

CS 4072 - Topics in CS Process Mining

Lecture # 24

May 25, 2022

Spring 2022

FAST - NUCES, CFD Campus

Dr. Rabia Maqsood

rabia.maqsood@nu.edu.pk

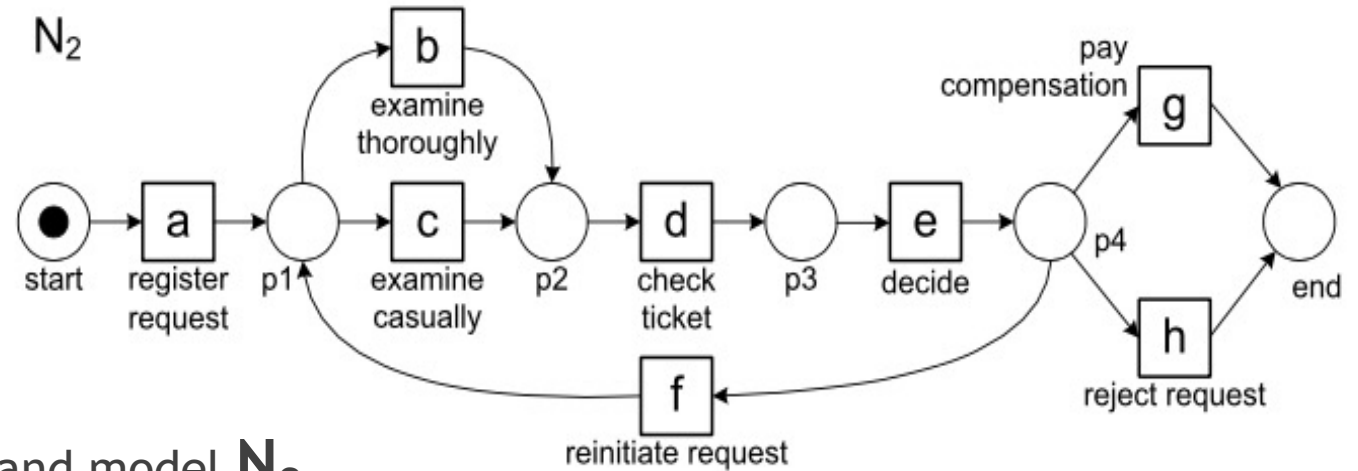
Today's Topics

- ▶ Conformance Checking
 - ▶ Sequence Alignment (continued)

Alignments

- ▶ **Alignments** were introduced to overcome the limitations of token-based replay.
- ▶ The objective is to find the *optimal* sequence alignment between two traces.

Alignments



- Consider a trace: $\sigma = \langle a, d, b, e, h \rangle$ and model N_2

- Followings are the possible alignments

Model move only

Log move only

$$\gamma_{2a} = \begin{array}{|c|c|c|c|c|c|} \hline a & \gg & d & b & e & h \\ \hline a & b & d & \gg & e & h \\ \hline \end{array}$$

$$\gamma_{2b} = \begin{array}{|c|c|c|c|c|c|} \hline a & \gg & d & b & e & h \\ \hline a & c & d & \gg & e & h \\ \hline \end{array}$$

$$\gamma_{2c} = \begin{array}{|c|c|c|c|c|c|} \hline a & d & b & \gg & e & h \\ \hline a & \gg & b & d & e & h \\ \hline \end{array}$$

