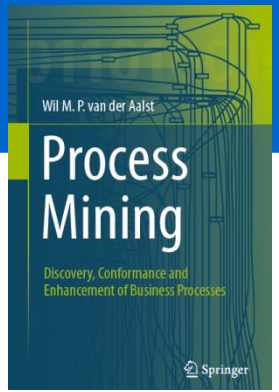
Process Mining: Data Science in Action

Different Types of Process Mining



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

TU/e Technische Universiteit Eindhoven University of Technology

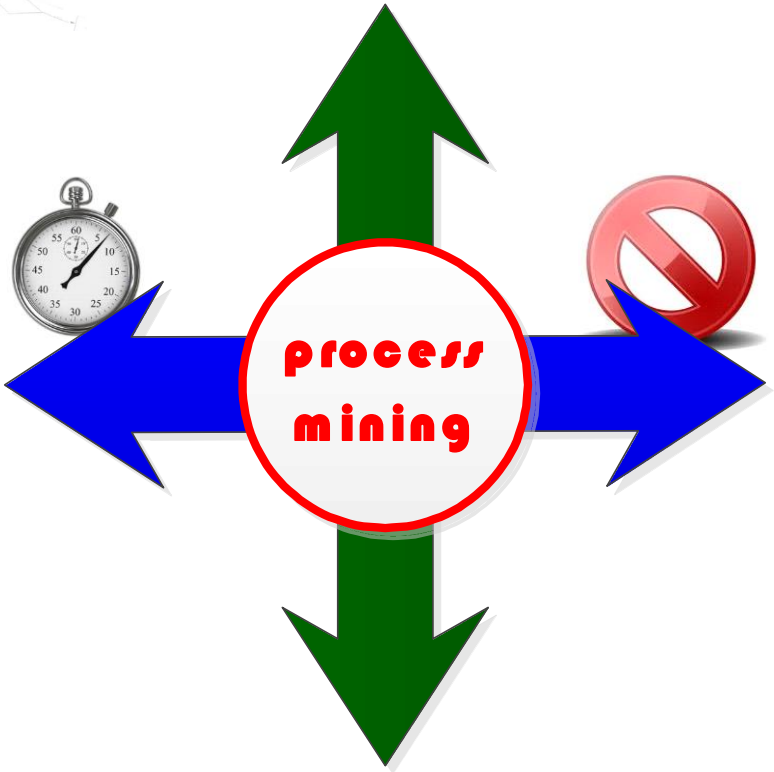
Where innovation starts

Process Mining: Data Science in Action

Positioning process mining

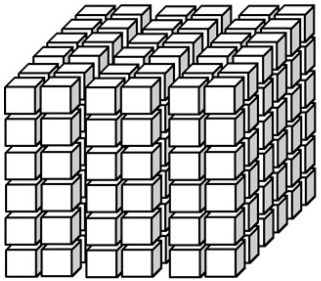
process model analysis

(simulation, verification, optimization, gaming, etc.)



performance-oriented questions, problems and solutions

compliance-oriented questions, problems and solutions



data-oriented analysis

(data mining, machine learning, business intelligence)

Starting point for process mining: Event data

every row is an event (here: an exam attempt)

student name	course name	exam date	mark
Peter Jones	Business Information systems	16-1-2014	8
Sandy Scott	Business Information systems	16-1-2014	5
Bridget White	Business Information systems	16-1-2014	9
John Anderson	Business Information systems	16-1-2014	8
Sandy Scott	BPM Systems	17-1-2014	7
Bridget White	BPM Systems	17-1-2014	8
Sandy Scott	Process Mining	20-1-2014	5
Bridget White	Process Mining	20-1-2014	9
John Anderson	Process Mining	20-1-2014	8
…	…	…	…

case id | activity name | timestamp | other data

TU/e

Another event log: order handling

order number	activity	timestamp	user	product	quantity
9901	register order	22-1-2014@09.15	Sara Jones	iPhone5S	1
9902	register order	22-1-2014@09.18	Sara Jones	iPhone5S	2
9903	register order	22-1-2014@09.27	Sara Jones	iPhone4S	1
9901	check stock	22-1-2014@09.49	Pete Scott	iPhone5S	1
9901	ship order	22-1-2014@10.11	Sue Fox	iPhone5S	1
9903	check stock	22-1-2014@10.34	Pete Scott	iPhone4S	1
9901	handle payment	22-1-2014@10.41	Carol Hope	iPhone5S	1
9902	check stock	22-1-2014@10.57	Pete Scott	iPhone5S	2
9902	cancel order	22-1-2014@11.08	Carol Hope	iPhone5S	2
…	…	…	…	…	…

case id **activity name** **timestamp** **resource** **other data**

TU/e

Another event log: patient treatment

patient	activity	timestamp	doctor	age	cost
5781	make X-ray	23-1-2014@10.30	Dr. Jones	45	70.00
5541	blood test	23-1-2014@10.18	Dr. Scott	61	40.00
5833	blood test	23-1-2014@10.27	Dr. Scott	24	40.00
5781	blood test	23-1-2014@10.49	Dr. Scott	45	40.00
5781	CT scan	23-1-2014@11.10	Dr. Fox	45	1200.00
5833	surgery	23-1-2014@12.34	Dr. Scott	24	2300.00
5781	handle payment	23-1-2014@12.41	Carol Hope	45	0.00
5541	radiation therapy	23-1-2014@13.57	Dr. Jones	61	140.00
5541	radiation therapy	23-1-2014@13.08	Dr. Jones	61	140.00
…	…	…	…	…	…

case id **activity name** **timestamp** **resource** **other data**

TU/e

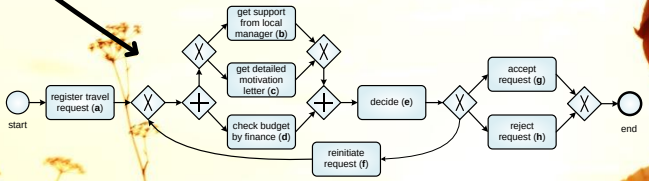
Let's play

Case	Activity	Timestamp	Resource
432	register travel request (a)	18-3-2014:9.15	John
432	get support from local manager (b)	18-3-2014:9.25	Mary
432	check budget by finance (d)	19-3-2014:8.55	John
432	decide (e)	19-3-2014:9.36	Sue
432	accept request (g)	19-3-2014:9.48	Mary

