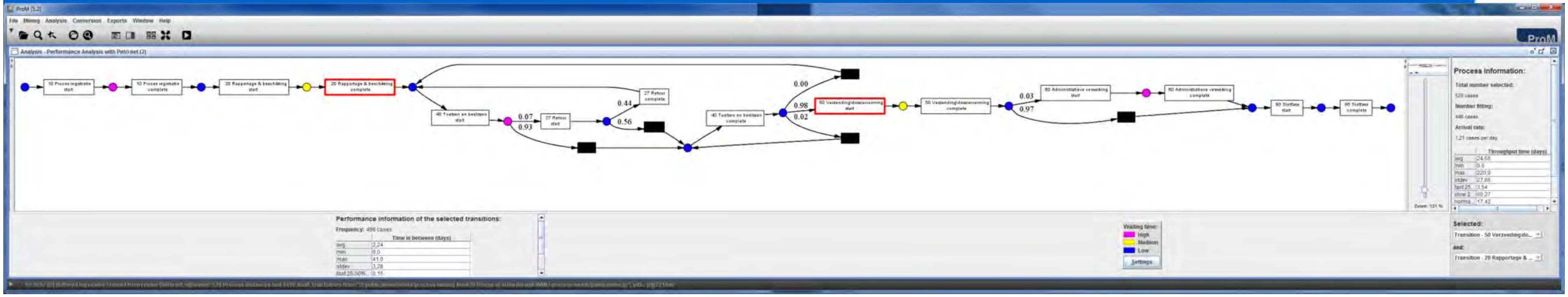
Performance information of the selected transitions:

Frequency: 496 cases

	Time in between (days)
avg	2.24
min	0.0
max	41.0
stdev	3.29
fast 25.00%	0.15



Bottleneck analysis WMO process (2/3)


ProM [5.2]
File Mining Analysis Conversion Exports Window Help
Analysis - Performance Analysis with Petri net (2)
10 Proces registratie start
10 Proces registratie complete
20 Rapportage & beschikking start
20 Rapportage & beschikking complete
40 Toetsen en beslissen start
0.07
0.93
27 Retour start
0.44
0.56
27 Retour complete
40 Toetsen en beslissen complete
0.00
0.98
0.02
50 Verzending\dossiervorming start
50 Verzending\dossiervorming complete
0.03
0.97
60 Administratieve verwerking start
60 Administratieve verwerking complete
90 Slotfase start
90 Slotfase complete
Zoom: 131 %
ProM
Settings

Process information:

Total number selected:

525 cases

Number fitting:

496 cases

Arrival rate:

1.21 cases per day

	Throughput time (days)
avg	24.66
min	0.0
max	220.0
stdev	27.88
fast 25	3.54
slow 2	60.27
norma...	17.42

Performance information of the selected transitions:

Frequency: 496 cases

	Time in between (days)
avg	2,24
min	0,0
max	41,0
stdev	3,28
fast 25,00%...	0,15

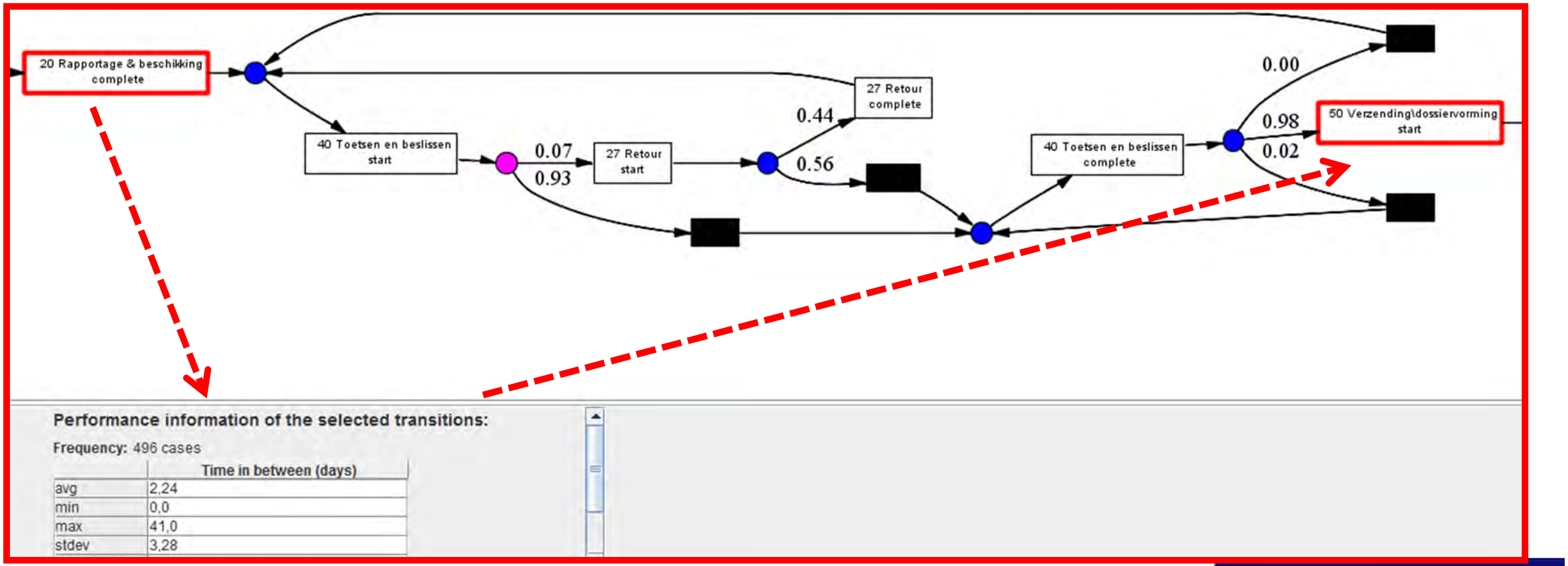
Waiting time:
- 🟪 High
- 🟨 Medium
- 🟦 Low

Selected:

Transition - 50 Verzending\do...

and:

Transition - 20 Rapportage & ...

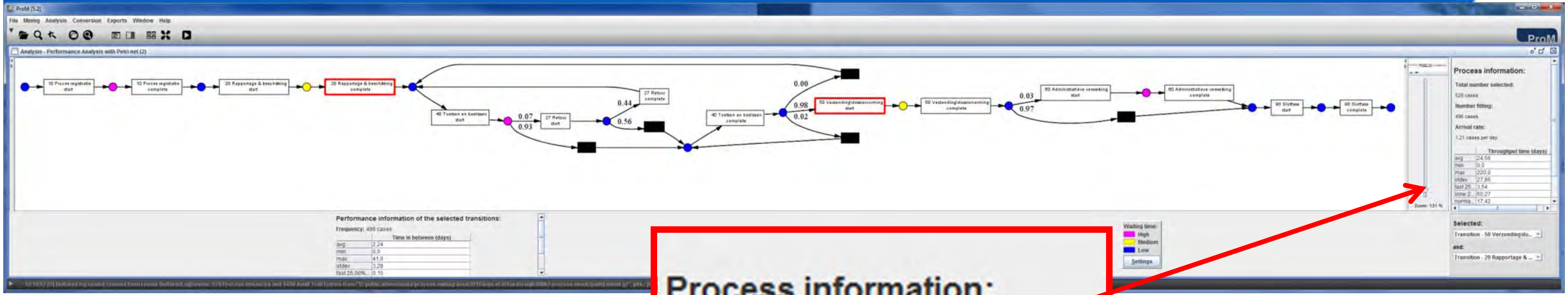

20 Rapportage & beschikking complete
40 Toetsen en beslissen start
0.07
0.93
27 Retour start
0.44
0.56
27 Retour complete
40 Toetsen en beslissen complete
0.00
0.98
0.02
50 Verzending\dossiervorming start

Performance information of the selected transitions:

Frequency: 496 cases

	Time in between (days)
avg	2,24
min	0,0
max	41,0
stdev	3,28

Bottleneck analysis WMO process (3/3)