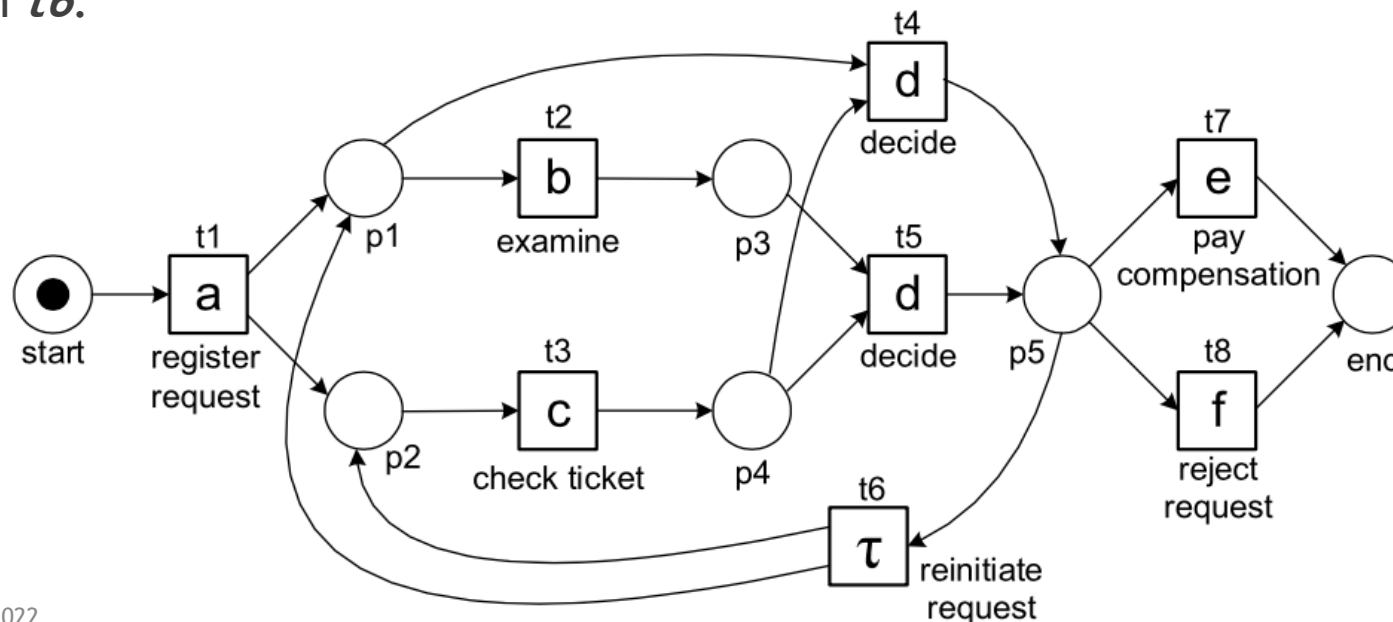
Matched or
Synchronous moves

Misalignment

Top row: projection on top row (excluding “no moves” or \gg) corresponds to trace in an event log
 Bottom row: projection on bottom row (excluding “no moves” or \gg) corresponds to a run of a model

Alignments

- ▶ Alignments can be defined *for any process notation*, including Petri nets having **duplicate** and **silent activities**.
- ▶ Consider the model N5 with duplicate transition d ($t4$ & $t5$) and a silent transition $t6$.

