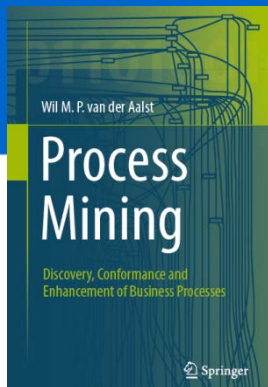


Process Mining: Data Science in Action

Mining Bottlenecks

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



TU/e

Technische Universiteit
Eindhoven
University of Technology

Where innovation starts



**process mining is most interesting for processes
having performance problems**

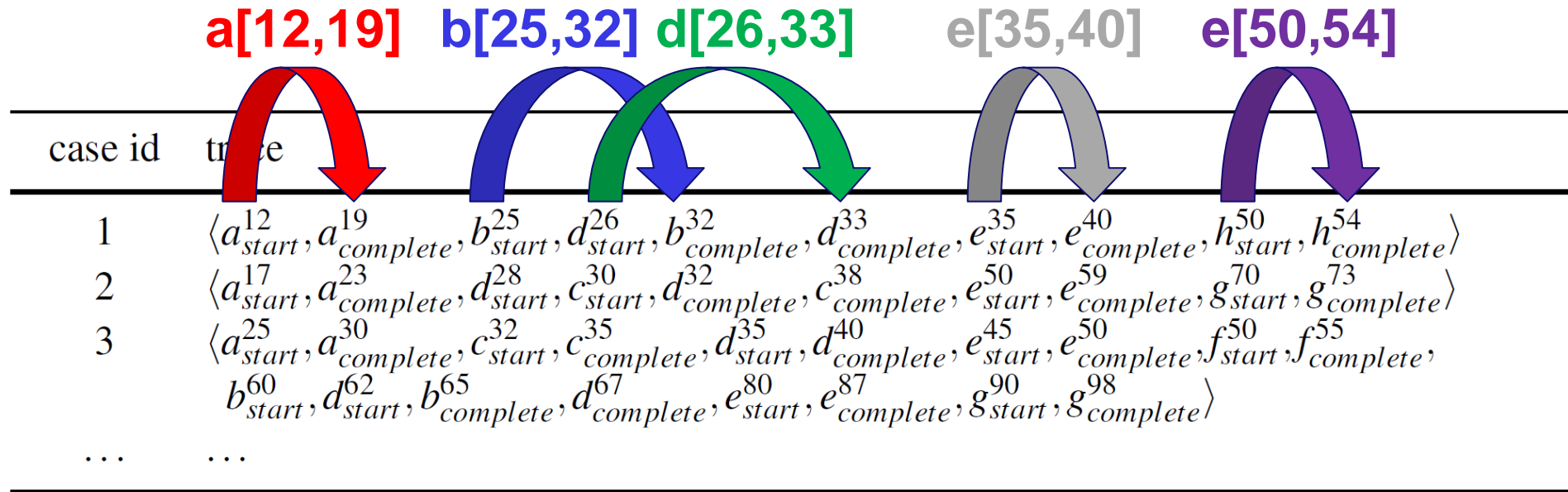
bottlenecks



where?

why?

