

A fresh look at Precision in Process Conformance

Jorge Muñoz-Gama

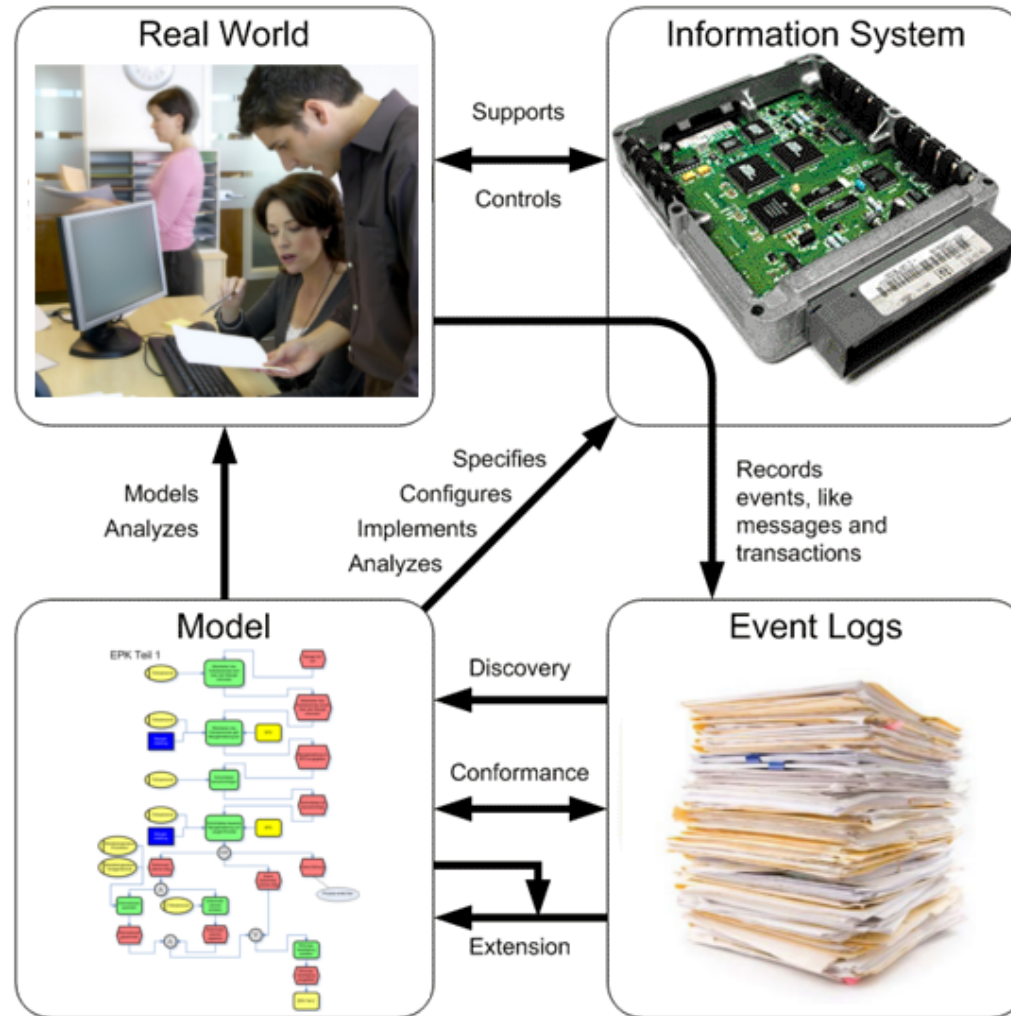
Josep Carmona

*Universitat Politècnica de Catalunya
(Barcelona, Spain)*

Outline

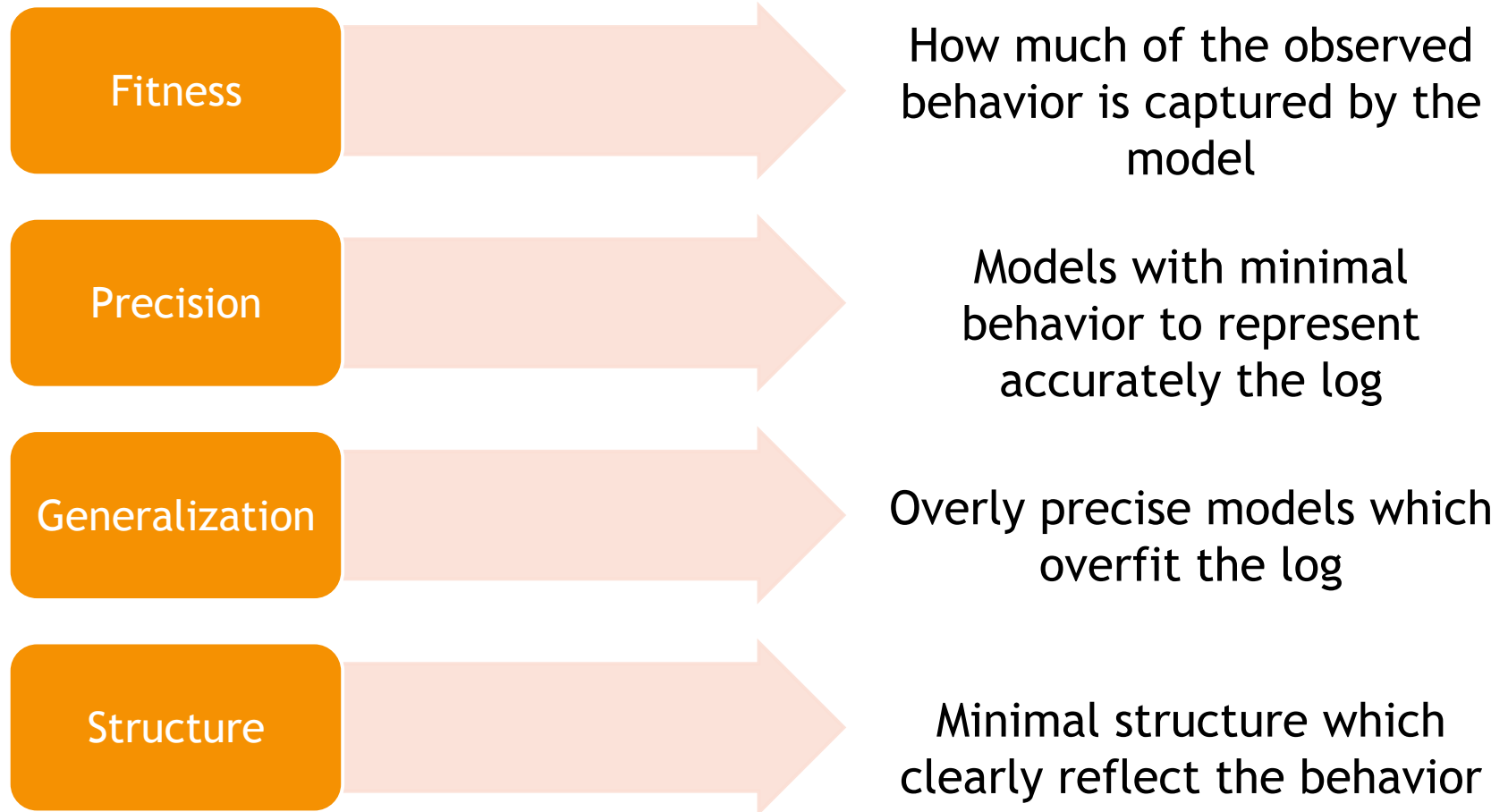
- ▶ Process Mining and Process Conformance
- ▶ Motivation
- ▶ Approach
 - ▶ General Approach
 - ▶ Implementation
 - ▶ Results
 - ▶ Extensions
- ▶ Future work
- ▶ Conclusions

