Selected slides from

Process Mining: Data Science in Action

Conformance Checking Using Token-Based Replay

Wil van der Aalst

Process
Mining

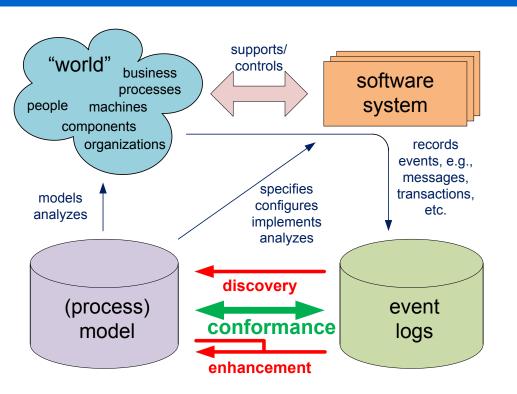
Data Science in Action
Second Edition

prof.dr.ir. Wil van der Aalst www.processmining.org



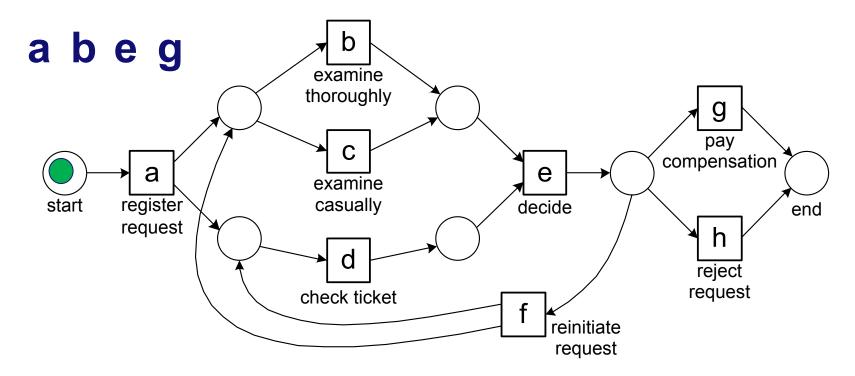
Where innovation starts

Conformance checking

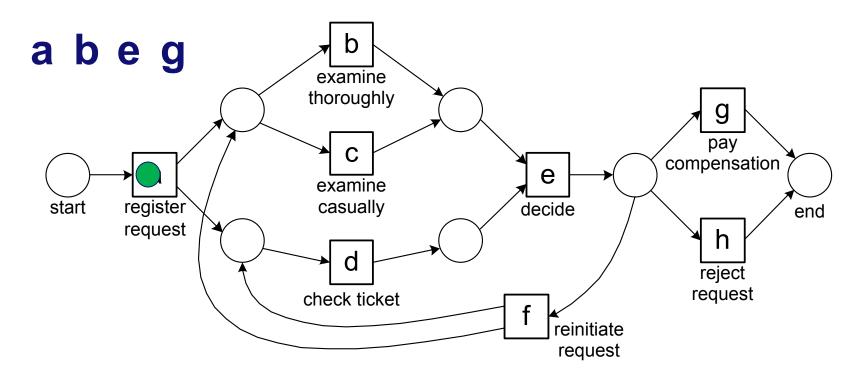


- 1. Conformance checking using causal footprints.
- 2. Conformance checking based on token-based replay.
- 3. Alignment-based conformance checking.

