CS 4072 – Topics in CS Process Mining

Lecture # 08

March 08, 2022

Spring 2022

FAST – NUCES, CFD Campus

Dr. Rabia Maqsood

rabia.maqsood@nu.edu.pk

Today's Topics

- Alpha algorithm
  - Activities ordering relation
  - Footprint matrix

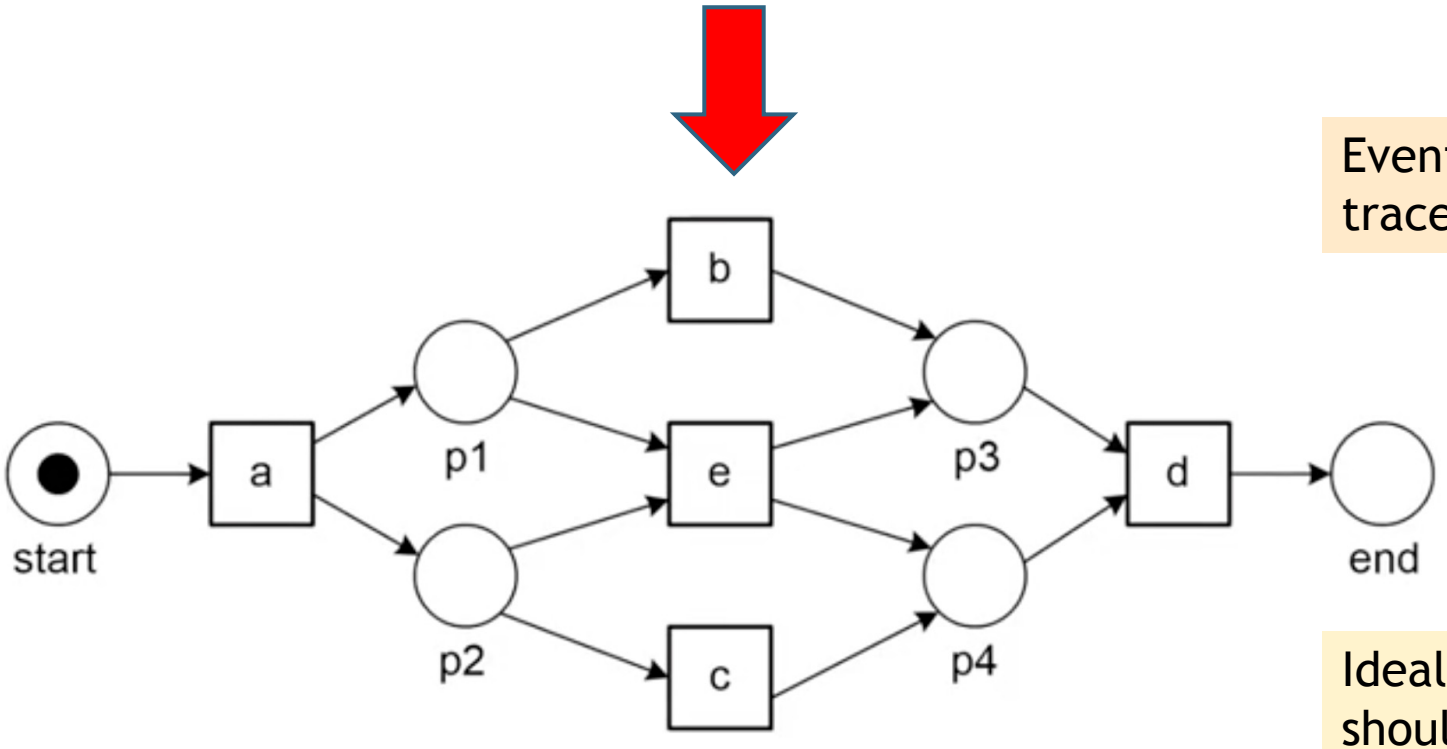
Event log

$$L_1 = [\langle a, b, c, d \rangle^3, \langle a, c, b, d \rangle^2, \langle a, e, d \rangle]$$

- An **event log** is a multiset of traces (same trace may appear multiple times).
- A **trace** is a sequence of activity names ordered by a timestamp.

