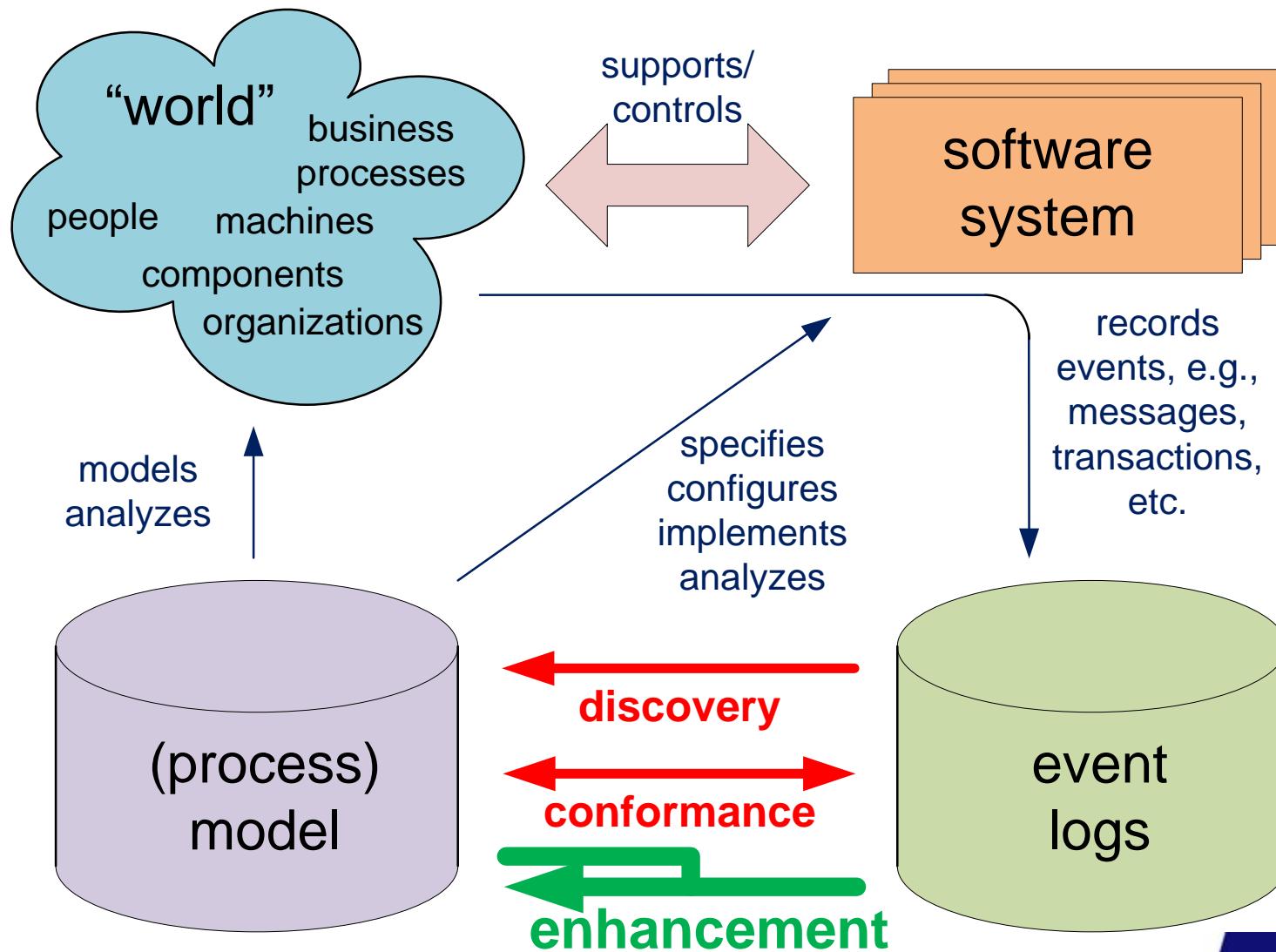
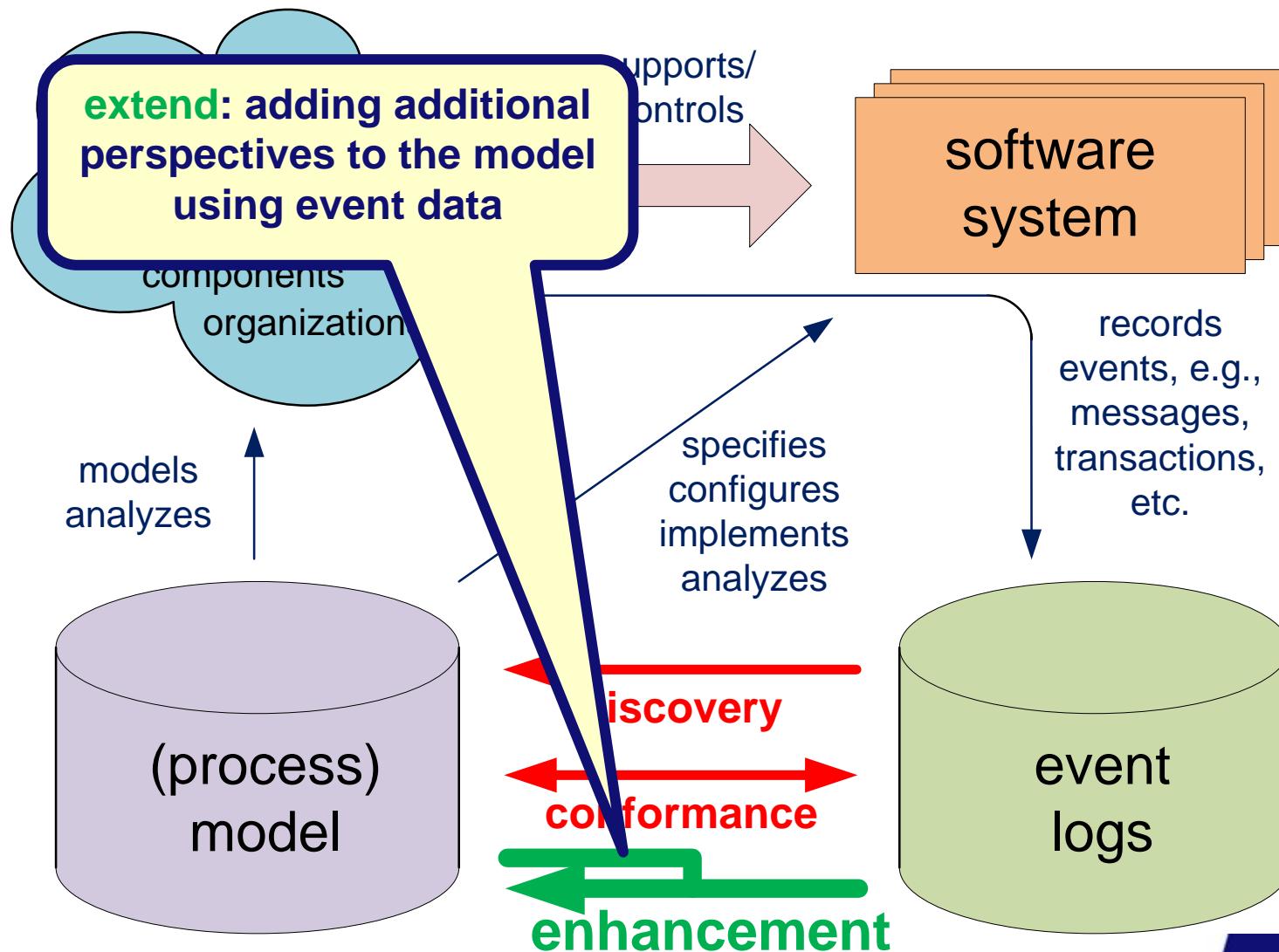


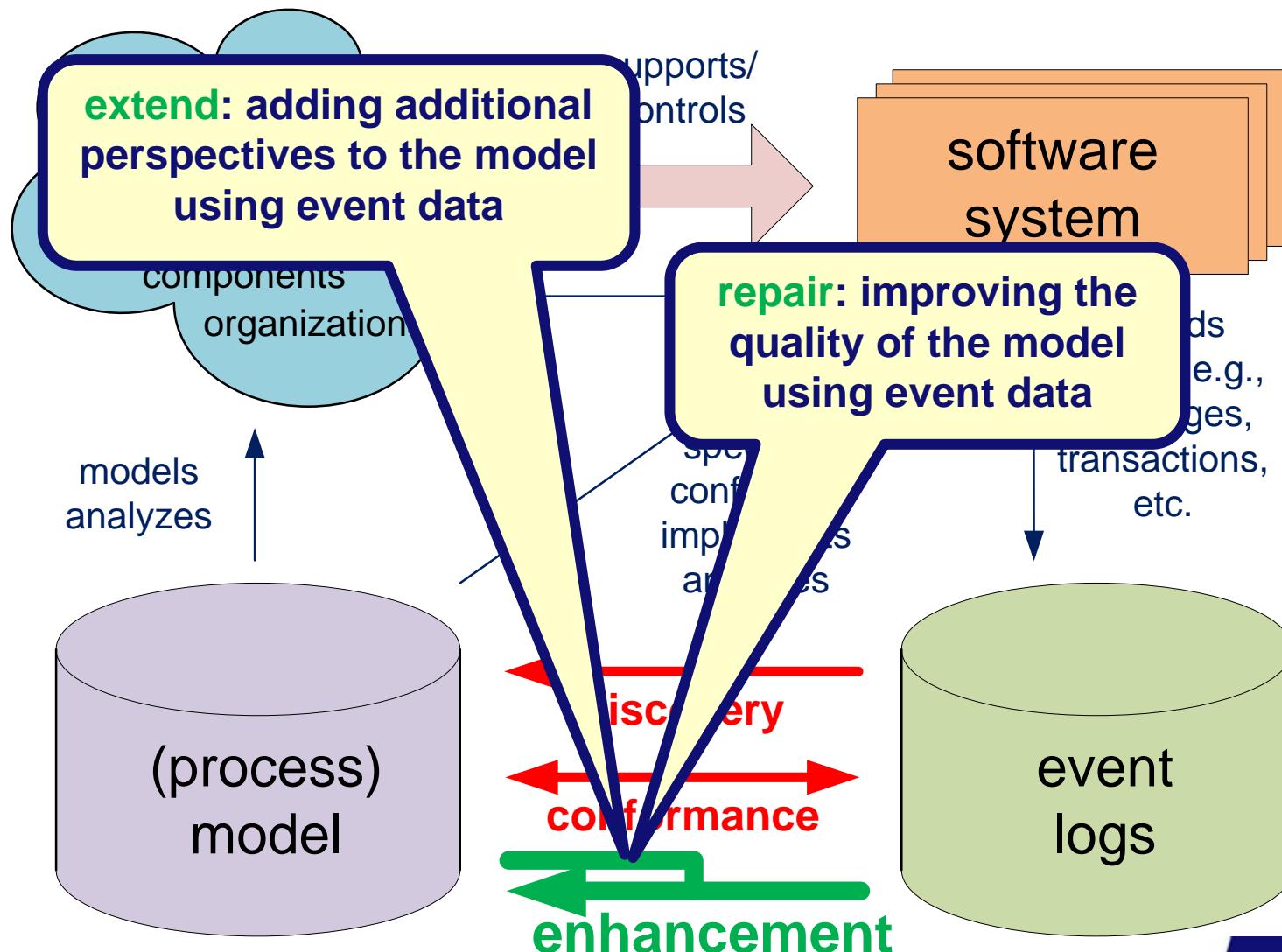
Enhancement: Extension and Repair



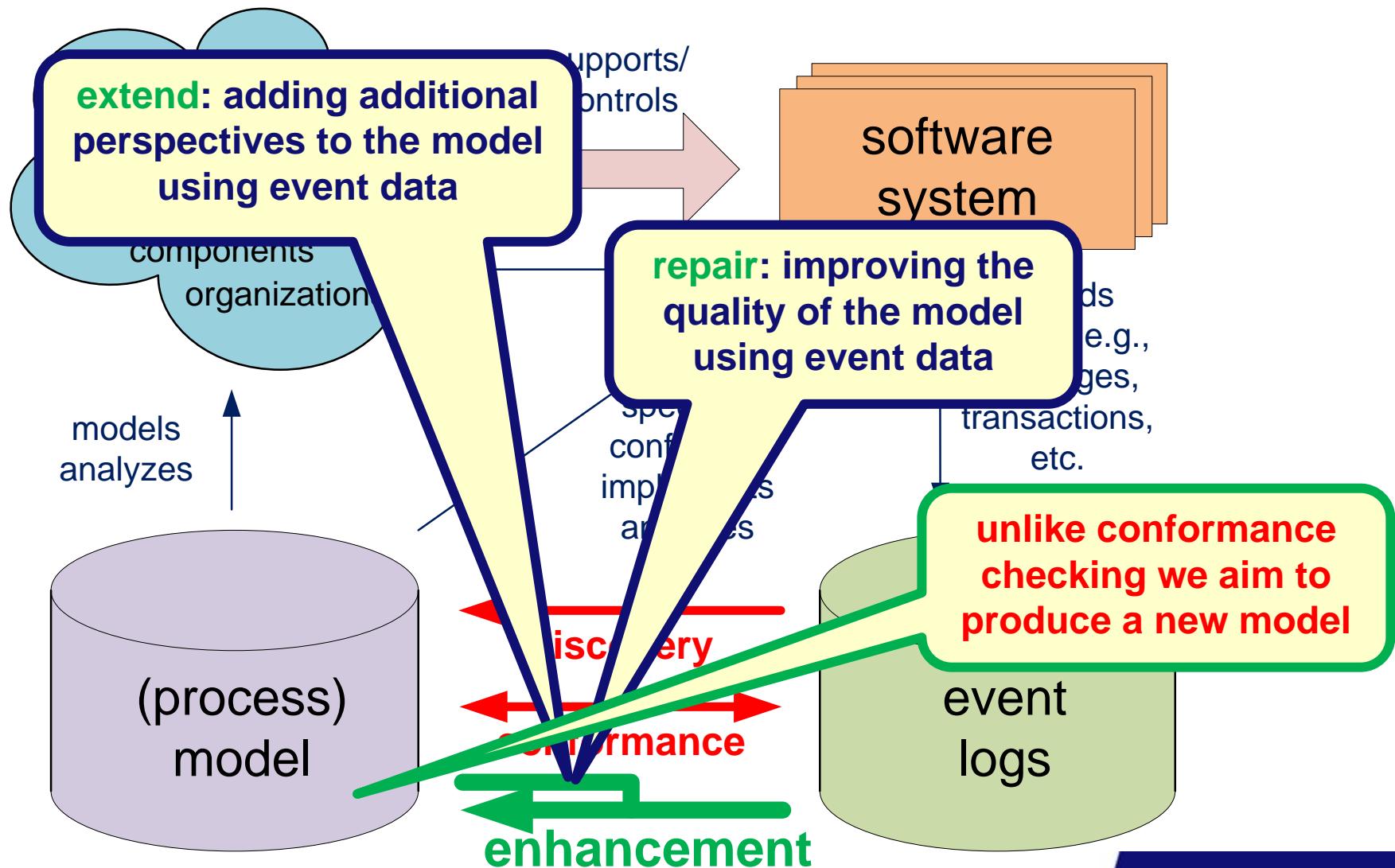
Enhancement: Extension and Repair



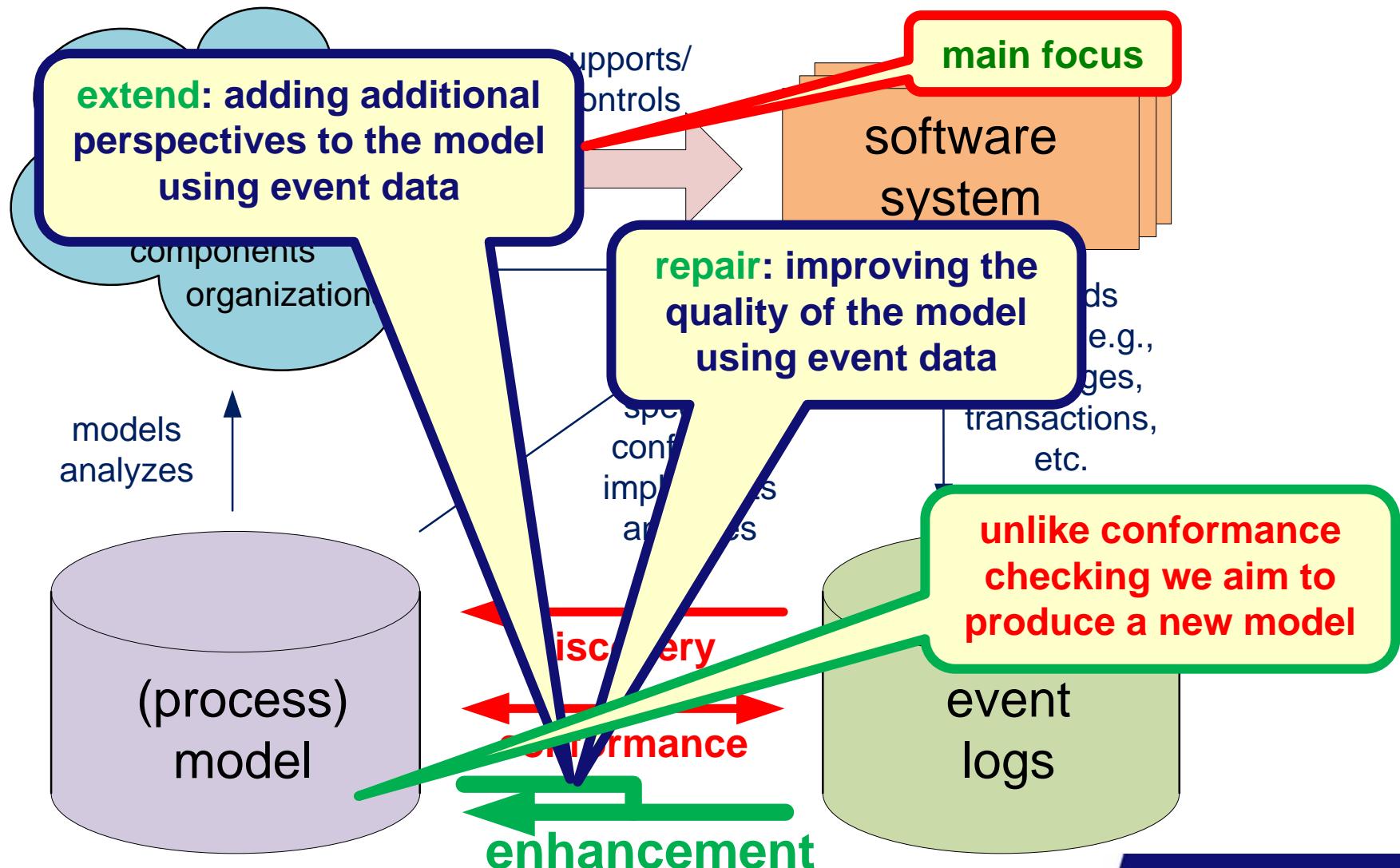
Enhancement: Extension and Repair



Enhancement: Extension and Repair



Enhancement: Extension and Repair



repair
"paving the cow paths"

Repairing a model based on event data

repair = pave the cow paths

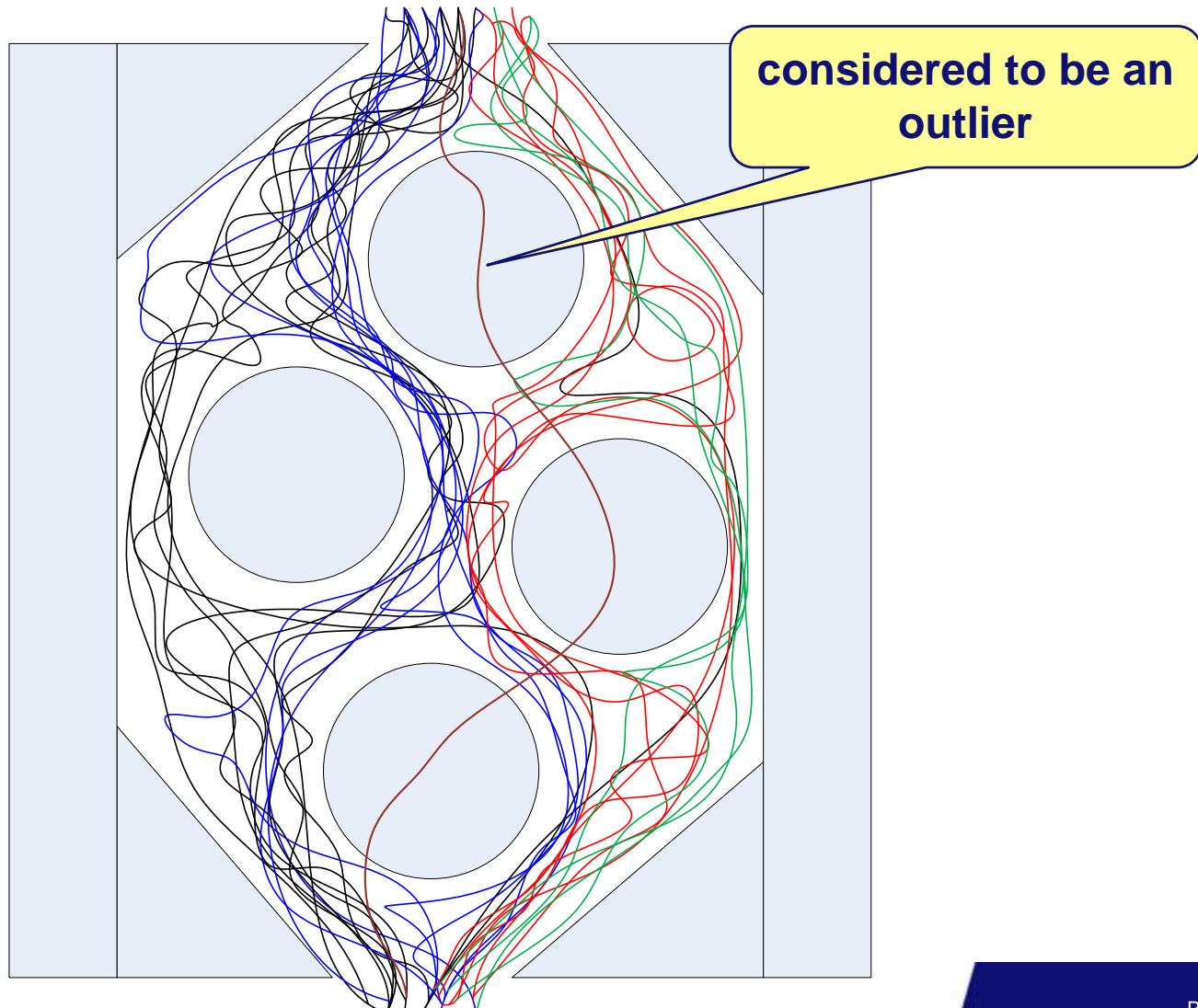


Repairing a model based on event data

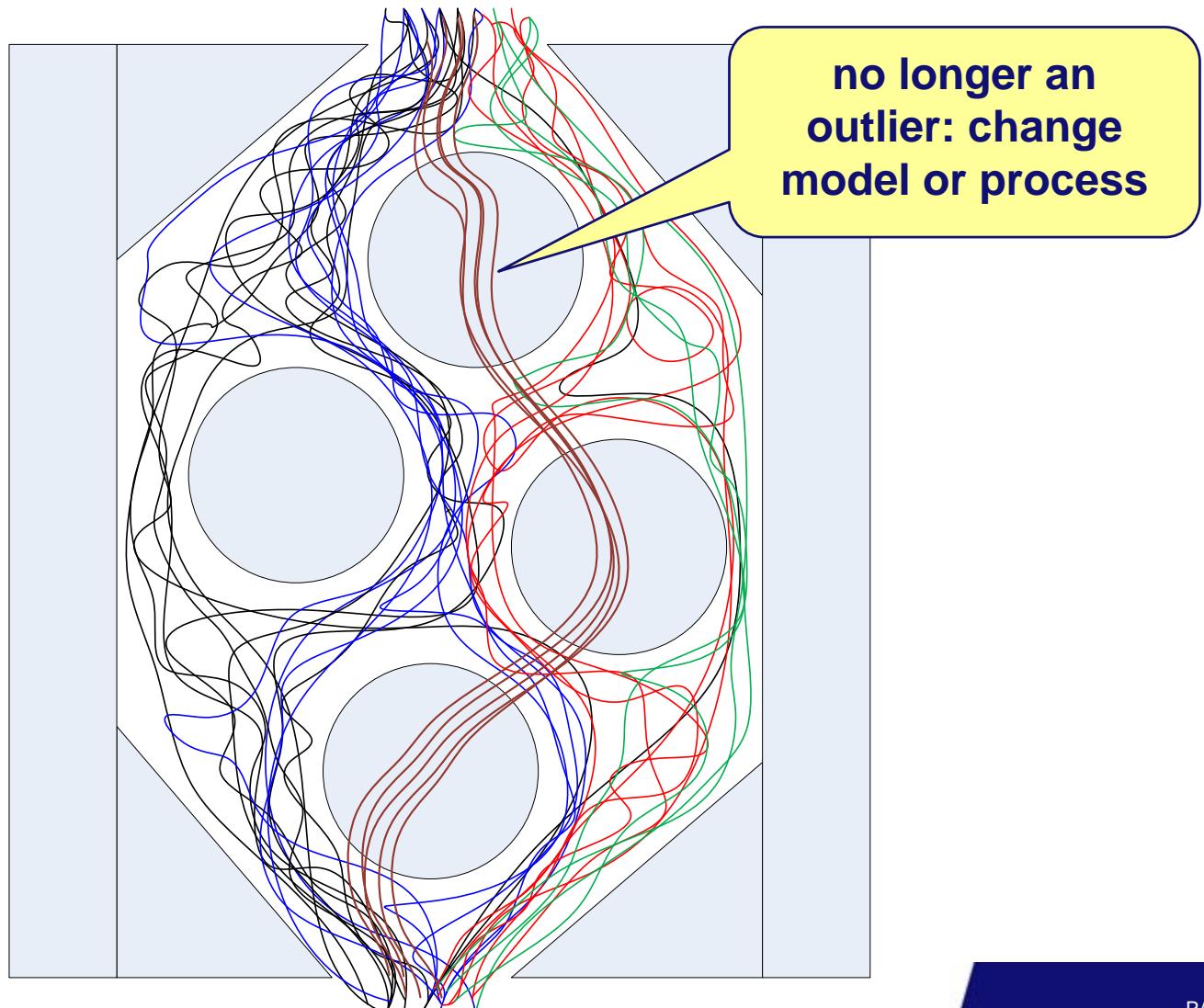
repair = pave the cow paths



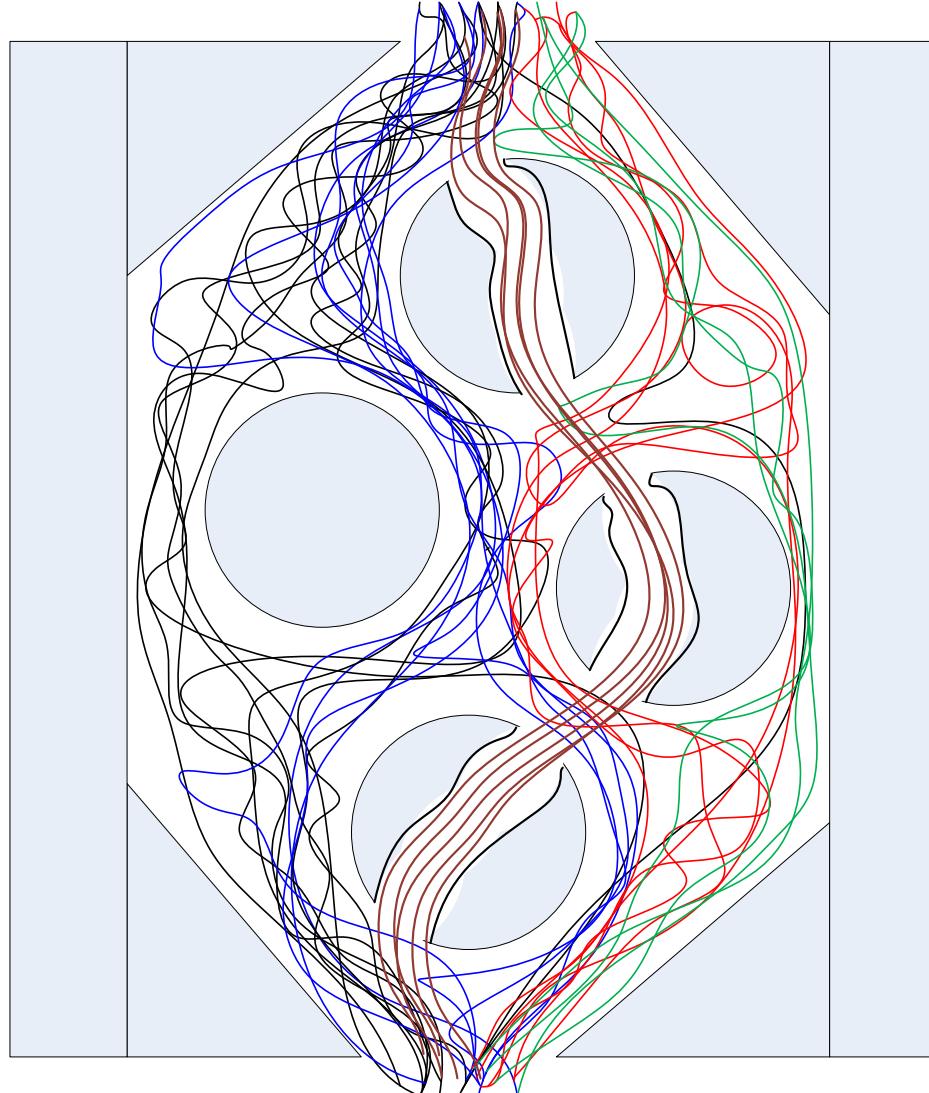
No need to repair



Repair model or influence reality ???



Repaired model

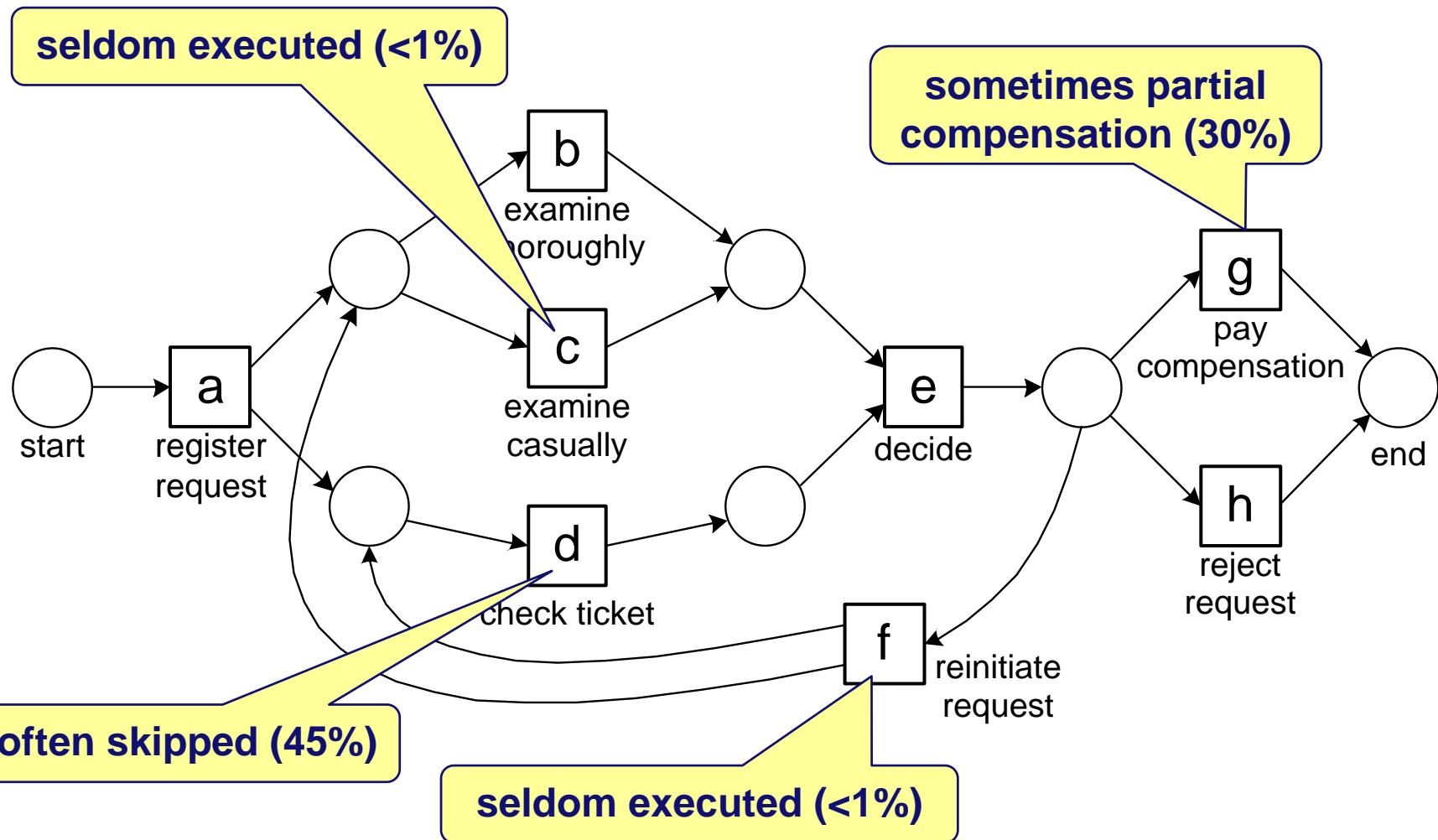


Metaphor: Improving passageways

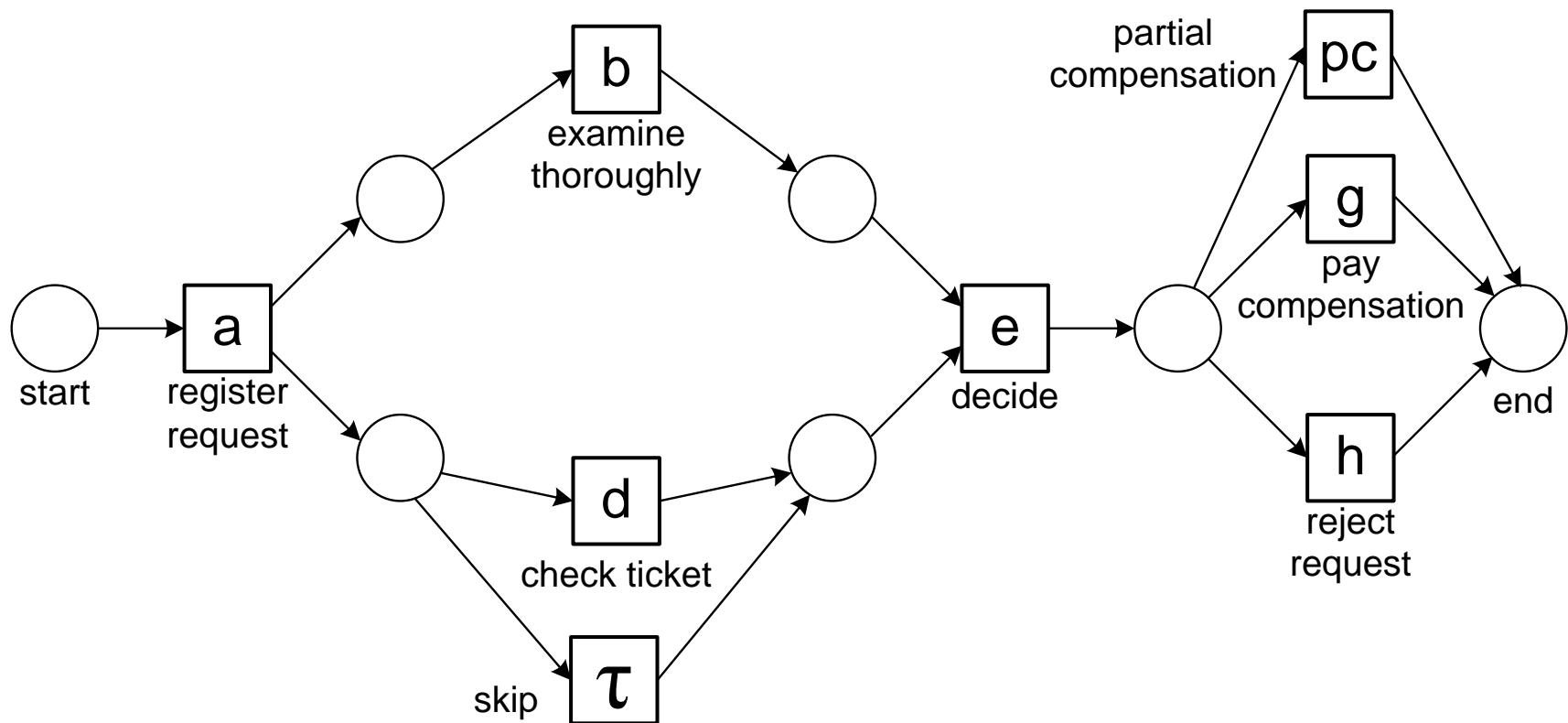
(what are the minimal changes in m² asphalt)



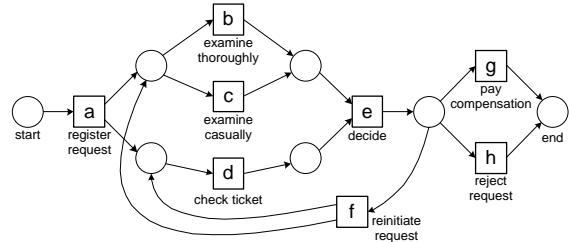
Assume the following replay results



Repaired model

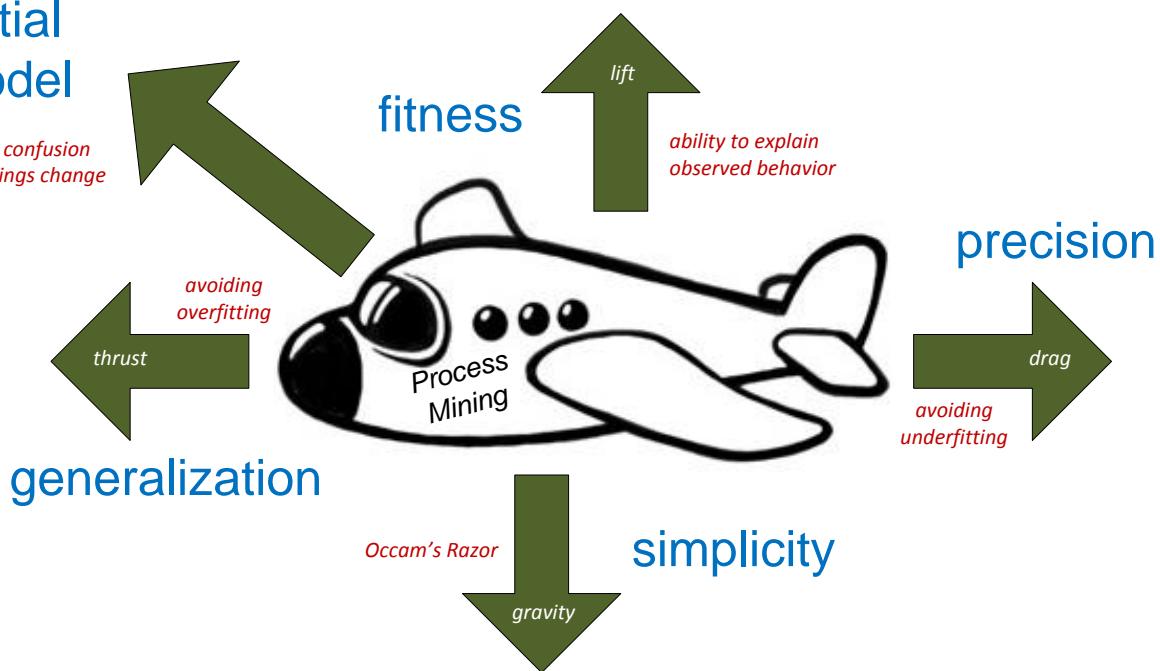


Original model is like a 5th force



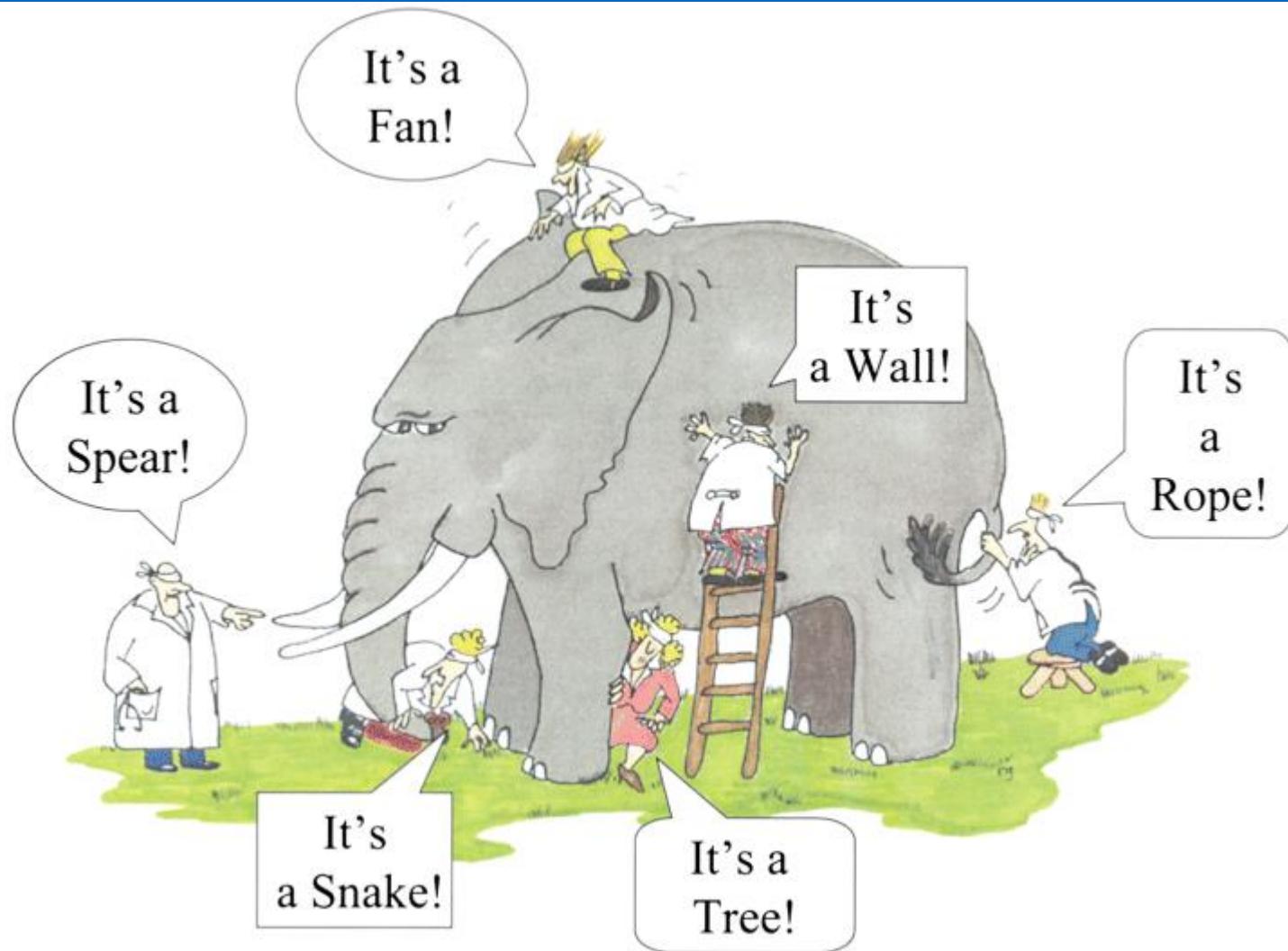
initial
model

*to avoid confusion
when things change*

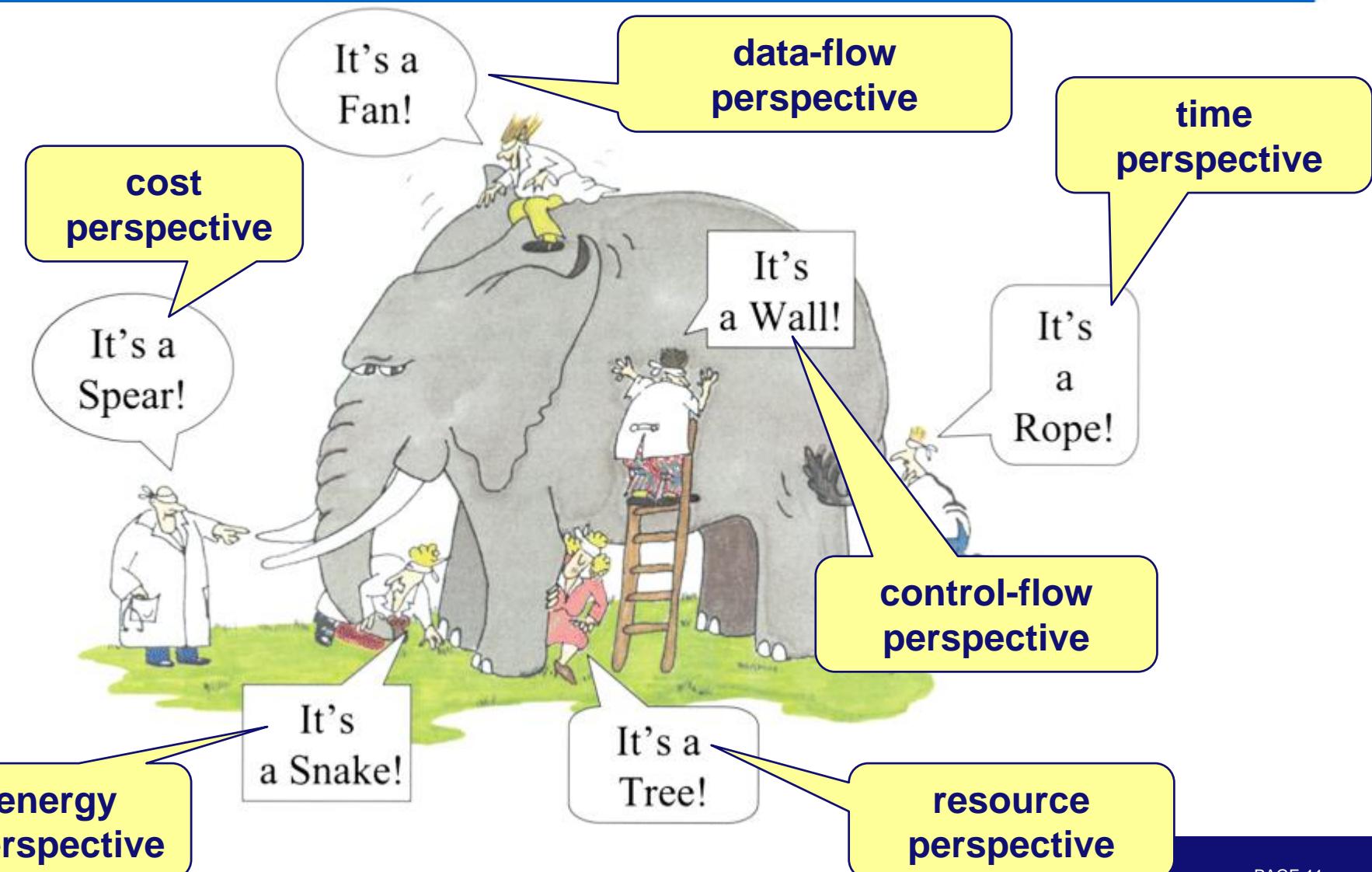


**extending process
models with additional
perspectives**

Process are not just about control-flow!