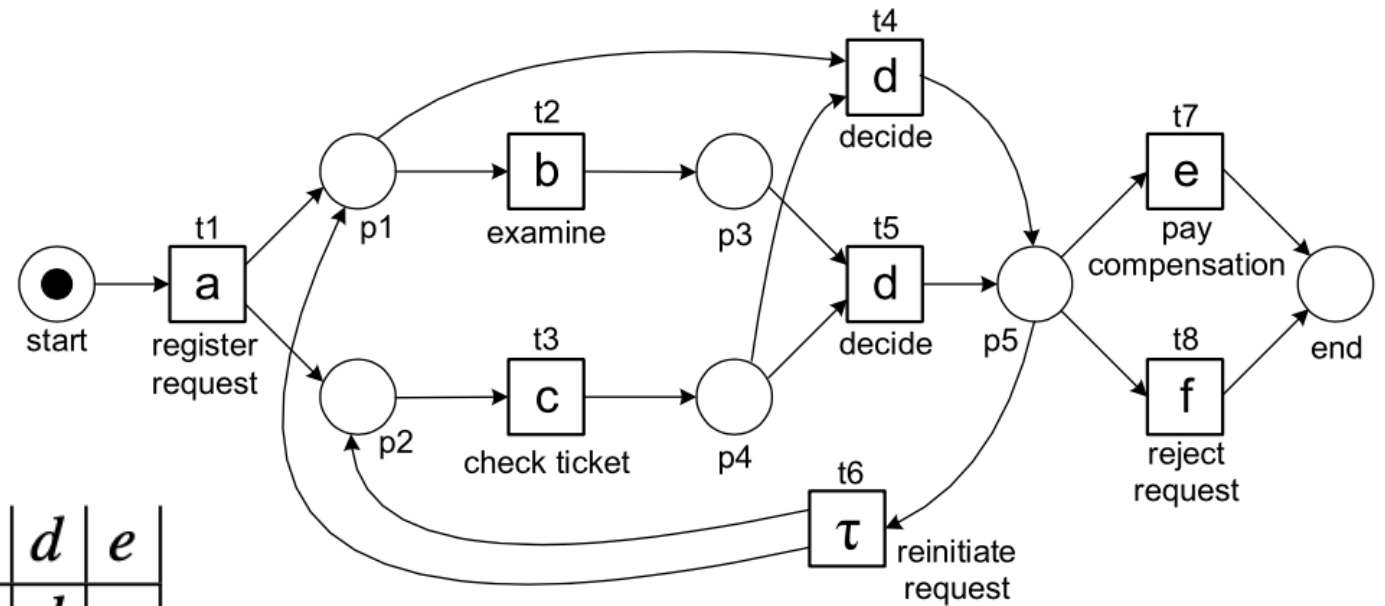


Alignments

- Consider $\sigma_1 = \langle a, c, d, e \rangle$ and N_5

- A possible alignment is: $\gamma_{5,1} =$

a	c	d	e
t_1	t_3	t_4	t_7



A move will be defined by a pair $(x, (y, t))$
 First element refers to the log
 Second element refers to the model

Examples:

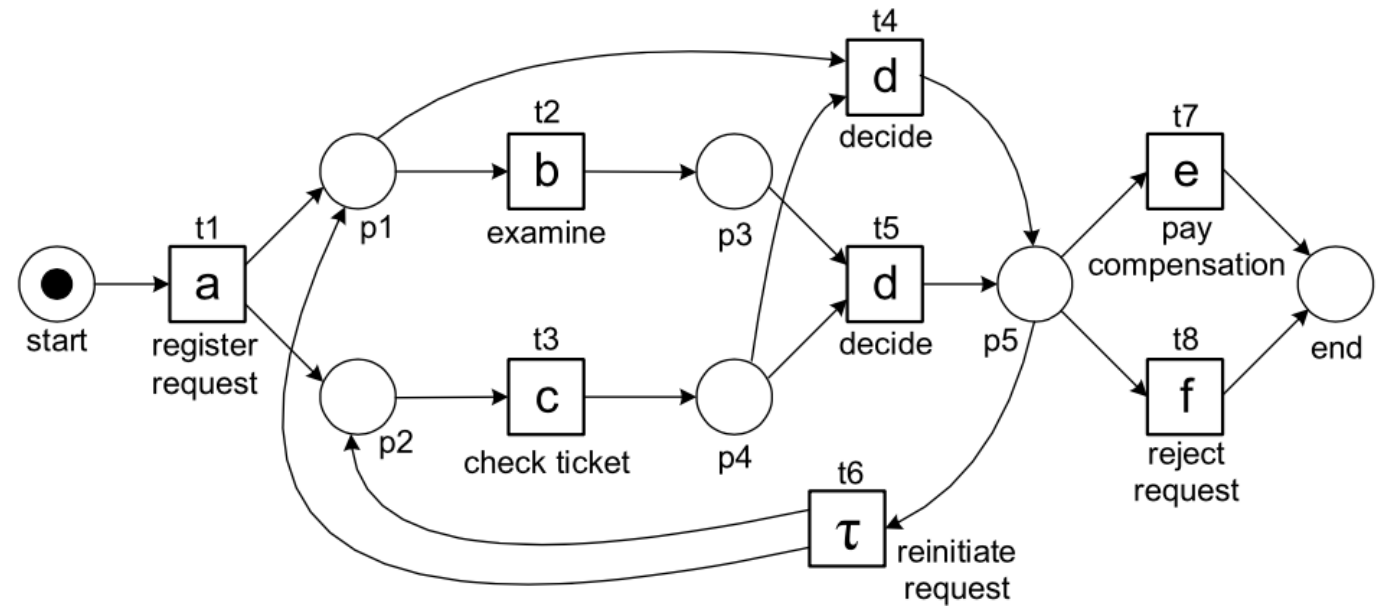
$(a, (a, t_1)) \rightarrow$ both log & model make “a” move using t_1

