

Process Mining: Extension Mining Algorithms

Ana Karla Alves de Medeiros

Eindhoven University of Technology

Department of Information Systems

a.k.medeiros@tue.nl

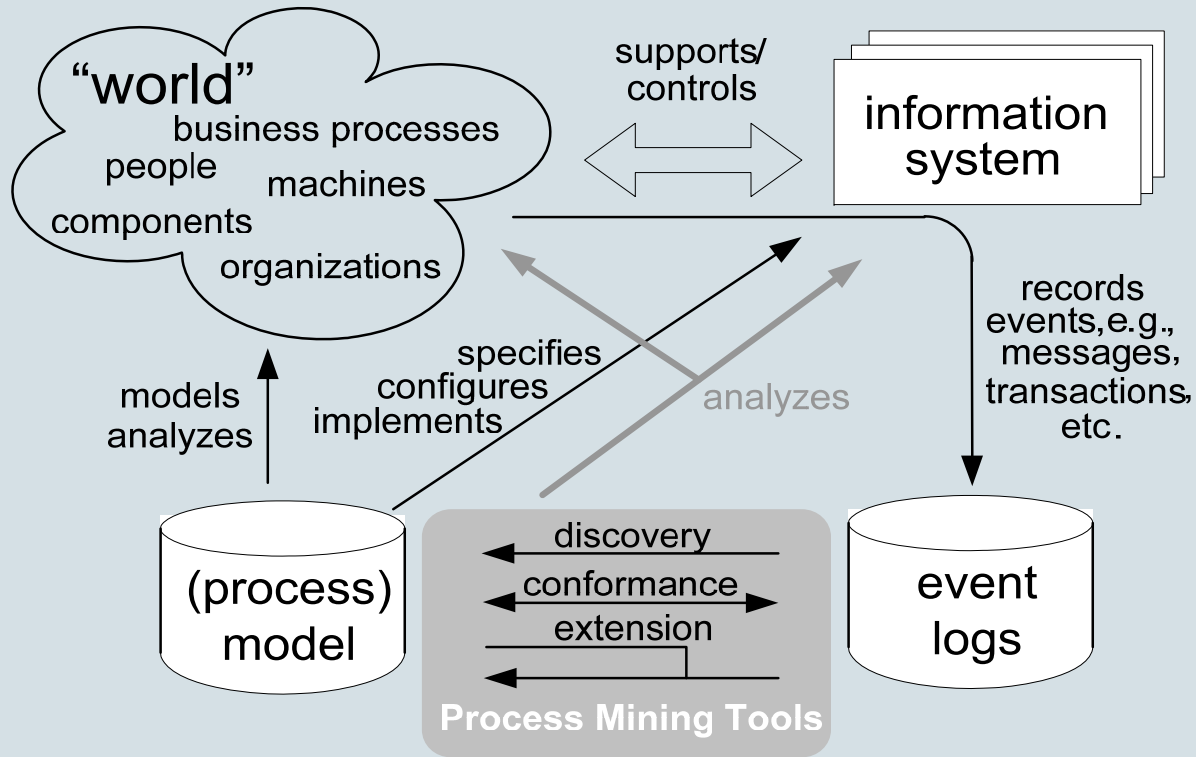
Process Mining

- Short Recap
- Extension Techniques
 - Decision Miner
 - Performance Analysis with Petri Nets
- Summary
- Announcements
- Presentation Futura Technology

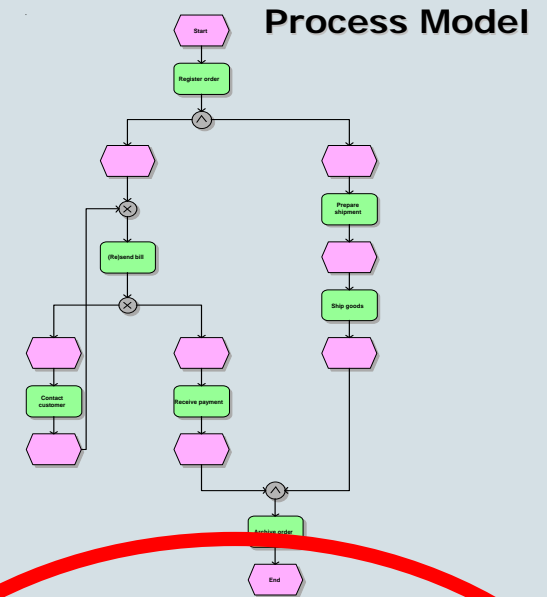
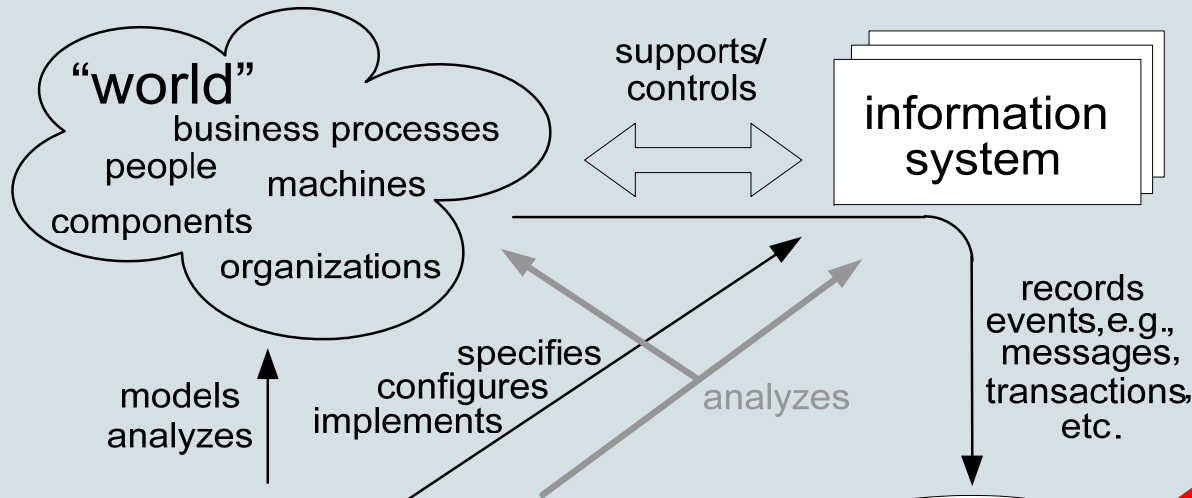
Process Mining

- **Short Recap**
- Extension Techniques
 - Decision Miner
 - Performance Analysis with Petri Nets
- Summary
- Announcements
- Presentation Futura Technology