Event data with timestamps and transactional information



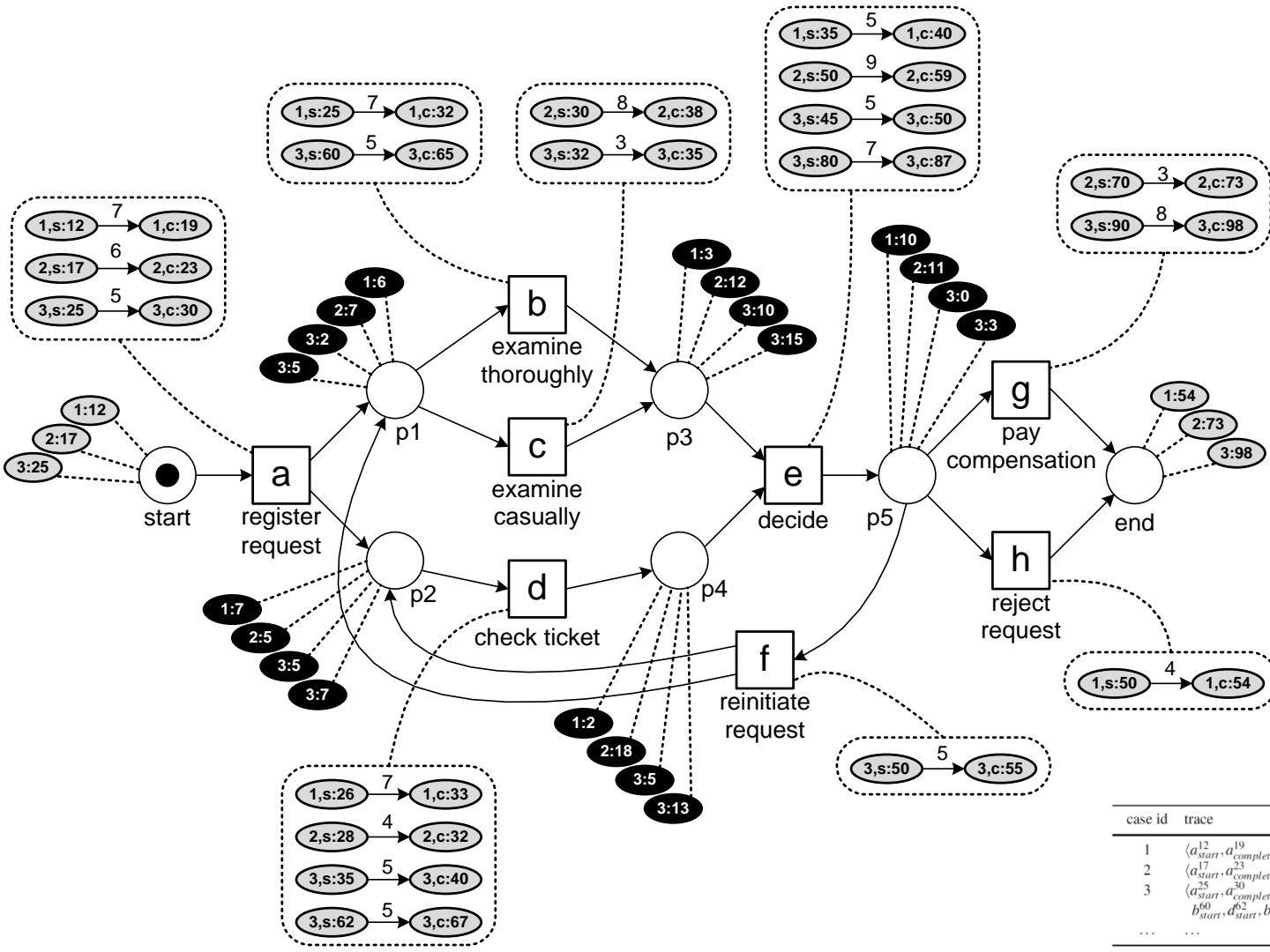
Learning time and probabilities

case id trace

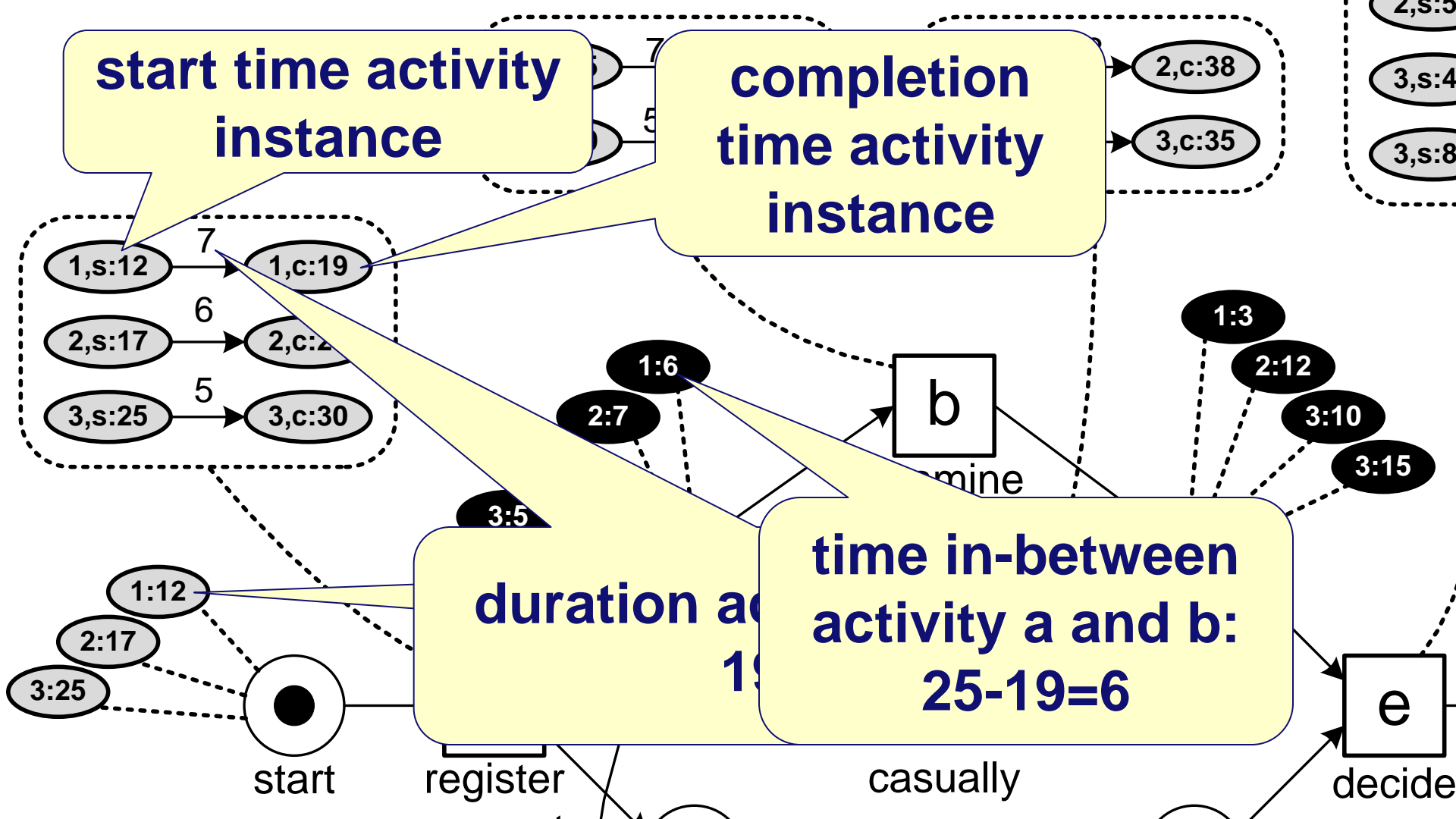
1	$\langle a_{start}^{12}, a_{complete}^{19}, b_{start}^{25}, d_{start}^{26}, b_{complete}^{32}, d_{complete}^{33}, e_{start}^{35}, e_{complete}^{40}, h_{start}^{50}, h_{complete}^{54} \rangle$
2	$\langle a_{start}^{17}, a_{complete}^{23}, d_{start}^{28}, c_{start}^{30}, d_{complete}^{32}, c_{complete}^{38},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

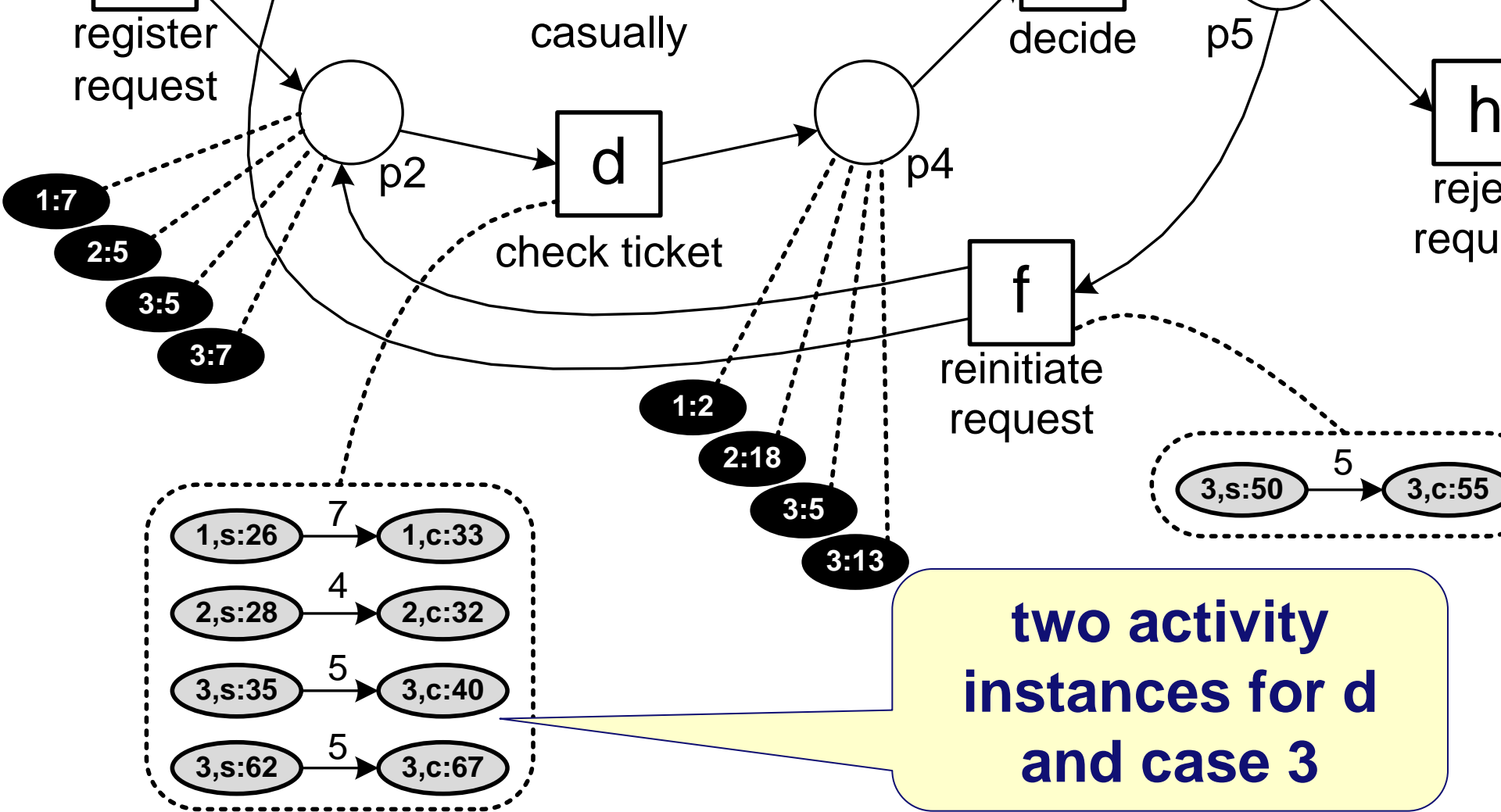
- **case 1** starts at time **12** and ends at time **54**
- **case 2** starts at time **17** and ends at time **73**
- **case 3** starts at time **25** and ends at time **98**