Play-In

Replay

Play-Out

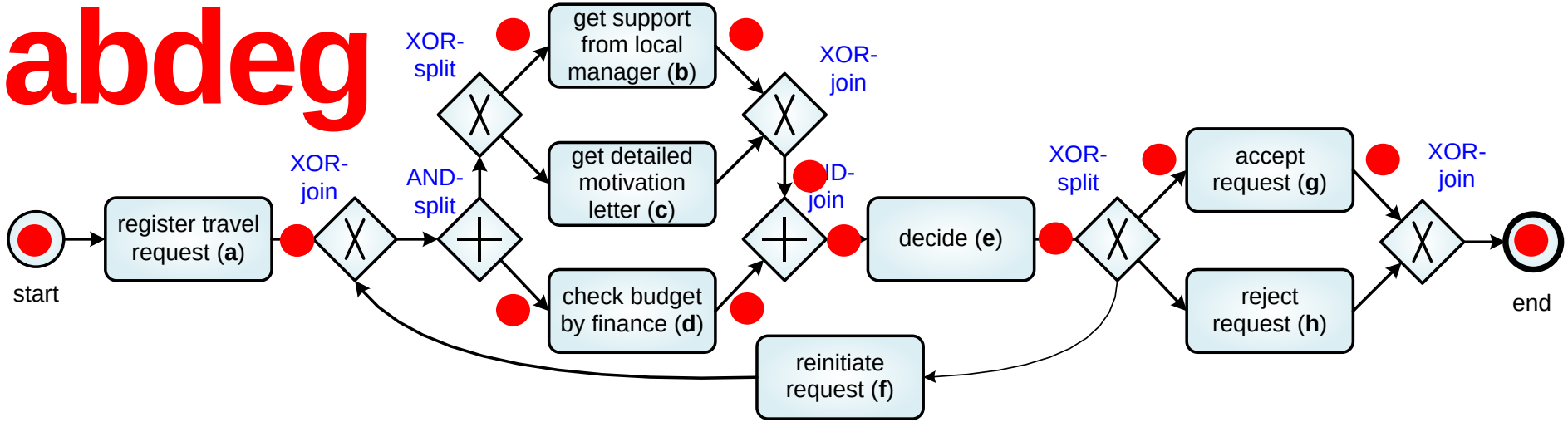


Play-Out

Case	Activity	Timestamp	Resource
432	register travel request (a)	18-3-2014:9.15	John
432	get support from local manager (b)	18-3-2014:9.25	Mary
432	check budget by finance (d)	19-3-2014:8.55	John
432	decide (e)	19-3-2014:9.36	Sue
432	accept request (g)	19-3-2014:9.48	Mary

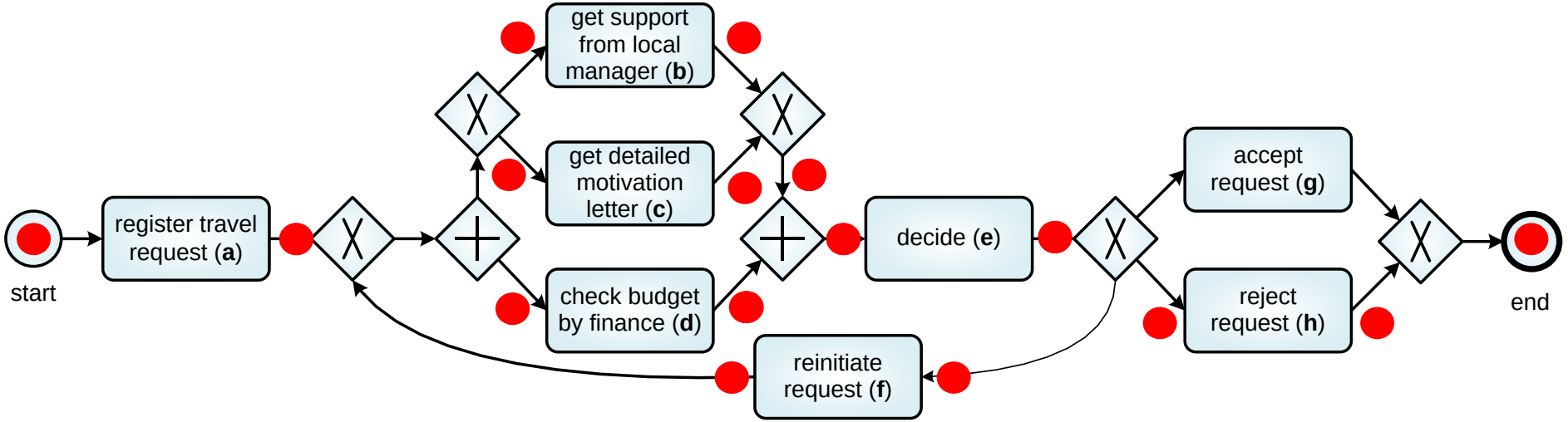
Play Out: A possible scenario

abdeg



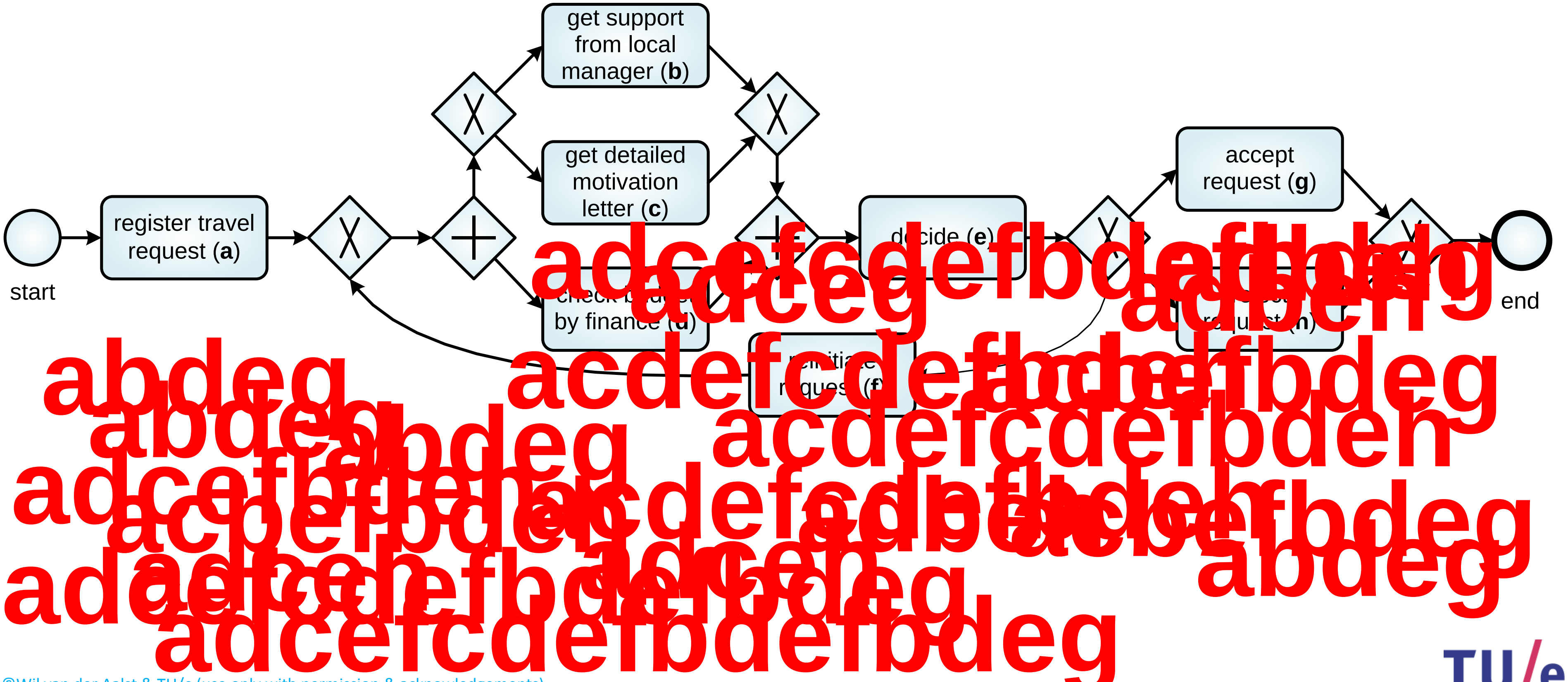
Case	Activity	Timestamp	Resource
432	register travel request (a)	18-3-2014:9.15	John
432	get support from local manager (b)	18-3-2014:9.25	Mary
432	check budget by finance (d)	19-3-2014:8.55	John
432	decide (e)	19-3-2014:9.36	Sue
432	accept request (g)	19-3-2014:9.48	Mary