ProM
File Mining Analysis Conversion Exports Window Help
Analysis - Performance Analysis with Petri net (2)
10 Proces registratie start
10 Proces registratie complete
20 Rapportage & beschikking start
20 Rapportage & beschikking complete
40 Toetsen en beslissen start
27 Retour start
27 Retour complete
40 Toetsen en beslissen complete
50 Verzending/dossiervorming start
50 Verzending/dossiervorming complete
60 Administratieve verwerking start
60 Administratieve verwerking complete
90 Slotfase start
90 Slotfase complete
0.07
0.93
0.44
0.56
0.00
0.98
0.02
0.03
0.97
Zoom: 131 %
Waiting time:
High
Medium
Low
Settings

Performance information of the selected transitions:

Frequency: 496 cases

	Time in between (days)
avg	2,24
min	0,0
max	41,0
stdev	3,28
fast 25.00%...	0,15

Selected:

Transition - 50 Verzending/do...

and:

Transition - 20 Rapportage & ...

flow time of approx. 25 days with a standard deviation of approx. 28

Process information:

Total number selected:

528 cases

Number fitting:

496 cases

Arrival rate:

1,21 cases per day

	Throughput time (days)
avg	24,66
min	0,0
max	220,0
stdev	27,86
fast 25...	3,54
slow 2...	60,27
norma...	17,42

Two additional Lasagna processes

RWS („Rijkswaterstaat”) process

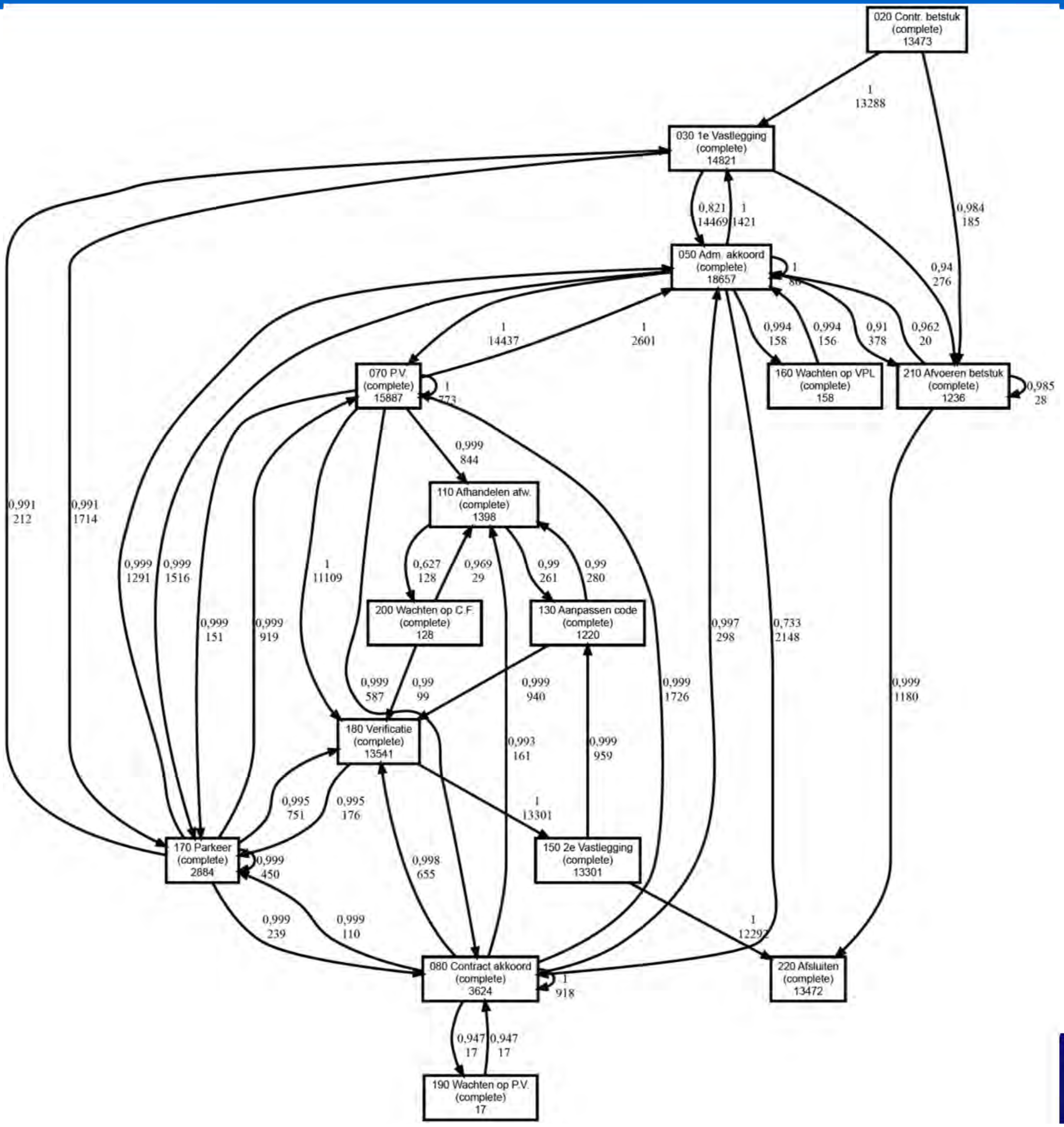
WOZ („Waardering Onroerende Zaken”) process

