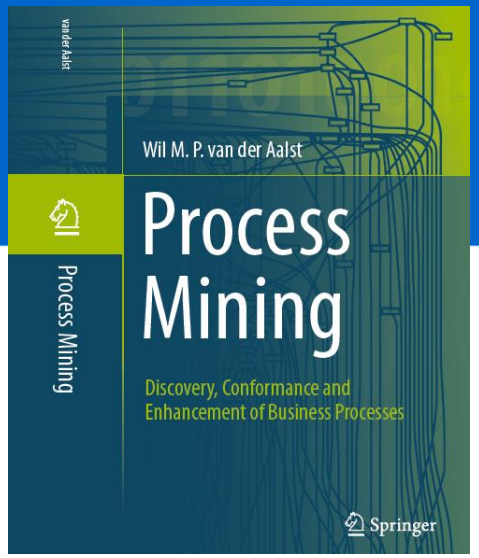
Bits&Chips 2012 Embedded Systems
Congrescentrum Brabanthallen, 's-Hertogenbosch, 8-11-2012

Process Mining

How are my systems used and when do they fail?

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



TU/e Technische Universiteit Eindhoven University of Technology

Where innovation starts

1. Desire Lines in Big Data
2. On the different roles of (process) models ...
3. Process Mining
4. Example Philips Healthcare
5. Conclusion

Process Mining

supports/
controls

software
system

records
events, e.g.,
messages,
transactions,
etc.

models
analyzes

specifies
configures
implements
analyzes

(process)
model

discovery

conformance

enhancement

event
logs

Desire Lines in Big Data

Big Data

"All of the world's music can be stored on a $600 disk drive."

"Enterprises globally stored more than 7 exabytes of new data on disk drives in 2010, while consumers stored more than 6 exabytes of new data on devices such as PCs and notebooks."

"Indeed, we are generating so much data today that it is physically impossible to store it all. Health care providers, for instance, discard 90 percent of the data that they generate."

Source: "Big Data: The Next Frontier for Innovation, Competition, and Productivity" McKinsey Global Institute, 2011.

What Happens in an **Internet Minute?**

And **Future Growth** is **Staggering**

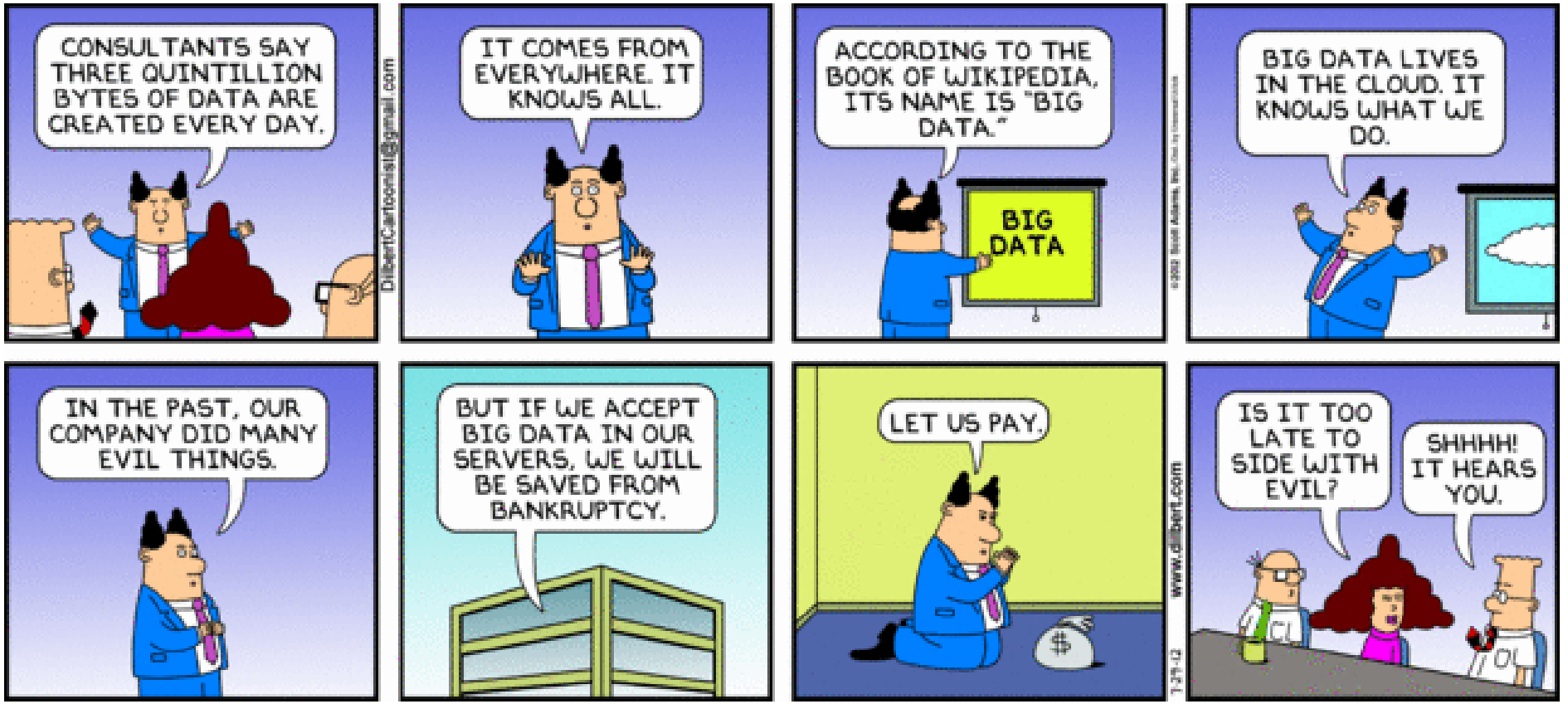
Today, the number of **networked devices** = the global population

By **2015**, the number of **networked devices** = **2x** the global population

In **2015**, it would take you **5 years**

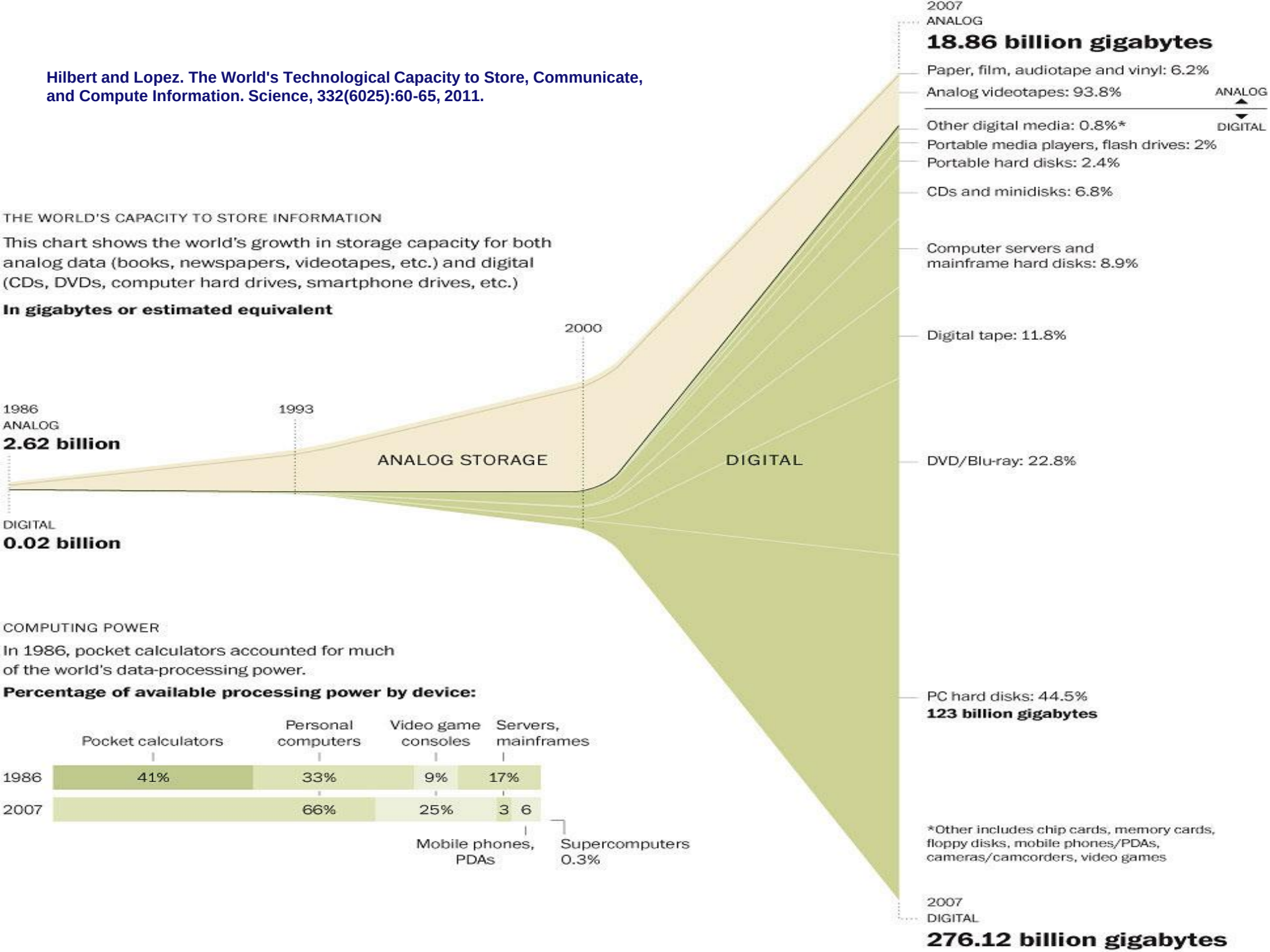
to view all video crossing IP networks each **second**

Big Data: Even Dilbert and the "pointy-haired boss" know about it …



http://dilbert.com/strips/comic/2012-07-29/

Hilbert and Lopez. The World's Technological Capacity to Store, Communicate, and Compute Information. Science, 332(6025):60-65, 2011.

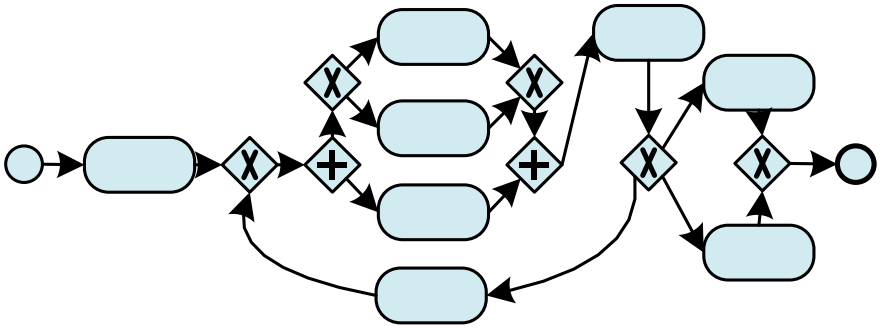



www.olifantenpaadjes.nl
01001100110101010
0100110101010
60

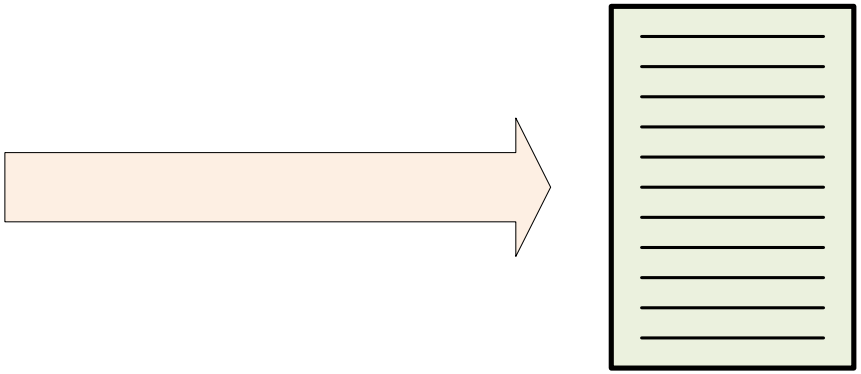

0101010101010

On the different roles of (process) models …

Play-Out

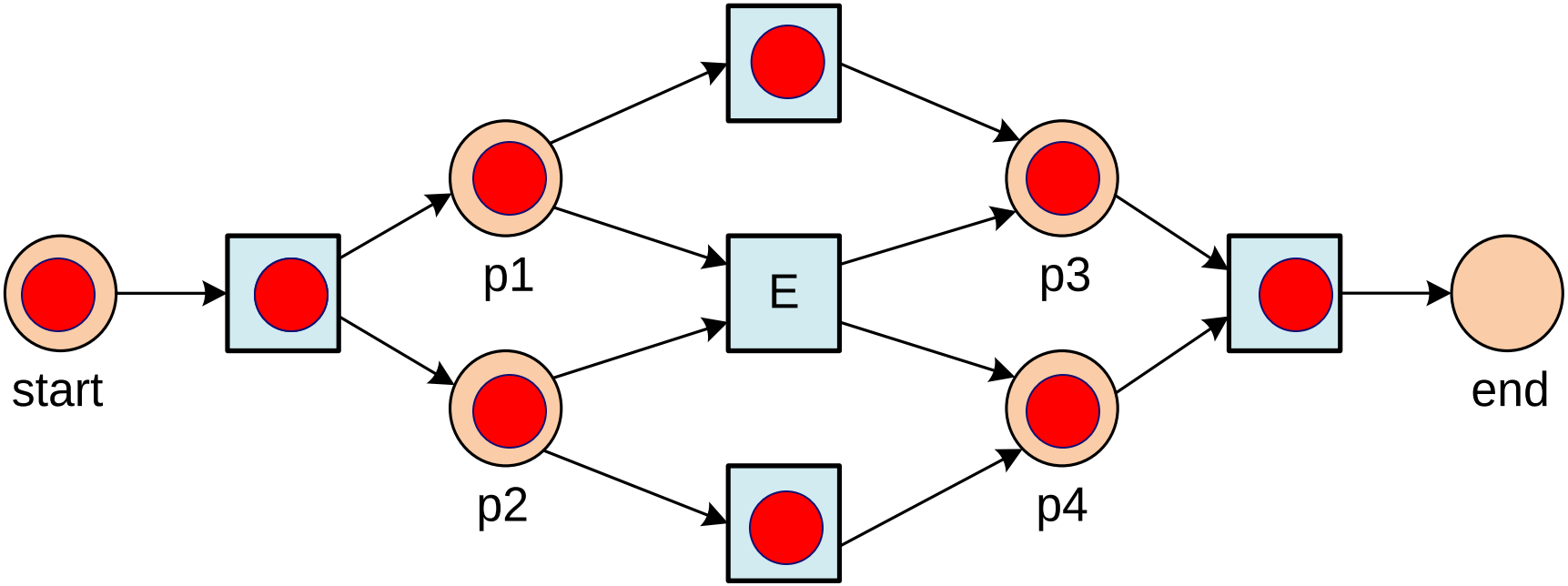


process model



event log

Play-Out (Classical use of models)