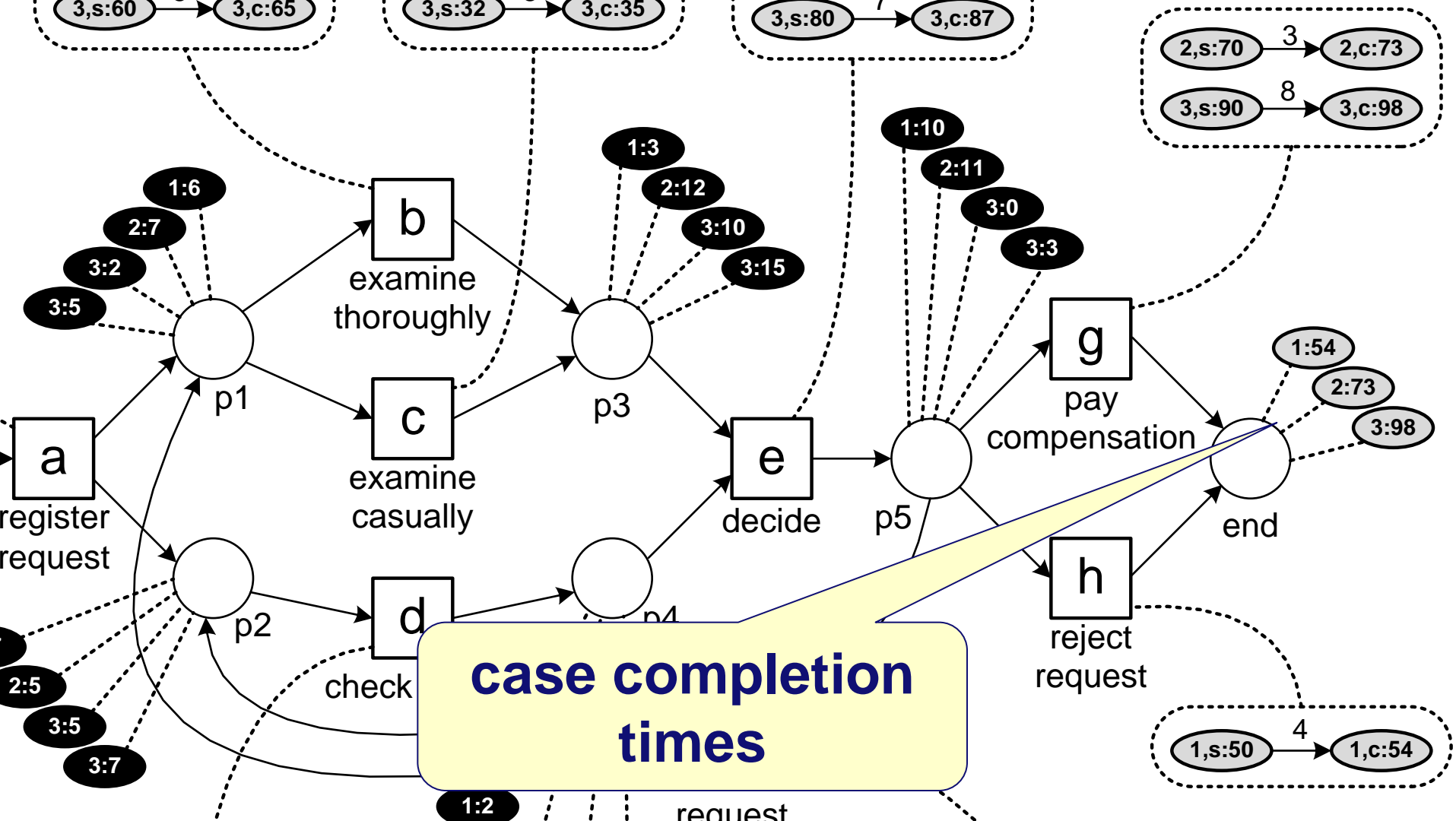
replaying the first three cases



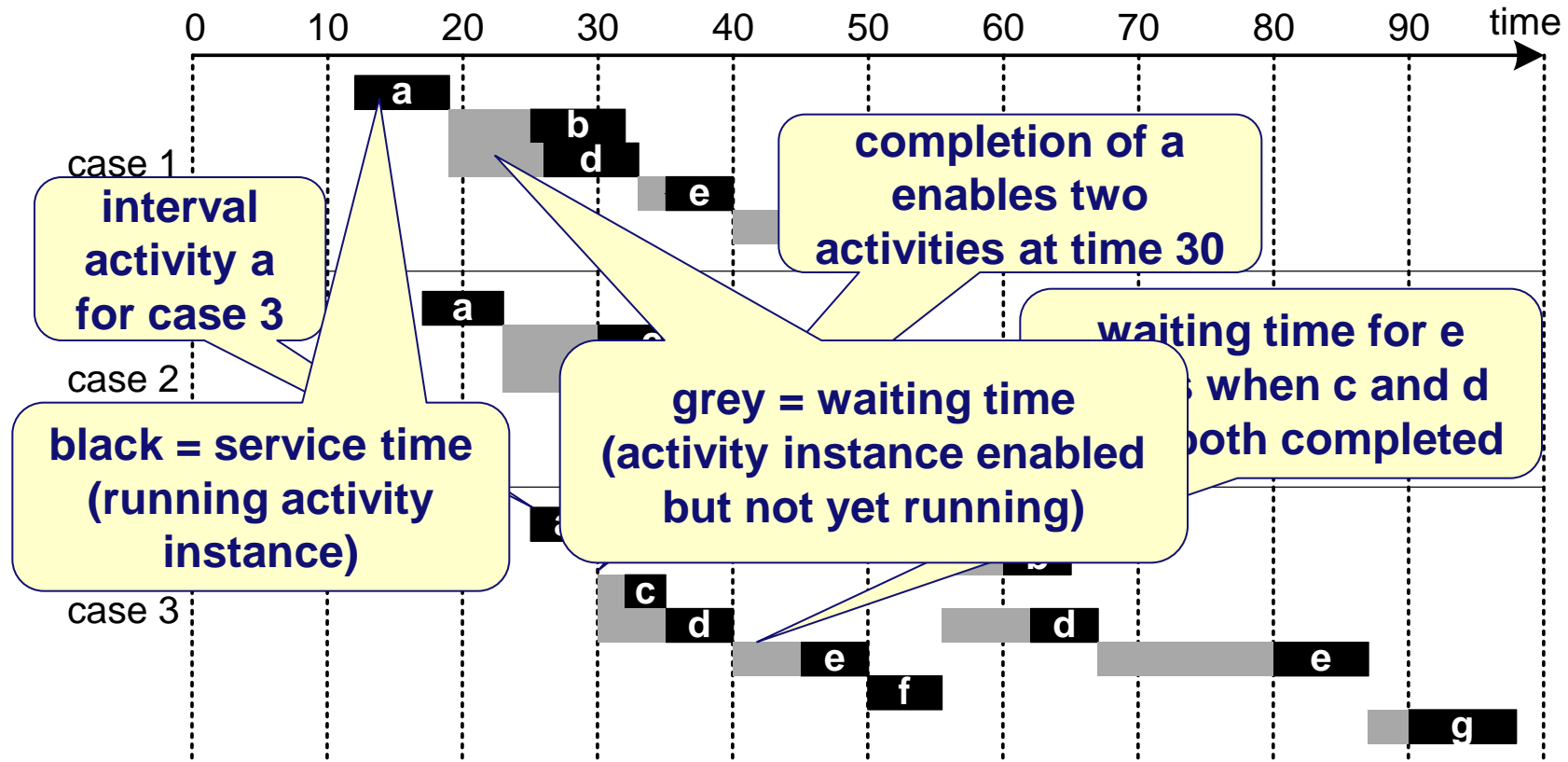
case id	trace
1	$\langle a_{start}^{12}, a_{complete}^{10}, b_{start}^{25}, d_{start}^{26}, b_{complete}^{32}, d_{complete}^{33}, c_{start}^{35}, e_{start}^{40}, h_{start}^{54}, h_{complete}^{54} \rangle$
2	$\langle a_{start}^{17}, a_{complete}^{28}, c_{start}^{30}, c_{complete}^{38}, e_{start}^{50}, e_{complete}^{59}, s_{start}^{70}, s_{complete}^{73} \rangle$
3	$\langle a_{start}^{25}, a_{complete}^{32}, c_{start}^{35}, c_{complete}^{40}, e_{start}^{45}, e_{complete}^{50}, f_{start}^{55}, f_{complete}^{60}, b_{start}^{62}, b_{complete}^{65}, d_{start}^{80}, d_{complete}^{87}, s_{start}^{90}, s_{complete}^{98} \rangle$
...	...