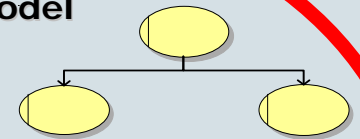
Types of Algorithms



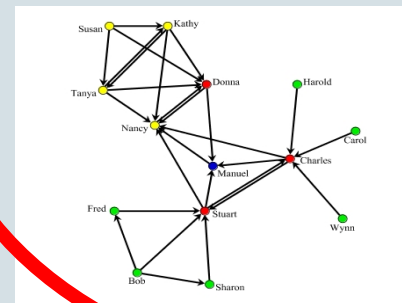
Types of Algorithms



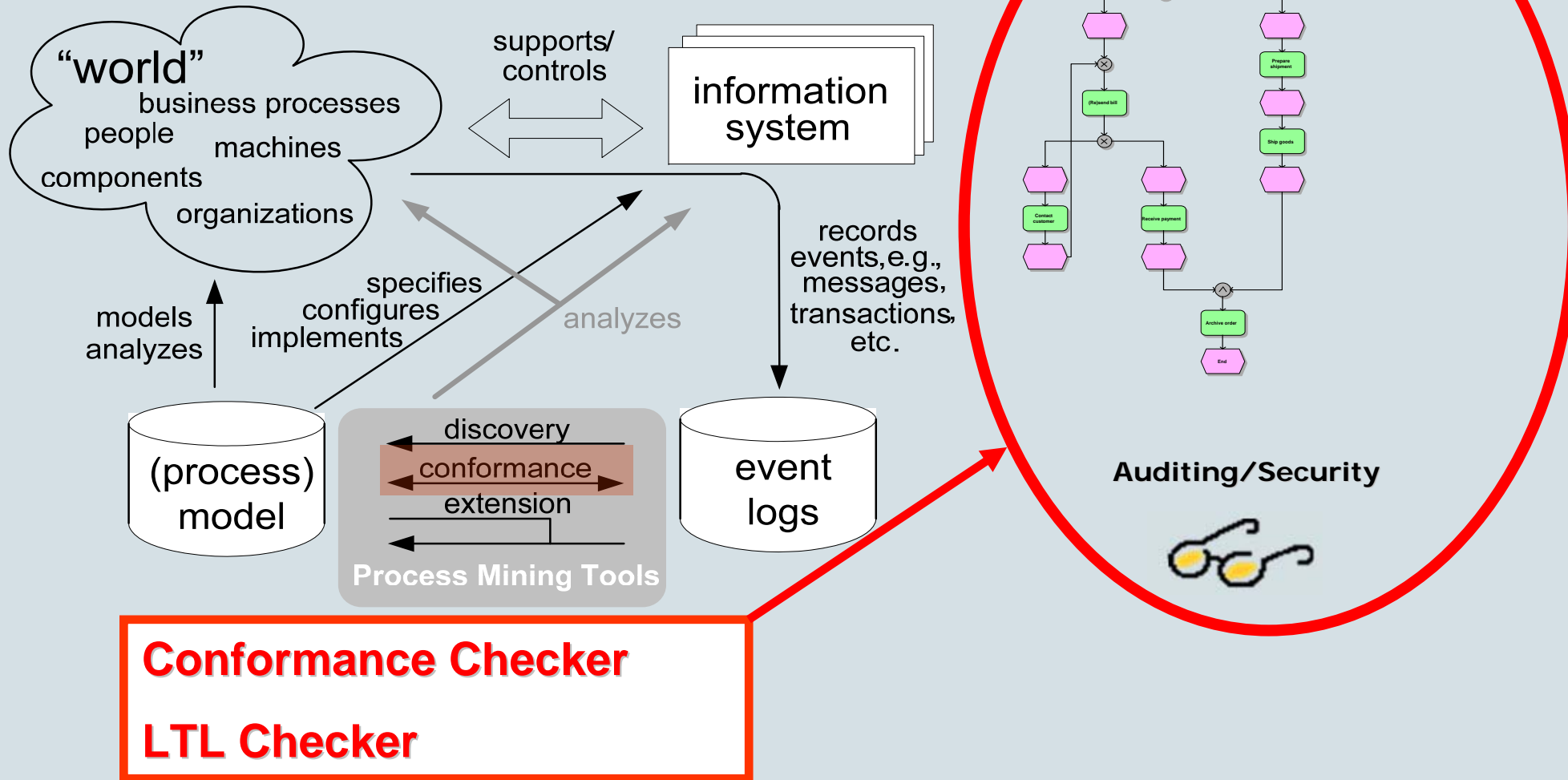
Organizational Model



Social Network



Types of Algorithms



Main Points Lecture 4

- Organizational mining plug-ins can discover
 - Roles/Teams in organizations
 - Social networks for originators
- Some metrics of social networks are based on ordering relations (e.g., the ordering relations used by the Alpha algorithm)
- Conformance Checker assesses how much a process model matches process instances
- LTL Checker uses logics to verify properties in event logs

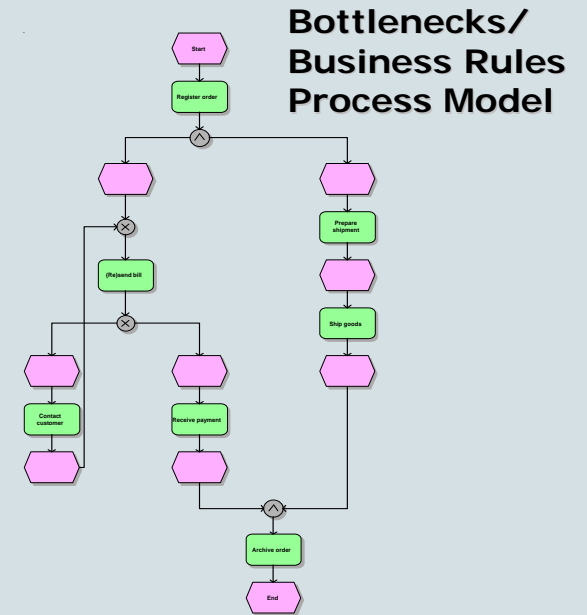
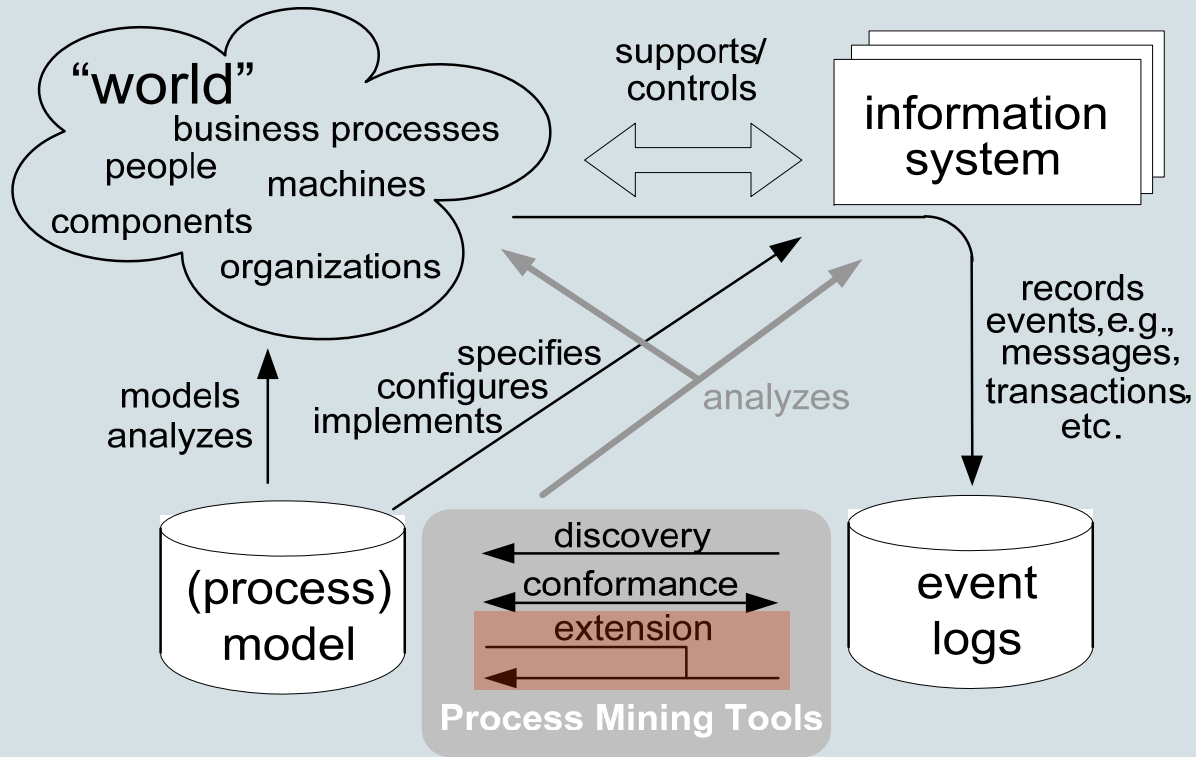
Process Mining