Process Mining



* www.processmining.org

Conformance Dimensions



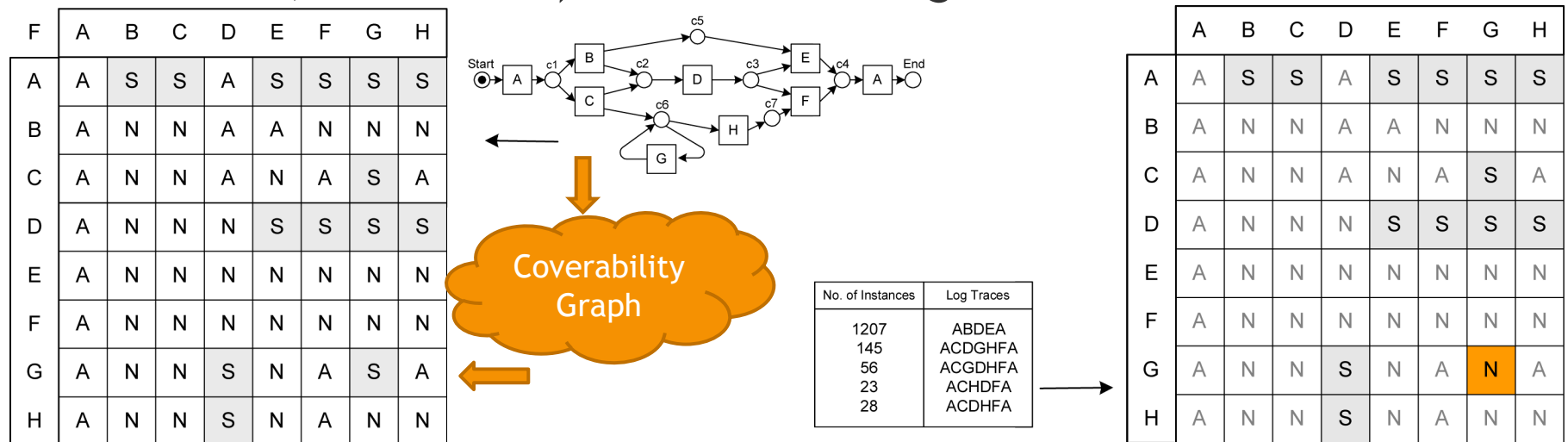
Outline

- ▶ Process Mining and Process Conformance
- ▶ ***Motivation***
- ▶ Approach
 - ▶ General Approach
 - ▶ Implementation
 - ▶ Results
 - ▶ Extensions
- ▶ Future work
- ▶ Conclusions

Related Work

► Precision in the literature

- Most related work *Rozinat et al. Information System 33 (2008)*
- Metric a'_B for Precision in Petri Nets
- Computation of **Follows** and **Precedes** relations (*Always, Never, Sometimes*) of Model and Log.

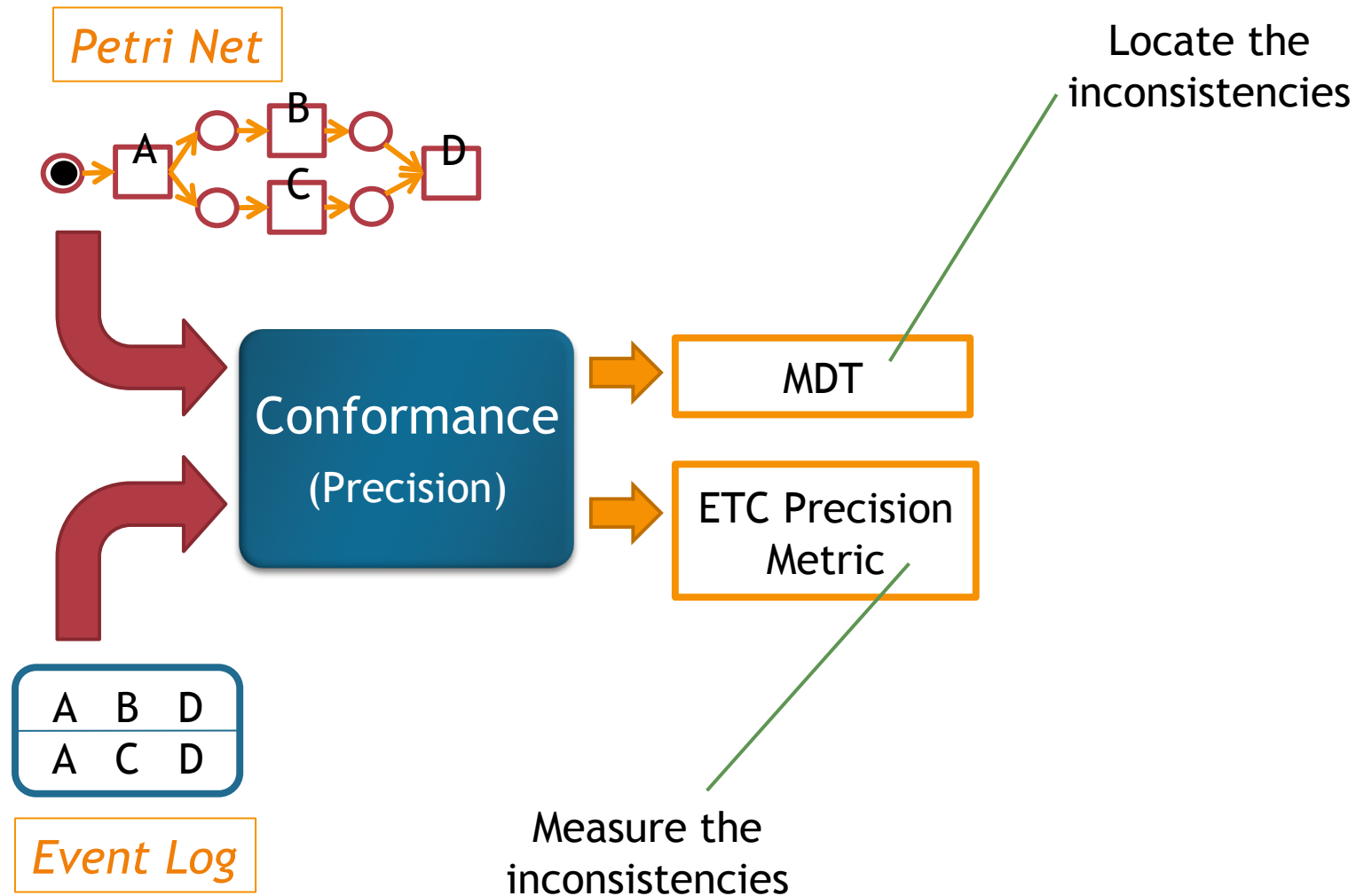


- Measurement based on discrepancies in Sometimes relations
- Model relations require a model space state exploration

Motivation

- ▶ **Goals and Requirements**
 - ▶ Precision Dimension
 - ▶ Petri Nets
 - ▶ Avoid the complete state space exploration
 - ▶ Effort needed to obtain an accurate model
 - ▶ Fine-level precision
 - ▶ Locate the precision inconsistencies