Process are not just about control-flow!



Attributes in event logs

case id	event id	properties				
		time	activity	trans	resource	cost
1	35654423	30-12-2010:11.02	register request	start	Pete	
	35654424	30-12-2010:11.08	register request	complete	Pete	50
	35654425	31-12-2010:10.06	examine thoroughly	start	Sue	
	35654427	31-12-2010:10.08	check ticket	start	Mike	
	35654428	31-12-2010:10.12	examine thoroughly	complete	Sue	400
	35654429	31-12-2010:10.20	check ticket	complete	Mike	100
	35654430	06-01-2011:11.18	decide	start	Sara	
	35654431	06-01-2011:11.22	decide	complete	Sara	200
	35654432	07-01-2011:14.24	reject request	start	Pete	
	35654433	07-01-2011:14.32	reject request	complete	Pete	200
2	35654483	30-12-2010:11.32	register request	start	Mike	
	35654484	30-12-2010:11.40	register request	complete	Mike	50
	35654485	30-12-2010:12.12	check ticket	start	Mike	
	35654486	30-12-2010:12.24	check ticket	complete	Mike	100
	35654487	30-12-2010:14.16	examine casually	start	Pete	
	35654488	30-12-2010:14.22	examine casually	complete	Pete	400
	35654489	05-01-2011:11.22	decide	start	Sara	
	35654490	05-01-2011:11.29	decide	complete	Sara	200
	35654491	08-01-2011:12.05	pay compensation	start	Ellen	
	35654492	08-01-2011:12.15	pay compensation	complete	Ellen	200
...						

As discussed before:

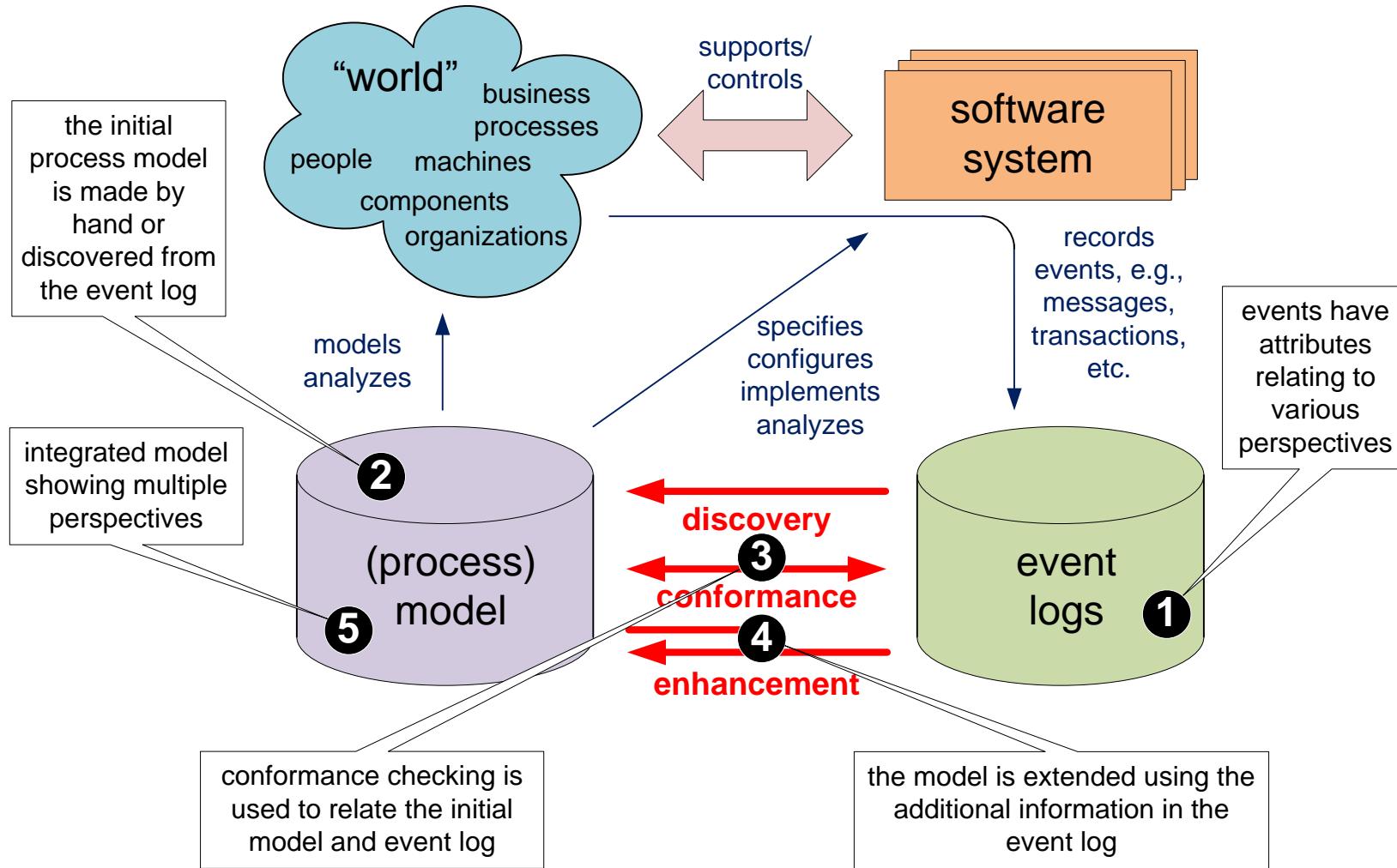
- A process consists of cases.
- A case consists of events such that each event relates to precisely one case.
- Events within a case are ordered.
- Events can have attributes.
- Examples of typical attribute names are activity, time, costs, and resource.



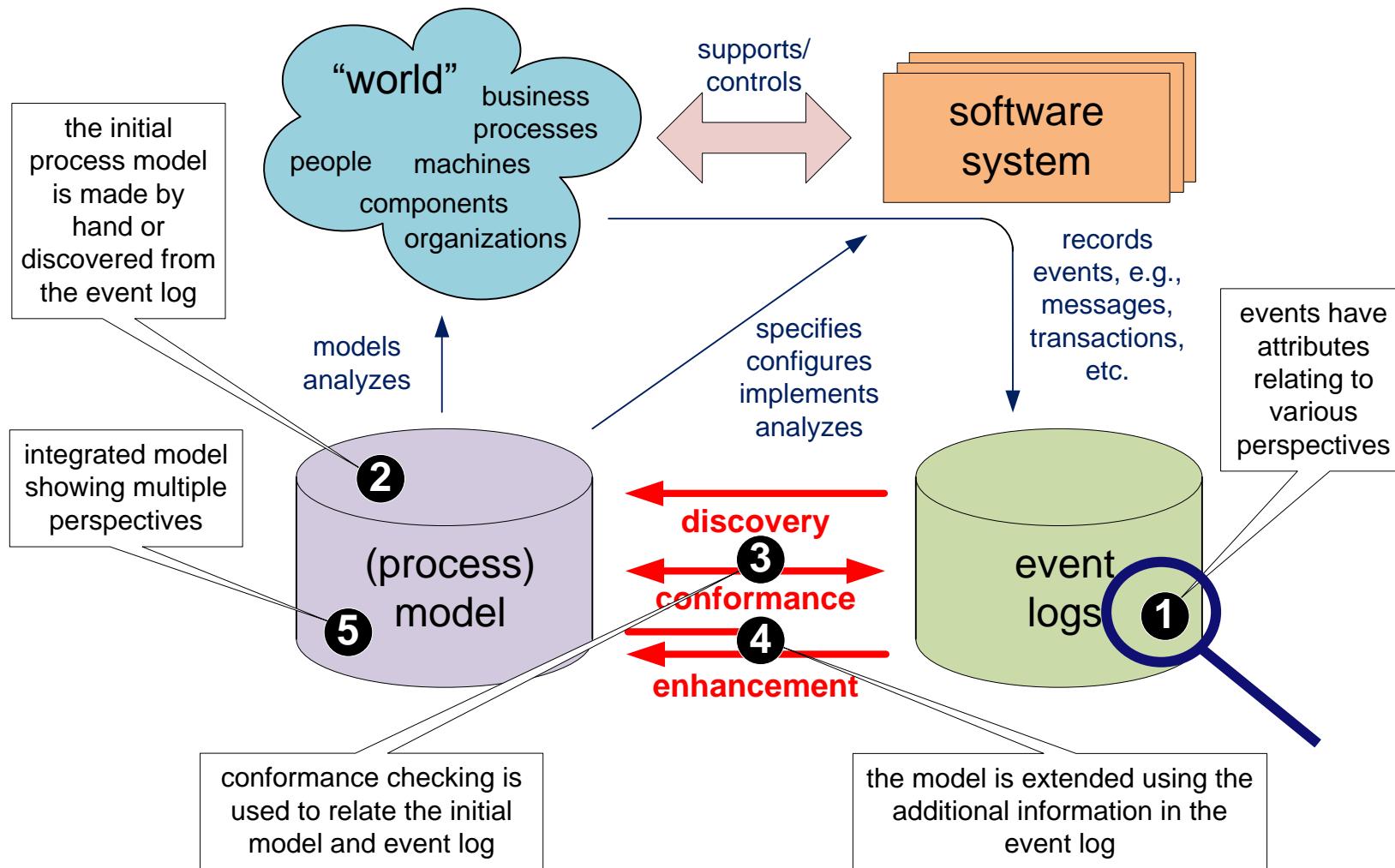
Cases may also have attributes

case id	custid	name	type	region	amount
1	9911	Smith	gold	south	989.50
2	9915	Jones	silver	west	546.00
3	9912	Anderson	silver	north	763.20
4	9904	Thompson	silver	west	911.70
5	9911	Smith	gold	south	812.10
6	9944	Baker	silver	east	788.00
7	9944	Baker	silver	east	792.80
8	9911	Smith	gold	south	544.70
...

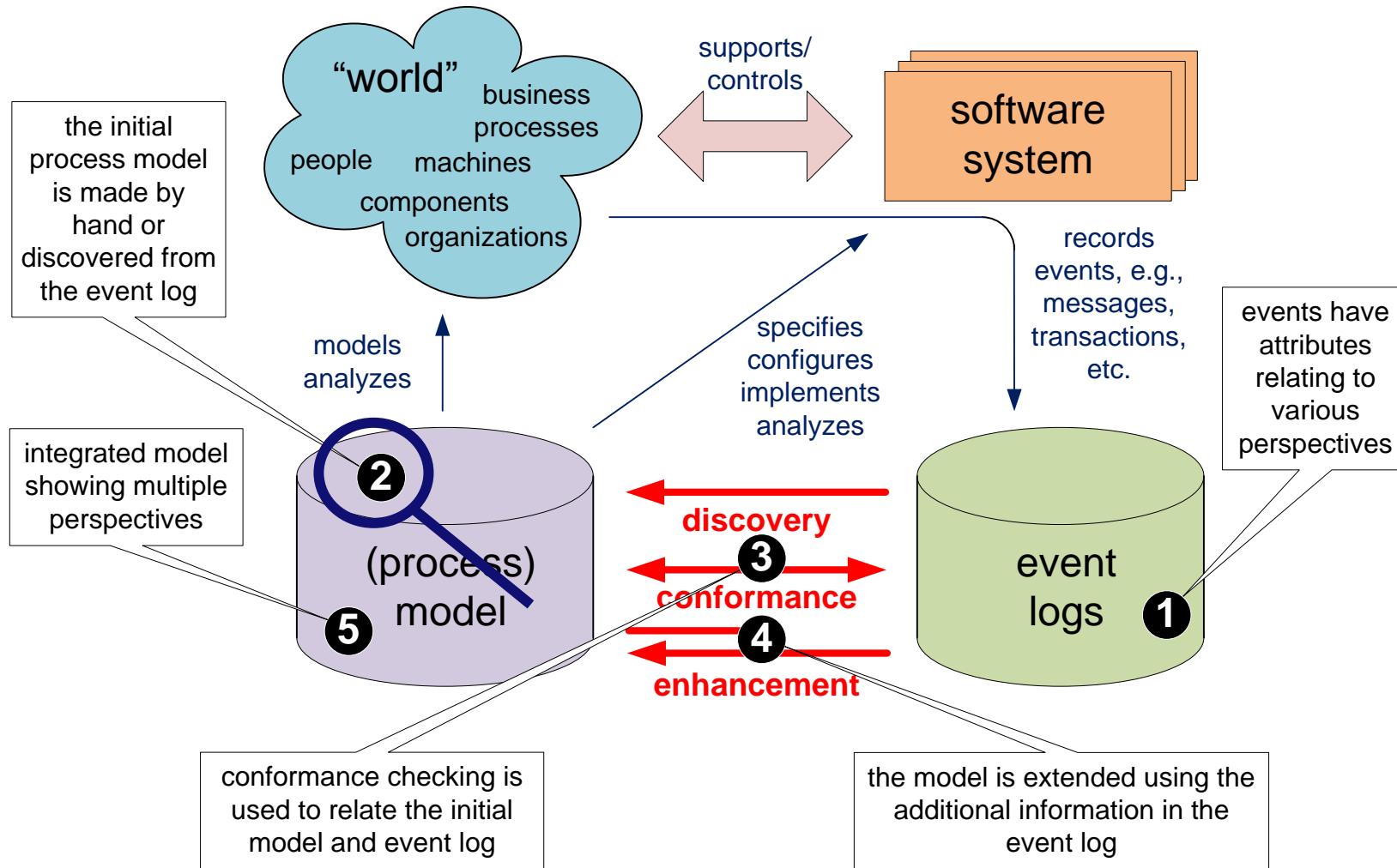
Extending process models using event data



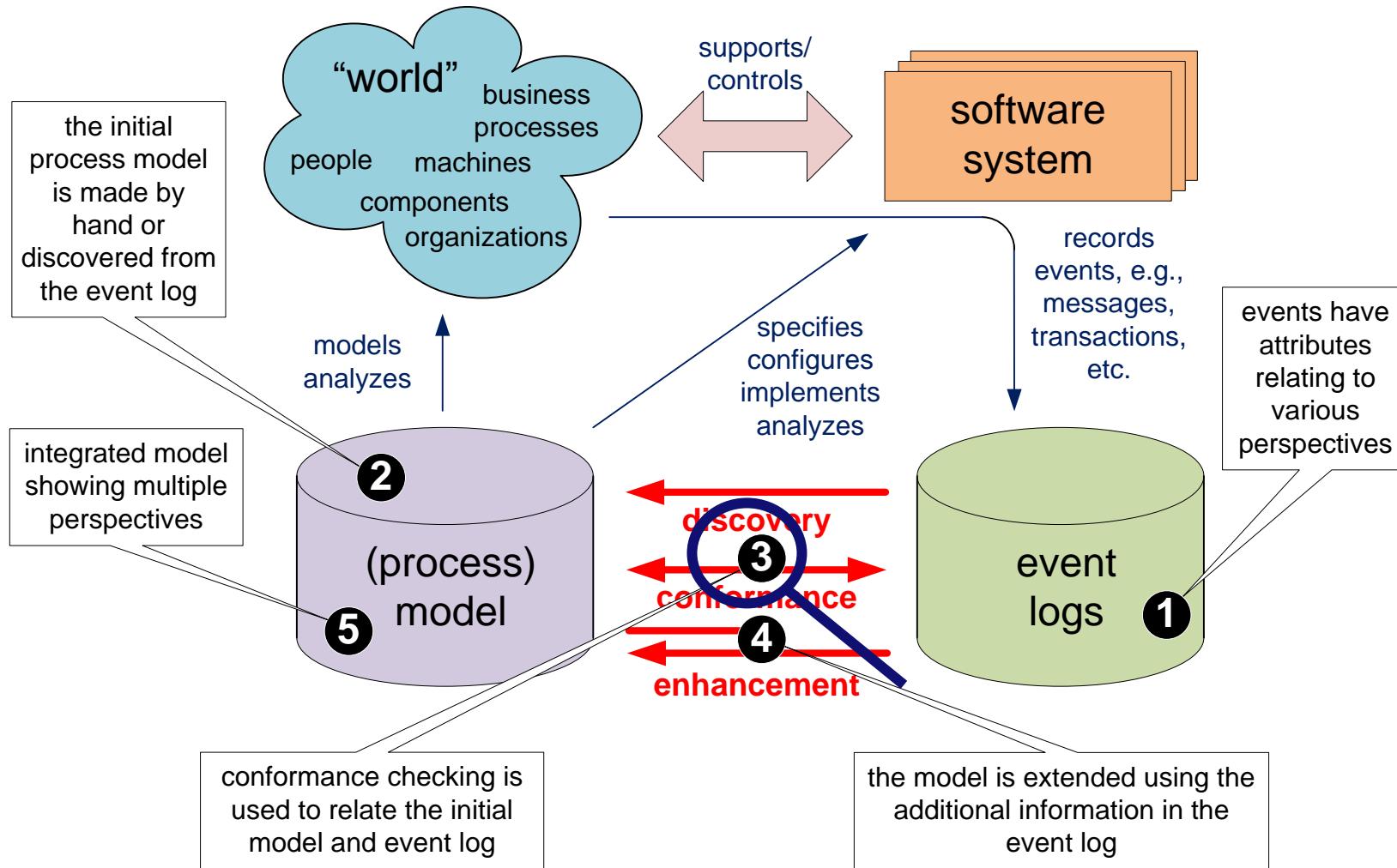
Extending process models using event data



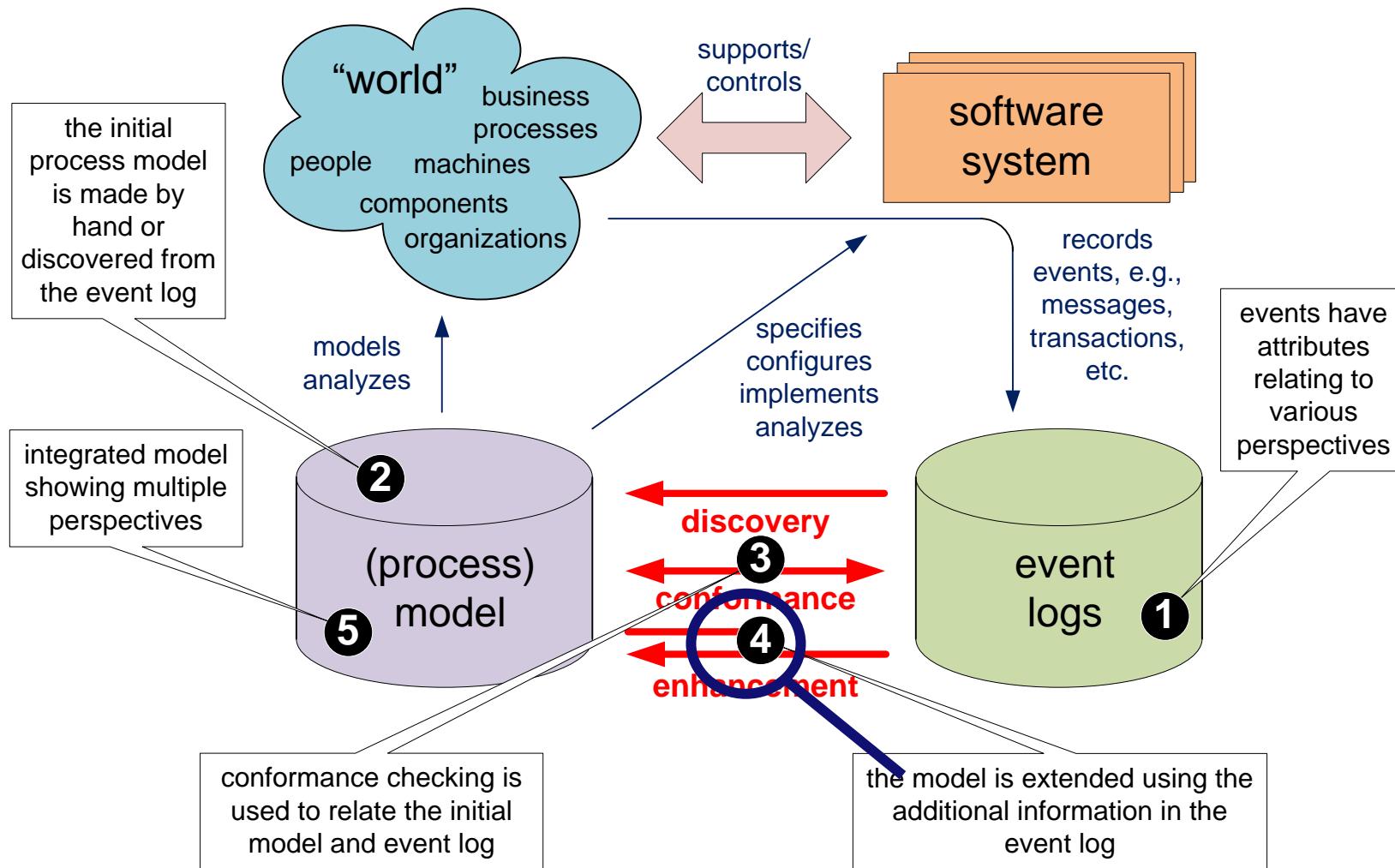
Extending process models using event data



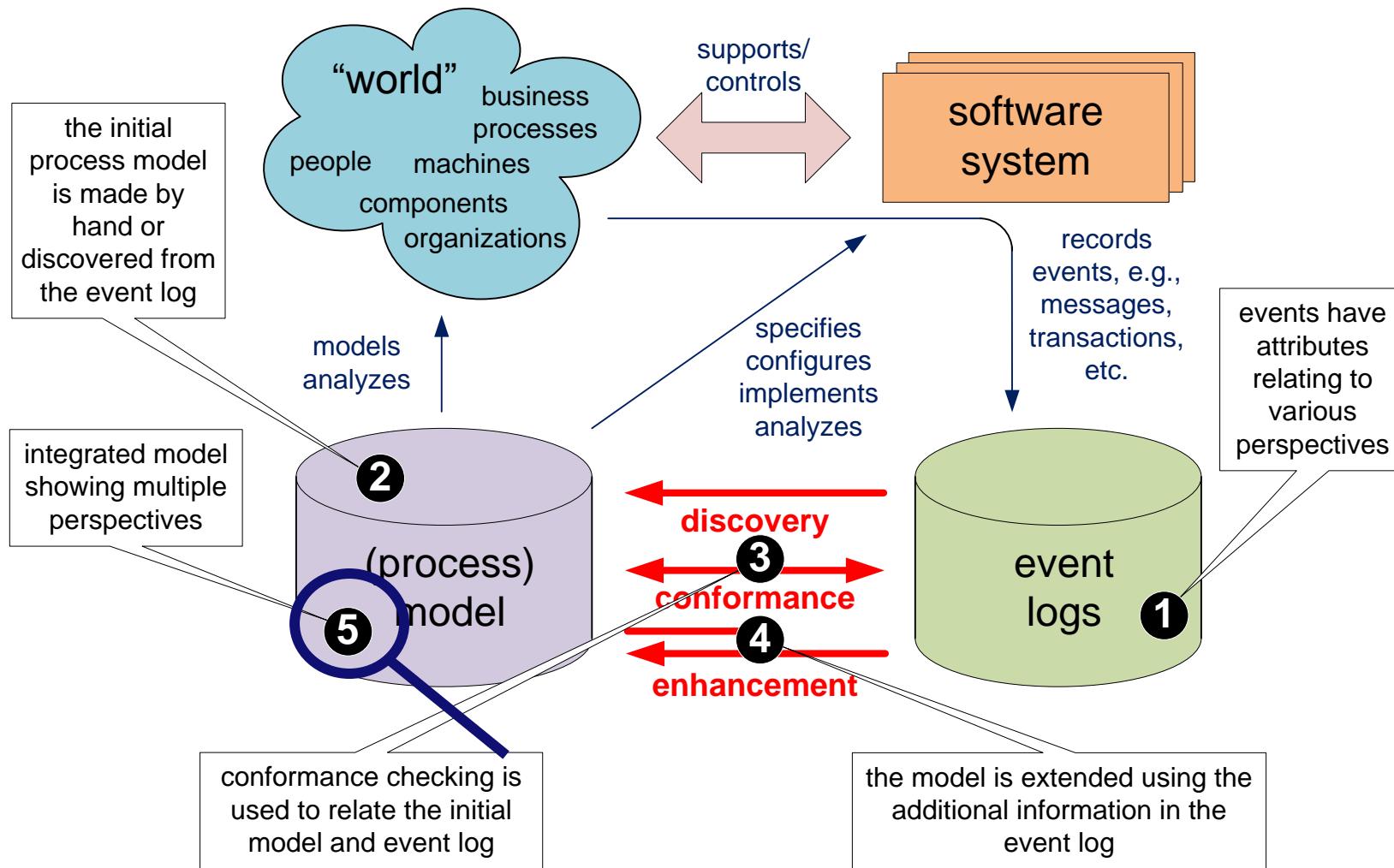
Extending process models using event data



Extending process models using event data



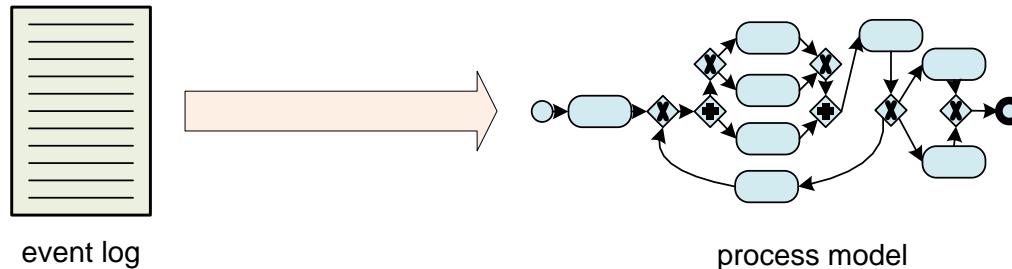
Extending process models using event data



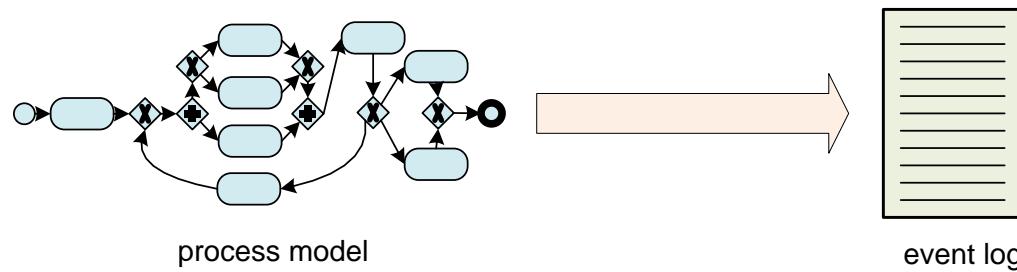
replay revisited

Connecting events to model elements is essential for model extension

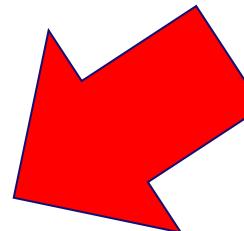
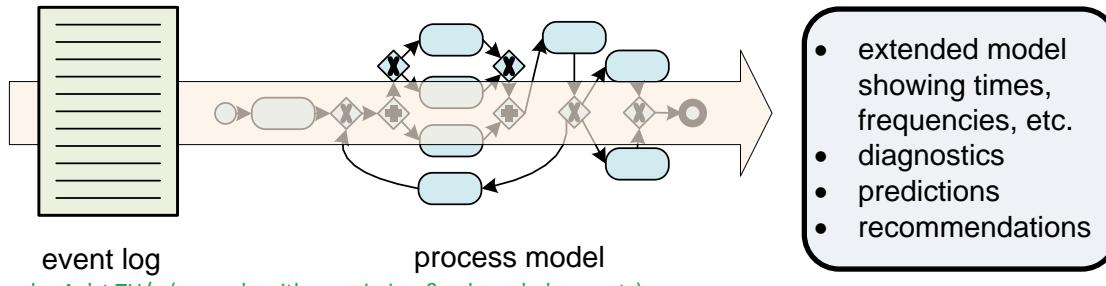
Play-In



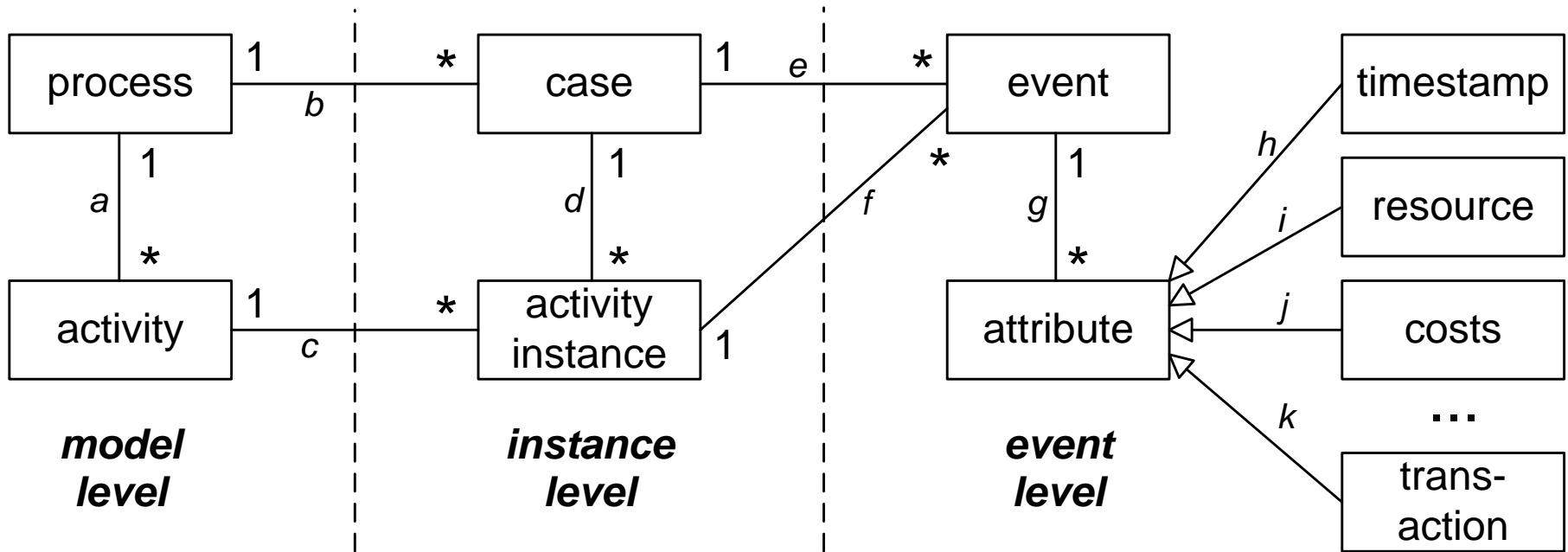
Play-Out



Replay



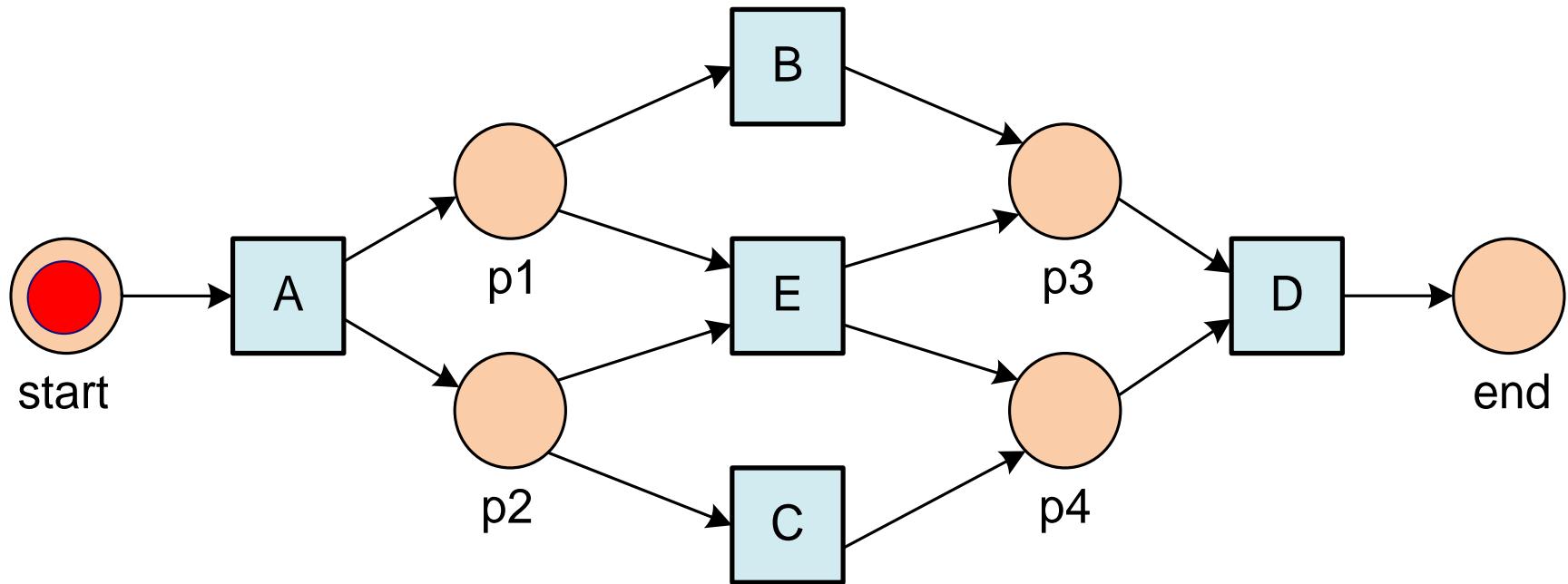
Connected event log and model through replay or alignments



trace in the event log is related to a path in the model

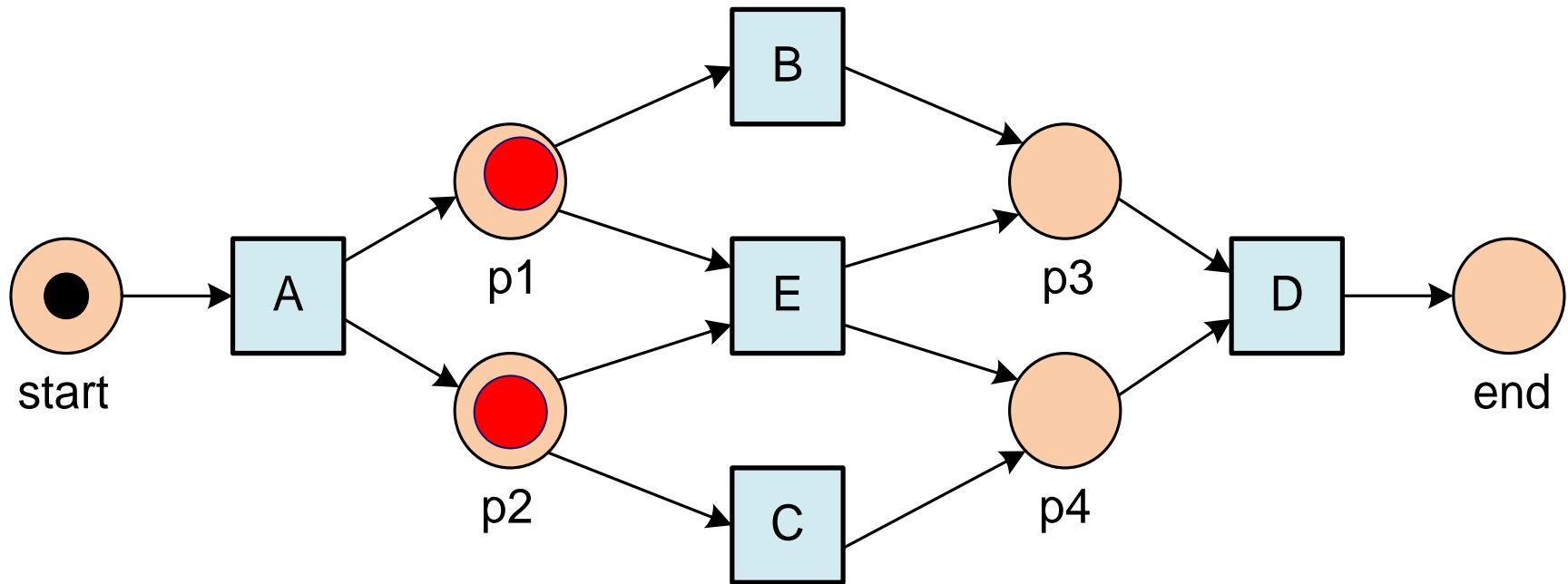
Remember: Replay!

A B C D



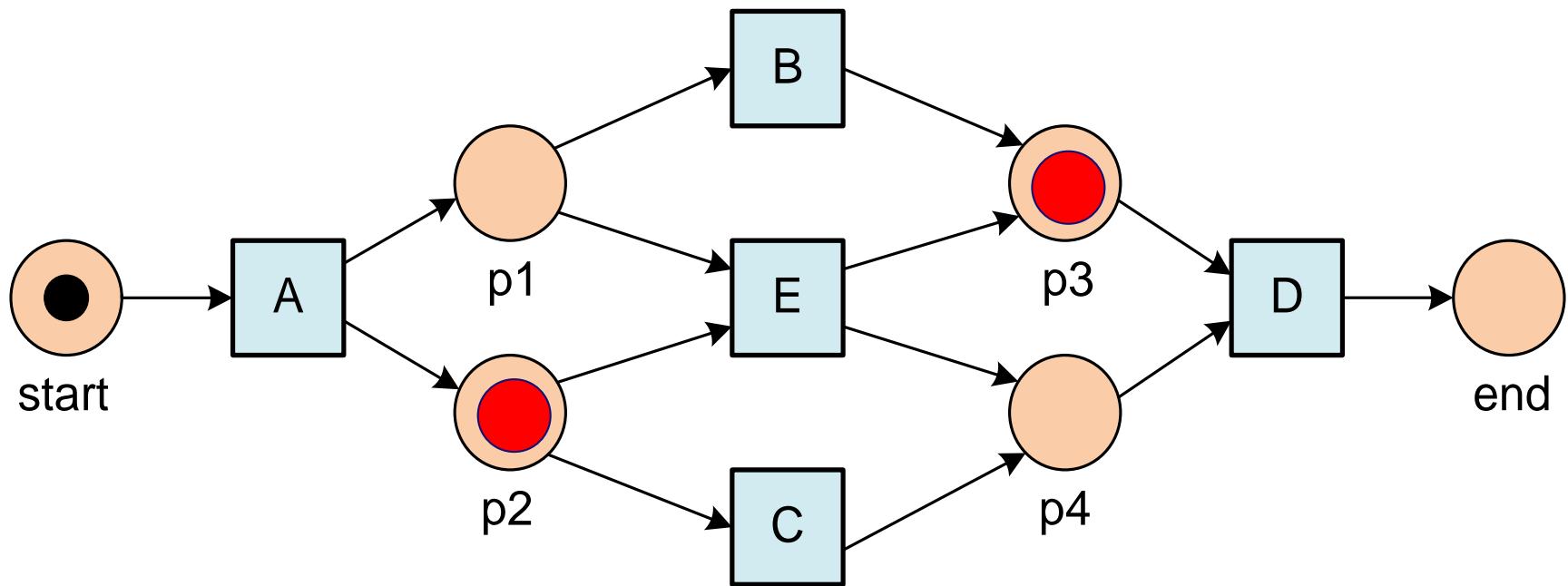
Remember: Replay!

B C D

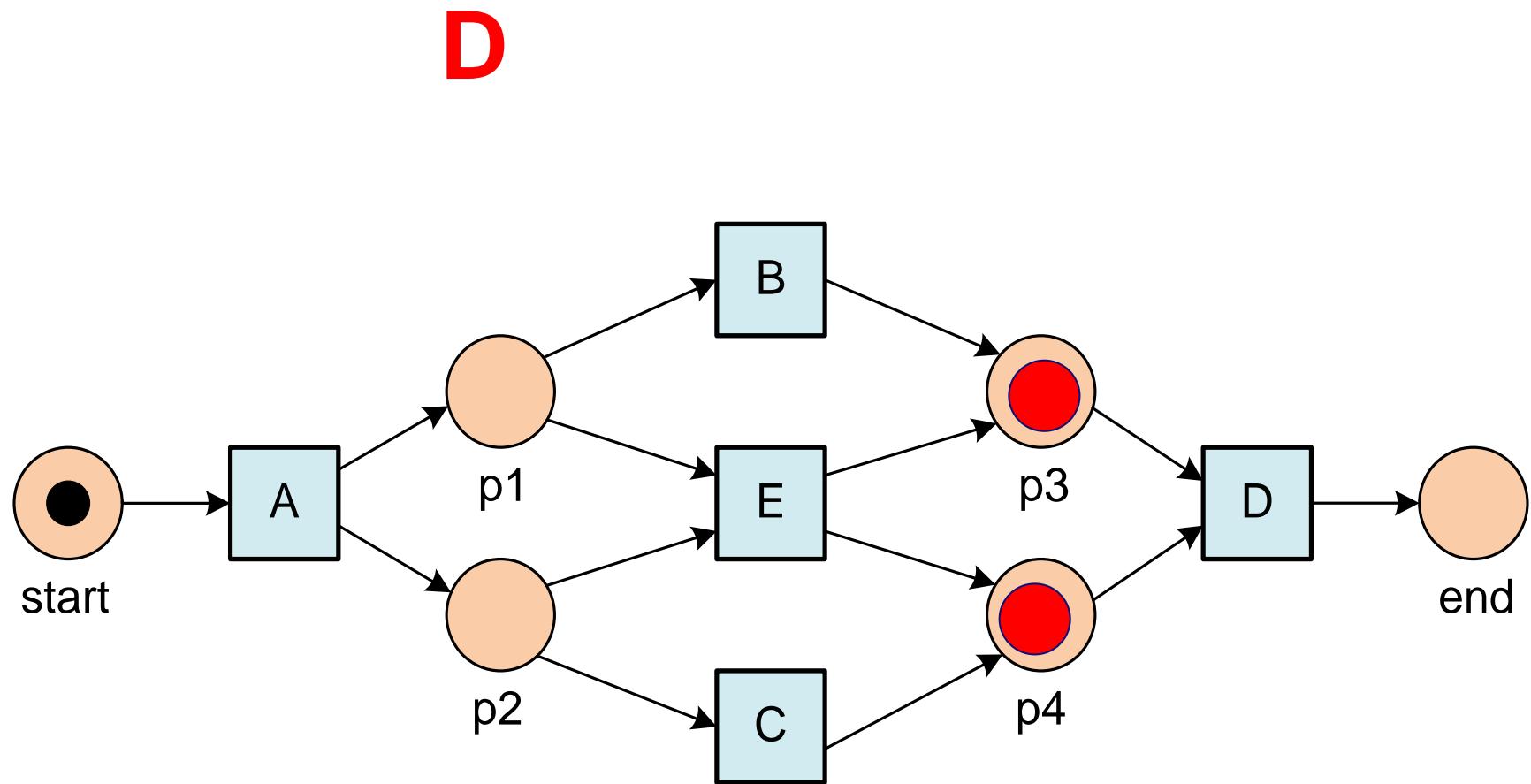


Remember: Replay!