- Short Recap
- Extension Techniques
 - Decision Miner
 - Performance Analysis with Petri Nets
- Summary
- Announcements
- Presentation Futura Technology

Process Mining

- Short Recap
- **Extension Techniques**
 - Decision Miner
 - Performance Analysis with Petri Nets
- Summary
- Announcements
- Presentation Futura Technology

Types of Algorithms



Performance Analysis



Process Mining

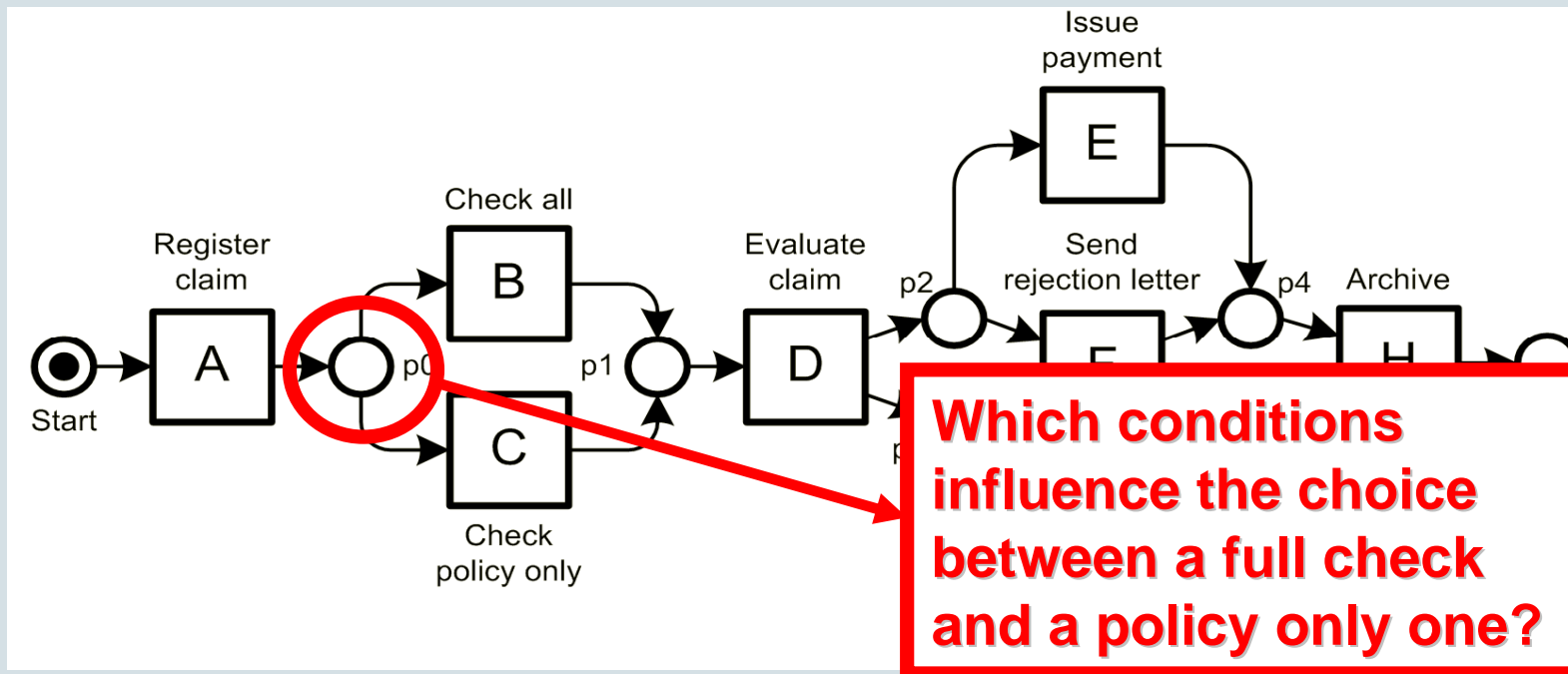
- Short Recap
- Extension Techniques
 - Decision Miner
 - Performance Analysis with Petri Nets
- Summary
- Announcements
- Presentation Futura Technology

Process Mining

- Short Recap
- Extension Techniques
 - Decision Miner
 - Performance Analysis with Petri Nets
- Summary
- Announcements
- Presentation Futura Technology