Play Out: Another scenario



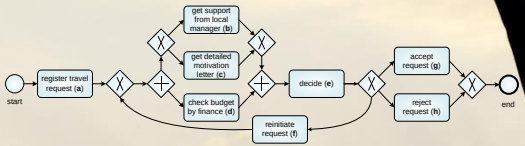
a d c e f b d e h

Play Out: Process model allows for many more scenarios



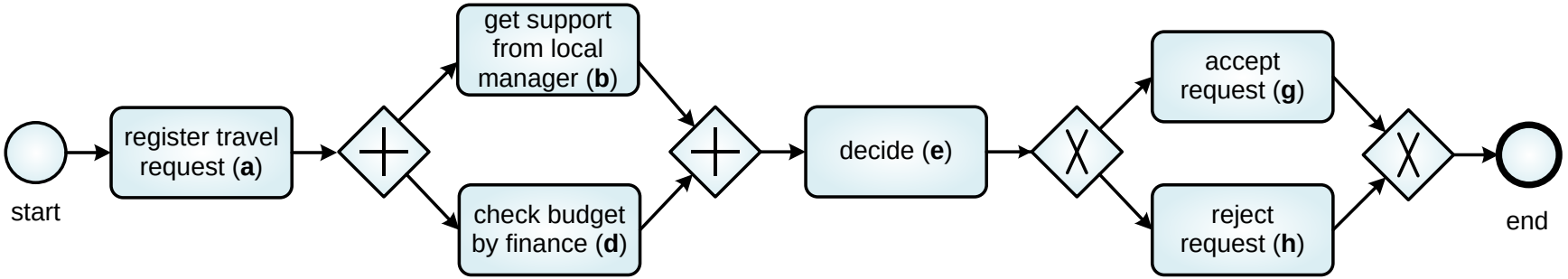
Play-In

Case	Activity	Timestamp	Resource
432	register travel request (a)	18-3-2014:9.15	John
432	get support from local manager (b)	18-3-2014:9.25	Mary
432	check budget by finance (d)	19-3-2014:8.55	John
432	decide (e)	19-3-2014:9.36	Sue
432	accept request (g)	19-3-2014:9.48	Mary



Play In: Simple process allowing for 4 traces

abdeg
adbeg
adbeg
adbeh
abdeh
abdeh
abdeg
abdeh
abdeh
adbeh
adbeh
adbeh



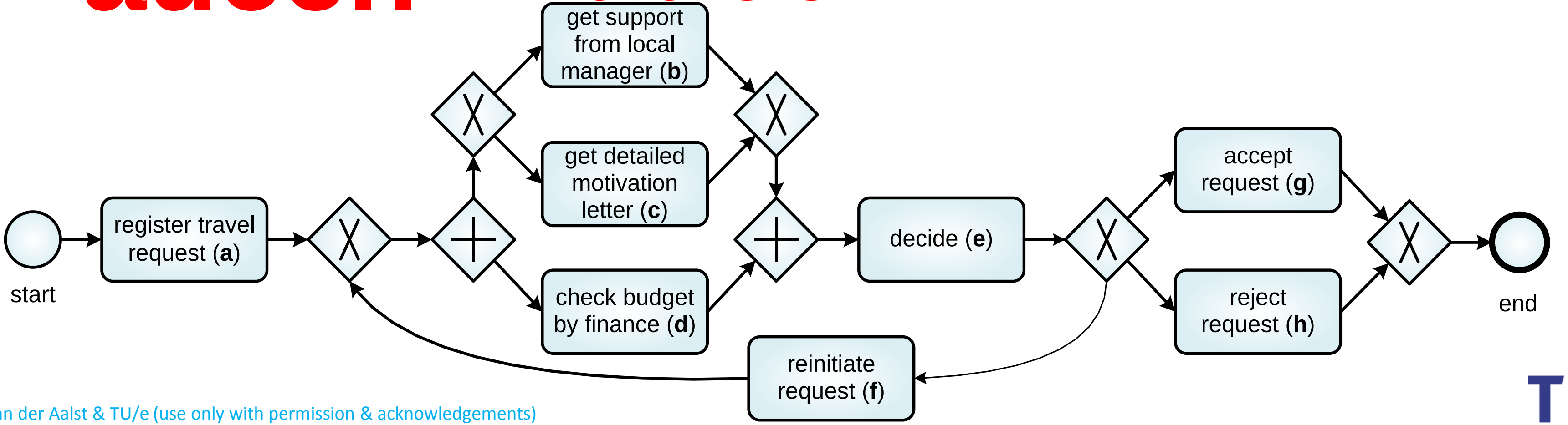
TU/e

Play In: Process allowing for more traces

abdeg

adcefcdefbdefbdeg

adceh




No modeling needed!

Question

abde

acde

adbf

adce

adbe

abdf

adbe

adcf

abdf

acdf

Create a process model that allows for the traces shown above.