$(\gg, (c, t_3)) \rightarrow$ “c” using t_3 is not mimicked by log

$(f, \gg) \rightarrow$ log makes “f” move not followed by model

Alignments

- Consider $\sigma_2 = \langle a, b, d, f \rangle$ and N_5



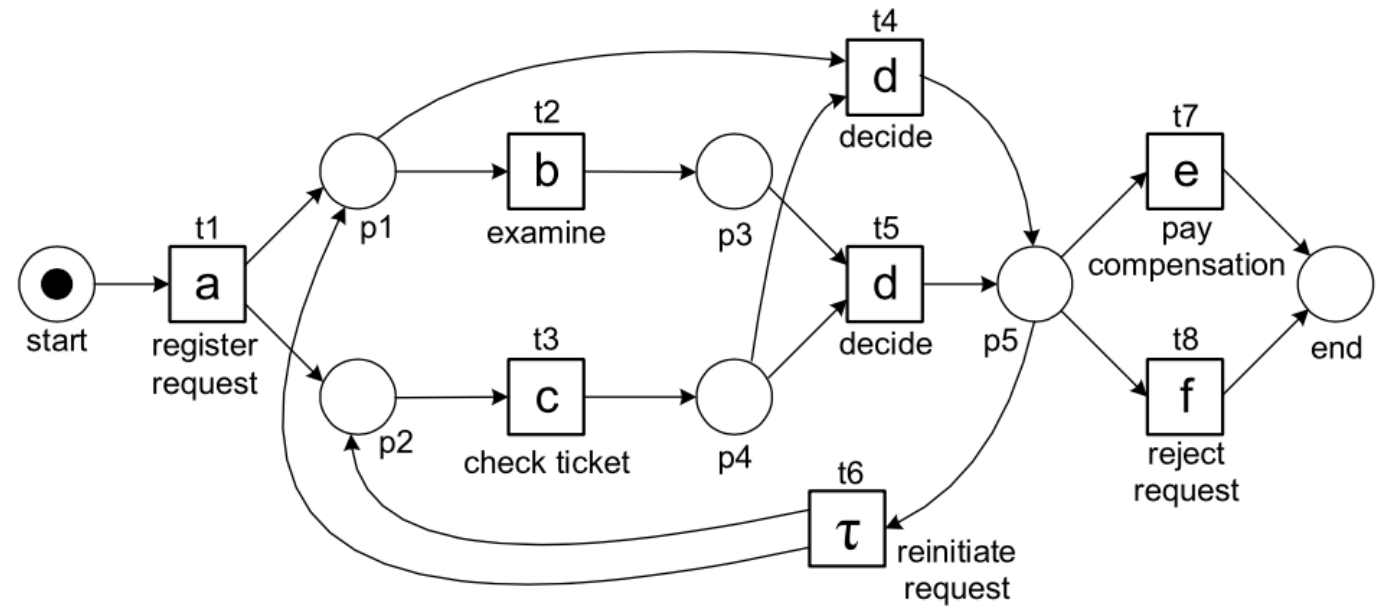
- Following are the possible alignments:

$$\gamma_{5,2a} = \begin{array}{c|c|c|c|c} a & b & \gg & d & f \\ \hline a & b & c & d & f \\ \hline t1 & t2 & t3 & t5 & t8 \end{array}$$

$$\gamma_{5,2b} = \begin{array}{c|c|c|c|c} a & \gg & b & d & f \\ \hline a & c & b & d & f \\ \hline t1 & t3 & t2 & t5 & t8 \end{array}$$

Alignments

- Consider $\sigma_3 = \langle a, c, d, e, f \rangle$ and N_5



- Following are the possible alignments:

$$\gamma_{5,3a} = \begin{array}{|c|c|c|c|} \hline a & c & d & e & f \\ \hline a & c & d & e & \gg \\ \hline t1 & t3 & t4 & t7 & \\ \hline \end{array}$$

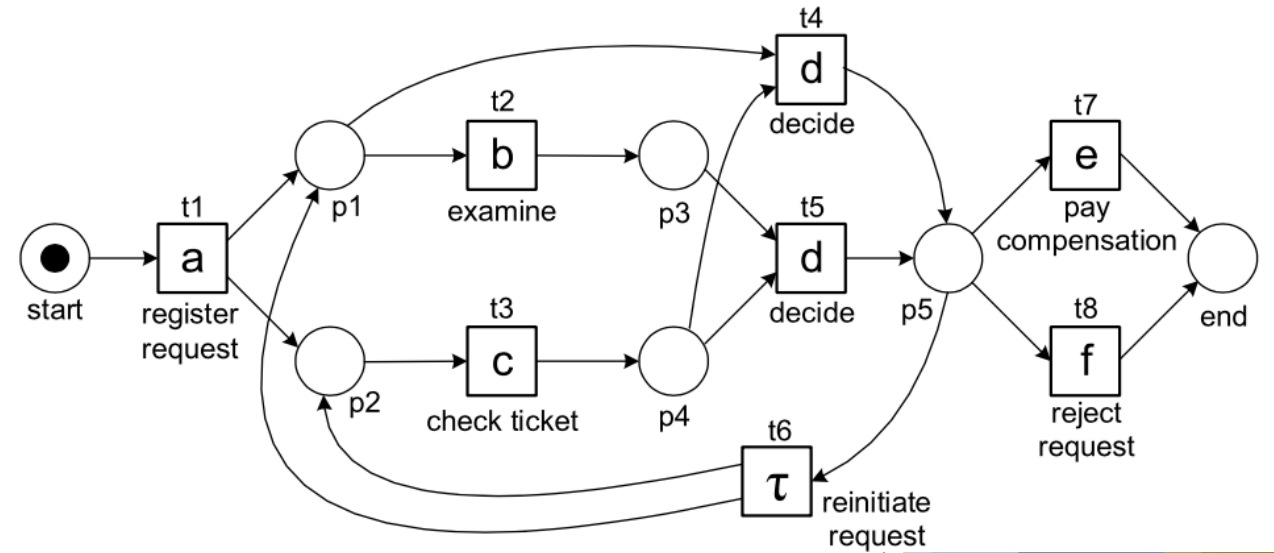
$$\gamma_{5,3b} = \begin{array}{|c|c|c|c|} \hline a & c & d & e & f \\ \hline a & c & d & \gg & f \\ \hline t1 & t3 & t4 & & t8 \\ \hline \end{array}$$