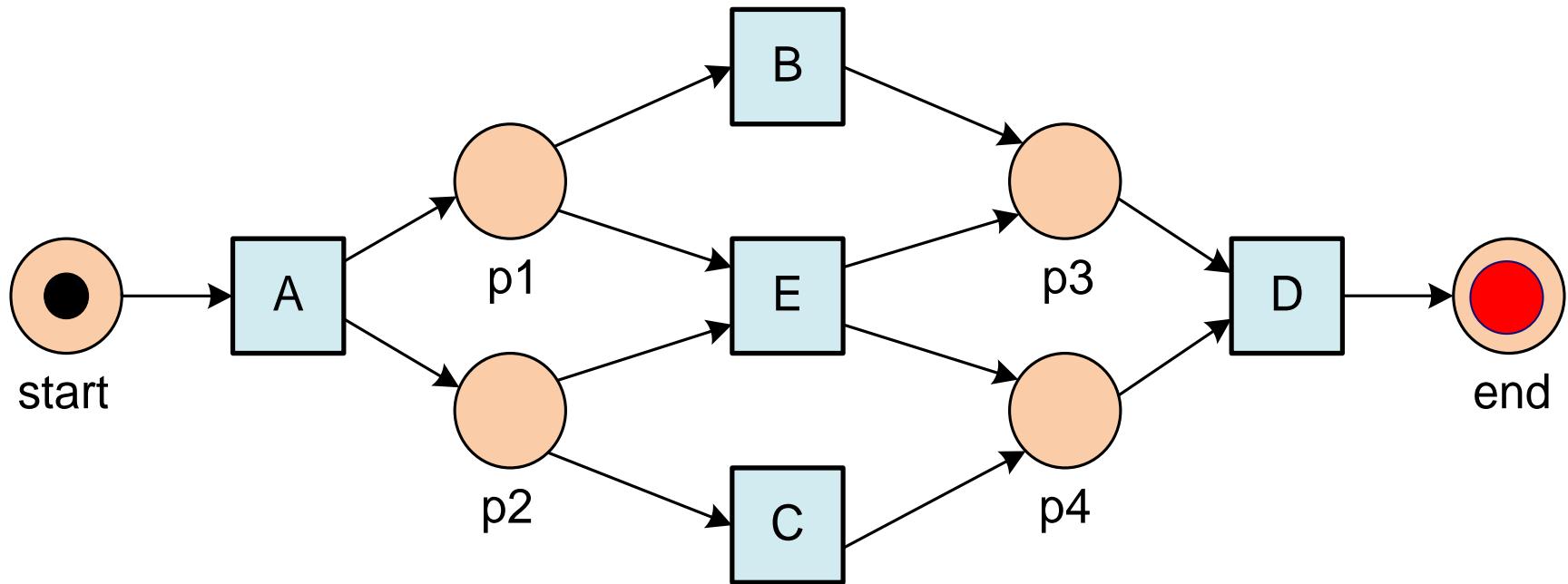
C D



Remember: Replay!

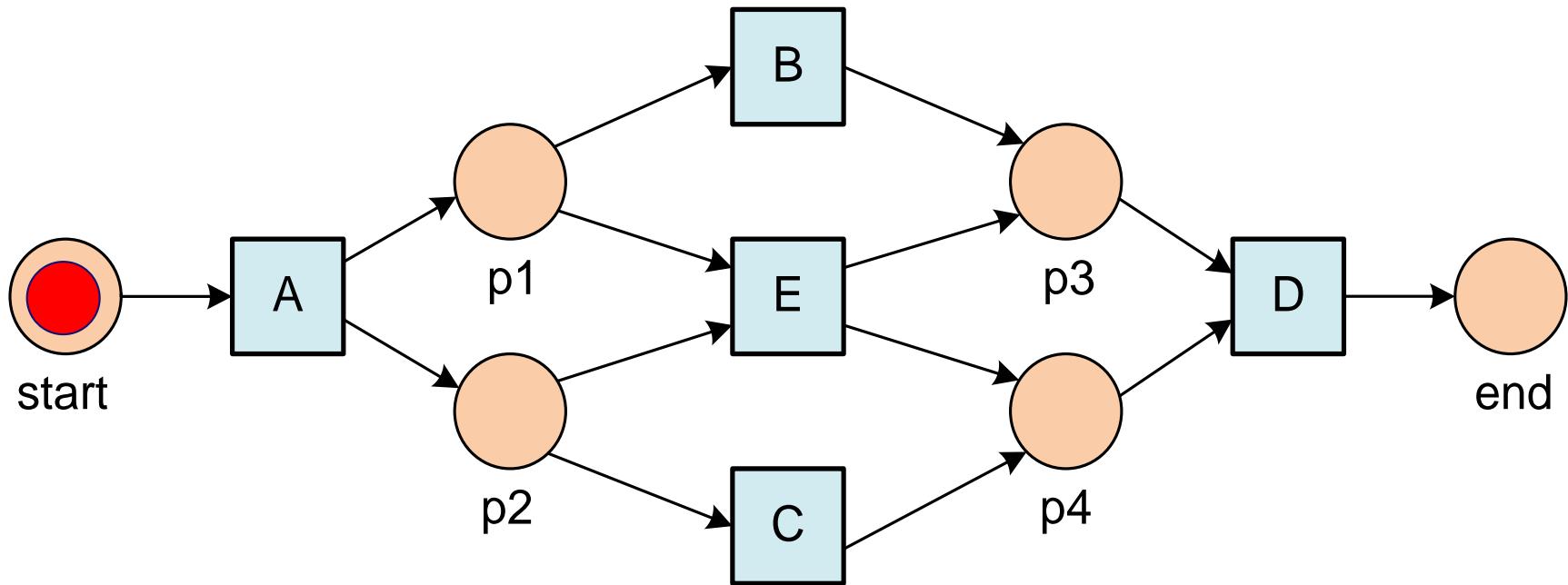


Remember: Replay!



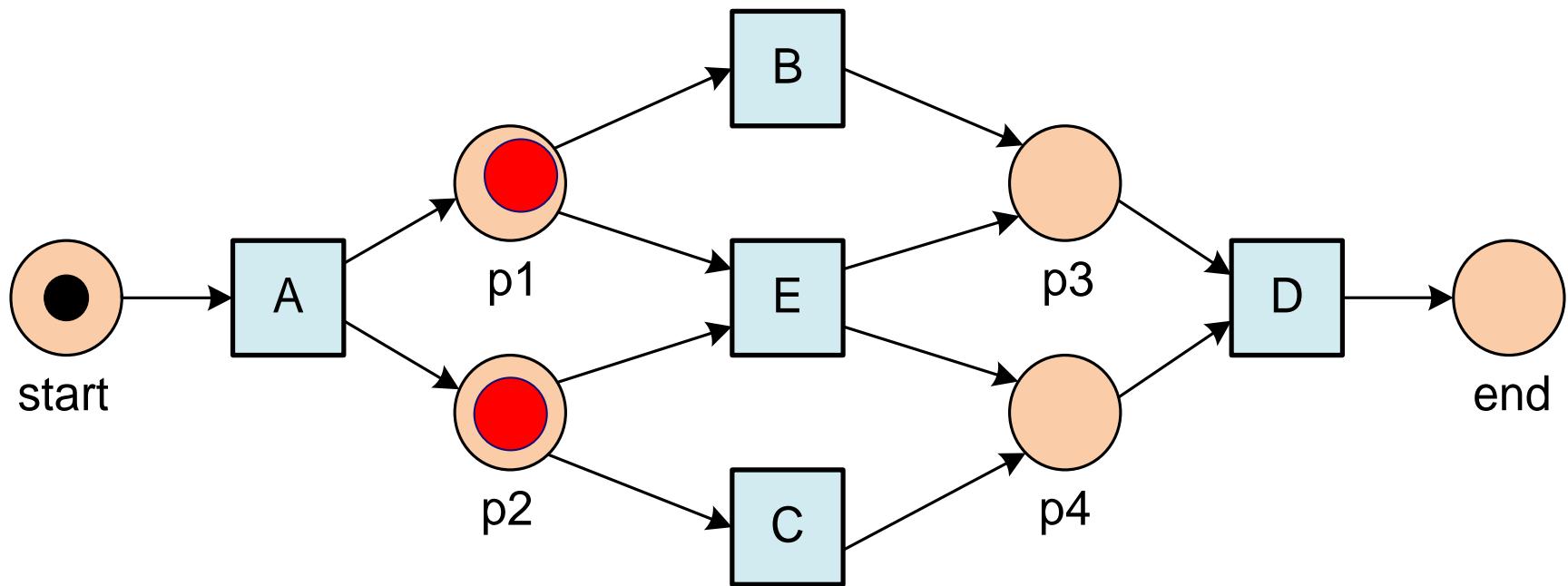
Replay can detect problems

A C D

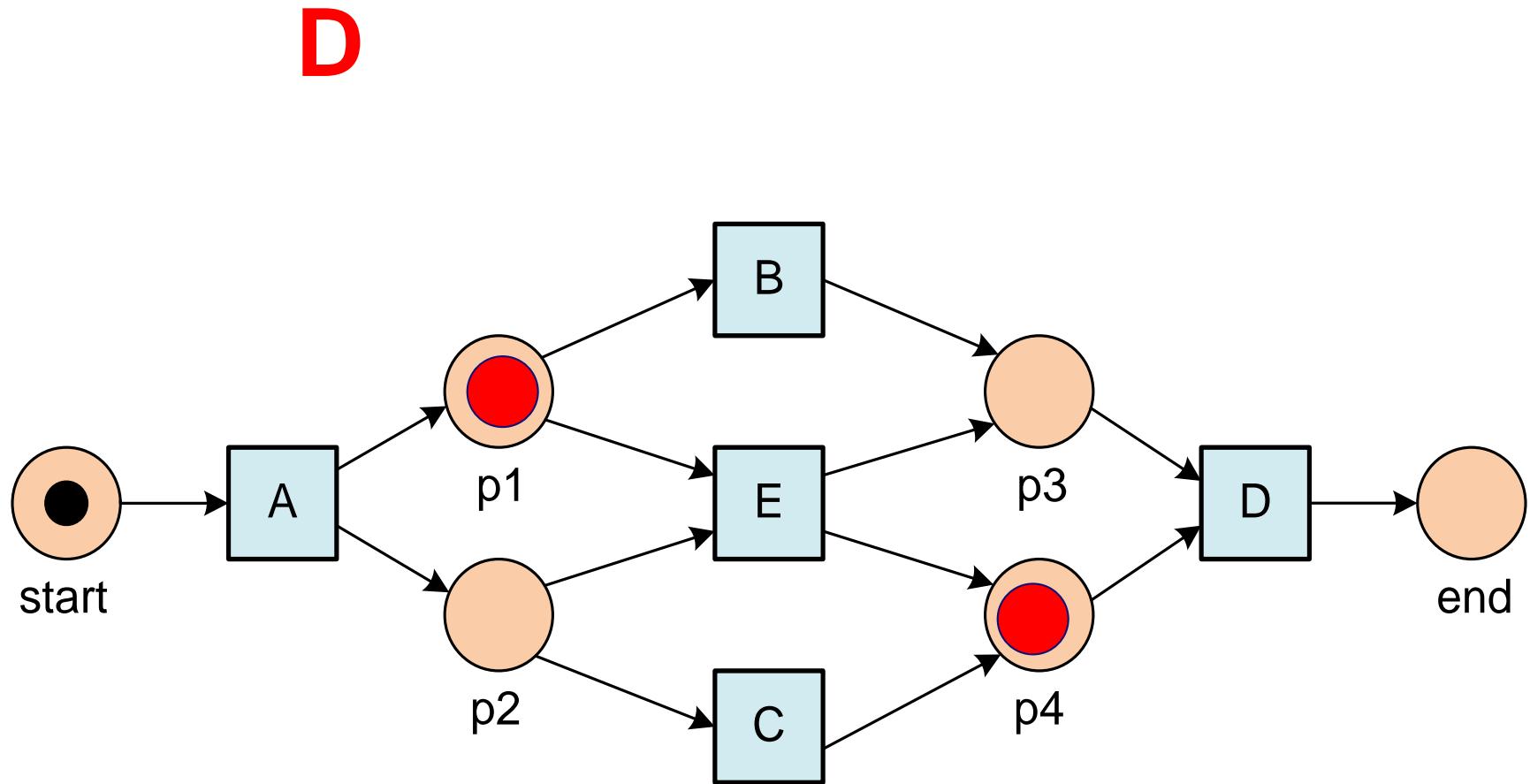


Replay can detect problems

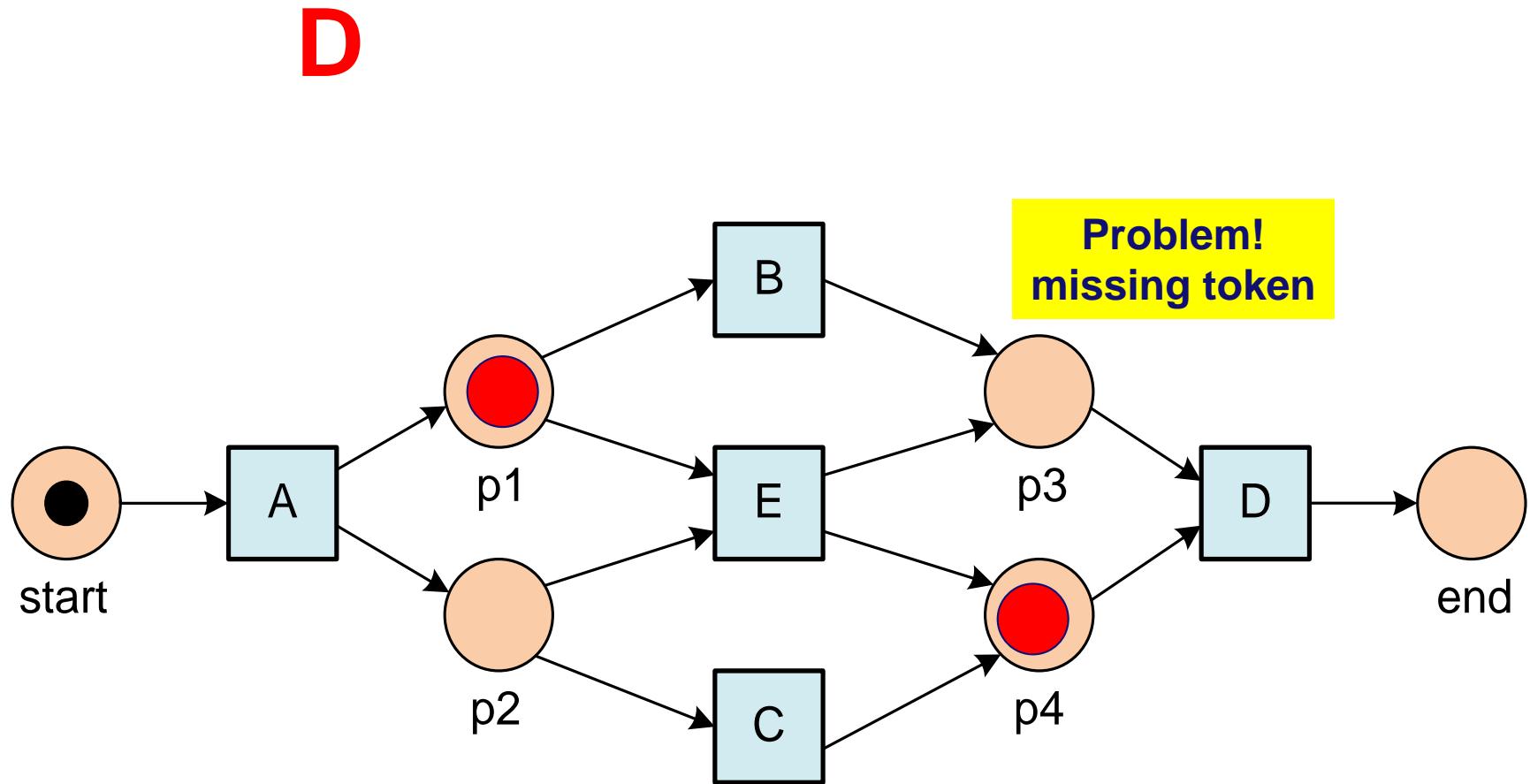
C D



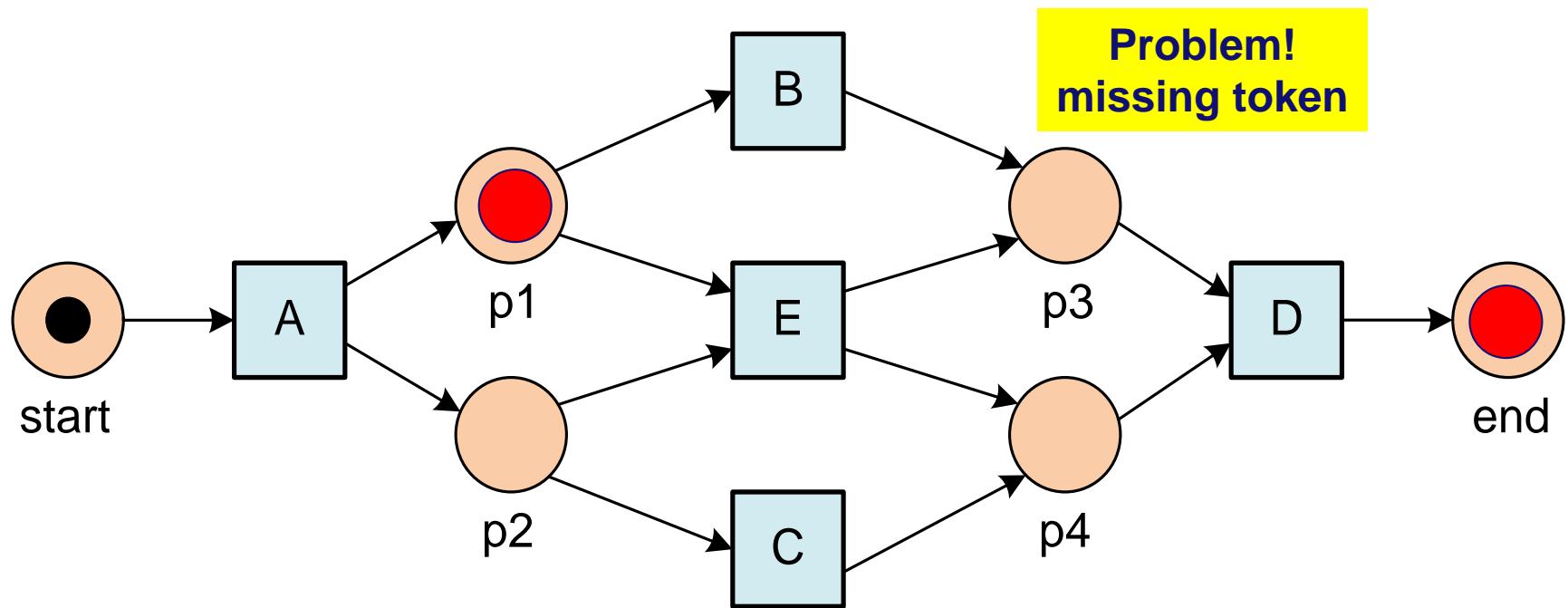
Replay can detect problems



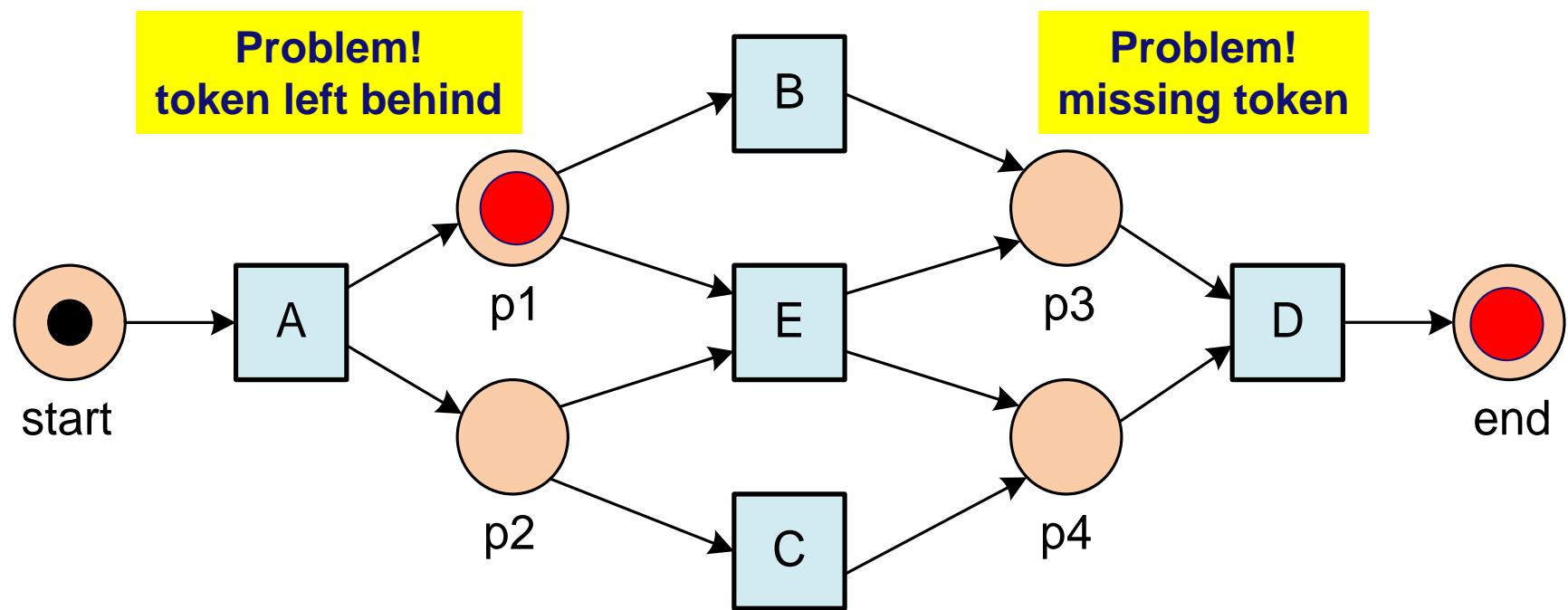
Replay can detect problems



Replay can detect problems

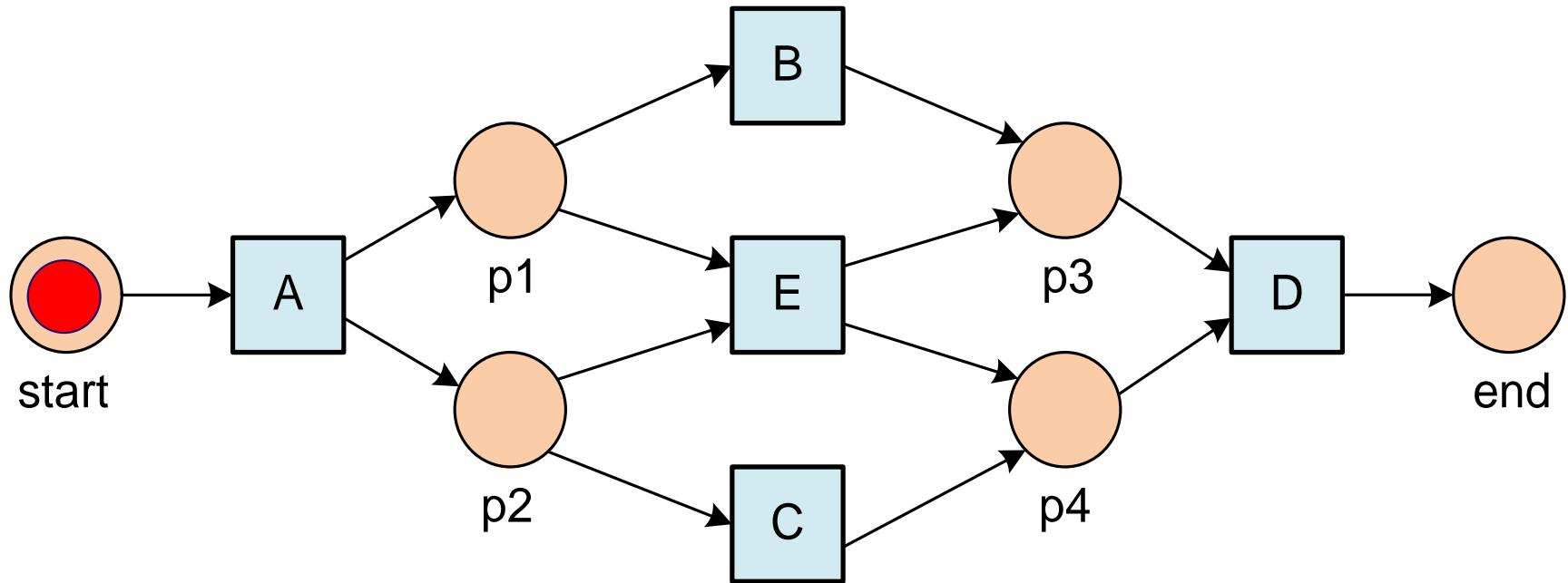


Replay can detect problems



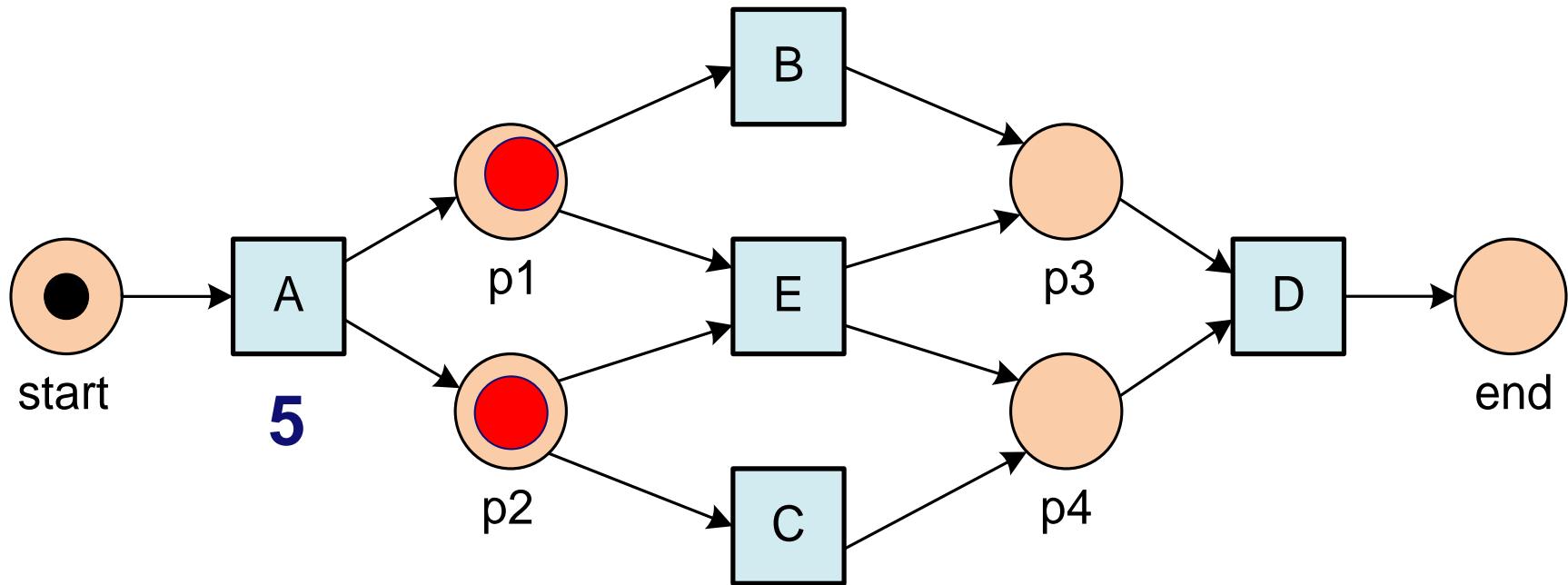
Replay can extract timing information

A⁵B⁸C⁹D¹³



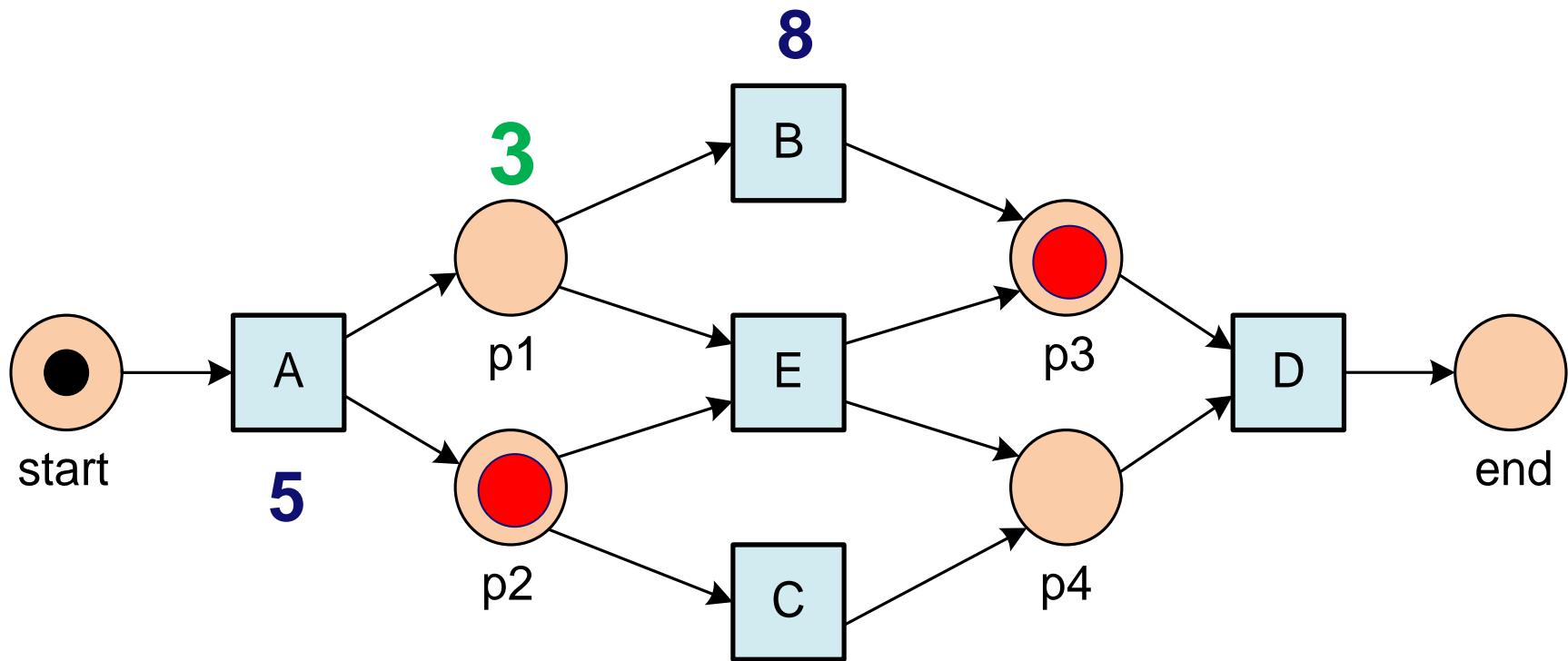
Replay can extract timing information

B⁸ C⁹ D¹³

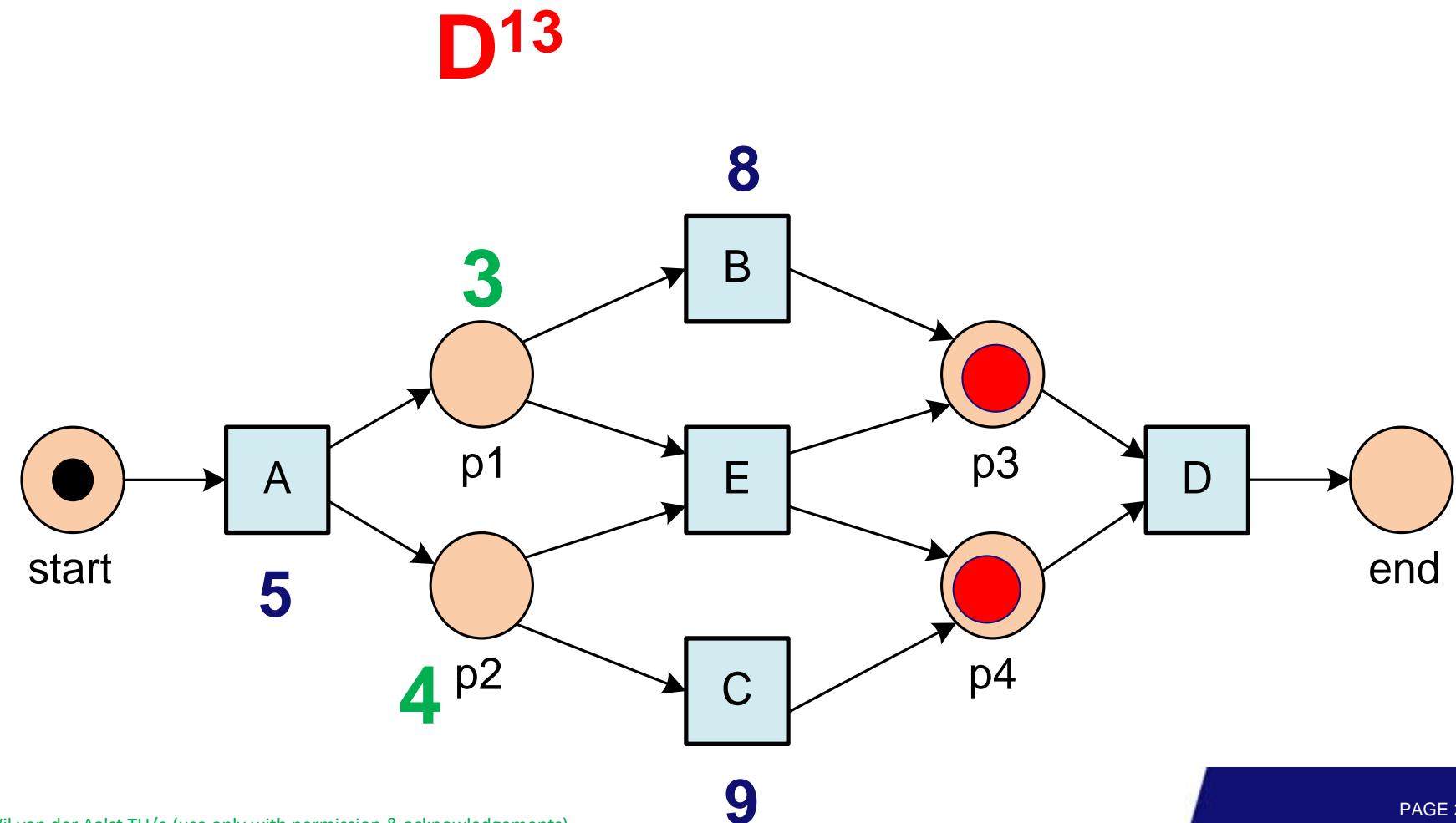


Replay can extract timing information

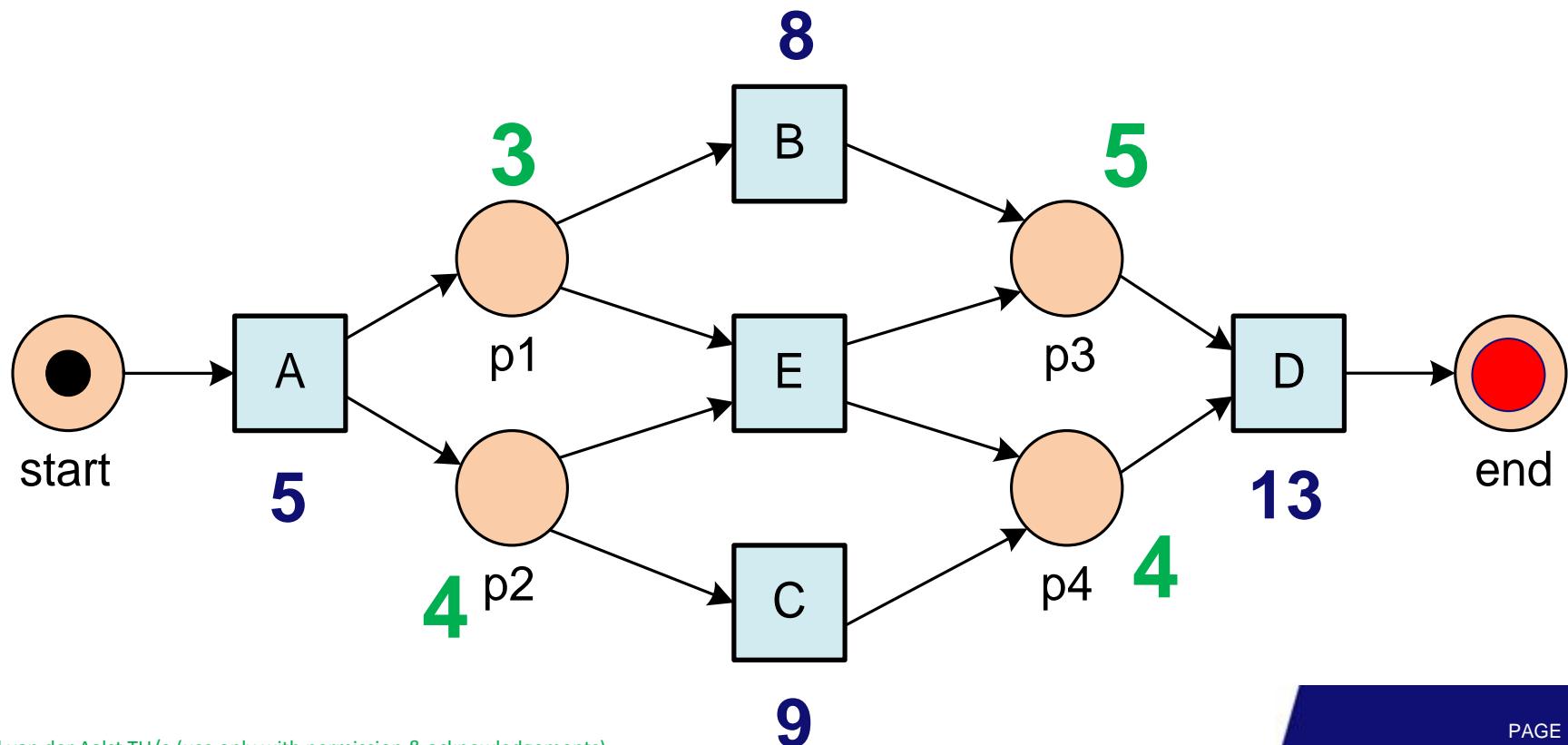
C⁹ D¹³



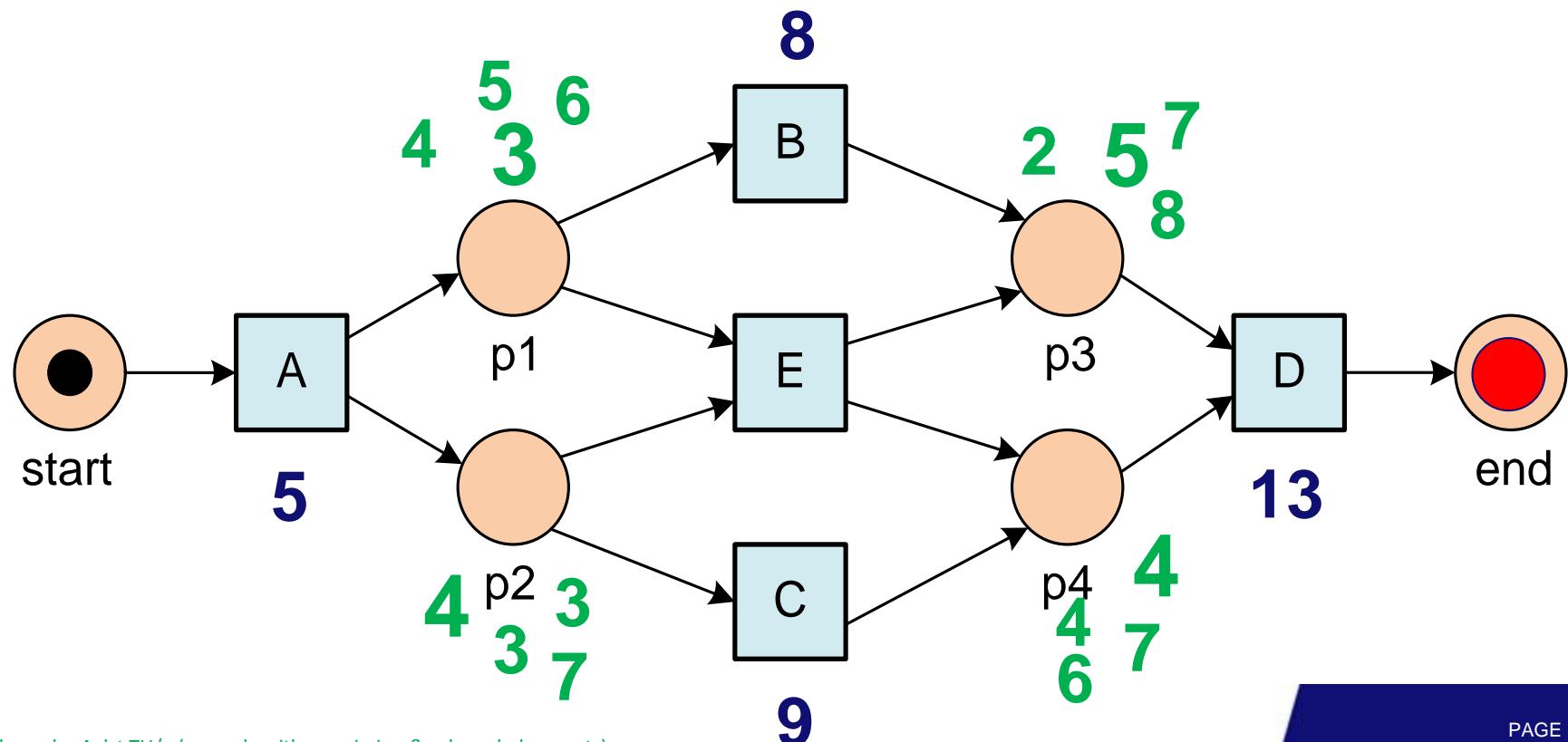
Replay can extract timing information



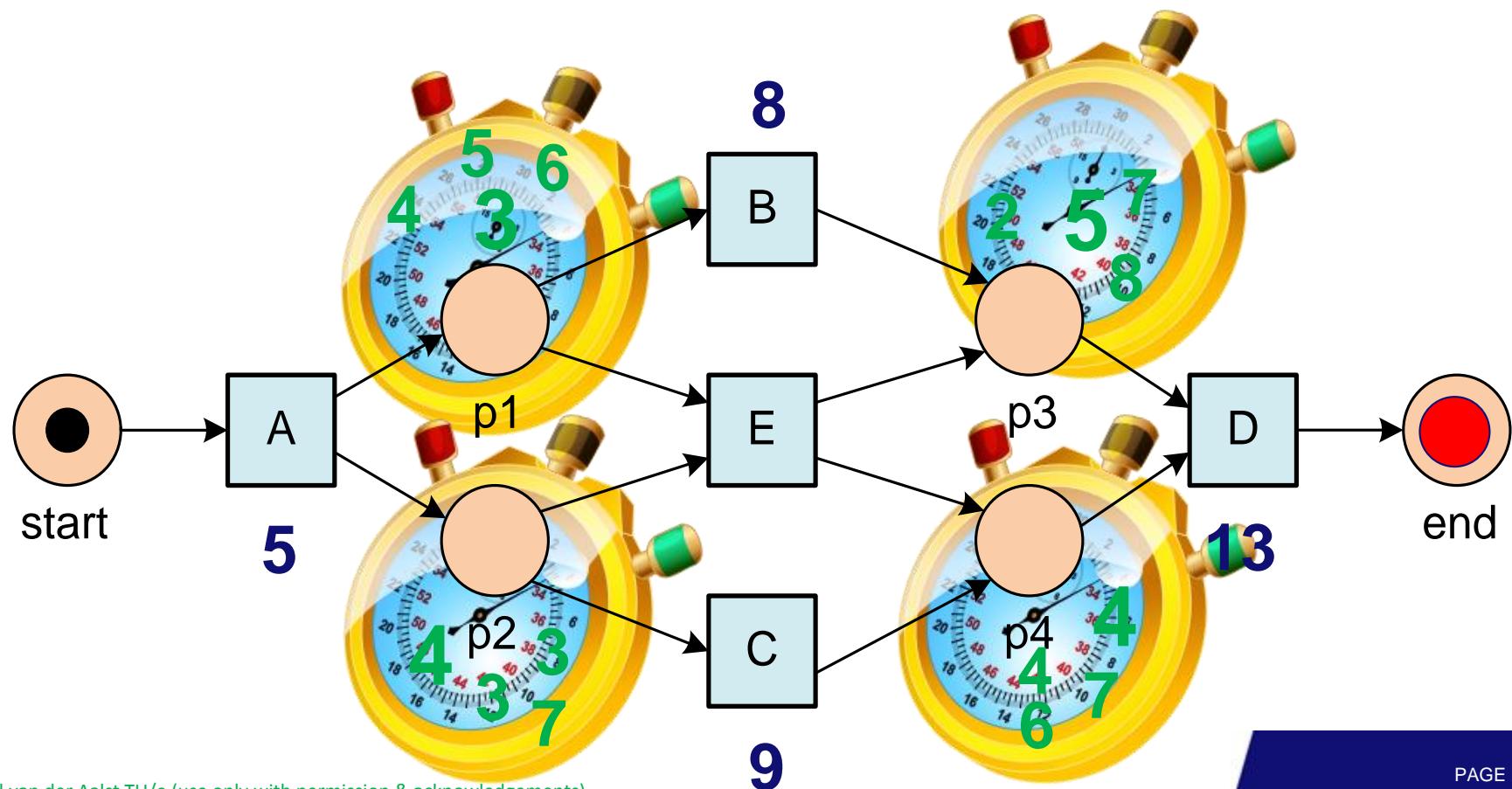
Replay can extract timing information



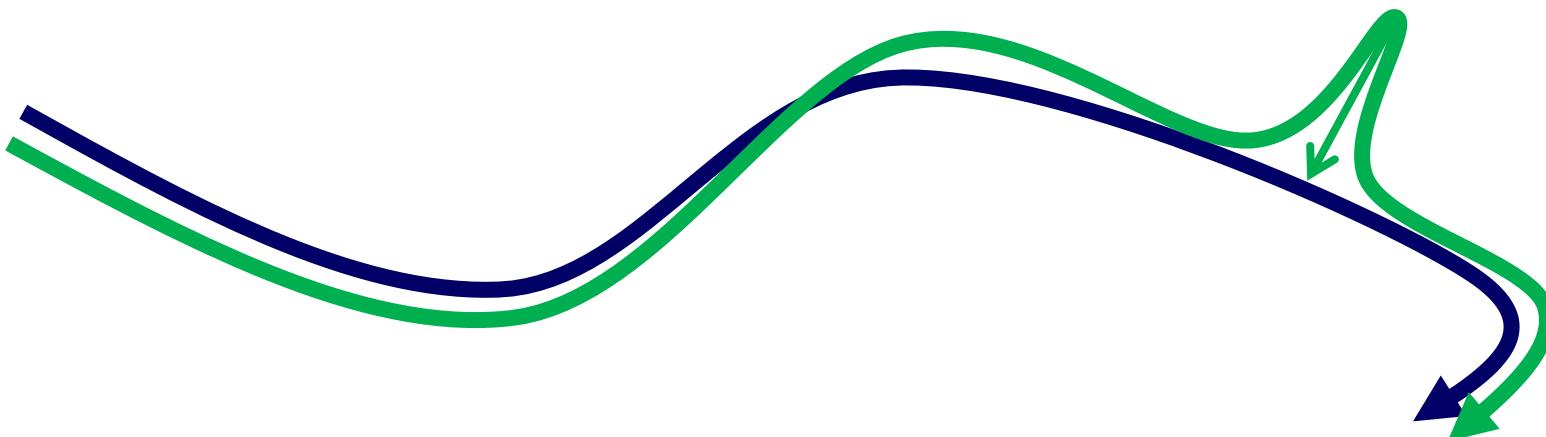
Replay can extract timing information



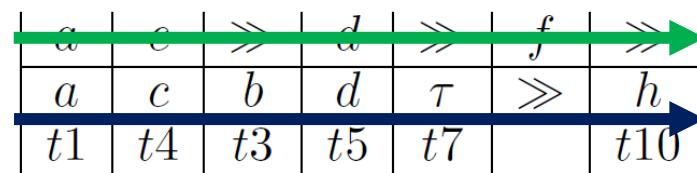
Replay can extract timing information



When there are deviation we need to squeeze reality into the model



- Traces may not perfectly fit the model.
- Often we cannot throw non-fitting traces (loosing most of the data and/or introducing a hidden bias).
- Conformance checking techniques help us to map traces onto the "nearest" path in the model.