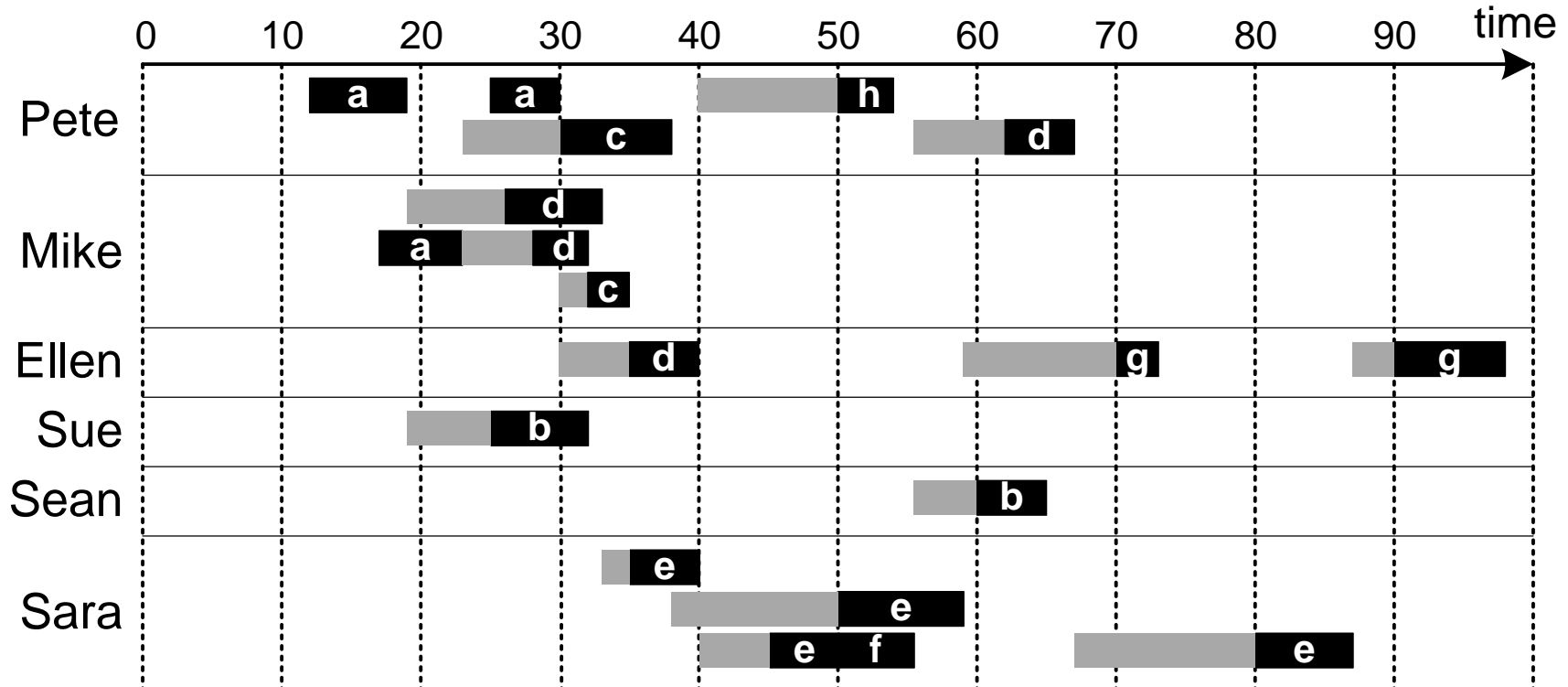




Another view on the timed replay of the first three cases

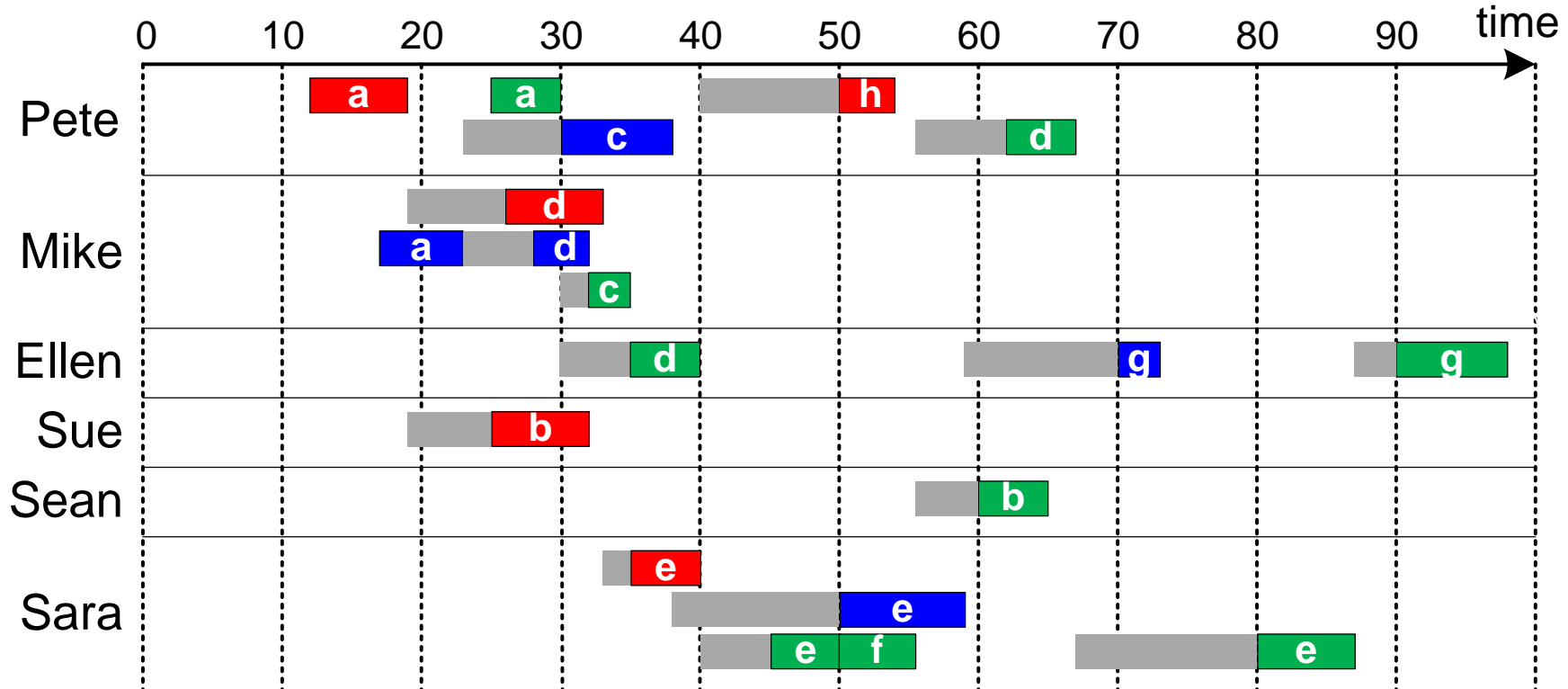


Timed replay projected onto resources

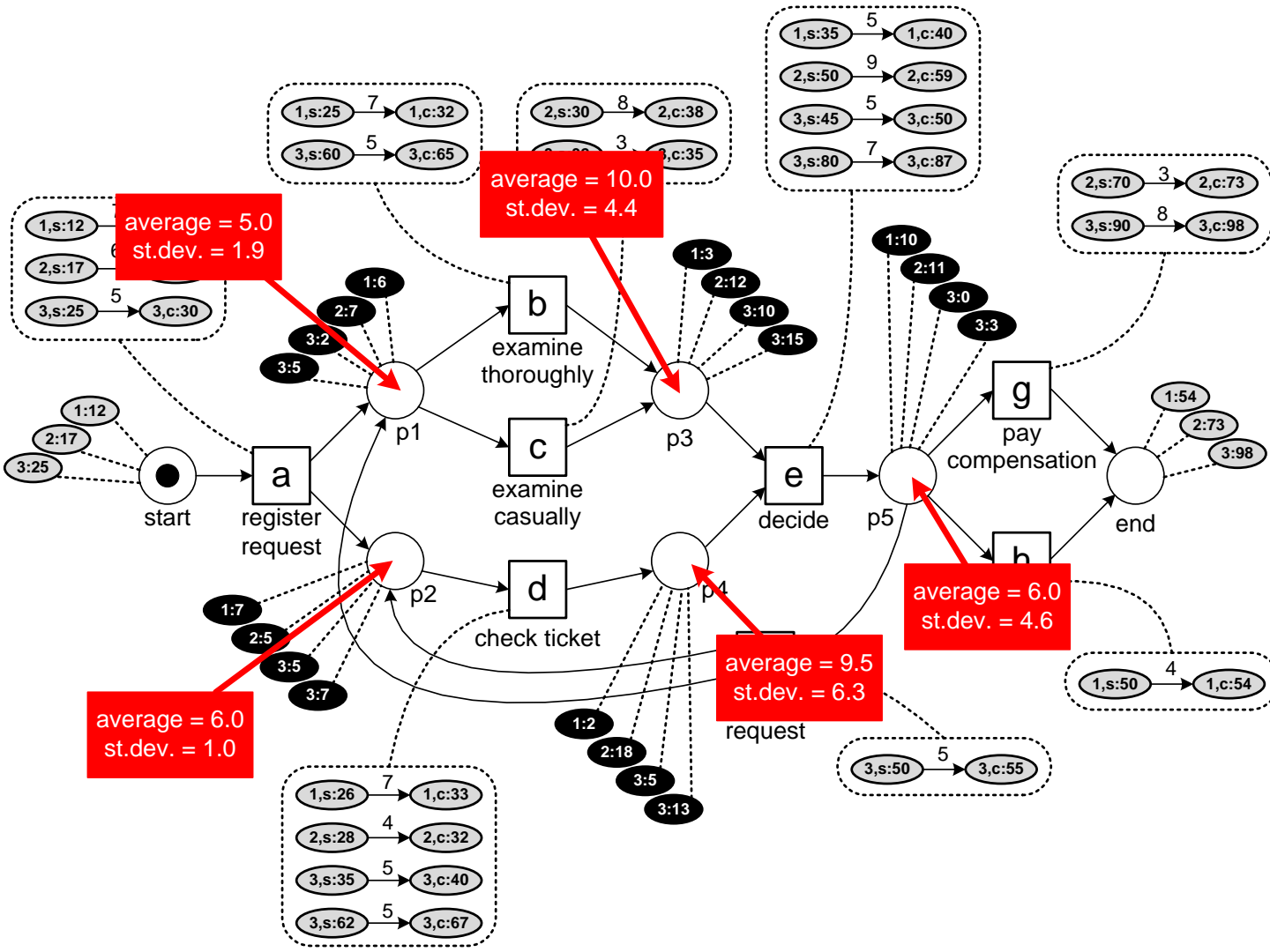