NOTE: silent transition leaves no trail in the event log

Alignments

- Consider

$\sigma_4 = \langle a, c, d, b, c, d, c, d, c, b, d, f \rangle$ and N_5



- Following are the possible alignments:

$$\gamma_{5,4} = \begin{array}{c|c|c|c|c|c|c|c|c|c|c|c|c|c|c} a & c & d & \gg & b & c & d & \gg & c & d & \gg & c & b & d & f \\ \hline a & c & d & \tau & b & c & d & \tau & c & d & \tau & c & b & d & f \\ \hline t1 & t3 & t4 & t6 & t2 & t3 & t5 & t6 & t3 & t4 & t6 & t3 & t2 & t5 & t8 \end{array}$$

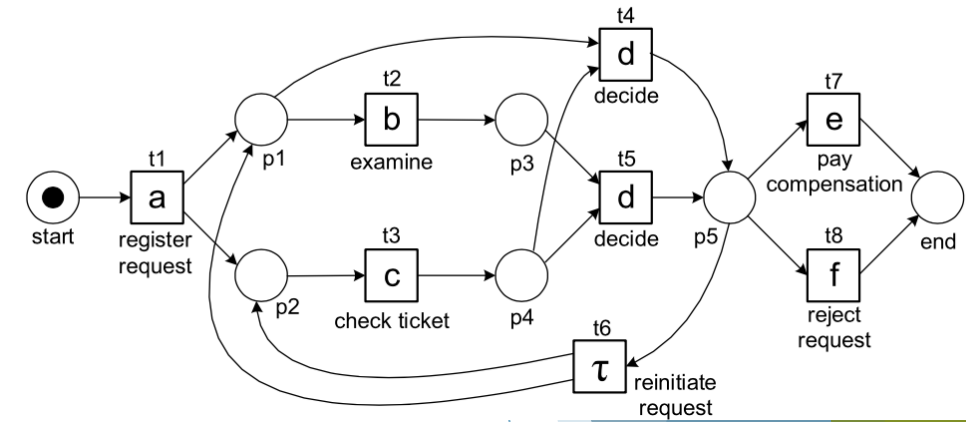
Alignments

- ▶ $(x, (y, t))$ is a *legal move* if one of the following four cases holds:
 - ▶ $x = y$ and y is the visible label of transition t (*synchronous move*)
 - ▶ $x = \gg$ and y is the visible label of transition t (*visible model move*)
 - ▶ $x = \gg$, $y = \tau$ and transition t is silent (*invisible model move*)
 - ▶ $x \neq \gg$ and $(y, t) = \gg$ (*log move*)
- ▶ Other moves such as (\gg, \gg) and $(x, (y, t))$ with $x \neq y$ are illegal moves.

Alignments

- ▶ Given a log trace and a process model, there may be many (if not infinitely many) alignments.
- ▶ The objective of sequence alignment is to find an *optimal* matching sequences.
- ▶ How to find the *optimal* alignment?
- ▶ What about the *worst* alignment?