Goal of Alpha algorithm

$L_1 = [\langle a, b, c, d \rangle^3, \langle a, c, b, d \rangle^2, \langle a, e, d \rangle]$

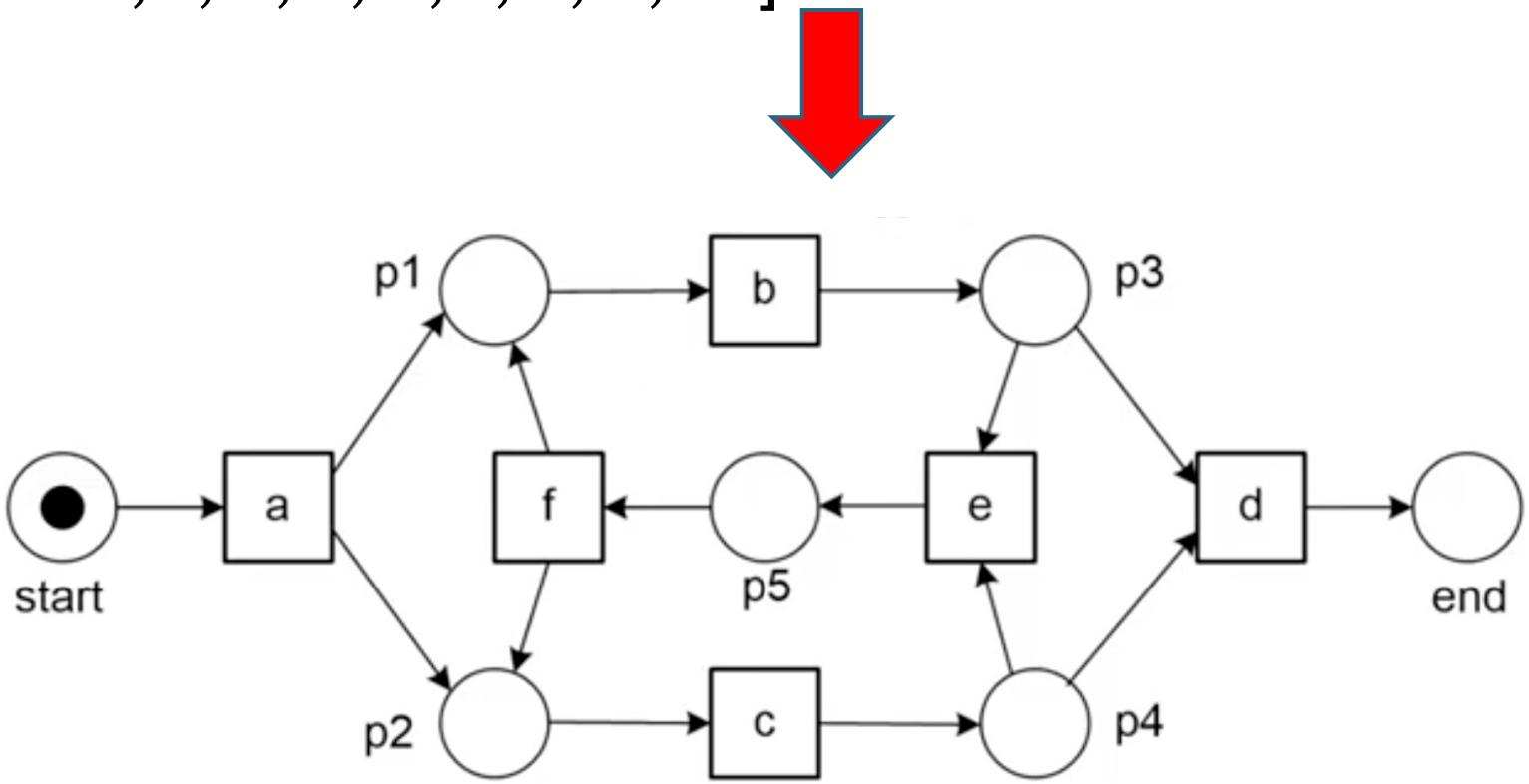


Event log contains all possible traces of model and vice versa.