Timed replay projected onto resources

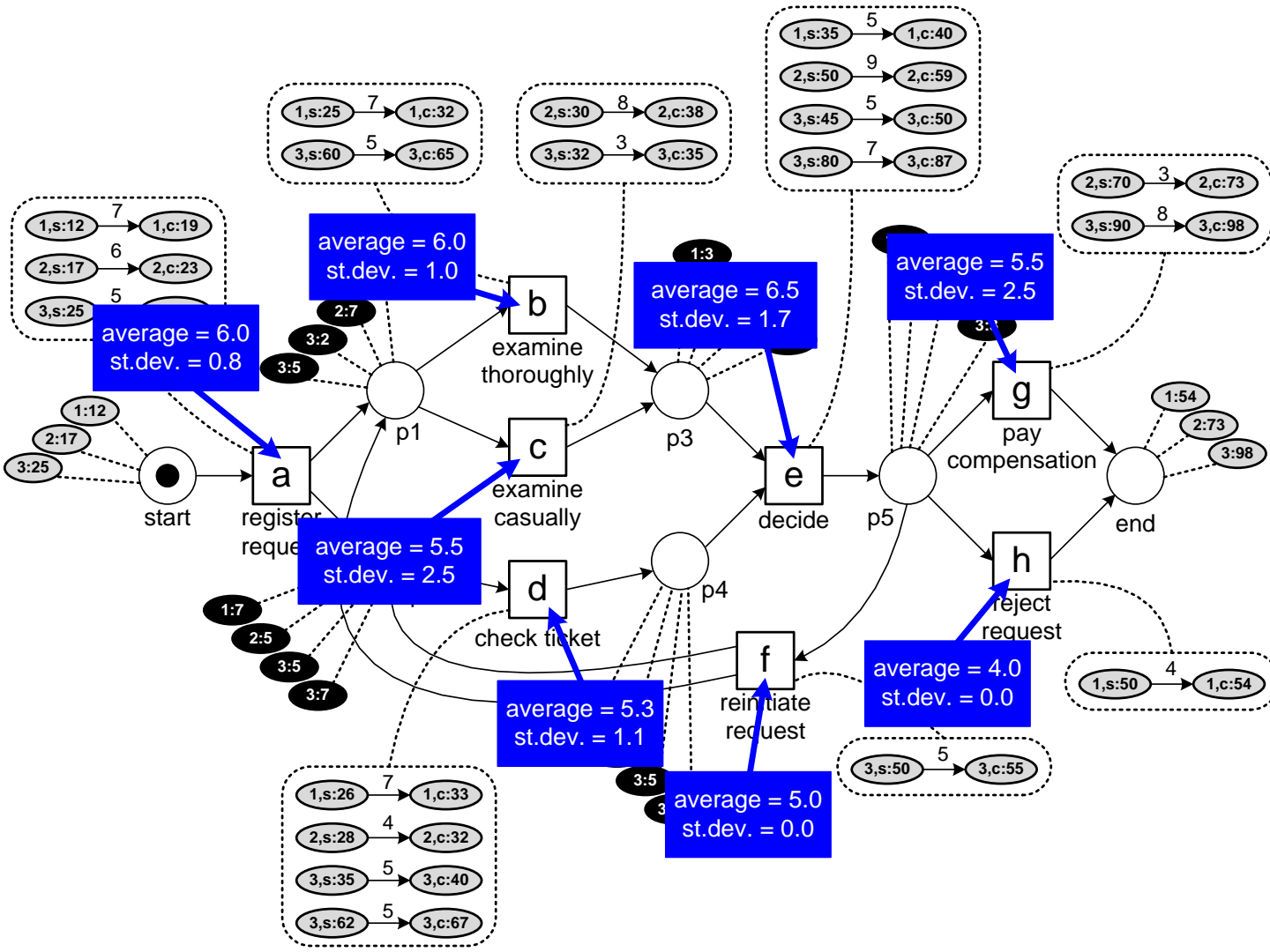
(activities colored by case) ① ② ③



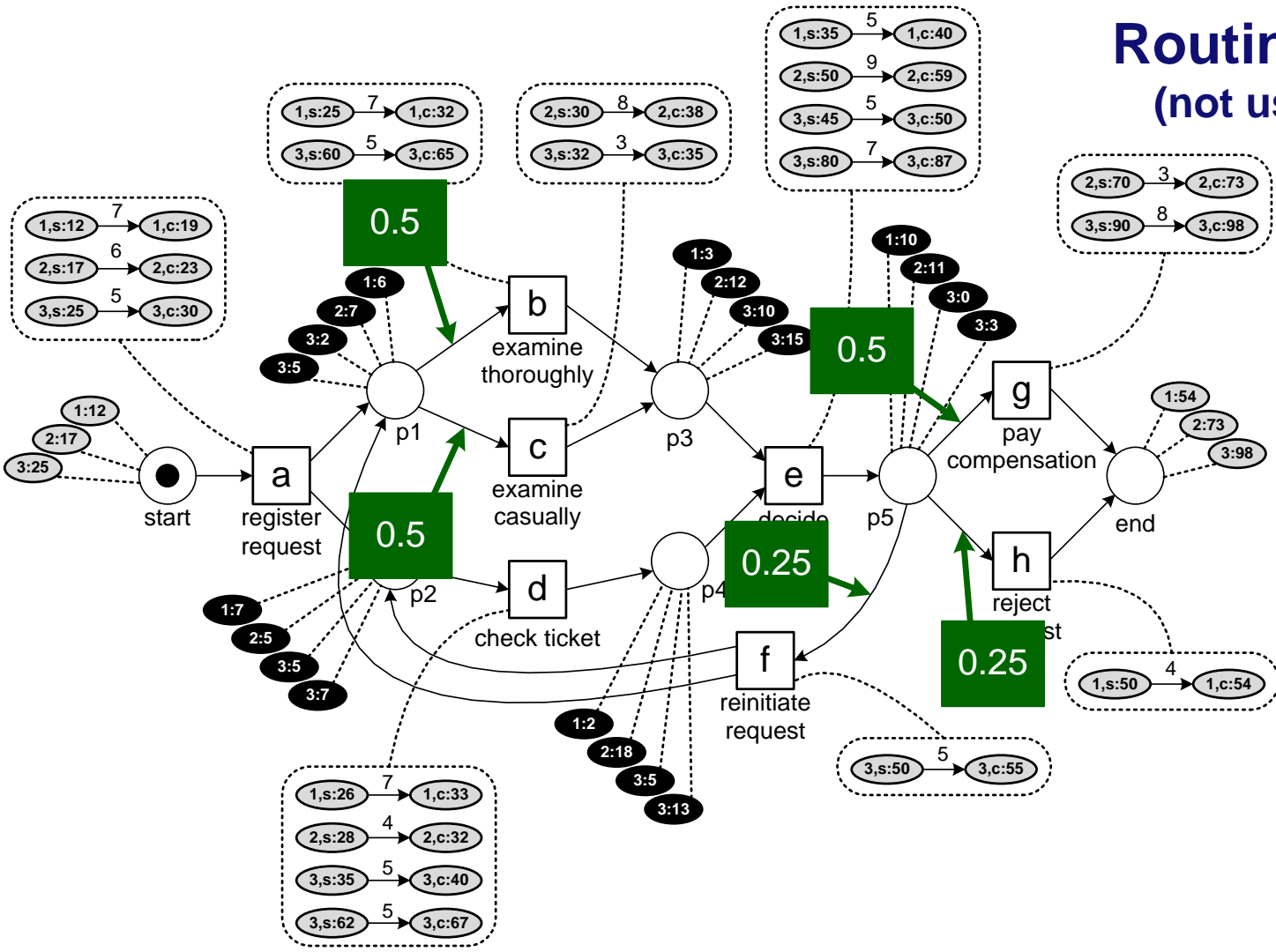
Waiting times (average and standard deviation)