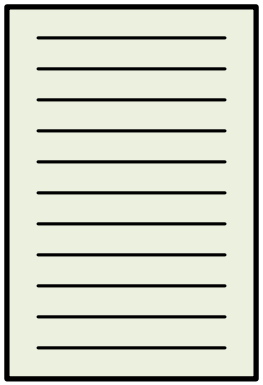


A B C D A E D A E D

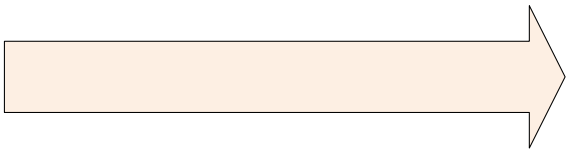
A B C D A C B D

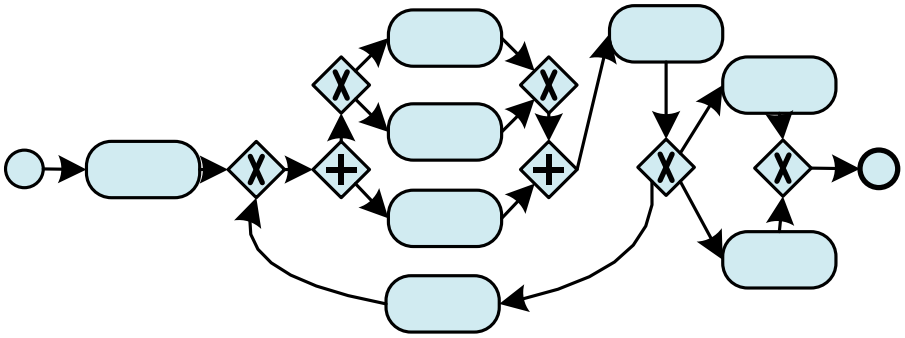
A C B D A E D A C B D

Play-In



event log





process model

Play-In

A B C D

A E D

A E D

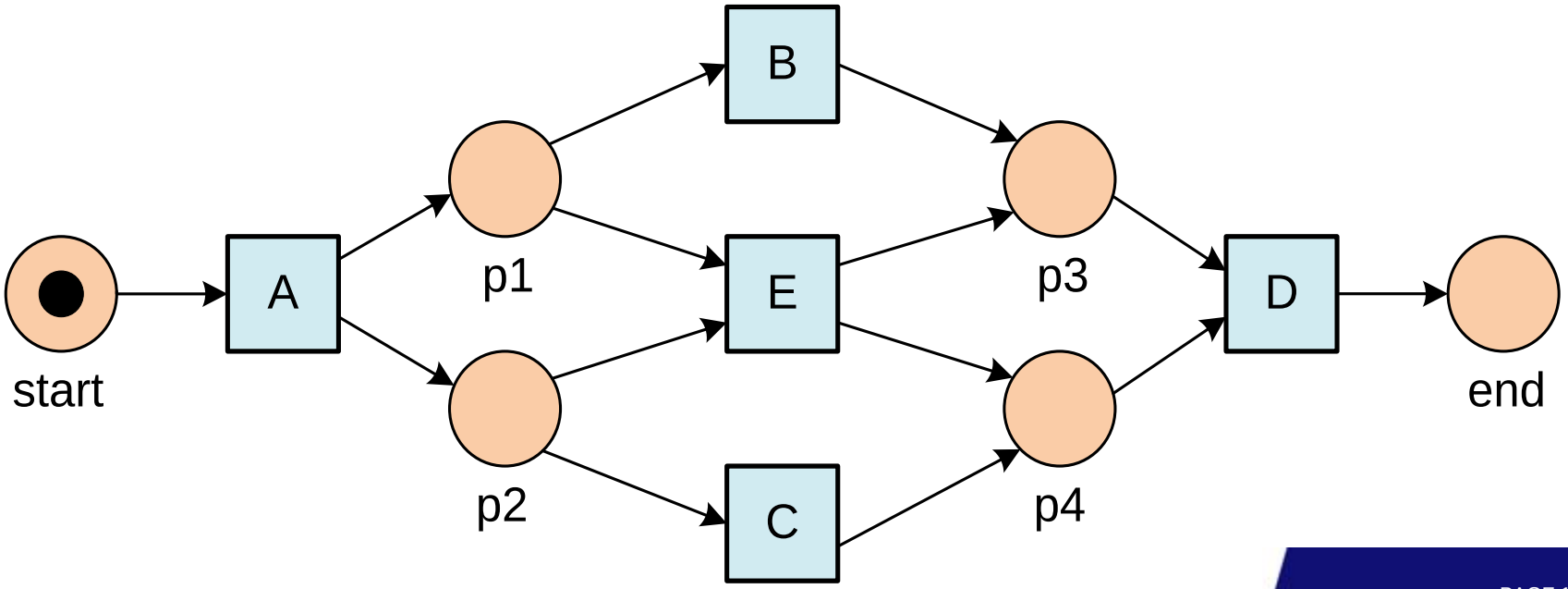
A B C D

A C B D

A C B D

A E D

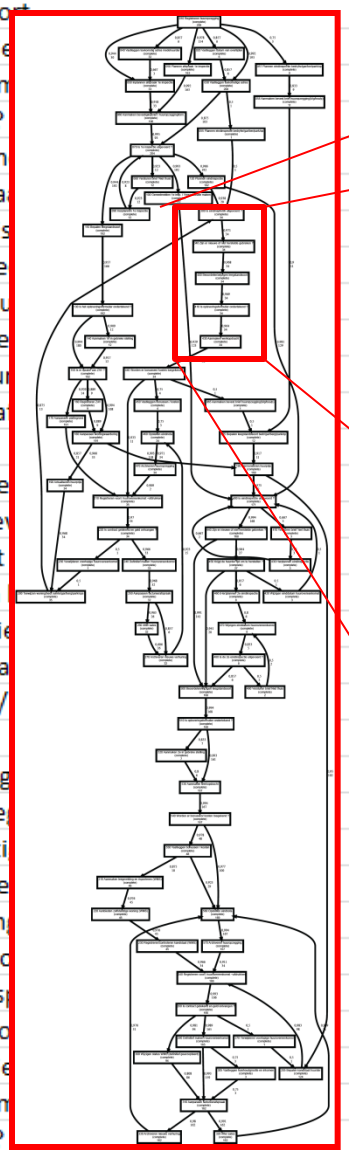
A C B D

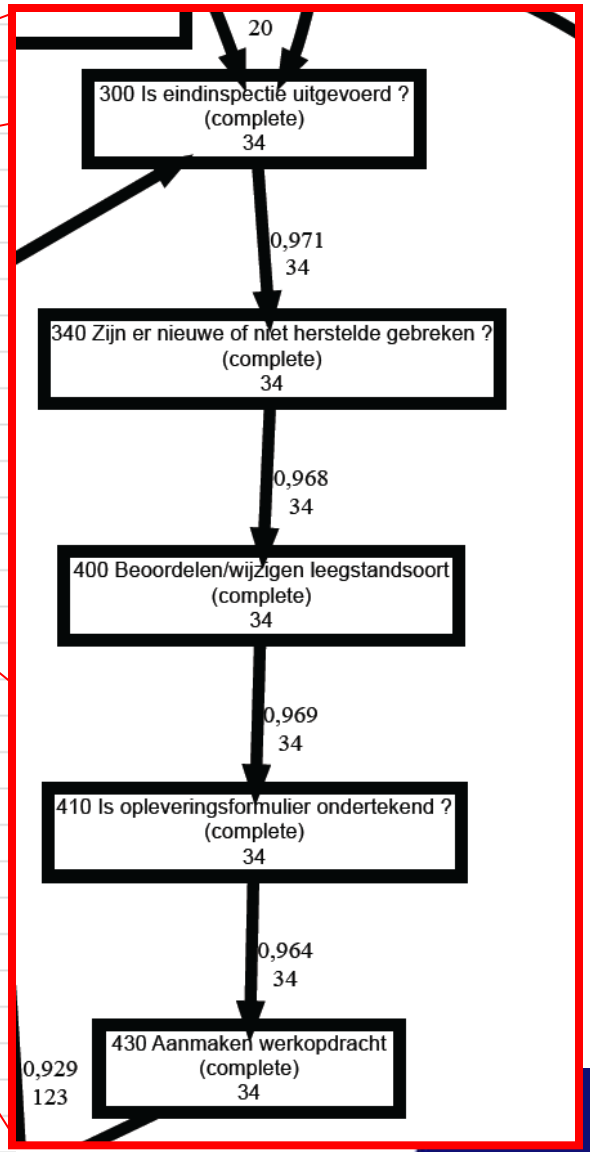


Example Process Discovery

(Vestia, Dutch housing agency, 208 cases, 5987 events)