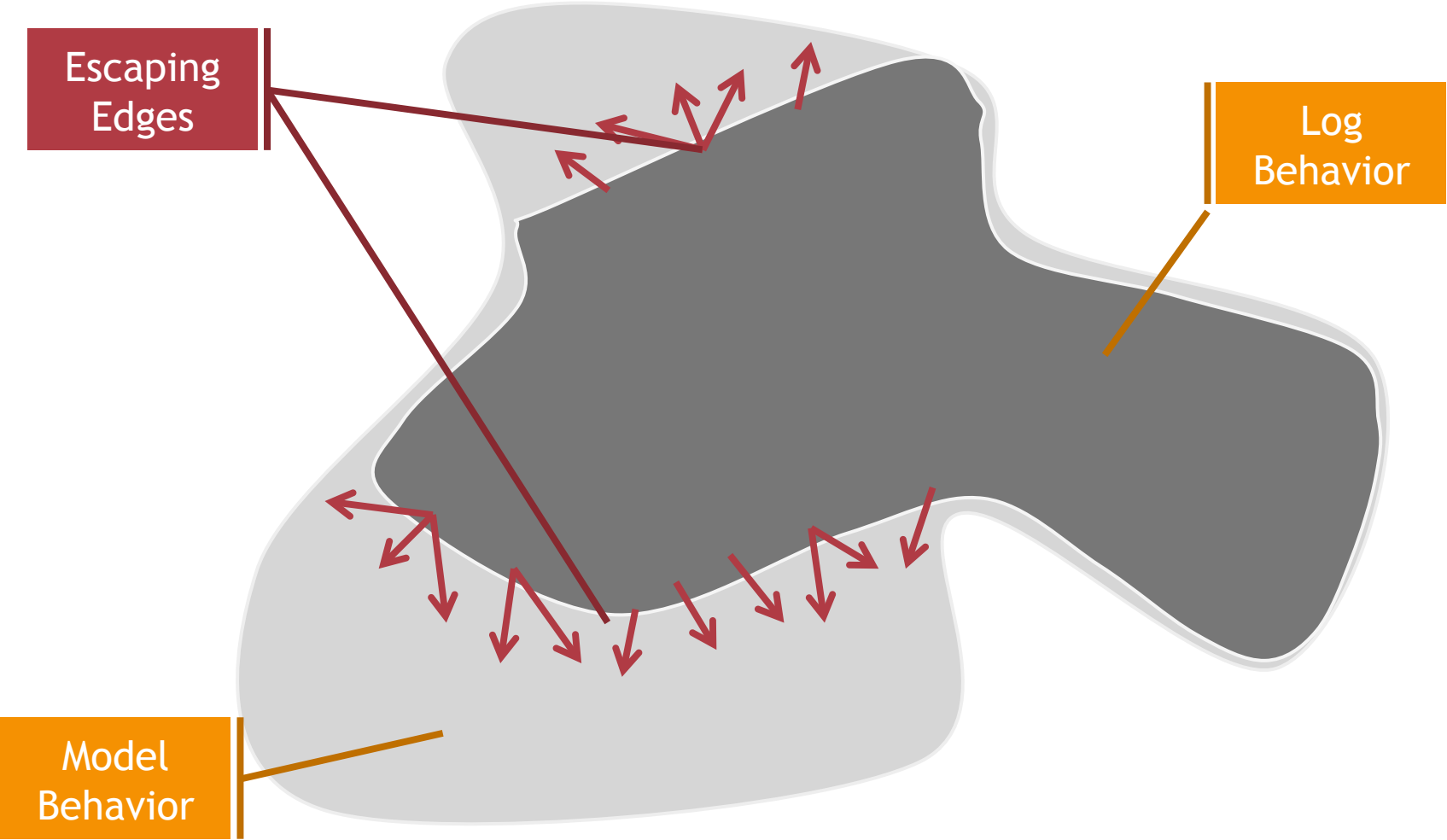
Process Conformance and Refinement



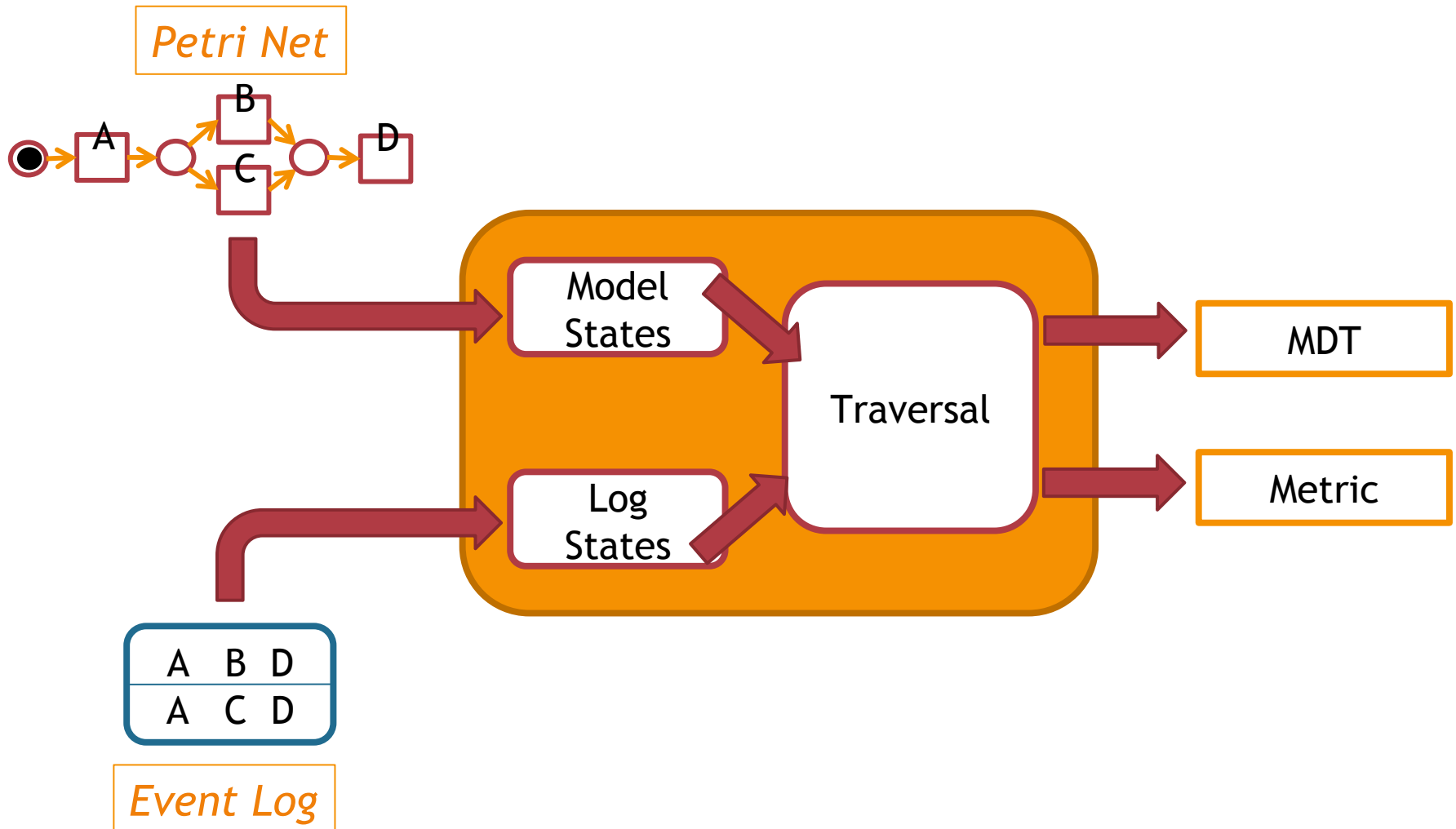
Outline

- ▶ Process Mining and Process Conformance
- ▶ Motivation
- ▶ ***Approach***
 - ▶ General Approach
 - ▶ Implementation
 - ▶ Results
 - ▶ Extensions
- ▶ Future work
- ▶ Conclusions

General Idea: Escaping Edges



Conformance Route Map

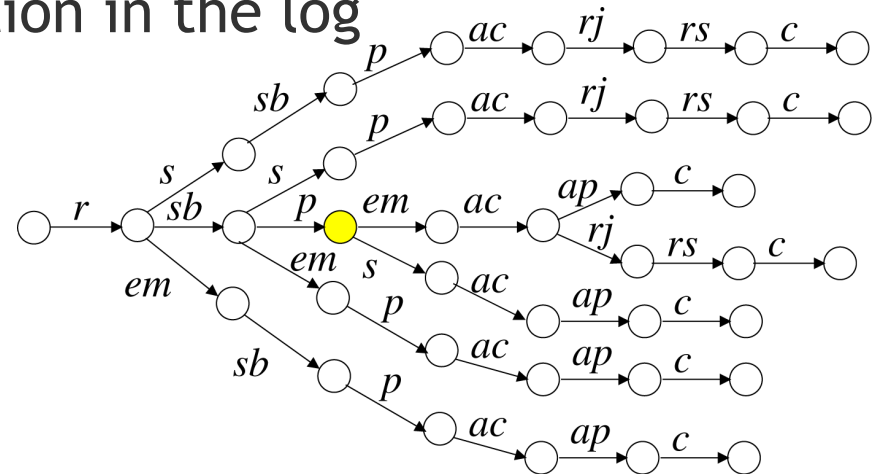


Log and Model States

Log

- ▶ Incorporate state information in the log

1	r, s, sb, p, ac, ap, c
2	r, sb, em, p, ac, ap, c
3	$r, sb, p, em, ac, rj, rs, c$
4	r, em, sb, p, ac, ap, c
5	$r, sb, s, p, ac, rj, rs, c$
6	r, sb, p, s, ac, ap, c
7	r, sb, p, em, ac, ap, c