Service times (average and standard deviation)



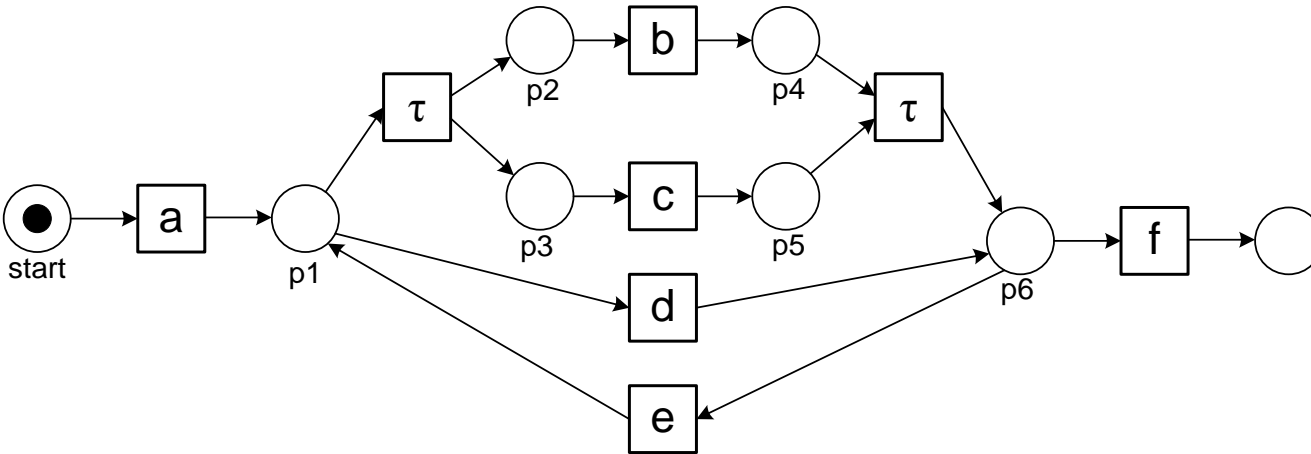
Routing probabilities (not using data attributes)