In the remainder we assume that through prepressing model and log are aligned



In the remainder we assume that through prepressing model and log are aligned

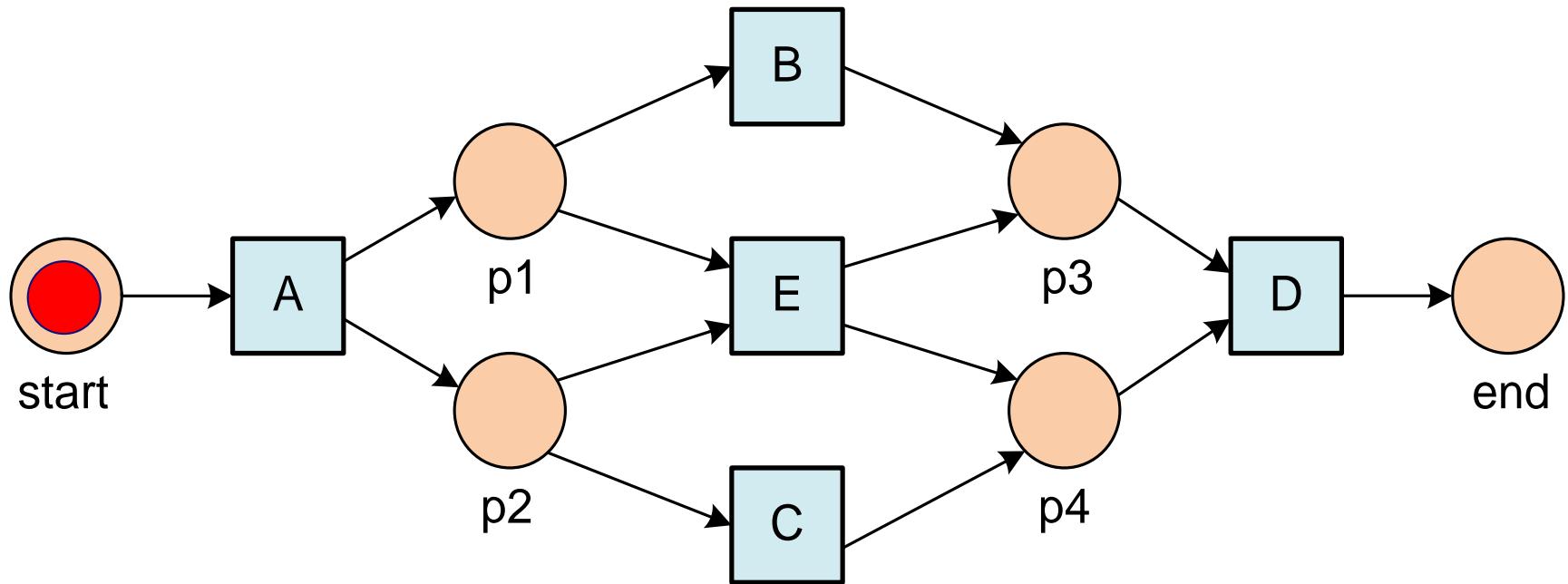


Assumption: every complete trace corresponds to a complete path through the model.

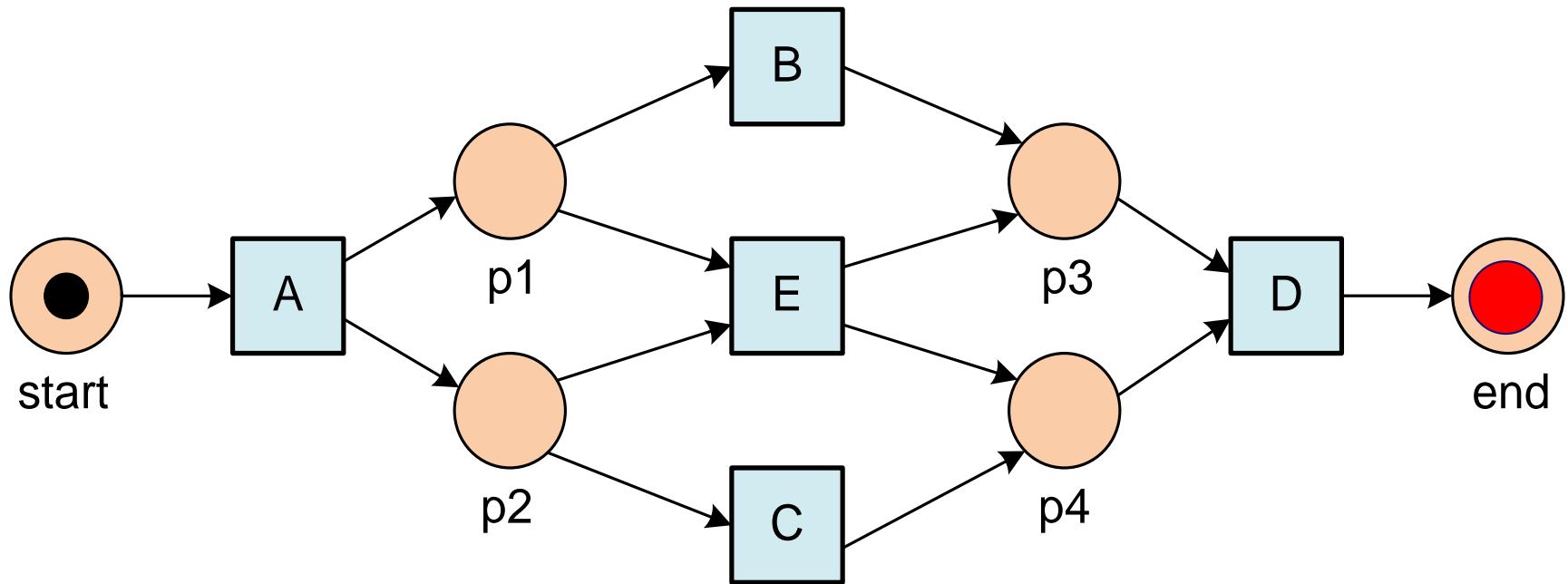
mining decision
points

Decision mining: “Red” cases

A B C D

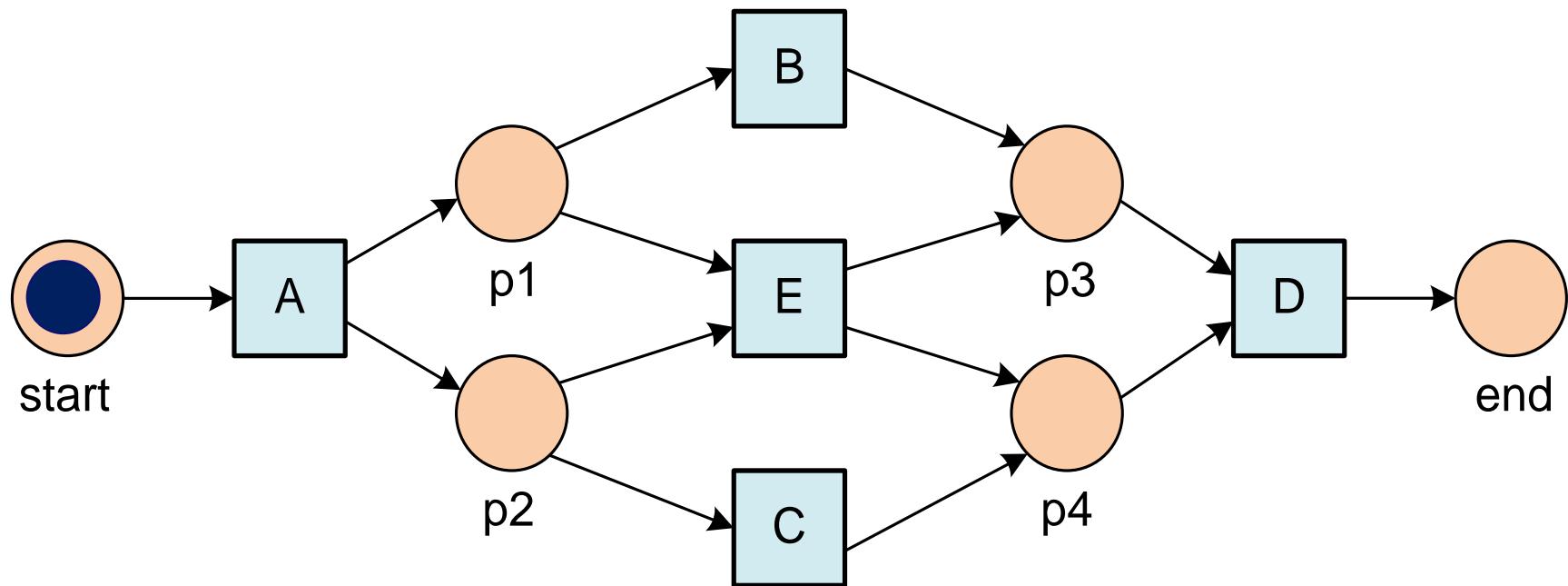


Decision mining: “Red” cases

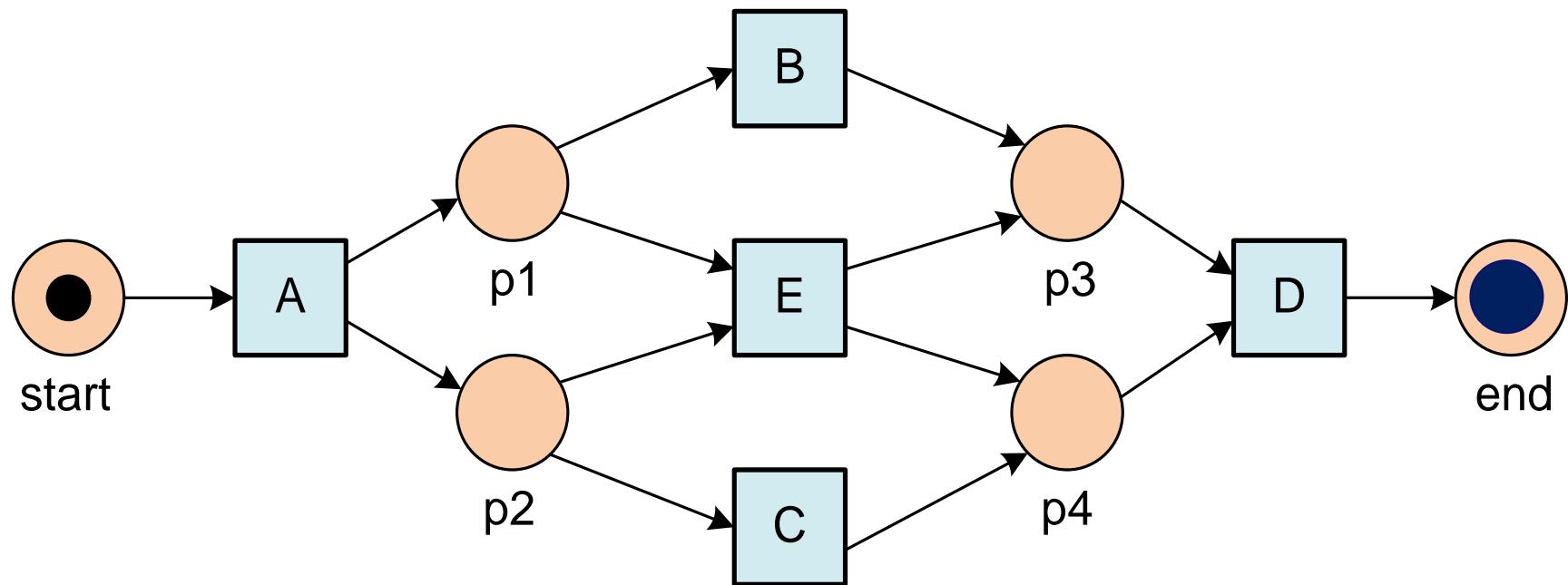


Decision mining: “Blue” cases

A E D

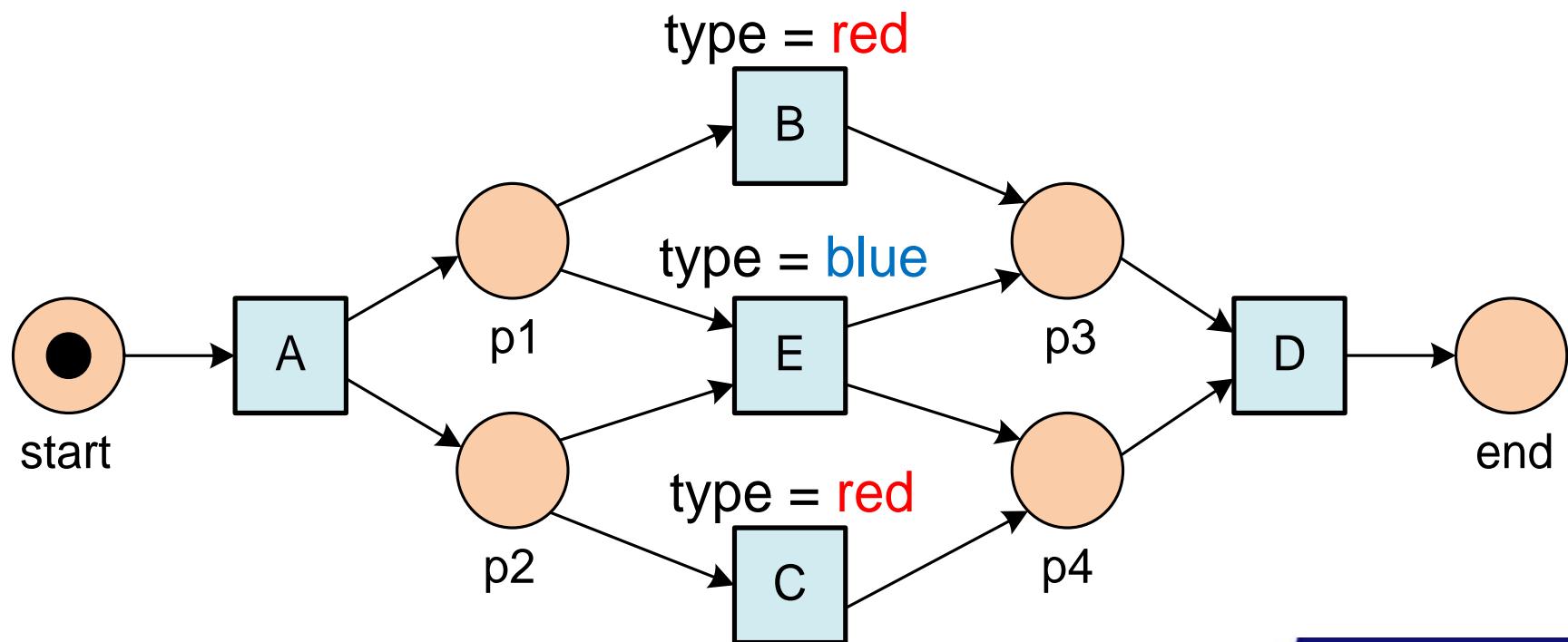


Decision mining: “Blue” cases

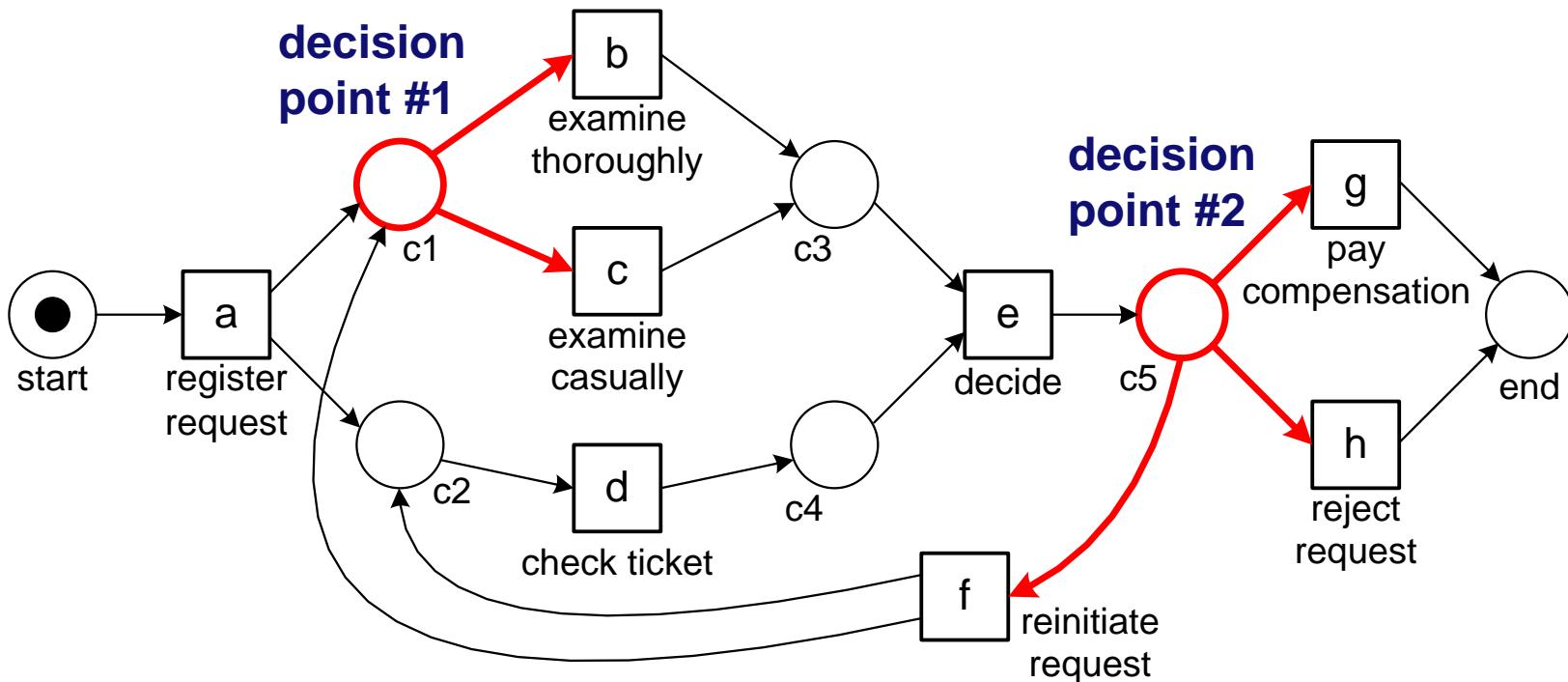


Guards ensure that the right path is taken

If red then B+C;
If blue then E;



Decision mining: places with multiple output arcs form decision points



In other notations often specific building blocks, e.g., XOR/OR gateways.

Remember: classification using decision trees

- **Response variable (dependent variable): claim (yes/no).**
- **Predictor variables (independent variables): gender, age, smoker, car brand.**

gender	age	smoker	car brand	claim
female	47	yes	Volvo	no
male	31	no	Alfa Romeo	yes
male	59	no	Alfa Romeo	yes
male	28	no	Fiat	no
male	44	no	BMW	no
female	27	no	Fiat	no
male	29	no	Subaru	no
male	44	yes	Subaru	yes
male	39	no	BMW	no
male	35	no	Subaru	yes

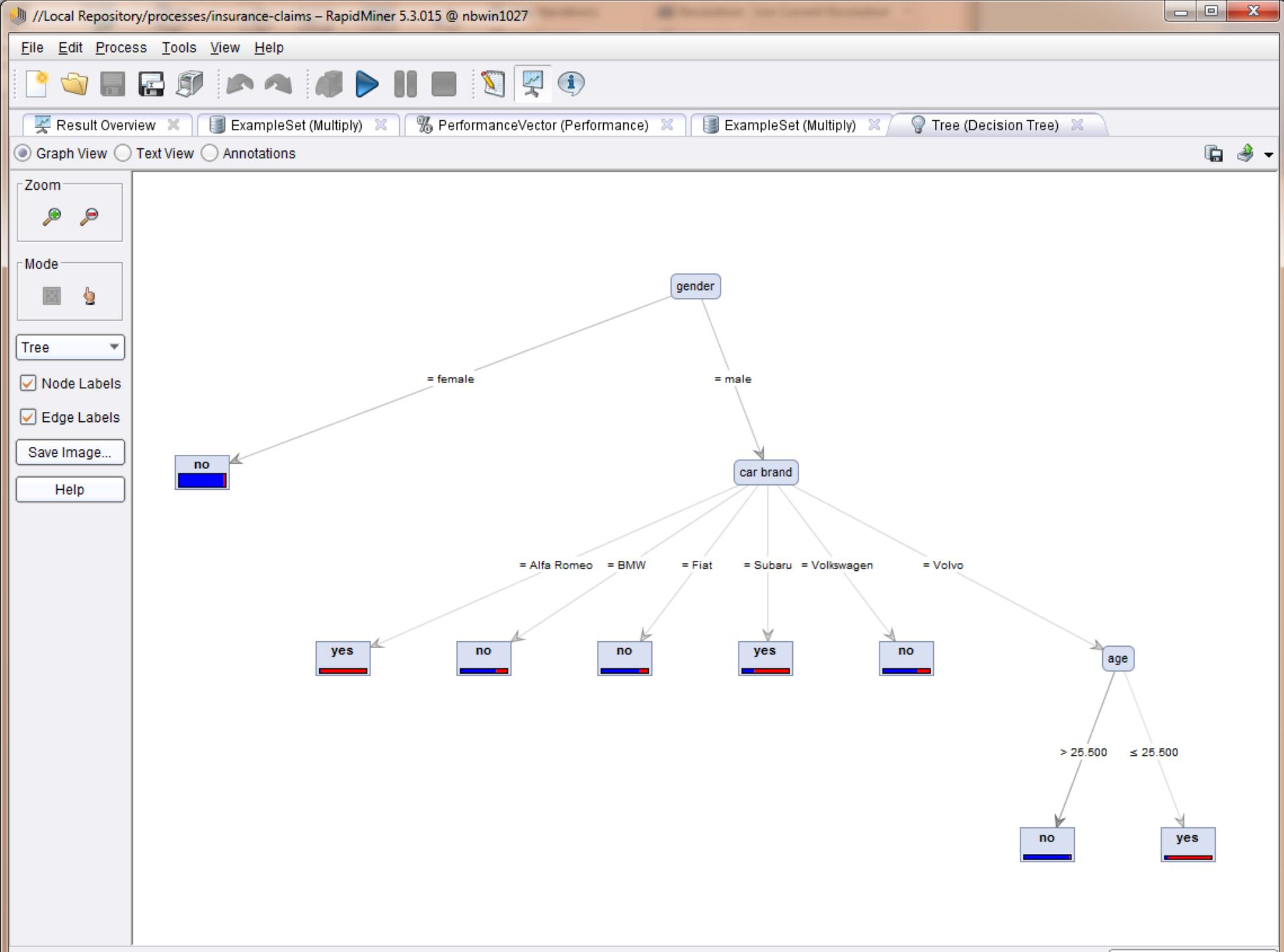
Remember: classification using decision trees

- Response variable (dependent variable): **claim (yes/no)**.
- Predictor variables (independent variables): **gender, age, smoker, car brand.**

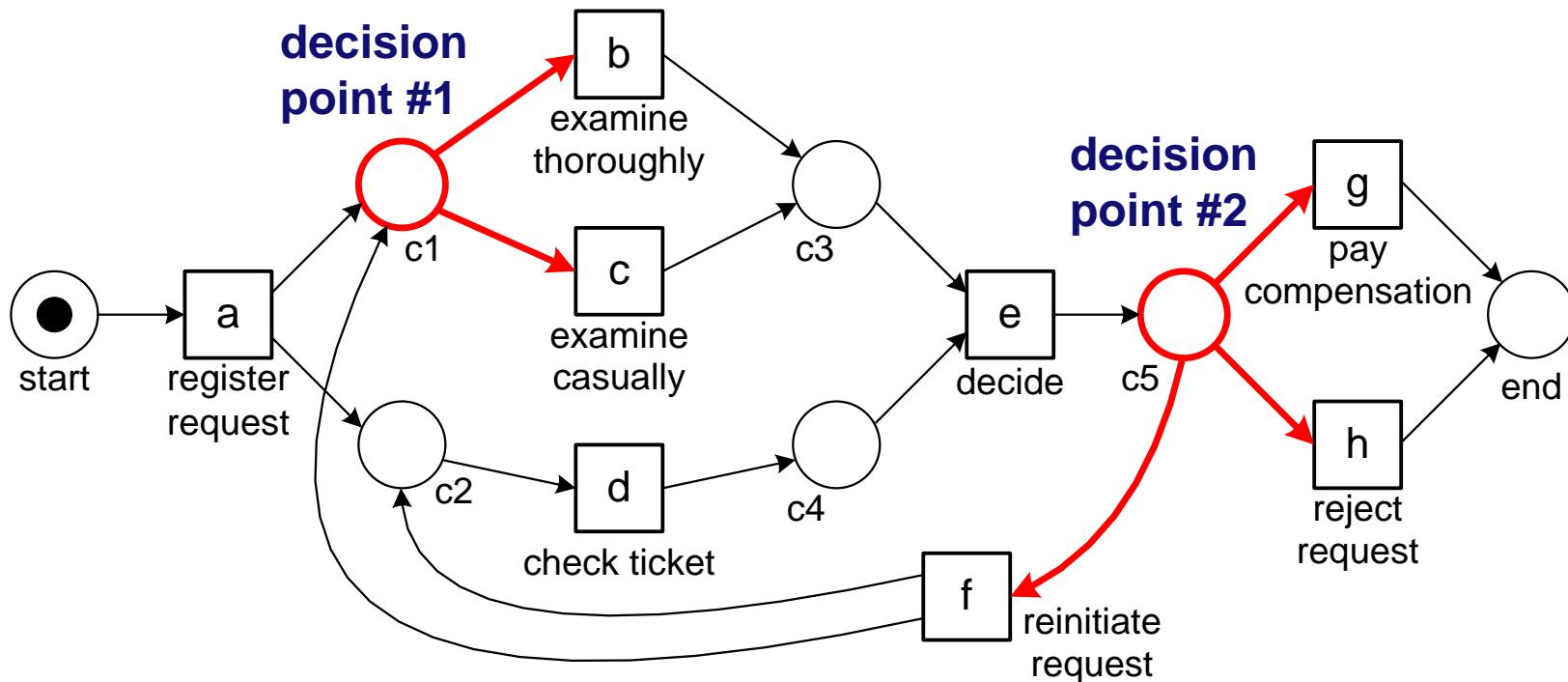


gender	age	smoker	car brand	claim
female	47	yes	Volvo	no
male	31	no	Alfa Romeo	yes
male	59	no	Alfa Romeo	yes
male	28	no	Fiat	no
male	44	no	BMW	no
female	27	no	Fiat	no
male	29	no	Subaru	no
male	44	yes	Subaru	yes
male	39	no	BMW	no
male	35	no	Subaru	yes

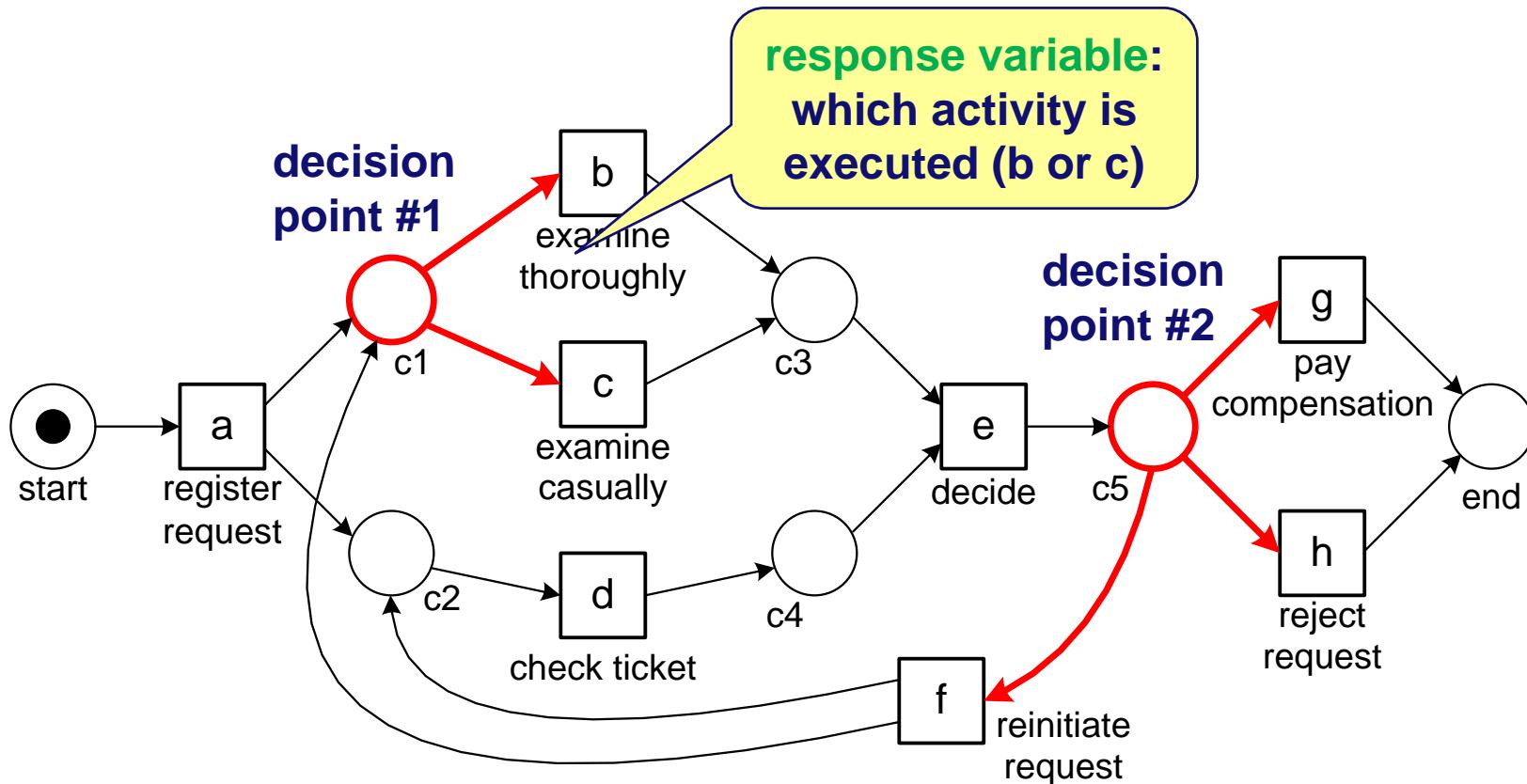
Goal: explain
response variable
in terms of
relevant predictor
variables.



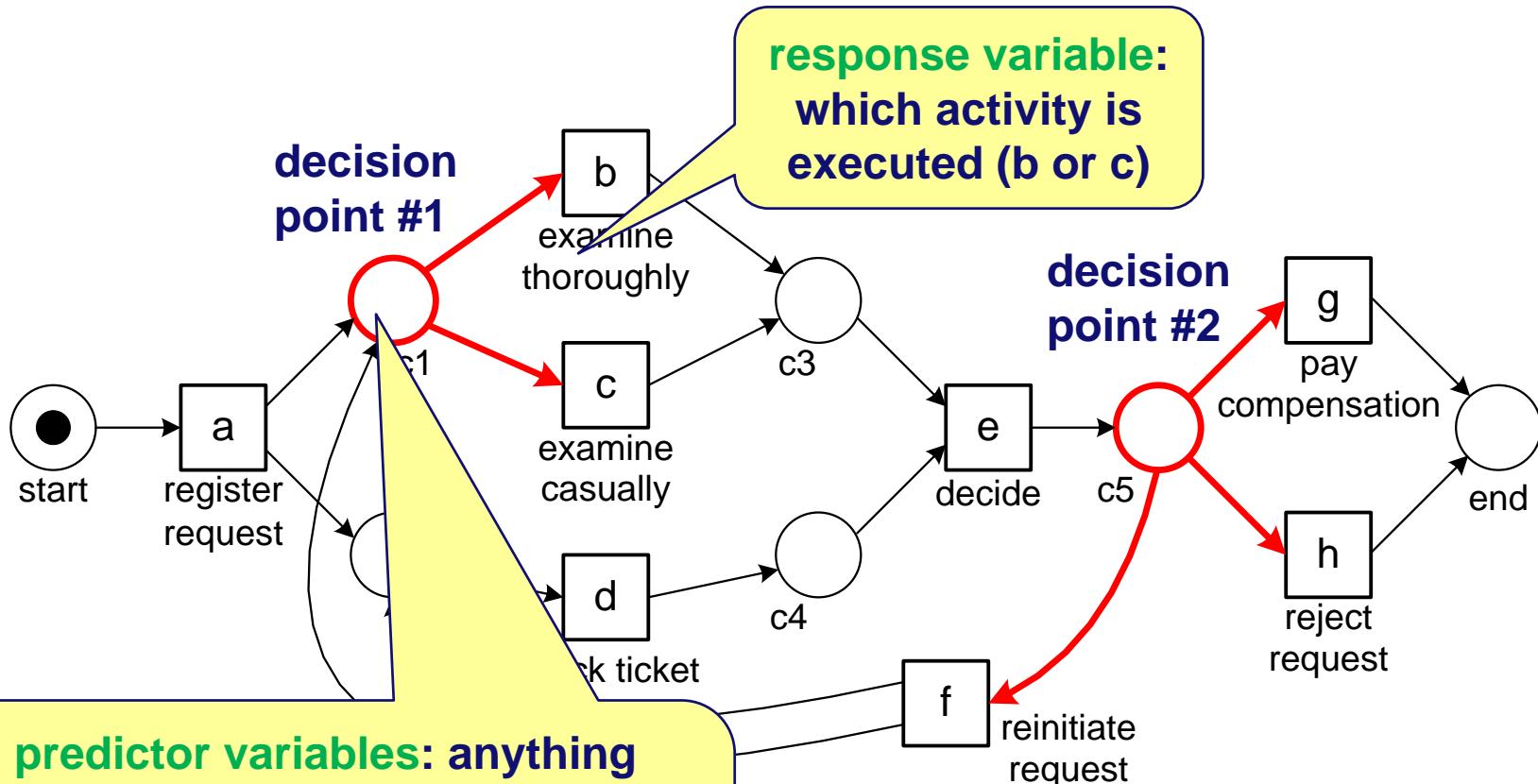
Creating a classification problem



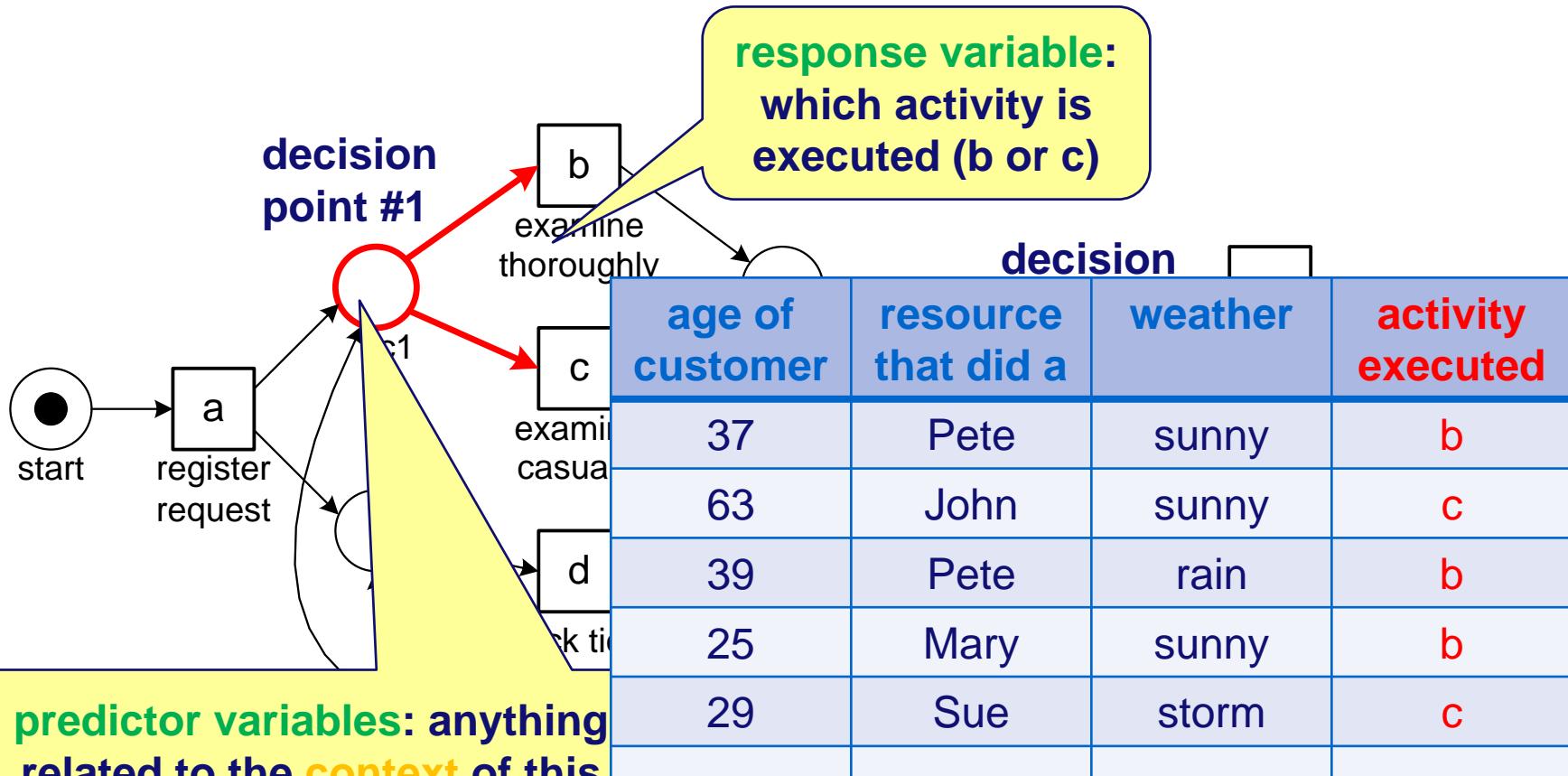
Creating a classification problem



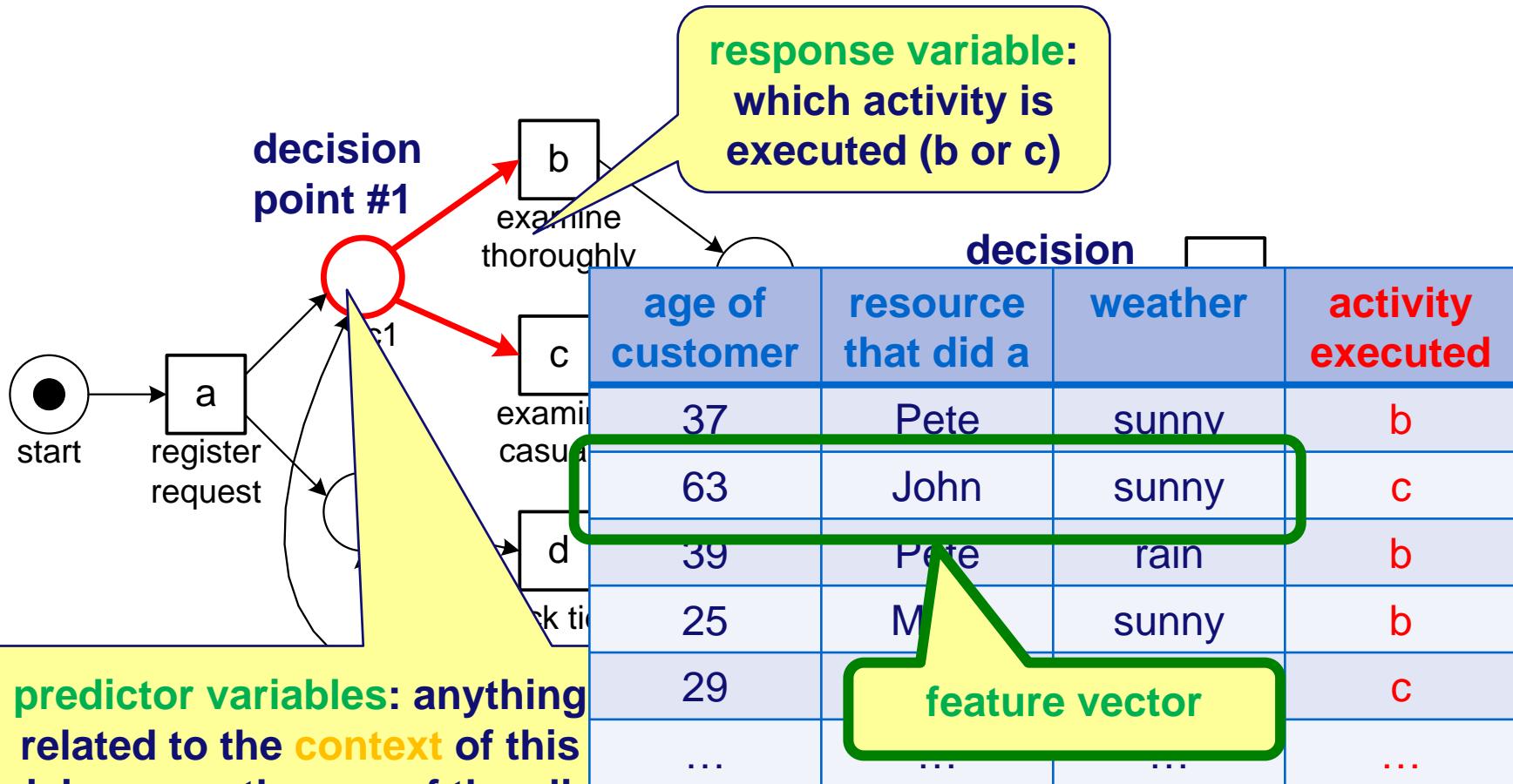
Creating a classification problem



Creating a classification problem

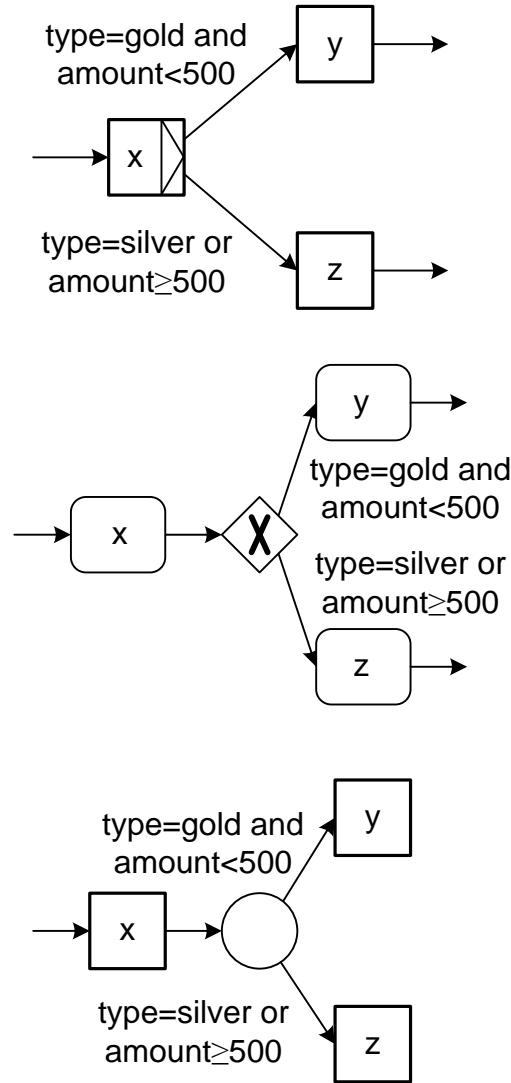


Creating a classification problem



Example: XOR-split

type	region	amount	activity
gold	south	987.30	z
silver	north	178.70	z
gold	south	211.50	y
silver	west	587.70	z
silver	east	224.70	z
silver	south	278.50	z
gold	north	488.50	y
silver	west	443.20	z
silver	south	673.70	z
gold	west	413.50	y
silver	south	687.70	z
gold	south	987.30	z
silver	north	378.80	z
gold	south	314.50	y
silver	north	537.70	z
silver	west	158.70	z
gold	east	344.50	y
...