TU/e

Answer

abde

acde

adbf

adce

adbe

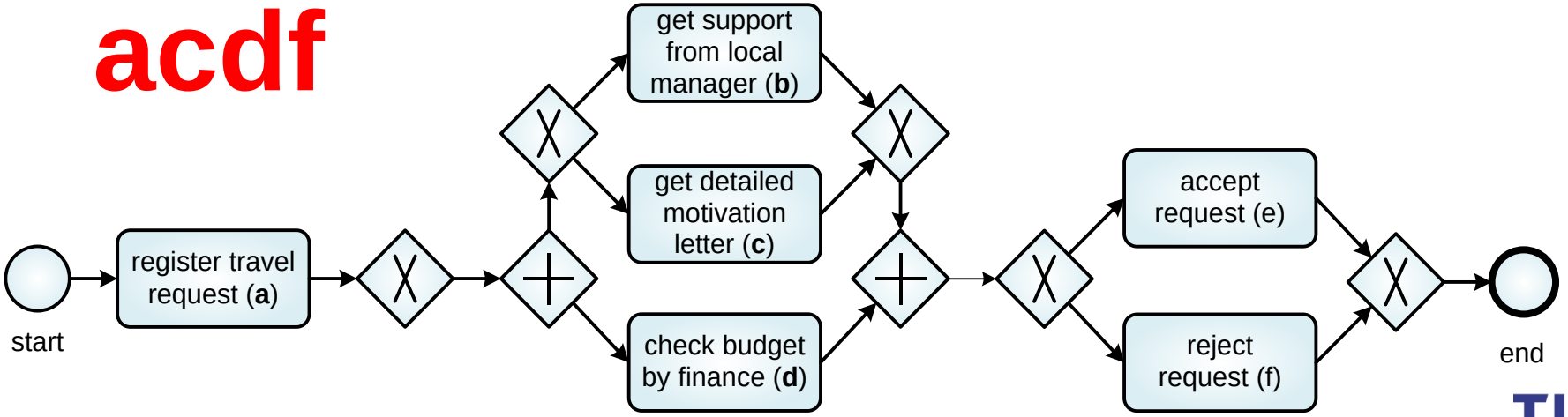
abdf

adbe

adcf

abdf

acdf

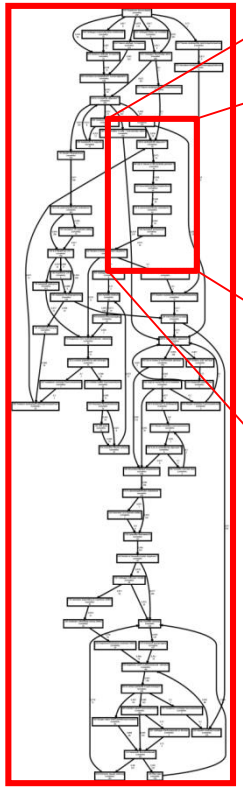


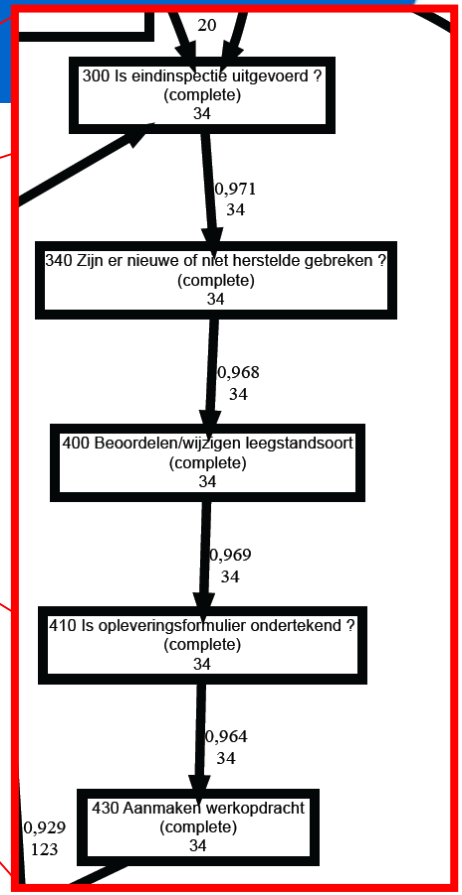
TU/e

Example Process Discovery

(Dutch housing agency, 208 cases, 5987 events)

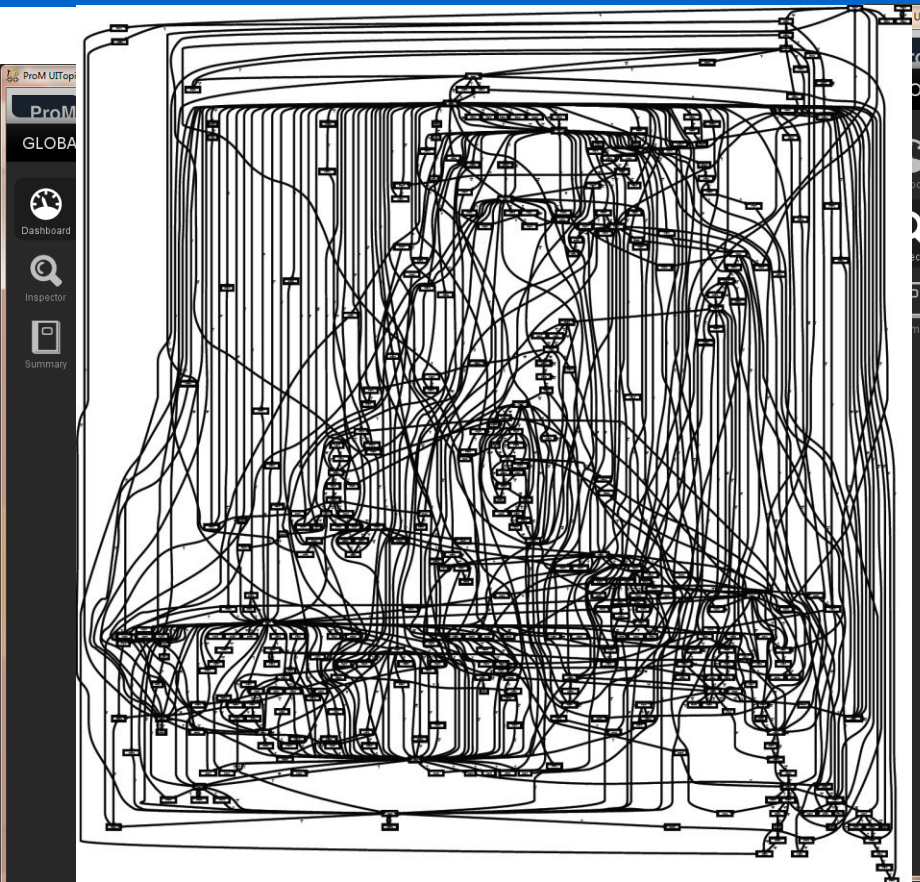
117315	110 Bepalen leegstandsoort	16.05.2007 14:06:23
117315	120 Plannen eindinspectie	16.05.2007 14:36:01
117315	130 Is het opleveringsformulier ondertekend ?	23.05.2007 09:41:40
117315	150 Is er sprake van ZAV ?	23.05.2007 09:41:51
117315	170 Aanpassen plattegrond	23.05.2007 11:57:18
117315	180 Aanpassen woningwaardering	23.05.2007 09:42:37
117315	190 Actualiseren huurprijs	23.05.2007 09:48:23
117315	200 Toewijzen woning/bedr.ruimte/gar/berg/park/ops	23.05.2007 09:48:29
117315	210 Registreren voorl. huurovereenkomst +afdrukken	10.09.2007 16:24:36
117315	220 Is contract getekend en geld ontvangen ?	11.09.2007 14:56:18
117315	240 Definitief maken Huurovereenkomst	31.03.2008 16:17:12
117315	250 Aanpassen factureerafspraak	09.09.2008 15:39:59
117315	260 After sales	09.09.2008 16:51:24
117315	270 Archiveren nieuwe verhuring	10.09.2008 07:52:08
117315	300 Is eindinspectie uitgevoerd ?	07.06.2007 14:47:04
117315	340 Zijn er nieuwe of niet herstelde gebreken ?	07.06.2007 14:47:06
117315	400 Beoordelen/wijzigen leegstandsoort	07.06.2007 14:51:16
117315	410 Is opleveringsformulier ondertekend ?	07.06.2007 14:51:26
117315	430 Aanmaken werkopdracht	11.06.2007 09:21:39
117315	440 Worden er bonussen/ kosten toegekend ?	11.06.2007 09:21:49
117315	460 Opstellen eindnota	08.08.2007 16:18:26
117315	470 Archiveren huuropzegging	09.08.2007 14:42:23
119763	010 Registreren huuropzegging	09.05.2007 11:19:14
119763	030 Vastleggen toekomstige adres	09.05.2007 12:25:01
119763	050 Inplannen afspraak 1e inspectie	09.05.2007 11:59:52
119763	060 Aanmaken bevestigingbrief / huuropzeggingform.	09.05.2007 12:31:57
119763	070 Is 1e inspectie uitgevoerd ?	16.05.2007 13:04:26
119763	100 Gereedmelden 1e insp. / Voorcalculatie maken	16.05.2007 13:43:39
119763	110 Bepalen leegstandsoort	16.05.2007 13:43:28
119763	120 Plannen eindinspectie	16.05.2007 13:42:58
119763	130 Is het opleveringsformulier ondertekend ?	16.05.2007 13:34:49
119763	150 Is er sprake van ZAV ?	16.05.2007 13:34:56