Ideally, the discovered Petri net should be a sound WF-net.

Another Example

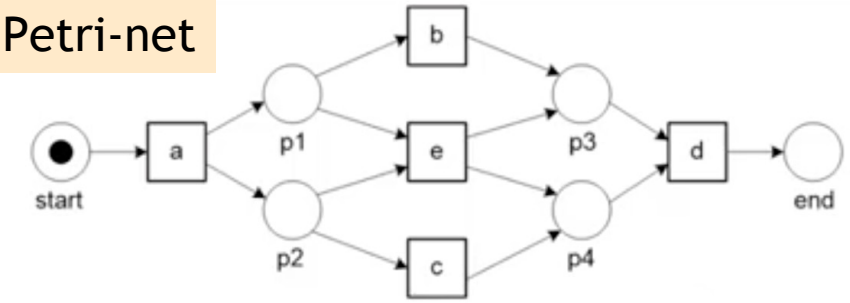
L_2 = [<a, b, c, d>^{3}, <a, c, b, d>^{4}, <a, b, c, e, f, b, c, d>^{2}, <a, b, c, e, f, c, b, d>, <a, c, b, e, f, b, c, d>^{2}, <a, c, b, e, f, b, c, e, f, c, b, d>]



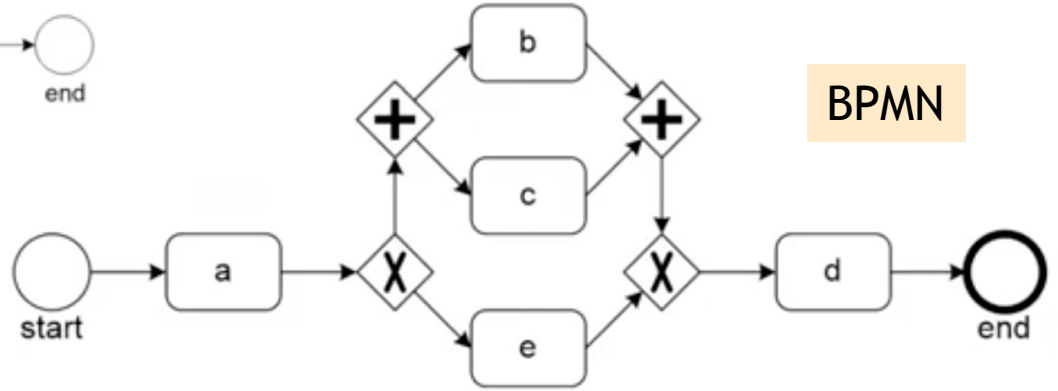
Generalization: event log contains only subset of all possible traces of model.

Notation is less relevant

Petri-net



BPMN



$$L_1 = [\langle a,b,c,d\rangle^3, \langle a,c,b,d\rangle^2, \langle a,e,d\rangle]$$