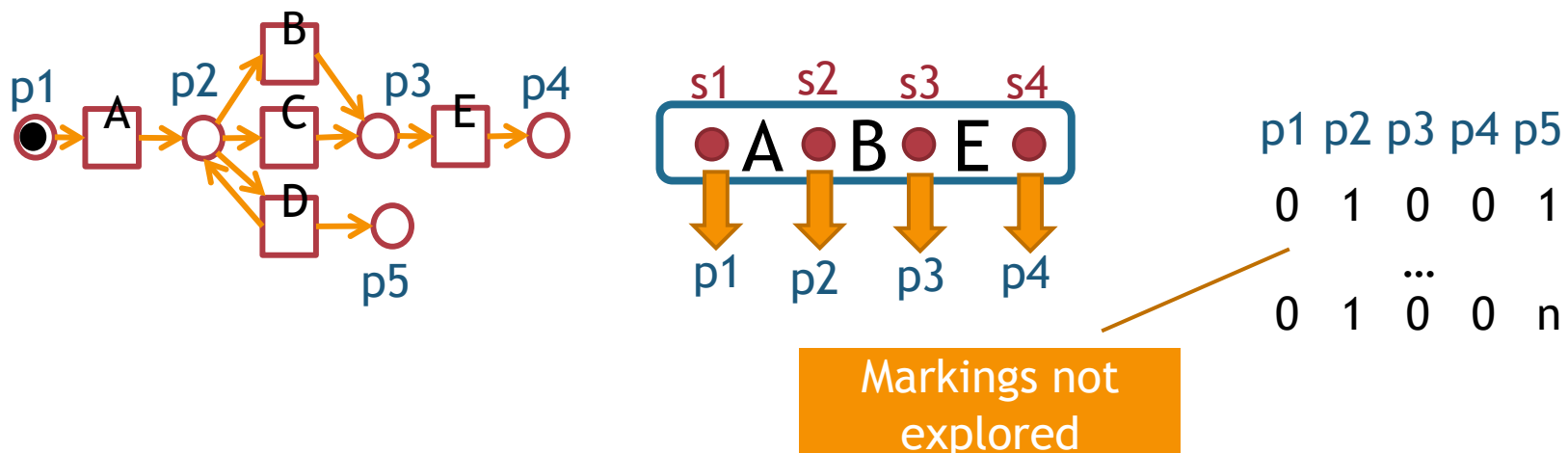
- ▶ (Aalst et al. *Software and Systems Modeling*, 2009)
 - ▶ Past, Unlimited and Sequence

Model

- ▶ Markings of the Petri Net

Model States and Mapping

- ▶ Not all the reachable markings (could be infinite)
- ▶ Only Markings with a Log State mapped on
- ▶ Log and Model States Mapping
 - ▶ $M \dashv\circ s$, if there exists a trace σ such that $s_0 \xrightarrow{\sigma} s$ in TS and $M_0[\sigma]M$
 - ▶ i.e., reached marking after replay state prefix

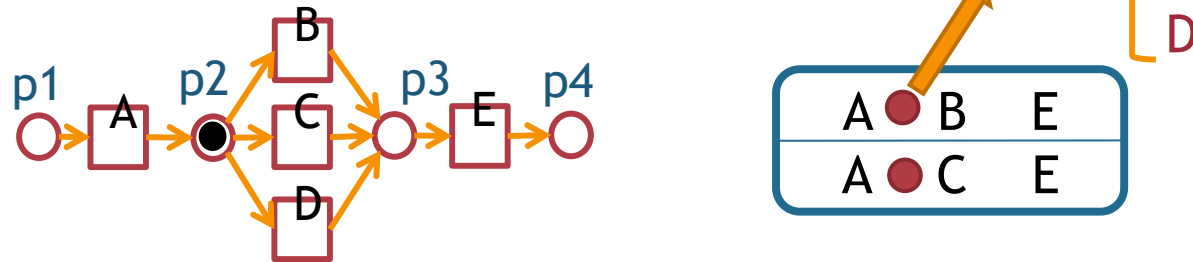


Log-guided Traversal

- ▶ Log-guided Traversal of Model Behavior

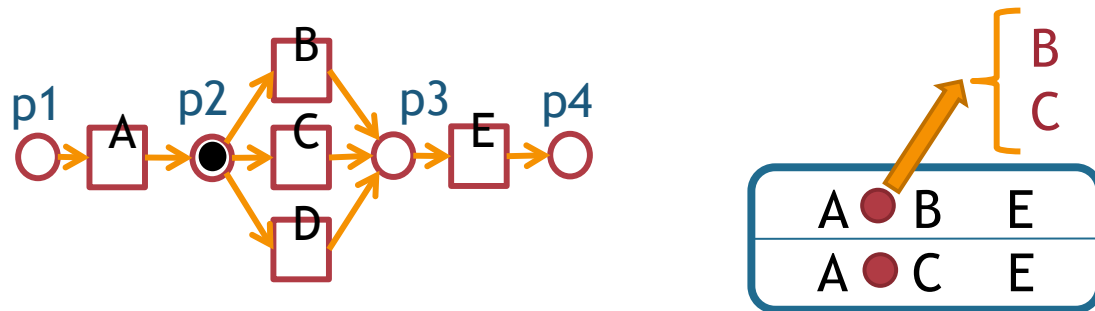
- ▶ **Allowed Tasks** : $A_T(s) = \{t \in T \mid M \xrightarrow{t} s \wedge M[t]M'\}$

- ▶ i.e., actions enabled in that moment



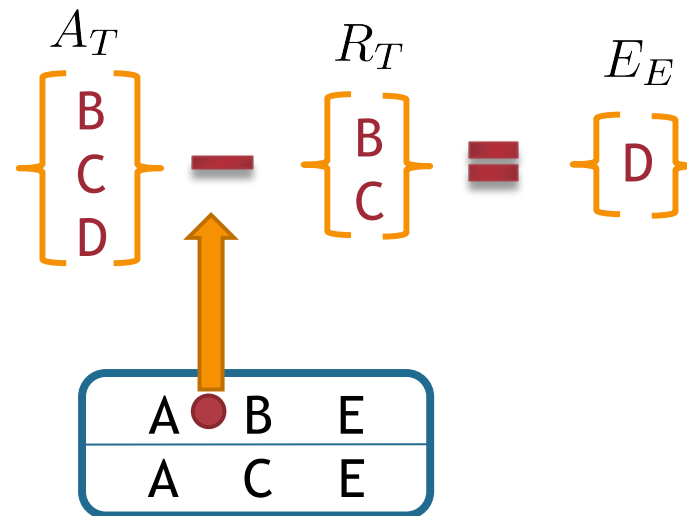
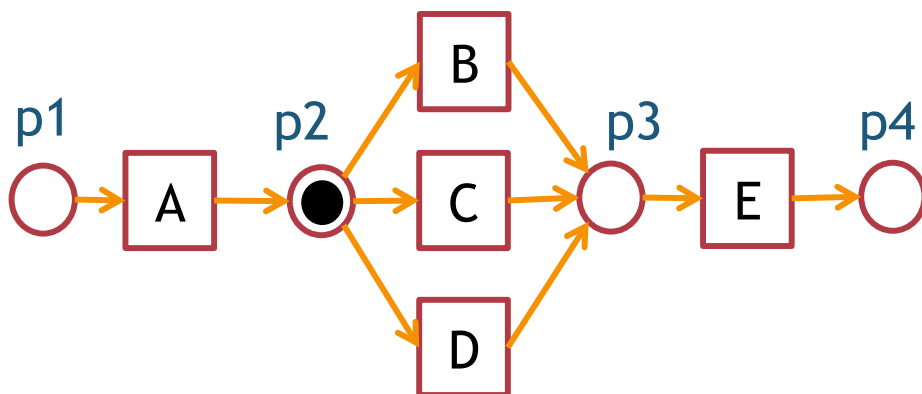
- ▶ **Reflected Tasks** : $R_T(s) = \{t \in T \mid s \xrightarrow{t} s'\}$

- ▶ i.e., actions really executed (thus, annotated in the log)



Traversal (2)

- ▶ **Escaping Edges** : $E_E(s) = A_T(s) \setminus R_T(s)$
 - ▶ i.e., enabled actions not executed
 - ▶ Precision discrepancies

