

Process Mining

Olifantenpaden in Assurance

prof. Wil van der Aalst
Technische Universiteit Eindhoven

www.processmining.org





007001101010101010





What Happens in an Internet Minute?

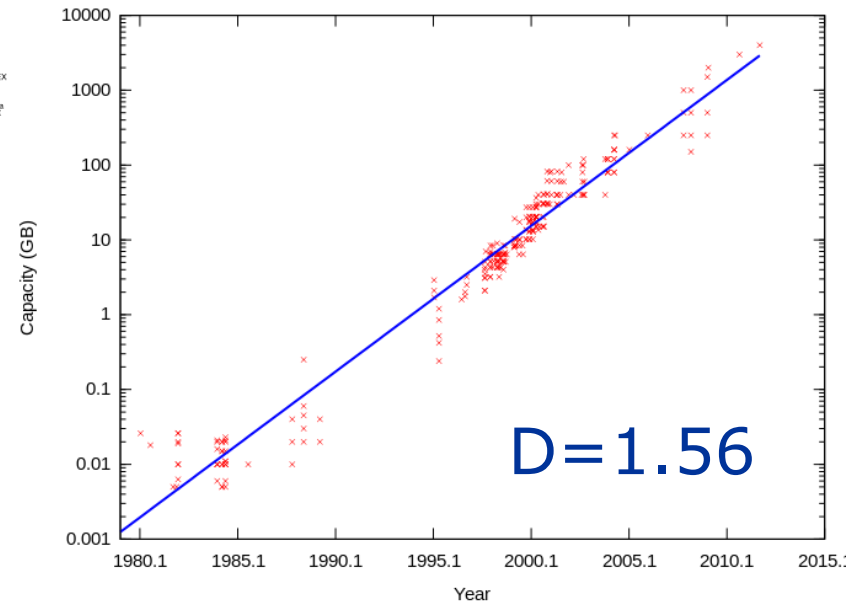
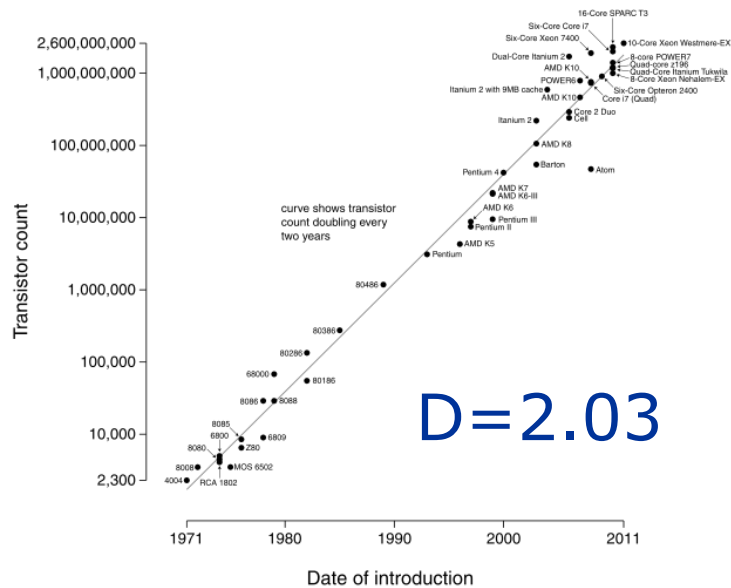


And Future Growth is Staggering



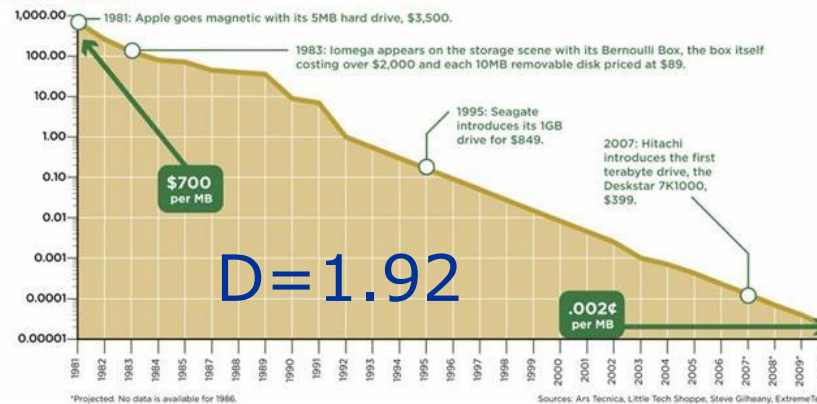
Moore's Law

Microprocessor Transistor Counts 1971-2011 & Moore's Law



STORAGE: FROM HIGHWAY ROBBERY TO RUNAWAY BARGAIN

\$ per megabyte



A simple calculation



- Starting point 2010:
 - Harddisk 1 Terabyte = 10^{12} bytes
 - Digital Universe 1.2 Zettabyte = 1.2×10^{21} bytes
(estimate in IDC's annual report, "The Digital Universe Decade – Are You Ready?" May 2010)
- Disk needs to grow $2^{30.16} = 1.2 \times 10^9 = 1.2 \times 10^{21} / 10^{12}$ times its current size.
- Assuming $D=1.56$ this takes $30.16 \times 1.56 = 47.05$ years.
- Hence, in 2060 your laptop can contain all of today's digital universe (internet, computer files, transaction logs, movies, photos, music, books, databases, etc.)!

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

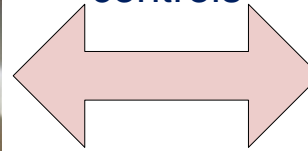


<http://dilbert.com/strips/comic/2012-07-29/>

Process Mining



supports/
controls



software
system

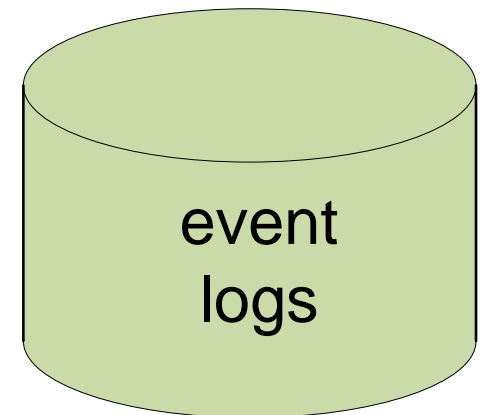
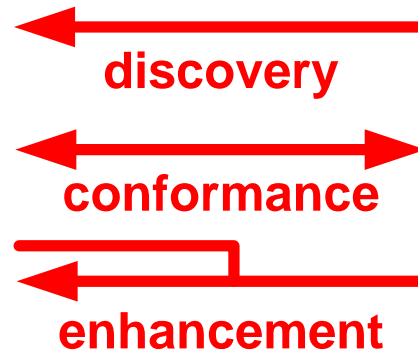
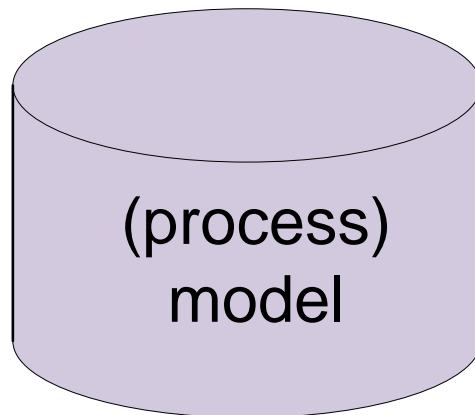
models
analyzes