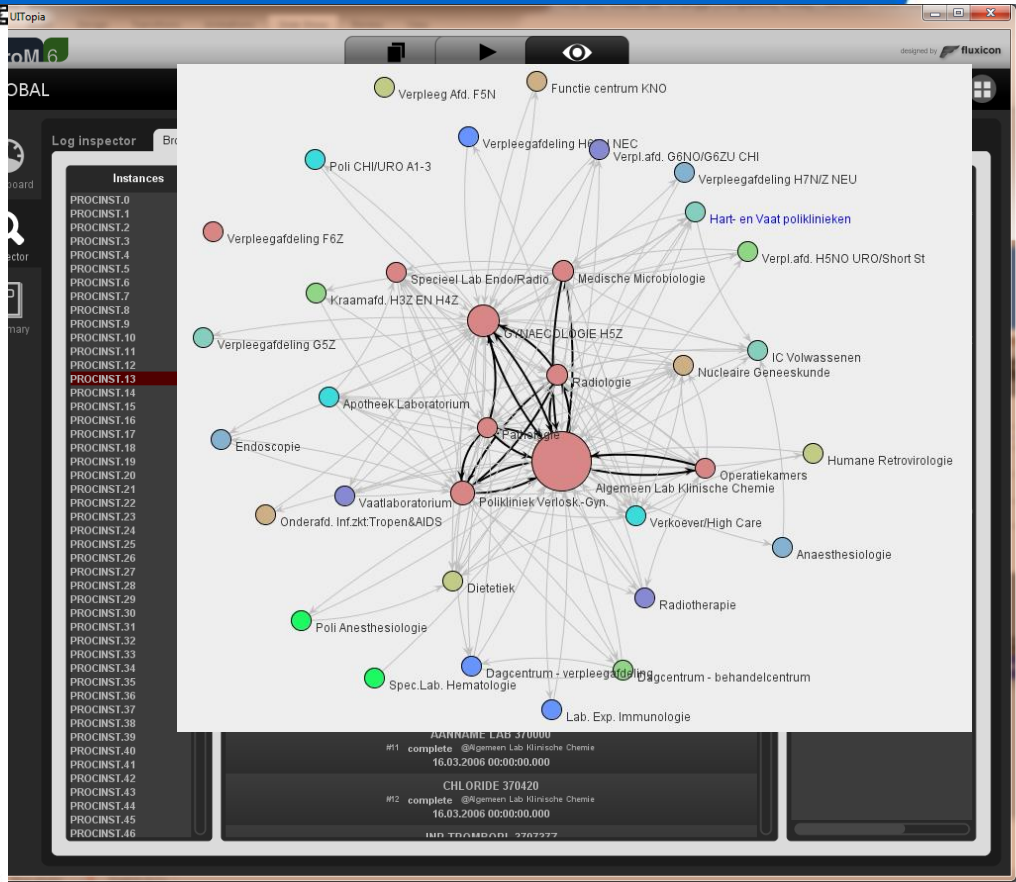




Example process discovery for hospital

(627 gynecological oncology patients, 24331 events)



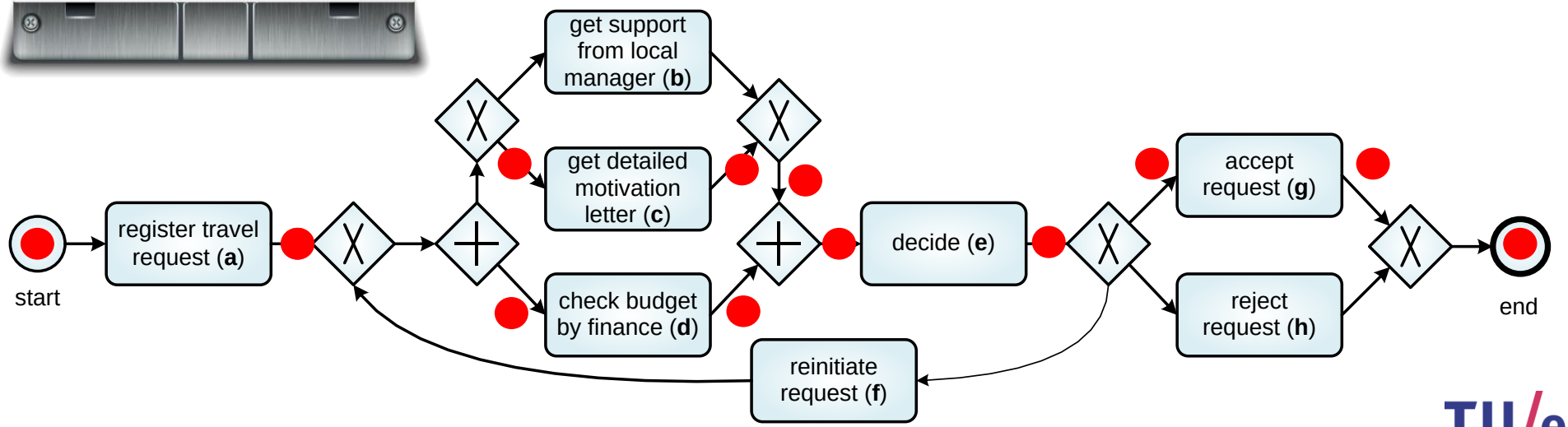


Replay

Case	Activity	Timestamp	Resource
432	register travel request (a)	18-3-2014:9.15	John
432	get support from local manager (b)	18-3-2014:9.25	Mary
432	check budget by finance (d)	19-3-2014:8.55	John
432	decide (e)	19-3-2014:9.36	Sue
432	accept request (g)	19-3-2014:9.48	Mary

Replay

a c d e g



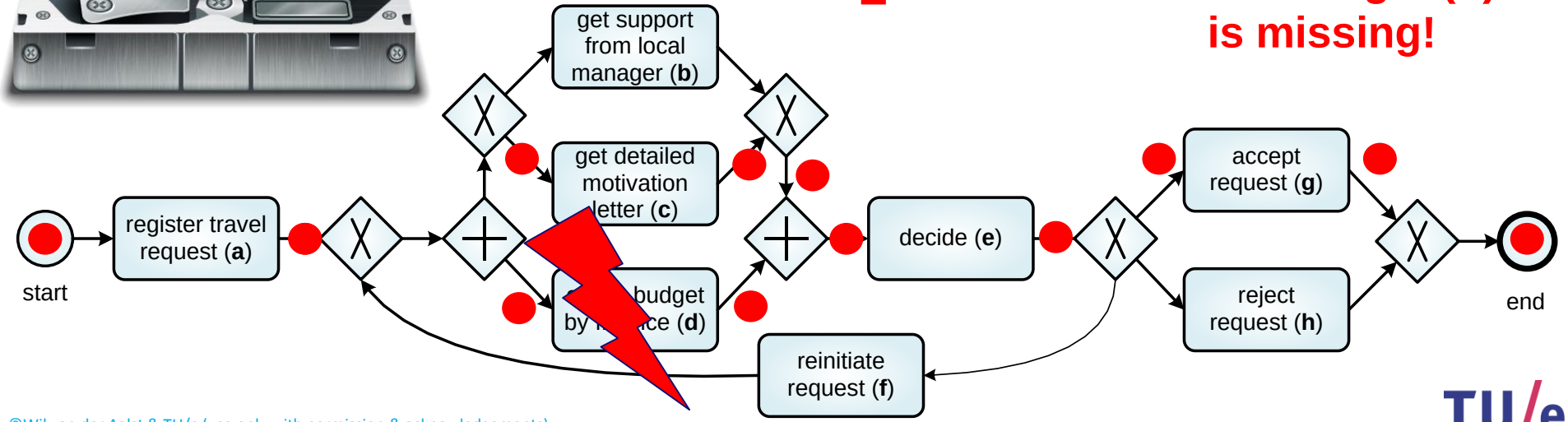
TU/e

Replay

a c e g

?

check budget (d) is missing!