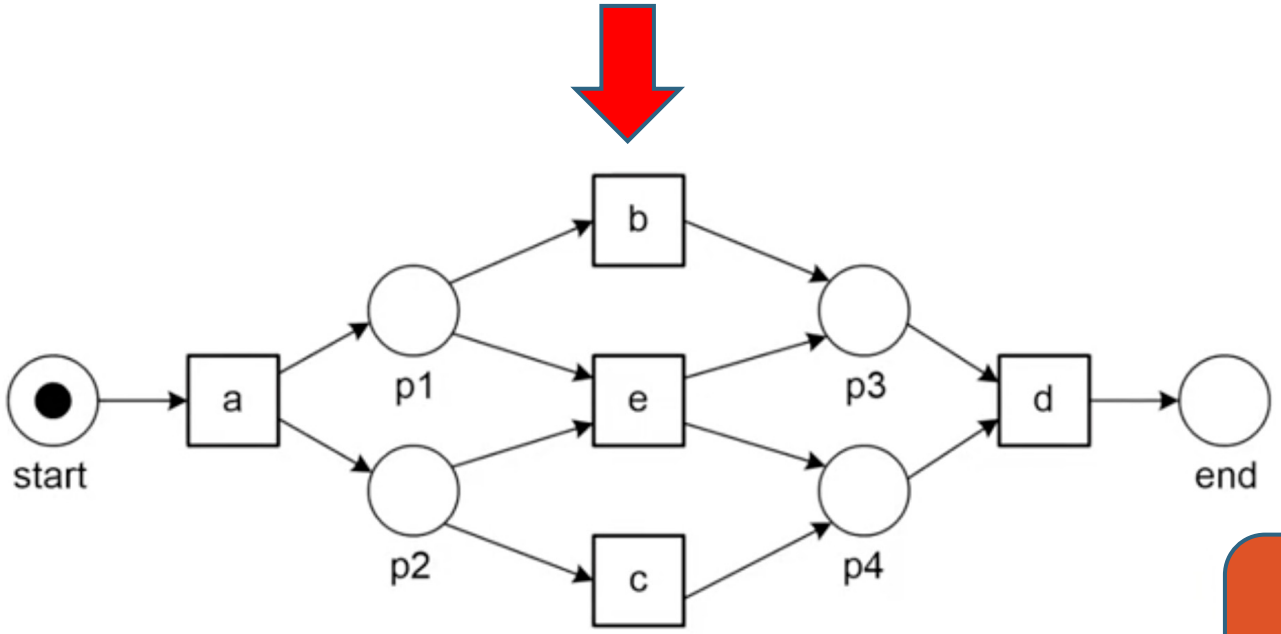
Process Discovery

Definition 6.2 (Specific process discovery problem) A *process discovery algorithm* is a function γ that maps a log $L \in \mathbb{B}(\mathcal{A}^*)$ onto a marked Petri net $\gamma(L) = (N,M)$.
Ideally, N is a *sound WF-net* and all traces in L correspond to possible firing sequences of (N,M).

Note that it is not necessary that an event log contains all firing sequences of (N,M).

Challenge for Process Discovery

$L_1 = [\langle a, b, c, d \rangle^3, \langle a, c, b, d \rangle^2, \langle a, e, d \rangle]$



Fitness
(ability to explain observed behavior)

Precision
(avoiding underfitting)

Generalization
(avoiding overfitting)

Simplicity
(Occam's razor principle)

Alpha Algorithm

Process discovery from event log

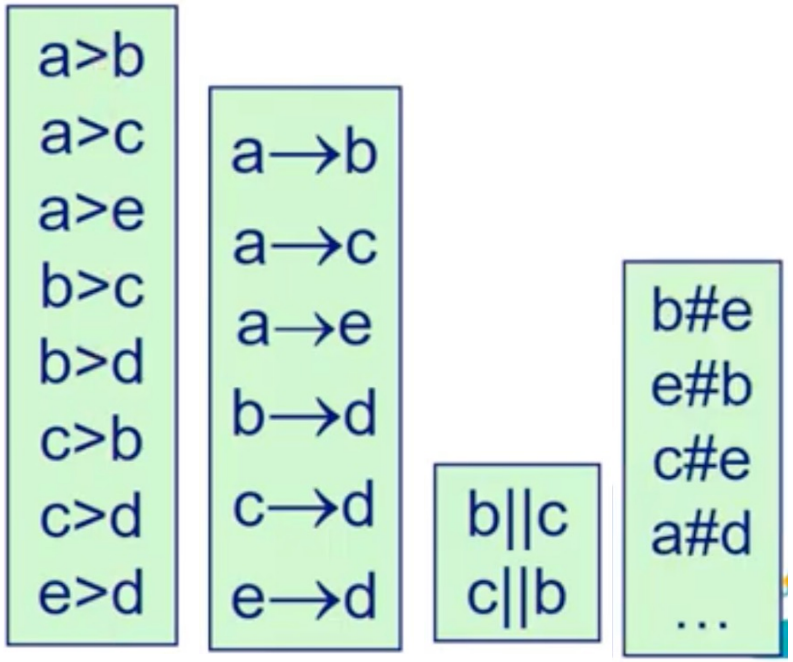
Log-based ordering relations (>, →, ||, #)