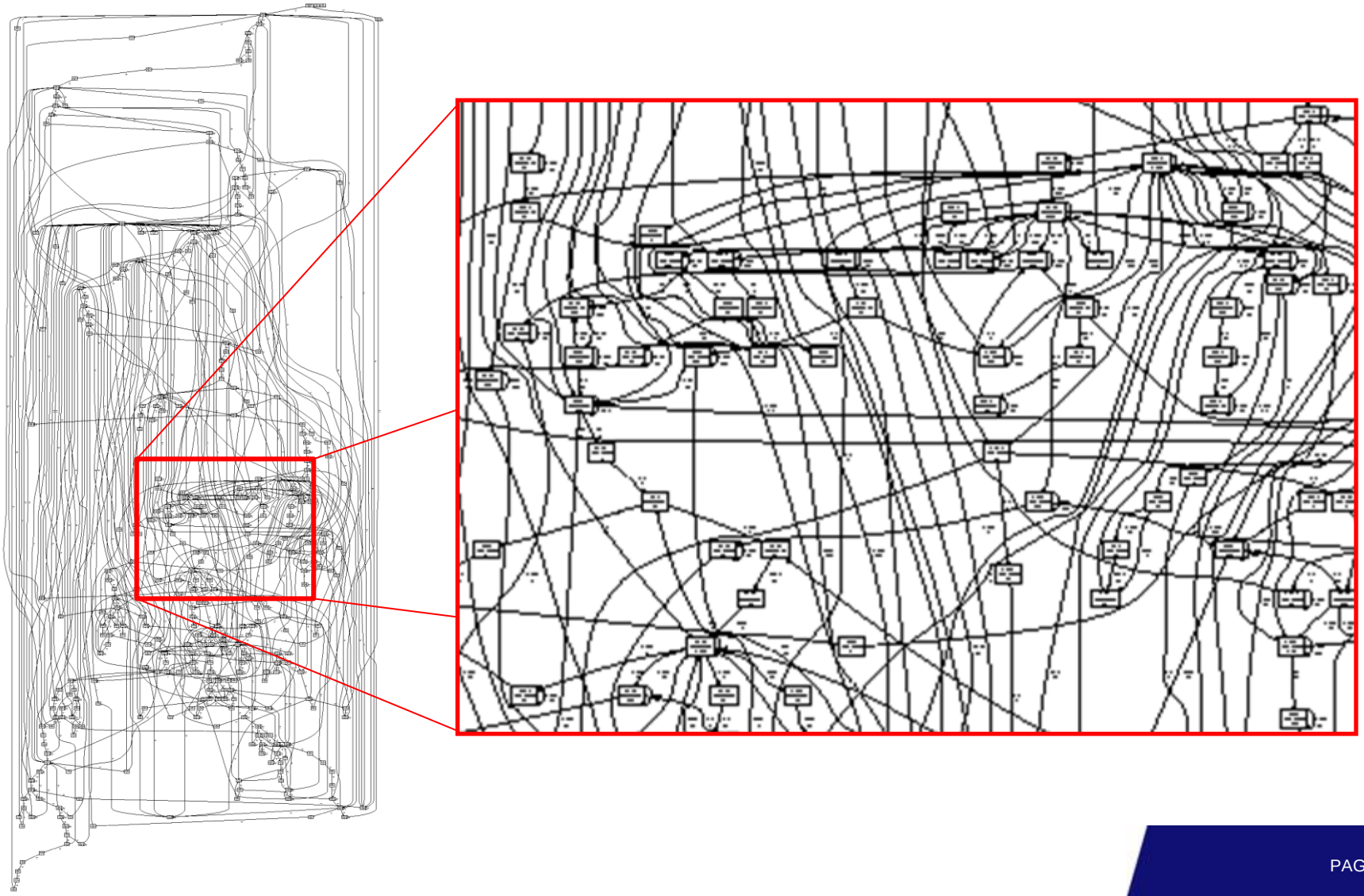
117315	110 Bepalen leegstandsoo	16.05.2007 14:06:23
117315	120 Plannen eindinspectie	16.05.2007 14:36:01
117315	130 Is het opleveringsform	23.05.2007 09:41:40
117315	150 Is er sprake van ZAV ?	23.05.2007 09:41:51
117315	170 Aanpassen plattegron	23.05.2007 11:57:18
117315	180 Aanpassen woningwa	23.05.2007 09:42:37
117315	190 Actualiseren huurprijs	23.05.2007 09:48:23
117315	200 Toewijzen woning/be	23.05.2007 09:48:29
117315	210 Registreren voorl. huu	10.09.2007 16:24:36
117315	220 Is contract getekend e	11.09.2007 14:56:18
117315	240 Definitief maken Huur	31.03.2008 16:17:12
117315	250 Aanpassen factureera	09.09.2008 15:39:59
117315	260 After sales	09.09.2008 16:51:24
117315	270 Archiveren nieuwe ve	10.09.2008 07:52:08
117315	300 Is eindinspectie uitgev	07.06.2007 14:47:04
117315	340 Zijn er nieuwe of niet	07.06.2007 14:47:06
117315	400 Beoordelen/wijzigen l	07.06.2007 14:51:16
117315	410 Is opleveringsformulie	07.06.2007 14:51:26
117315	430 Aanmaken werkopdra	11.06.2007 09:21:39
117315	440 Worden er bonussen/	11.06.2007 09:21:49
117315	460 Opstellen eindnota	08.08.2007 16:18:26
117315	470 Archiveren huuropzeg	09.08.2007 14:42:23
119763	010 Registreren huuropze	09.05.2007 11:19:14
119763	030 Vastleggen toekomsti	09.05.2007 12:25:01
119763	050 Inplannen afspraak 1e	09.05.2007 11:59:52
119763	060 Aanmaken bevestigin	09.05.2007 12:31:57
119763	070 Is 1e inspectie uitgevo	16.05.2007 13:04:26
119763	100 Gereedmelden 1e insp	16.05.2007 13:43:39
119763	110 Bepalen leegstandsoo	16.05.2007 13:43:28
119763	120 Plannen eindinspectie	16.05.2007 13:42:58
119763	130 Is het opleveringsform	16.05.2007 13:34:49
119763	150 Is er sprake van ZAV ?	16.05.2007 13:34:56





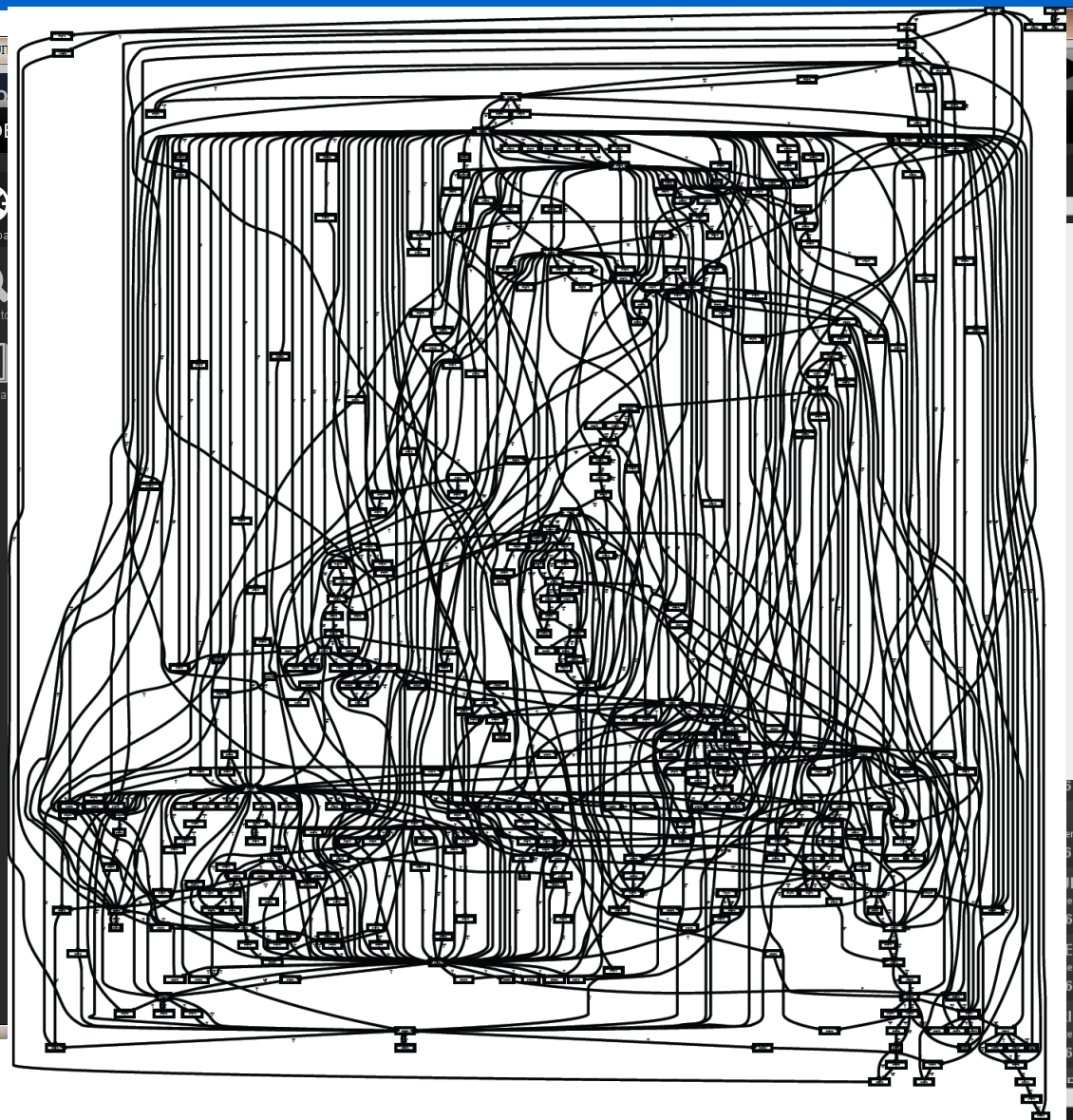
Example Process Discovery

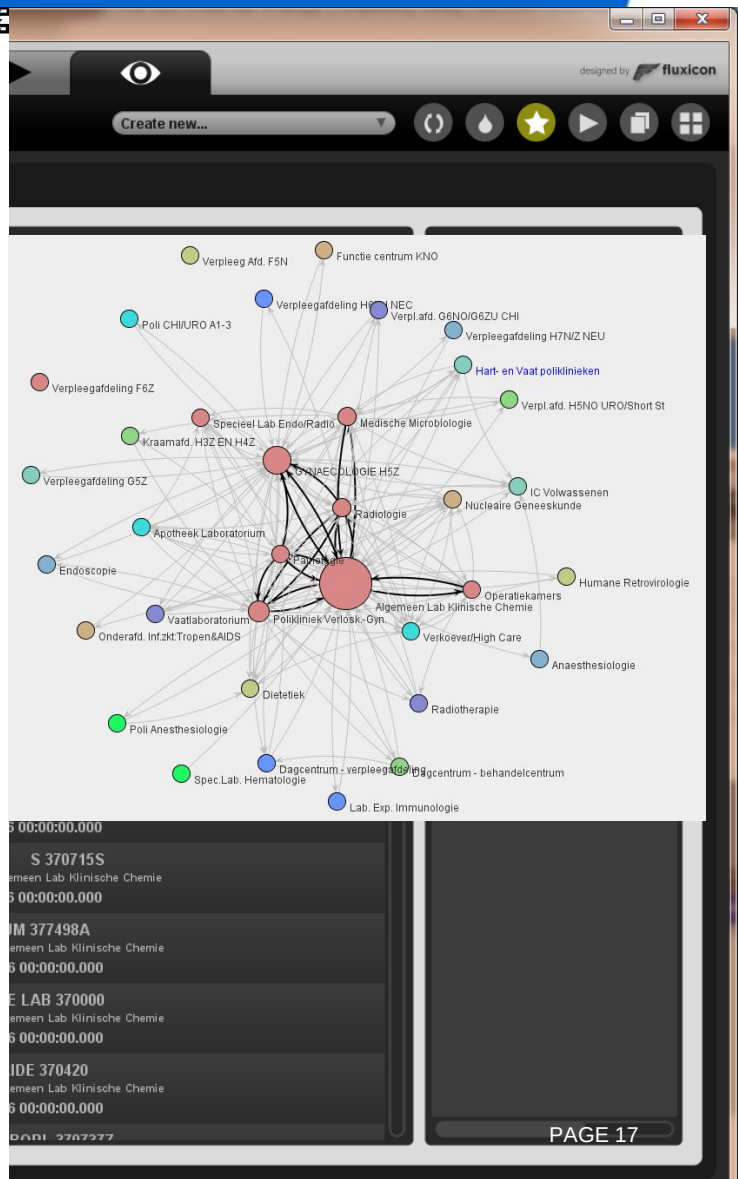
(ASML, test process lithography systems, 154966 events)



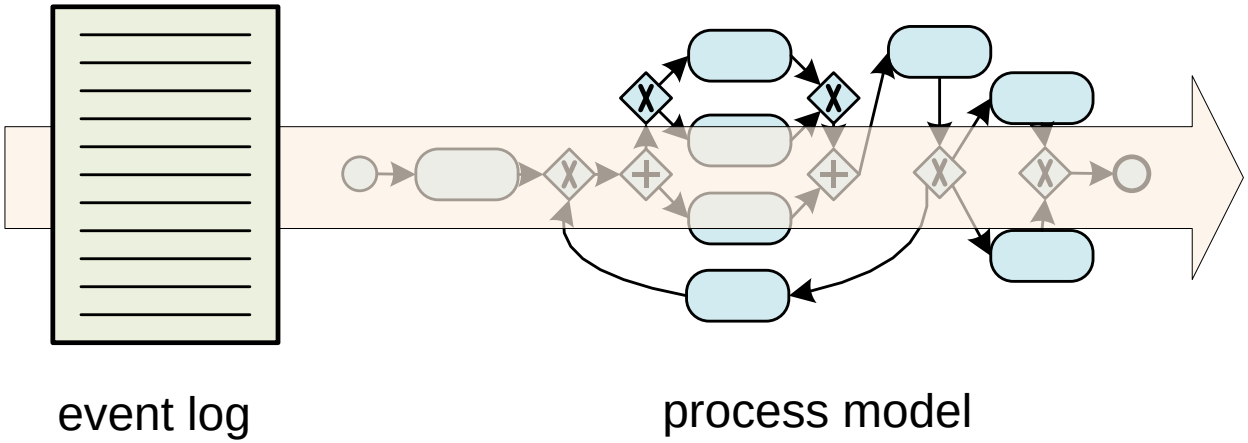
Example Process Discovery

(AMC, 627 gynecological oncology patients, 24331 events)