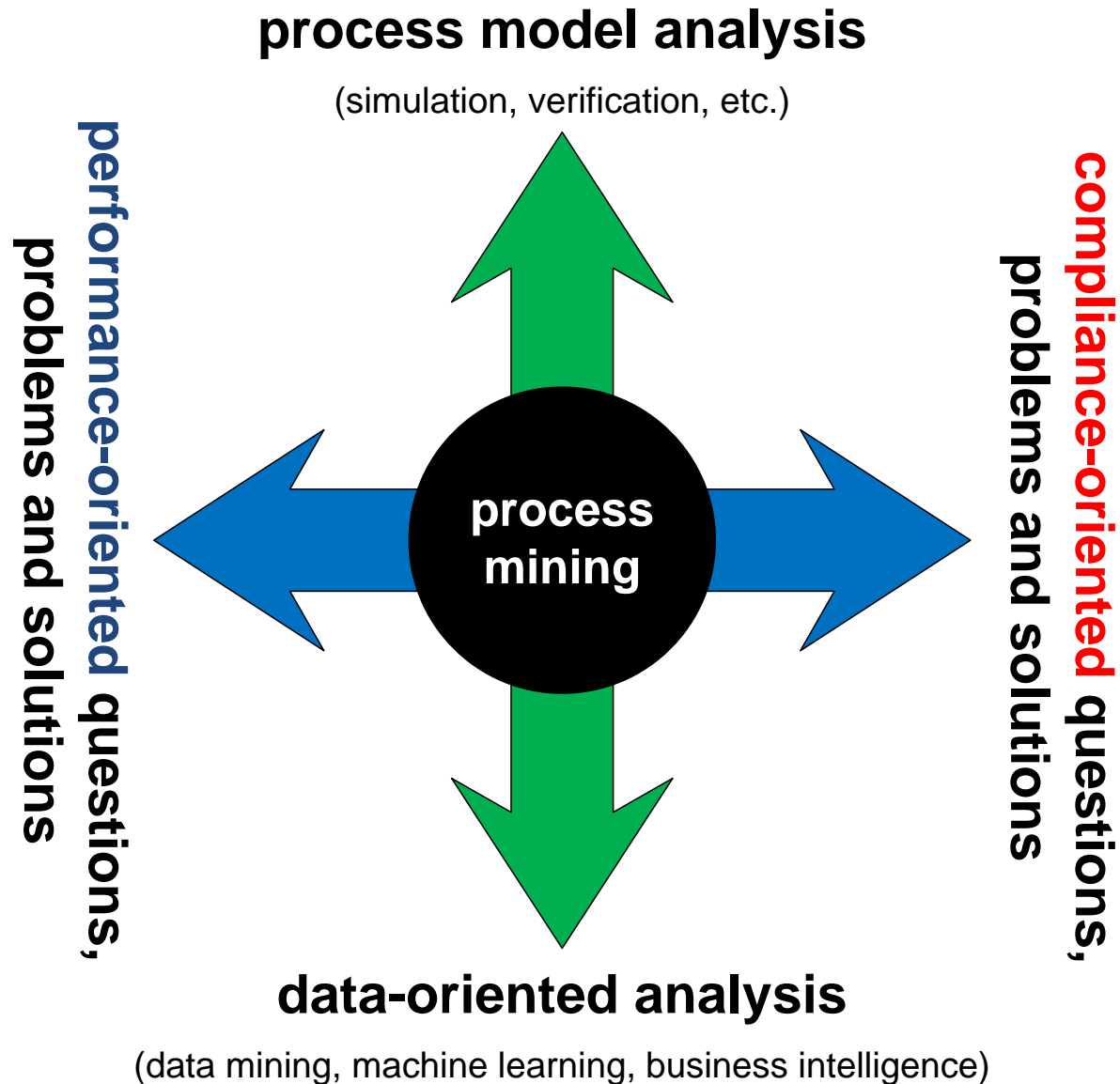
specifies
configures
implements
analyzes



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




Positioning Process Mining



**Front-
office**

**Back-
office**



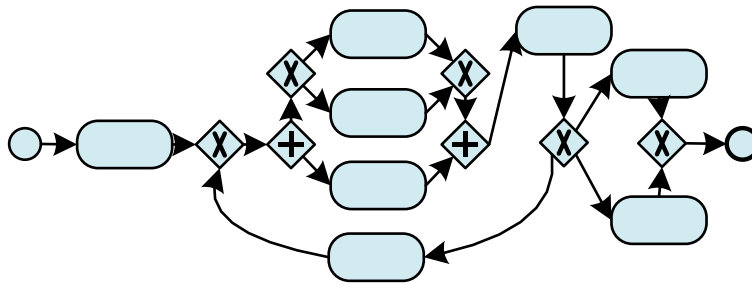


Assurance: Part of corporate governance in which a management provides **accurate and current information** to the stakeholders about the **efficiency and effectiveness of its policies and operations** and its compliance with **statutory obligations** (businessdictionary.com).

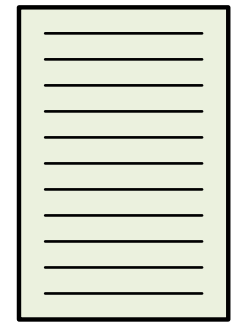
Let's play!



Play-Out

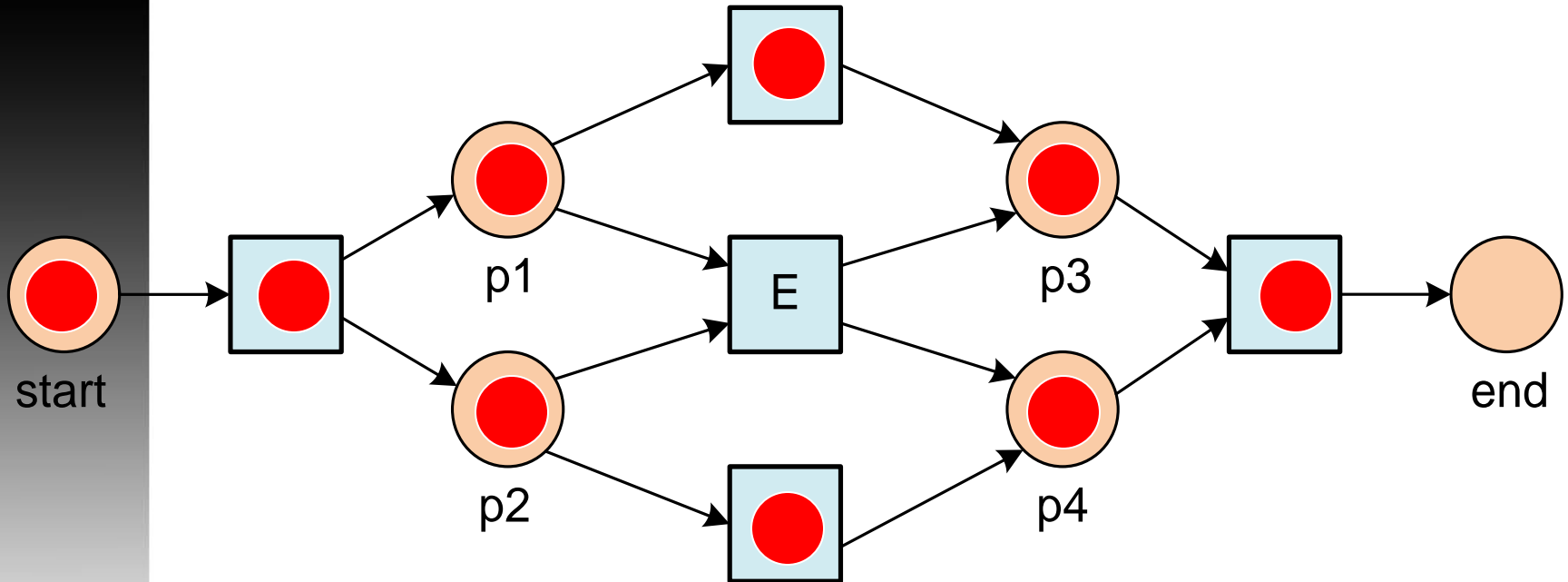


process model



event log

Play-Out (Classical use of models)