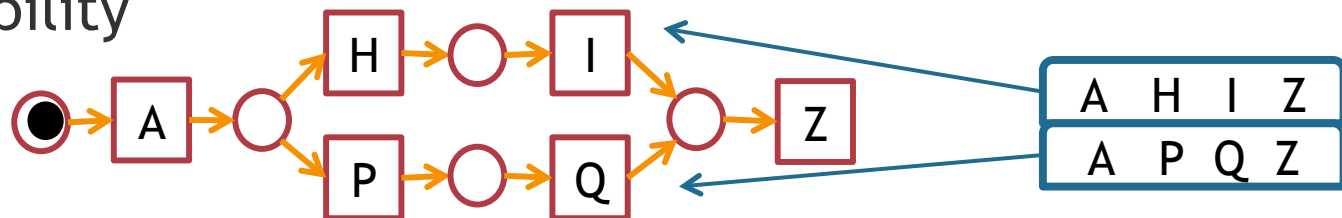


Precision Metric

- ▶ Take into account the Escaping Edges

$$etc_P(EL, PN) = 1 - \frac{\sum_{i=1}^{|\text{EL}|} \sum_{j=1}^{|\sigma_i|+1} |E_E(s_j^i)|}{\sum_{i=1}^{|\text{EL}|} \sum_{j=1}^{|\sigma_i|+1} |A_T(s_j^i)|}$$

- ▶ Between 0 (*imprecise*) and 1 (*precise*)
- ▶ More frequent traces, more weight in the metric
- ▶ Independent of Structural dimension
- ▶ Global precision
- ▶ Localizability

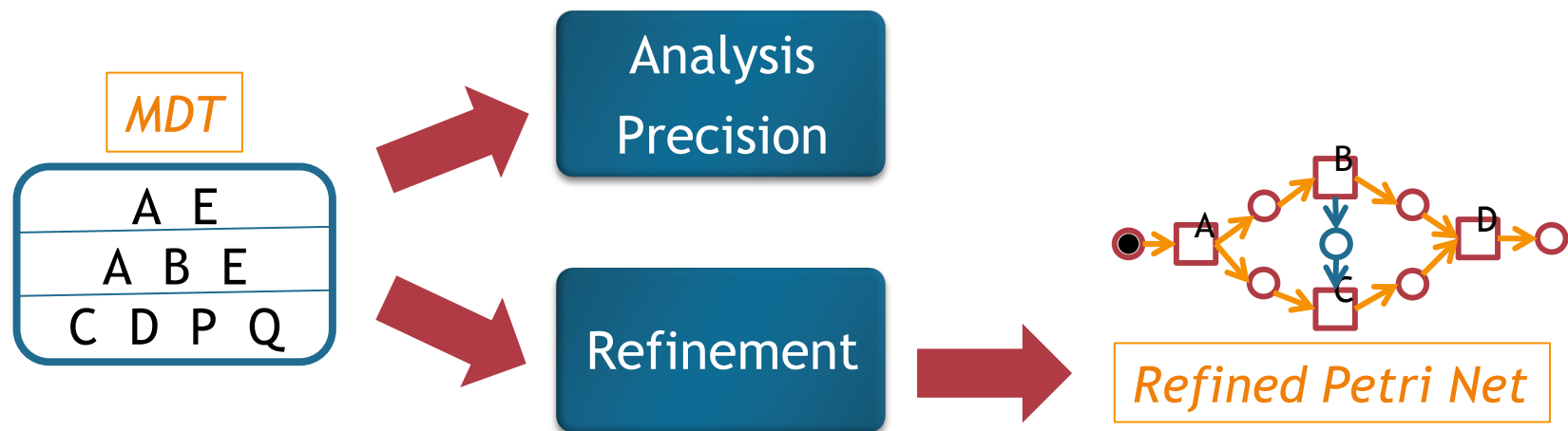


Minimal Disconformant Traces (MDT)

- ▶ Localizability of precision inconsistencies

Set of traces $\sigma = \sigma' t$ such that $M_0[\sigma]M$, $\sigma' \in Pref(EL)$ and $\sigma \notin Pref(EL)$

- ▶ i.e., Minimal traces indicating where the model starts to deviate from the log
- ▶ Algorithm to compute all **MDT** using **Escaping Edges**



Outline

- ▶ Process Mining and Process Conformance
- ▶ Motivation
- ▶ Approach
 - ▶ General Approach
 - ▶ ***Implementation***
 - ▶ Results
 - ▶ Extensions
- ▶ Future work
- ▶ Conclusions

Implementation



► ProM 6 Framework

Conformance Plug-In

Outline

- ▶ Process Mining and Process Conformance
- ▶ Motivation
- ▶ Approach
 - ▶ General Approach
 - ▶ Implementation
 - ▶ ***Results***
 - ▶ Extensions
- ▶ Future work
- ▶ Conclusions

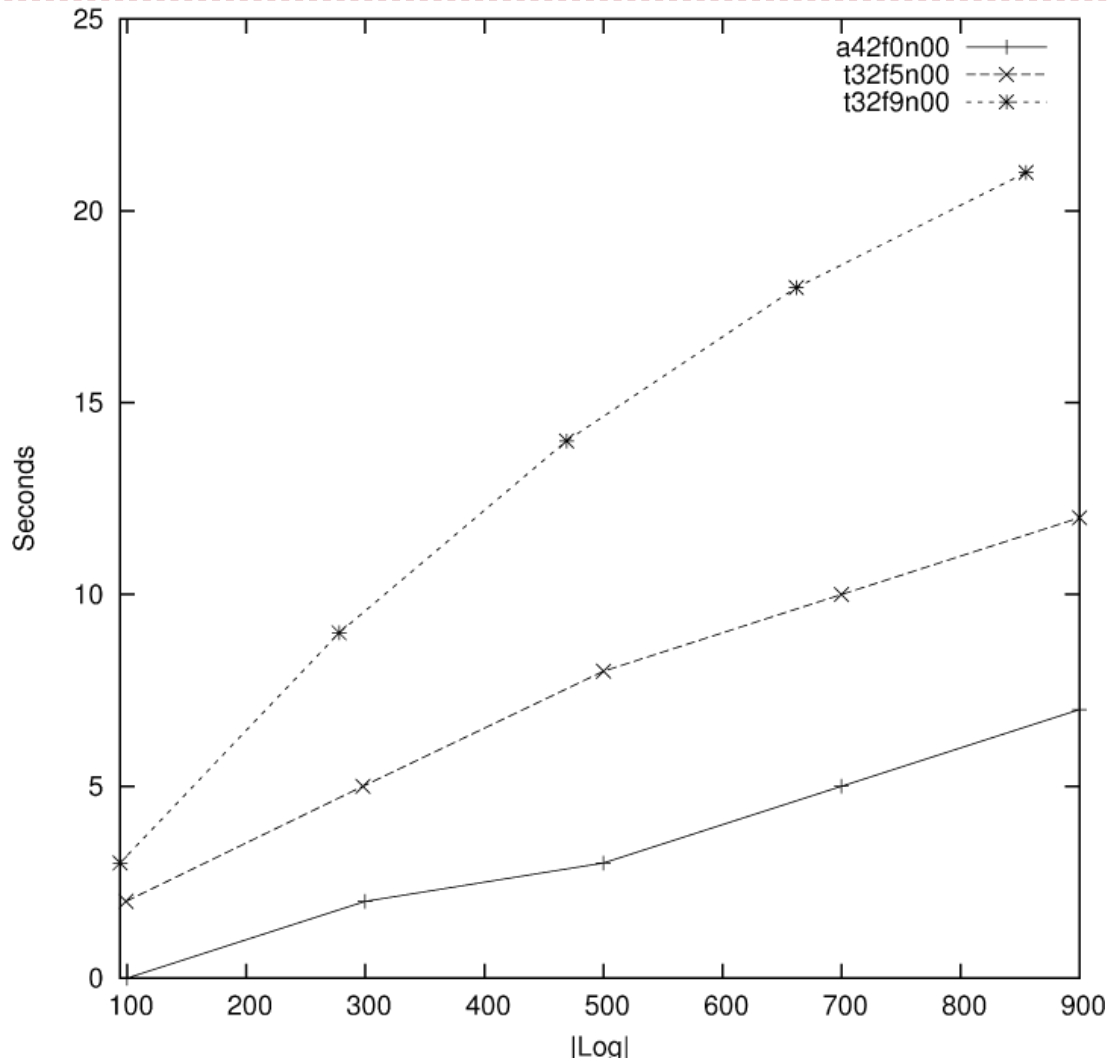
Results

Benchmark	a'_B	etc_P	MDT	Benchmark	a'_B	etc_P	MDT
GFA6NTC	1.00	1.00	0	GF12lOpt	1.00	0.85	7
GFA7	1.00	1.00	0	GFAL2	0.86	0.90	391
GFA8	1.00	1.00	0	GFDrivers	0.78	0.89	2
GFA12	1.00	1.00	0	GFBN3	0.71	0.88	181
GFChoice	1.00	1.00	0	GFBN2	0.59	0.96	19
GFBN1	1.00	1.00	0	GFA5	0.50	0.57	35
GFParallel5	1.00	0.99	11	GF12l	0.47	0.75	11
GFAL1	1.00	0.88	251	GF12lSkip	0.30	0.74	10