Replay



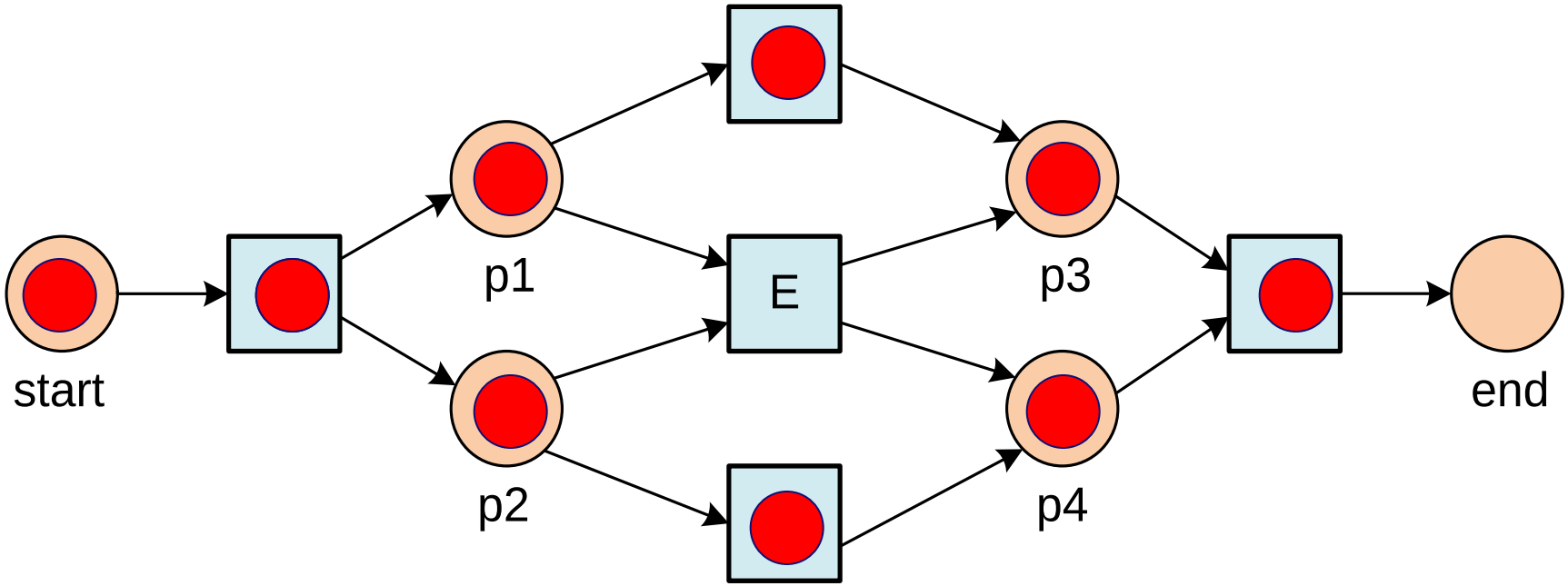
event log

process model

- extended model showing times, frequencies, etc.
- diagnostics
- predictions
- recommendations