Alignments



- For $\sigma_2 = \langle a, b, d, f \rangle$ and N_5 other possible alignments are:

$$\gamma_{5,2c} = \begin{array}{|c|c|c|c|c|c|c|c|c|} \hline a & b & d & f & \gg & \gg & \gg & \gg & \gg \\ \hline \gg & \gg & \gg & \gg & a & b & c & d & f \\ \hline & & & & t1 & t2 & t3 & t5 & t8 \\ \hline \end{array}$$

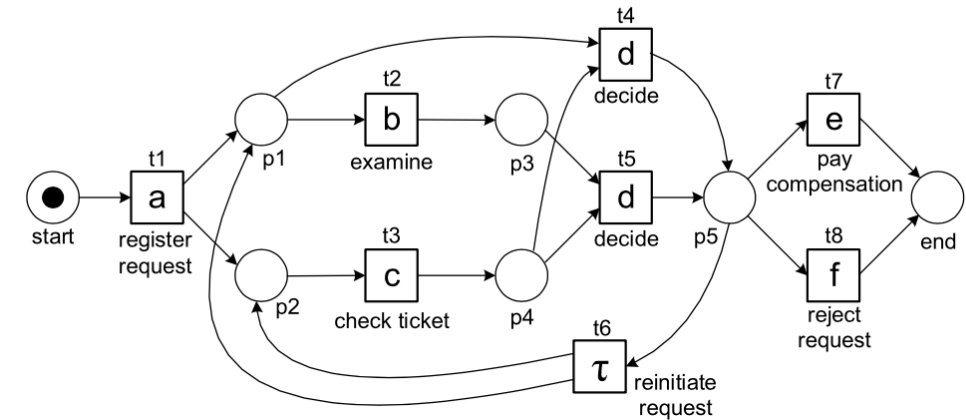
$$\gamma_{5,2d} = \begin{array}{|c|c|c|c|c|c|c|} \hline a & b & d & f & \gg & \gg & \gg \\ \hline a & \gg & \gg & \gg & c & d & e \\ \hline t1 & & & & t3 & t4 & t7 \\ \hline \end{array}$$

- Previously identified alignments were:

$$\gamma_{5,2a} = \begin{array}{|c|c|c|c|c|} \hline a & b & \gg & d & f \\ \hline a & b & c & d & f \\ \hline t1 & t2 & t3 & t5 & t8 \\ \hline \end{array}$$

$$\gamma_{5,2b} = \begin{array}{|c|c|c|c|c|} \hline a & \gg & b & d & f \\ \hline a & c & b & d & f \\ \hline t1 & t3 & t2 & t5 & t8 \\ \hline \end{array}$$

Alignments



- For $\sigma_2 = \langle a, b, d, f \rangle$ and N_5 other possible alignments are:

$$\gamma_{5,2c} = \begin{array}{|c|c|c|c|c|c|c|c|c|} \hline a & b & d & f & \gg & \gg & \gg & \gg & \gg \\ \hline \gg & \gg & \gg & \gg & a & b & c & d & f \\ \hline & & & & t1 & t2 & t3 & t5 & t8 \\ \hline \end{array}$$

9 misalignments

$$\gamma_{5,2d} = \begin{array}{|c|c|c|c|c|c|c|} \hline a & b & d & f & \gg & \gg & \gg \\ \hline a & \gg & \gg & \gg & c & d & e \\ \hline t1 & & & & t3 & t4 & t7 \\ \hline \end{array}$$

6 misalignments

- Previously identified alignments were:

There can be multiple optimal alignments with same **cost**.

$$\gamma_{5,2a} = \begin{array}{|c|c|c|c|c|} \hline a & b & \gg & d & f \\ \hline a & b & c & d & f \\ \hline t1 & t2 & t3 & t5 & t8 \\ \hline \end{array}$$

1 misalignment

$$\gamma_{5,2b} = \begin{array}{|c|c|c|c|c|} \hline a & \gg & b & d & f \\ \hline a & c & b & d & f \\ \hline t1 & t3 & t2 & t5 & t8 \\ \hline \end{array}$$