A B C D A E D A E D

A B C D

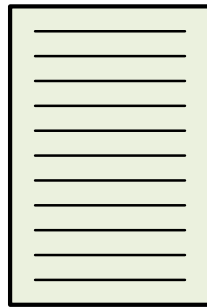
A C B D

A C B D

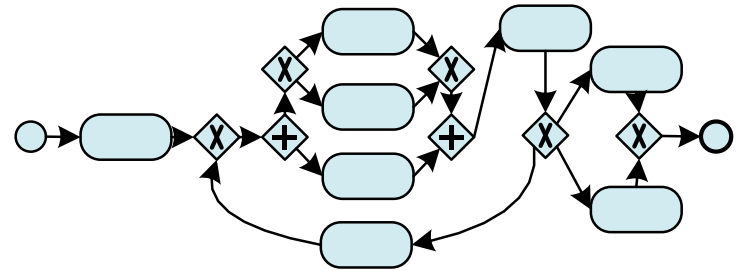
A E D

A C B D

Play-In



event log



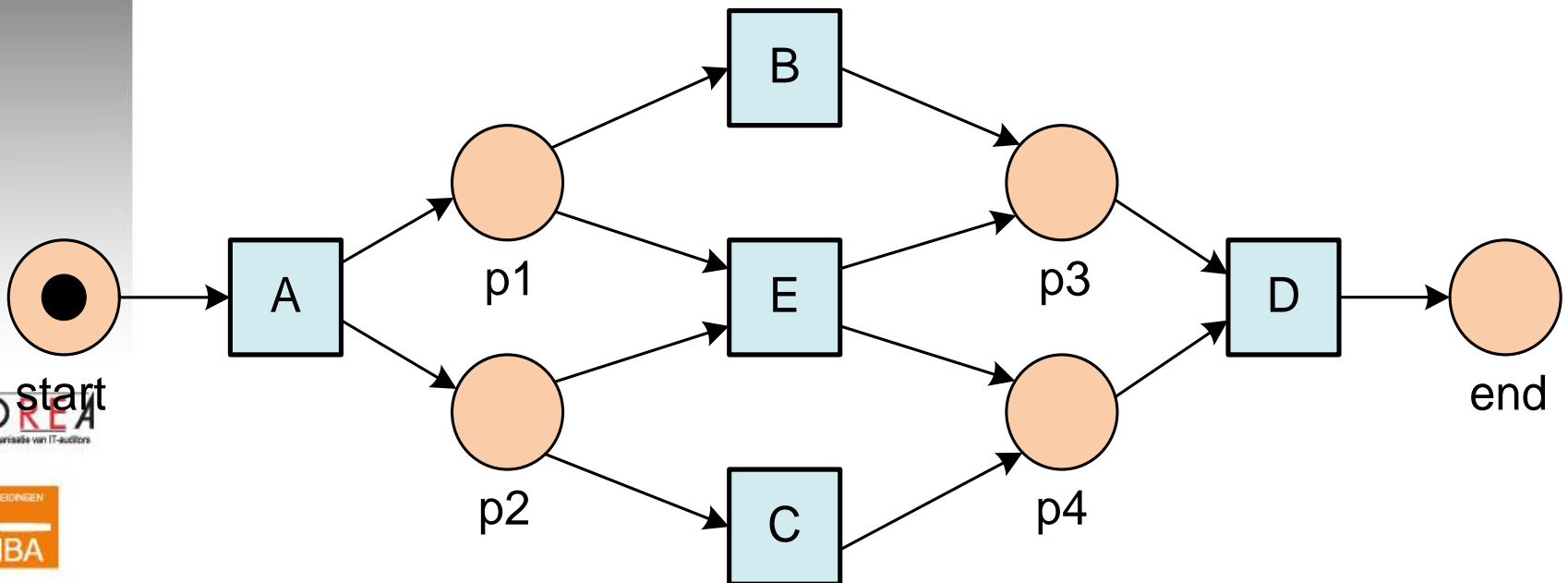
process model

Play-In

A B C D A E D A E D

A C B D A B C D

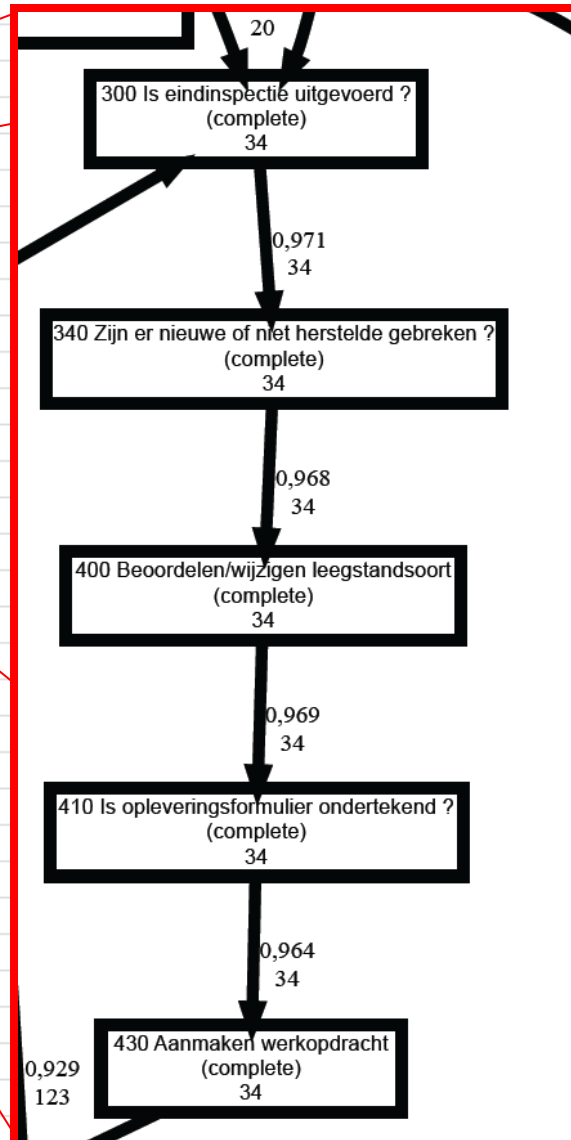
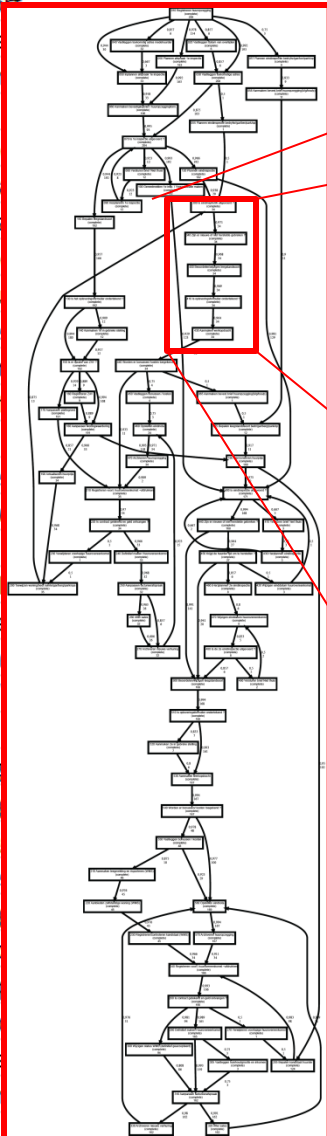
A C B D A E D A C B D



Example Process Discovery

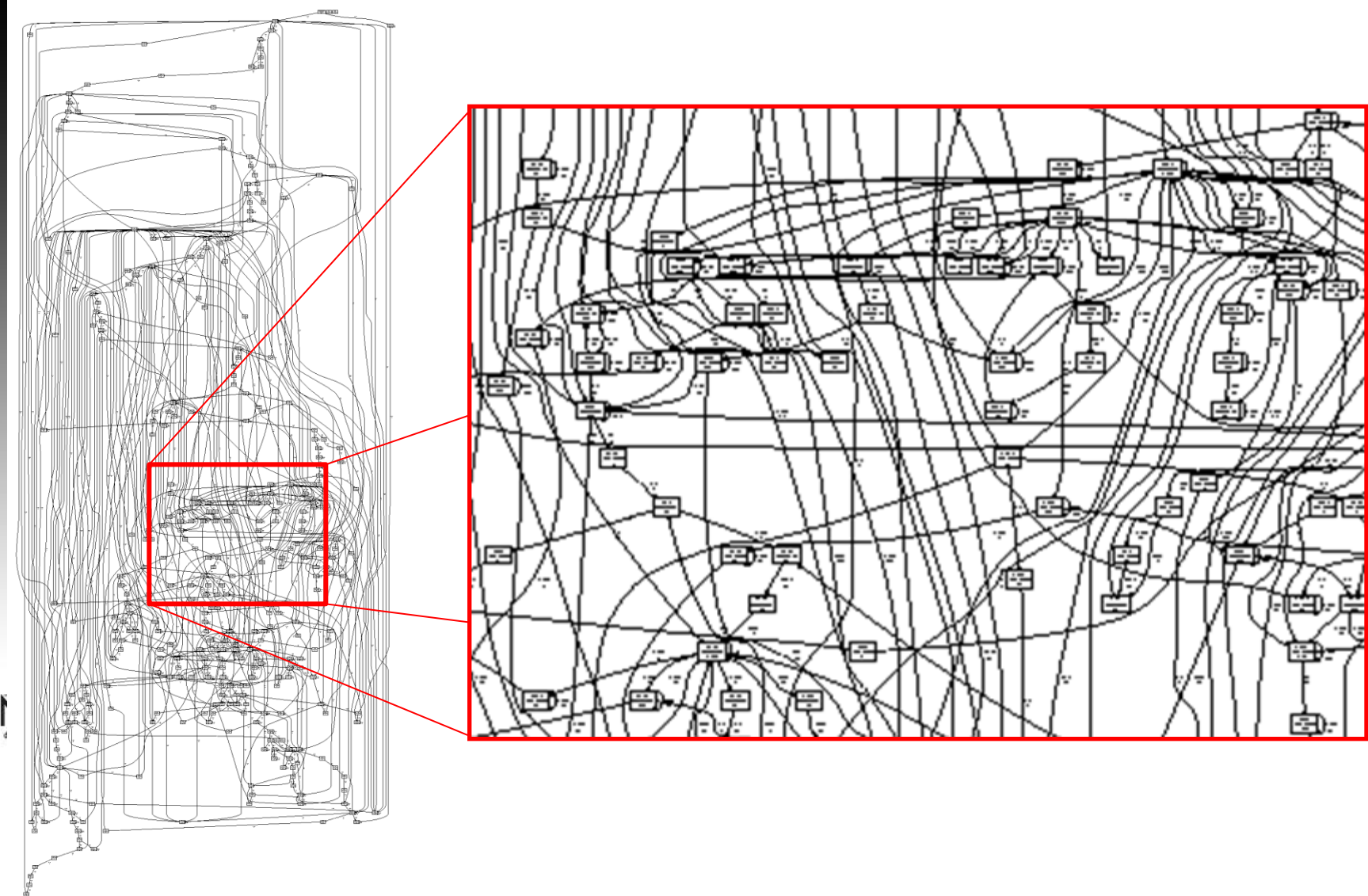
(Vestia, Dutch housing agency, 208 cases, 5987 events)