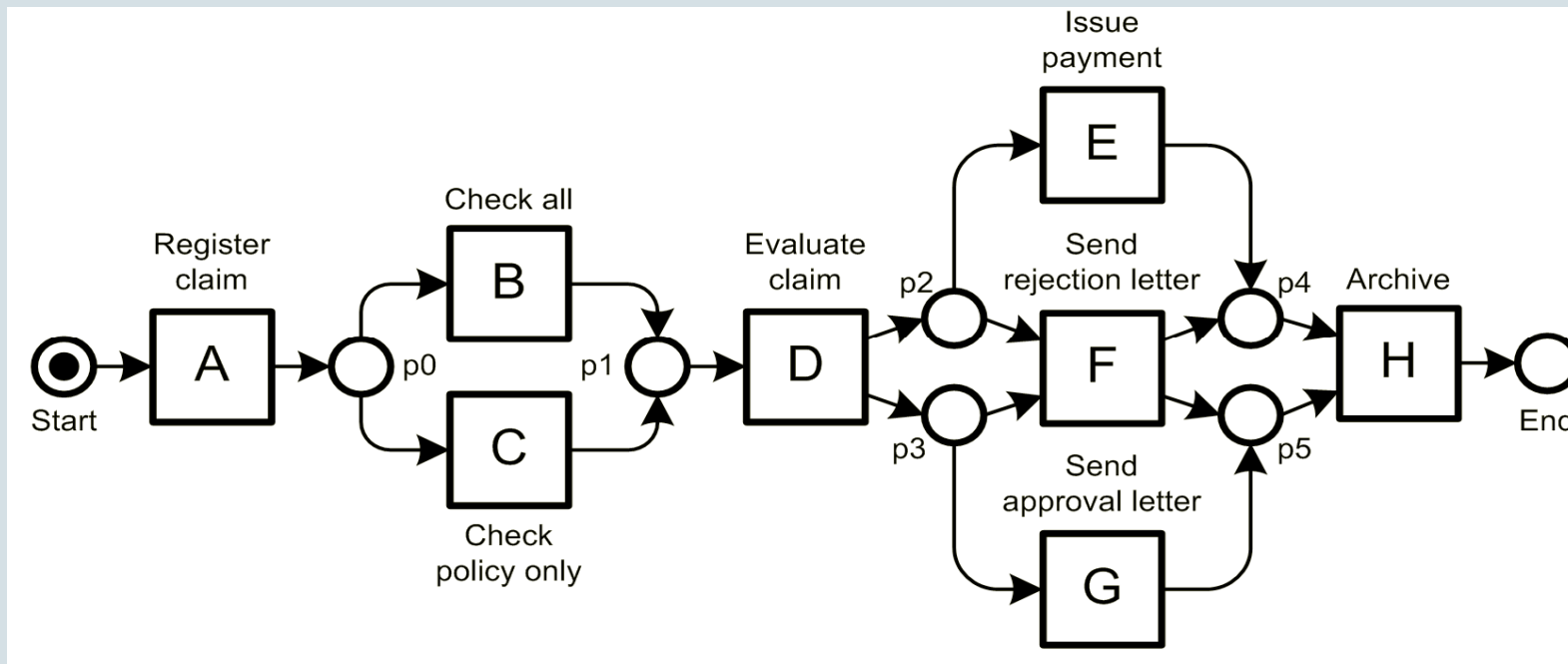
Decision Point Analysis: Main Idea

- Detection of data dependencies that affect the routing the routing of process instances



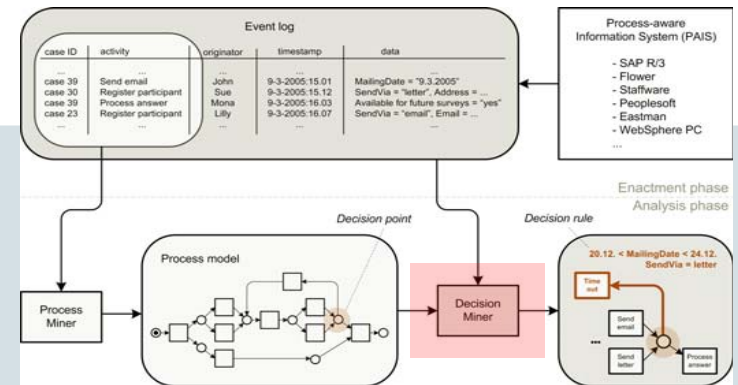
Decision Point Analysis: Motivation

- Make tacit knowledge explicit
- Better understand the process model



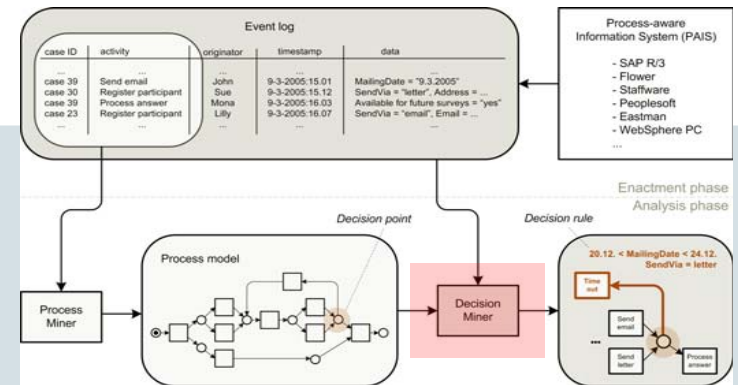
Decision Point Analysis: Algorithm's Main Steps

1. Read a log + model
2. Identify the decision points in a model
3. Find out which alternative branch has been taken for a given process instance and decision point
4. Discover the rules for each decision point
5. Return the enhanced model with the discovered rules



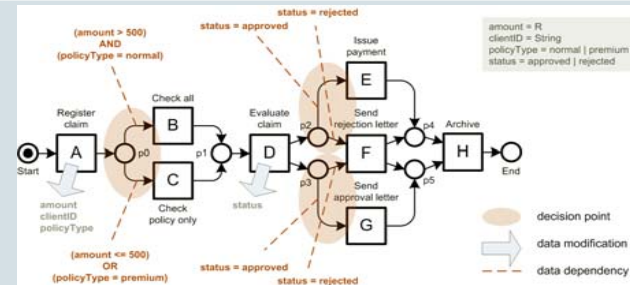
Decision Point Analysis: Algorithm's Main Steps

1. Read a log + model
2. Identify the decision points in a model
3. Find out which alternative branch has been taken for a given process instance and
4. How can we spot the decision points in a Petri net
5. Model with the discovered rules



Decision Point Analysis: Algorithm's Main Steps

1. Read a log + model
2. Identify the decision points in a model
3. Find out which alternative branch has been taken for a given process instance and decision point
4. Discover the rules for each decision point
5. Return the enhanced model with the discovered rules



Quick Recap Lecture 1: Decision Trees

Illustration (10 learning examples):

	Hair	Length	Weight	Suntan cream	Burned
1	blond	medium	light	yes	no
2	blond	medium	light	no	yes
			light		
			heavy		
5	blond	long	medium	yes	no
6	brown	long	light	no	no
7	red	small	heavy	no	yes
8	brown	long	light	yes	no
9	blond	medium	heavy	no	yes
10	brown	small	heavy	no	no

Attributes

Classes: Yes/No

New (test) examples:

1	red	medium	light	yes	yes
2	blond	medium	medium	no	yes
3	brown	small	light	yes	no