- Let L be an event log over $\mathcal{A}$, i.e., $L \in \mathbb{B}(\mathcal{A}^*)$. Let a, b $\in \mathcal{A}$
  - **$a >_L b$** if and only if there is a trace $\sigma = \langle t_1, t_2, t_3, \ldots, t_n \rangle$ and $i \in \{1, \ldots, n-1\}$ such that $\sigma \in L$ and $t_i = a$ and $t_{i+1} = b$
  - **$a \rightarrow_L b$** if and only if $a >_L b$ and $b \not>_L a$
  - **$a \#_L b$** if and only if $a \not>_L b$ and $b \not>_L a$
  - **$a \|_L b$** if and only if $a >_L b$ and $b >_L a$

Log-based ordering relations: (>, →, ||, #)

$L_1 = [\langle a, b, c, d\rangle^3, \langle a, c, b, d\rangle^2, \langle a, e, d\rangle]$

- Direct succession: **x > y**
  - Iff for some case **x is directly followed by y**
- Causality: **x → y**
  - Iff **x > y and not y > x**
- Parallel: **x || y**
  - Iff **x > y and y > x**
- Choice: **x # y**
  - Iff **not x > y and not y > x**



abcd
acbd
aed

Footprint of a log

- A matrix capturing relations between all the activities

Table 6.1 Footprint of L_1: $a \#_{L_1} a$, $a \rightarrow_{L_1} b$, $a \rightarrow_{L_1} c$, etc.

	a	b	c	d	e
a	$\#_{L_1}$	$\rightarrow_{L_1}$	$\rightarrow_{L_1}$	$\#_{L_1}$	$\rightarrow_{L_1}$
b	$\leftarrow_{L_1}$	$\#_{L_1}$	$\|_{L_1}$	$\rightarrow_{L_1}$	$\#_{L_1}$
c	$\leftarrow_{L_1}$	$\|_{L_1}$	$\#_{L_1}$	$\rightarrow_{L_1}$	$\#_{L_1}$
d	$\#_{L_1}$	$\leftarrow_{L_1}$	$\leftarrow_{L_1}$	$\#_{L_1}$	$\leftarrow_{L_1}$
e	$\leftarrow_{L_1}$	$\#_{L_1}$	$\#_{L_1}$	$\rightarrow_{L_1}$	$\#_{L_1}$

Answer

- Create a footprint matrix for this log.

  L_2 = [<a, b, c, d>^{3}, <a, c, b, d>^{4}, <a, b, c, e, f, b, c, d>^{2}, <a, b, c, e, f, c, b, d>, <a, c, b, e, f, b, c, d>^{2}, <a, c, b, e, f, b, c, e, f, c, b, d>]

Table 6.2 Footprint of $L_2 = [\langle a, b, c, d\rangle^3, \langle a, c, b, d\rangle^4, \langle a, b, c, e, f, b, c, d\rangle^2, \langle a, b, c, e, f, c, b, d\rangle, \langle a, c, b, e, f, b, c, d\rangle^2, \langle a, c, b, e, f, b, c, e, f, c, b, d\rangle]$

	a	b	c	d	e	f
a	#	→	→	#	#	#
b	←	#	‖	→	→	←
c	←	‖	#	→	→	←
d	#	←	←	#	#	#
e	#	←	←	#	#	→
f	#	→	→	#	←	#

Answer

- Create a footprint matrix for this log.

  L_3 = ⟨a,b,c,d,e,f,b,d,c,e,g⟩, ⟨a,b,d,c,e,g⟩2, ⟨a,b,c,d,e,f,b,c,d,e,f,b,d,c,e,g⟩

Table 6.3 Footprint of L_3

	a	b	c	d	e	f	g
a	#	→	#	#	#	#	#
b	←	#	→	→	#	←	#
c	#	←	#	‖	→	#	#
d	#	←	‖	#	→	#	#
e	#	#	←	←	#	→	→
f	#	→	#	#	←	#	#
g	#	#	#	#	←	#	#

Reading Material

- Chapter 6: Aalst