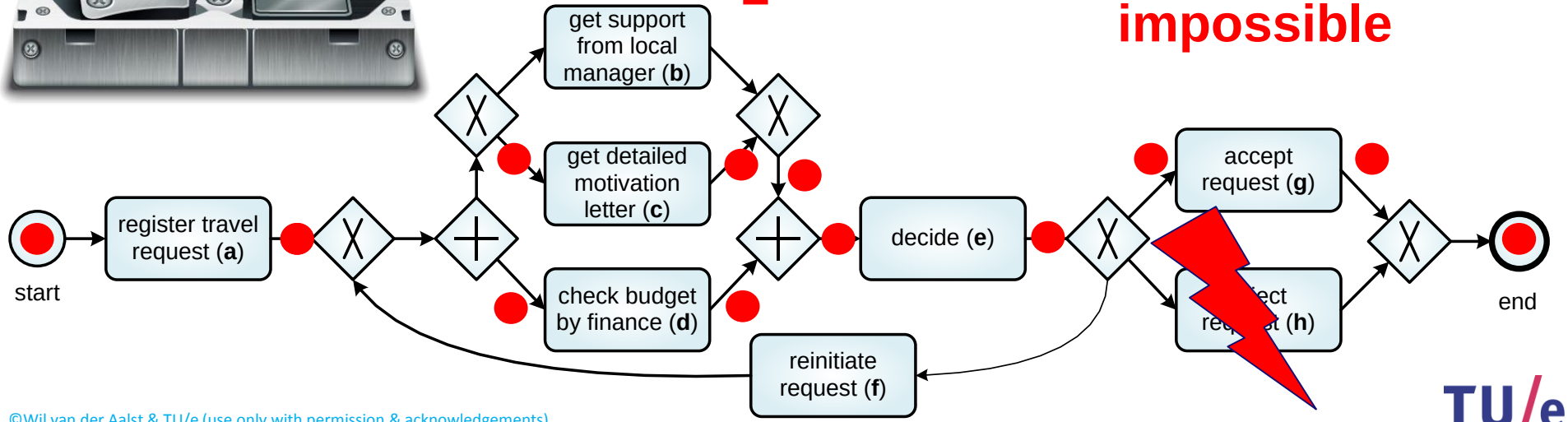
TU/e

Replay

a c h d e g

?

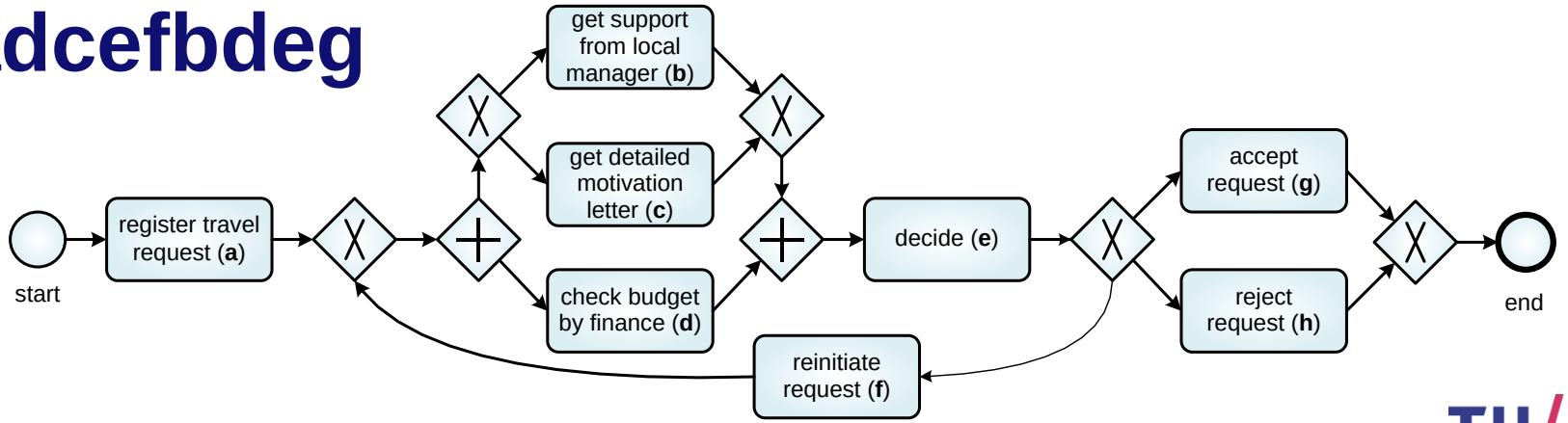
reject request (h) is impossible



TU/e

Question

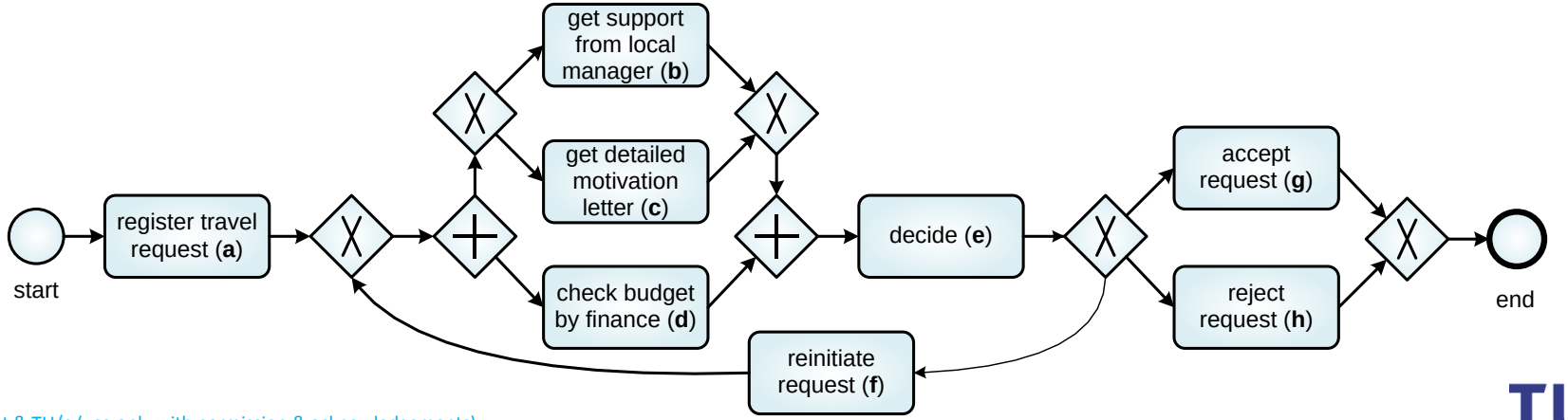
- **Which traces are possible in the model?**
  - **abdceh**
  - **aceh**
  - **adcefbdeg**