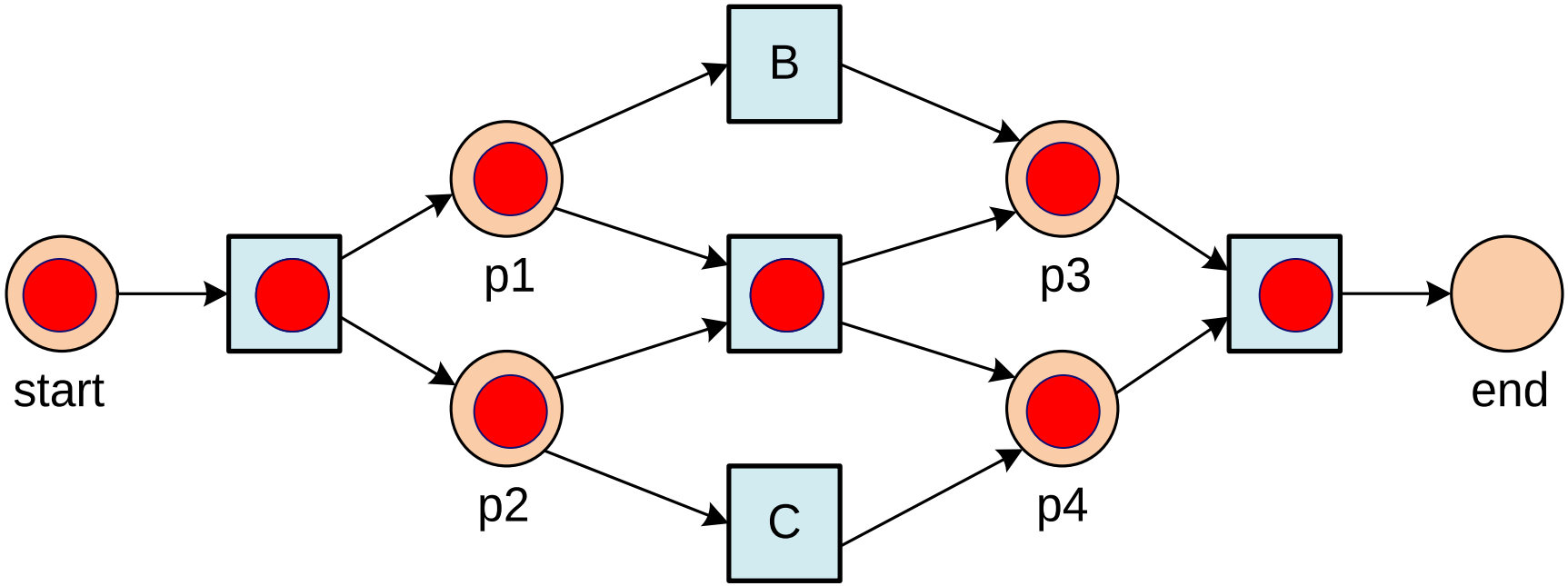
Replay

A B C D



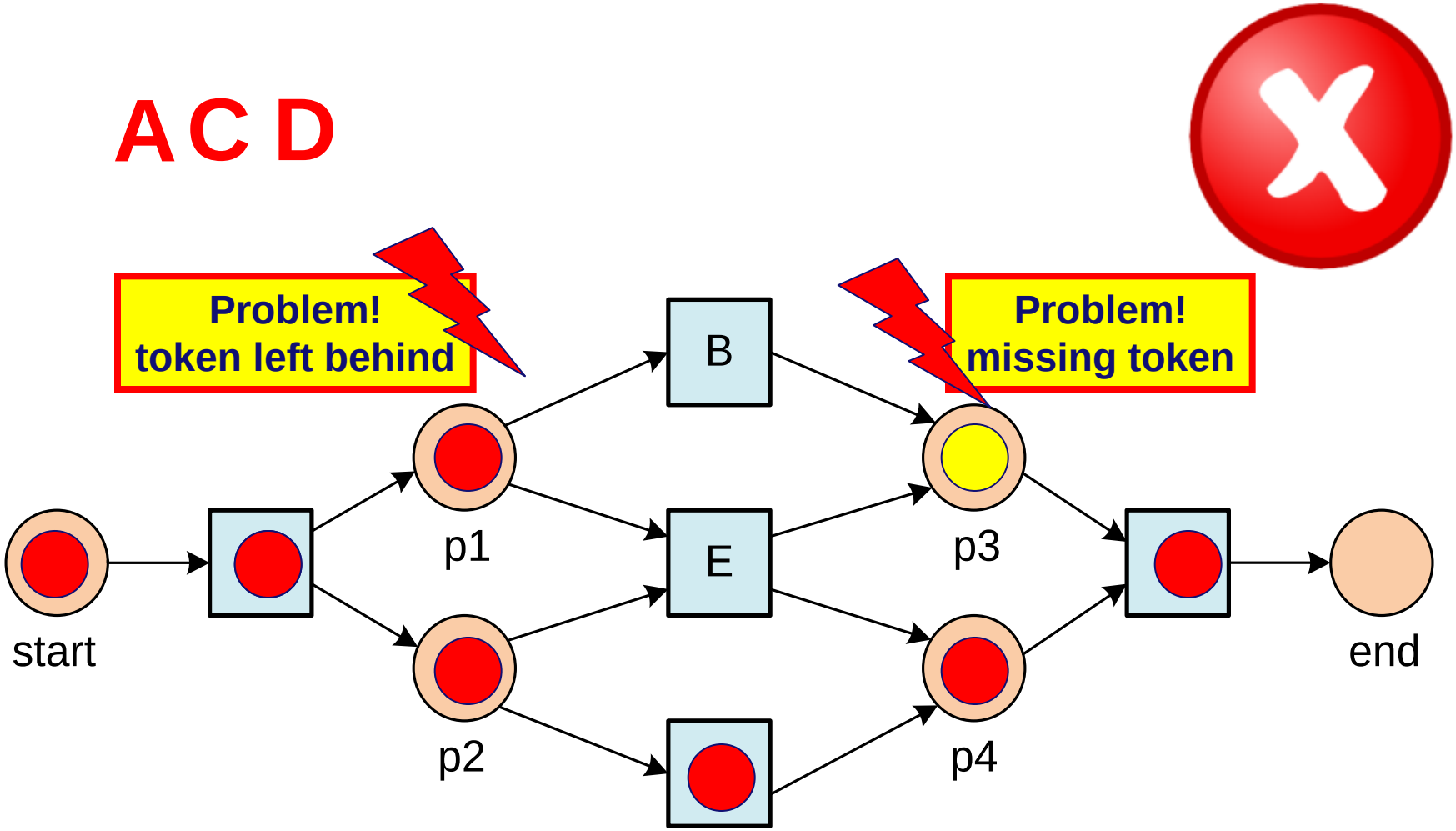
Replay

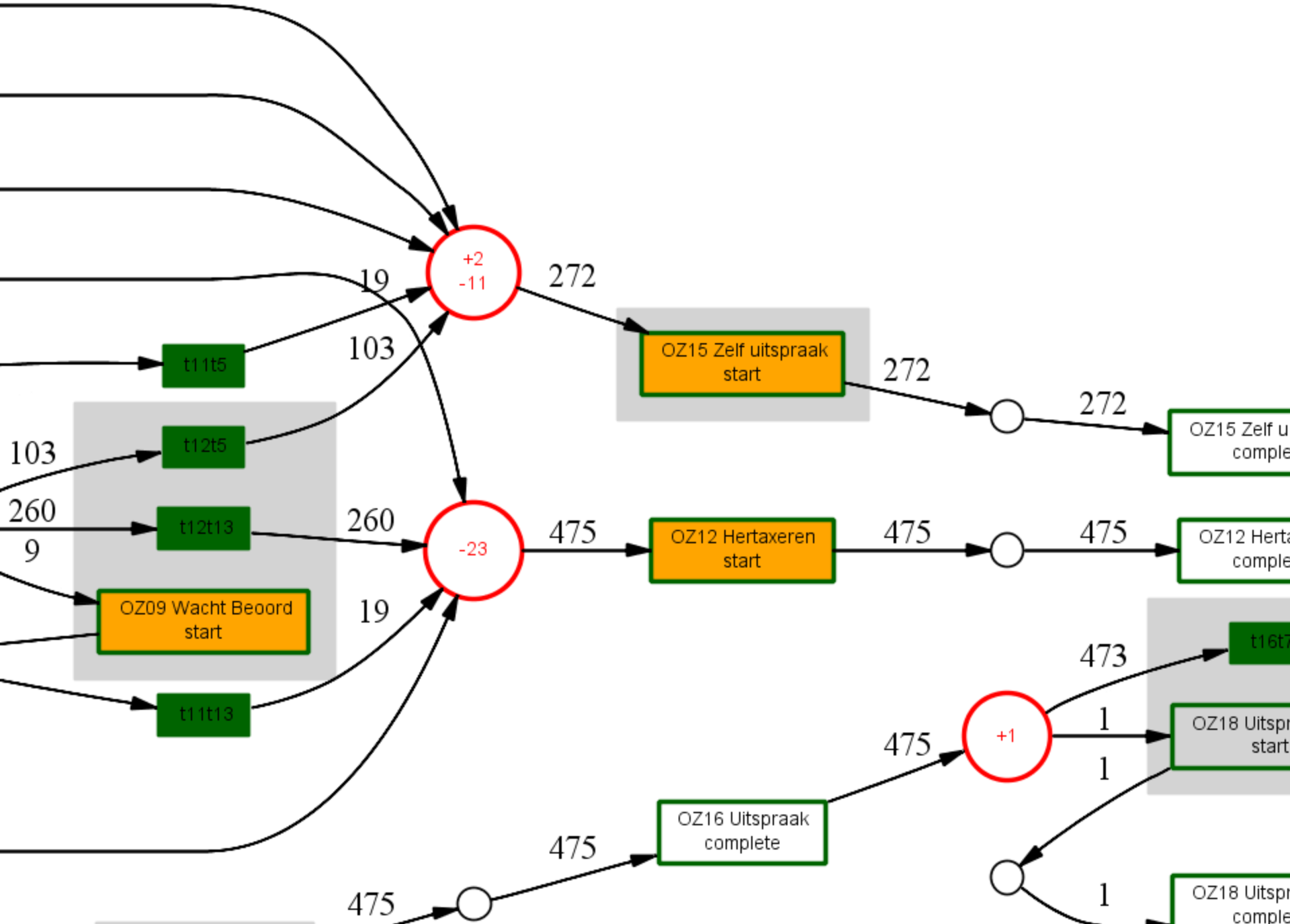
A E D



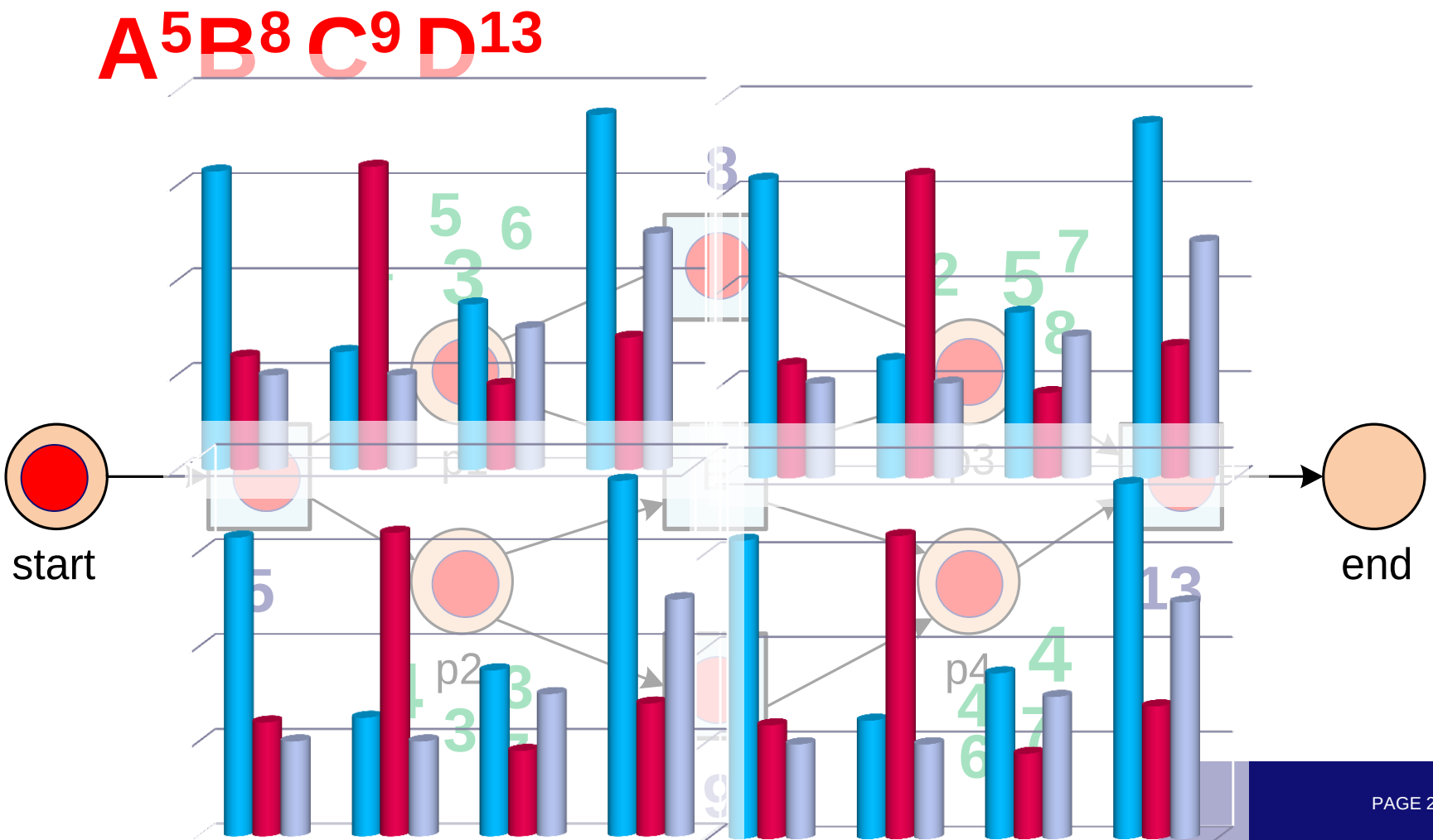
Replay can detect problems

A C D



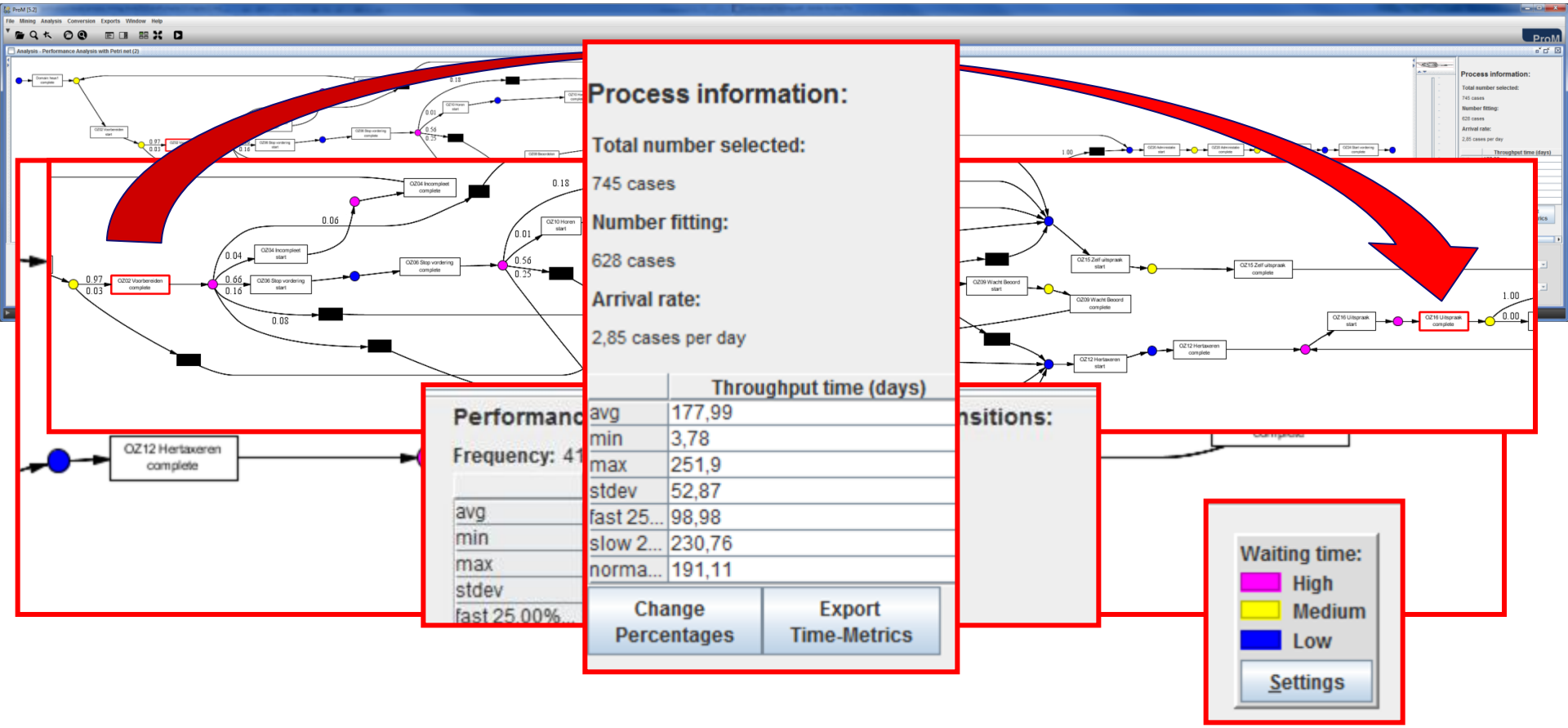

+2
-11
272
19
103
t11t5
t12t5
103
260
t12t13
260
9
OZ09 Wacht Beoord start
19
t11t13
-23
475
OZ15 Zelf uitspraak start
272
272
OZ15 Zelf u
comple
OZ12 Hertaxeren start
475
475
475
OZ12 Herta
comple
473
1
1
+1
475
OZ16 Uitspraak complete
475
475
1
OZ18 Uitspr
start
OZ18 Uitspr
comple

Replay can extract timing information



Performance Analysis Using Replay

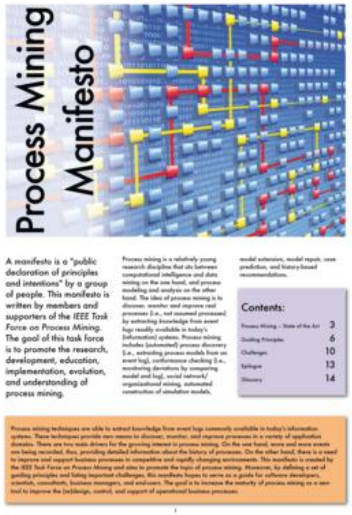
(WOZ objections Dutch municipality, 745 objections, 9583 event, f= 0.988)



Process Mining

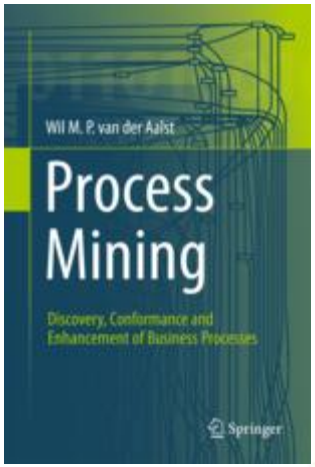
Advances in Process Mining

- **Many process discovery and conformance checking algorithms and tools are available (cf. the various ProM packages).**
- **Also commercial software based on these ideas:**
  Disco (Fluxicon), Reflect (Futura/Percentive), BPMOne (Pallas Athena/Perceptive), ARIS Process Performance Manager (Software AG), Futura Reflect (Futura Technology), Interstage Automated Process Discovery (Fujitsu), QPR ProcessAnalyzer/Analysis (QPR Software), flow (fourspark), Discovery Analyst (StereoLOGIC), etc.
- **We applied process mining in over 100 organizations.**

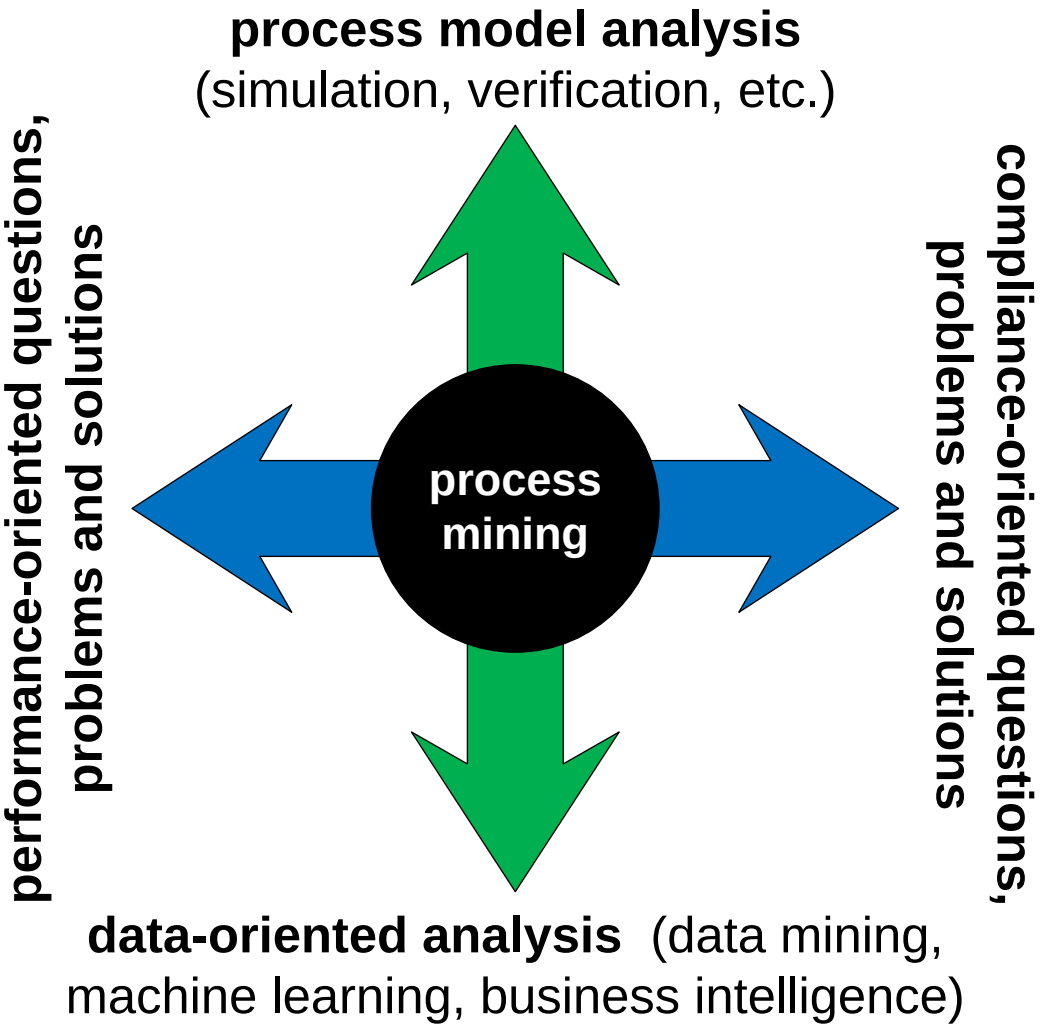


More than 75 people involving more than 50 organizations created the Process Mining Manifesto in the context of the IEEE Task Force on Process Mining.