Another event log

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

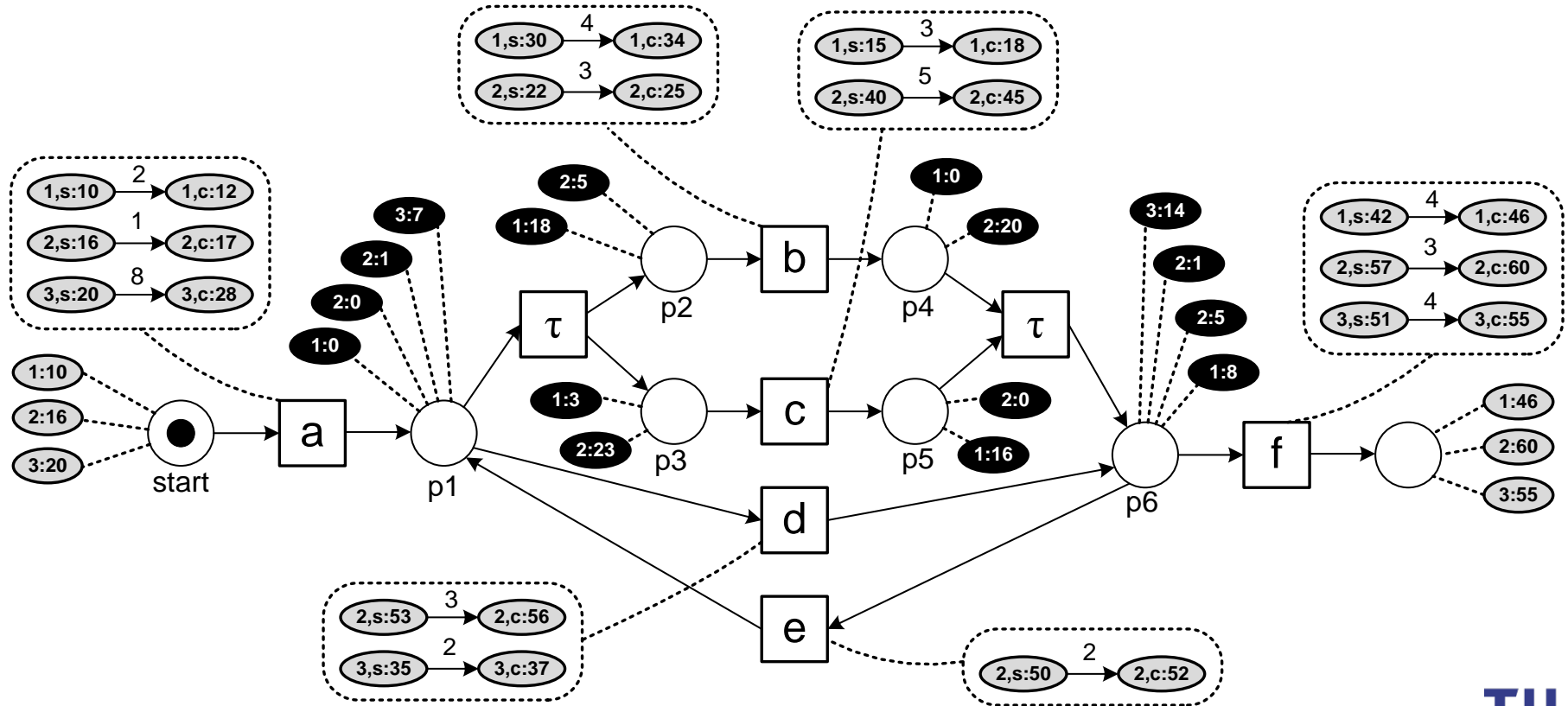
The corresponding model



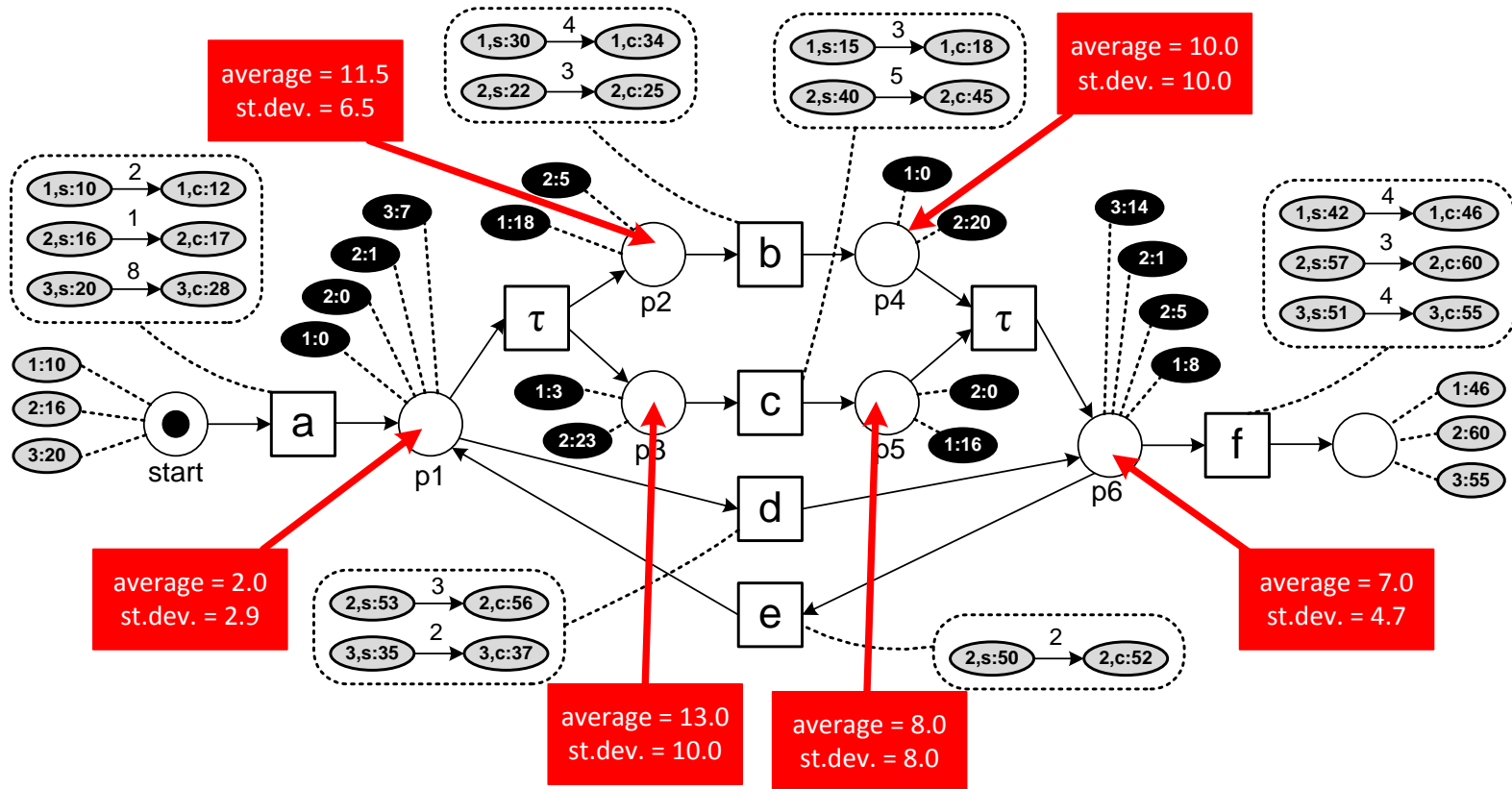
**Let's estimate service times,
waiting times, and routing
probabilities**