Decision Point Analysis: Algorithm's Main Steps

1. Read a log + model

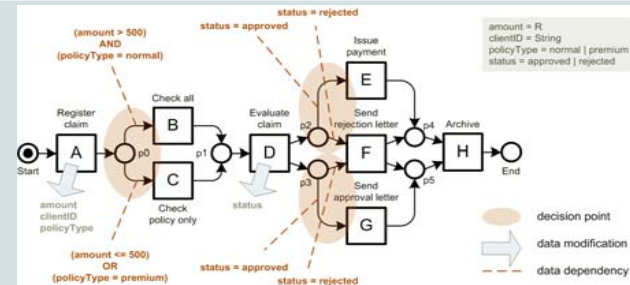
2. Identify

3. Find

taken for
decision point

4. Discover the rules for each decision point

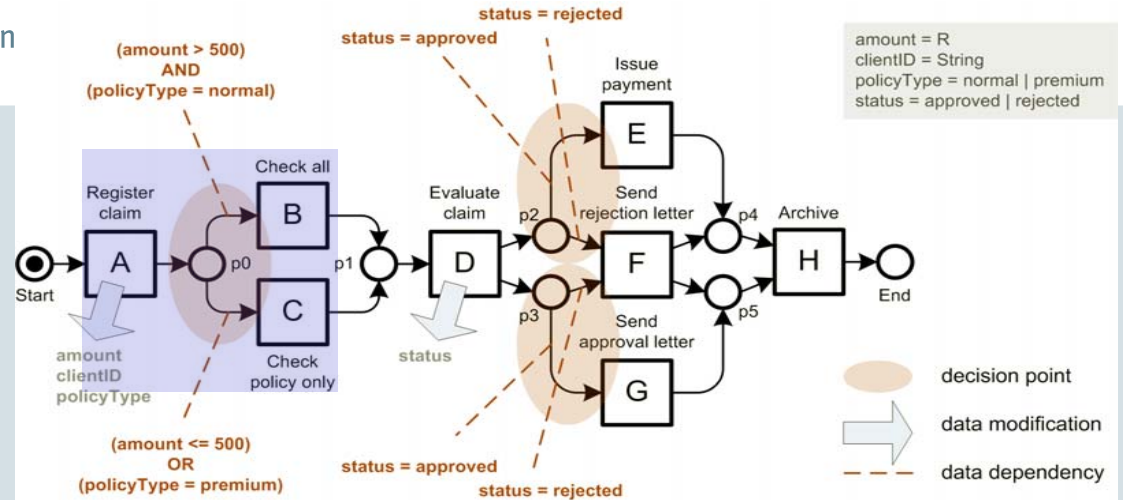
5. Return the enhanced model with the discovered rules



Which elements are the classes and which are the attributes?

model
has been

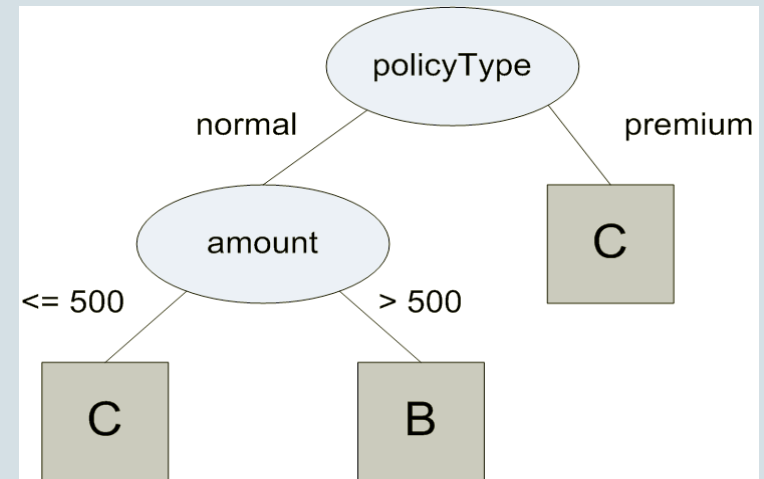
Step 4



Training examples for decision point "p0"

amount	clientID	policyType	class
1000	C567894938	premium	C
700	C938609223	normal	B
550	C135697567	normal	B
500	C568120443	normal	C
50	C493823084	normal	C
200	C945675110	premium	C

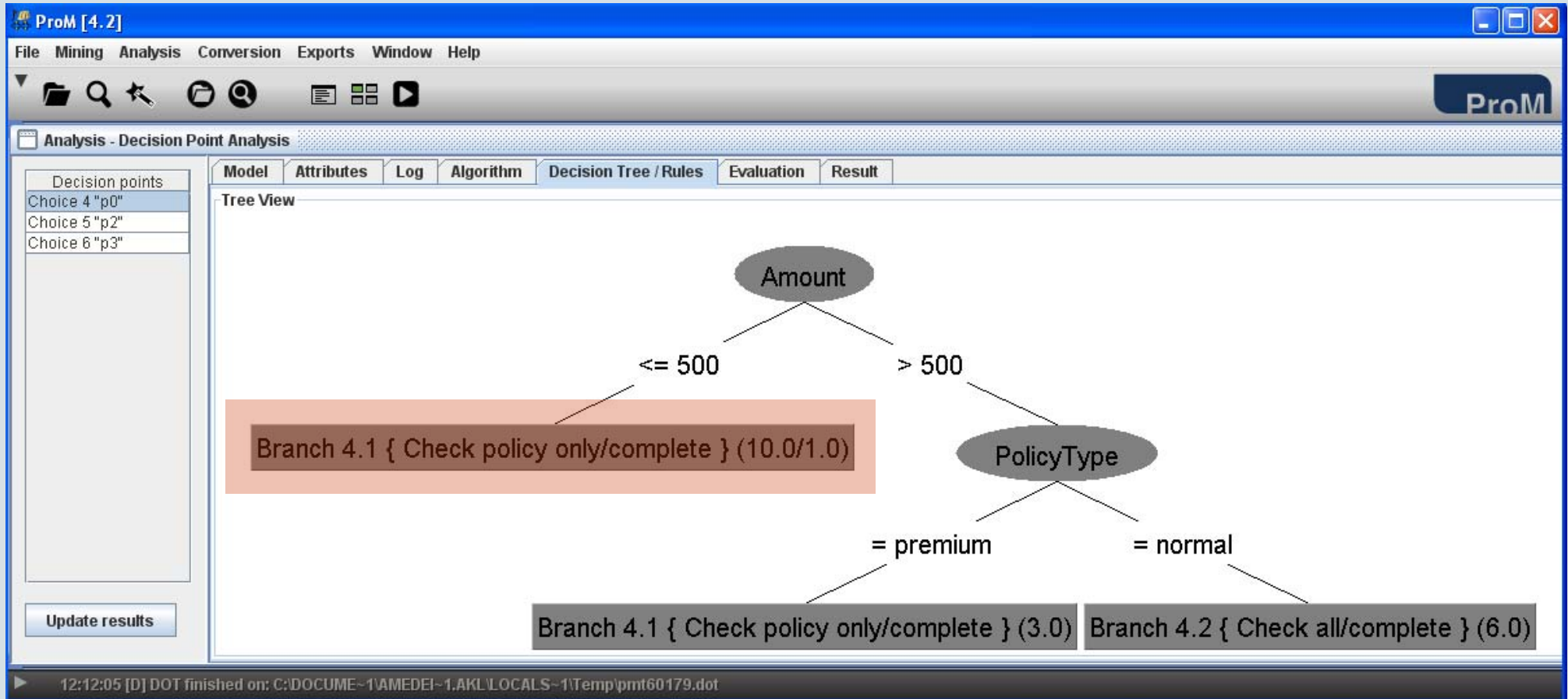
Discovered decision tree for point "p0"