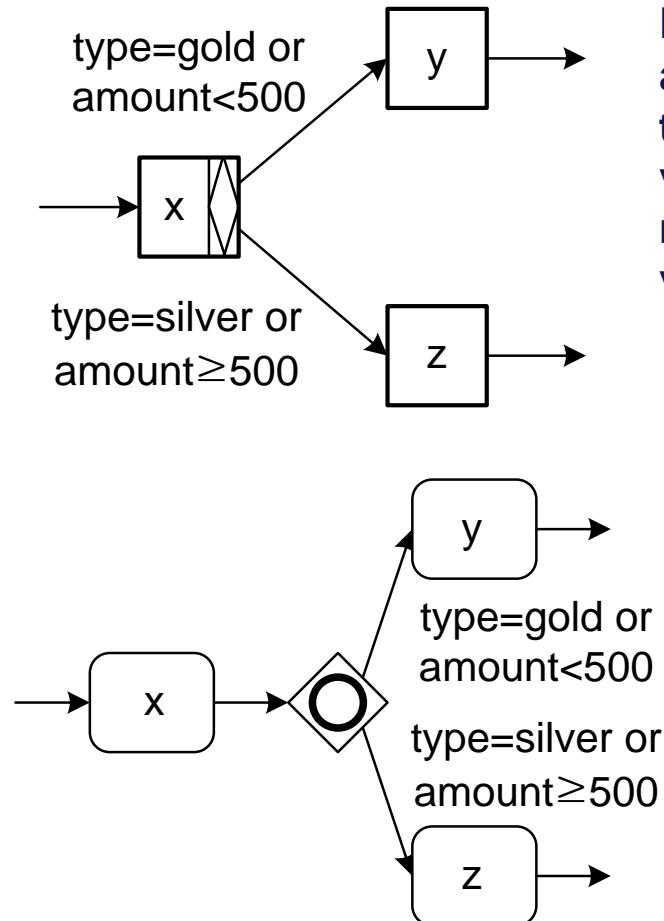


What are the “**features**” (predictor variables) influencing the decision?

A **classification technique** like decision tree learning can be used to find such rules: explain response variable (dependent variable) in terms of predictor variables (independent variables).

Example: OR-split

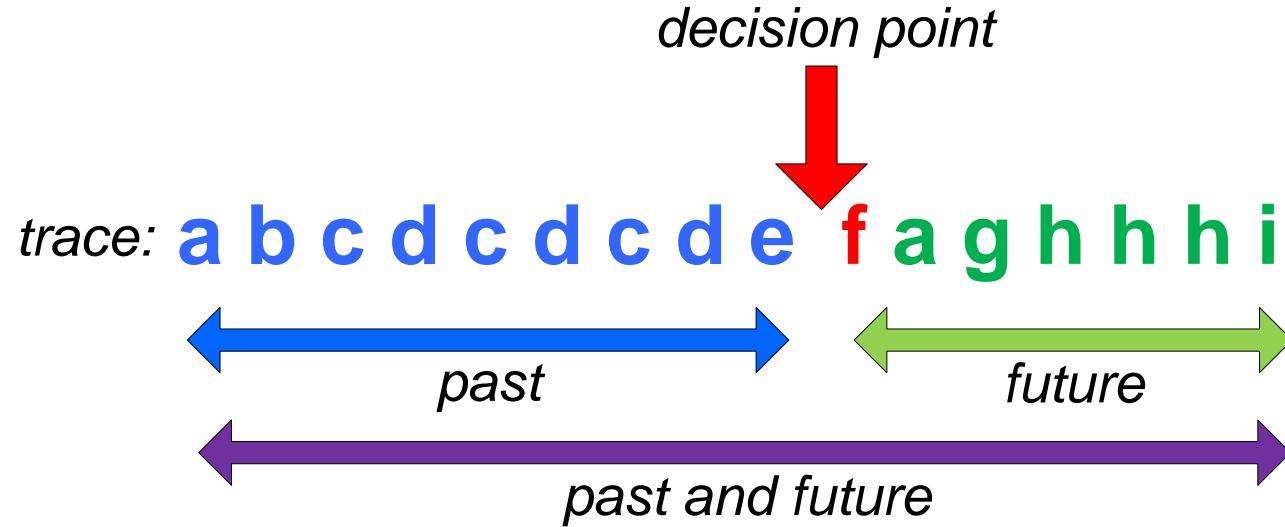
type	region	amount	activity
gold	south	987.30	y and z
silver	north	178.70	y and z
gold	south	211.50	just y
silver	west	587.70	just z
silver	east	224.70	y and z
silver	south	278.50	y and z
gold	north	488.50	just y
silver	west	443.20	y and z
silver	south	673.70	just z
...



Replay/ alignment will tell what the value of the response variable is.

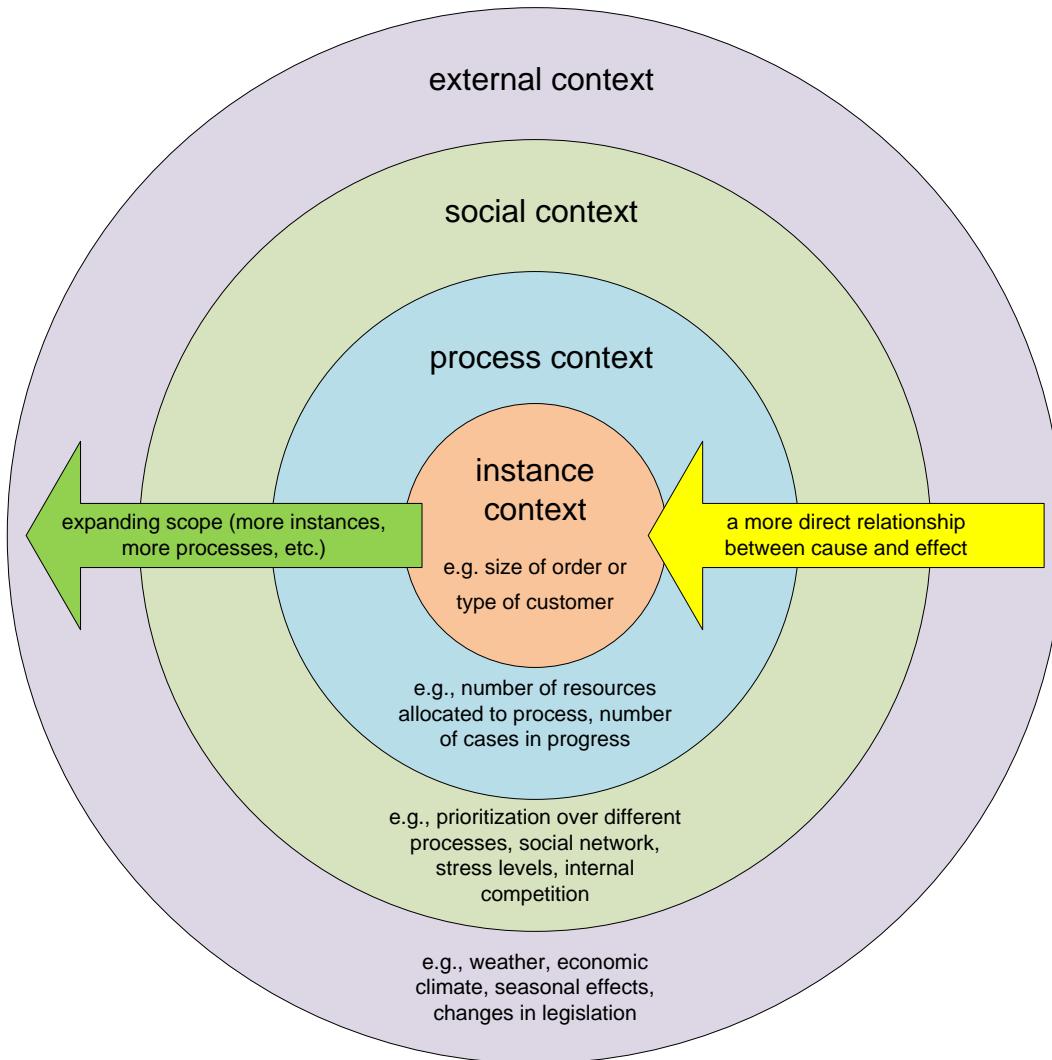
Response variable (dependent variable) is now an activity set rather than a single activity.

Predictor variables based on process instance only



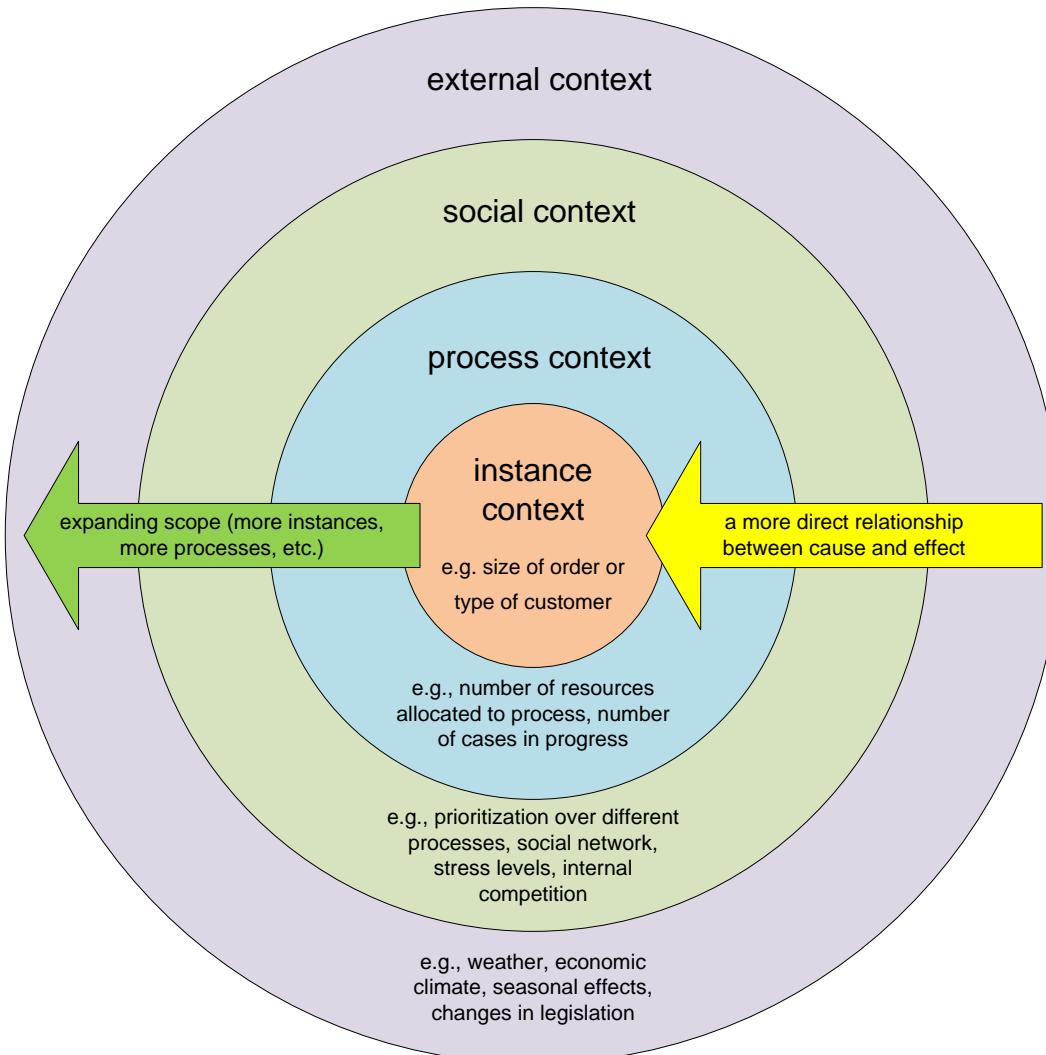
- Case variables
- Attributes (data, resource, etc.) of all past events
- Attributes (data, resource, etc.) of last event only
- Attributes (data, resource, etc.) of all events (including future).

Predictor variables based on context of the process instance



- Number of cases running (e.g. skip check if busy).
- Number of resources present.
- Day of the week.
- Weather.

Predictor variables based on context of the process instance

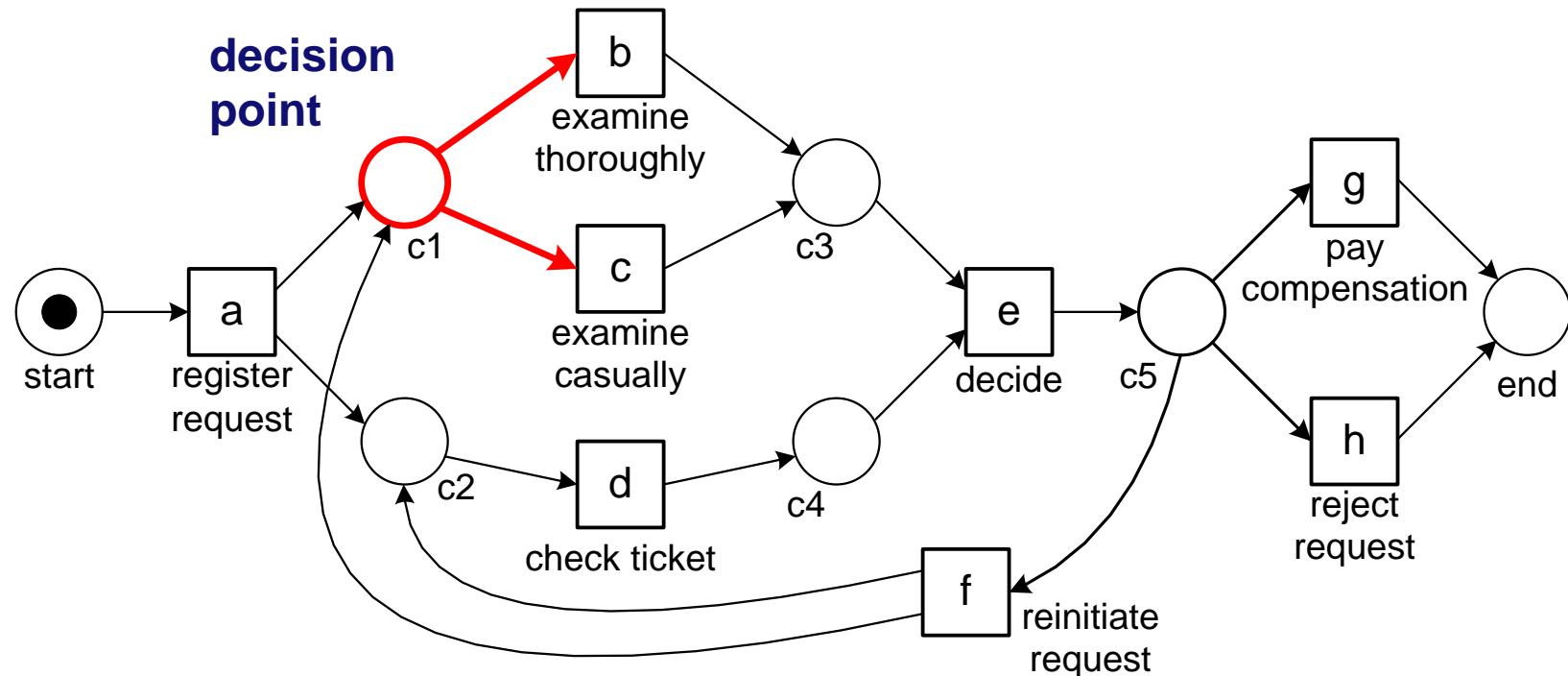


- Number of cases running (e.g. skip check if busy).
- Number of resources present.
- Day of the week.
- Weather.



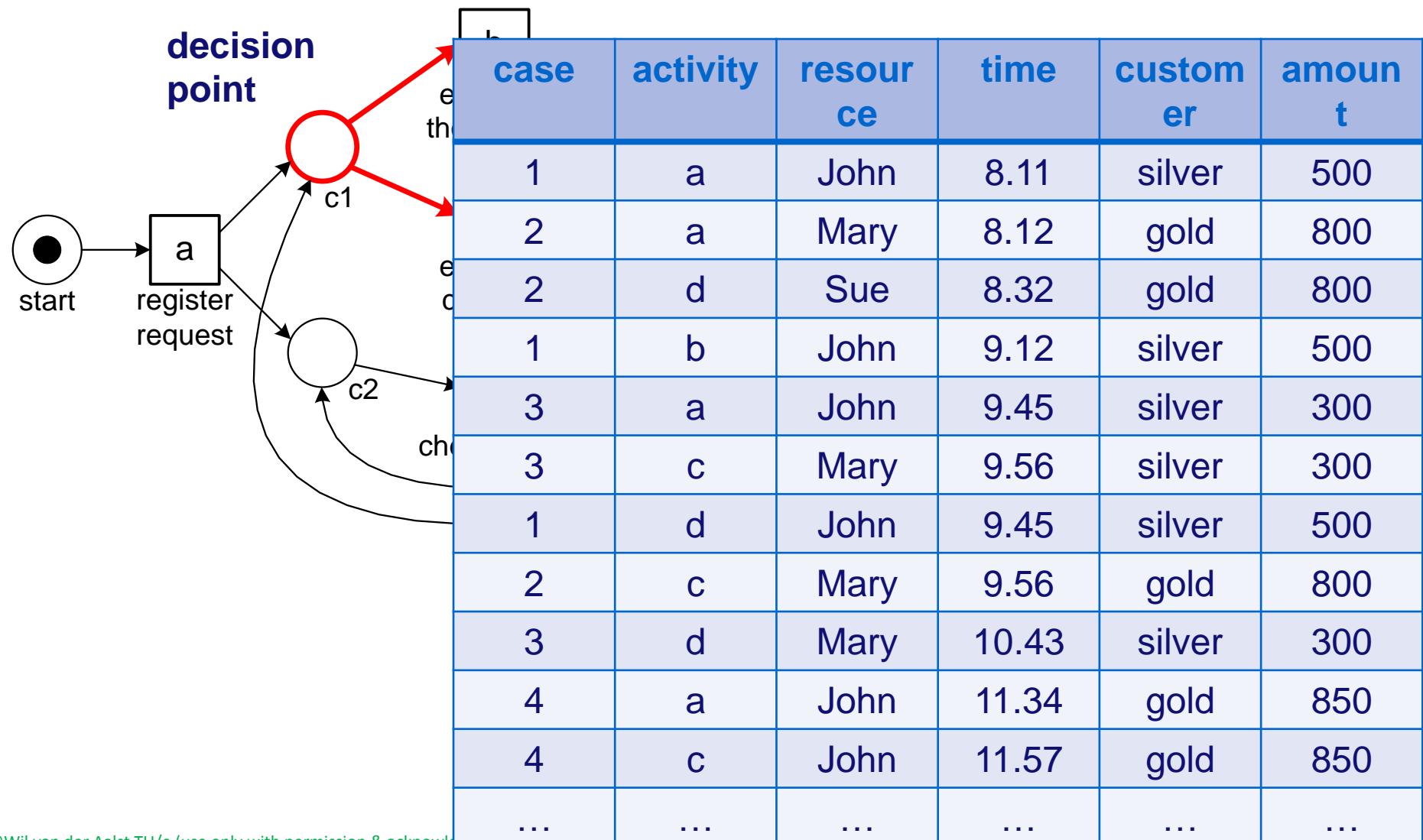
Problem: curse of dimensionality!

Question: Build classification problem to learn the decision point

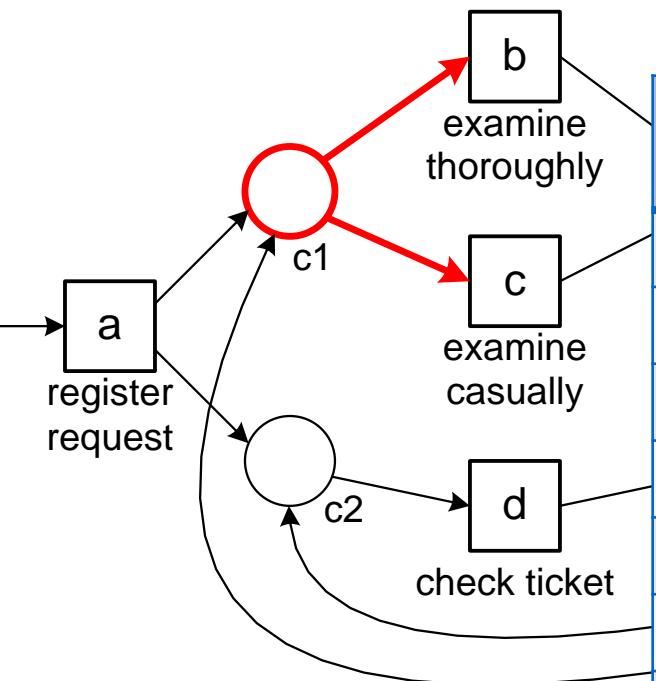


case	activity	resource	time	customer	amount
1	a	John	8.11	silver	500
2	a	Mary	8.12	gold	800
3	a	S	8.22	II	200

Question: Build classification problem to learn the decision point



Question: Build classification problem to learn the decision point



case	activity	resource	time	customer	amount
1	a	John	8.11	silver	500
2	a	Mary	8.12	gold	800
2	d	Sue	8.32	gold	800
1	b	John	9.12	silver	500
3	a	John	9.45	silver	300
3	c	Mary	9.56	silver	300
1	d	John	9.45	silver	500
2	c	Mary	9.56	gold	800
3	d	Mary	10.43	silver	300
4	a	John	11.34	gold	850
4	c	John	11.57	gold	850
...

Classification problem (example, not all possible attributes included)

Diagram illustrating a classification problem based on a process model and a database.

The database table contains the following data:

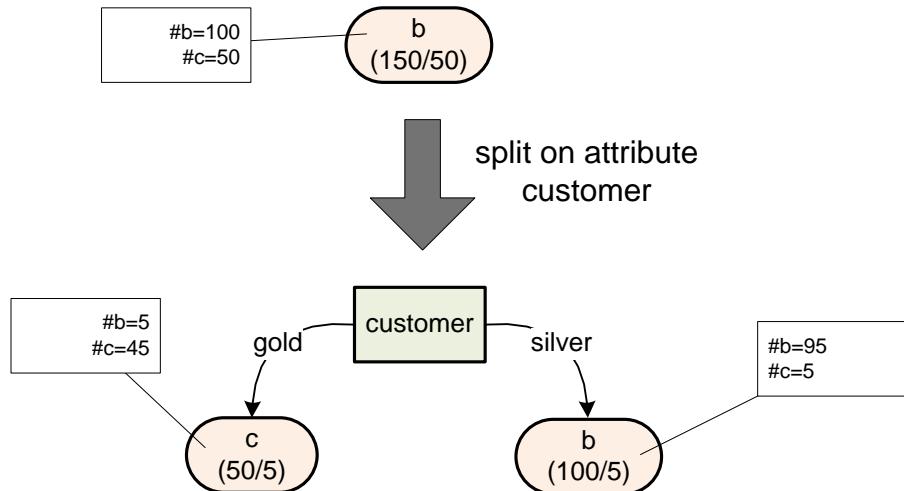
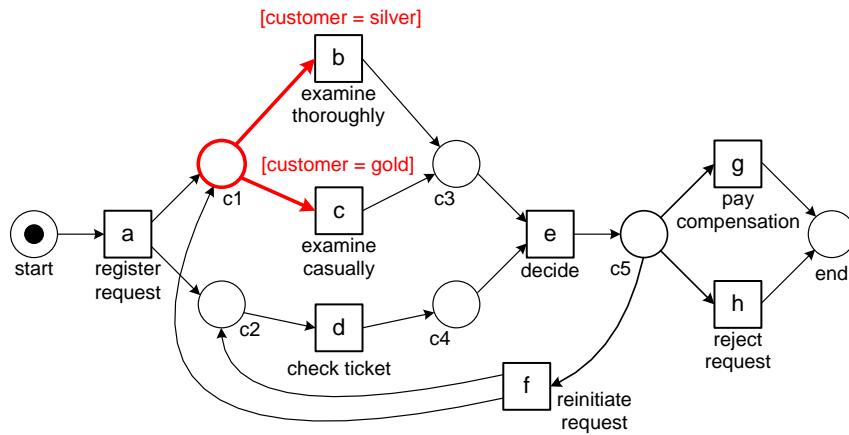
case	activity	resource	time	customer	amount
1	a	John	8.11	silver	500
2	a	Mary	8.12	gold	800
2	d	Sue	8.32	gold	800
1	b	John	9.12	silver	500
3	a	John	9.45	silver	300
3	c	Mary	9.56	silver	300
1	d	John	9.45	silver	500
2	a	Mary	9.56	gold	800

Below the main table, there is a secondary table with the following data:

case	resource executing a	customer	amount	class
1	John	silver	500	b
2	Mary	gold	800	c
3	John	silver	300	b
4	John	gold	850	c

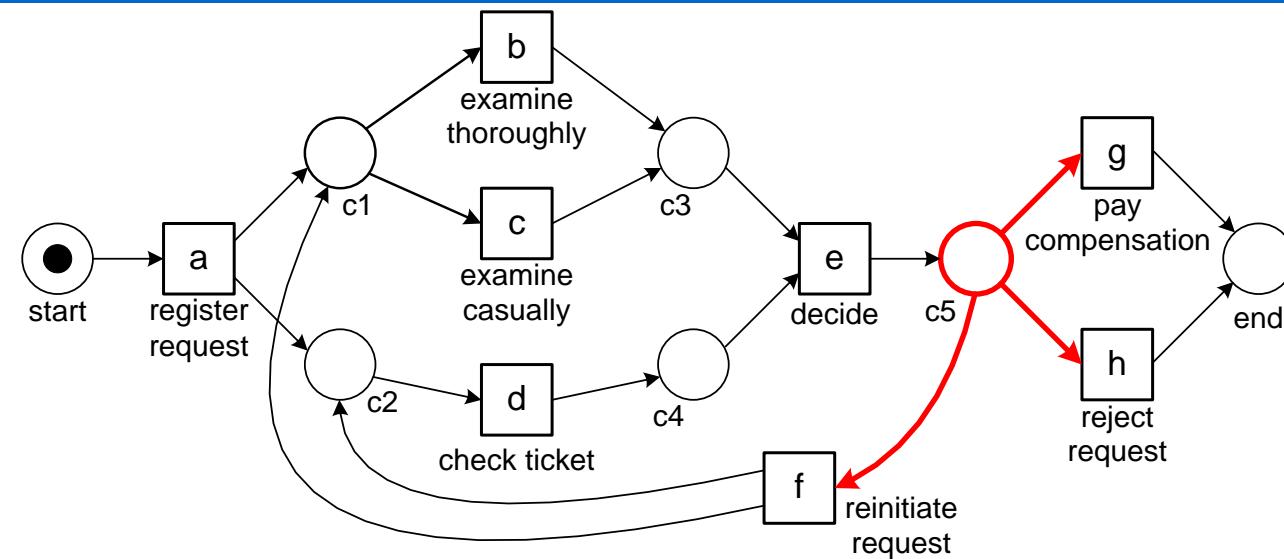
PAGE 37

Decision tree and resulting guards

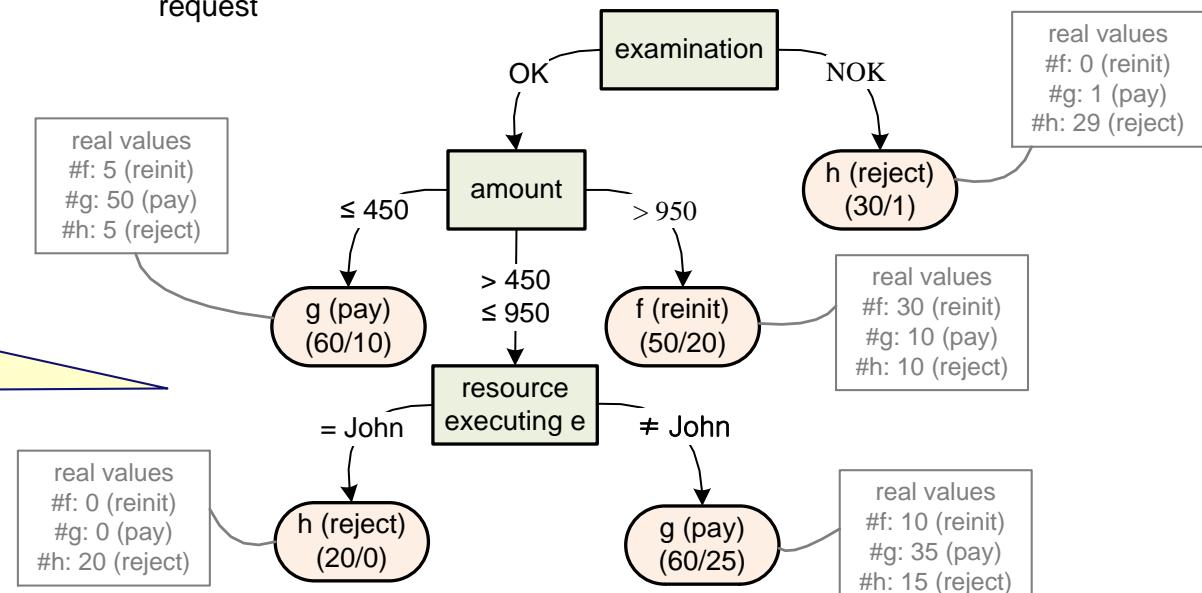


case	resource executing a	customer	amount	class
1	John	silver	500	b
2	Mary	gold	800	c
3	John	silver	300	b
4	John	gold	850	c

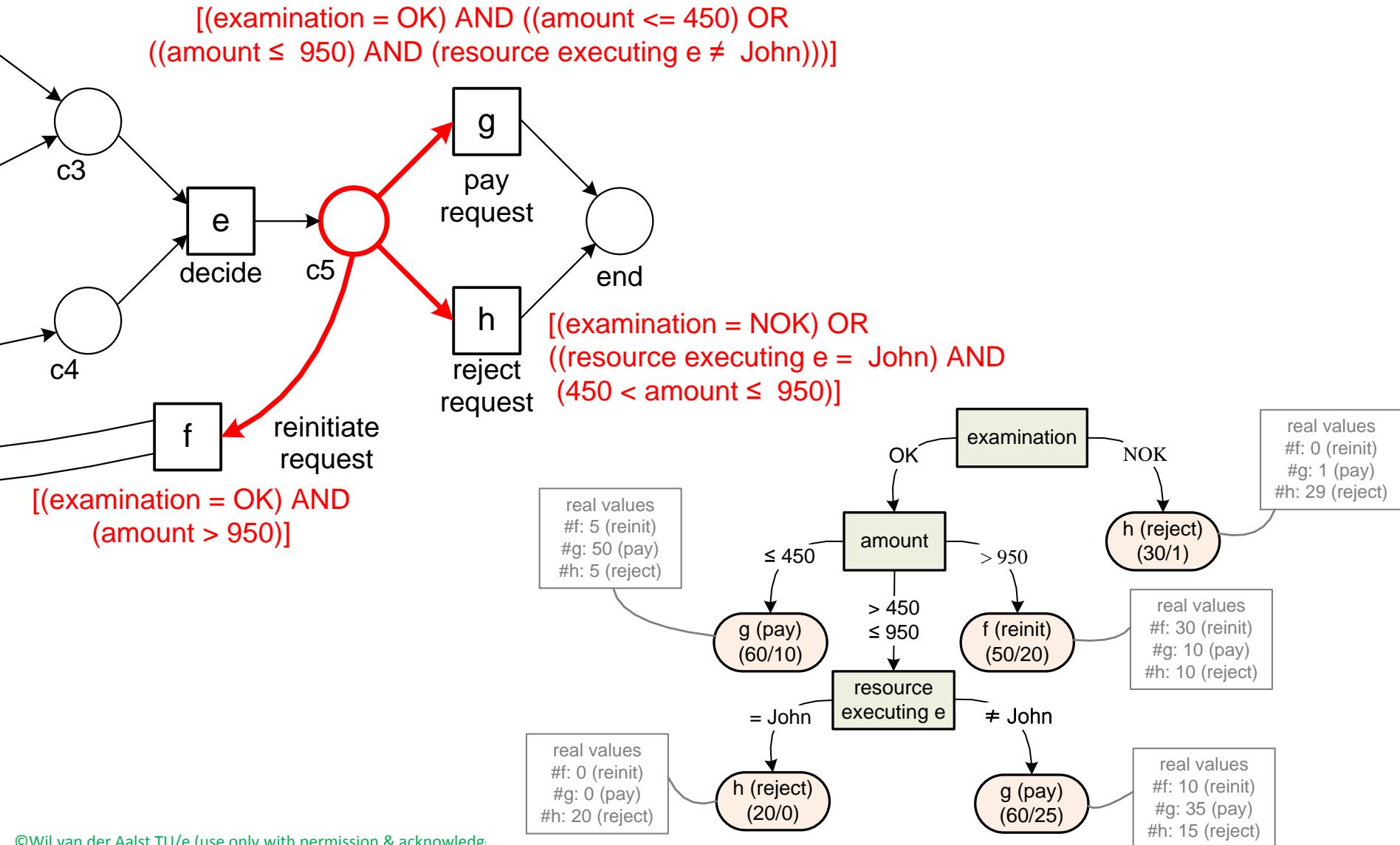
Question: create guards based on decision tree



assume that this decision tree was learned for decision point

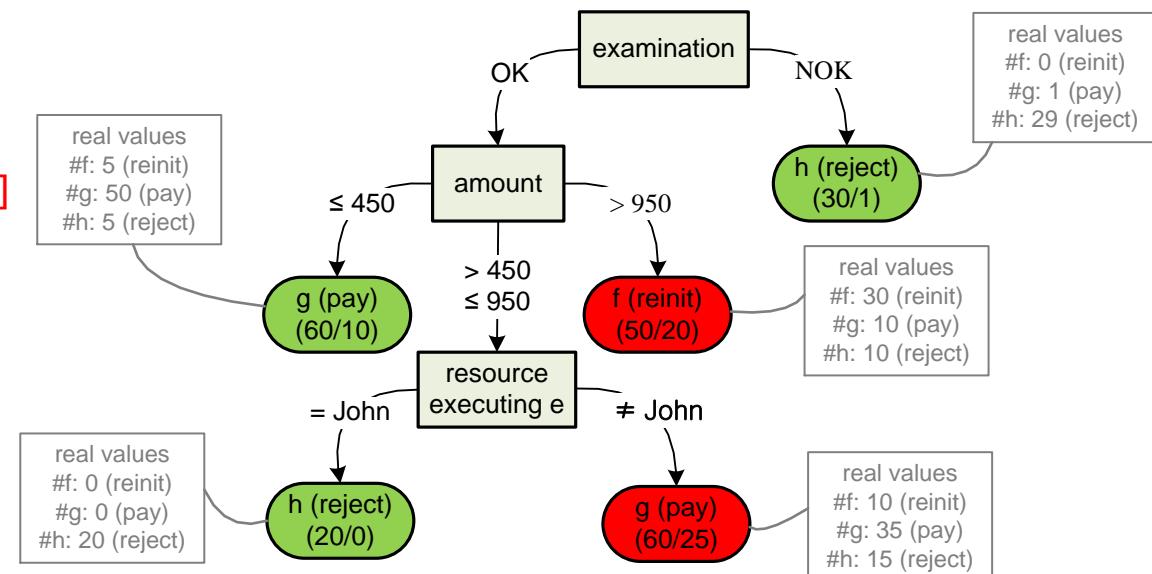
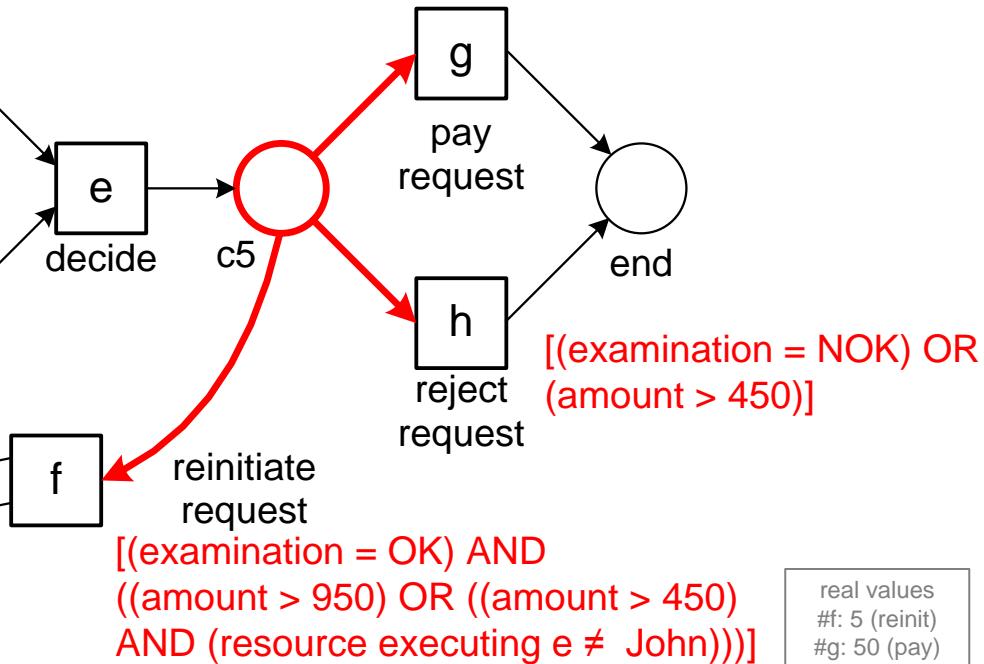


Deterministic guards



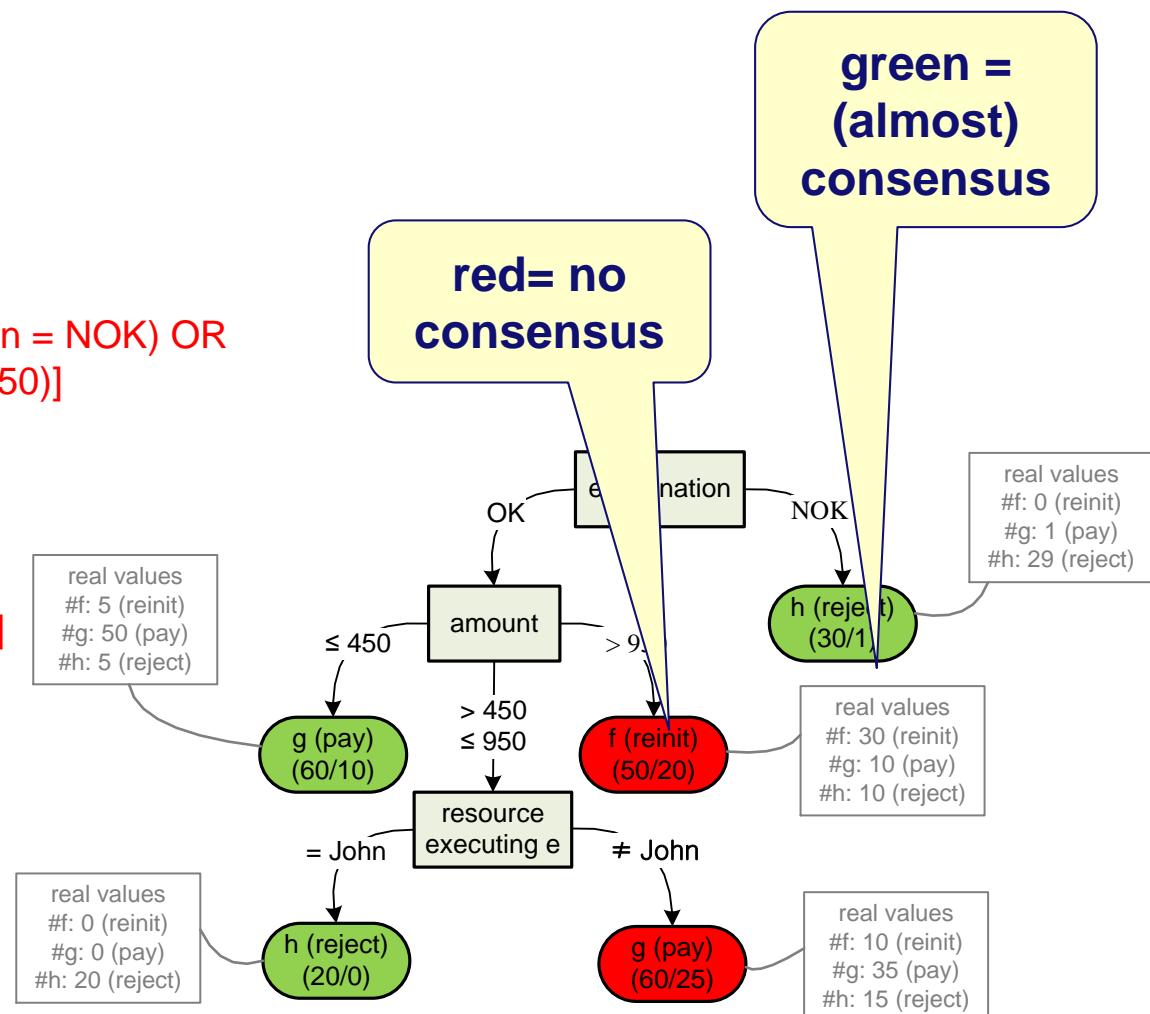
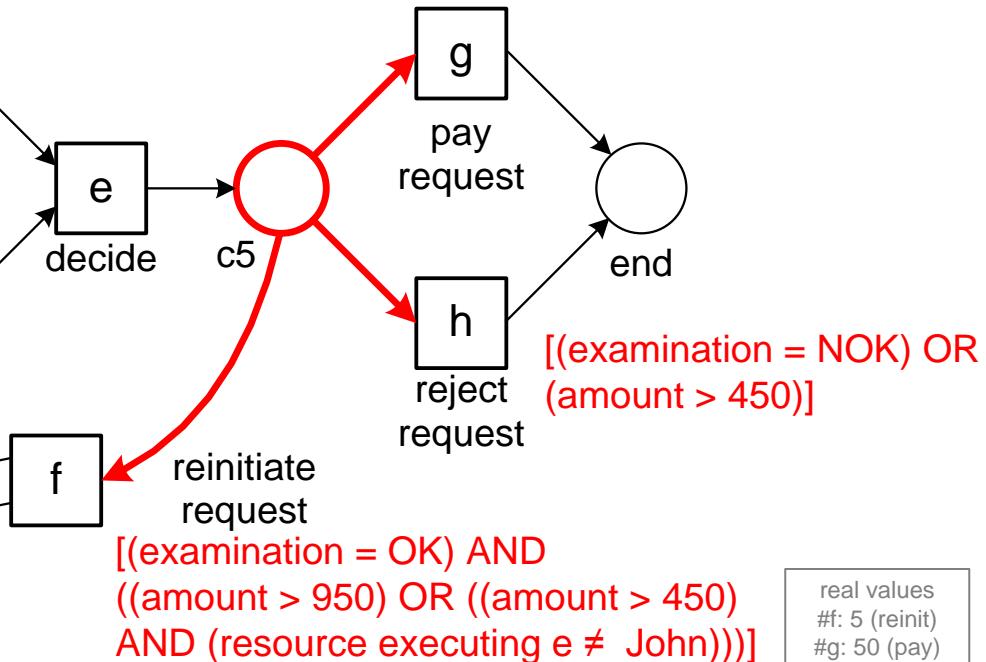
Non-deterministic guards

$[(\text{examination} = \text{OK}) \text{ AND } ((\text{amount} \leq 450) \text{ OR } (\text{amount} > 950)) \text{ OR } (\text{resource executing } e \neq \text{John})]]$



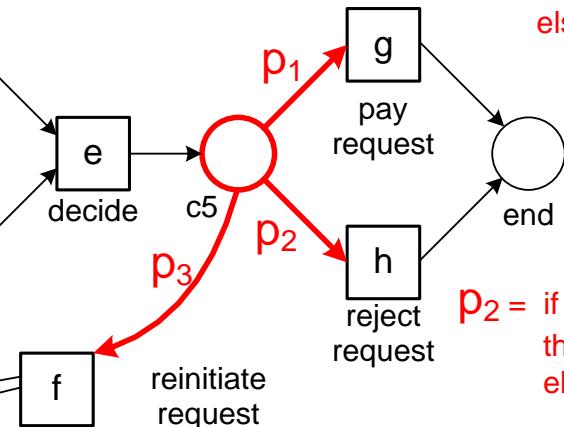
Non-deterministic guards

$[(\text{examination} = \text{OK}) \text{ AND } ((\text{amount} \leq 450) \text{ OR } (\text{amount} > 950)) \text{ OR } (\text{resource executing } e \neq \text{John})]]$



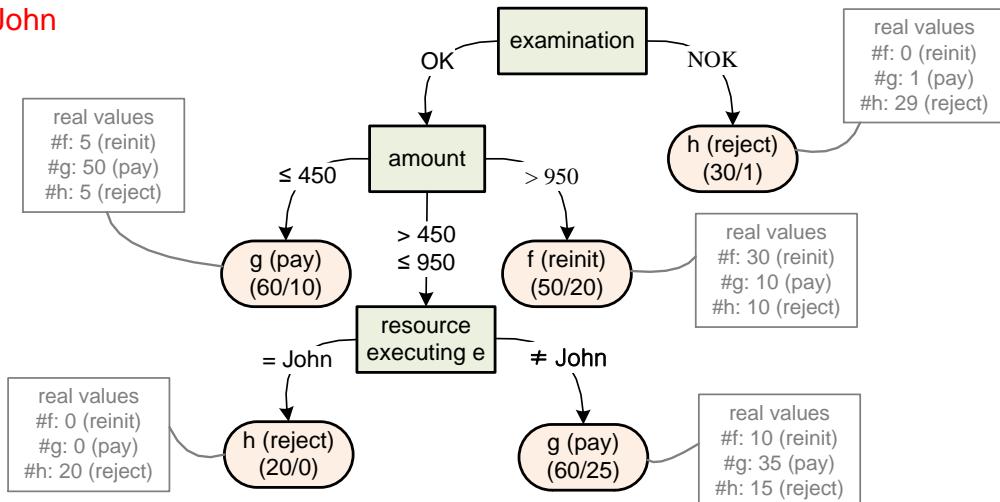
Data-dependent probabilities rather than guards

$p_1 = \begin{cases} \text{if examination} = \text{NOK} \\ \text{then } 1/30 \\ \text{else if amount} \leq 450 \\ \text{then } 50/60 \\ \text{else if amount} > 950 \\ \text{then } 10/50 \\ \text{else if resource executing } a = \text{John} \\ \text{then } 0/20 \\ \text{else } 35/60 \end{cases}$



$p_2 = \begin{cases} \text{if examination} = \text{NOK} \\ \text{then } 29/30 \\ \text{else if amount} \leq 450 \\ \text{then } 5/60 \\ \text{else if amount} > 950 \\ \text{then } 10/50 \\ \text{else if resource executing } a = \text{John} \\ \text{then } 20/20 \\ \text{else } 15/60 \end{cases}$

$p_3 = \begin{cases} \text{if examination} = \text{NOK} \\ \text{then } 0/30 \\ \text{else if amount} \leq 450 \\ \text{then } 5/60 \\ \text{else if amount} > 950 \\ \text{then } 30/50 \\ \text{else if resource executing } a = \text{John} \\ \text{then } 0/20 \\ \text{else } 10/60 \end{cases}$



So we can mine decision points ...