117315	110 Bepalen leegstandsoort	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. hu	10.09.2007 16:24:36
117315	220 Is contract getekend e	11.09.2007 14:56:18
117315	240 Definitief maken Huu	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	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 toekomst	09.05.2007 12:25:01
119763	050 Inplannen afspraak 1e	09.05.2007 11:59:52
119763	060 Aanmaken bevestiging	09.05.2007 12:31:57
119763	070 Is 1e inspectie uitgev	16.05.2007 13:04:26
119763	100 Gereedmelden 1e ins	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



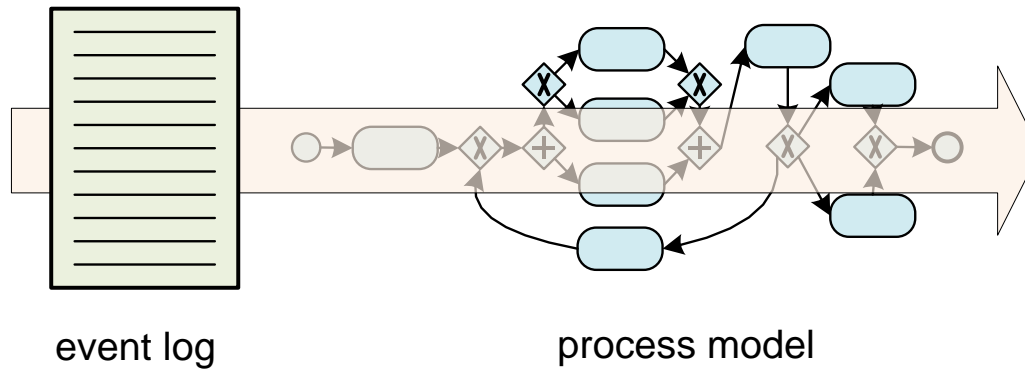
Example Process Discovery

(ASML, test process lithography systems, 154966 events)