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

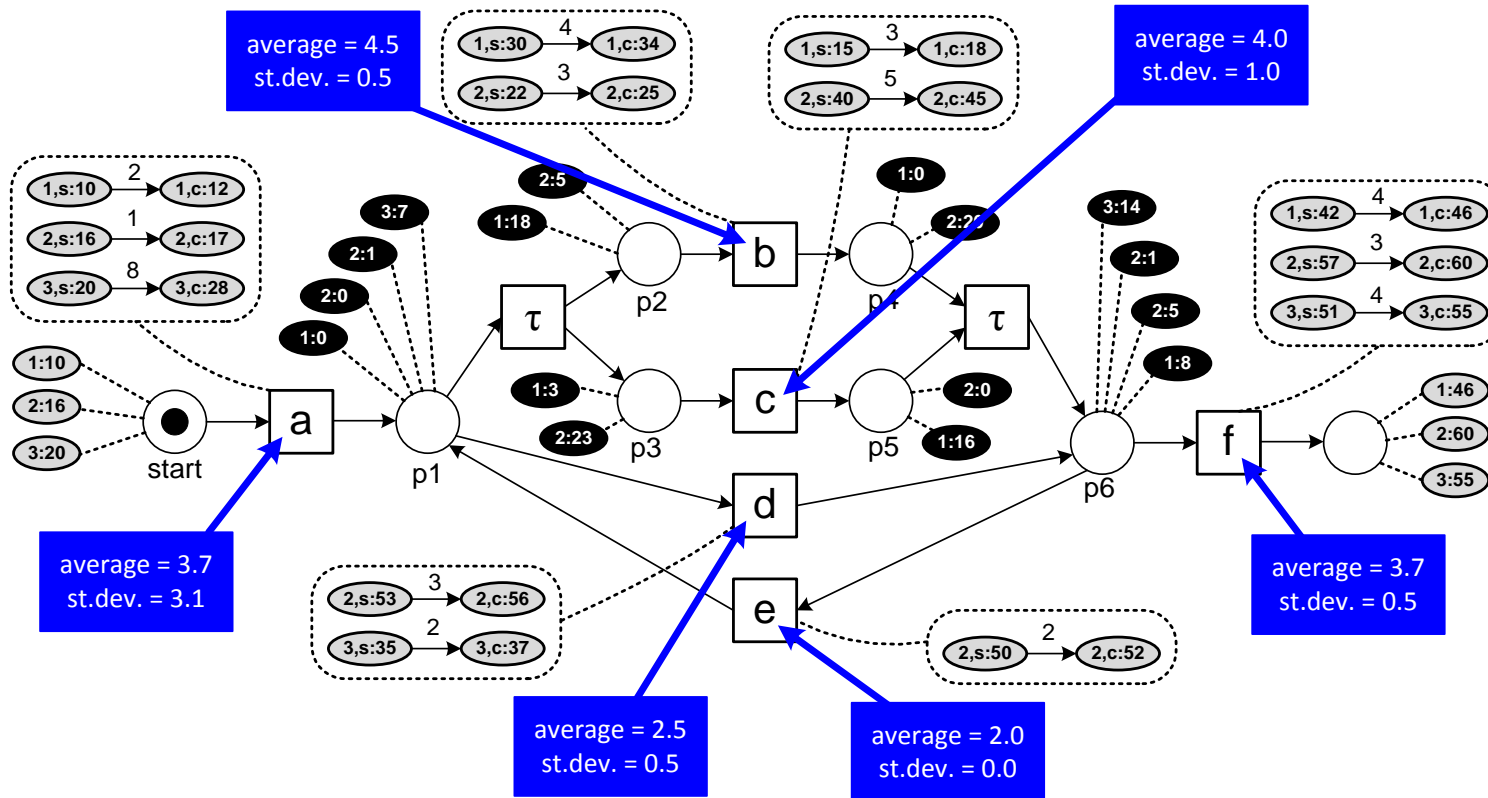
Times recorded during replay



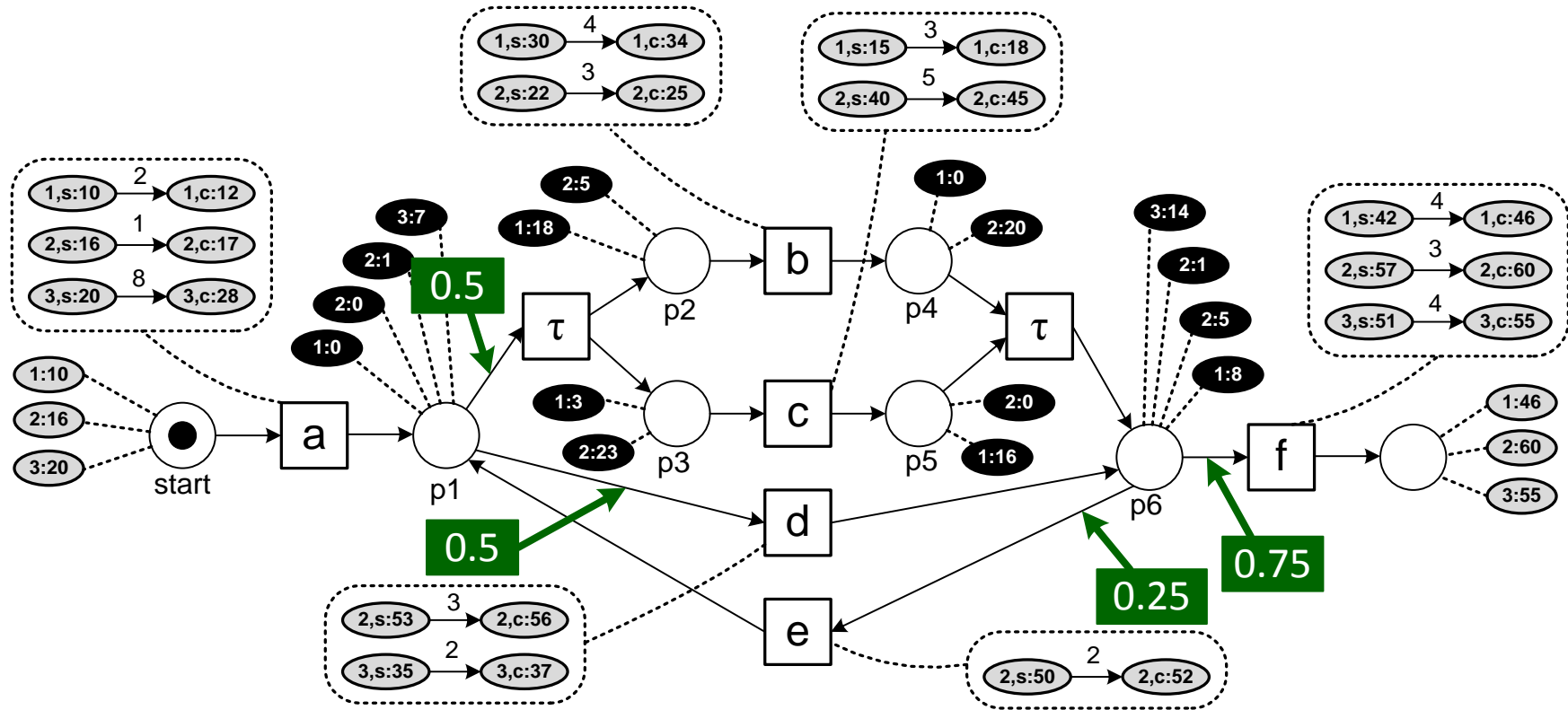
Waiting times



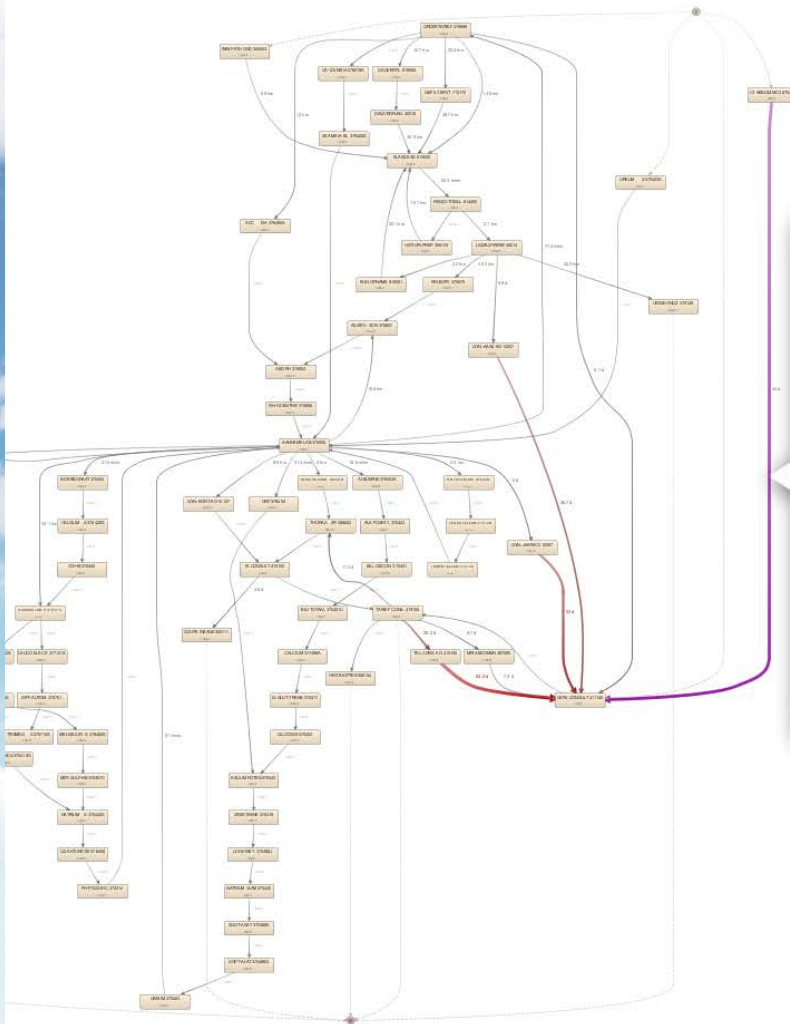
Service times



Routing probabilities



we have seen this before ...



Disco

CTABDOM.MC 387042A → VERV.
CONSULT 411100

Frequency

Absolute frequency	20
Case frequency	20
Max. repetitions	1

Performance

Total duration	22.4 mths
Median duration	7.8 d
Mean duration	34 d
Max. duration	17.6 wks

[Filter this path...](#)

Inspector

Info Display Filter Export

Legend

View

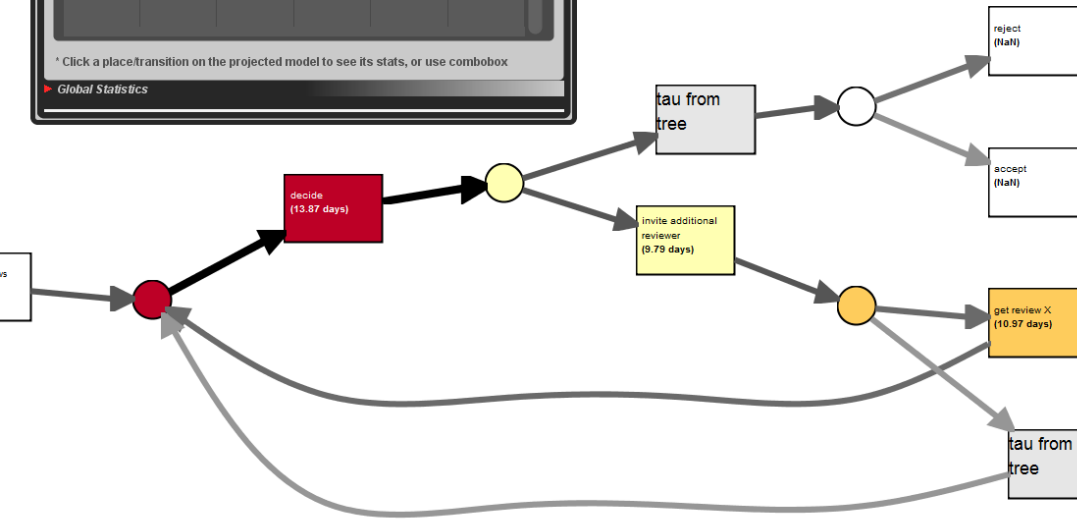
▼ Element Statistics

Selected elements* 2 sink

Property	Min.	Max.	Avg.	Std. Dev	Freq.
Waiting time	0.00 ms	5.03 months	13.87 days	18.84 days	884
Synchroniz...	0.00 ms	0.00 ms	0.00 ms	0.00 ms	884
Sojourn time	0.00 ms	5.03 months	13.87 days	18.84 days	884

* Click a place/transition on the projected model to see its stats, or use combobox

Global Statistics



ProM





Using process mining we can automatically :

- **discover process models,**
- **align log and model,**
- **replay the alignments, and**
- **identify bottlenecks and other performance related problems.**

Starting point for process improvement and operational support (e.g., predictions)!

Part I: Preliminaries

Chapter 1
Introduction

Chapter 2
Process Modeling and
Analysis

Chapter 3
Data Mining

Part II: From Event Logs to Process Models

Chapter 4
Getting the Data

Chapter 5
Process Discovery: An
Introduction

Chapter 6
Advanced Process
Discovery Techniques

Part III: Beyond Process Discovery

Chapter 7
Conformance
Checking

Chapter 8
Mining Additional
Perspectives

Chapter 9
Operational Support

Part IV: Putting Process Mining to

Chapter 10
Tool Support

Chapter 11
Analyzing “Lasagna
Processes”

Chapter 12
Analyzing “Spaghetti
Processes”

Part V: Reflection

Chapter 13
Cartography and
Navigation

Chapter 14
Epilogue



Wil M. P. van der Aalst

Process Mining

Discovery, Conformance and
Enhancement of Business Processes

 Springer