1 misalignment

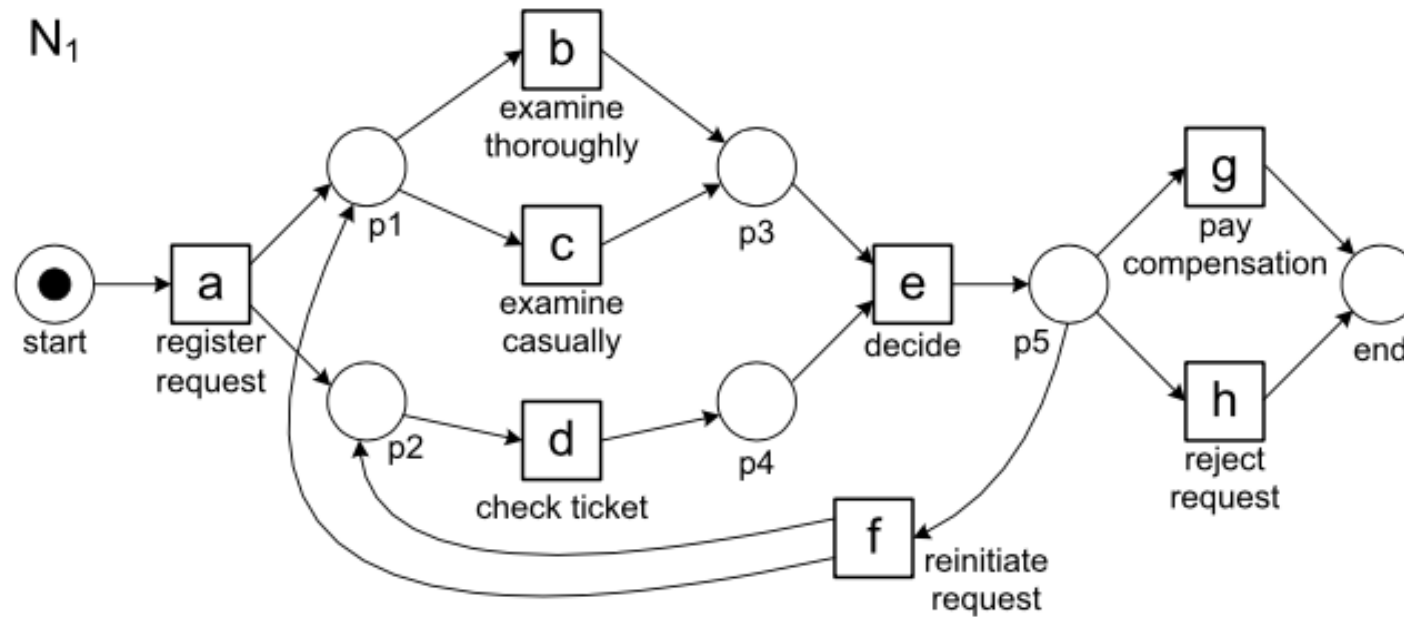
The count of misalignments can be taken as cost of aligning a log and a model.

Alignments

- ▶ To select the most appropriate alignment, we associate **costs** to undesirable moves and select an alignment with the **lowest total costs**.
- ▶ Generic cost function:
 - ▶ Cost function δ assigns costs to **legal moves**.
 - ▶ Moves where log and model agree have no costs, i.e., $\delta(x, (y, t)) = 0$ for *synchronous moves* (with $x = y$).
 - ▶ Moves in model only have no costs if the transition is invisible, i.e., $\delta(\gg, (\tau, t)) = 0$ for *invisible model moves*.
 - ▶ $\delta(\gg, (y, t)) > 0$ is the cost when the model makes a “y move” without a corresponding move of the log (*visible model move*).
 - ▶ $\delta(x, \gg) > 0$ is the cost for an “x move” in just the log (*log move*).