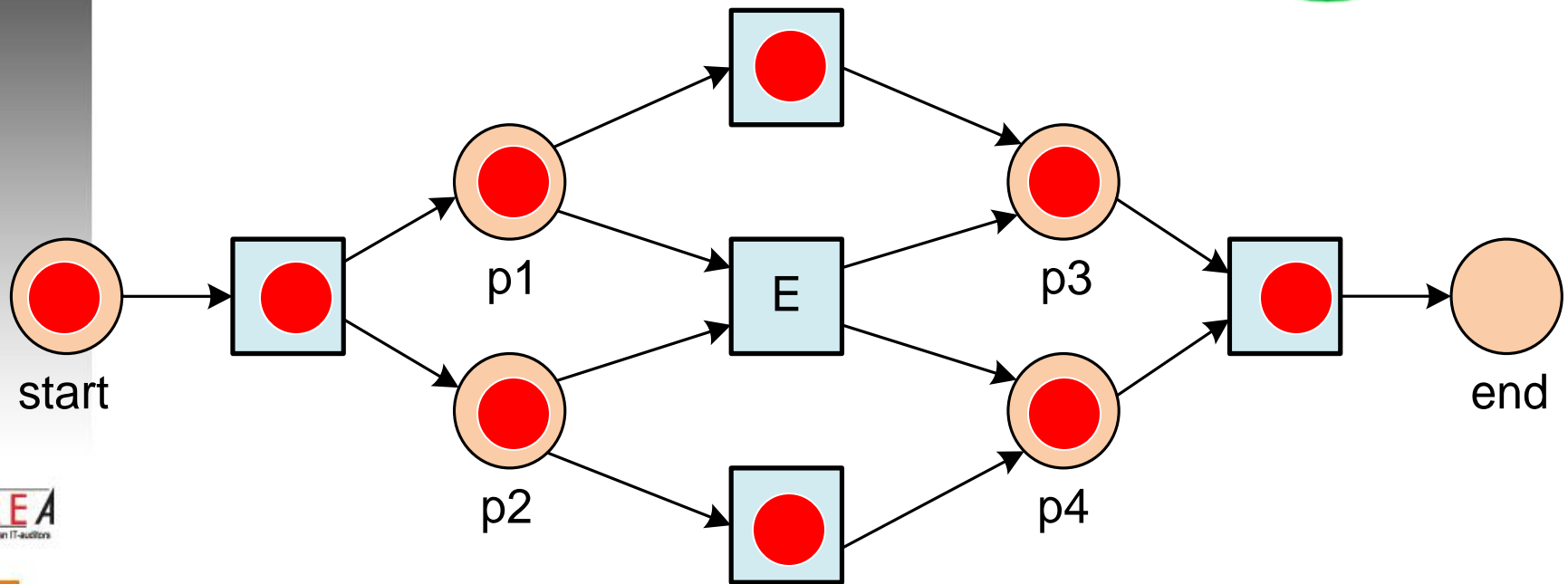
Replay



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

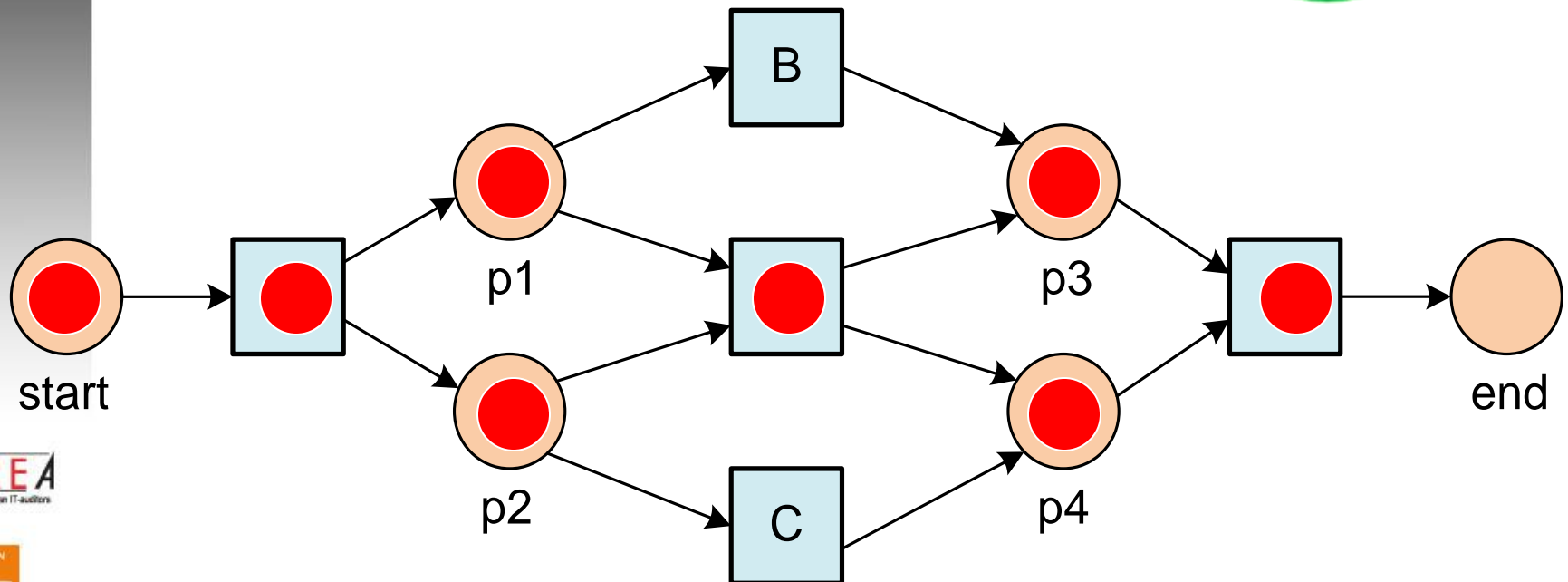
Replay

A B C D



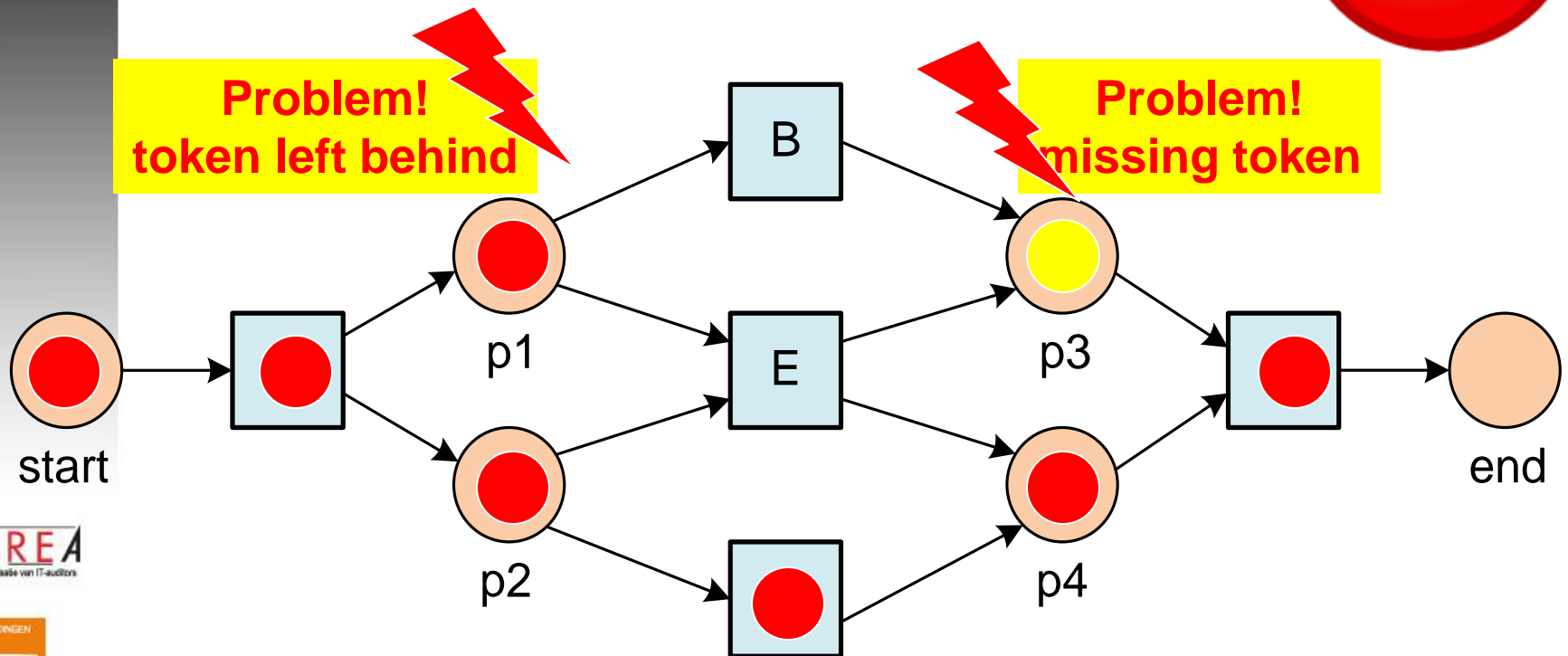
Replay

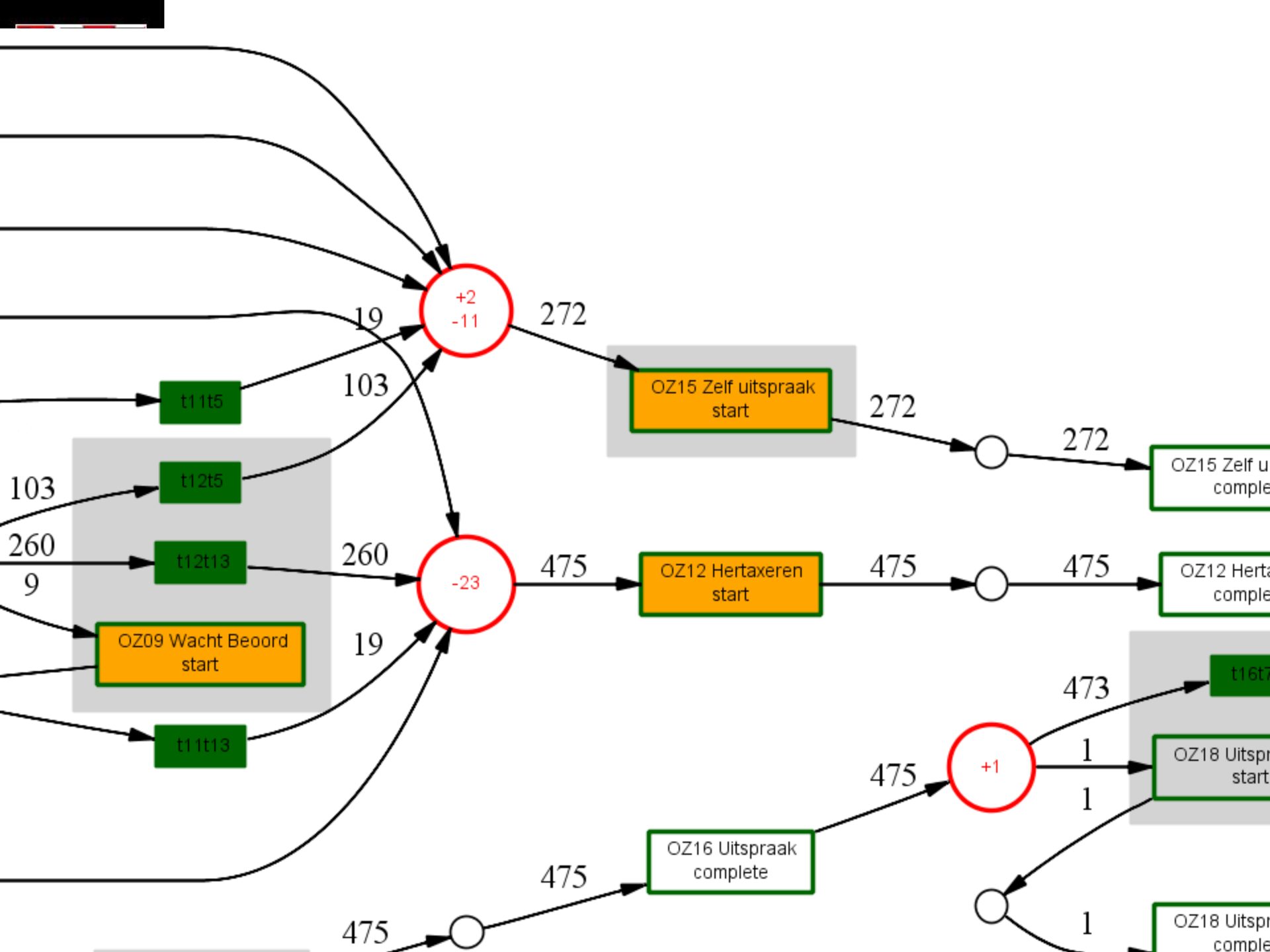
A E D



Replay can detect problems