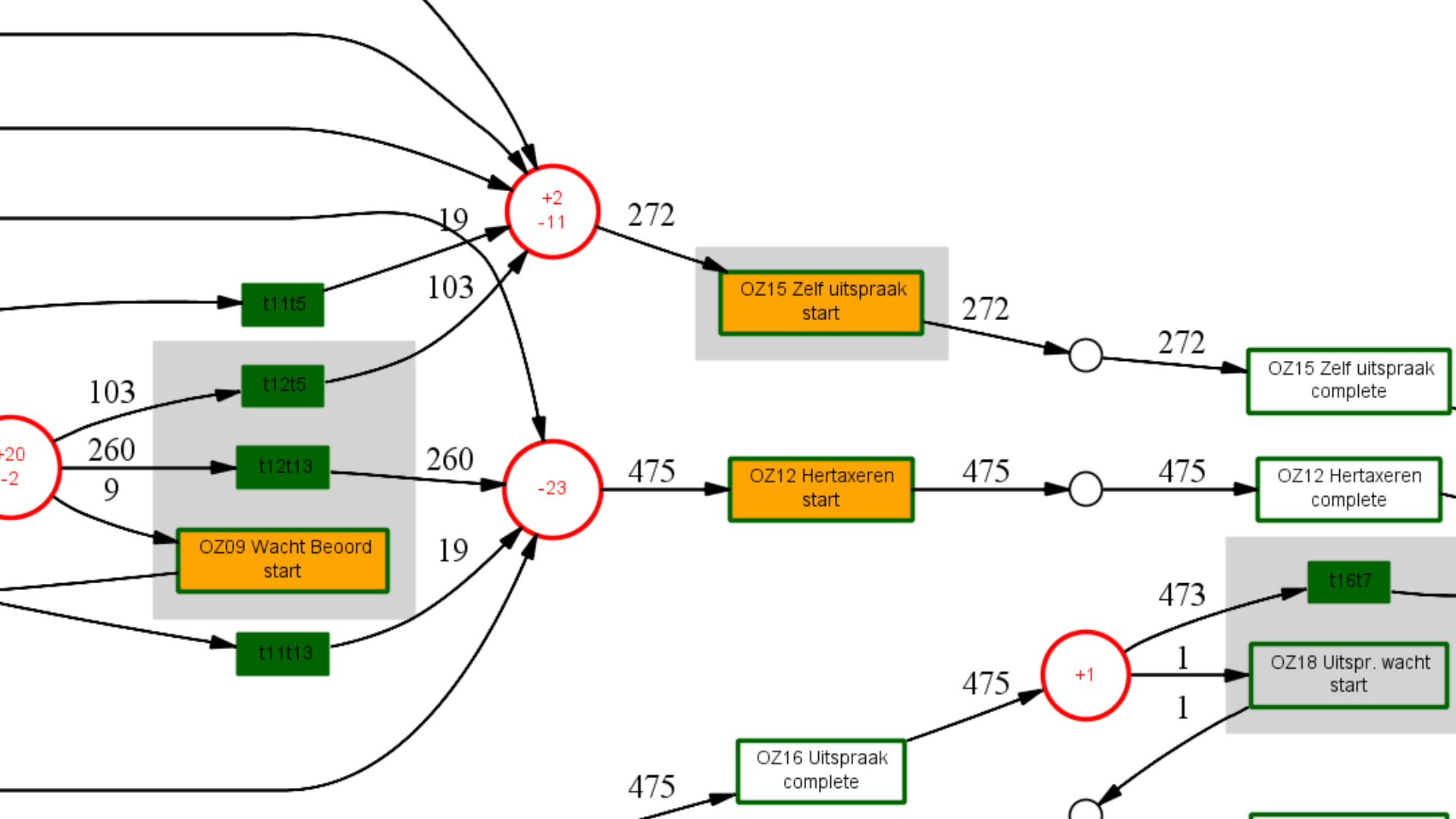
TU/e

Answer

- ✗ abdceh **No!** (cannot do both b and c)
- ✗ aceh **No!** (missing d)
- ✓ adcefbdeg **Yes!**

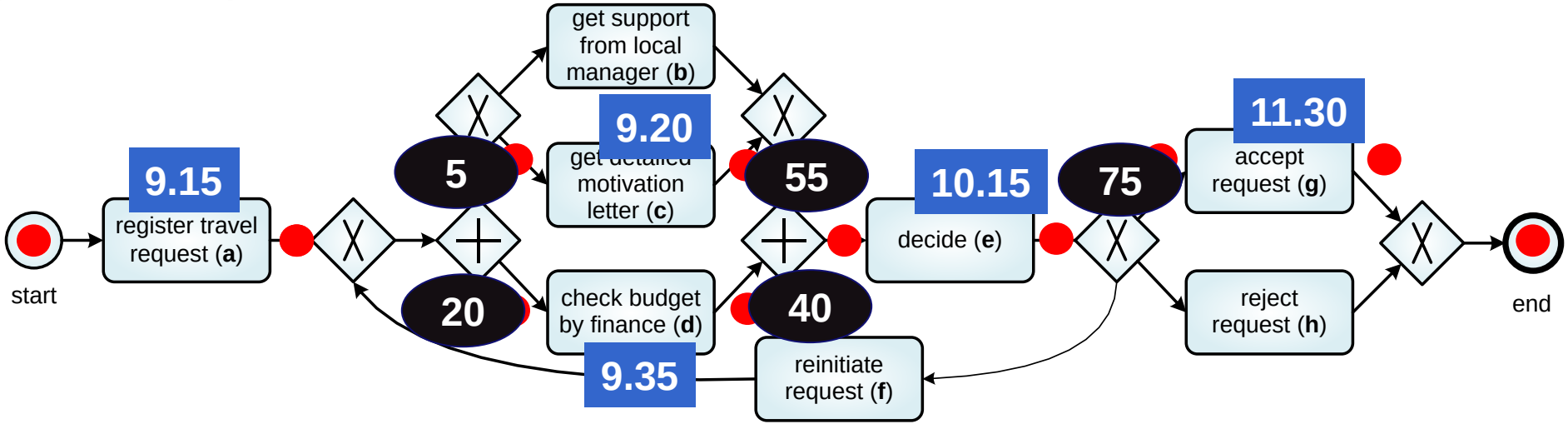


TU/e


+2
-11
19
103
272
t11t5
OZ15 Zelf uitspraak start
272
272
OZ15 Zelf uitspraak complete
103
t12t5
260
t12t13
260
9
-23
475
OZ12 Hertaxeren start
475
475
OZ12 Hertaxeren complete
OZ09 Wacht Beoord start
19
t11t13
473
t16t7
475
+1
1
OZ18 Uitspr. wacht start
1
OZ16 Uitspraak complete
475

Replay with timestamps

$a^{9.15}$ $c^{9.20}$ $d^{9.35}$ $e^{10.15}$ $g^{11.30}$



TU/e

Replay with timestamps for many traces

frequencies of paths

frequencies of activities

waiting times and other delays between activities

durations of activities

start

register travel request (**a**)

get support from local manager (**b**)

get detailed motivation letter (**c**)

check budget by finance (**d**)

decide (**e**)

accept request (**g**)

reject request (**h**)