RWS Process

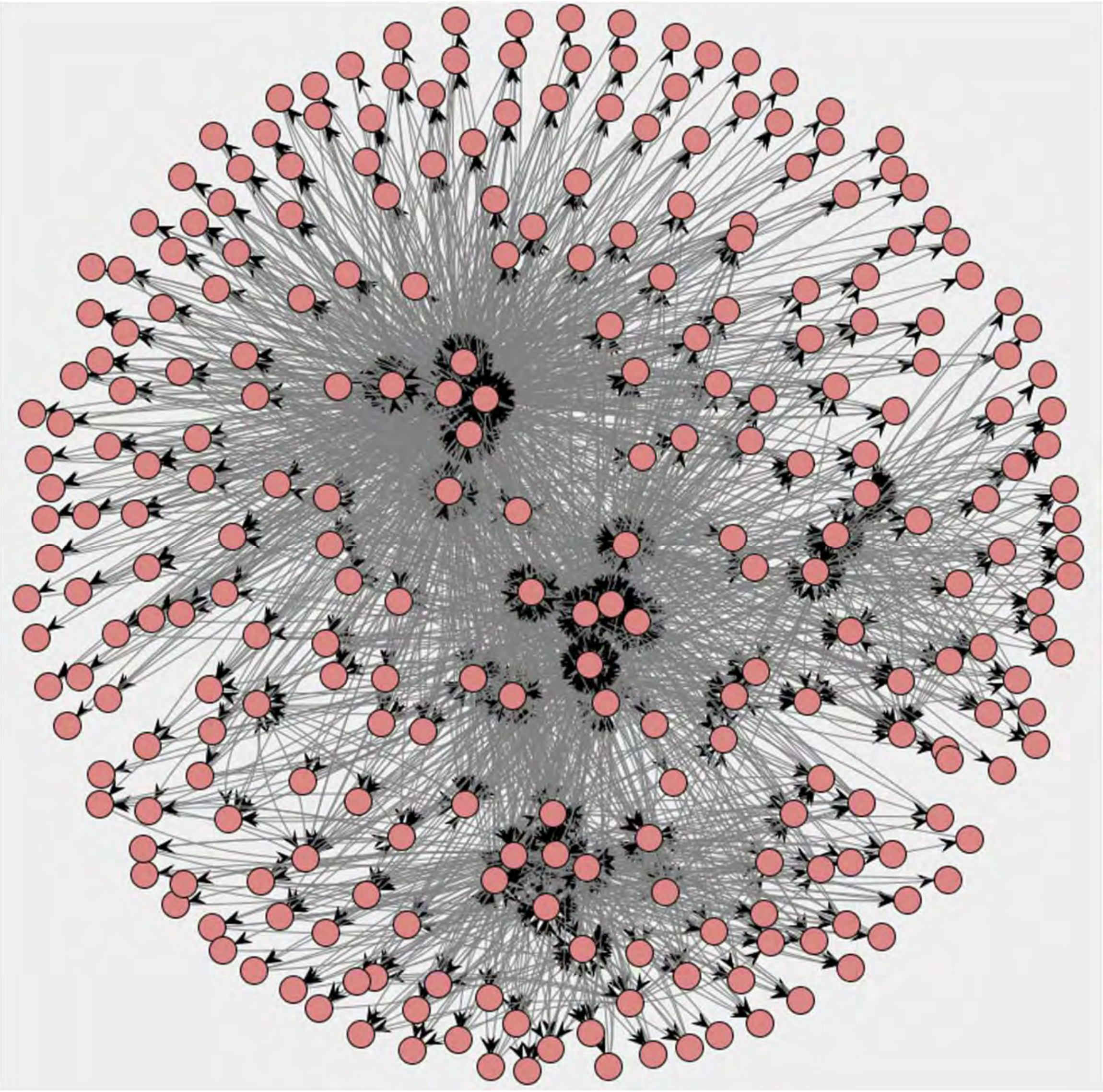
Rijkswaterstaat

- The Dutch national public works department, called “Rijkswaterstaat” (RWS), has twelve provincial offices. We analyzed the handling of invoices in one of these offices.
- The office employs about 1,000 civil servants and is primarily responsible for the construction and maintenance of the road and water infrastructure in its province.
- To perform its functions, the RWS office subcontracts various parties such as road construction companies, cleaning companies, and environmental bureaus. Also, it purchases services and products to support its construction, maintenance, and administrative activities.

C-net discovered using heuristic miner

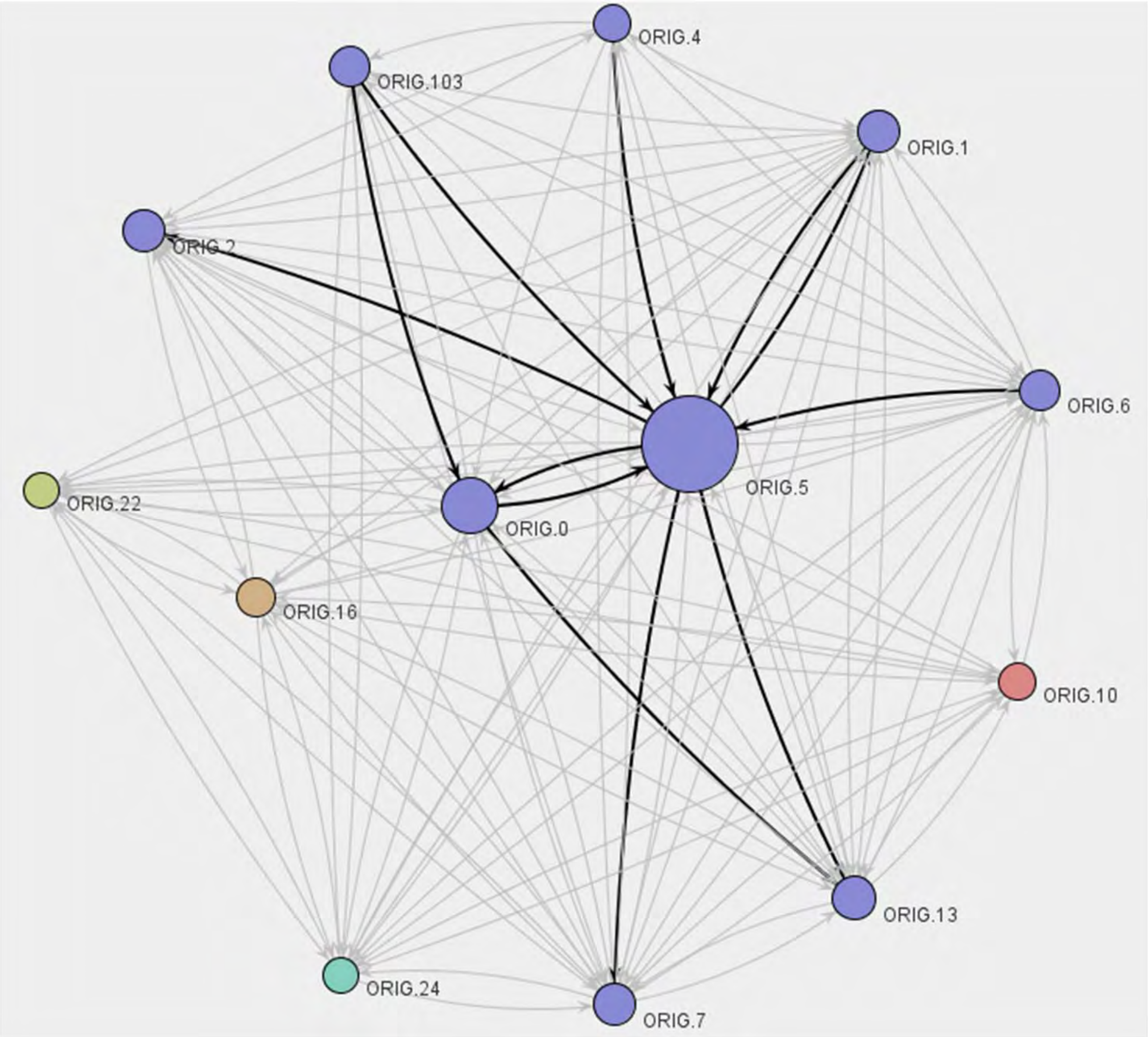

020 Contr. betstuk (complete) 13473
1 13288
0,984 185
030 1e Vastlegging (complete) 14821
0,821 14469
1 1421
0,94 276
050 Adm. akkoord (complete) 18657
1 80
1 14437
1 2601
0,994 158
0,994 156
0,91 378
0,962 20
160 Wachten op VPL (complete) 158
210 Afvoeren betstuk (complete) 1236
0,985 28
070 P.V. (complete) 15887
1 773
0,999 844
110 Afhandelen afw. (complete) 1398
0,991 212
0,991 1714
0,999 1291
0,999 1516
1 11109
0,627 128
0,969 29
0,99 261
0,99 280
200 Wachten op C.F. (complete) 128
130 Aanpassen code (complete) 1220
0,999 151
0,999 919
0,997 298
0,733 2148
0,999 587
0,99 99
0,999 940
0,999 1726
0,999 1180
180 Verificatie (complete) 13541
0,993 161
0,999 959
0,995 751
0,995 176
1 13301
170 Parkeer (complete) 2884
0,999 450
0,998 655
150 2e Vastlegging (complete) 13301
0,999 239
0,999 110
1 12292
080 Contract akkoord (complete) 3624
1 918
220 Afsluiten (complete) 13472
0,947 17
0,947 17
190 Wachten op P.V. (complete) 17

Social network constructed based on handovers of work



Each of the 271 nodes corresponds to a civil servant. Two civil servants are connected if one executed an activity causally following an activity executed by the other civil servant

Social network consisting of civil servants that executed more than 2000 activities in a 9 month period.