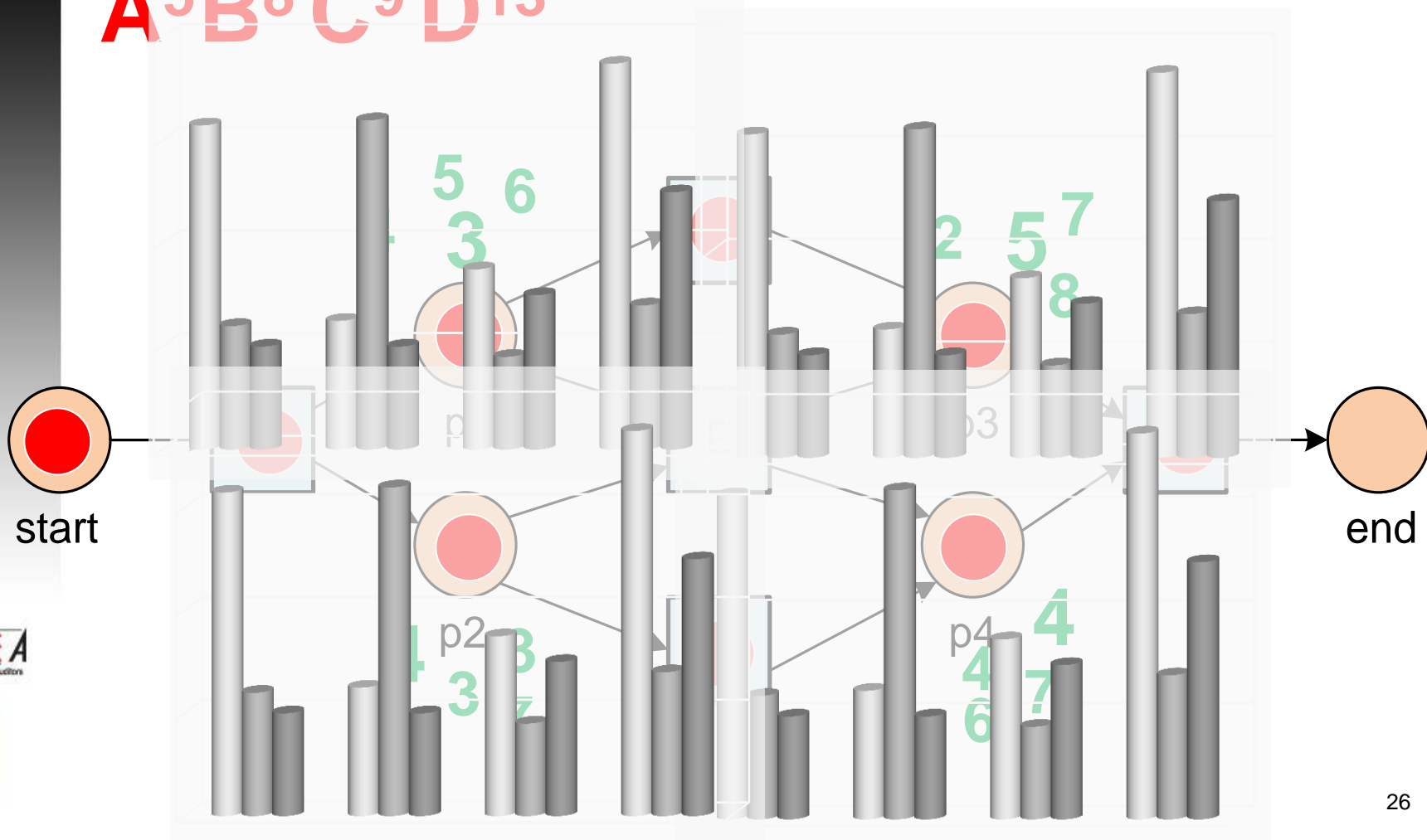
A C D





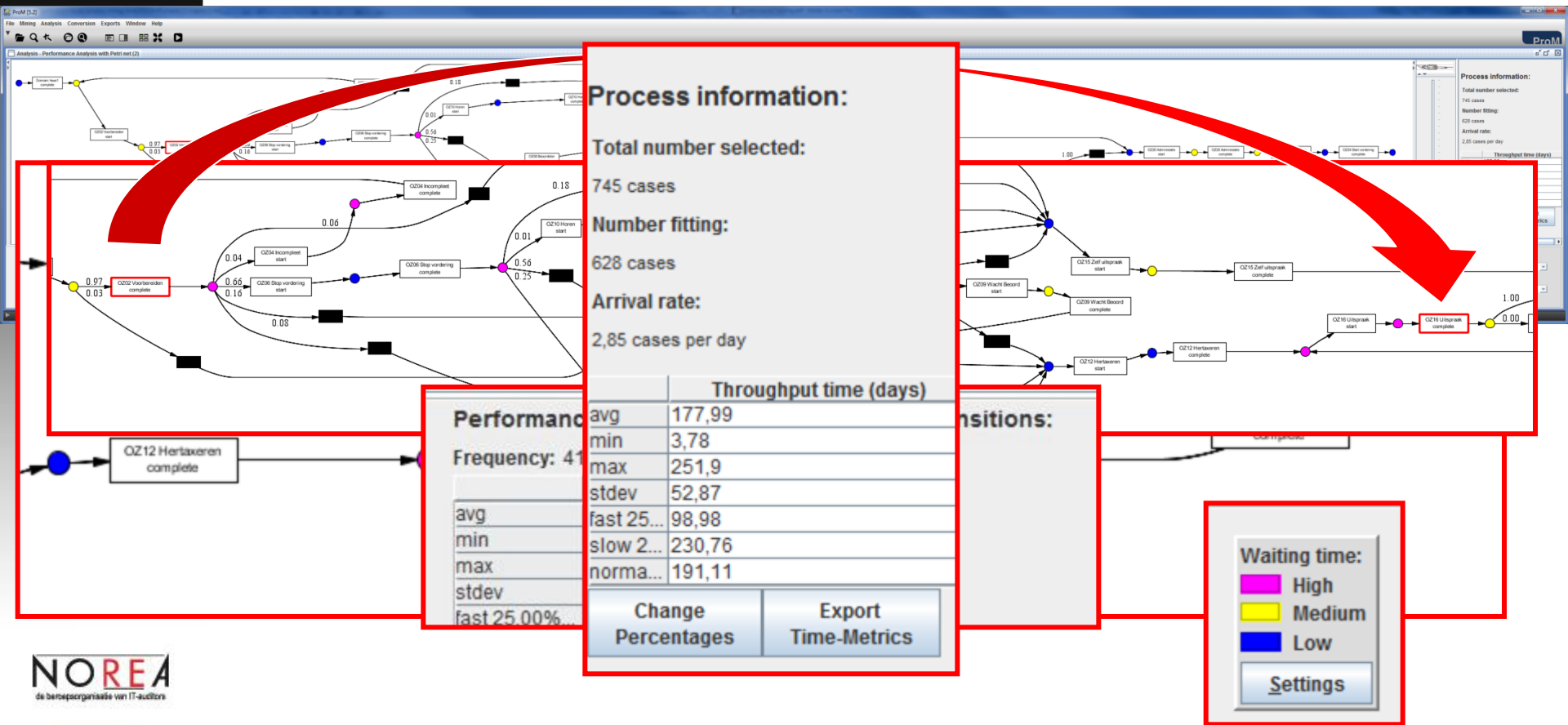
Replay can extract timing information

A⁵B⁸C⁹D¹³



Performance Analysis Using Replay

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





Process Discovery (small selection)