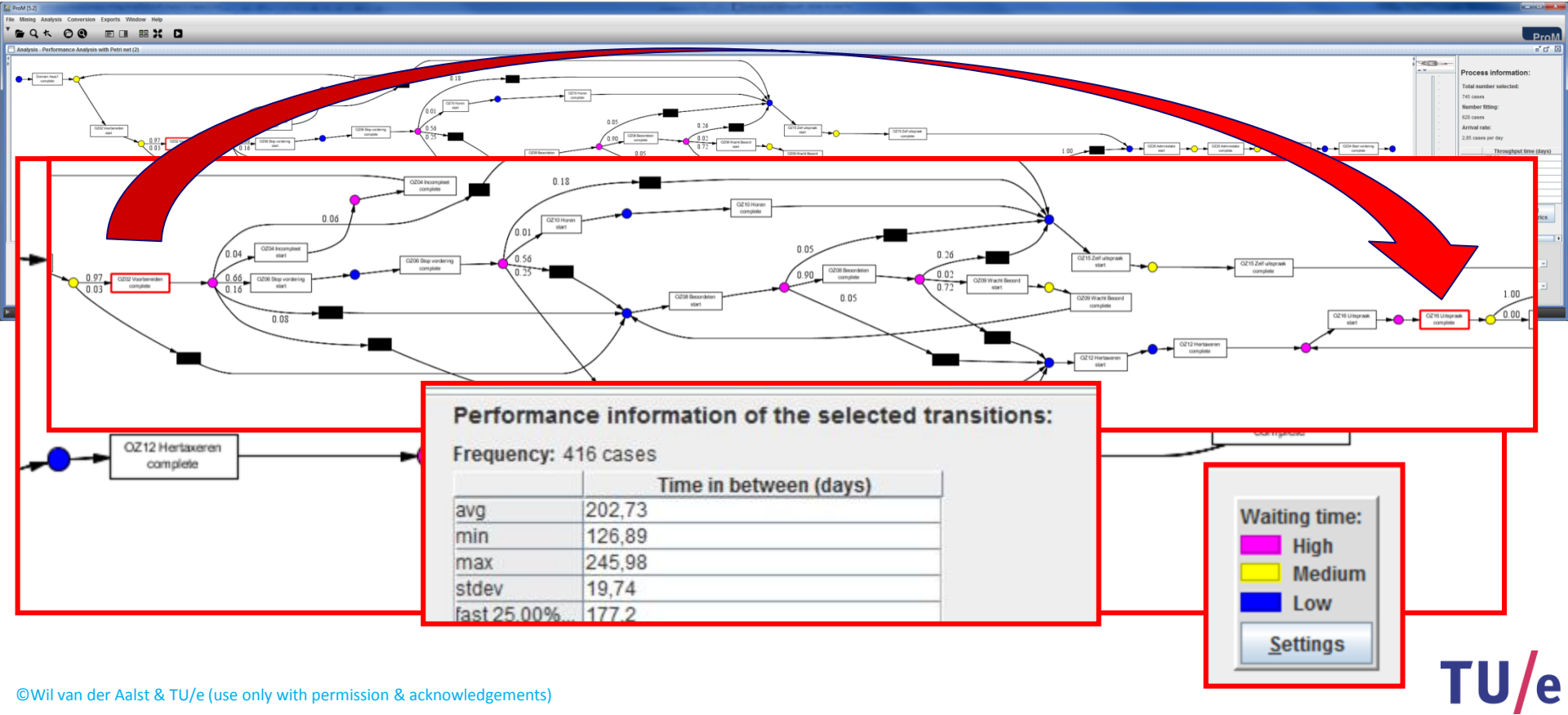
reinitiate request (**f**)

end

TU/e

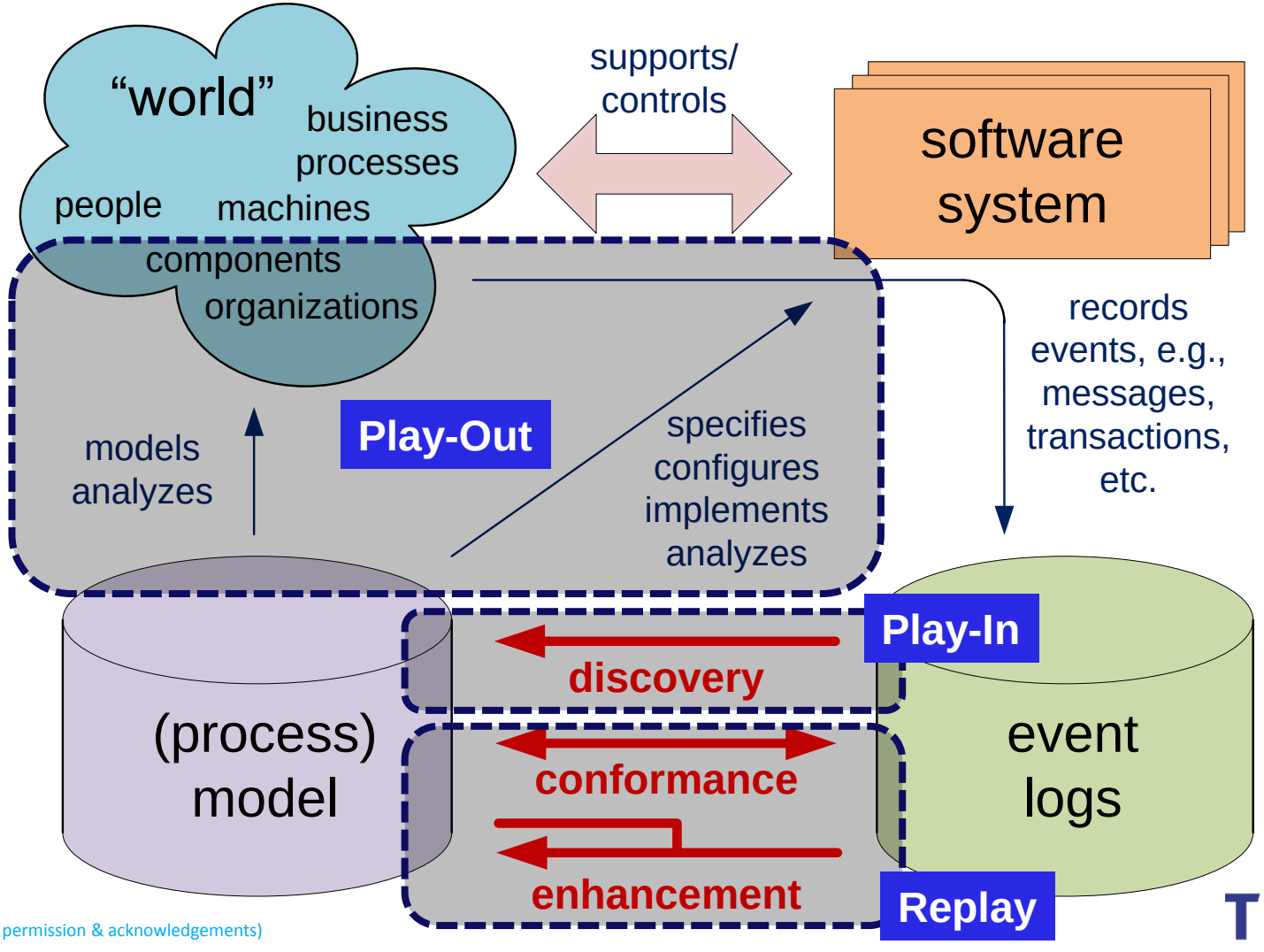
Performance Analysis Using Replay

(WOZ objections Dutch municipality, 745 objections, 9583 event, f= 0.988)



TU/e

Overview



TU/e

Part I: Preliminaries

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

Part II: From Event Logs to Process Models

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

Part III: Beyond Process Discovery

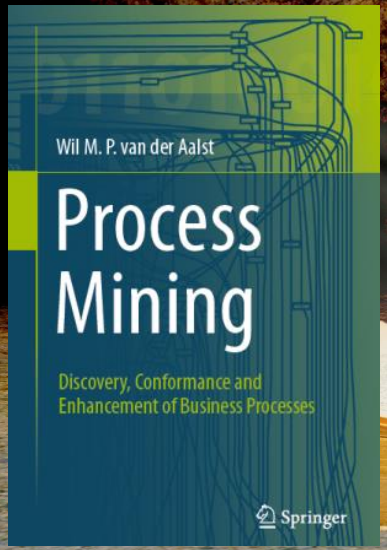
- Chapter 7: Conformance Checking
- Chapter 8: Mining Additional Perspectives
- Chapter 9: Operational Support

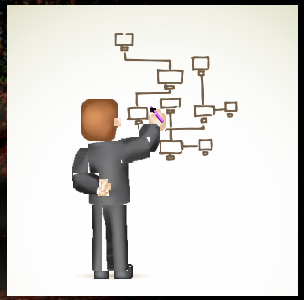
Part IV: Putting Process Mining to Work

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

Part V: Reflection

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





TU/e