		M_T	Parikh					RBMiner				
Benchmark	TS	etc_P	P	T	etc_P	MDT	CPU	P	T	etc_P	MDT	CPU
a22f0n00_1	1309	0.06	19	22	0.63	1490	0(0)	19	22	0.63	1490	0(0)
a22f0n00_5	9867	0.07	19	22	0.73	9654	0(3)	19	22	0.73	9654	0(4)
a32f0n00_1	2011	0.04	31	32	0.52	2945	0(0)	32	32	0.52	2944	0(1)
a32f0n00_5	16921	0.05	31	32	0.59	22750	2(10)	31	32	0.59	22750	2(11)
a42f0n00_1	2865	0.03	44	42	0.35	7761	0(2)	52	42	0.37	7228	0(2)
a42f0n00_5	24366	0.04	44	42	0.42	60042	5(28)	46	42	0.42	60040	6(29)
t32f0n00_1	7717	0.03	30	33	0.37	15064	1(15)	31	33	0.37	15062	1(12)
t32f0n00_5	64829	0.04	30	33	0.39	125429	9(154)	30	33	0.39	125429	8(160)

Results (2)

CPU vs. Log size

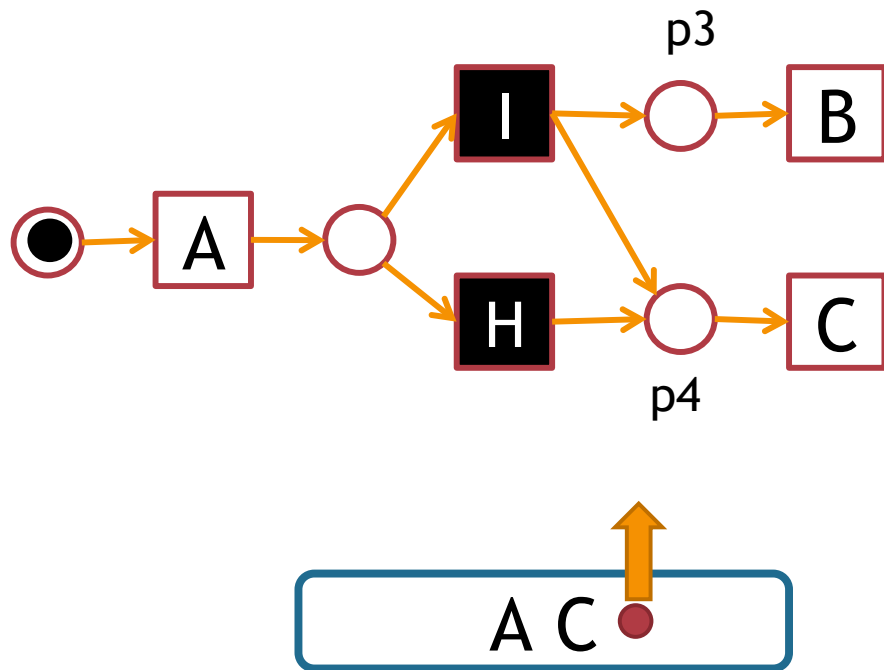


Outline

- ▶ Process Mining and Process Conformance
- ▶ Motivation
- ▶ Approach
 - ▶ General Approach
 - ▶ Implementation
 - ▶ Results
 - ▶ ***Extensions***
 - ▶ Invisible Tasks done
 - ▶ Duplicate Tasks done
 - ▶ States as Markings in progress
 - ▶ Non fitting in progress
- ▶ Future work
- ▶ Conclusions

Invisible Tasks

(Transitions associated with no event)



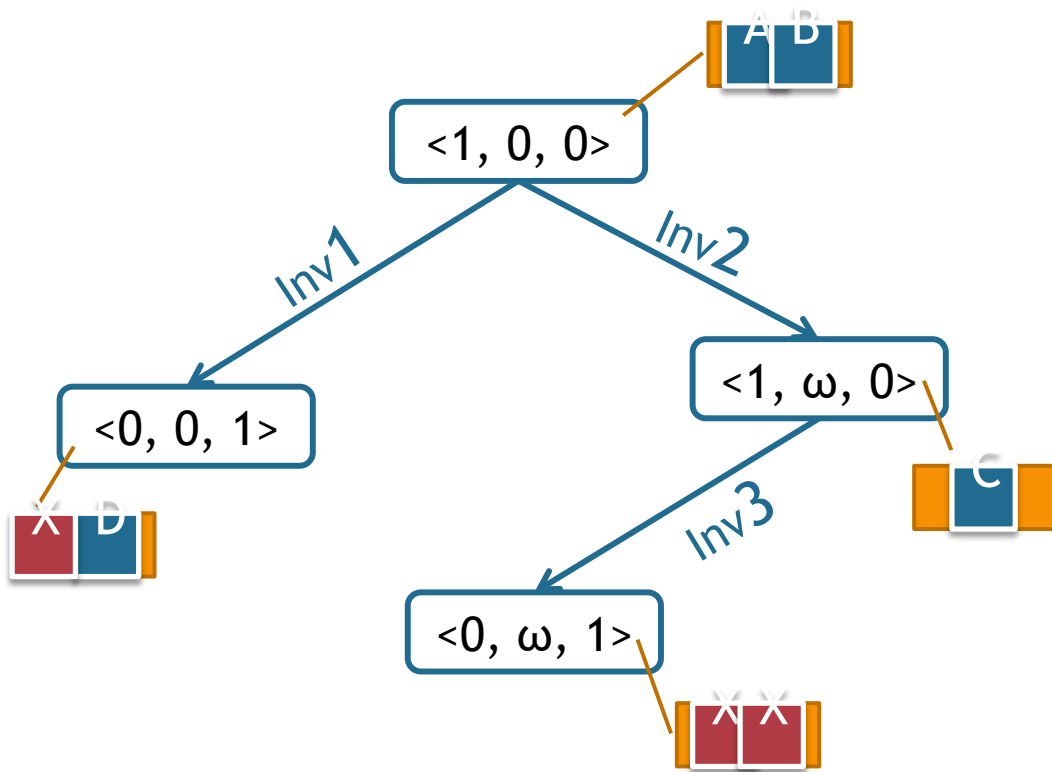
▶ Which Sequence?

▶ A H C ?

▶ A I C ?

▶ INDETERMINISM

Invisible Tasks (2)



- ▶ Invisible Coverability Graph

- ▶ Solutions

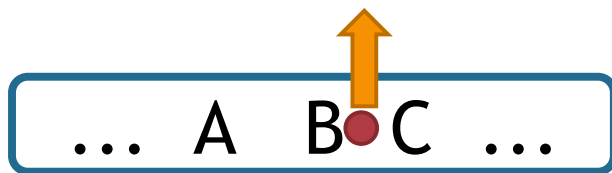
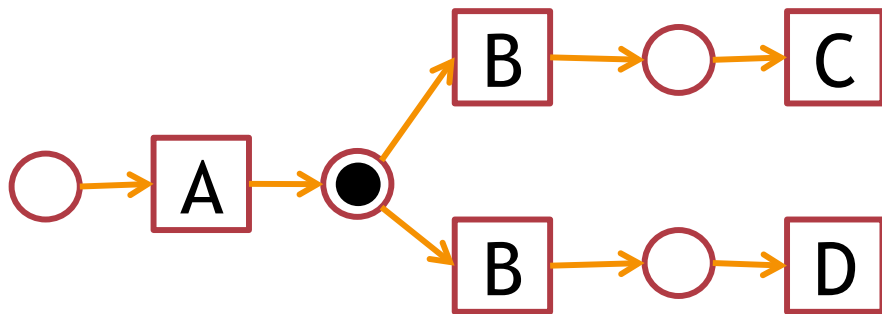
- ▶ Union of Enabled
- ▶ Lazy Invisibles *
- ▶ One path only
- ▶ Shortest Invisible Path *

*Rozinat et al.