automata-based learning

heuristic mining

genetic mining

stochastic task graphs

fuzzy mining

mining block structures

α algorithm

$\alpha\#$ algorithm

$\alpha++$ algorithm

distributed genetic mining

language-based regions

state-based regions

LTL mining

neural networks

hidden Markov models



multi-phase mining

conformal process graph

partial-order based mining

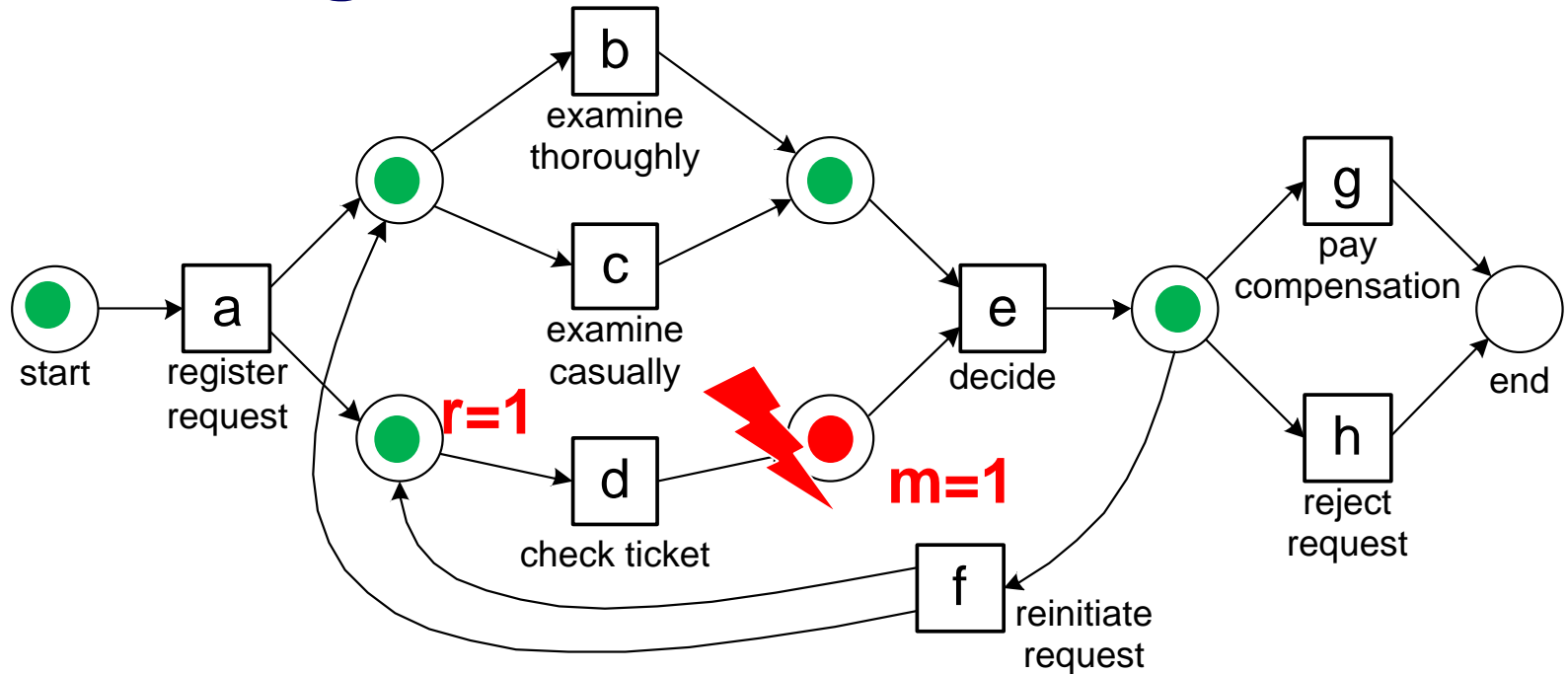
ILP mining

Conformance Checking



Replaying trace "abeg"

a b e g



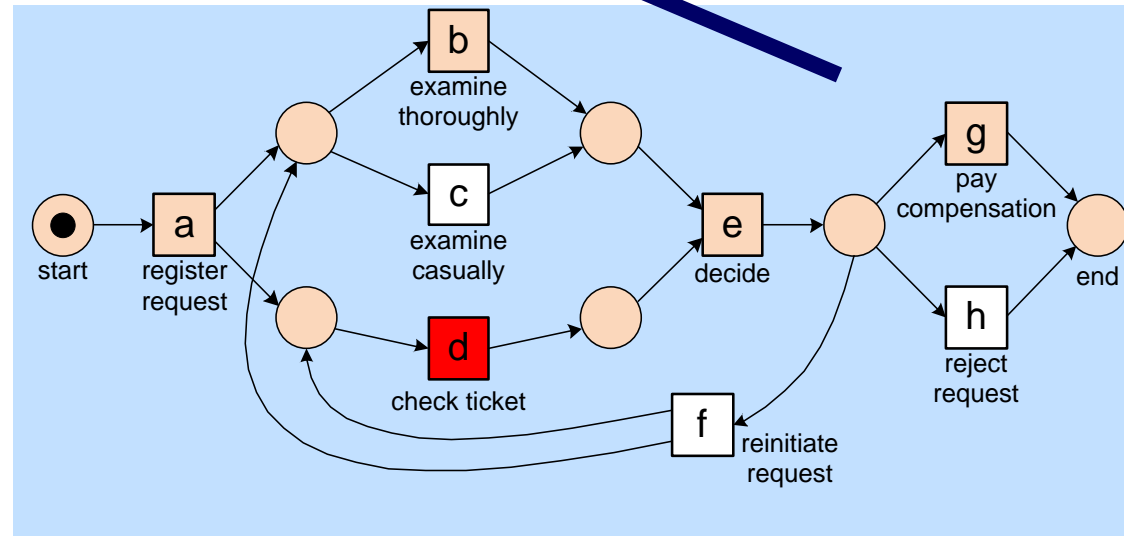
$$fitness(\sigma, N) = \frac{1}{2} \left(1 - \frac{1}{6} \right) + \frac{1}{2} \left(1 - \frac{1}{6} \right) = 0.83333$$

From “playing the token game” to optimal alignments ...

observed trace: “abeg”

a	b	»	e	g
a	b	d	e	g

move in
model only

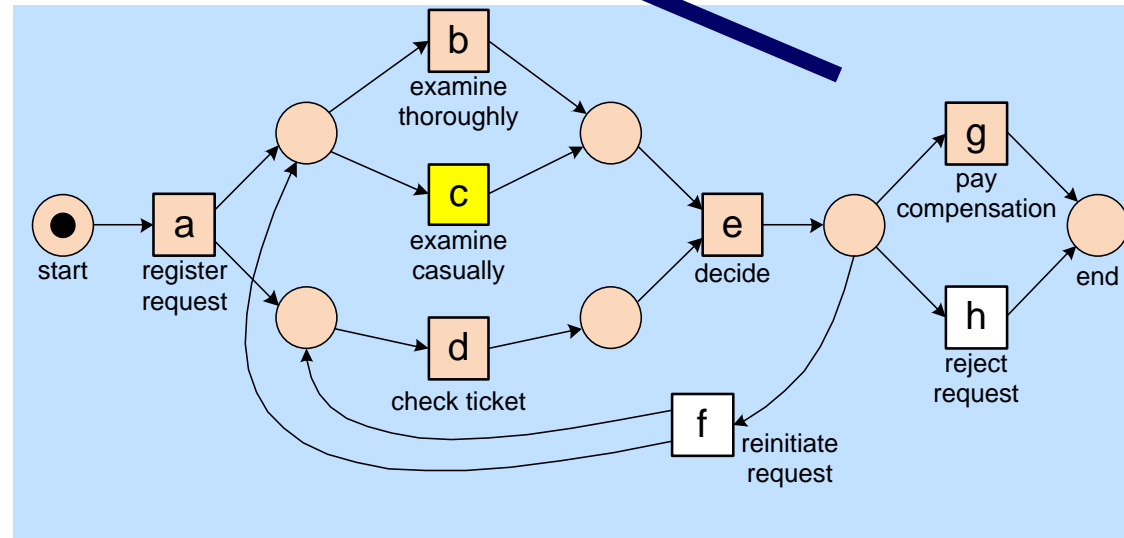


Another alignment

observed trace: "abcdeg"

a	b	c	d	e	g
a	b	»	d	e	g

move in
log only



Moves have costs

...	a	...
...	»	...

...	»	...
...	a	...

...	a	...
...	a	...