Question

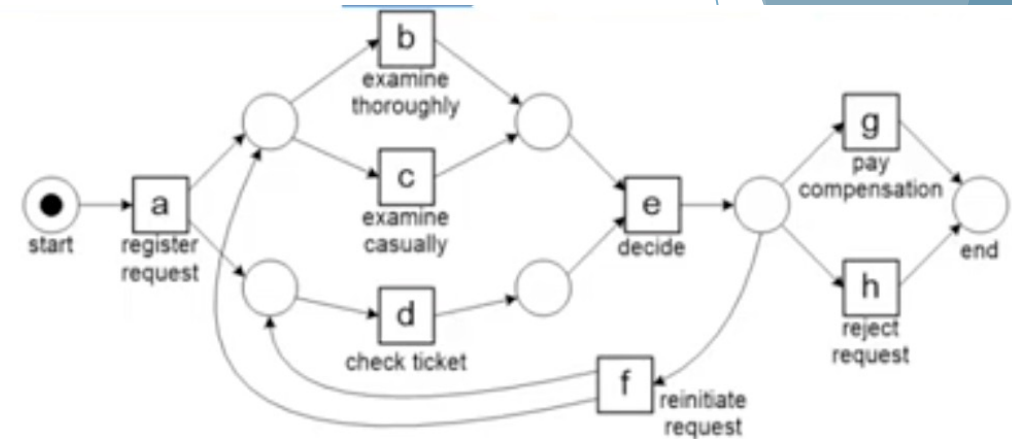
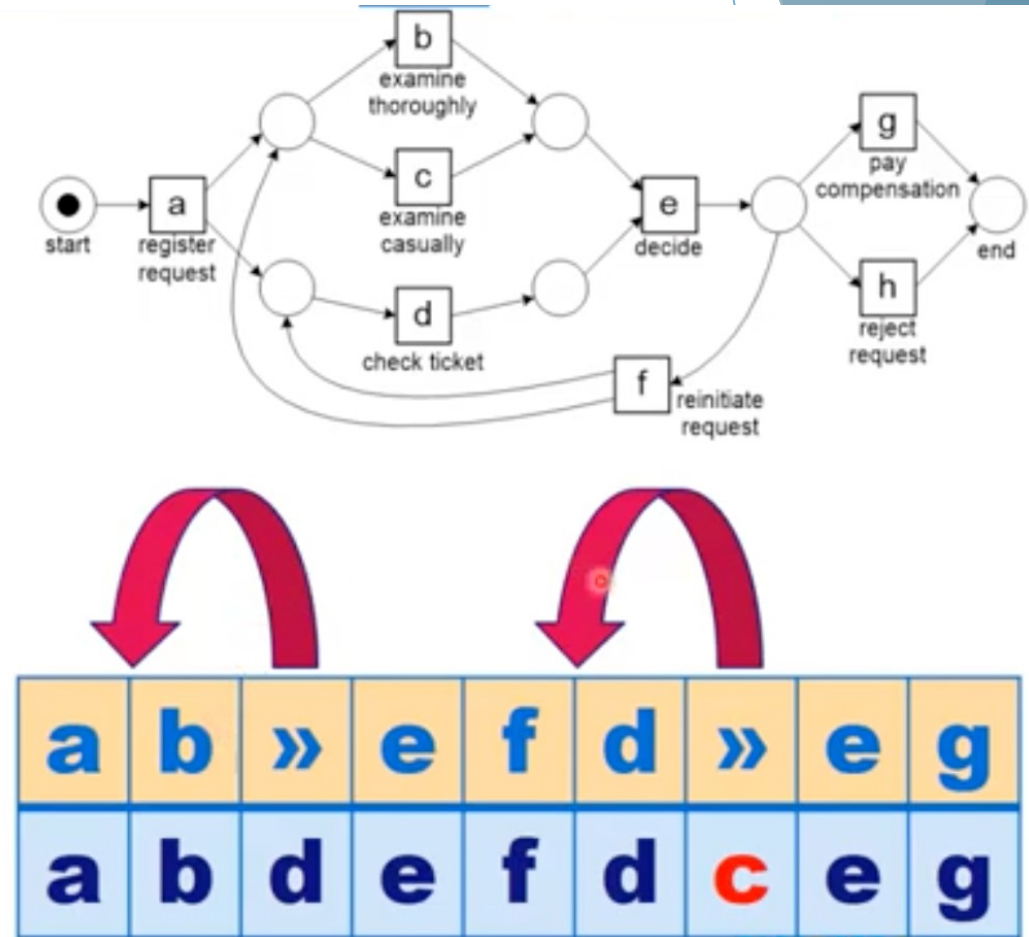
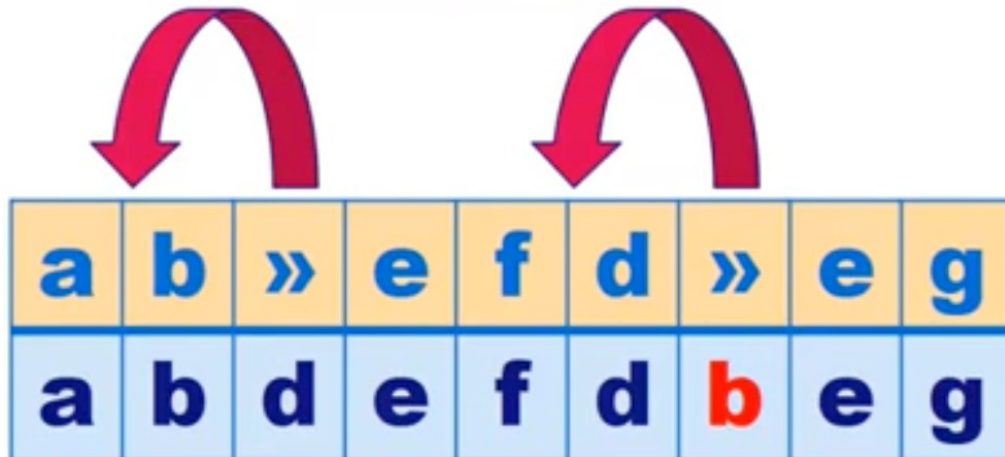
- ▶ Given N_1 , how many optimal alignments are there for $\langle a, b, e, f, d, e, g \rangle$?



Answer



move in model can be
reordered in concurrent part



Reading Material

- ▶ Chapter 8: Aalst