Available in 13 languages



Missing link



Process Discovery (small selection)

automata-based learning

distributed genetic mining

language-based regions

heuristic mining

genetic mining

state-based regions

stochastic task graphs

LTL mining

neural networks

fuzzy mining

hidden Markov models

mining block structures

α algorithm

conformal process graph

multi-phase mining

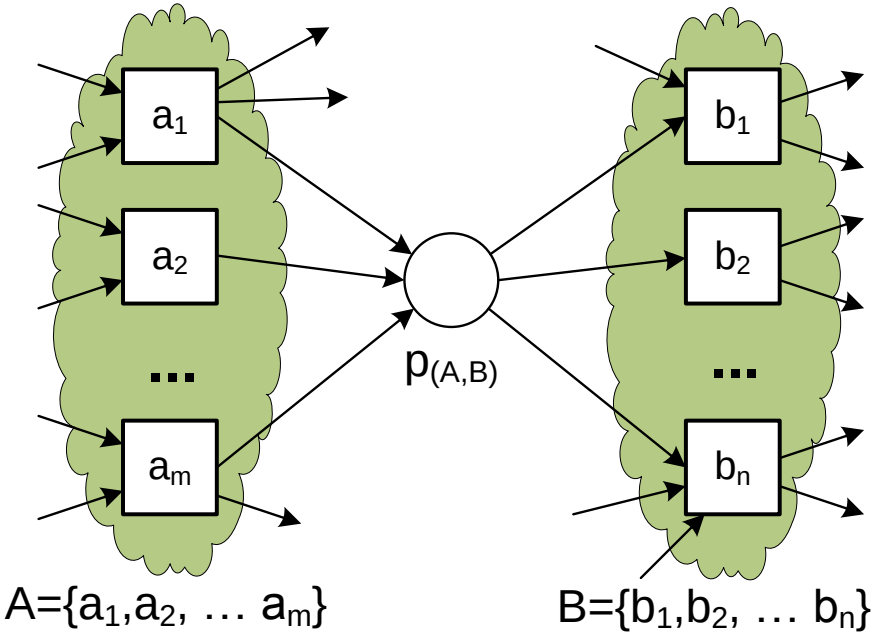
partial-order based mining

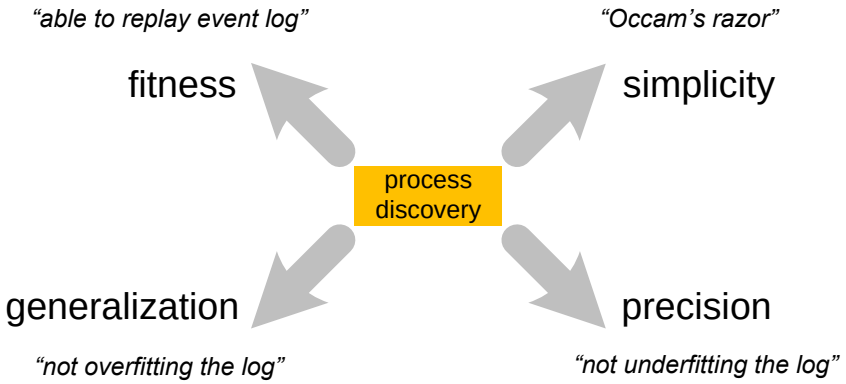
α# algorithm

ILP mining

α++ algorithm

Petri net view: Just discover the places …



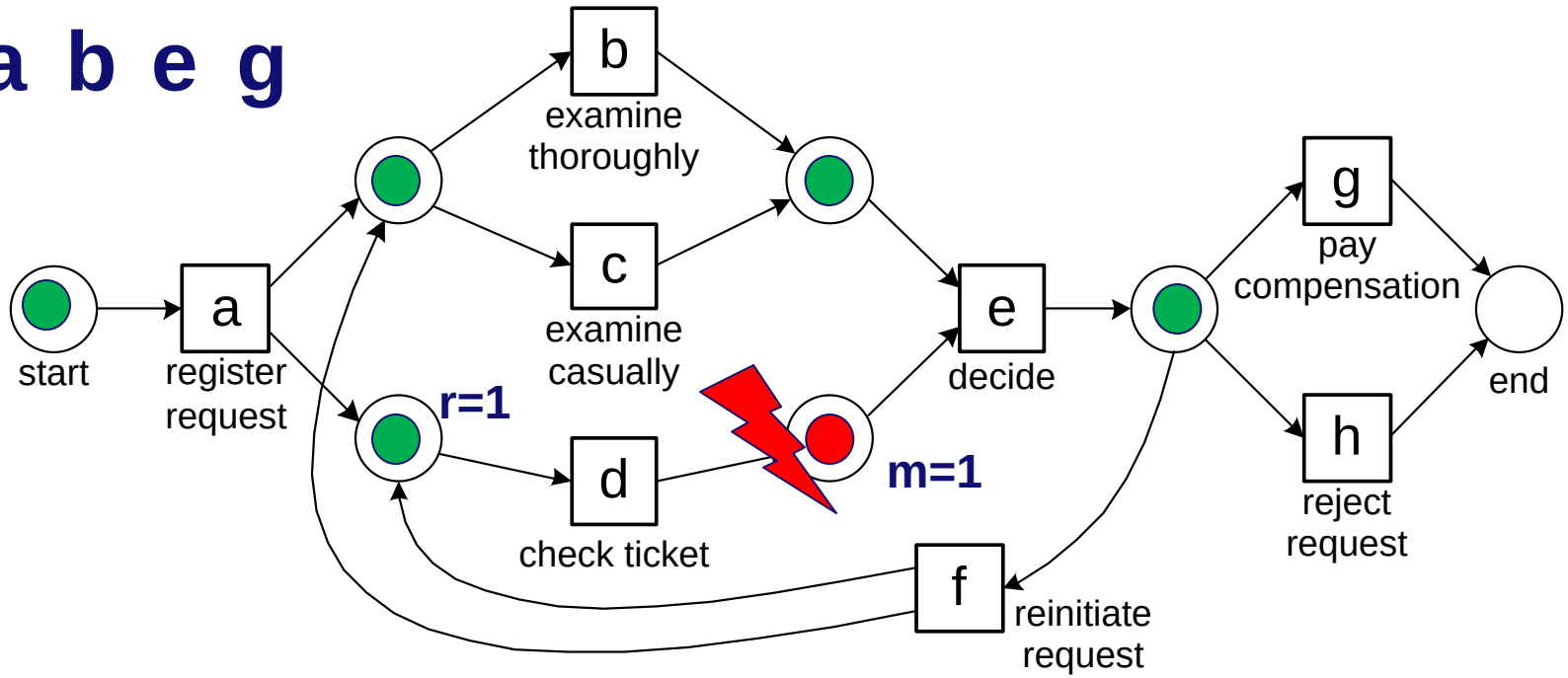


Adding a place limits behavior:

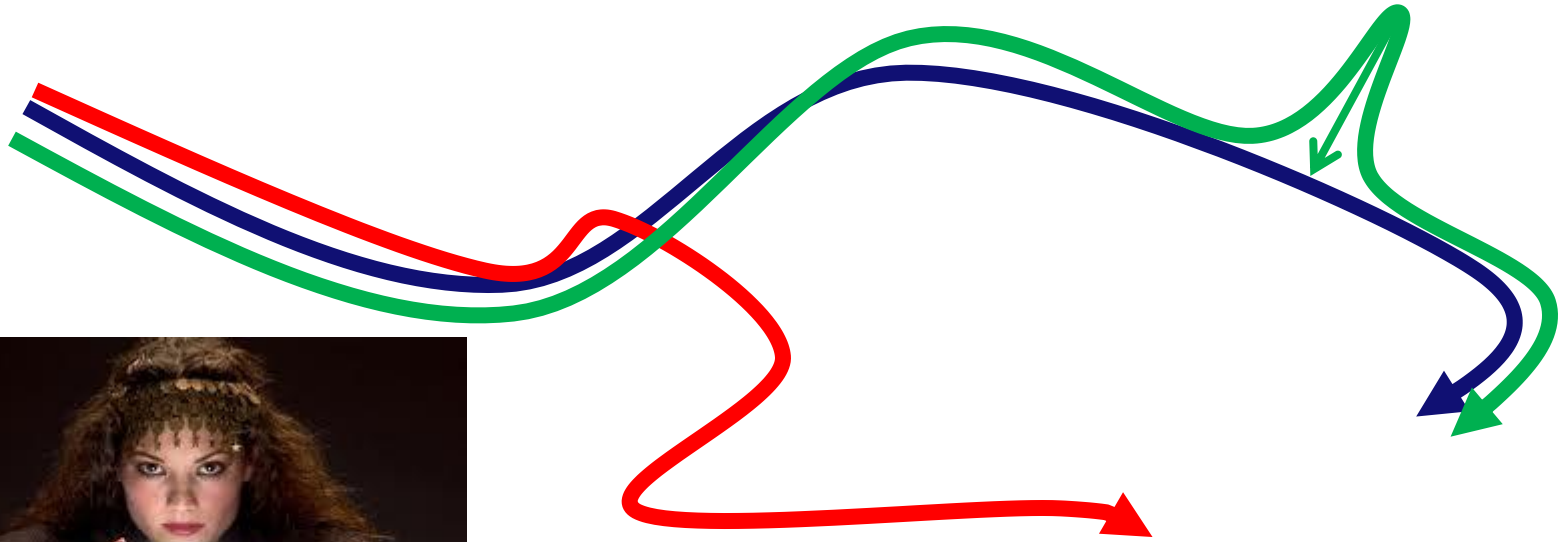
- **overfitting ≈ adding too many places**
- **underfitting ≈ adding too few places**

Replaying trace “abeg”

a b e g



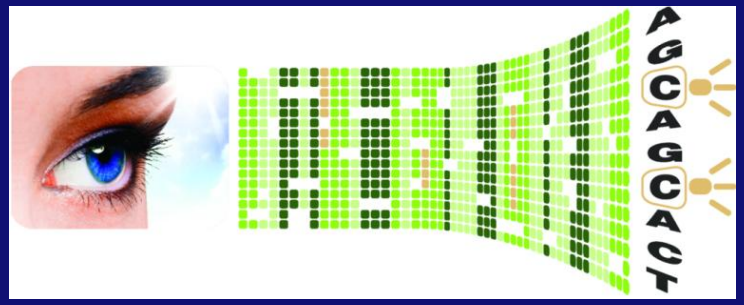
$$fitness(\sigma, N) = \frac{1}{2}\left(1 - \frac{1}{6}\right) + \frac{1}{2}\left(1 - \frac{1}{6}\right) = 0.83333$$



Alignments are essential!

- **conformance checking to diagnose deviations**
- **squeezing reality into the model to do model-based analysis**
- **not just control-flow (also data, resources, time, …) !**
- **also at run-time (predictions and recommendations) !**

Example Philips Healthcare
(work of JC Bose)



http://www.win.tue.nl/~jcbose/

Data Collection Process

RADAR

Data warehouse

~2500 CV Systems

What is Being Logged?

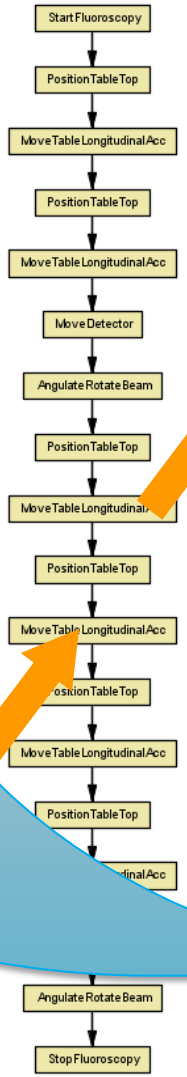
- Errors
- Warnings
- Information
- User commands
- Run information

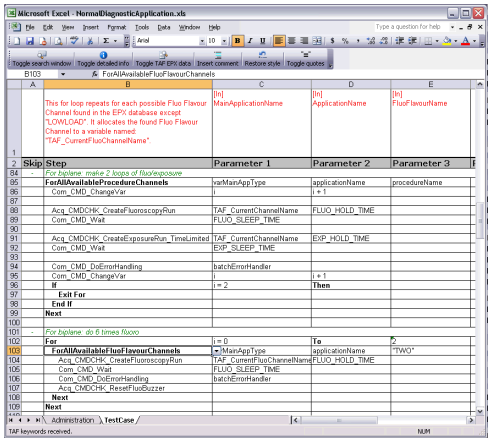
Challenges

- **Event Granularity**
  - Fine grained and too detailed
  - Often differs from the granularity at which an end-user/analyst is interested in
  - Lack of any standard/guideline on logging specifications
- **Case Heterogeneity**
  - Heterogeneous mix of usage scenarios
  - Wide range of functionality and flexibility in usage
- **Voluminous Data**
  - Hundreds of events can be triggered within a short span of time (even within a second)
  - Challenges the availability and scalability of analysis techniques
- **Unreliable Timestamps**
  - Ordering of events is not always reliable
  - An outcome of the current logging mechanism
    - Local clocks and local buffers
  - Clocks are not synchronized
- **Scoping**
  - Single log file captures *everything* related to a particular system on a single day
  - Appropriate scoping is essential based on the perspective of analysis

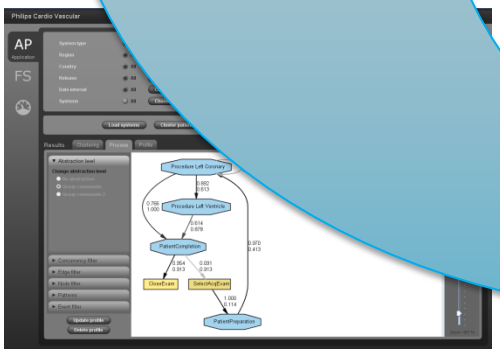
Test Automation Framework:
Testing CV systems under realistic circumstances