Information System 33 (2008)

Duplicate Tasks

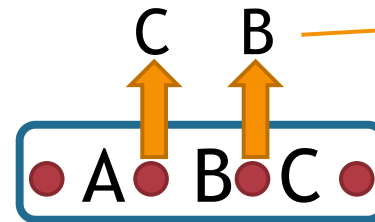
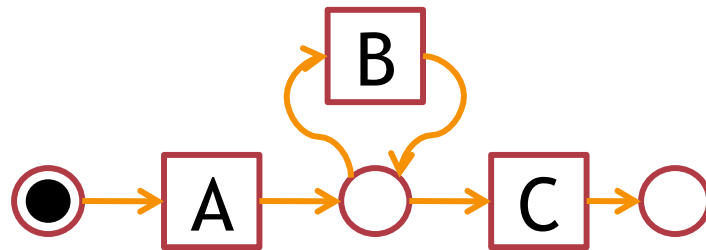
(Several Transitions associated with the same event)



- ▶ Which Task?
 - ▶ B ?
 - ▶ B ?
- ▶ INDETERMINISM
- ▶ Solutions
 - ▶ e.g. Look-ahead

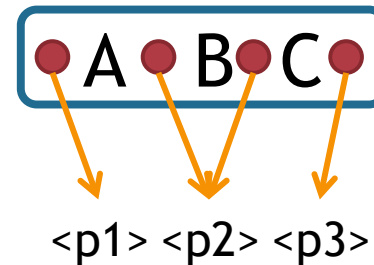
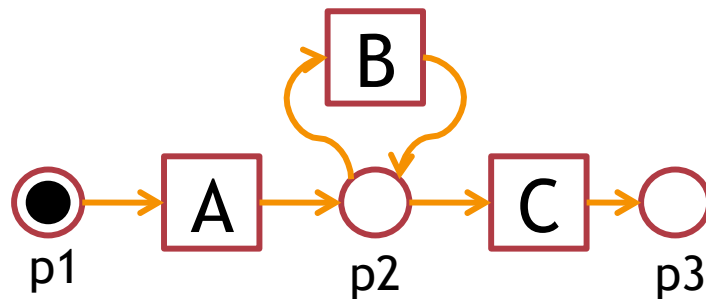
Variant: States as Markings

▶ States as Prefix



2
Escaping
Edges

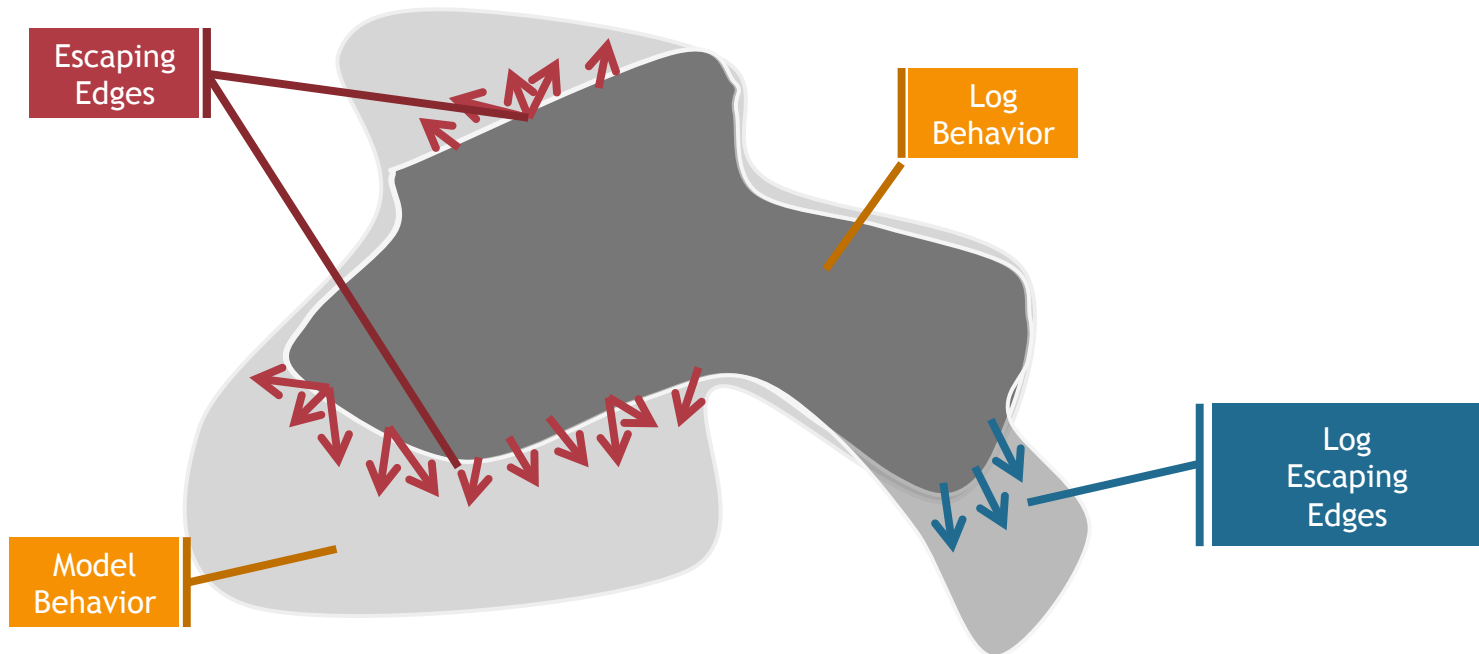
▶ States as Markings



NO
Escaping
Edges

Variant: Non fitting models

- ▶ Symmetric to the *Escaping Edges (Ee)*
- ▶ **Log Escaping Edges (LEe)**: The points where the log deviates from the model