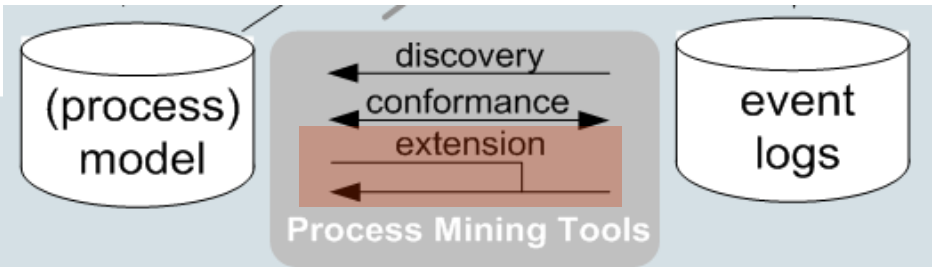
Decision Point Analysis: Example in ProM

The screenshot displays the ProM 4.2 interface. The main window shows a decision tree for the task 'Register Claim complete'. The tree starts with a root node that branches into two paths based on the 'PolicyType' attribute. The top path leads to a node with the condition $(\#Amount\ data > 500) \text{ andalso } (\#PolicyType\ data = \text{normal})$, which then leads to the 'Check all complete' task. The bottom path leads to a node with the condition $(\#Amount\ data \leq 500) \text{ or else } (\#Amount\ data > 500) \text{ andalso } (\#PolicyType\ data = \text{premium})$, which then leads to the 'Check policy only complete' task. A list of decision points on the left includes 'Choice 4 "p0"', 'Choice 5 "p2"', and 'Choice 6 "p3"'. The status bar at the bottom indicates the analysis was finished on 12:12:05 [D] at the path C:\DOCUME~1\AMEDEI~1\AKL\LOCALS~1\Temp\pmt60179.dot.

Decision Point Analysis: Example in ProM



Decision Point Analysis



The screenshot shows the ProM 4.2 interface with the 'Analysis - Decision Point Analysis' window open. The 'Model' tab is selected, showing a decision tree. The root node is 'Register Claim complete', which branches into two paths based on the 'PolicyType' attribute. The left path leads to a circular node with two outgoing edges: one to 'Check all complete' with the rule $(\#Amount\ data > 500)\ andalso\ (\#PolicyType\ data = normal)$, and another to 'Check policy only complete' with the rule $(\#Amount\ data > 500)\ andalso\ (\#PolicyType\ data = premium)$. The right path leads to a circular node with one outgoing edge to 'Check all complete' with the rule $(\#Amount\ data \le 500)$. The 'Result' tab is also visible, showing a list of decision points: Choice 4 "p0", Choice 5 "p2", and Choice 6 "p3".

Process Mining

- Short Recap
- Extension Techniques
 - Decision Miner
 - Performance Analysis with Petri Nets
- Summary
- Announcements
- Presentation Futura Technology

Process Mining

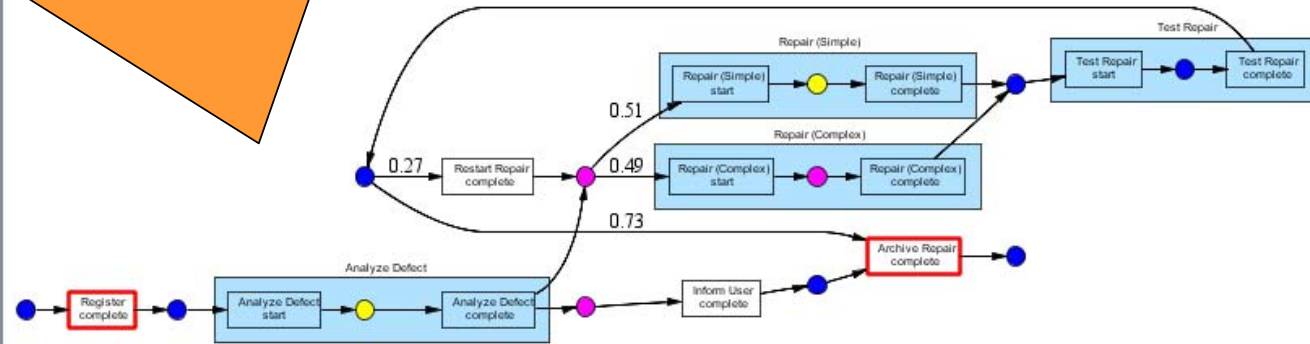
- Short Recap
- Extension Techniques
 - Decision Miner
 - Performance Analysis with Petri Nets
- Summary
- Announcements
- Presentation Futura Technology

Performance Analysis with Petri Nets

- Motivation
 - Provide different Key Performance Indicators (KPIs) relating to the execution of processes
- Main idea
 - Replay the log in a model and detect
 - Bottlenecks
 - Throughput times
 - Execution times
 - Waiting times
 - Synchronization times
 - Path probabilities etc

Bottlenecks – Waiting Times and Execution Times

How can we spot the difference between waiting and execution times?



Process information:

Total number selected:
1000 cases

Number fitting:
1000 cases

Arrival rate:
0.03 cases per minute

Throughput time (minutes)	
avg	66.59
min	29.0
max	159.0
stdev	19.45
fast 25...	45.7
slow 2...	93.87
norma...	63.39

Change Percentages Export Time-Metrics

Performance information of the selected transitions:

Frequency: 1000 cases

Time in between (minutes)	
avg	66.59
min	29.0
max	159.0
stdev	19.45
fast 25.00%...	45.7

Waiting time:

- High (pink)
- Medium (yellow)
- Low (blue)

Settings

Selected:

Transition - Register complete

and:

Transition - Archive Repair co...