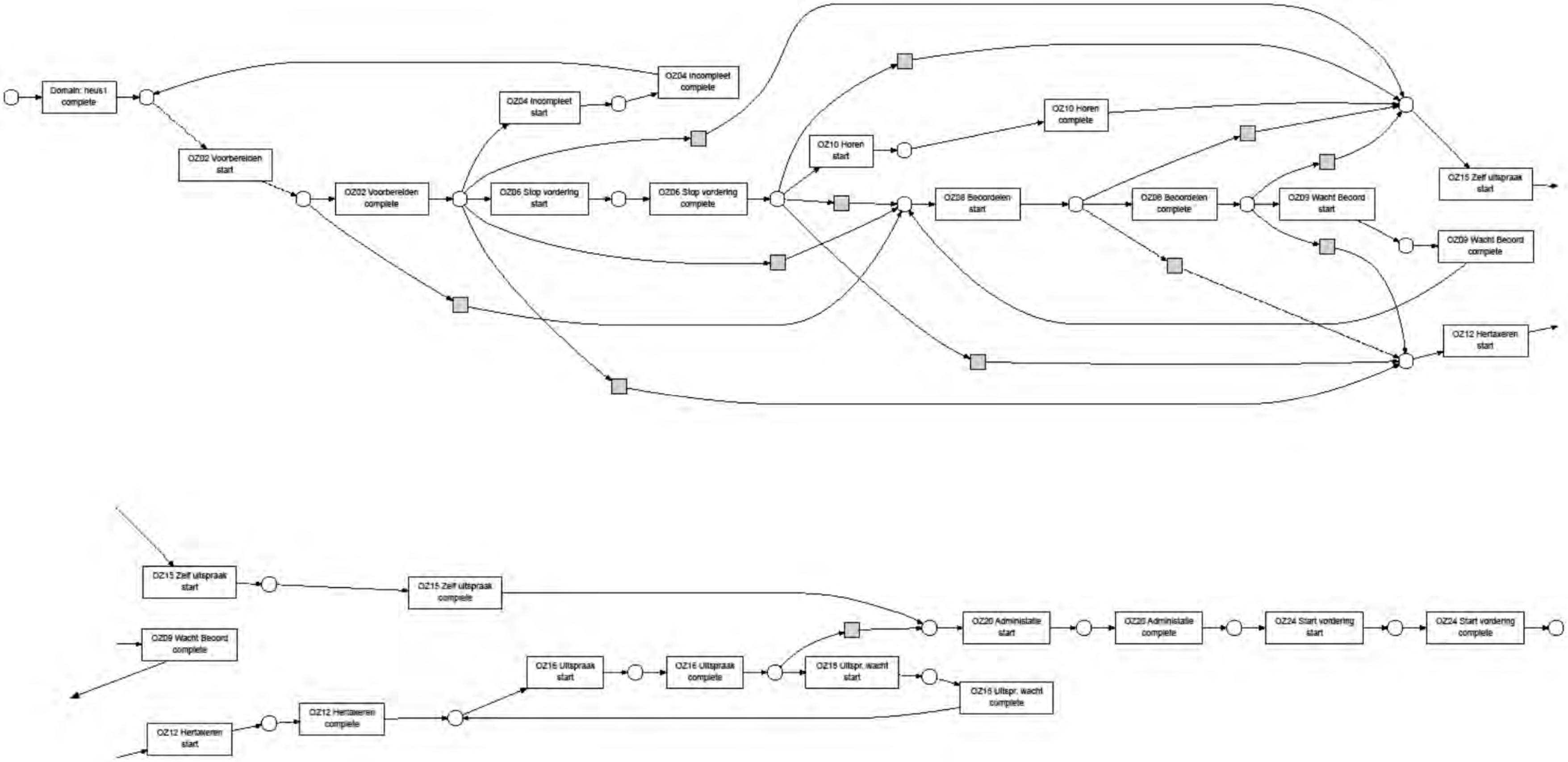


The darker arcs indicate the strongest relationships in the social network. Nodes having the same color belong to the same clique.

WOZ process

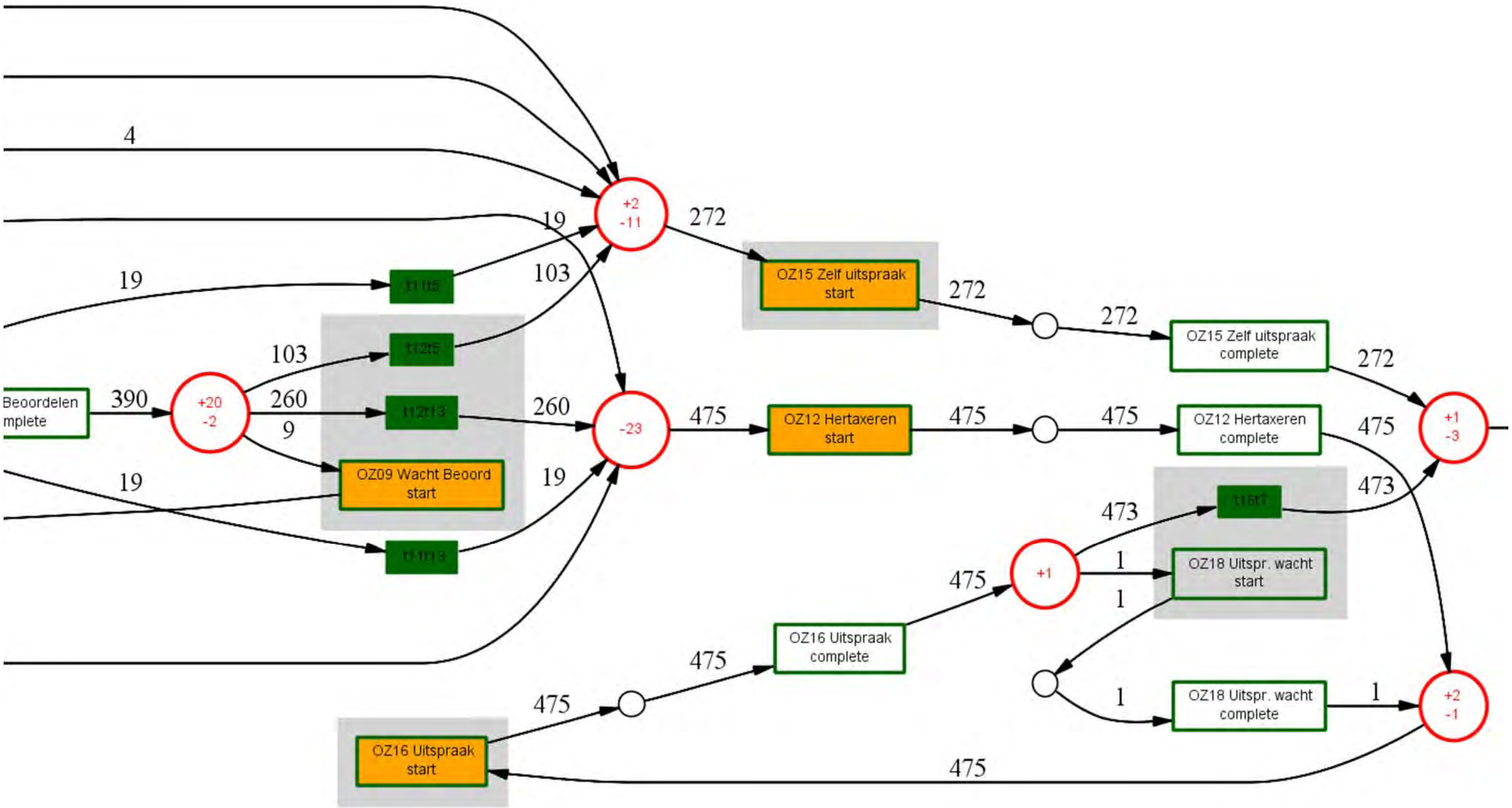
- Event log containing information about 745 objections against the so-called WOZ (“Waardering Onroerende Zaken”) valuation.
- Dutch municipalities need to estimate the value of houses and apartments. The WOZ value is used as a basis for determining the real-estate property tax.
- The higher the WOZ value, the more tax the owner needs to pay. Therefore, there are many objections (i.e., appeals) of citizens that assert that the WOZ value is too high.
- “WOZ process” discovered for another municipality (i.e., different from the one for which we analyzed the WMO process).