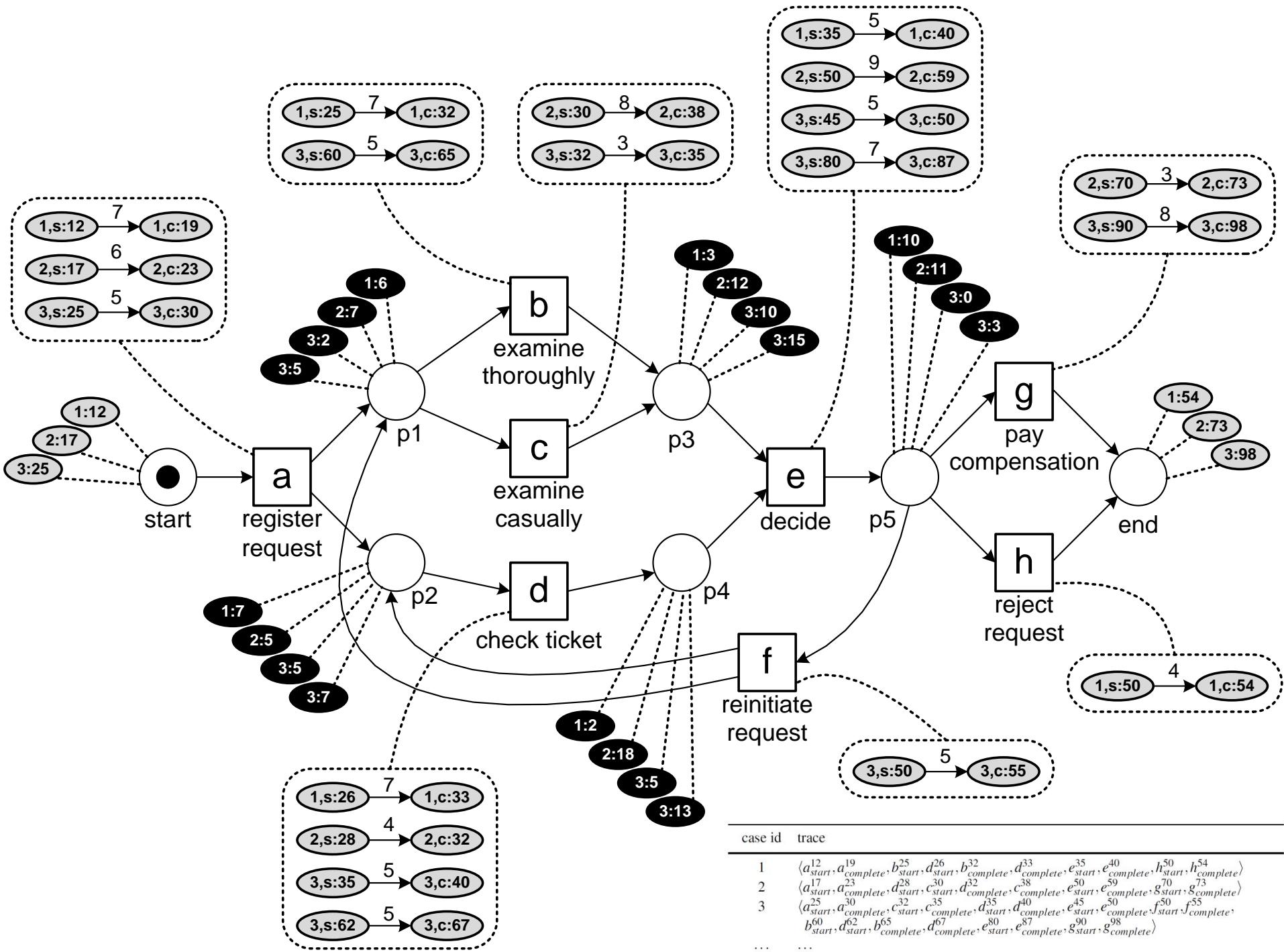


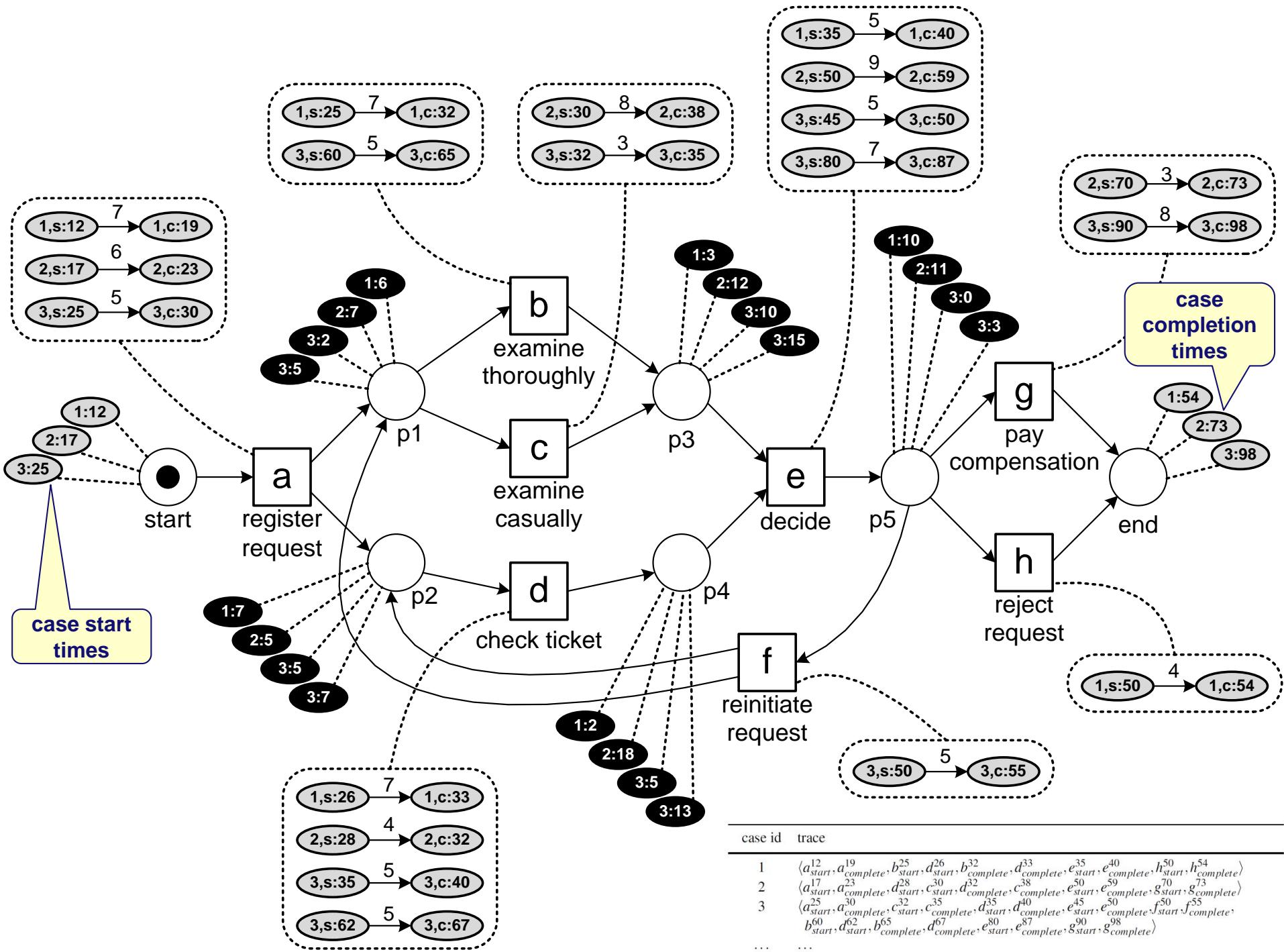
mining bottlenecks

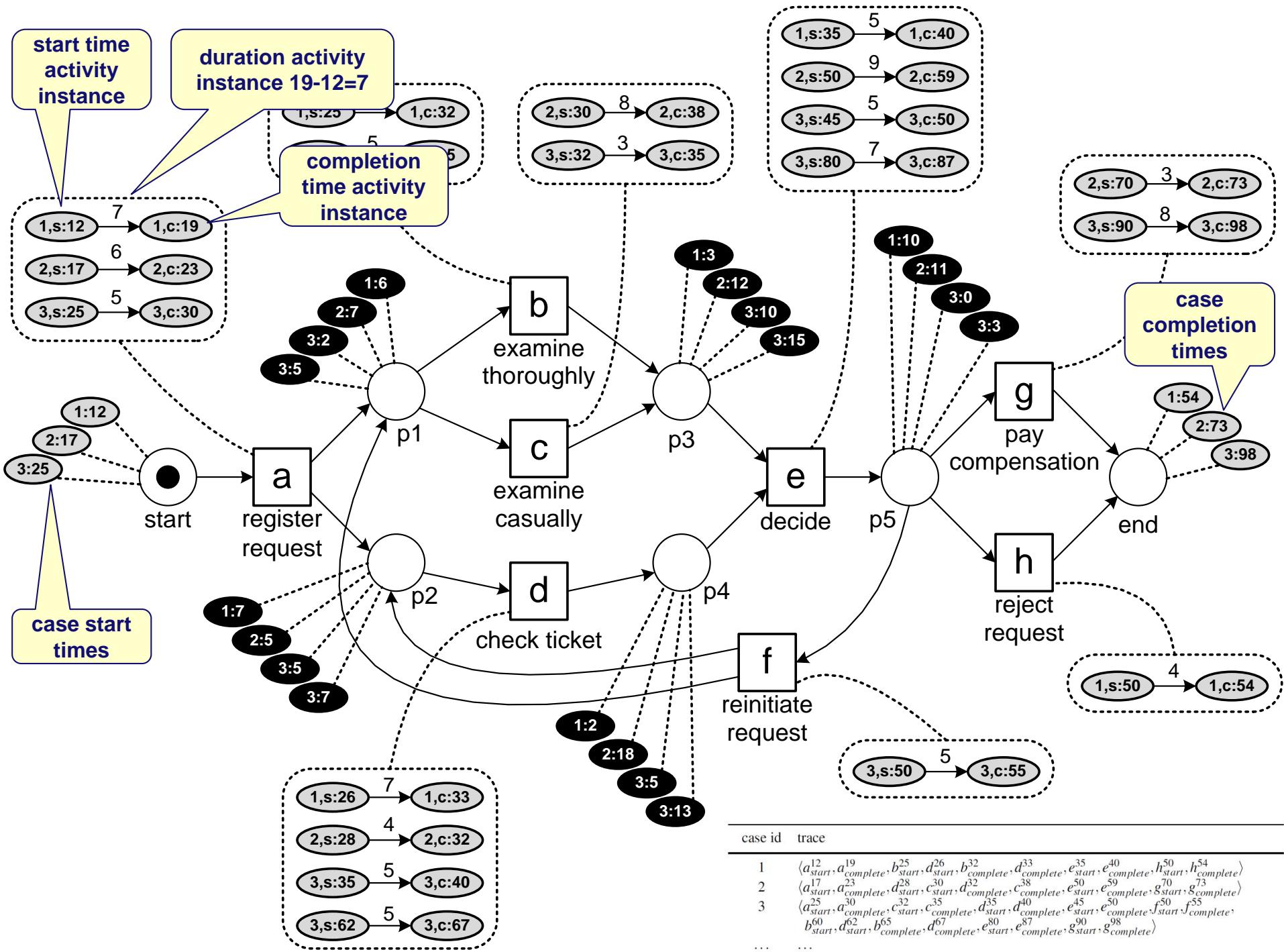
Learning time and probabilities

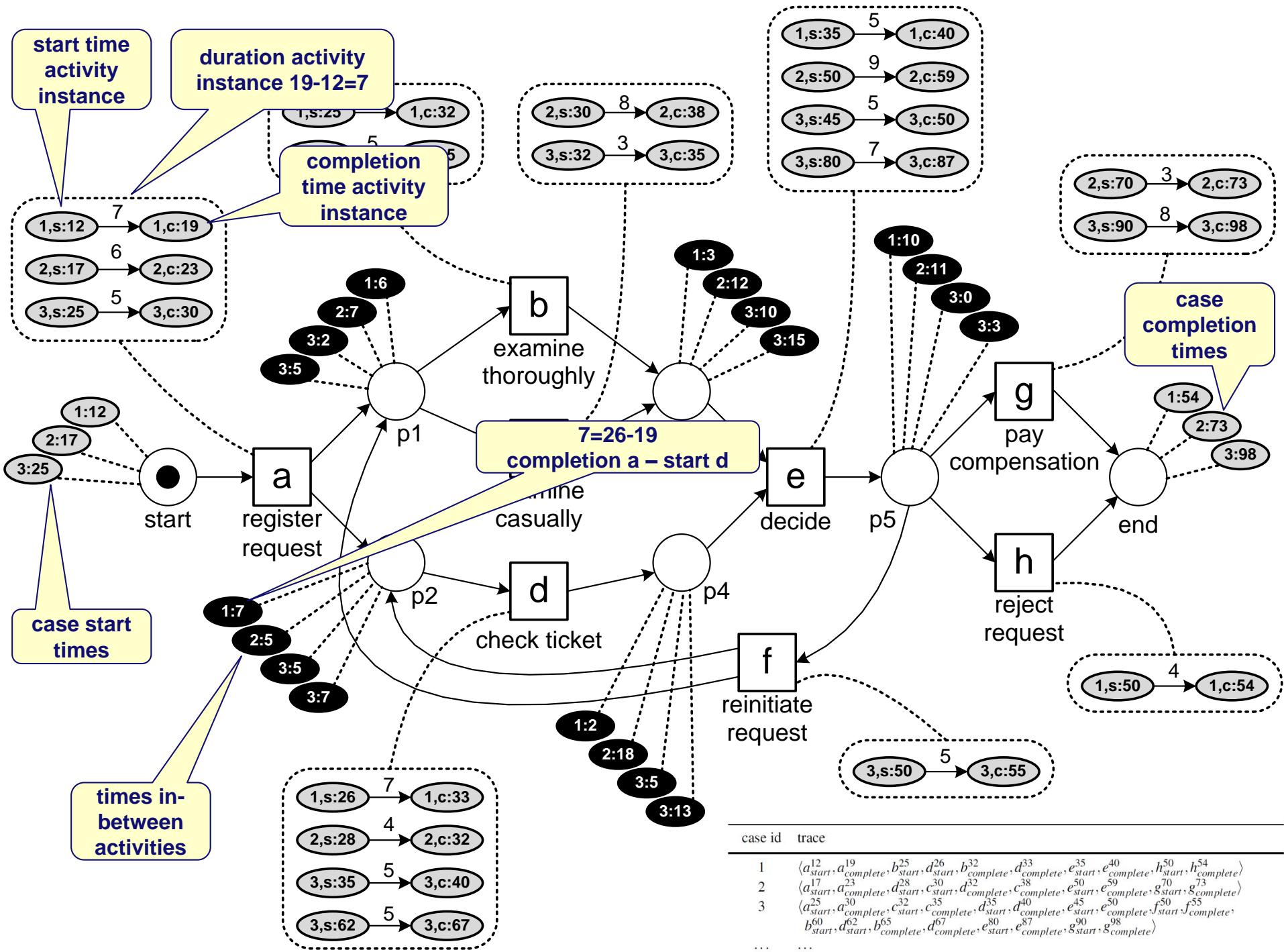
case id	trace
1	$\langle a_{start}^{12}, a_{complete}^{19}, b_{start}^{25}, d_{start}^{26}, b_{complete}^{32}, d_{complete}^{33}, e_{start}^{35}, e_{complete}^{40}, h_{start}^{50}, h_{complete}^{54} \rangle$
2	$\langle a_{start}^{17}, a_{complete}^{23}, d_{start}^{28}, c_{start}^{30}, d_{complete}^{32}, c_{complete}^{38}, e_{start}^{50}, e_{complete}^{59}, g_{start}^{70}, g_{complete}^{73} \rangle$
3	$\langle a_{start}^{25}, a_{complete}^{30}, c_{start}^{32}, c_{complete}^{35}, d_{start}^{35}, d_{complete}^{40}, e_{start}^{45}, e_{complete}^{50}, f_{start}^{50}, f_{complete}^{55}, b_{start}^{60}, d_{start}^{62}, b_{complete}^{65}, d_{complete}^{67}, e_{start}^{80}, e_{complete}^{87}, g_{start}^{90}, g_{complete}^{98} \rangle$
...	...

- Replay, as before, but now considering timestamps.
- Let us replay the first three cases in the event log:
 - case 1 starts at time 12 and ends at time 54,
 - case 2 starts at time 17 and ends at time 73,
 - case 3 starts at time 25 and ends at time 98.

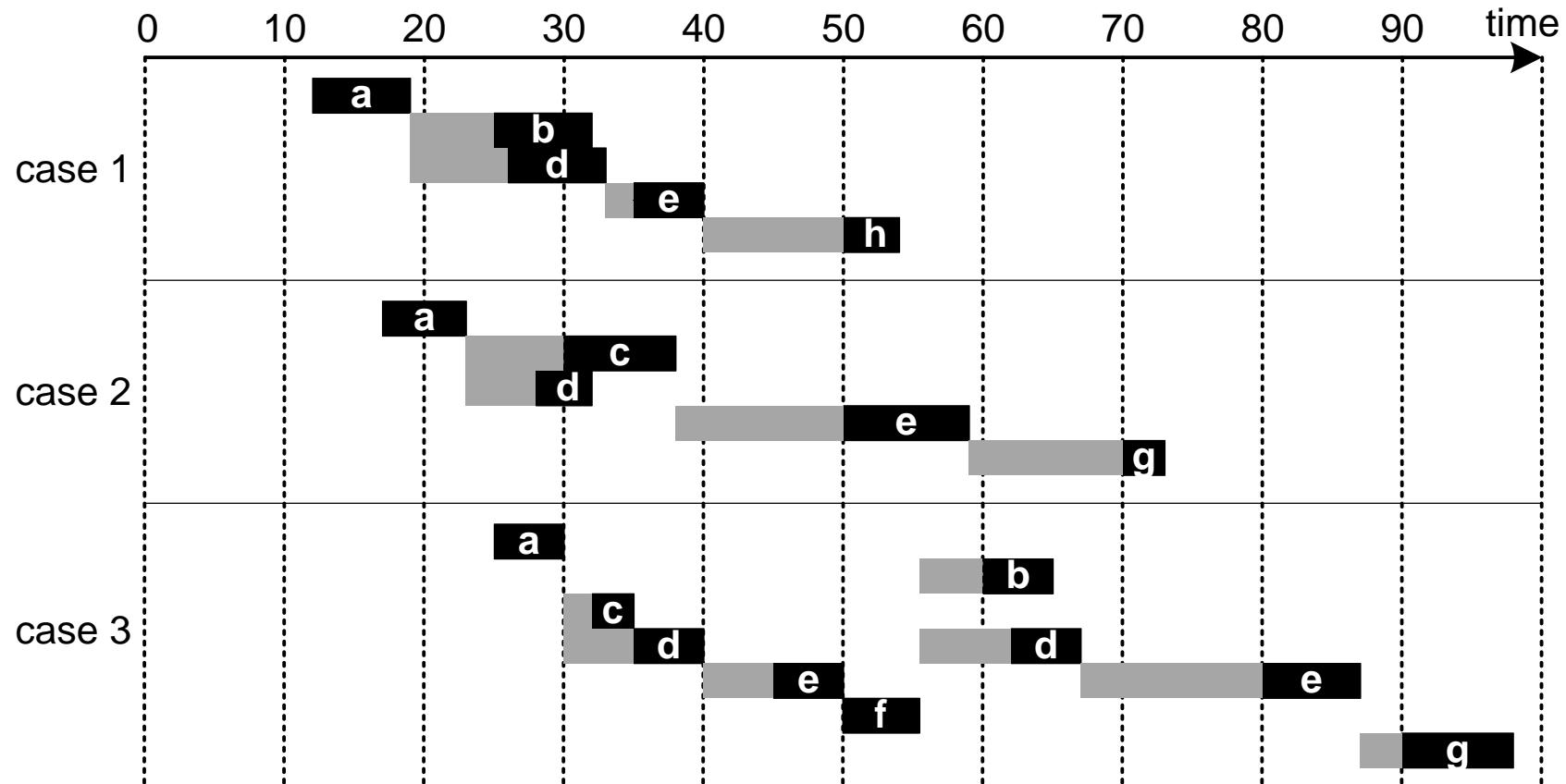




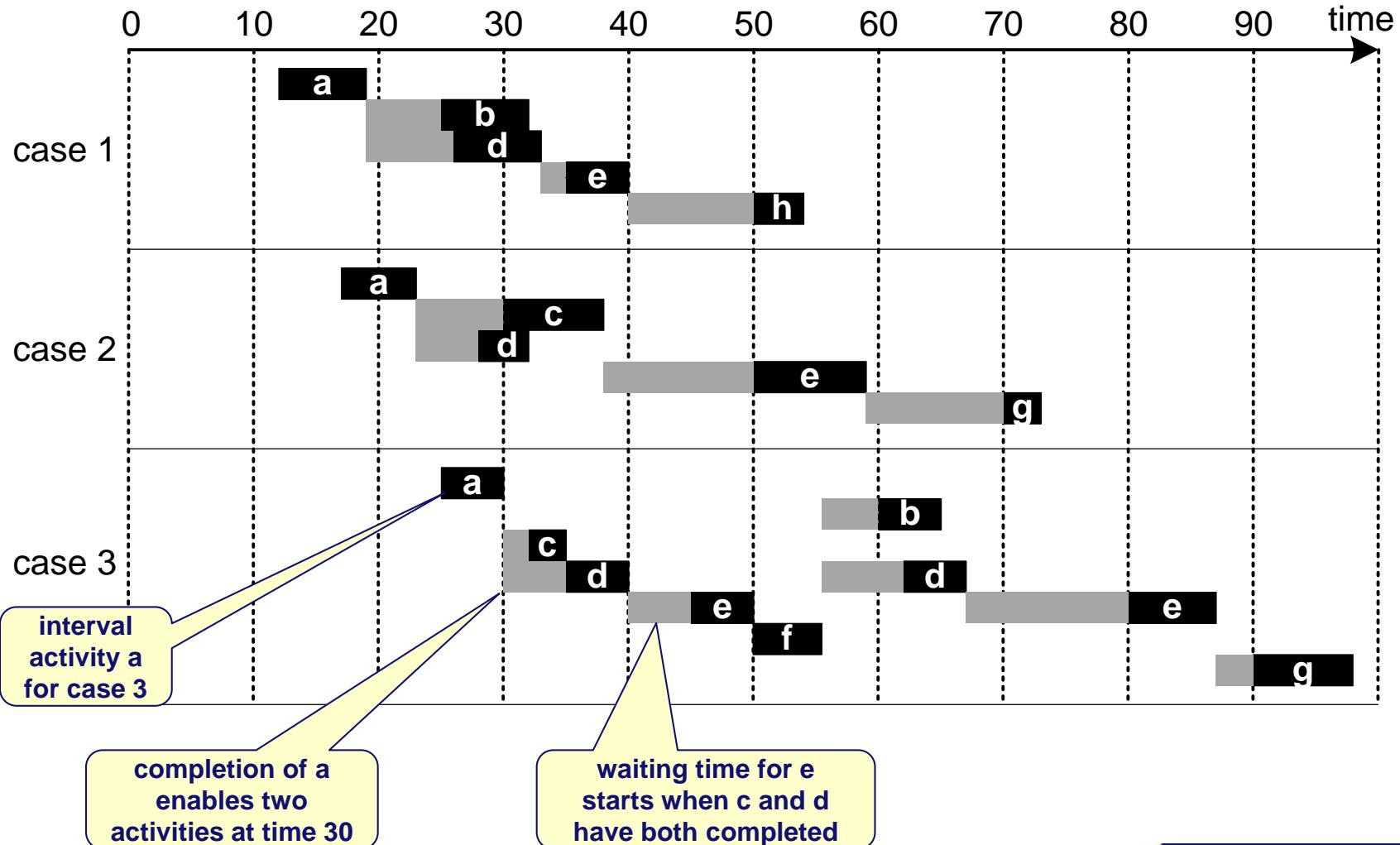




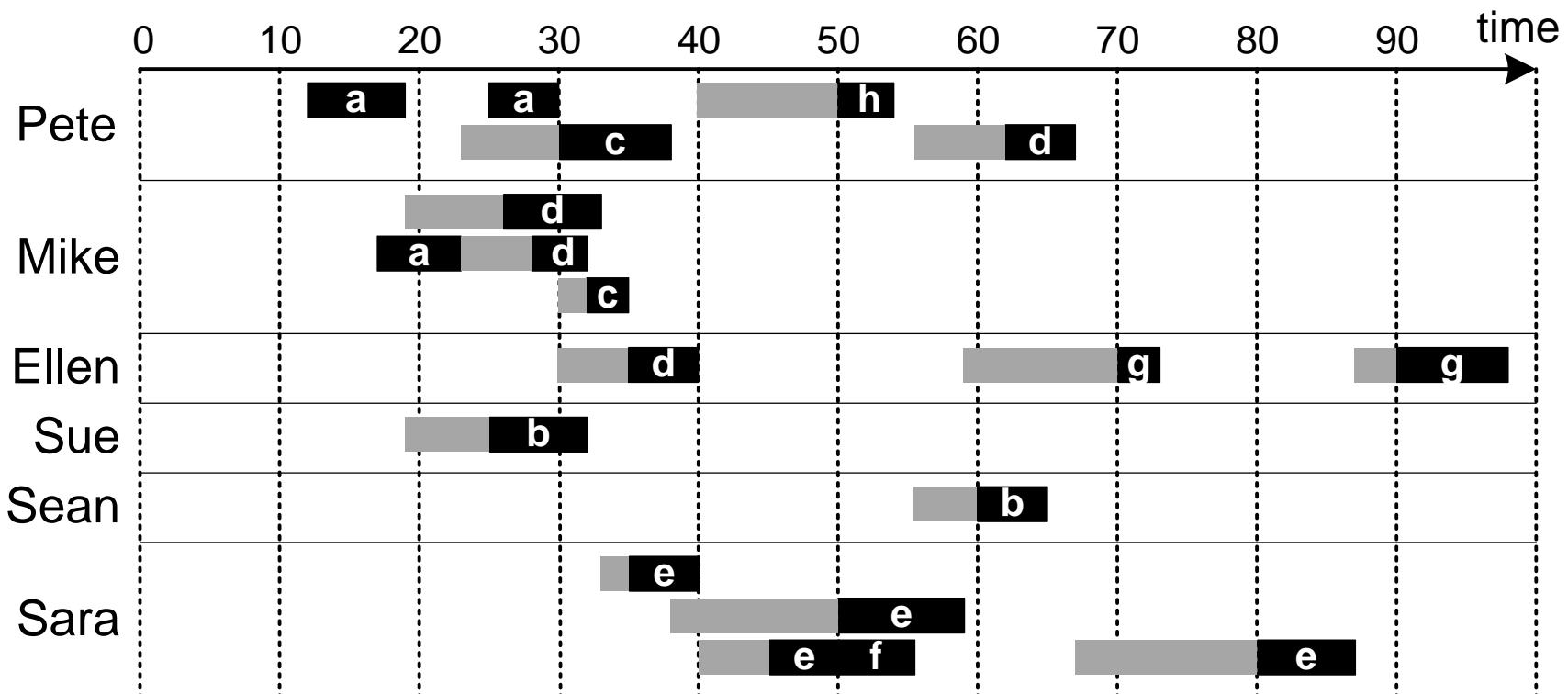
Another view on the timed replay of the first three cases



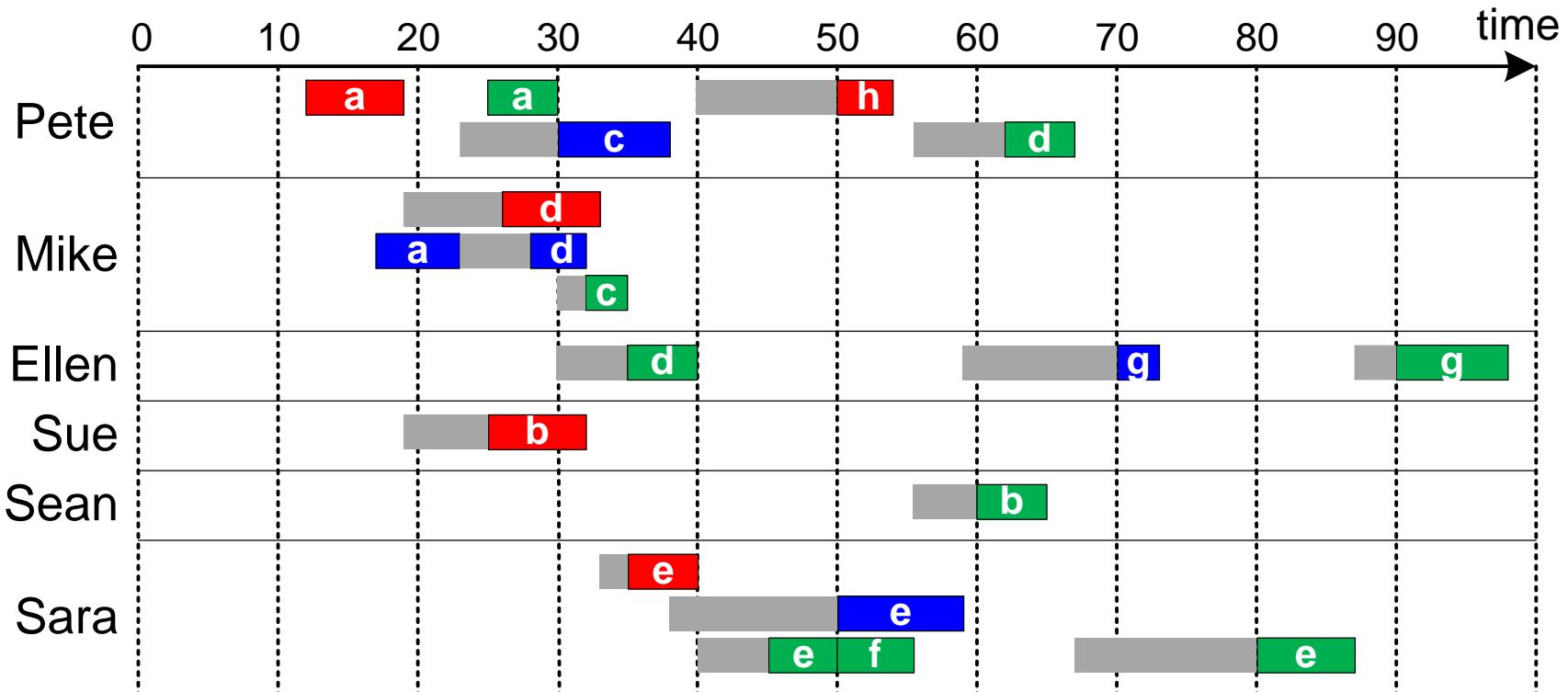
Another view on the timed replay of the first three cases



Timed replay projected onto resources

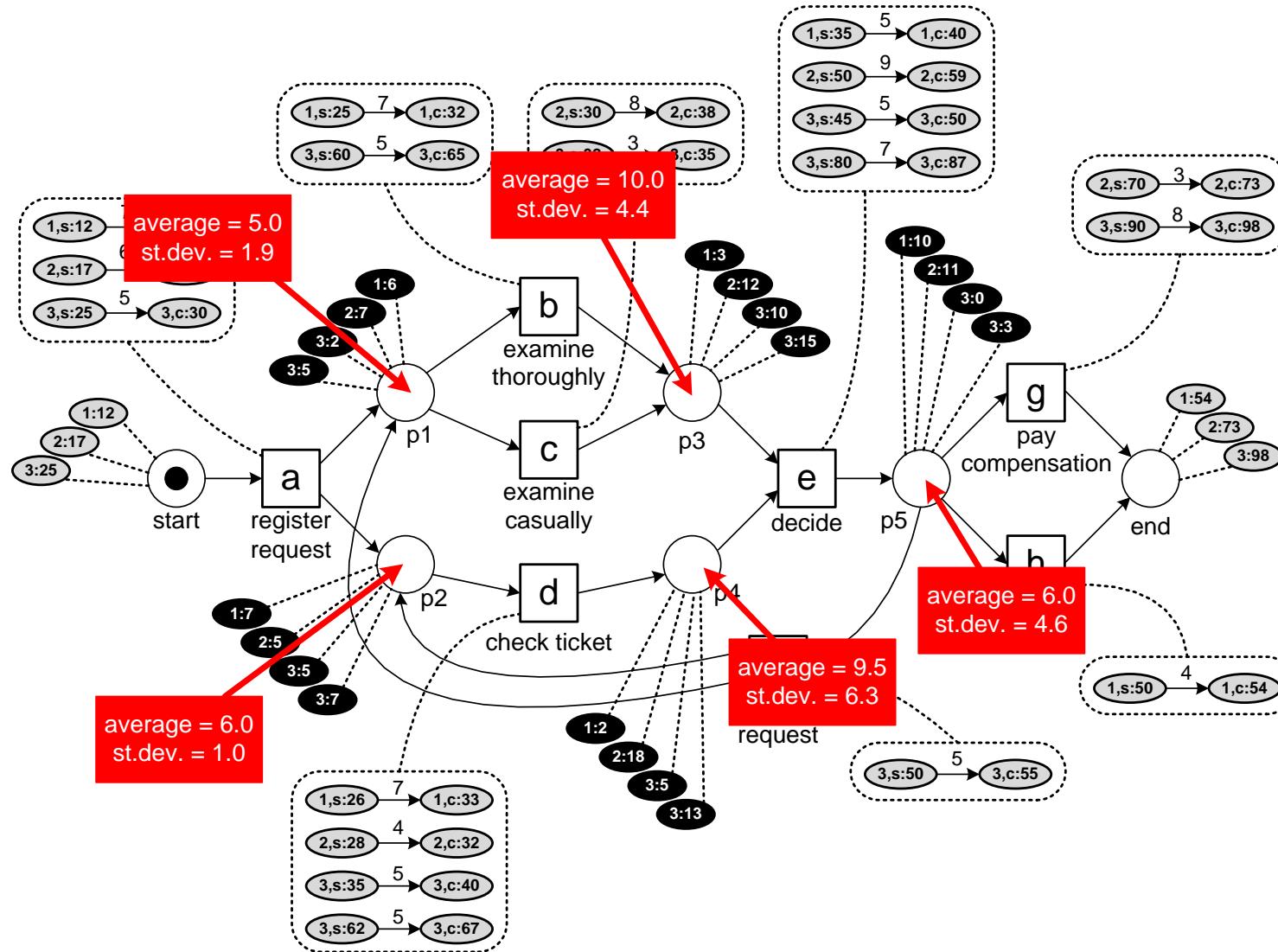


Timed replay projected onto resources (activities colored by case)

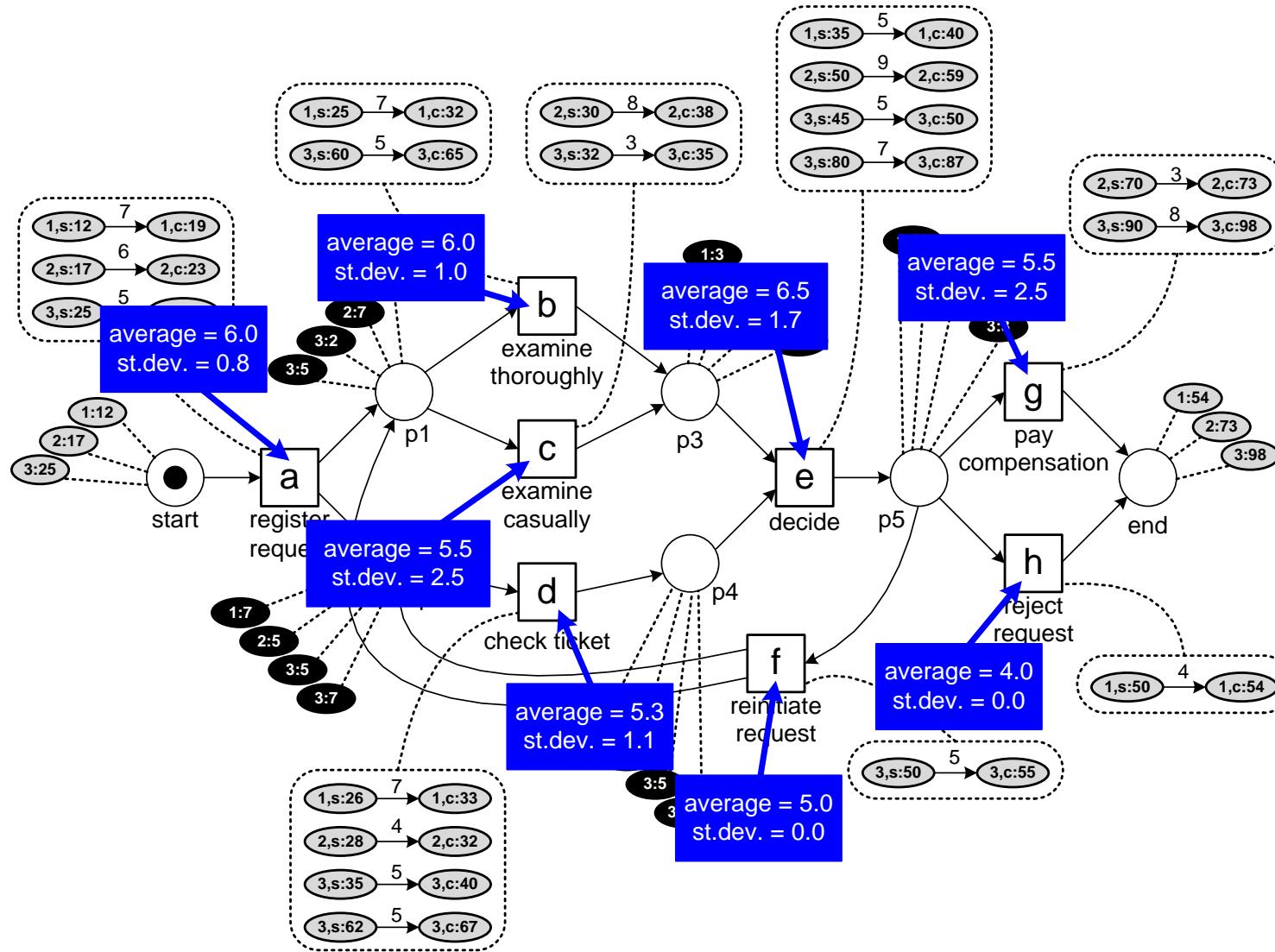


case id	trace
1	$\langle a_{start}^{12}, a_{complete}^{19}, b_{start}^{25}, d_{start}^{26}, b_{complete}^{32}, d_{complete}^{33}, e_{start}^{35}, e_{complete}^{40}, h_{start}^{50}, h_{complete}^{54} \rangle$
2	$\langle a_{start}^{17}, a_{complete}^{23}, d_{start}^{28}, c_{start}^{30}, d_{complete}^{32}, c_{complete}^{38}, e_{start}^{50}, e_{complete}^{59}, g_{start}^{70}, g_{complete}^{73} \rangle$
3	$\langle a_{start}^{25}, a_{complete}^{30}, c_{start}^{32}, c_{complete}^{35}, d_{start}^{35}, d_{complete}^{40}, e_{start}^{45}, e_{complete}^{50}, f_{start}^{50}, f_{complete}^{55}, b_{start}^{60}, d_{start}^{62}, b_{complete}^{65}, d_{complete}^{67}, e_{start}^{80}, e_{complete}^{87}, g_{start}^{90}, g_{complete}^{98} \rangle$
...	...