Bottlenecks – Throughput Times

ProM [4.2]

File Mining Analysis Conversion Exports Window Help

Analysis - Performance Analysis with Petri net

Log Traces

1
10
100
1000
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115

Update
Invert Selection

Process information:

Total number selected:
1000 cases

Number fitting:
1000 cases

Arrival rate:
0.03 cases per minute

Throughput time (minutes)	
avg	66.59
min	29.0
max	159.0
stdev	19.45
fast 25...	45.7
slow 2...	93.87
norma...	63.39

Change Percentages Export Time-Metrics

Zoom: 84 %

Performance information of the selected transitions:
Frequency: 1000 cases

Time in between (minutes)	
avg	66.59
min	29.0
max	159.0
stdev	19.45
fast 25.00%...	45.7

Waiting time:
■ High
■ Medium
■ Low

Selected:
 Transition - Register complete
 and:
 Transition - Archive Repair co...

12:48:53 [D] Buffered log reader created from reader buffered log reader, pitk.: [!@17fe3f

Bottlenecks – Synchronization Times

ProM [4.2]

File Mining Analysis Conversion Exports Window Help

Analysis - Performance Analysis with Petri net

Log Traces

1
10
100
1000
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115

Update
Invert Selection

Process information:

Total number selected:
1000 cases

Number fitting:
1000 cases

Arrival rate:
0.03 cases per minute

Throughput time (minutes)	
avg	66.59
min	29.0
max	159.0
stdev	19.45
fast 25...	45.7
slow 2...	93.87
norma...	63.39

Change Percentages Export Time-Metrics

Zoom: 84 %

Performance information of the selected transitions:

Frequency: 1000 cases

Time in between (minutes)	
avg	66.59
min	29.0
max	159.0
stdev	19.45
fast 25.00%	45.7

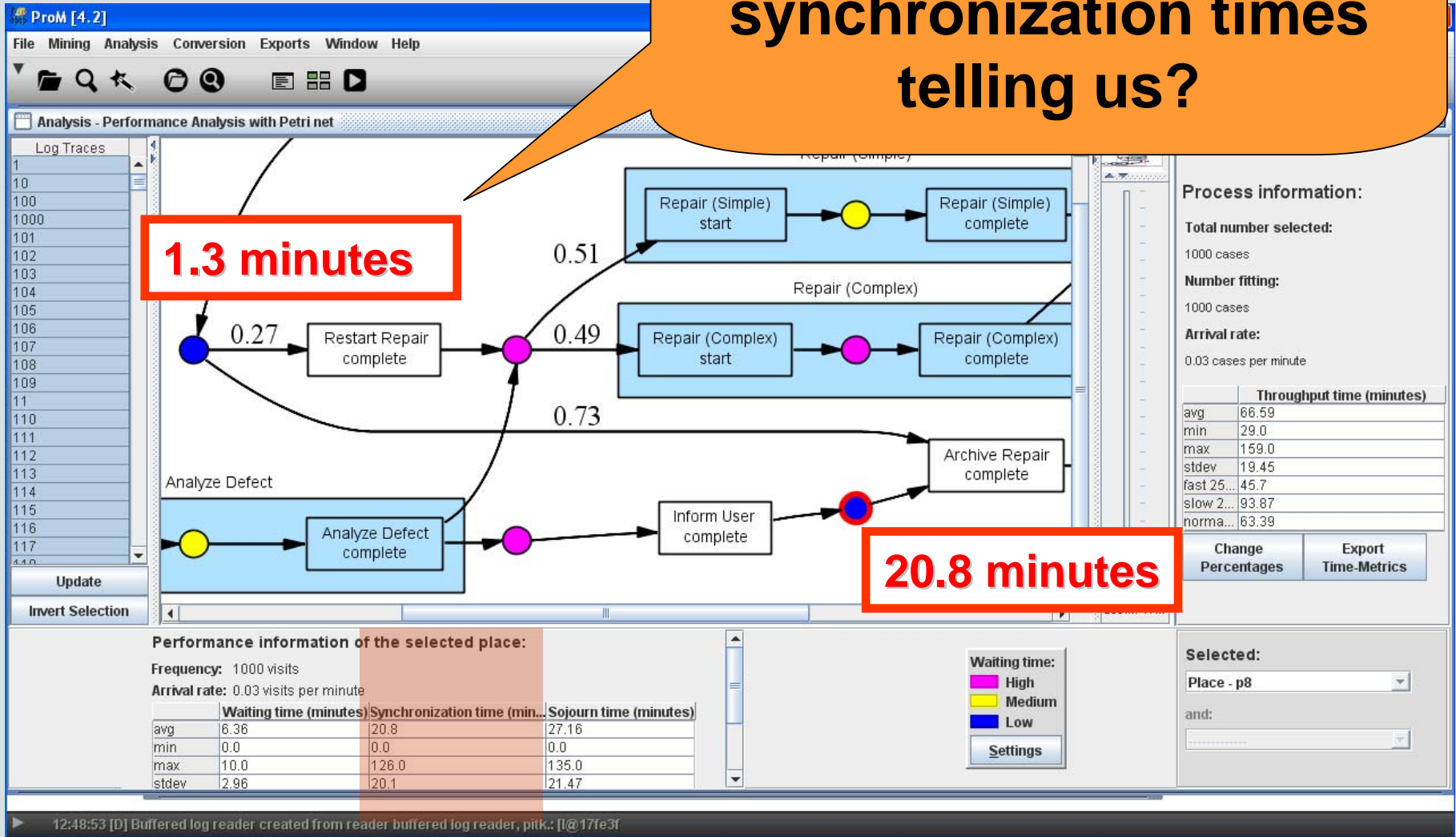
Waiting time:
■ High
■ Medium
■ Low
 Settings

Selected:
 Transition - Register complete
 and:
 Transition - Archive Repair co...

12:48:53 [D] Buffered log reader created from reader buffered log reader, pitk.: [!@17fe3f