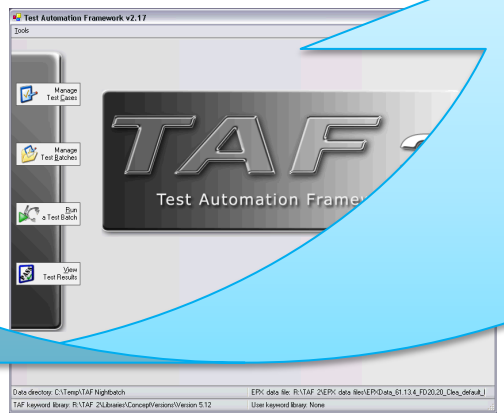
Hospital



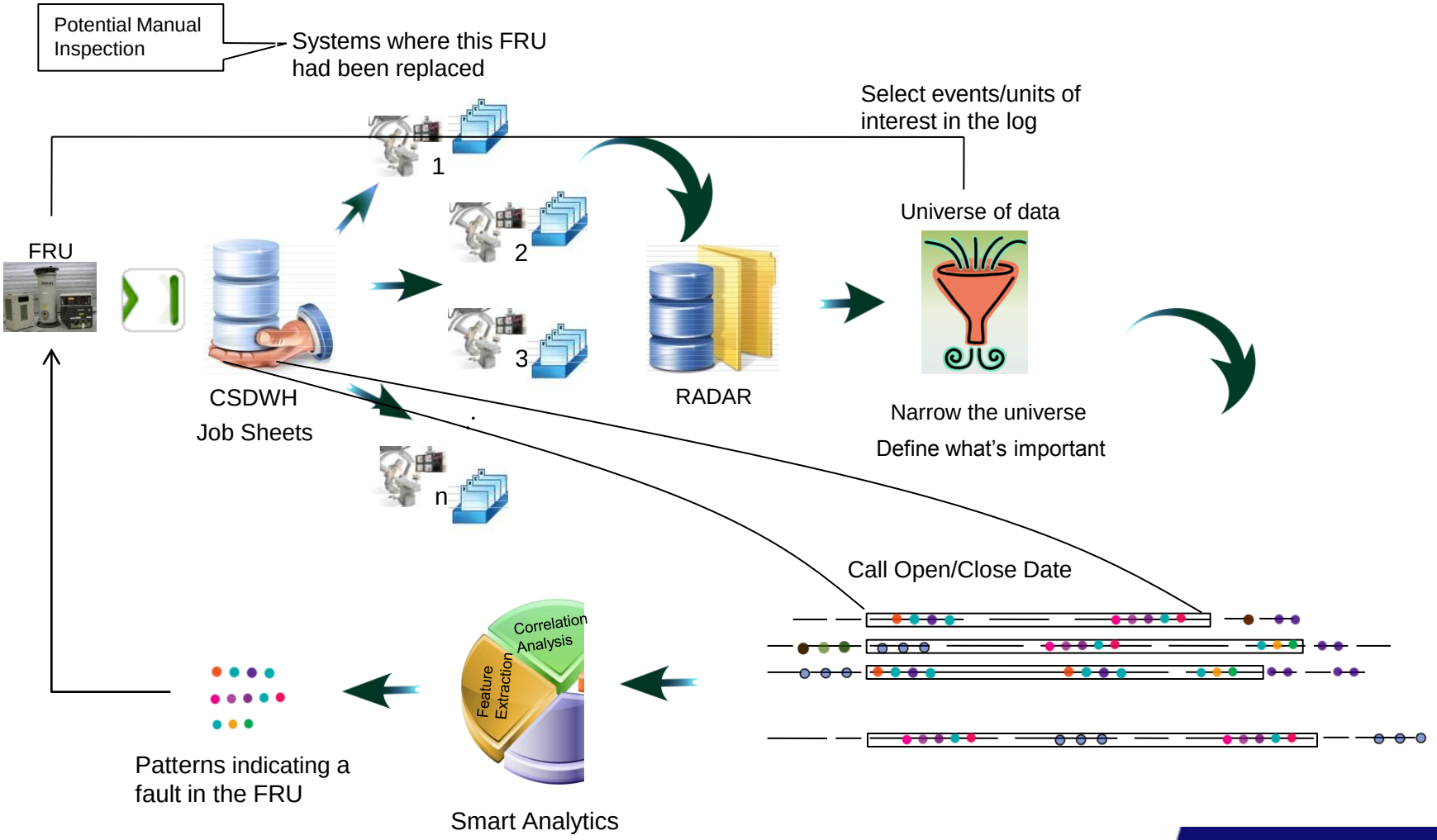


Philips Healthcare

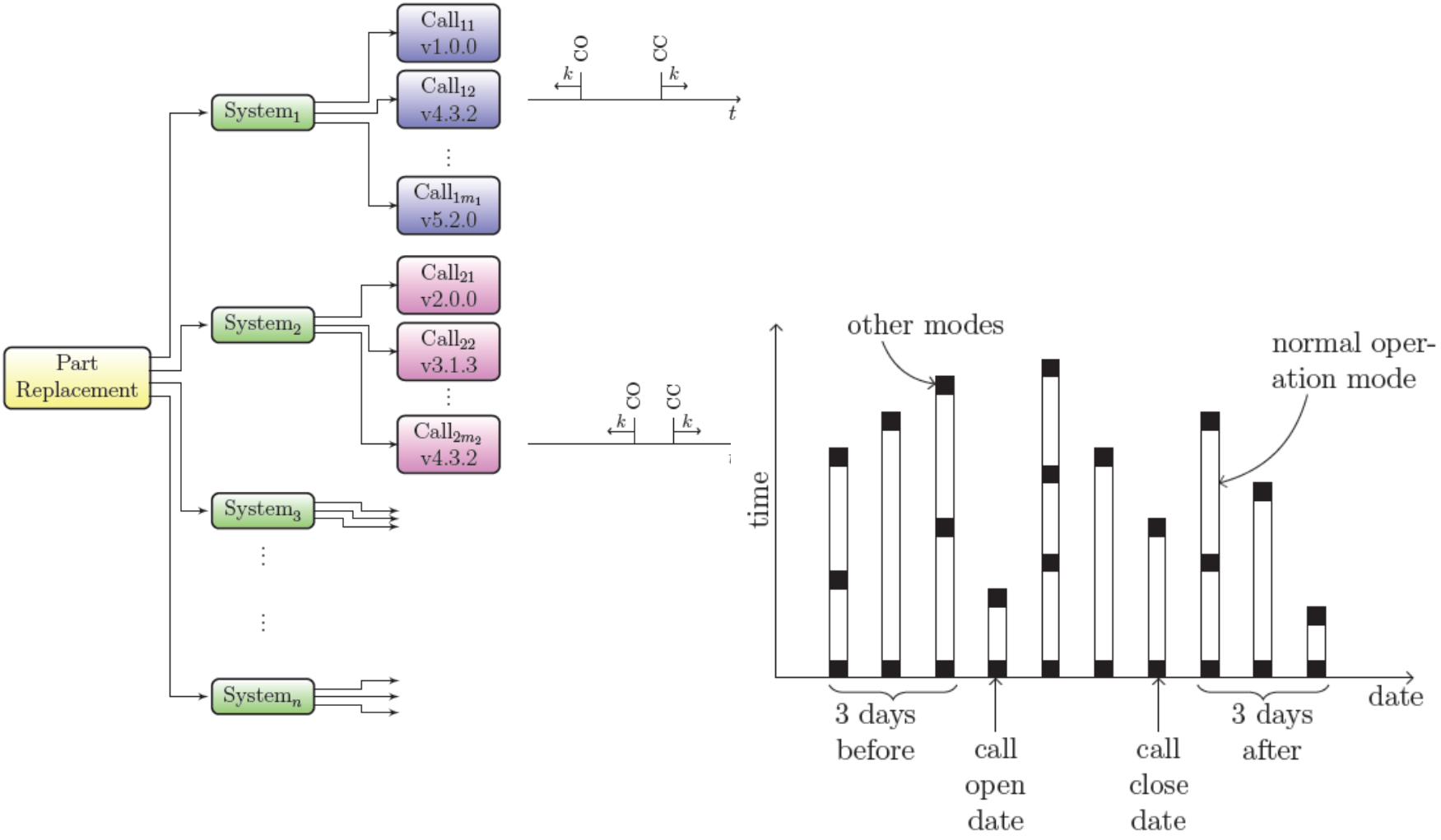




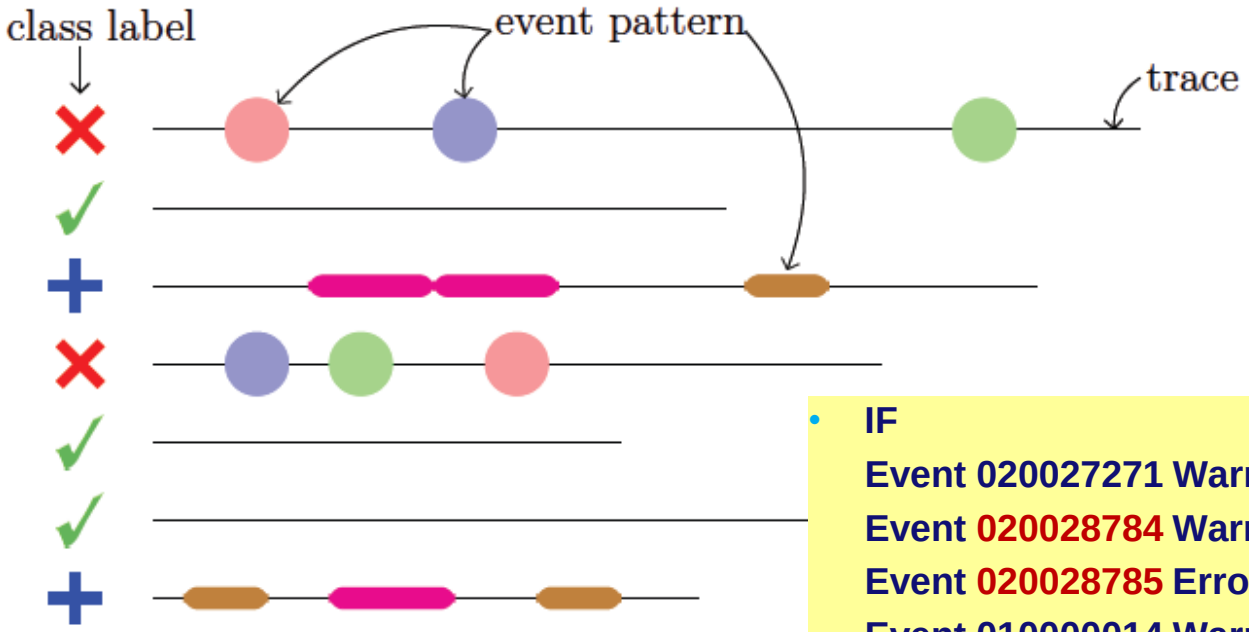
Signature Discovery



Building a classification problem

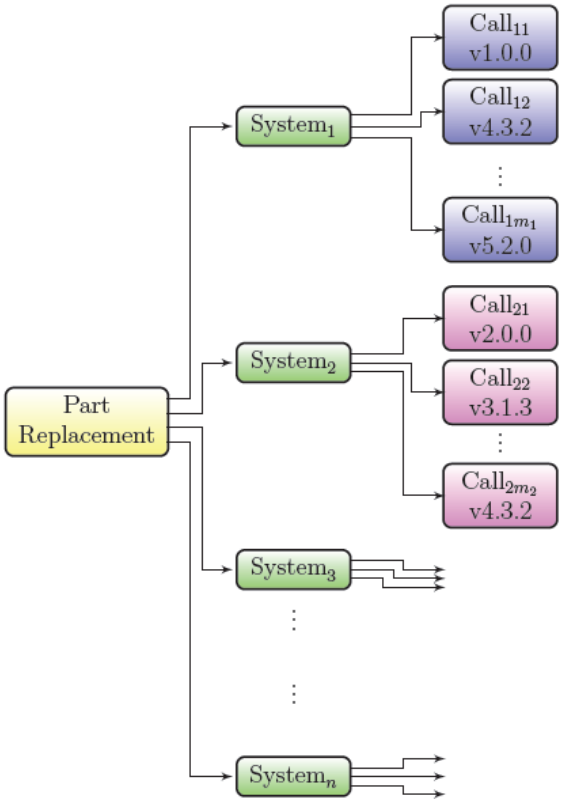


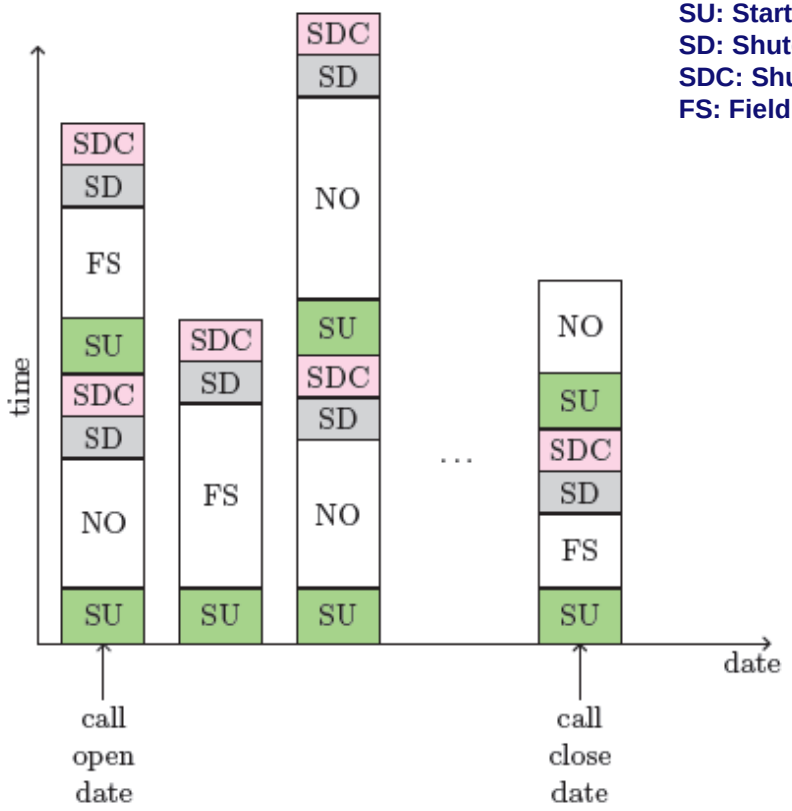
Signature Discovery



- **IF**
  Event 020027271 Warning from X-Ray Control AND
  Event 020028784 Warning from X-Ray Control AND
  Event 020028785 Error from X-Ray Control AND
  Event 010000014 Warning from X-Ray Generator AND
  Event 510032788 Error from Acquisition
- **OR**
  Event 020027271 Warning from X-Ray Control AND
  Event 020028786 Warning from X-Ray Control AND
  Event 020028787 Error from X-Ray Control AND
  Event 010000014 Warning from X-Ray Generator AND
  Event 510032788 Error from Acquisition

Field Service Engineer Workflow Analysis