Discovered process model

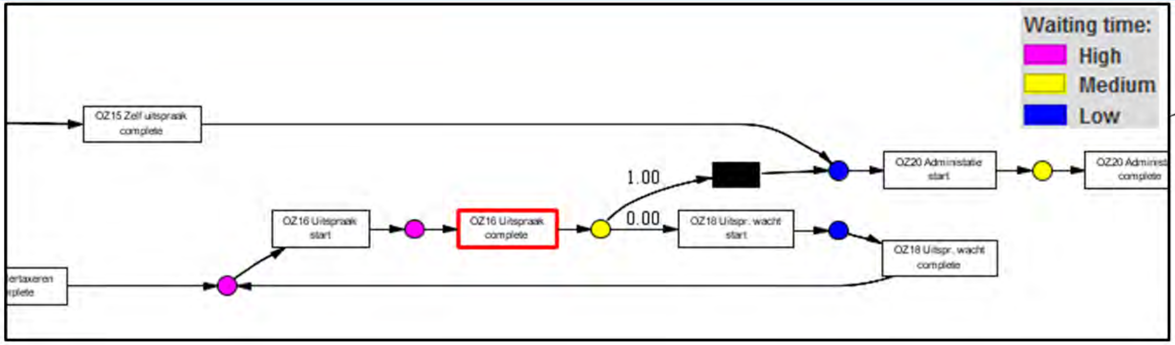


The log contains events related to 745 objections against the so-called WOZ valuation. These 745 objections generated 9583 events. There are 13 activities. For 12 of these activities both start and complete events are recorded. Hence, the WF-net has 25 transitions.

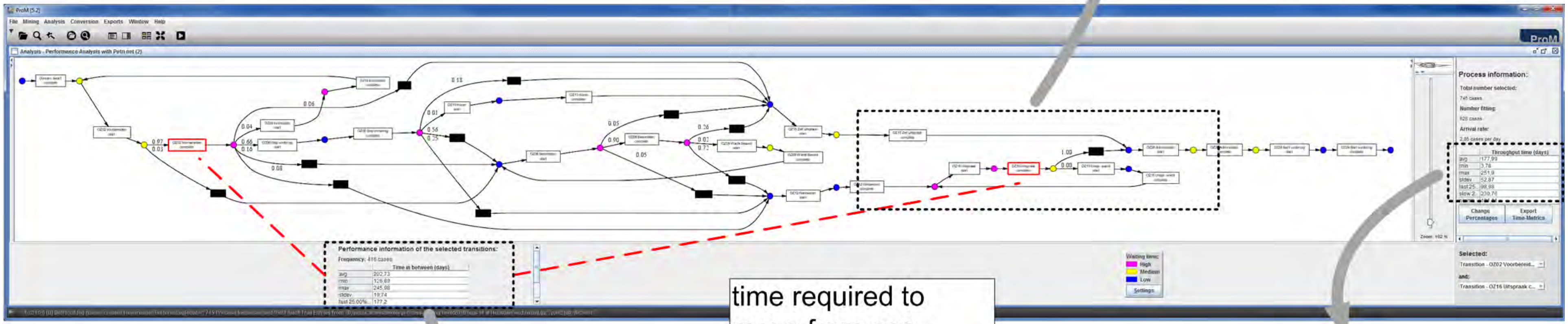
Conformance checker:
(fitness is 0.98876214)



Performance analysis

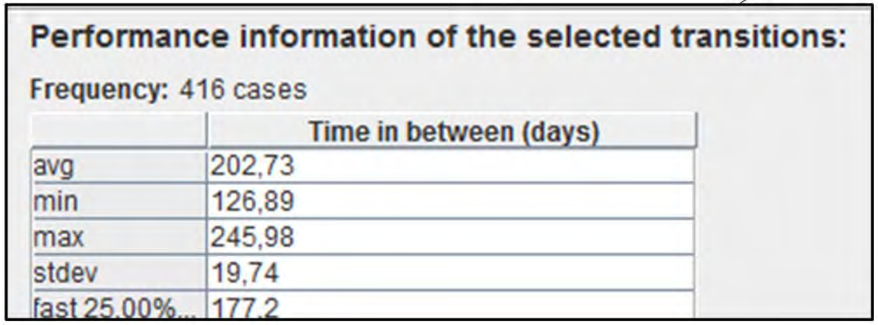


bottleneck detection: places are colored based on average durations



ProM 5.2
File Mining Analysis Conversion Exports Window Help
Analysis - Performance Analysis with Petri net (2)
Process information:
Waiting time:
High
Medium
Low
Settings
Change Percentages
Export Time-Metrics
Selected:
Transition - OZ02 Voorbereid...
and:
Transition - OZ16 Uitspraak c...

time required to move from one activity to another

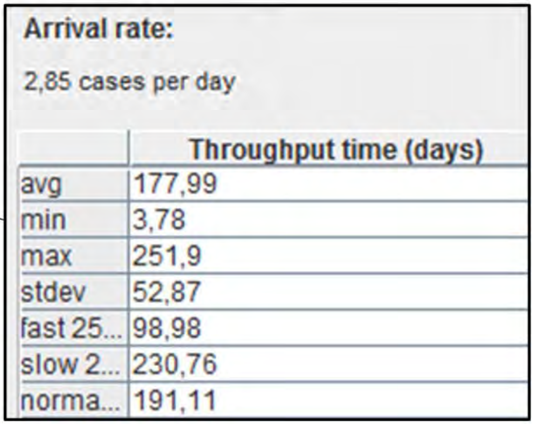


Performance information of the selected transitions:

Frequency: 416 cases

	Time in between (days)
avg	202,73
min	126,89
max	245,98
stdev	19,74
fast 25.00%...	177,2

information on total flow time



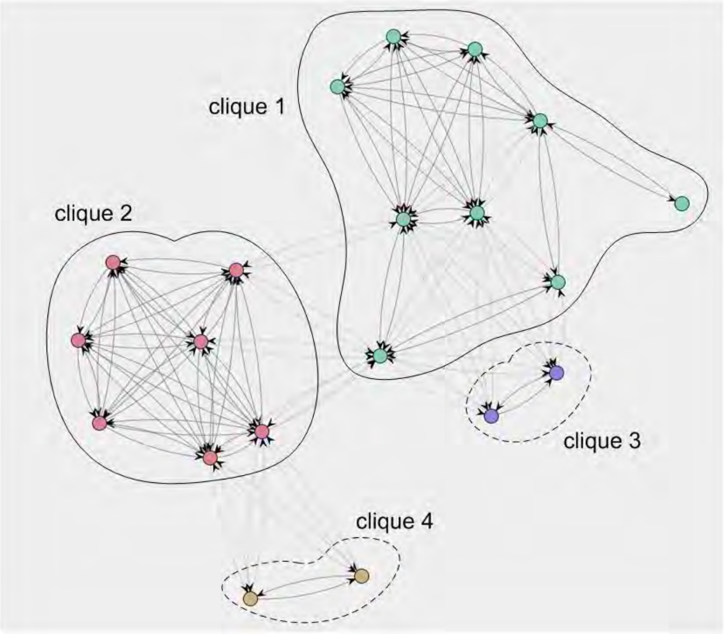
Arrival rate:

2,85 cases per day

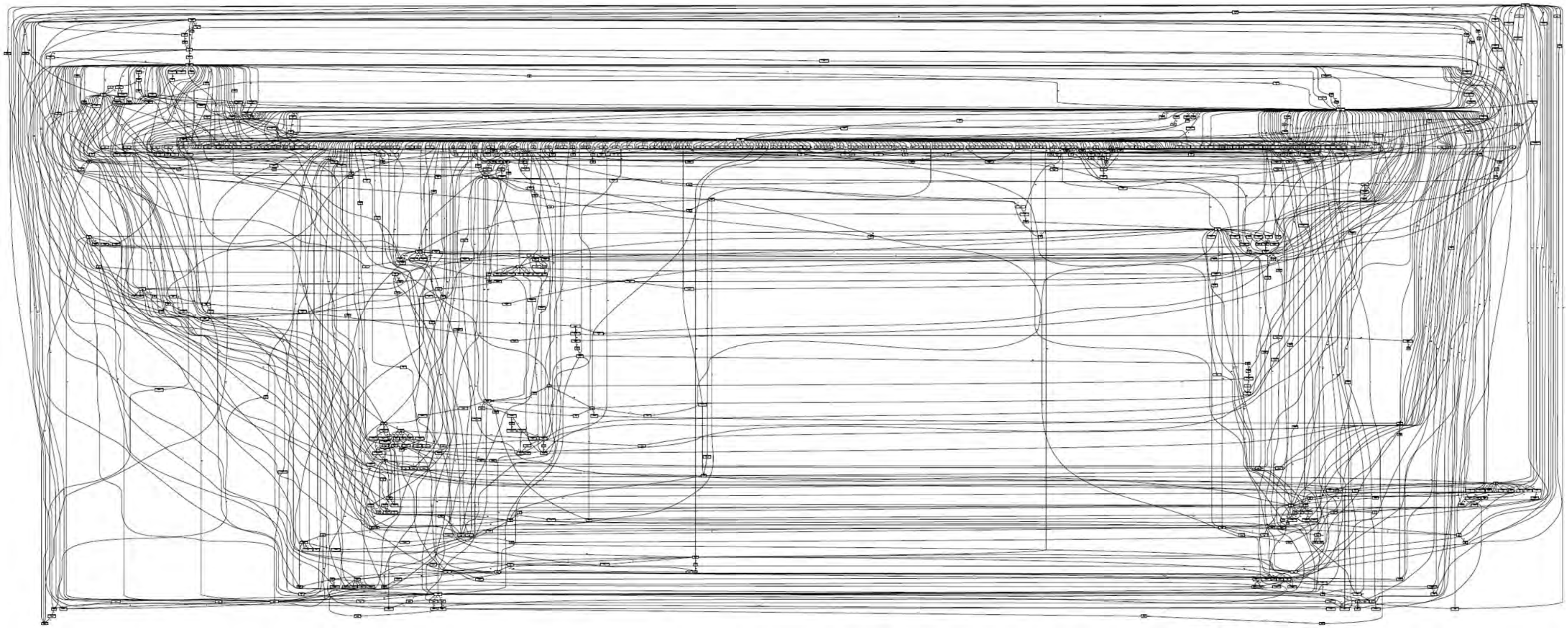
	Throughput time (days)
avg	177,99
min	3,78
max	251,9
stdev	52,87
fast 25...	98,98
slow 2...	230,76
norma...	191,11

Resource-activity matrix (four groups discovered)

user	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}	a_{12}	a_{13}
user 1	0	0	51	0	0	0	0	0	0	0	0	0	0
user 2	1	2	0	0	2	0	0	0	0	38	0	69	0
user 3	0	9	0	0	0	0	0	0	0	0	0	0	0
user 4	2	0	0	0	0	0	0	0	0	0	0	0	0
user 5	117	0	4	0	3	0	0	0	0	1	0	20	6
user 6	172	6	14	0	7	3	0	0	1	2	0	48	53
user 7	1	41	8	14	275	8	8	865	55	180	0	128	5
user 8	2	868	7	6	105	0	0	79	266	441	0	844	3
user 9	90	0	2	0	1	2	0	0	1	2	0	27	28
user 10	0	0	0	899	0	0	0	0	0	0	0	0	1019
user 11	336	1	3	1	4	2	0	0	0	1	0	18	23
user 12	1	645	13	21	419	3	0	3	217	281	1	334	9
user 13	0	1	0	0	0	0	0	0	0	0	0	0	0
user 14	0	0	0	0	0	0	0	0	0	1	0	0	0
user 15	0	0	0	0	0	0	0	2	2	0	0	2	0
user 16	1	3	3	2	1	0	0	1	2	3	1	0	0
user 17	0	4	0	0	0	0	0	0	0	0	0	0	0
user 18	9	0	0	0	0	0	0	0	0	0	0	0	0
user 19	13	1	0	0	1	0	0	0	0	0	0	4	0
user 20	0	0	0	21	0	0	0	0	0	0	0	0	258