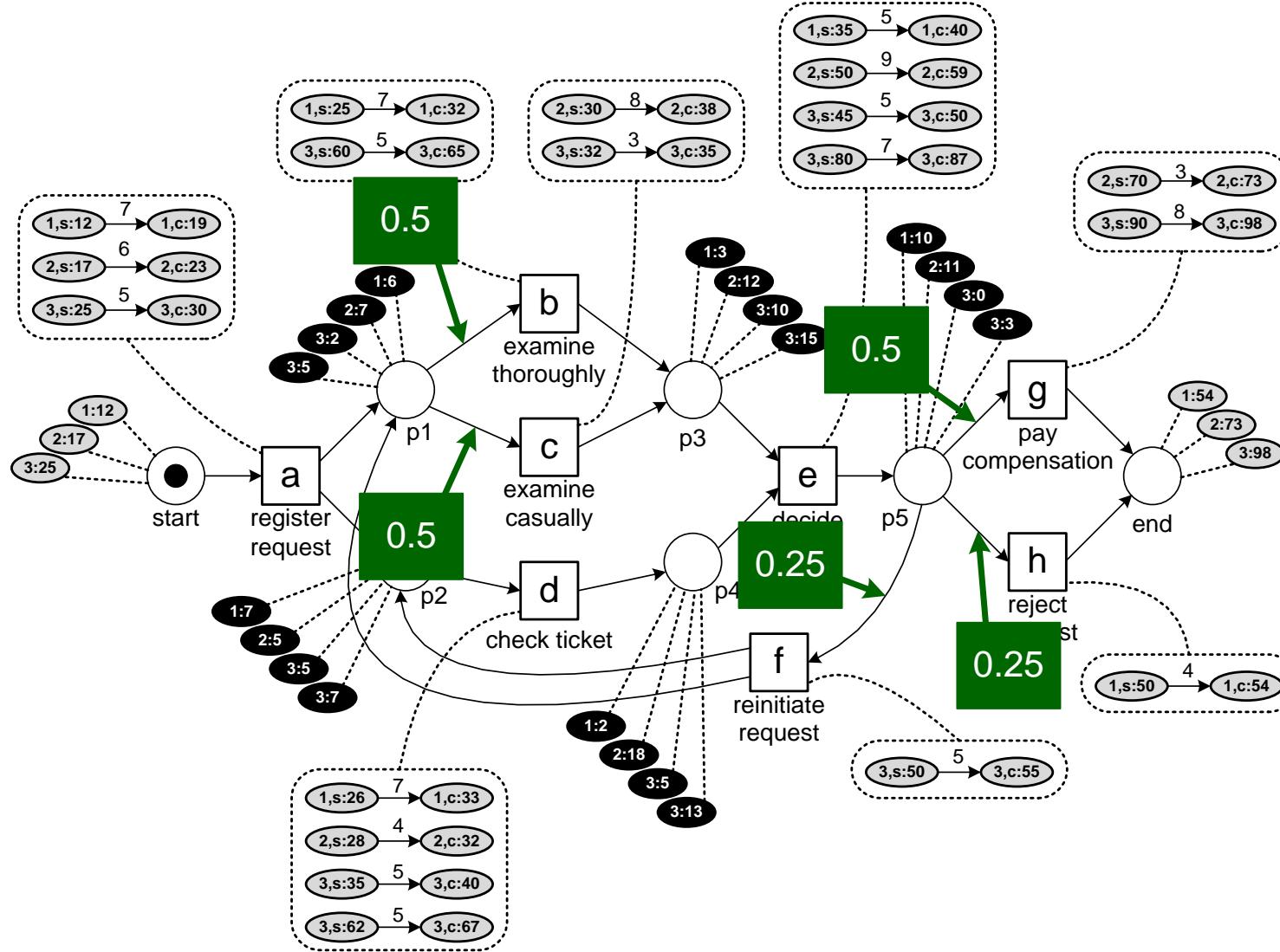
Waiting times



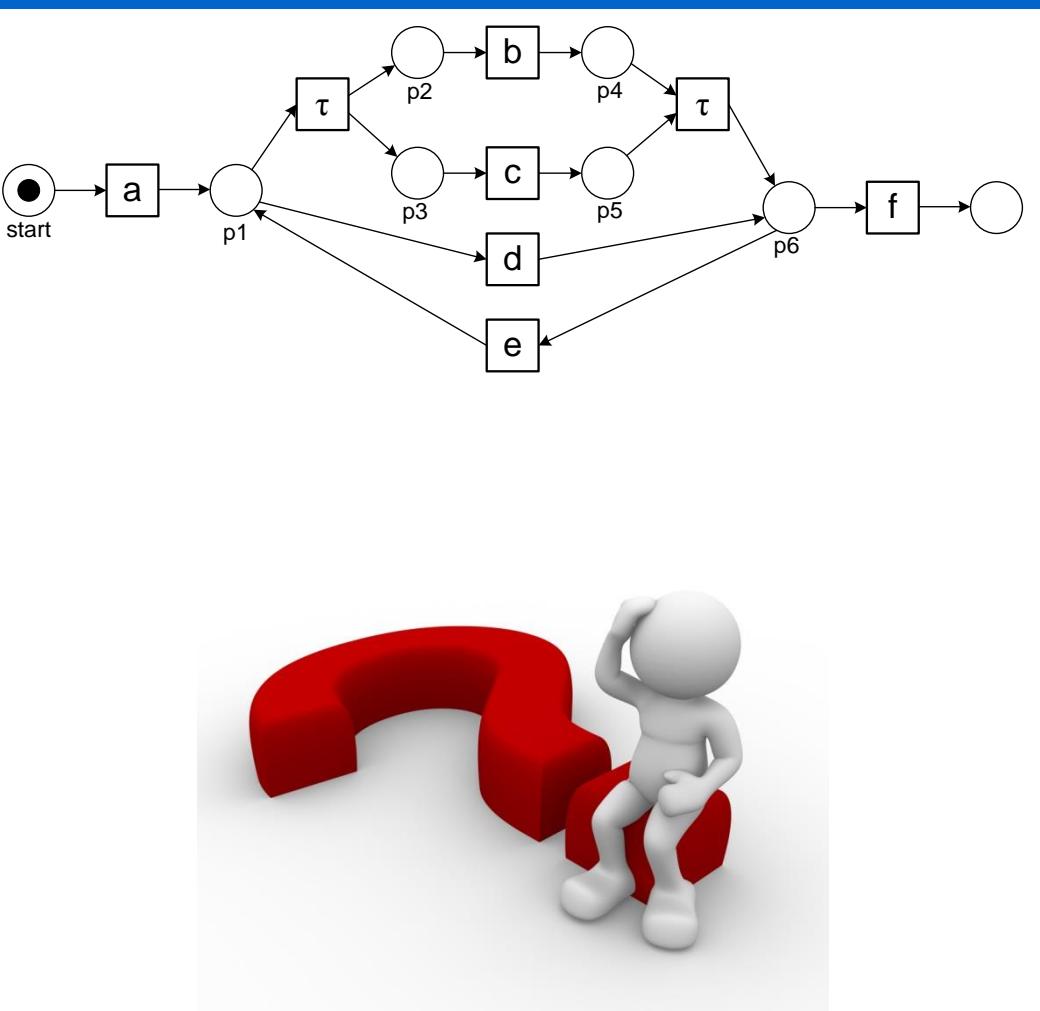
Service times



Routing probabilities

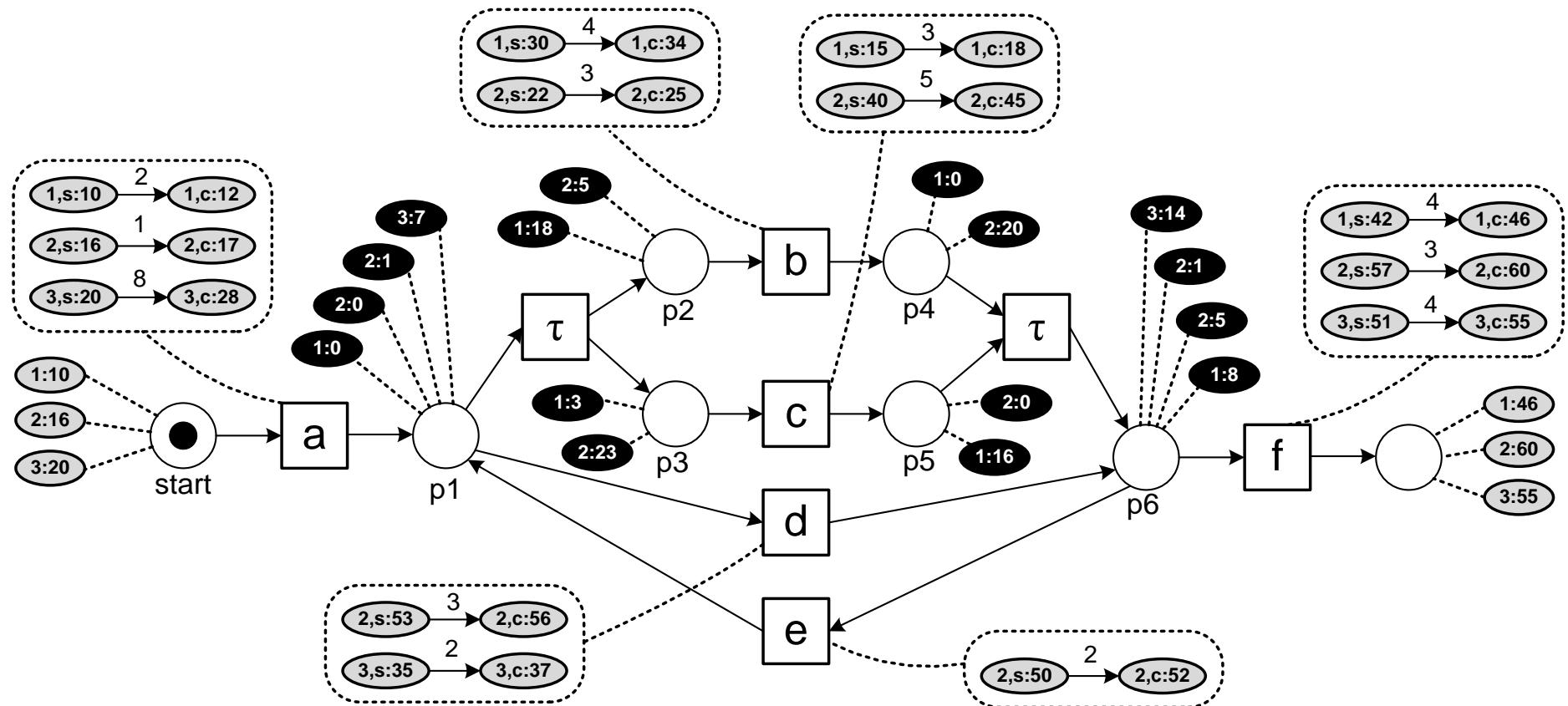


Estimate service times, waiting times, and routing probabilities



case id	activity	type	time	resource
1	a	start	10	Pete
1	a	complete	12	Pete
1	c	start	15	Sue
2	a	start	16	Pete
2	a	complete	17	Pete
1	c	complete	18	Sue
3	a	start	20	Pete
2	b	start	22	Mary
2	b	complete	25	Mary
3	a	complete	28	Pete
1	b	start	30	Mary
1	b	complete	34	Mary
3	d	start	35	Mary
3	d	complete	37	Mary
2	c	start	40	Sue
1	f	start	42	Carol
2	c	complete	45	Sue
1	f	complete	46	Carol
2	e	start	50	Kirsten
3	f	start	51	Carol
2	e	complete	52	Kirsten
2	d	start	53	Mary
3	f	complete	55	Carol
2	d	complete	56	Mary
2	f	start	57	Carol
2	f	complete	60	Carol

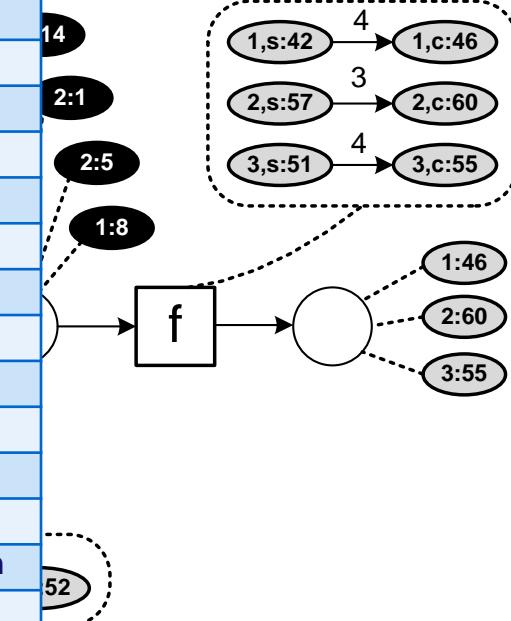
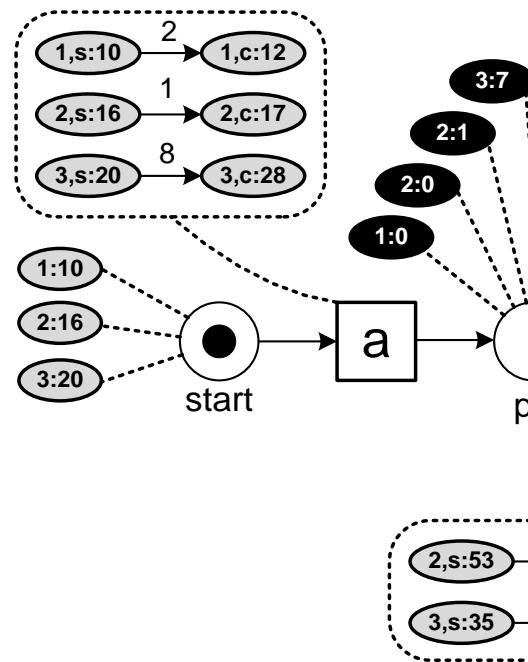
Times recorded during replay



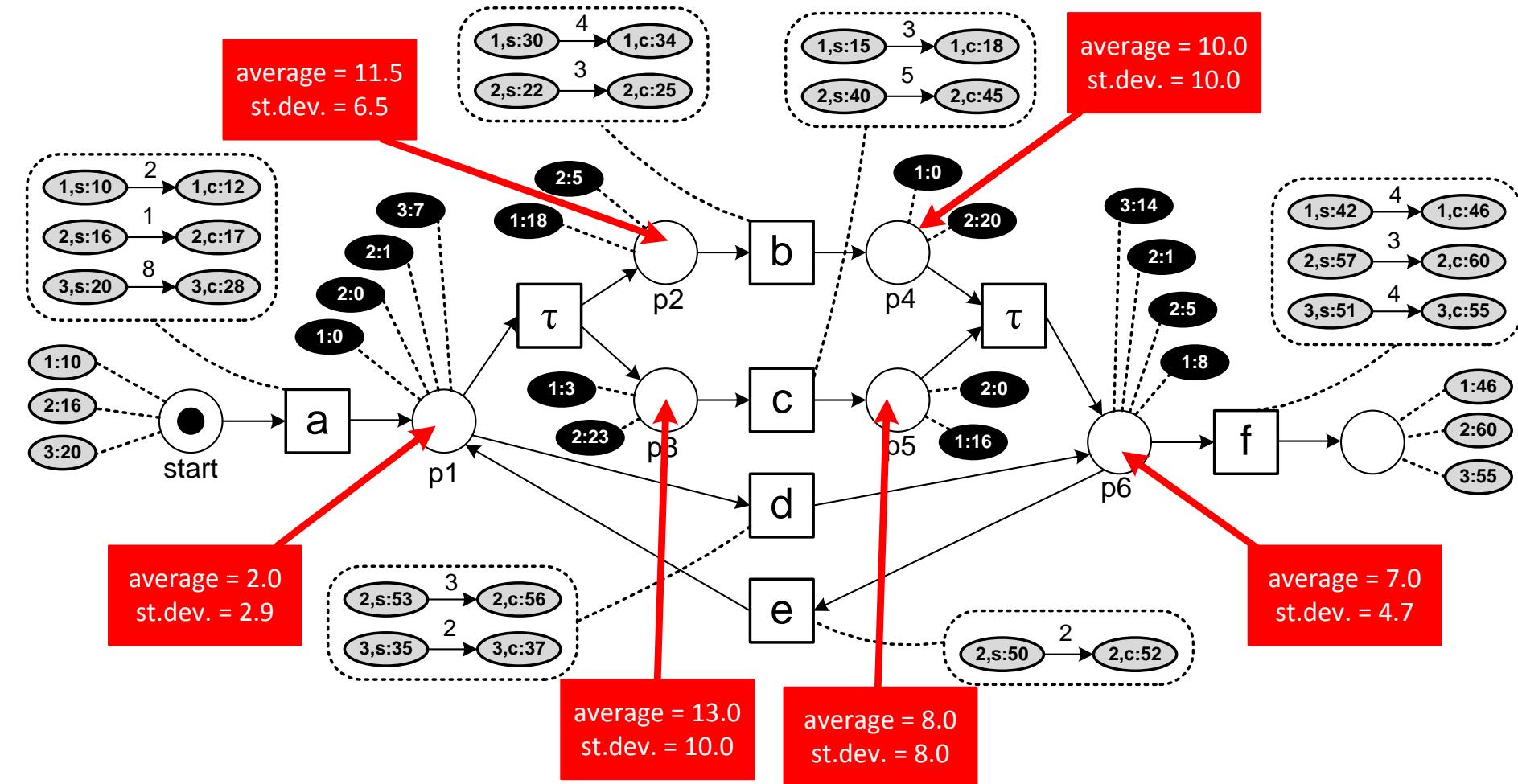
case id	activity	type	time	resource
1	a	start	10	Pete
1	a	complete	12	Pete
1	c	start	15	Sue
2	a	start	16	Pete
2	a	complete	17	Pete

Times recorded during replay

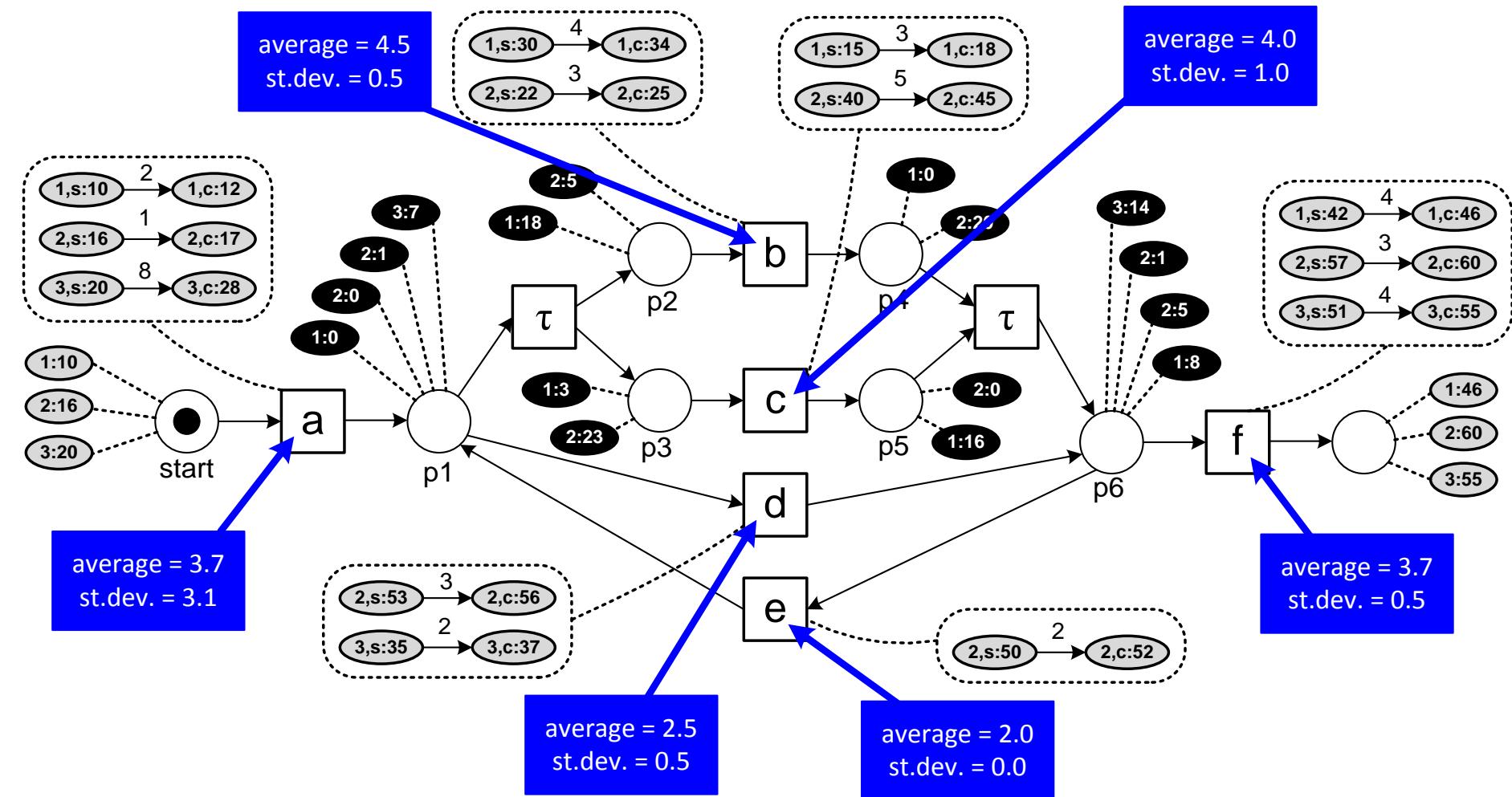
case id	activity	type	time	resource
1	a	start	10	Pete
1	a	complete	12	Pete
1	c	start	15	Sue
2	a	start	16	Pete
2	a	complete	17	Pete
1	c	complete	18	Sue
3	a	start	20	Pete
2	b	start	22	Mary
2	b	complete	25	Mary
3	a	complete	28	Pete
1	b	start	30	Mary
1	b	complete	34	Mary
3	d	start	35	Mary
3	d	complete	37	Mary
2	c	start	40	Sue
1	f	start	42	Carol
2	c	complete	45	Sue
1	f	complete	46	Carol
2	e	start	50	Kirsten
3	f	start	51	Carol
2	e	complete	52	Kirsten
2	d	start	53	Mary
3	f	complete	55	Carol
2	d	complete	56	Mary
2	f	start	57	Carol
2	f	complete	60	Carol



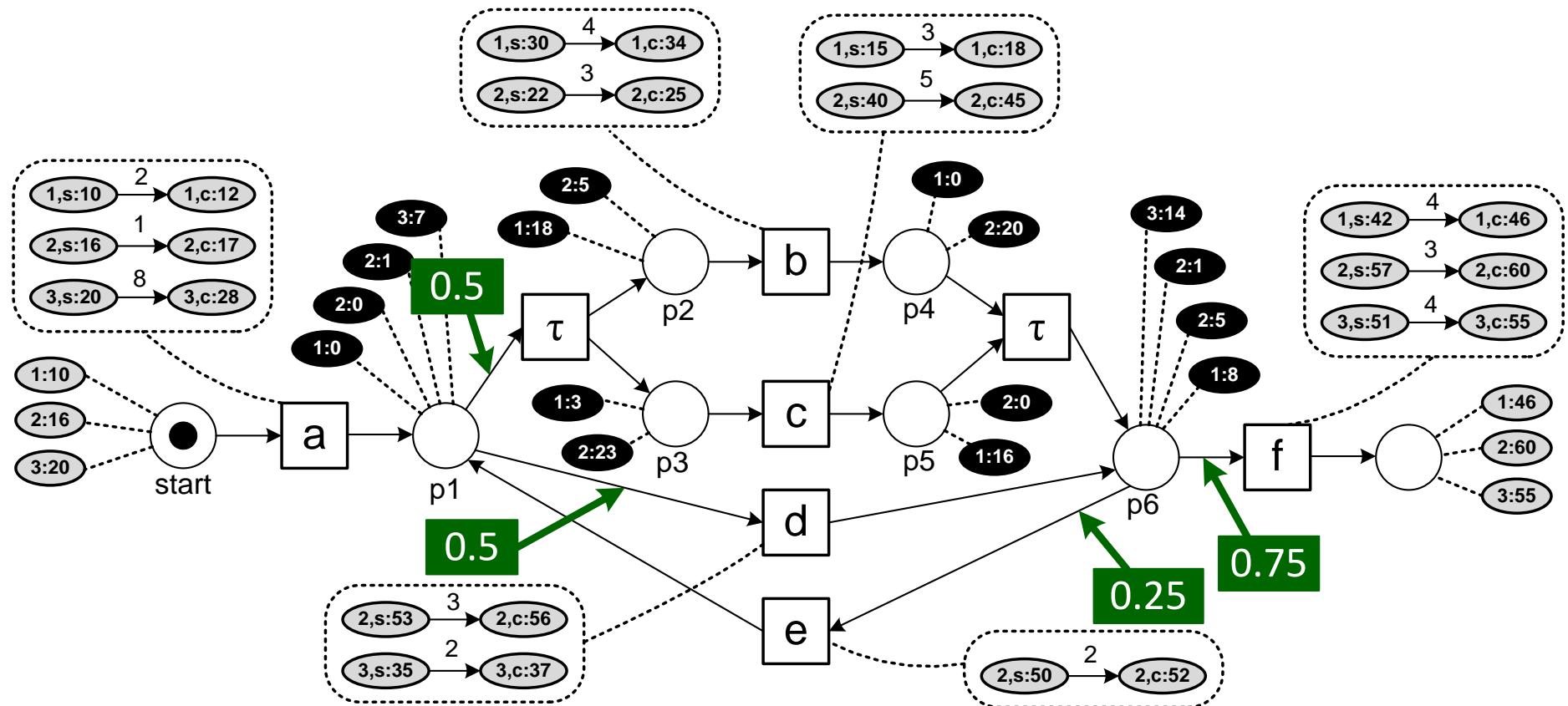
Waiting times



Service times



Routing probabilities



So we can mine for bottlenecks and other performance related properties ...



mining social
networks

Organizational mining

case id trace

1	$\langle a^{Pete}, b^{Sue}, d^{Mike}, e^{Sara}, h^{Pete} \rangle$
2	$\langle a^{Mike}, d^{Mike}, c^{Pete}, e^{Sara}, g^{Ellen} \rangle$
3	$\langle a^{Pete}, c^{Mike}, d^{Ellen}, e^{Sara}, f^{Sara}, b^{Sean}, d^{Pete}, e^{Sara}, g^{Ellen} \rangle$
4	$\langle a^{Pete}, d^{Mike}, b^{Sean}, e^{Sara}, h^{Ellen} \rangle$
5	$\langle a^{Ellen}, c^{Mike}, d^{Pete}, e^{Sara}, f^{Sara}, d^{Ellen}, c^{Mike}, e^{Sara}, f^{Sara}, b^{Sue}, d^{Pete}, e^{Sara}, h^{Mike} \rangle$
6	$\langle a^{Mike}, c^{Ellen}, d^{Mike}, e^{Sara}, g^{Mike} \rangle$
...	...

($a = \text{register request}$, $b = \text{examine thoroughly}$, $c = \text{examine casually}$, $d = \text{check ticket}$, $e = \text{decide}$, $f = \text{reinitiate request}$, $g = \text{pay compensation}$, and $h = \text{reject request}$)

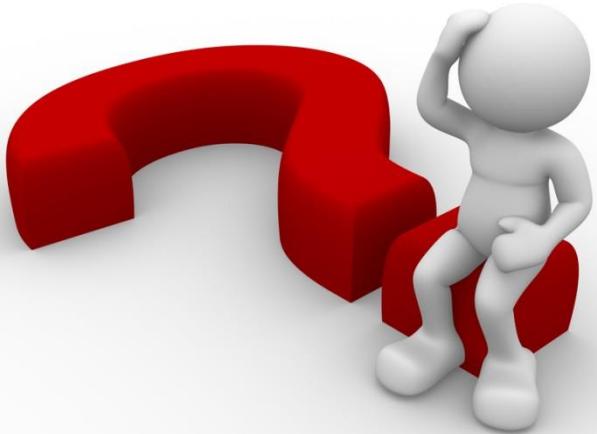
Resource-activity matrix

mean number of times a resource performs an activity per case

	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>	<i>g</i>	<i>h</i>
Pete	0.3	0	0.345	0.69	0	0	0.135	0.165
Mike	0.5	0	0.575	1.15	0	0	0.225	0.275
Ellen	0.2	0	0.23	0.46	0	0	0.09	0.11
Sue	0	0.46	0	0	0	0	0	0
Sean	0	0.69	0	0	0	0	0	0
Sara	0	0	0	0	2.3	1.3	0	0

Activity a is executed exactly once for each case (hence the sum of the first column is 1). Pete, Mike, and Ellen are the only ones executing this activity. In 30% of the cases, a is executed by Pete, 50% is executed by Mike, and 20% is executed by Ellen. Activities e and f are always executed by Sara. Activity e is executed, on average, 2.3 times per case. Etc.

Create resource-activity matrix



case id	activity	type	time	resource
1	a	start	10	Pete
1	a	complete	12	Pete
1	c	start	15	Sue
2	a	start	16	Pete
2	a	complete	17	Pete
1	c	complete	18	Sue
3	a	start	20	Pete
2	b	start	22	Mary
2	b	complete	25	Mary
3	a	complete	28	Pete
1	b	start	30	Mary
1	b	complete	34	Mary
3	d	start	35	Mary
3	d	complete	37	Mary
2	c	start	40	Sue
1	f	start	42	Carol
2	c	complete	45	Sue
1	f	complete	46	Carol
2	e	start	50	Kirsten
3	f	start	51	Carol
2	e	complete	52	Kirsten
2	d	start	53	Mary
3	f	complete	55	Carol
2	d	complete	56	Mary
2	f	start	57	Carol
2	f	complete	60	Carol

Resource-activity matrix

case id	activity	type	time	resource							
					a	b	c	d	e	f	
1	a	start	10	Pete							
1	a	complete	12	Pete							
1	c	start	15	Sue							
2	a	start	16								
2	a	complete	17								
1	c	complete	18	Pete	1.00	0.00	0.00	0.00	0.00	0.00	
3	a	start	20								
2	b	start	22	Mary	0.00	0.67	0.00	0.67	0.00	0.00	
2	b	complete	25								
3	a	complete	28	Sue	0.00	0.00	0.67	0.00	0.00	0.00	
1	b	start	30								
1	b	complete	34	Kirsten	0.00	0.00	0.00	0.00	0.33	0.00	
3	d	start	35	Carol	0.00	0.00	0.00	0.00	0.00	1.00	
3	d	complete	37	Mary							
2	c	start	40	Sue							
1	f	start	42	Carol							
2	c	complete	45	Sue							
1	f	complete	46	Carol							
2	e	start	50	Kirsten							
3	f	start	51	Carol							
2	e	complete	52	Kirsten							
2	d	start	53	Mary							
3	f	complete	55	Carol							
2	d	complete	56	Mary							
2	f	start	57	Carol							
2	f	complete	60	Carol							

mean number of times a resource performs an activity per case

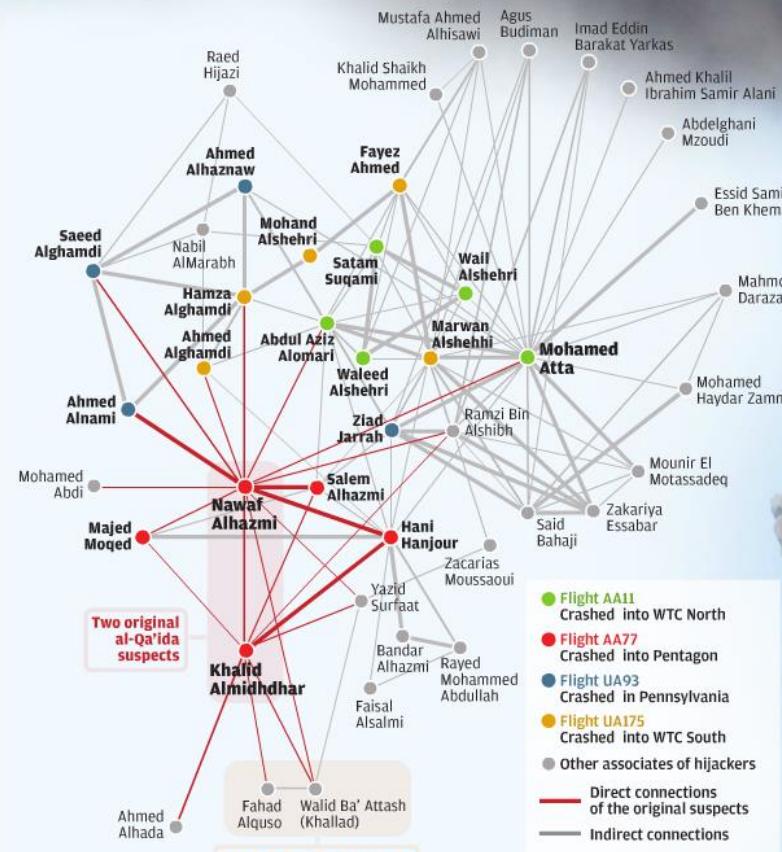
Social network analysis

How two names and a sheaf of newspaper cuttings revealed the 9/11 team

This social network of the 19 hijackers behind the 9/11 attacks in the United States, and their associates, was drawn up at the end of 2001. Valdis Krebs, a commercial consultant in network analysis, started with newspaper reports of the two original terrorist suspects, Nawaf Alhazmi and Khalid Almihdhar. He then plotted the position of the other hijackers and associates. His analysis highlighted the central role played by Mohamed Atta. It also shows the close associations between the "Hamburg cell" that Atta set up, as well as the close links with the two original suspects – critical information that may have helped to avert an attack had it been known.

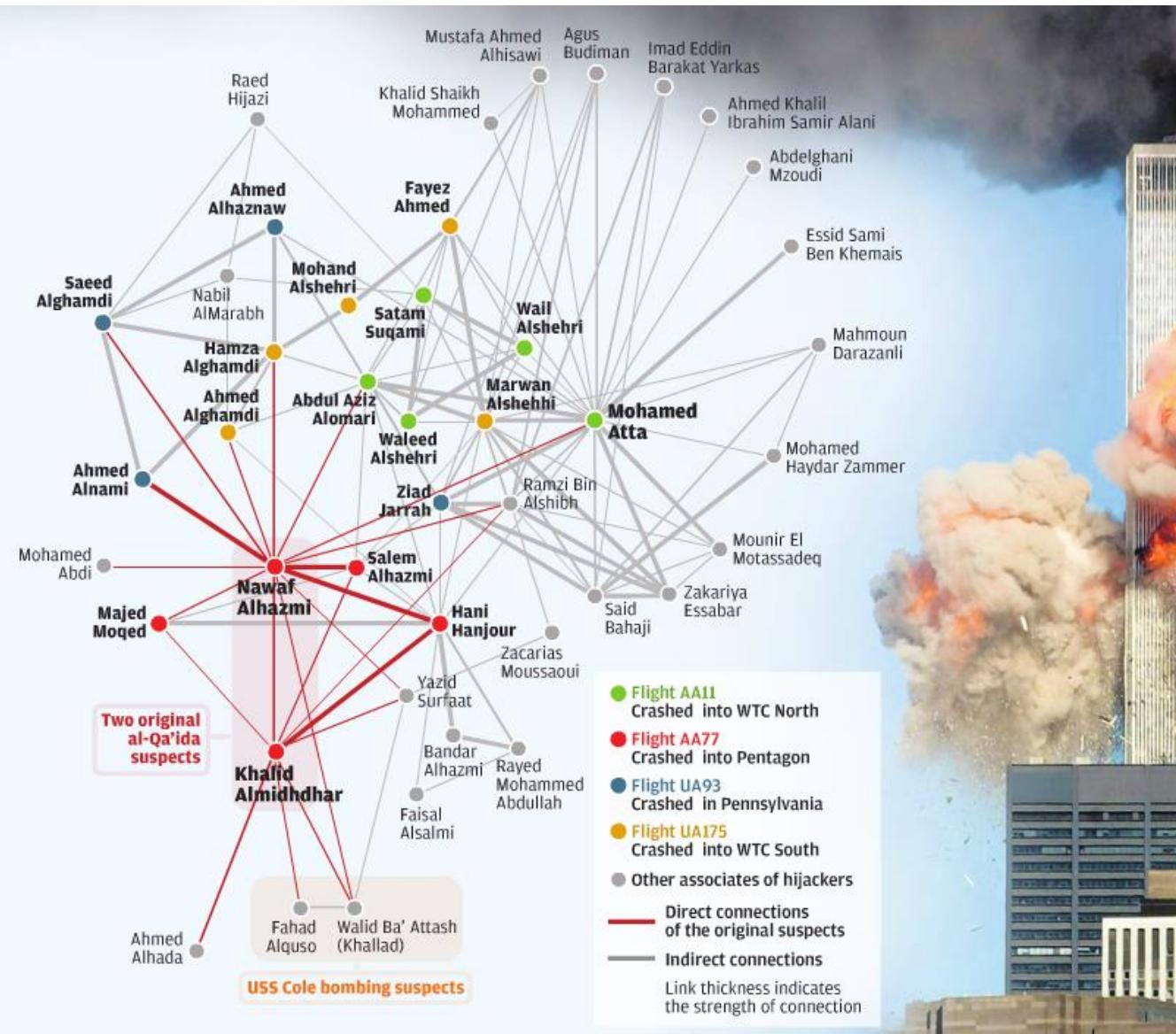


Emergency services attend the scene after Flight AA77 crashes into the Pentagon



The second plane, Flight UA175, crashes into the South Tower of the World Trade Centre

Social network analysis



- **Sociometry:** present data on interpersonal relationships in graph or matrix form.
- Jacob Levy Moreno used such techniques in the 1930s to better assign students to residential cottages.
- Arcs: weights or (inverted) distance.
- Metrics to denote importance:
 - centrality,
 - closeness,
 - betweenness,
 - ...
- Identification of cliques.

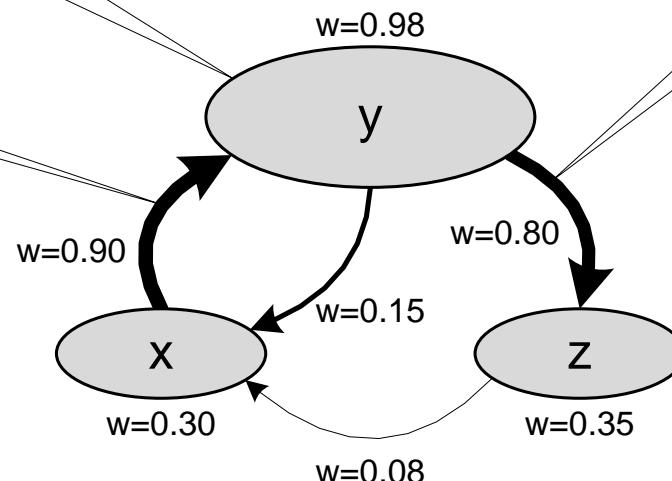
Social network

organizational entity (resource, person, role, department, etc.)

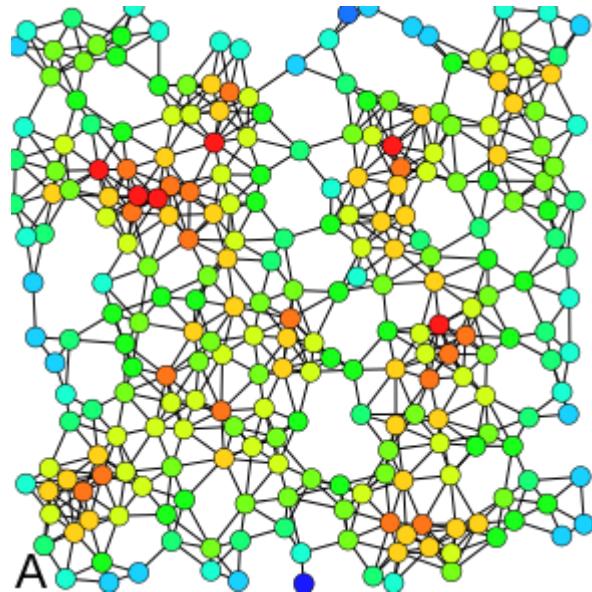
the thickness of the arc indicates the weight of the relationship

relationship

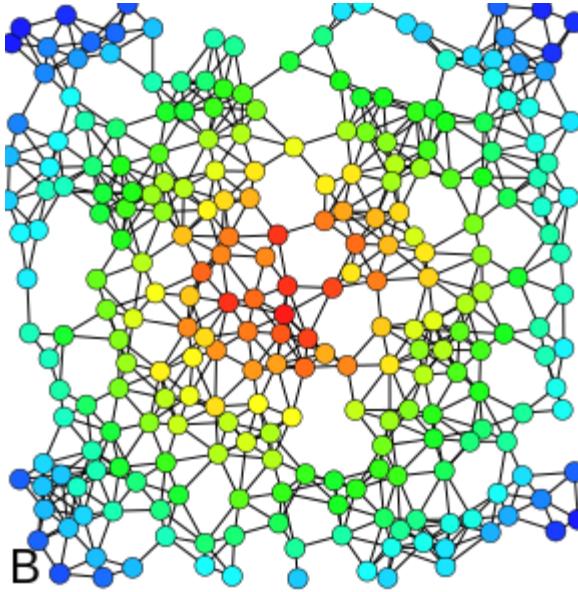
the size of the oval indicates the weight of the entity



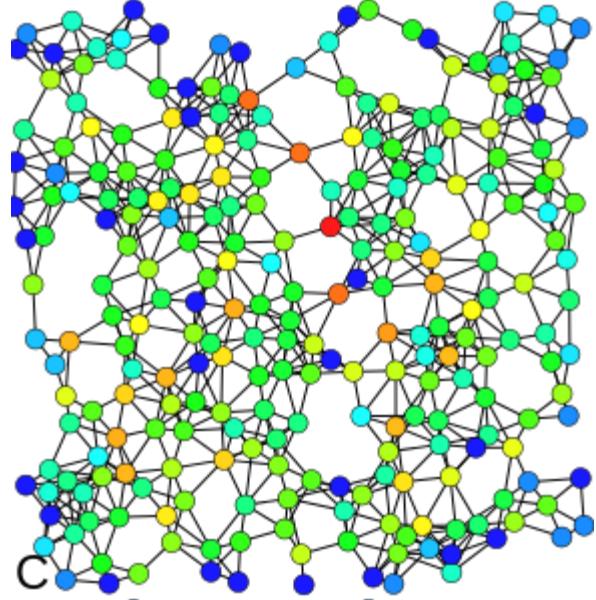
"Importance" of nodes in a social network



degree centrality:
number of connections
a particular node has



closeness centrality:
1 divided by the sum of
all shortest paths to a
particular node



betweenness centrality:
fraction of shortest
paths between any two
nodes passing a
particular node

A wide variety of definitions exist for "importance".

Figures by Claudio Rocchini

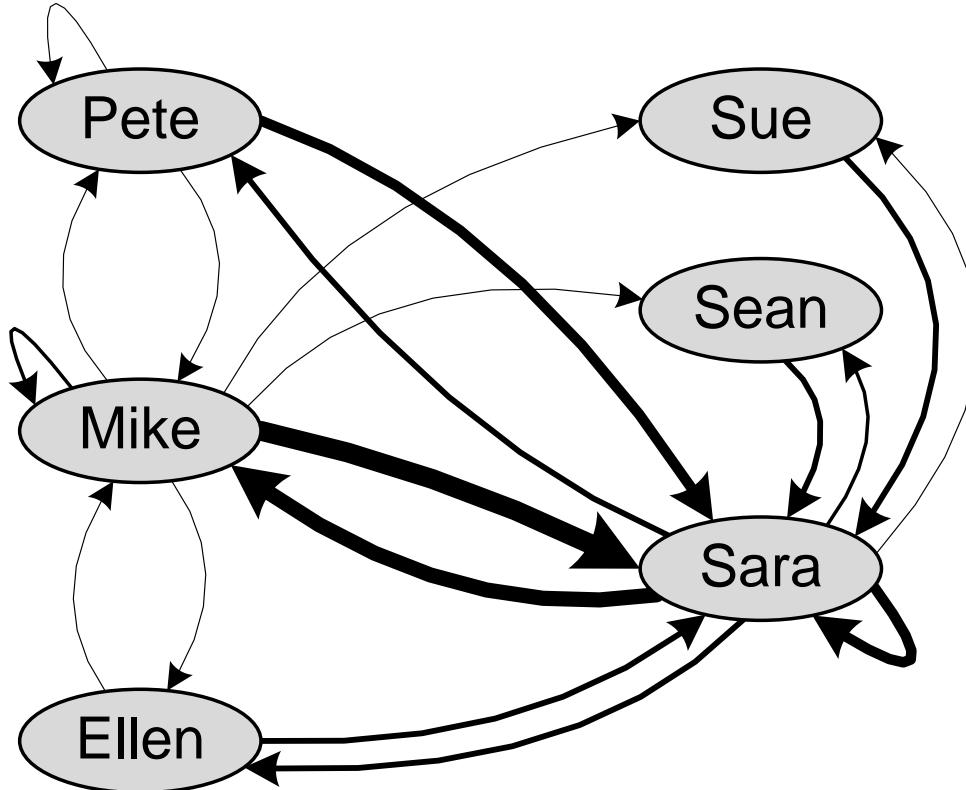
Handover of work matrix

	Pete	Mike	Ellen	Sue	Sean	Sara
Pete	0.135	0.225	0.09	0.06	0.09	1.035
Mike	0.225	0.375	0.15	0.1	0.15	1.725
Ellen	0.09	0.15	0.06	0.04	0.06	0.69
Sue	0	0	0	0	0	0.46
Sean	0	0	0	0	0	0.69
Sara	0.885	1.475	0.59	0.26	0.39	1.3

Count the number of times
work is handed over from one
resource to another (on
average per case).

The causal dependencies in
the process model are used
to count handovers in the
event log.

Social network based on handover of work (threshold of 0.1)

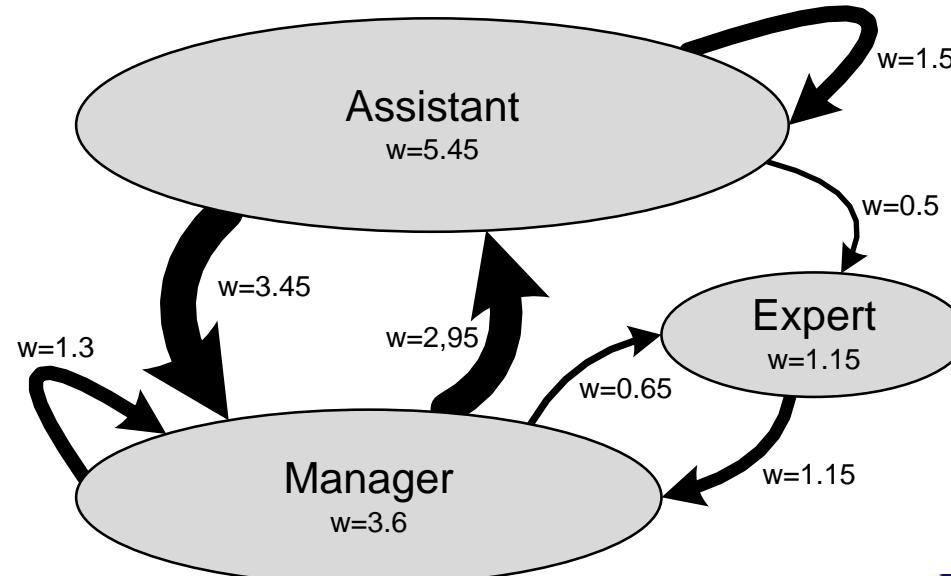


In this figure only the thickness of the arcs is based on frequencies.