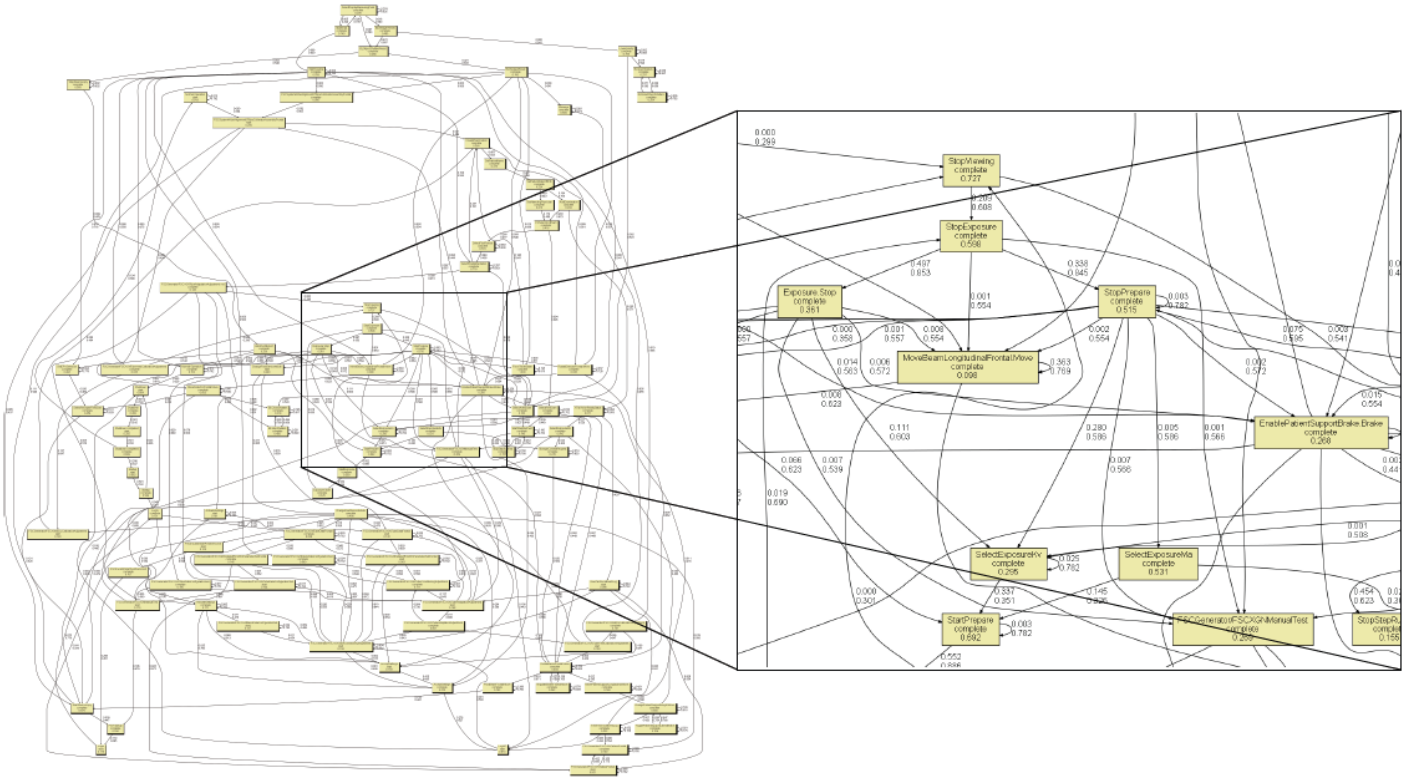




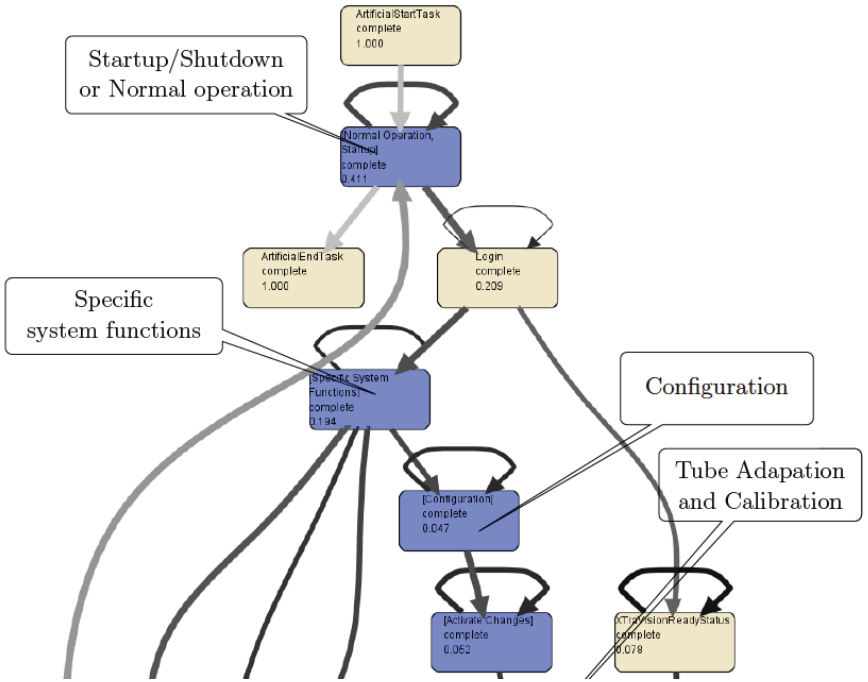
NO: Normal Operation
SU: Startup
SD: Shutdown
SDC: Shutdown Completed
FS: Field Service

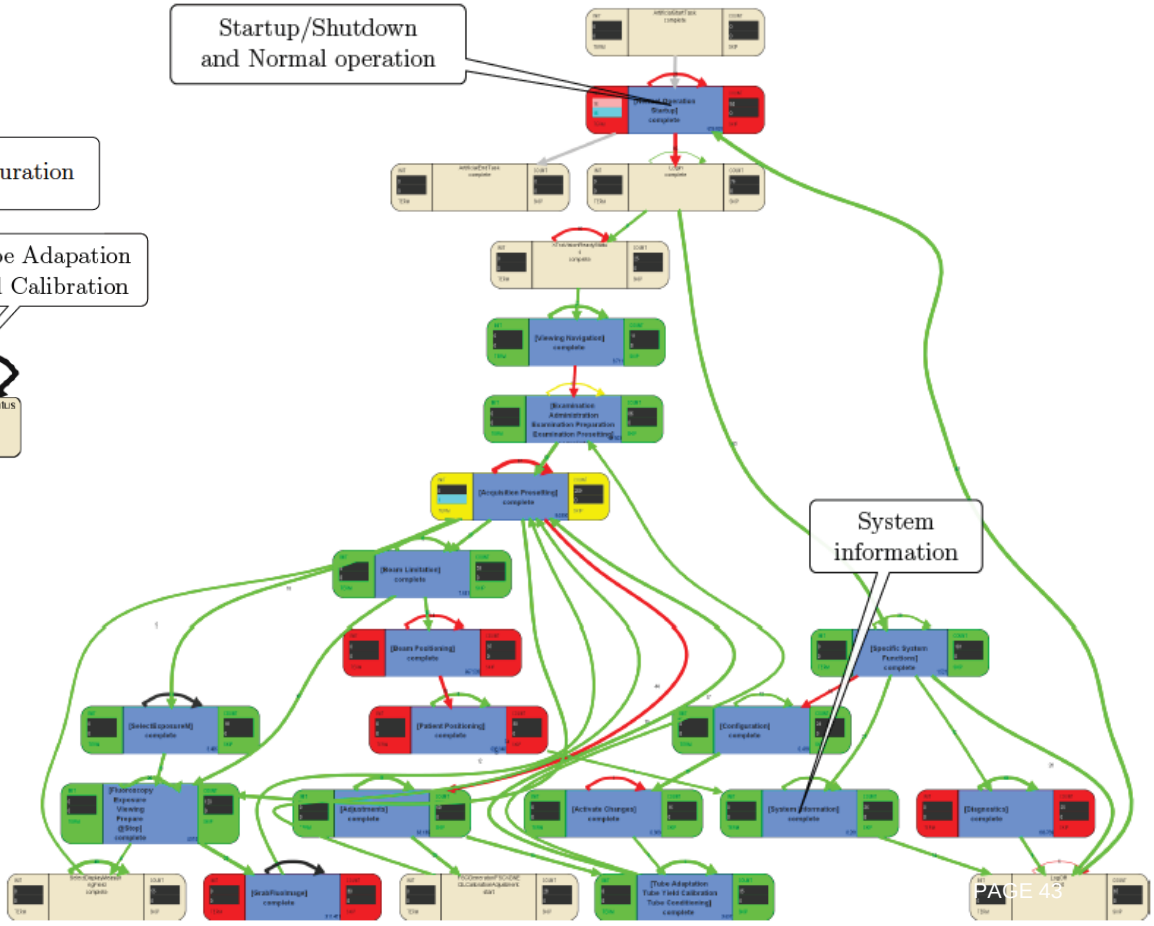
Conformance Checking and Discovering the Actual Workflow

- **Many deviations between instructions and real work of engineers.**
- **Event data are not at the right abstraction level.**



Using abstractions and time information

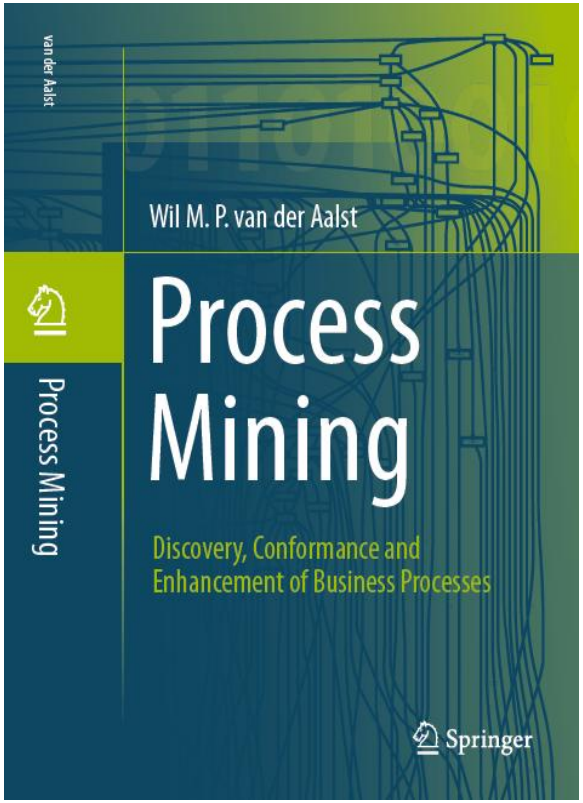




Conclusion

Conclusion

- Process Mining (PM): connecting event data and process models
- PM shows how systems are really being used.
- PM can be used for conformance checking and discovery.
- There are mature PM tools (e.g., ProM, Disco, Reflect).
- If there is enough high-quality data, PM can also be used for prediction and recommendation.
- Interested: Contact us!



www.processmining.org