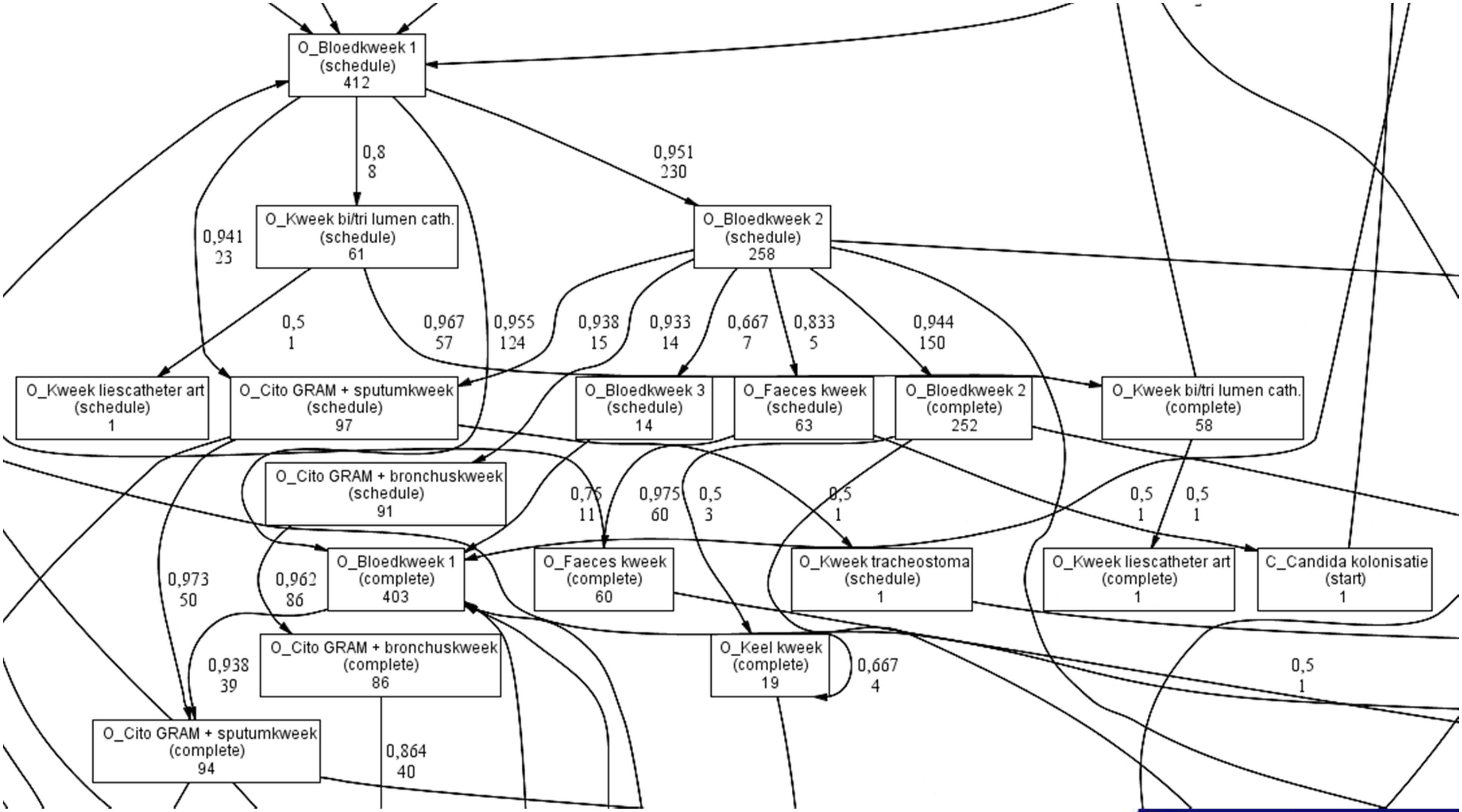
Example of a Spaghetti process



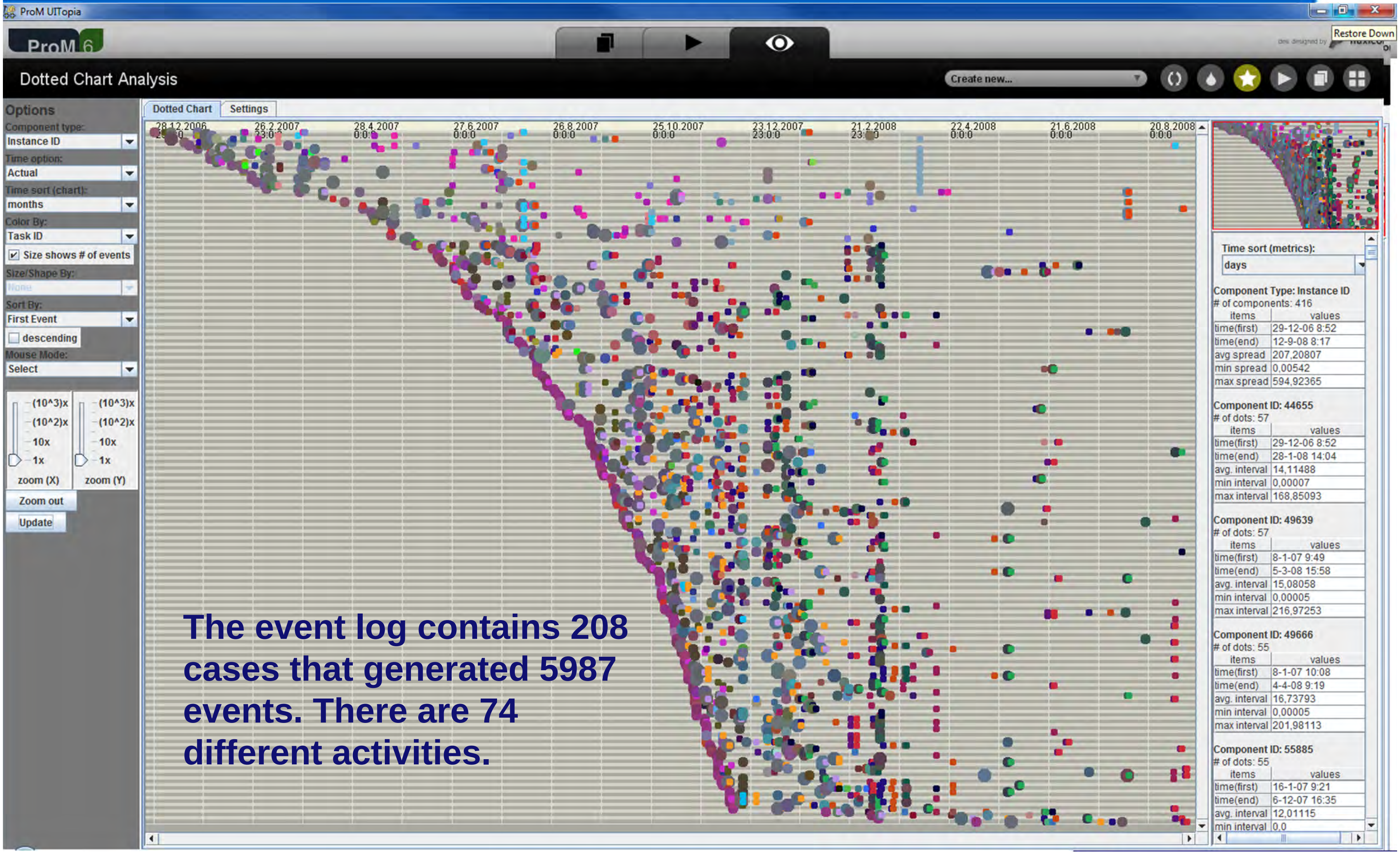
Spaghetti process describing the diagnosis and treatment of 2765 patients in a Dutch hospital. The process model was constructed based on an event log containing 114,592 events. There are 619 different activities (taking event types into account) executed by 266 different individuals (doctors, nurses, etc.).

Fragment

18 activities of the 619 activities (2.9%)



Another example
(event log of Dutch housing agency)



Options

Component type: Instance ID

Time option: Actual

Time sort (chart): months

Color By: Task ID

☑ Size shows # of events

Size/Shape By: None

Sort By: First Event

☐ descending

Mouse Mode: Select

Time sort (metrics): days

Component Type: Instance ID
# of components: 416

items	values
time(first)	29-12-06 8:52
time(end)	12-9-08 8:17
avg spread	207,20807
min spread	0,00542
max spread	594,92365

Component ID: 44655
# of dots: 57

items	values
time(first)	29-12-06 8:52
time(end)	28-1-08 14:04
avg. interval	14,11488
min interval	0,00007
max interval	168,85093

Component ID: 49639
# of dots: 57

items	values
time(first)	8-1-07 9:49
time(end)	5-3-08 15:58
avg. interval	15,08058
min interval	0,00005
max interval	216,97253

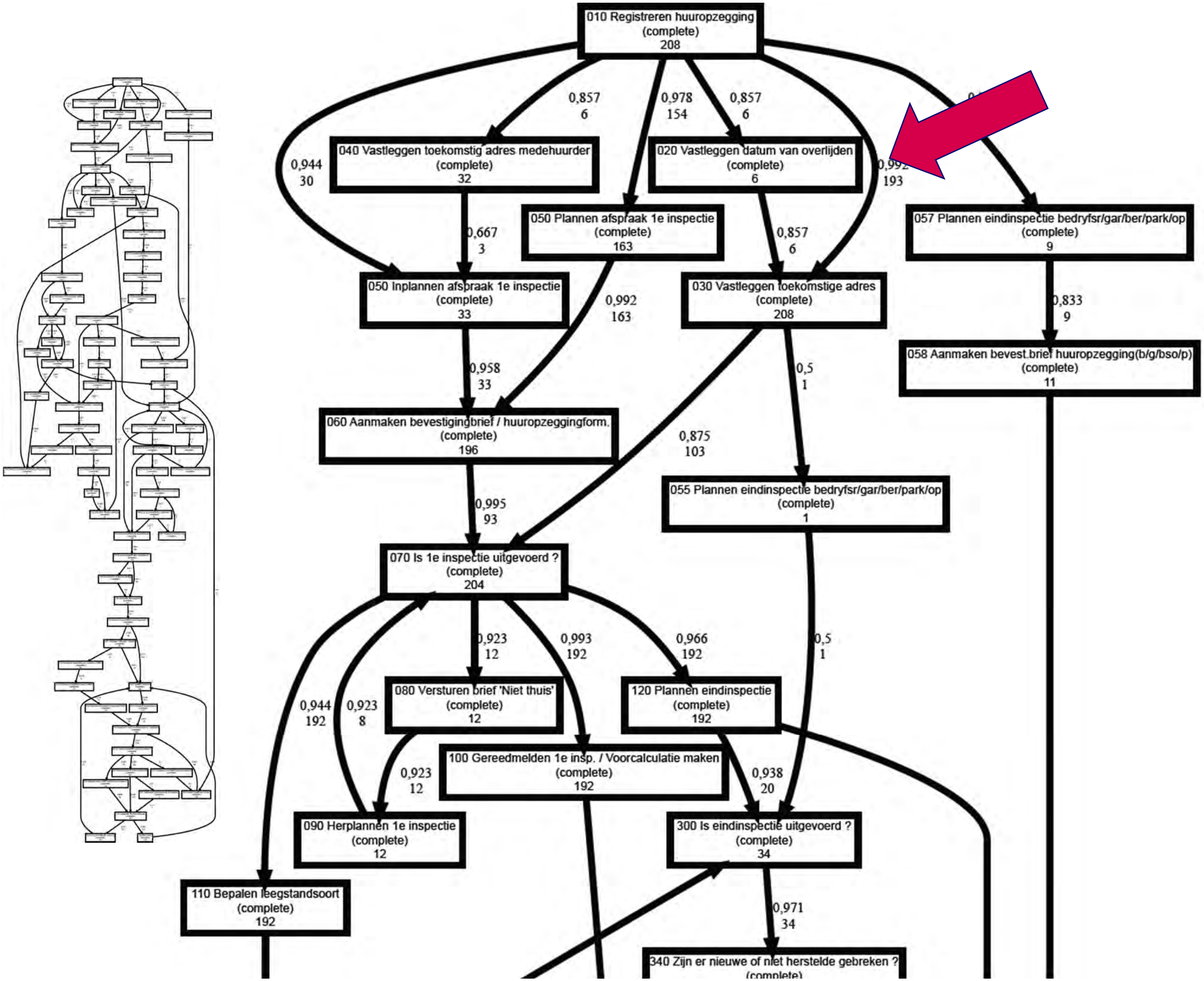
Component ID: 49666
# of dots: 55

items	values
time(first)	8-1-07 10:08
time(end)	4-4-08 9:19
avg. interval	16,73793
min interval	0,00005
max interval	201,98113