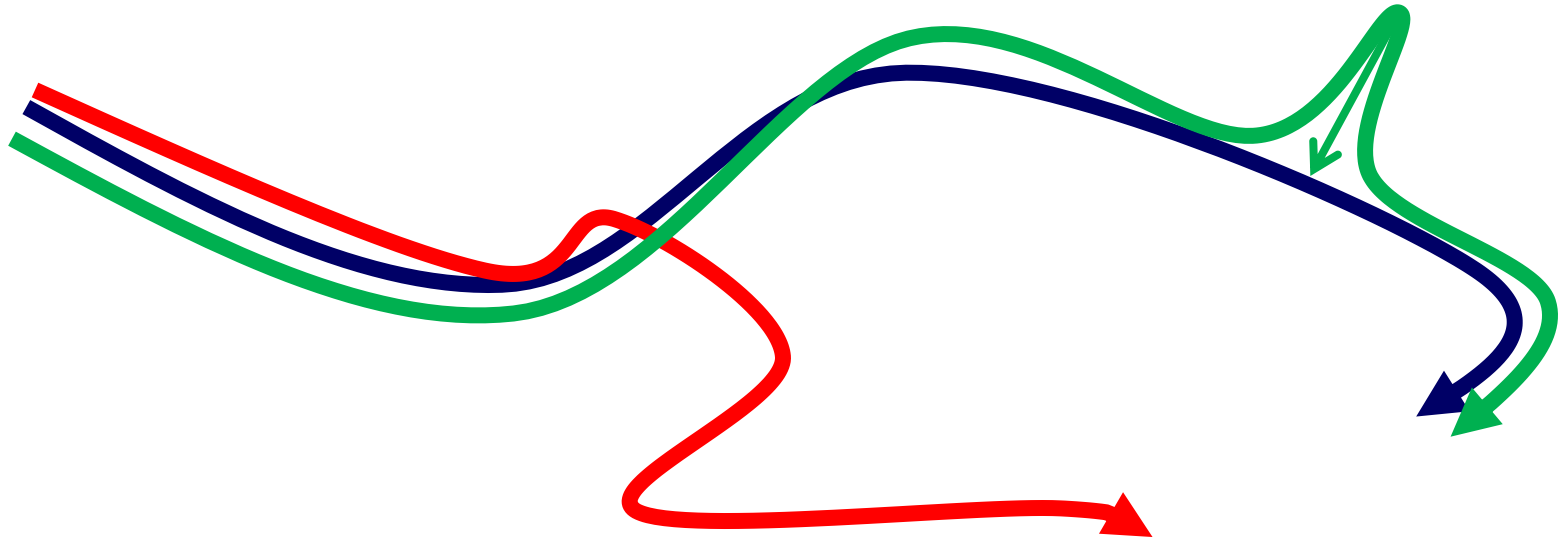
...	a	...
...	b	...

- Standard cost function:
 - $c(x, \gg) = 1$
 - $c(\gg, y) = 1$
 - $c(x, y) = 0$, if $x = y$
 - $c(x, y) = \infty$, if $x \neq y$

Any cost structure is possible

...	send-letter(John,2 weeks, \$400)	...
...	send-email(Sue,3 weeks,\$500)	...

- **Similar activities** (more similarity implies lower costs).
- **Resource conformance** (done by someone that does not have the specified role).
- **Data conformance** (path is not possible for this customer).
- **Time conformance** (missed the legal deadline)



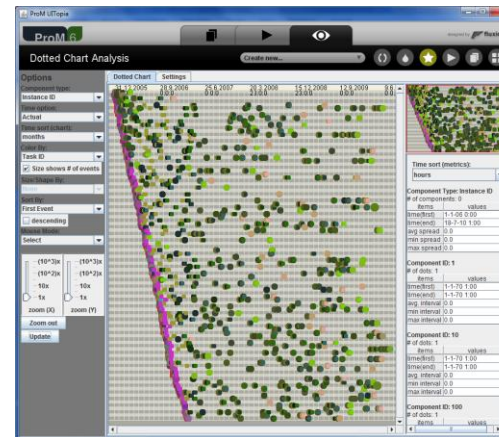
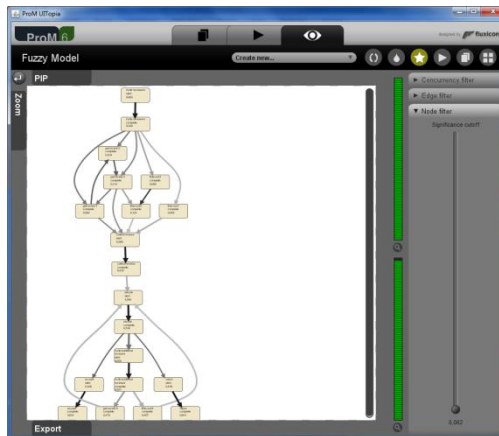
Alignments

- **conformance checking to diagnose deviations**
- **squeezing reality into the model to do model-based analysis**

How to get started?

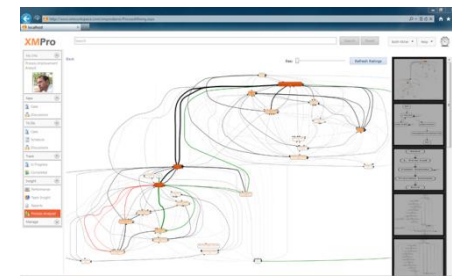
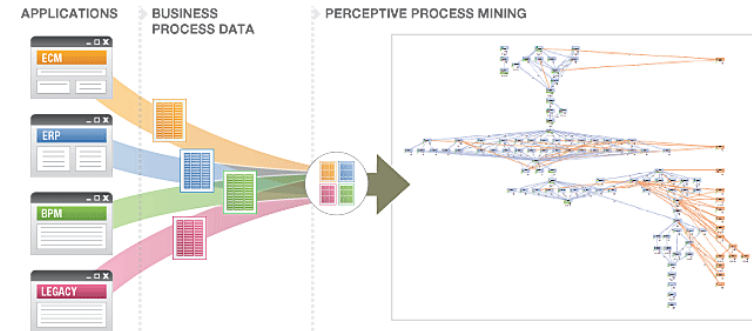
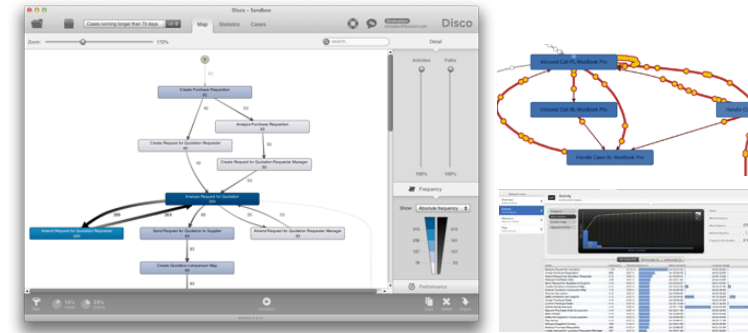


Hundreds of plug-ins available covering the whole process mining spectrum



Commercial Alternatives

- Disco (Fluxicon)
- Perceptive Process Mining (before Futura Reflect and BPM|one)
- ARIS Process Performance Manager
- QPR ProcessAnalyzer
- Interstage Process Discovery (Fujitsu)
- Discovery Analyst (StereoLOGIC)
- XMAalyzer (XMPro)
- ...

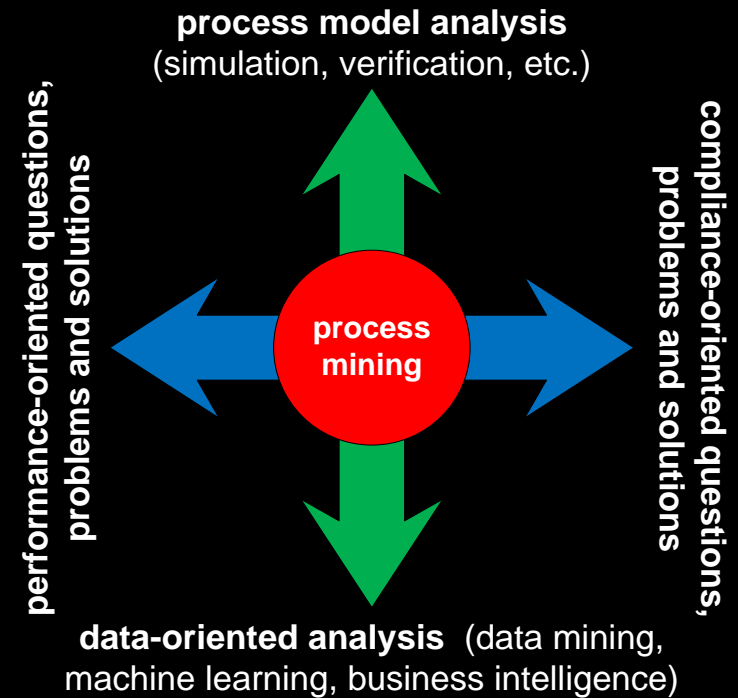


How to Get Started?

Collect event data ↔ Collect questions

- Minimal requirement: events referring to an activity name and a process instance.
- Good to have: timestamps, resource information, additional data elements.
- Challenges: scoping and sometimes correlation.
- What kind problems would you like to address (cost, time, risk, compliance, service, etc.)?
- Related to discovery, conformance, enhancement?
- Iterative process: can be “curiosity driven” initially.

Conclusion



Wil M. P. van der Aalst
Process Mining

Discovery, Conformance and Enhancement of Business Processes

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

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

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

Features and Benefits:

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

Computer Science

ISBN 978-3-642-19344-6



► springer.com

van der Aalst



Process Mining

Wil M. P. van der Aalst

Process Mining

Discovery, Conformance and
Enhancement of Business Processes

www.processmining.org

www.win.tue.nl/ieeetfpm/

 Springer