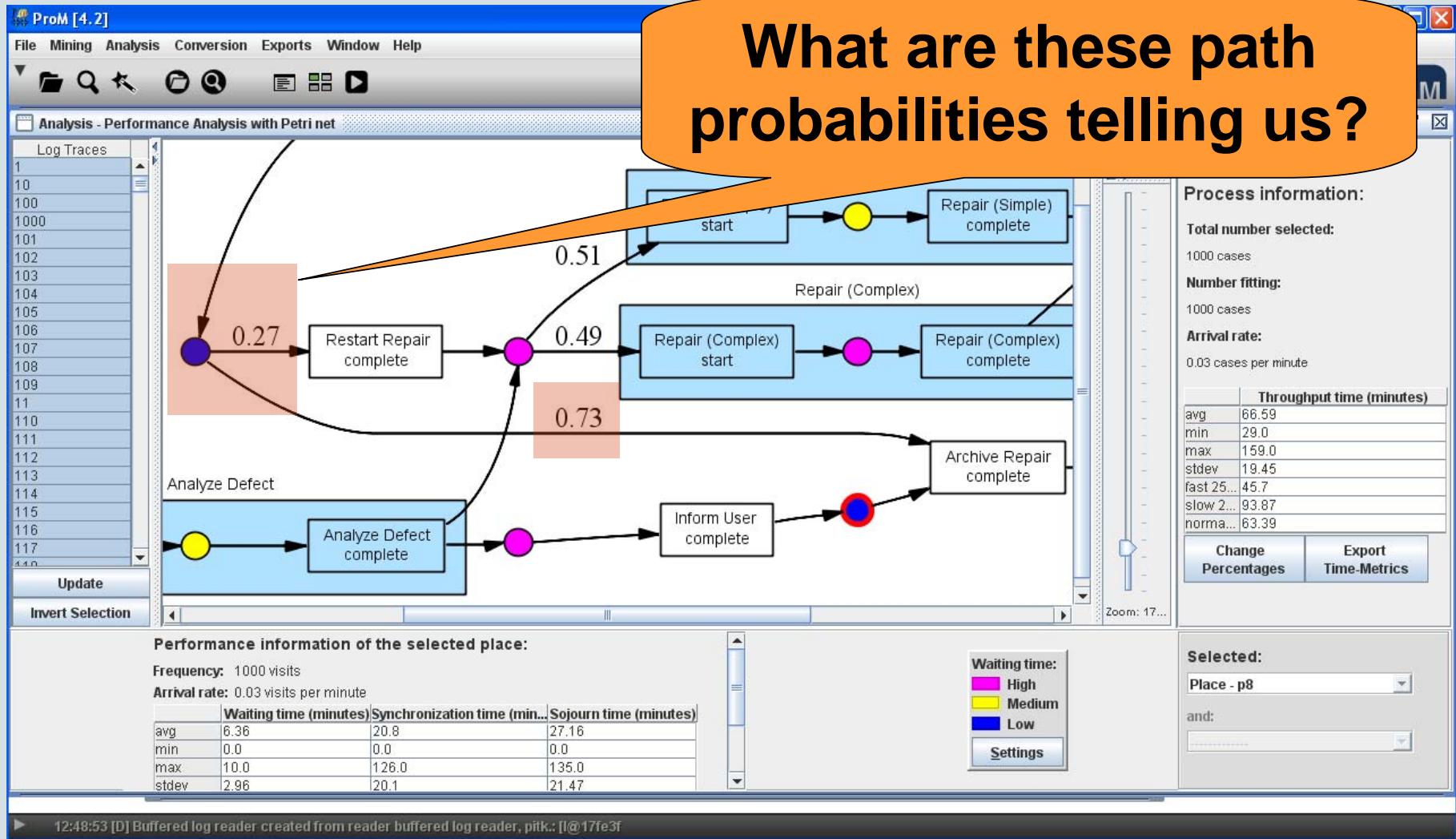
Bottlenecks – Synchron

What are these average synchronization times telling us?

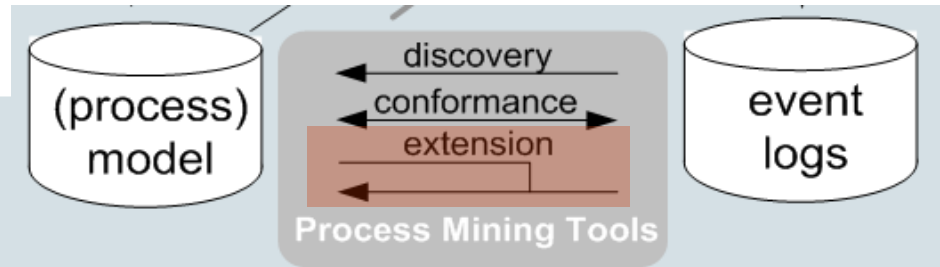


Bottlenecks – Path Probabilities

What are these path probabilities telling us?



Performance Analysis with Petri Nets



ProM [4.2]

File Mining Analysis Conversion Exports Window Help

Petri net model
Performance Simulation Model No.2 CPN Tools 2.0

Analysis - Performance Analysis

Log Traces

1
10
100
1000
101
102
103
104
105
106
107
108
109
11
110
111
112
113
114
115
116
117
4.4.0

Update
Invert Selection

Repair (Simple)
Repair (Simple) start → [Yellow Circle] → Repair (Simple) complete

Repair (Complex)
Repair (Complex) start → [Pink Circle] → Repair (Complex) complete

Analyze Defect
[Yellow Circle] → Analyze Defect complete

Restart Repair complete → [Pink Circle]

0.27

0.51

0.49

0.73

Archive Repair complete

Inform User complete

Process information:

Total number selected:
1000 cases

Number fitting:
1000 cases

Arrival rate:
0.03 cases per minute

Throughput time (minutes)	
avg	66.59
min	29.0
max	159.0
stdev	19.45
fast 25...	45.7
slow 25...	93.87
norma...	63.39

Change Percentages Export Time-Metrics

Performance information of the selected place:

Frequency: 1369 visits
Arrival rate: 0.04 visits per minute

	Waiting time (minutes)	Synchronization time (min...)	Sojourn time (minutes)
avg	6.7	1.37	8.06
min	0.0	0.0	5.0
max	10.0	31.0	31.0
stdev	2.74	4.17	2.86

Waiting time:
High (Pink)
Medium (Yellow)
Low (Blue)

Settings

Selected:
Place - p7

and:

13:12:18 [D] Buffered log reader created from reader buffered log reader, pitk: [!@49b9ad]

Process Mining

- Short Recap
- Extension Techniques
 - Decision Miner
 - Performance Analysis with Petri Nets
- Summary
- Announcements
- Presentation Futura Technology

Process Mining

- Short Recap
- Extension Techniques
 - Decision Miner
 - Performance Analysis with Petri Nets
- **Summary**
- Announcements
- Presentation Futura Technology

Summary

- Extension techniques enhance existing models with information discovered from event logs
- The *Decision Point Analysis* plug-in can discover the “business rules” for the moments of choice in a process model
- The *Performance Analysis with Petri Nets* plug-in provides various KPIs w.r.t. the execution of processes
- The results of both techniques can be used to create simulation models for CPN Tools

Process Mining

- Short Recap
- Extension Techniques
 - Decision Miner
 - Performance Analysis with Petri Nets
- Summary
- Announcements
- Presentation Futura Technology

Process Mining

- Short Recap
- Extension Techniques
 - Decision Miner
 - Performance Analysis with Petri Nets
- Summary
- **Announcements**
- Presentation Futura Technology

Announcements

- Assignment 5
 - Individual assignment
 - Q&A session during Instruction 5
 - Posting of Report with Answers
 - Digital version at StudyWeb (folder Assignment 5)
 - Printed version to be delivered at secretary's office of IS group (room Pav D3)
 - There will be a box on the desk
 - Deadline: March 14th, 2008 at 6pm
- Invited talk after the break!