Component ID: 55885
# of dots: 55

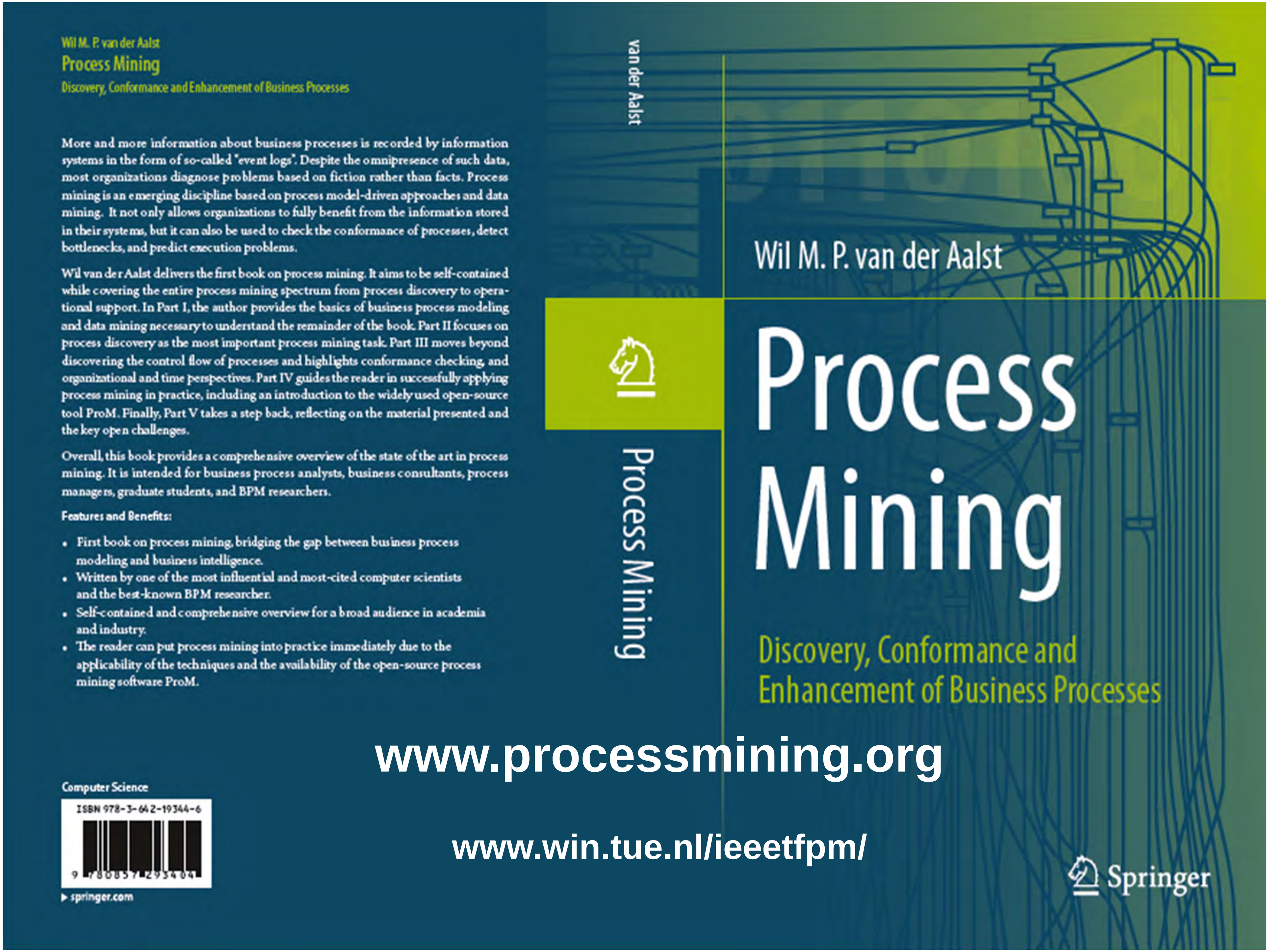
items	values
time(first)	16-1-07 9:21
time(end)	6-12-07 16:35
avg. interval	12,01115
min interval	0,0

The event log contains 208 cases that generated 5987 events. There are 74 different activities.


010 Registreren huuropzegging
(complete)
208
0,857
6
0,978
154
0,857
6
0,944
30
0,992
193
040 Vastleggen toekomstig adres medehuurder
(complete)
32
020 Vastleggen datum van overlijden
(complete)
6
050 Plannen afspraak 1e inspectie
(complete)
163
057 Plannen eindinspectie bedryfsr/gar/ber/park/op
(complete)
9
0,667
3
0,857
6
050 Inplannen afspraak 1e inspectie
(complete)
33
0,992
163
030 Vastleggen toekomstige adres
(complete)
208
0,833
9
058 Aanmaken bevest.brief huuropzegging(b/g/bso/p)
(complete)
11
0,958
33
0,5
1
060 Aanmaken bevestigingbrief / huuropzeggingform.
(complete)
196
0,875
103
055 Plannen eindinspectie bedryfsr/gar/ber/park/op
(complete)
1
0,995
93
070 Is 1e inspectie uitgevoerd ?
(complete)
204
0,923
12
0,993
192
0,966
192
0,5
1
0,944
192
0,923
8
080 Versturen brief 'Niet thuis'
(complete)
12
120 Plannen eindinspectie
(complete)
192
100 Gereedmelden 1e insp. / Voorcalculatie maken
(complete)
192
0,923
12
0,938
20
090 Herplannen 1e inspectie
(complete)
12
300 Is eindinspectie uitgevoerd ?
(complete)
34
110 Bepalen leegstandsoort
(complete)
192
0,971
34
340 Zijn er nieuwe of niet herstelde gebreken ?
(complete)

Conclusion

- Many concurrency-related BPM challenges.
- Process leave traces in event logs. So, if you are interested in processes, use them!
- Process mining: challenging and highly relevant.
- Process discovery challenge
  - balancing between different objectives
  - only example behavior
- Conformance checking challenge
  - finding the most likely trace
  - dealing with silent/duplicate steps
- Eldorado for exciting concurrency research!



Wil M. P. van der Aalst

Process Mining

Discovery, Conformance and Enhancement of Business Processes

More and more information about business processes is recorded by information systems in the form of so-called "event logs". Despite the omnipresence of such data, most organizations diagnose problems based on fiction rather than facts. Process mining is an emerging discipline based on process model-driven approaches and data mining. It not only allows organizations to fully benefit from the information stored in their systems, but it can also be used to check the conformance of processes, detect bottlenecks, and predict execution problems.

Wil van der Aalst delivers the first book on process mining. It aims to be self-contained while covering the entire process mining spectrum from process discovery to operational support. In Part I, the author provides the basics of business process modeling and data mining necessary to understand the remainder of the book. Part II focuses on process discovery as the most important process mining task. Part III moves beyond discovering the control flow of processes and highlights conformance checking, and organizational and time perspectives. Part IV guides the reader in successfully applying process mining in practice, including an introduction to the widely used open-source tool ProM. Finally, Part V takes a step back, reflecting on the material presented and the key open challenges.

Overall, this book provides a comprehensive overview of the state of the art in process mining. It is intended for business process analysts, business consultants, process managers, graduate students, and BPM researchers.

Features and Benefits:

- First book on process mining, bridging the gap between business process modeling and business intelligence.
- Written by one of the most influential and most-cited computer scientists and the best-known BPM researcher.
- Self-contained and comprehensive overview for a broad audience in academia and industry.
- The reader can put process mining into practice immediately due to the applicability of the techniques and the availability of the open-source process mining software ProM.

Computer Science

ISBN 978-3-642-19344-6

▸ springer.com

van der Aalst

Process Mining

Wil M. P. van der Aalst

Process Mining

Discovery, Conformance and Enhancement of Business Processes

Springer

www.processmining.org

www.win.tue.nl/ieeetfpm/