	Pete	Mike	Ellen	Sue	Sean	Sara
Pete	0.135	0.225	0.09	0.06	0.09	1.035
Mike	0.225	0.375	0.15	0.1	0.15	1.725
Ellen	0.09	0.15	0.06	0.04	0.06	0.69
Sue	0	0	0	0	0	0.46
Sean	0	0	0	0	0	0.69
Sara	0.885	1.475	0.59	0.26	0.39	1.3

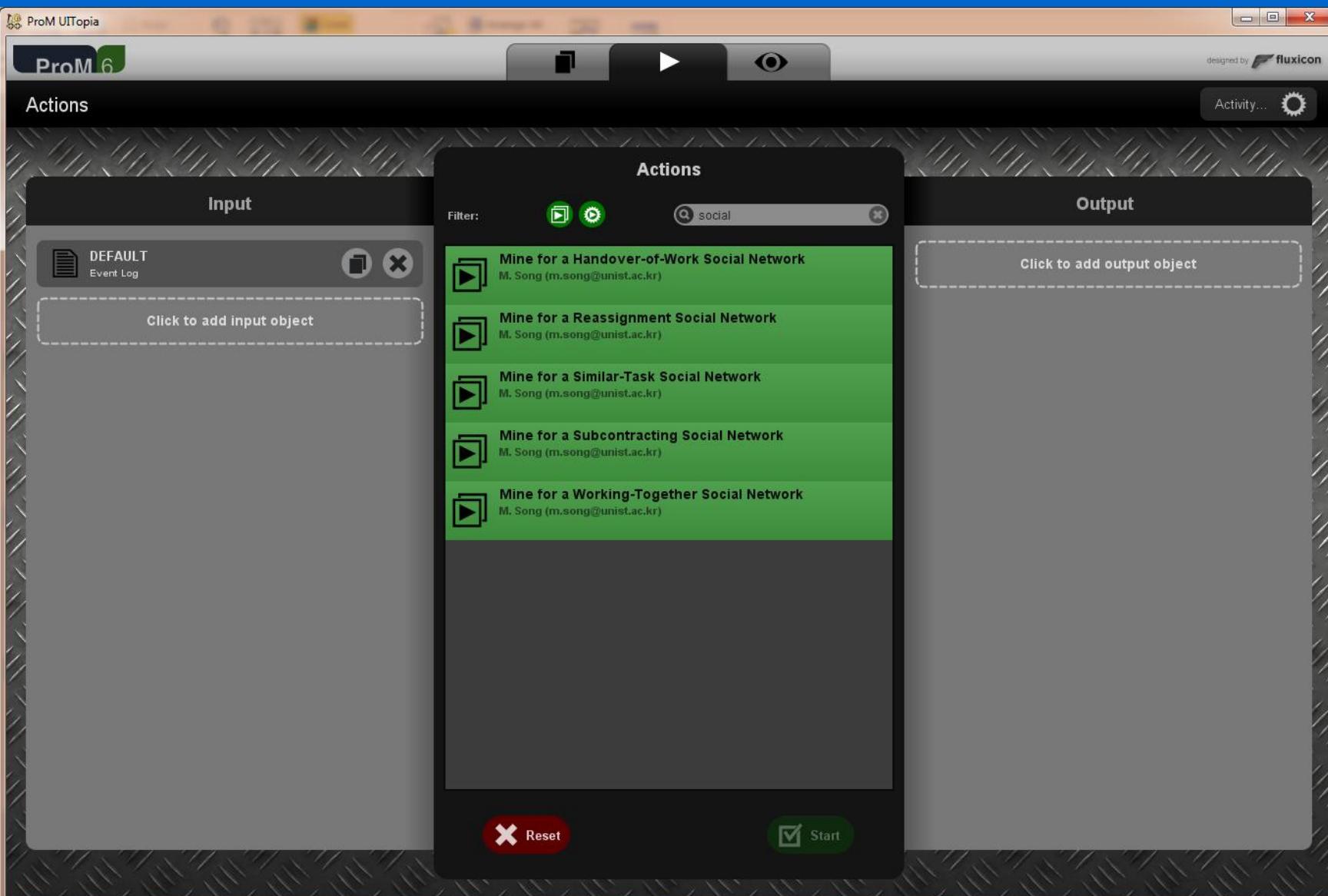
Handover of work at role level

	Assistant	Expert	Manager
Assistant	1.5	0.5	3.45
Expert	0	0	1.15
Manager	2.95	0.65	1.3

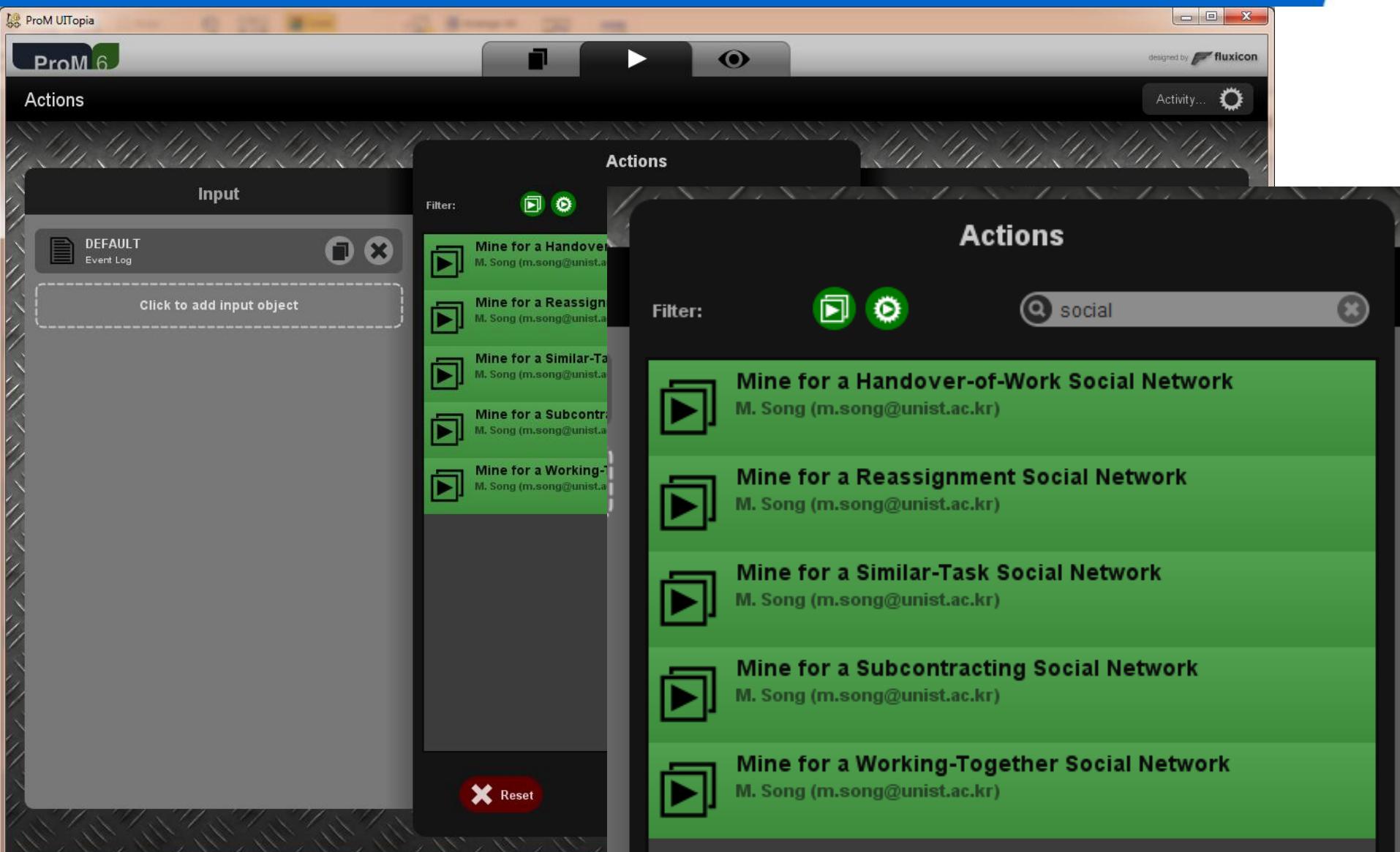


In this figure also the size of each node is based on frequencies.

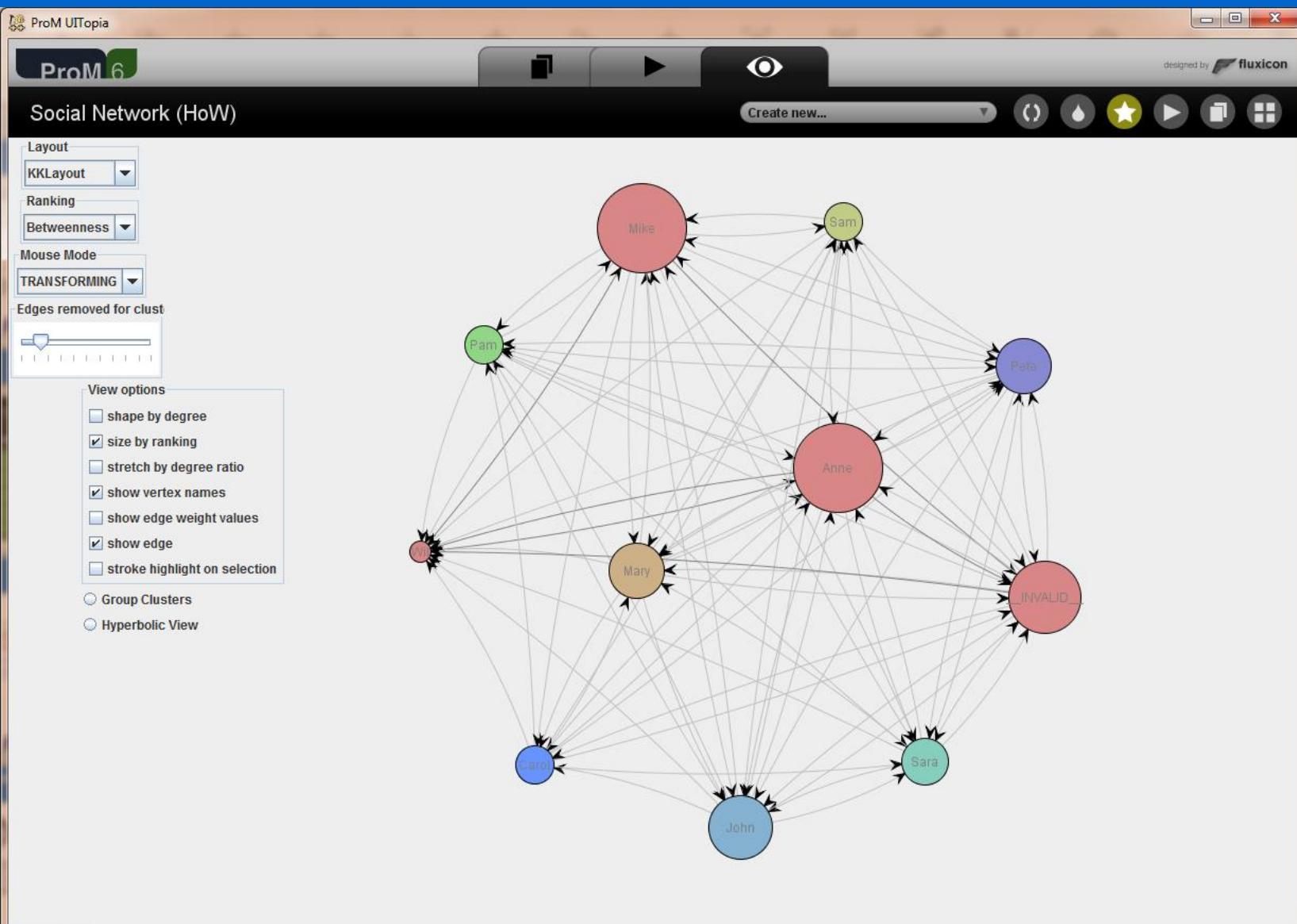
Social network miner in ProM



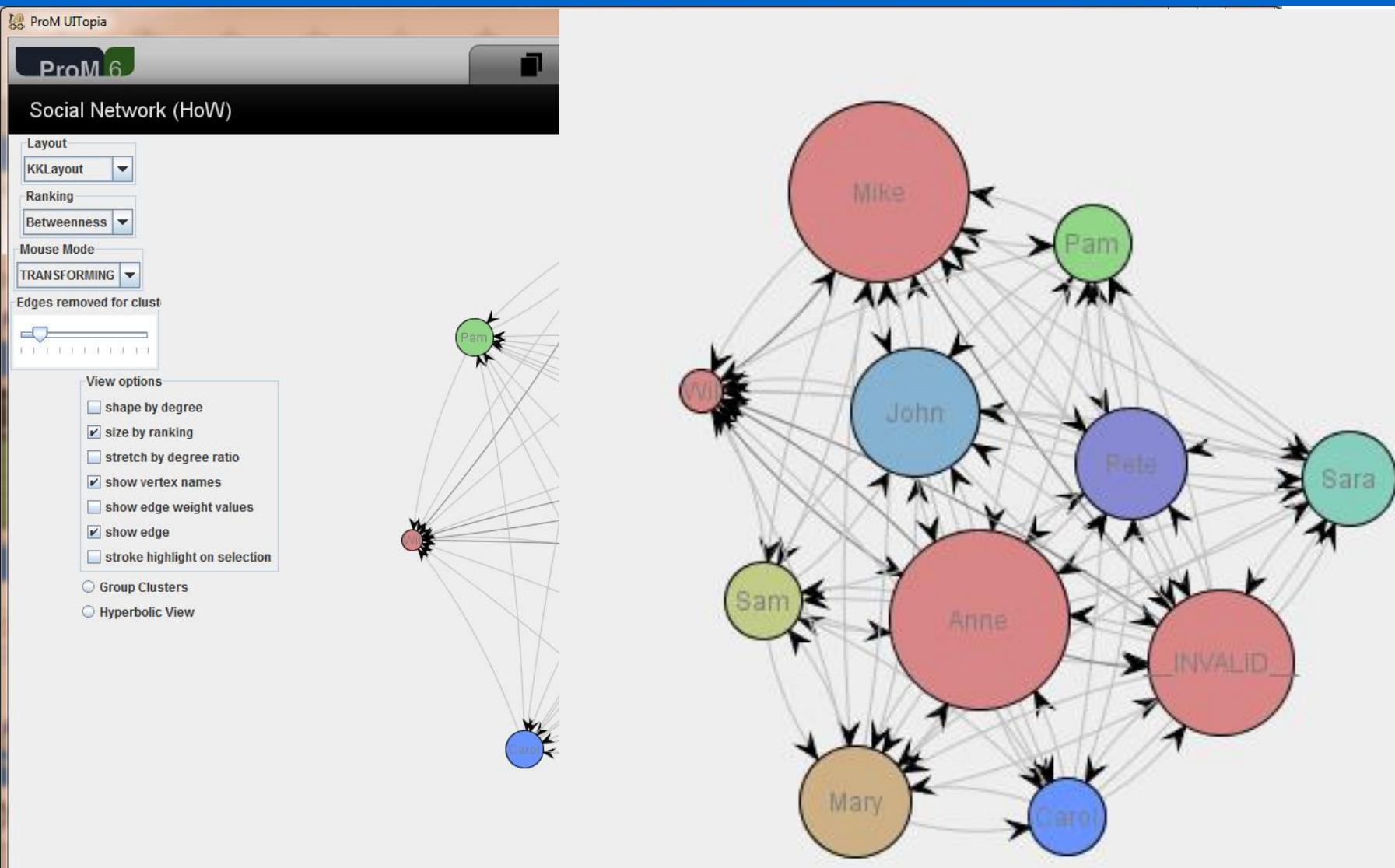
Social network miner in ProM



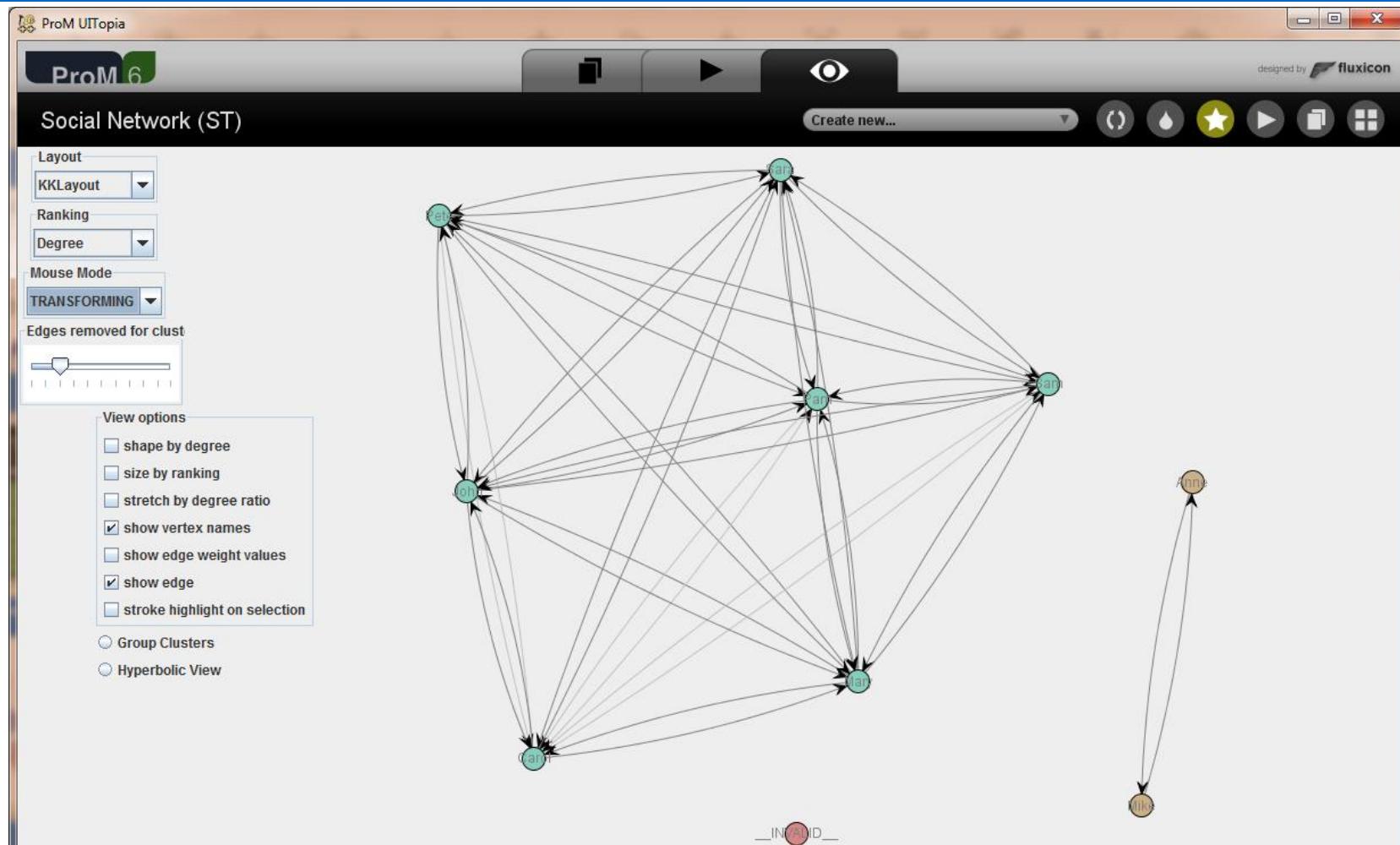
Social network based on hand-over of work



Social network based on hand-over of work



Social network based on similar tasks



Create handover of work matrix

case id	activity	type	time	resource
1	a	start	10	Pete
1	a	complete	12	Pete
1	c	start	15	Sue
2	a	start	16	Pete
2	a	complete	17	Pete
1	c	complete	18	Sue
3	a	start	20	Pete
2	b	start	22	Mary
2	b	complete	25	Mary
3	a	complete	28	Pete
1	b	start	30	Mary
1	b	complete	34	Mary
3	d	start	35	Mary
3	d	complete	37	Mary
2	c	start	40	Sue
1	f	start	42	Carol
2	c	complete	45	Sue
1	f	complete	46	Carol
2	e	start	50	Kirsten
3	f	start	51	Carol
2	e	complete	52	Kirsten
2	d	start	53	Mary
3	f	complete	55	Carol
2	d	complete	56	Mary
2	f	start	57	Carol
2	f	complete	60	Carol



- Count the number of times work is handed over from one resource to another (on average per case).
- Ignore process model, no need to consider causalities and concurrency.

Handover of work matrix

case id	activity	type	time	resource
1	a	start	10	Pete
1	a	complete	12	Pete
1	c	start	15	Sue
2	a	start	16	Pete
2	a	complete	17	Pete
1	c	complete	18	Sue
3	a	start	20	Pete
2	b	start	22	Mary
2	b	complete	25	Mary
3	a	complete	28	Pete
1	b	start	30	Mary
1	b	complete	34	Mary
3	d	start	35	Mary
3	d	complete	37	Mary
2	c	start	40	Sue
1	f	start	42	Carol
2	c	complete	45	Sue
1	f	complete	46	Carol
2	e	start	50	Kirsten
3	f	start	51	Carol
2	e	complete	52	Kirsten
2	d	start	53	Mary
3	f	complete	55	Carol
2	d	complete	56	Mary
2	f	start	57	Carol
2	f	complete	60	Carol

- 1. Pete, Sue, Mary, Carol**
- 2. Pete, Mary, Sue, Kirsten, Mary, Carol**
- 3. Pete, Mary, Carol**

Handover of work matrix

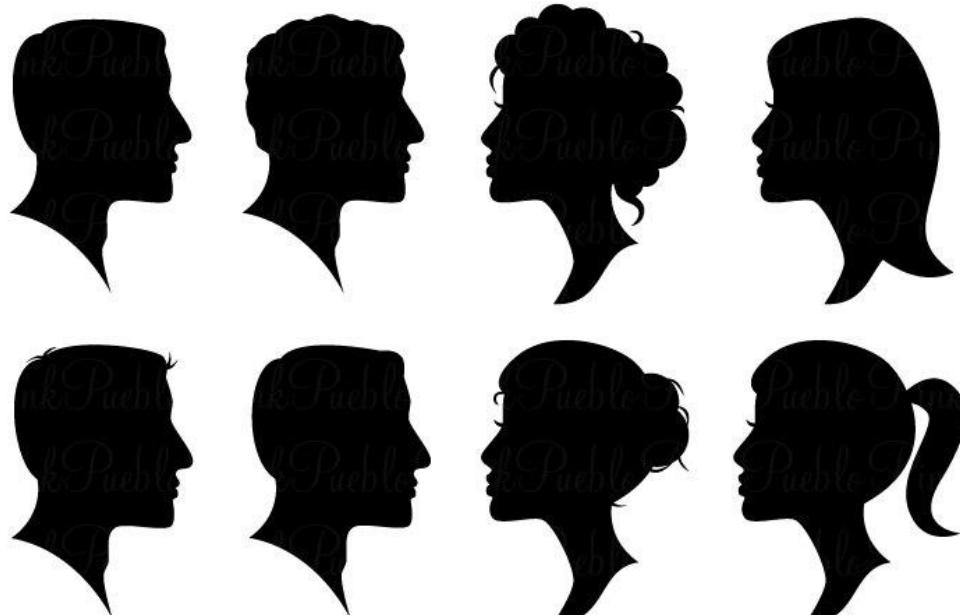
case id	activity	type	time	resource
1	a	start	10	Pete
1	a	complete	12	Pete
1	c	start	15	Sue
2	a	start	16	Pete
2	a	complete	17	Pete
1	c	complete	18	Sue
3	a	start	20	Pete
2	b	start	22	Mary
2	b	complete	25	Mary
3	a	complete	28	Pete
1	b	start	30	Mary
1	b	complete	34	Mary
3	d	start	35	Mary
3	d	complete	37	Mary
2	c	start	40	Sue
1	f	start	42	Carol
2	c	complete	45	Sue
1	f	complete	46	Carol
2	e	start	50	Kirsten
3	f	start	51	Carol
2	e	complete	52	Kirsten
2	d	start	53	Mary
3	f	complete	55	Carol
2	d	complete	56	Mary
2	f	start	57	Carol
2	f	complete	60	Carol

1. **Pete, Sue, Mary, Carol**
2. **Pete, Mary, Sue, Kirsten, Mary, Carol**
3. **Pete, Mary, Carol**

	Pete	Mary	Sue	Kirsten	Carol
Pete	0.00	0.66	0.33	0.00	0.00
Mary	0.00	0.00	0.33	0.00	1.00
Sue	0.00	0.33	0.00	0.33	0.00
Kirsten	0.00	0.33	0.00	0.00	0.00
Carol	0.00	0.00	0.00	0.00	0.00

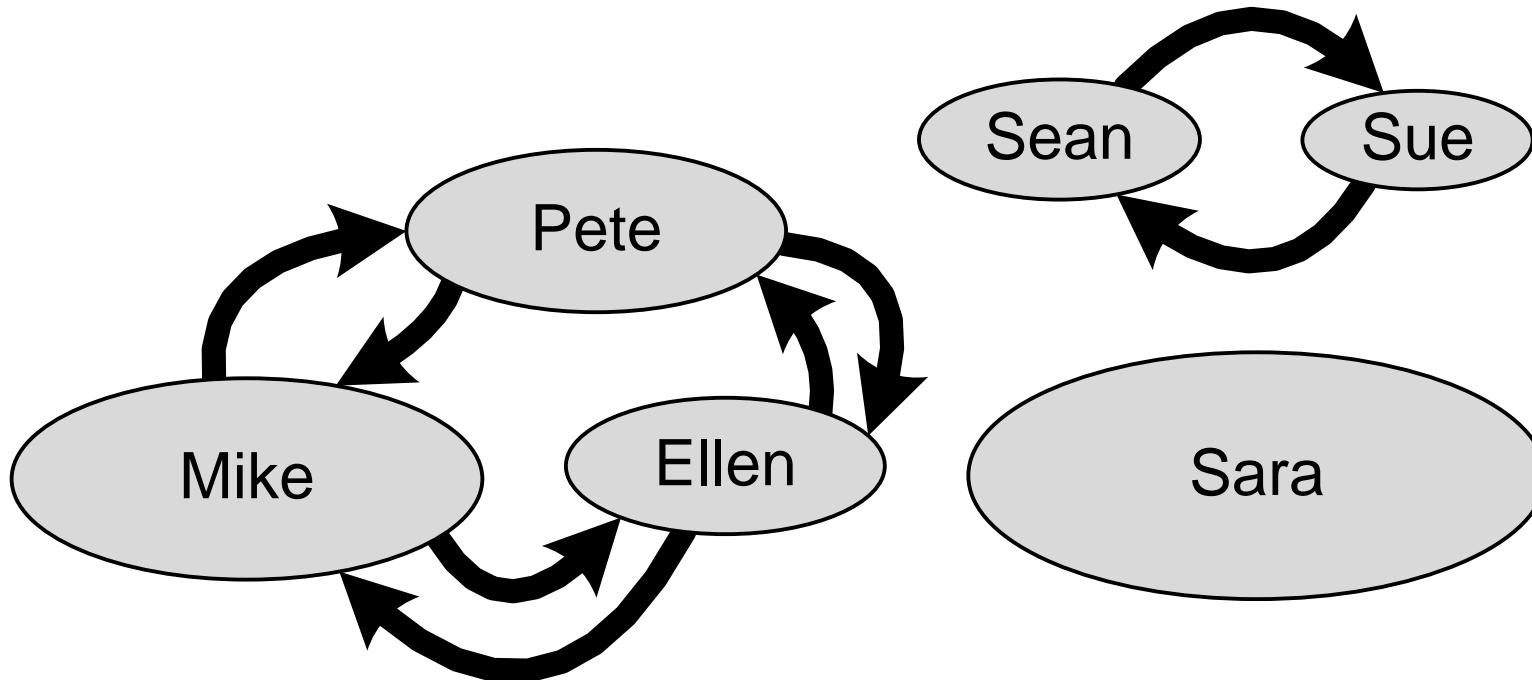
Profiles based on resource-activity matrix

	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>	<i>g</i>	<i>h</i>
Pete	0.3	0	0.345	0.69	0	0	0.135	0.165
Mike	0.5	0	0.575	1.15	0	0	0.225	0.275
Ellen	0.2	0	0.23	0.46	0	0	0.09	0.11
Sue	0	0.46	0	0	0	0	0	0
Sean	0	0.69	0	0	0	0	0	0
Sara	0	0	0	0	2.3	1.3	0	0



find the three
profiles

Social network based on similarity of profiles



Resources that execute similar collections of activities are related. Sara is the only resource executing e and f. Therefore, she is not connected to other resources. Self-loops are suppressed as they contain no information (self-similarity).

Clustering of resources

resource name	average nof times a was executed	average nof times b was executed	average nof times c was executed
Pete	0.15	0.00	0.05
Sue	0.16	0.00	0.04
Mary	0.07	0.01	0.14
Clare	0.09	0.02	0.13
John	0.00	0.80	0.00
...

Clustering of resources

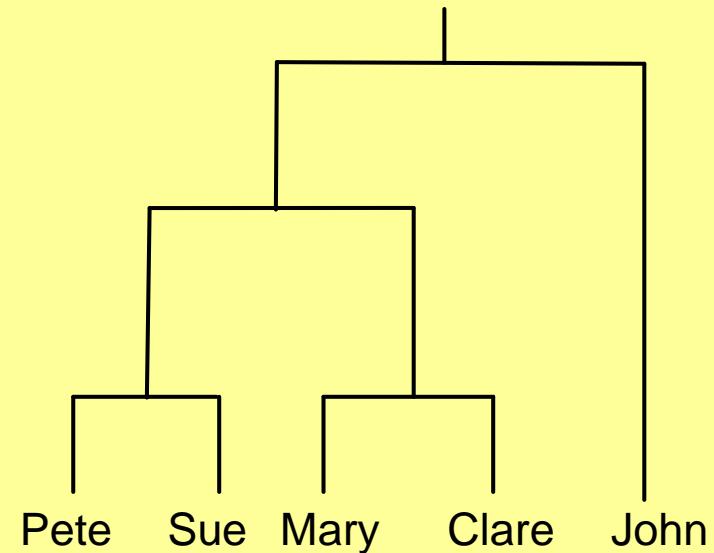
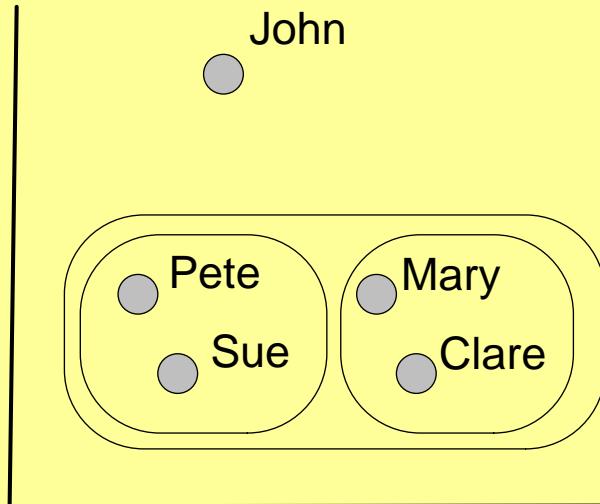
resource name	average nof times a was executed	average nof times b was executed	average nof times c was executed
Pete	0.15	0.00	0.05
Sue	0.16	0.00	0.04
Mary	0.07	0.01	0.14
Clare	0.09	0.02	0.13
John	0.00	0.80	0.00
...

feature vector

Clustering of resources

resource name	average nof times a was executed	average nof times b was executed	average nof times c was executed
Pete	0.15	0.00	0.05
Sue	0.16	0.00	0.04
Mary	0.07	0.01	0.14
Clare	0.09	0.02	0.13

feature vector

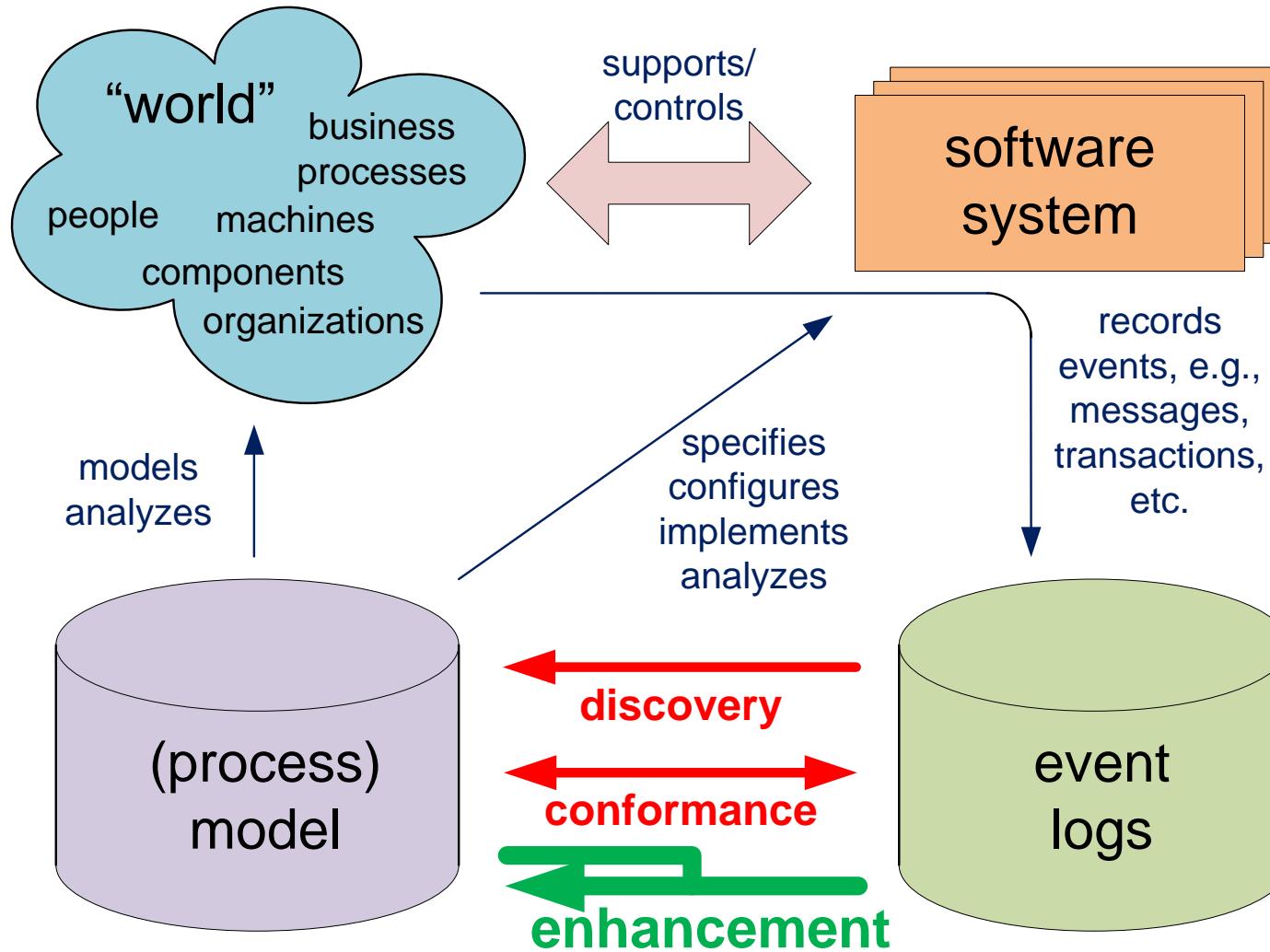


So we can mine social networks based on handover of work matrix or resource-activity matrix ...

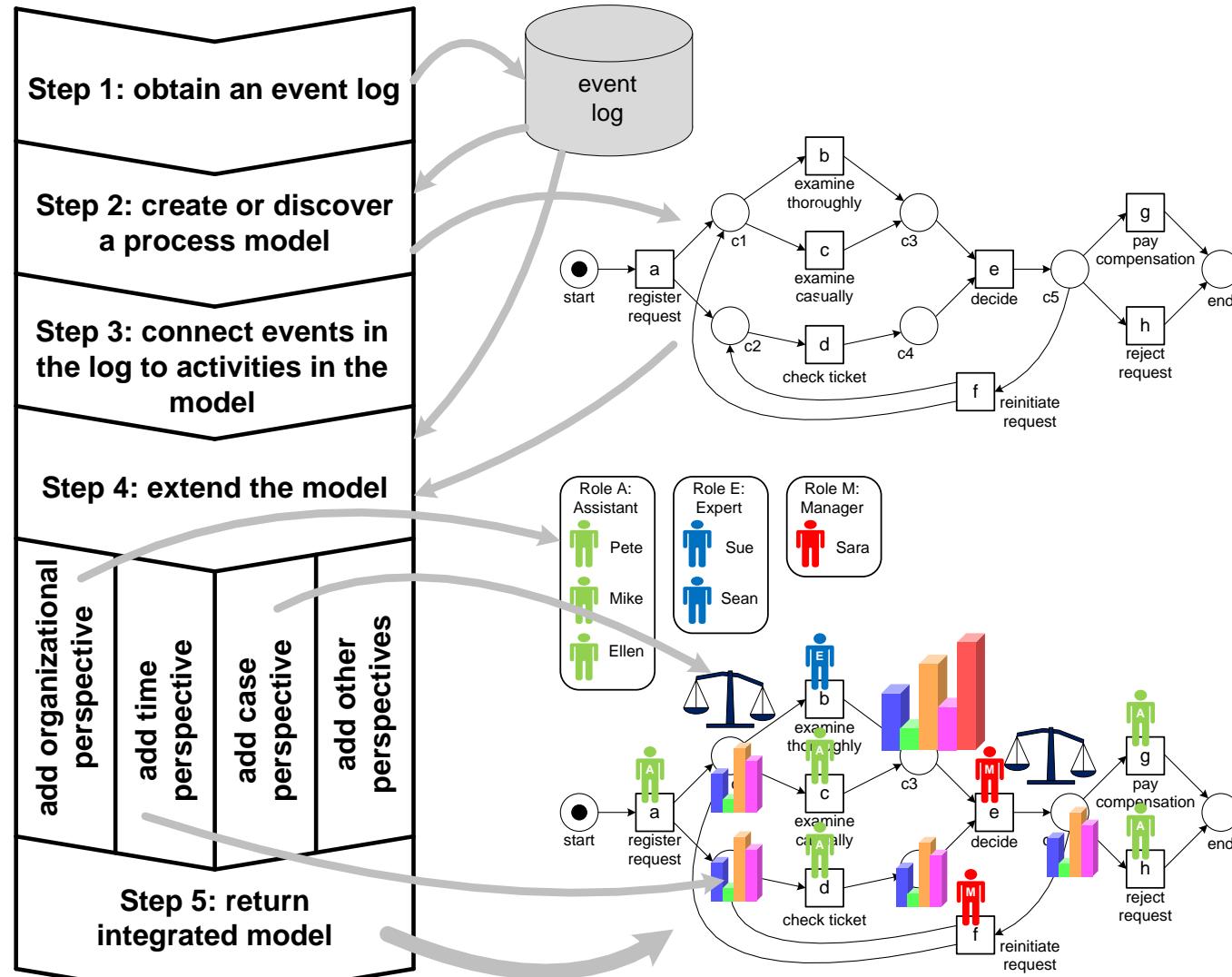


conclusion and outlook

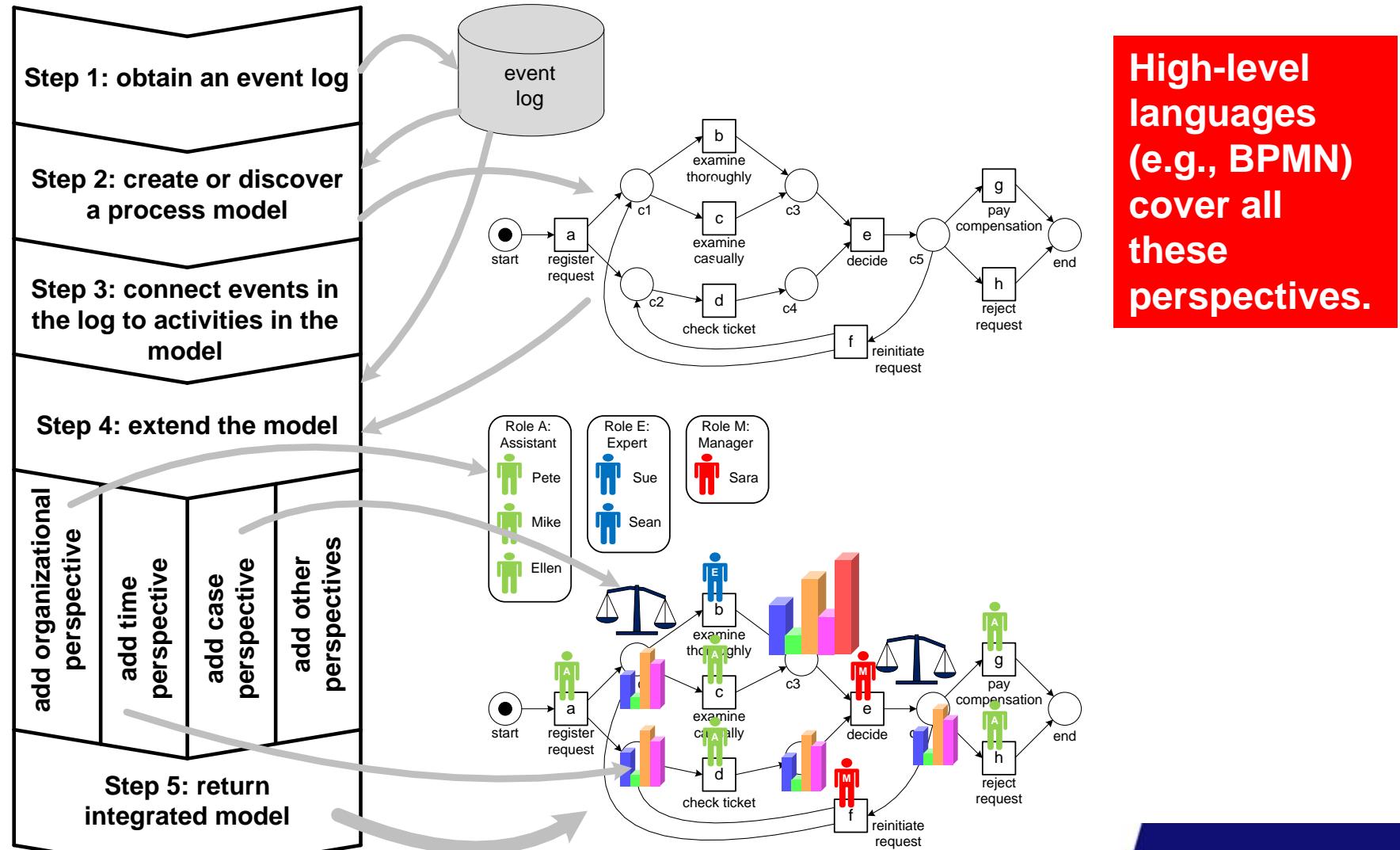
Enhancement: making better /more useful process models



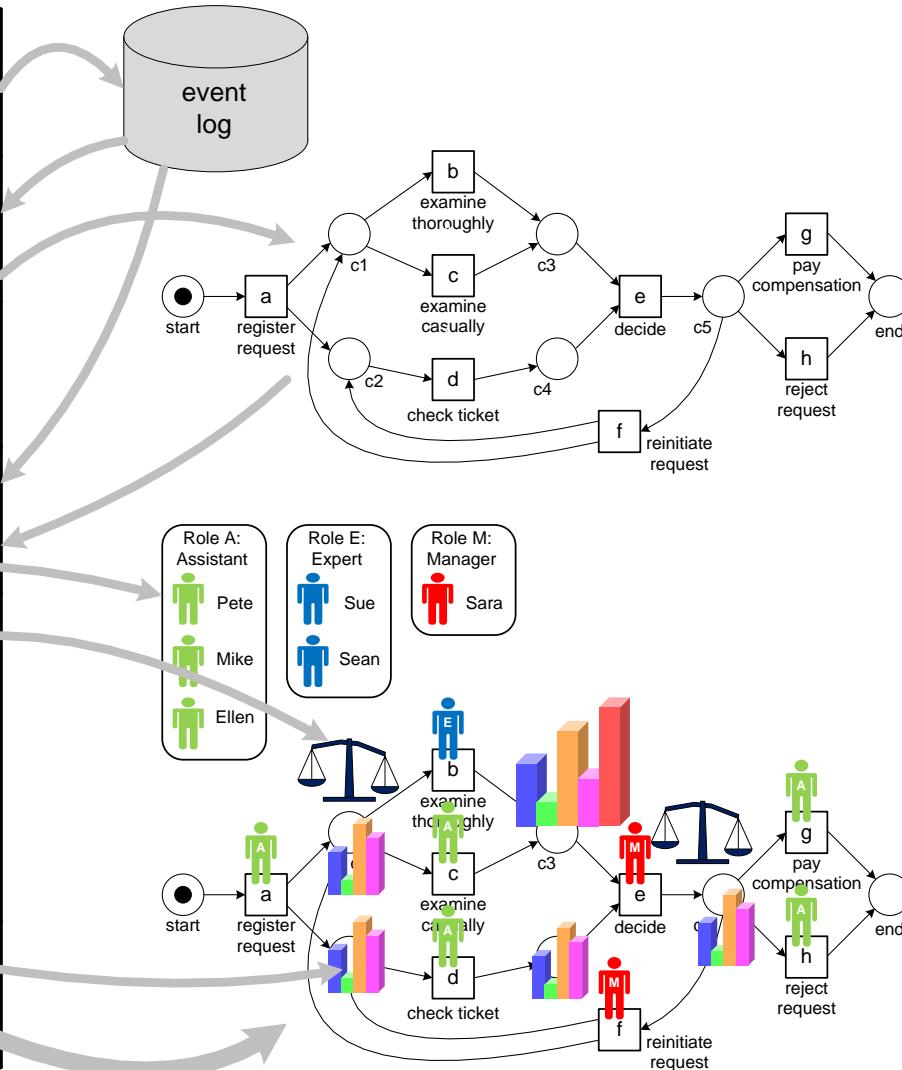
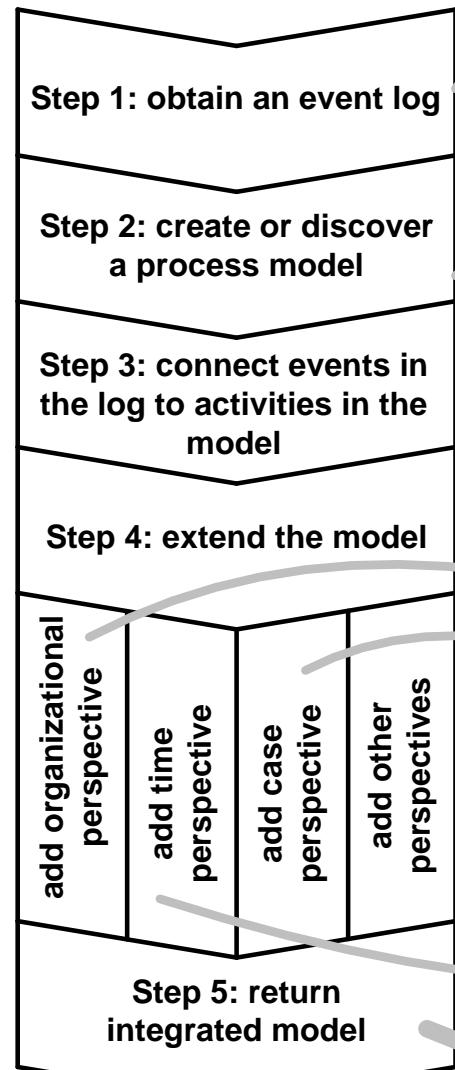
Bringing it all together



Bringing it all together



Bringing it all together



High-level languages (e.g., BPMN) cover all these perspectives.

Even possible to automatically learn simulation models!



Process Mining Data Science in Action

Starts
every
4
weeks

Info MOOC: www.coursera.org/course/procmin

This is the second edition of Wil van der Aalst's seminal book on process mining, which now discusses the field also in the broader context of data science and big data approaches. It includes several additions and updates, e.g. on inductive mining techniques, the notion of alignments, a considerably expanded section on software tools and a completely new chapter on process mining in the large. It is self-contained, while at the same time covering the entire process-mining spectrum from process discovery to predictive analytics.

After a general introduction to data science and process mining in Part I, Part II provides the basics of business process modeling and data mining necessary to understand the remainder of the book. Next, Part III focuses on process discovery as the most important process mining task, while Part IV moves beyond discovering the control flow of processes, highlighting conformance checking, and organizational and time perspectives. Part V offers a guide to successfully applying process mining in practice, including an introduction to the widely used open-source tool ProM and several commercial products. Lastly, Part VI 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 data mining within the context of process mining
- Comprehensive material, e.g. on data quality, inductive mining techniques and the notion of alignments
- Written for computer scientists and business analysts and audience in academia and industry
- Tools and the exploitation of process mining results

**Data science
Big data
Alignments
Data quality
Inductive mining
Tooling**

ISBN 978-3-662-49850-7



► springer.com

van der Aalst

Wil van der Aalst



Process Mining

Process Mining

Data Science in Action

Second Edition

<http://www.springer.com/9783662498507>

2nd Ed.

Springer