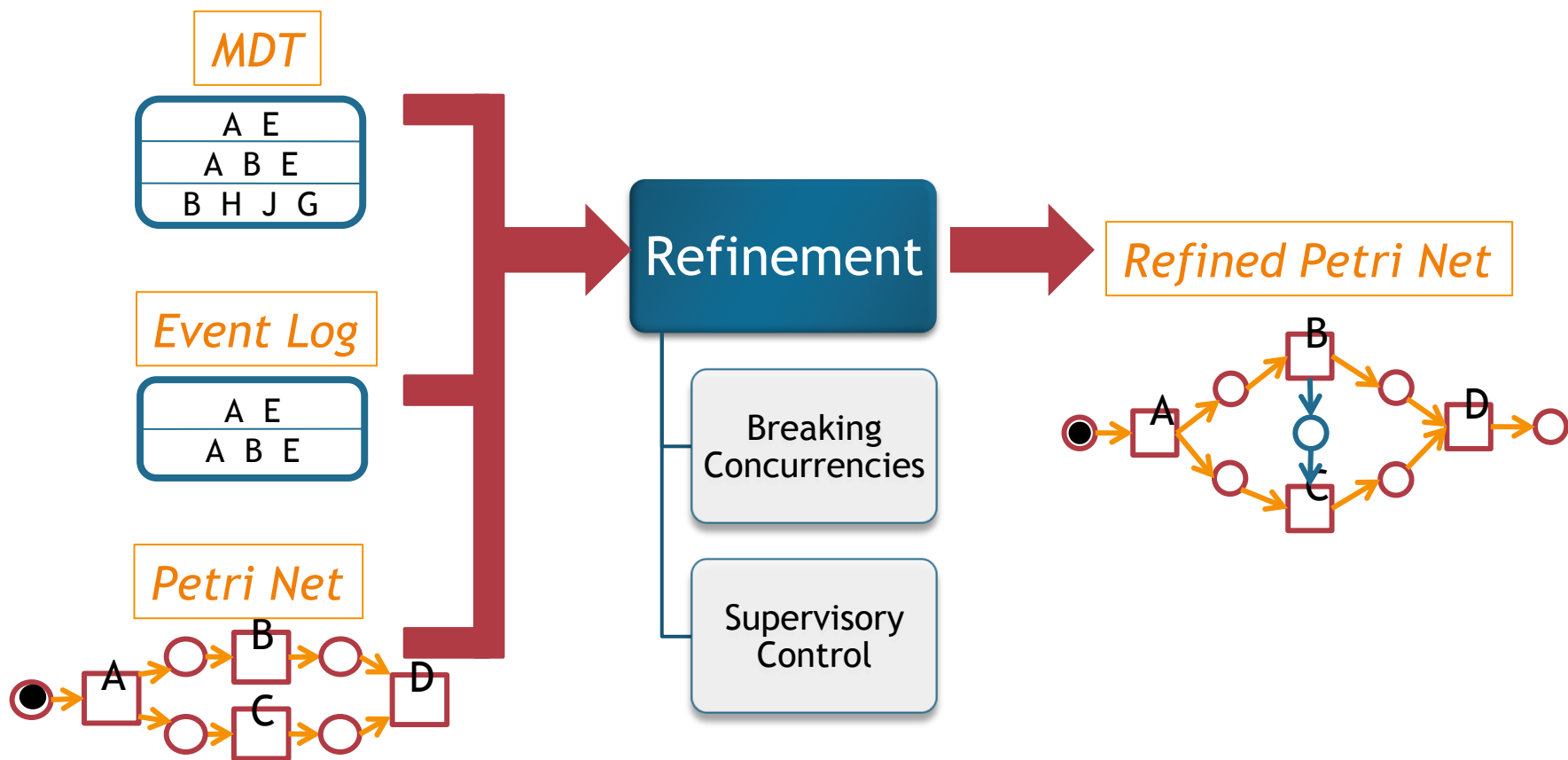


- ▶ Fitness instead of Precision

Outline

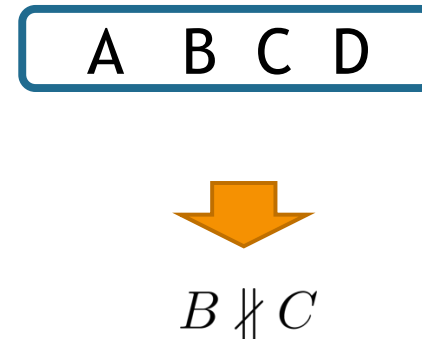
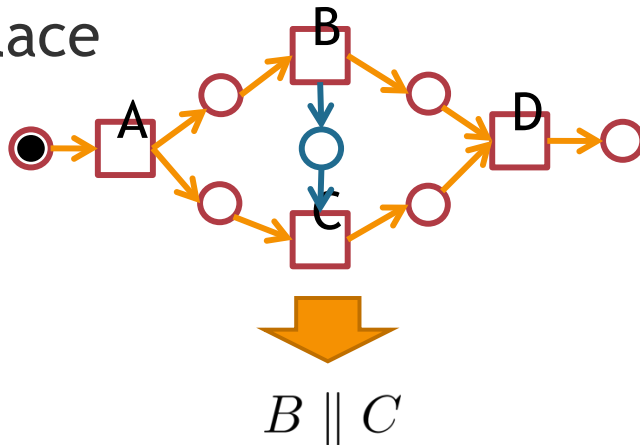
- ▶ Process Mining and Process Conformance
- ▶ Motivation
- ▶ Approach
 - ▶ General Approach
 - ▶ Implementation
 - ▶ Results
 - ▶ Extensions
- ▶ ***Future work***
- ▶ Conclusions

Future Work: Refinement



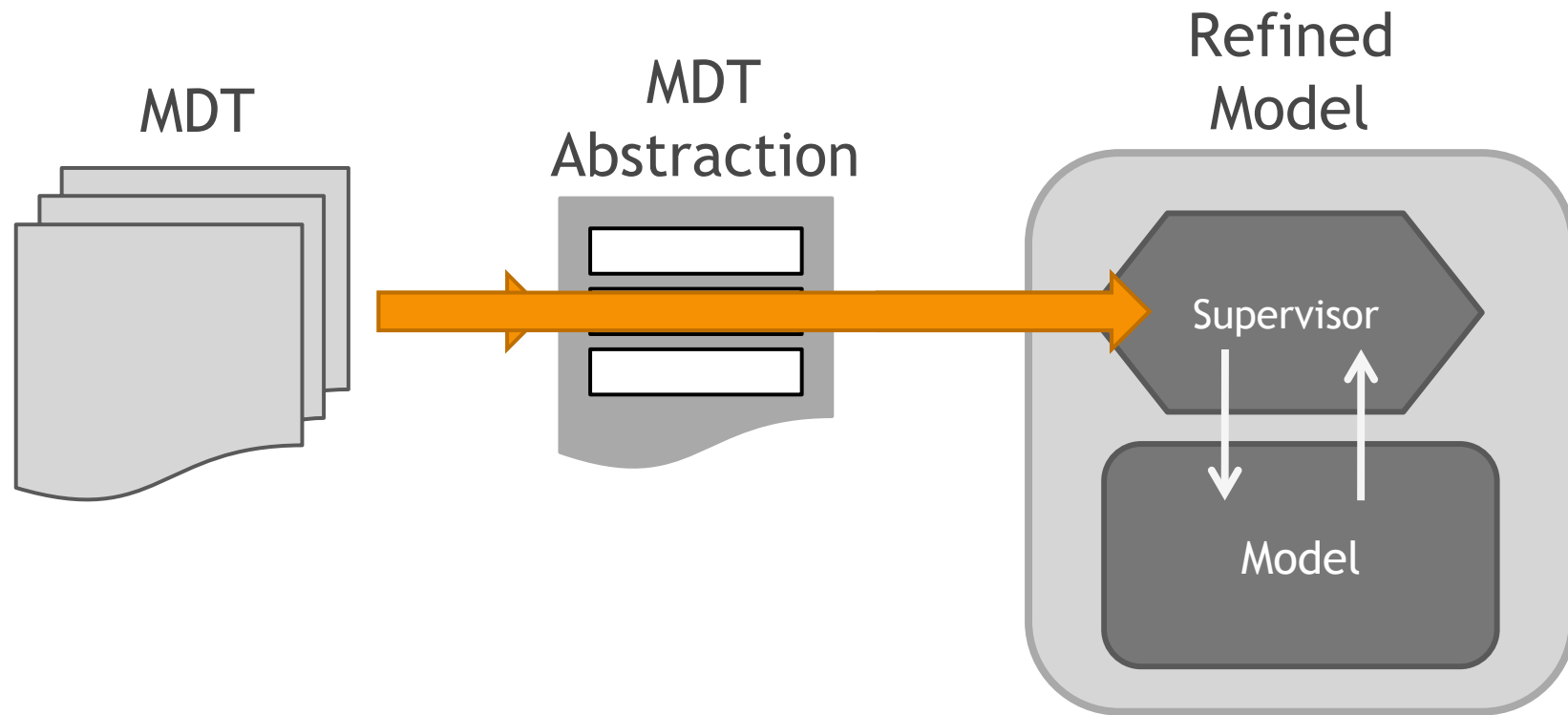
Future Work: Breaking Concurrency

- ▶ Concurrency in the model but not in the log
 - ▶ Break the model concurrency with a restriction, e.g. a place



- ▶ **Structural Concurrency**
 - ▶ Best effort overapproximation for general Petri Nets
 - ▶ Exact for live and bounded Free Choice systems
 - ▶ Polynomial Algorithm
 - ▶ Kovalyov and Esparza , *Proc. Intl. Workshop on Discrete Event Systems*, 1996

Supervisory Control



- ▶ Supervisory Control in Process Mining
 - ▶ *Santos et al. Supervisory Control Service (2010)*

Conclusions

- ▶ New technique for precision between Petri nets and Log.
- ▶ Avoids complete models state space exploration.
- ▶ Metric based on the effort needed to obtain a precise model.
- ▶ MDT, indicating the points where the model starts to deviates from the log.
- ▶ Approach implemented as plug-in of ProM 6.

Thank You

Thank You for Your Attention