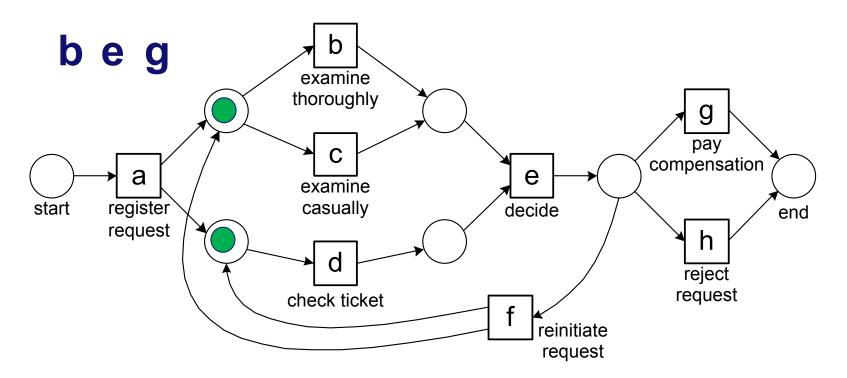




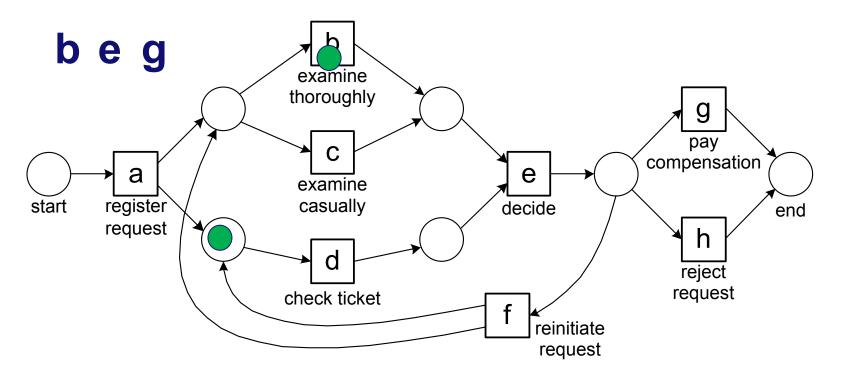




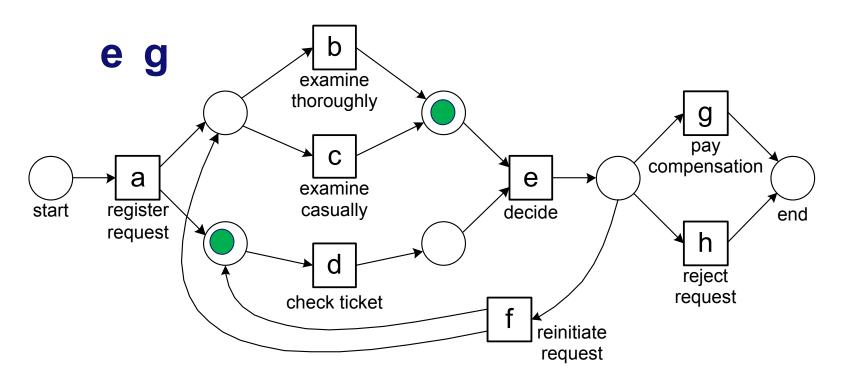




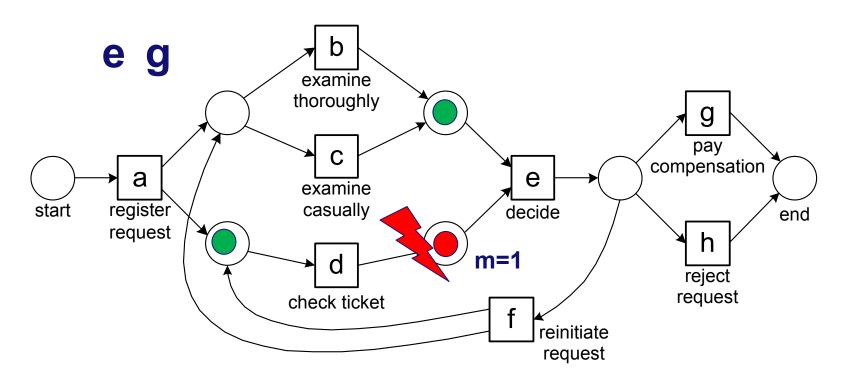




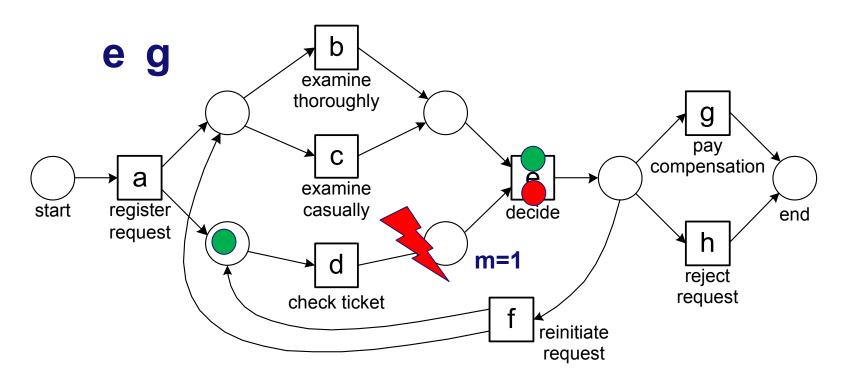




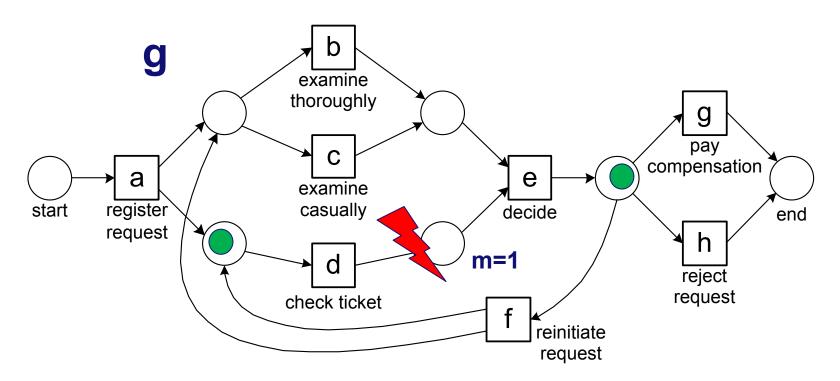




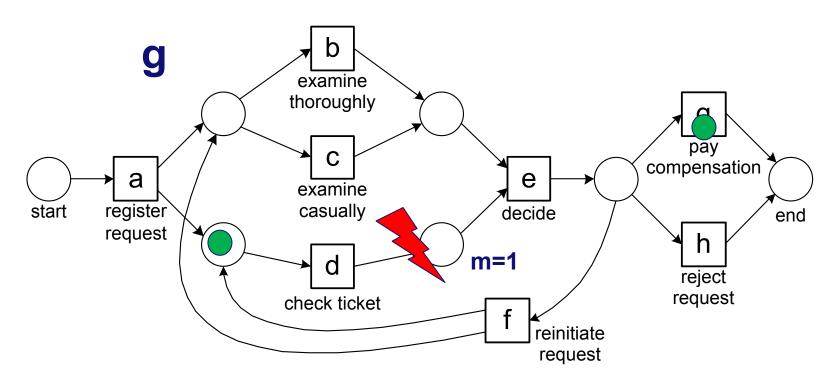




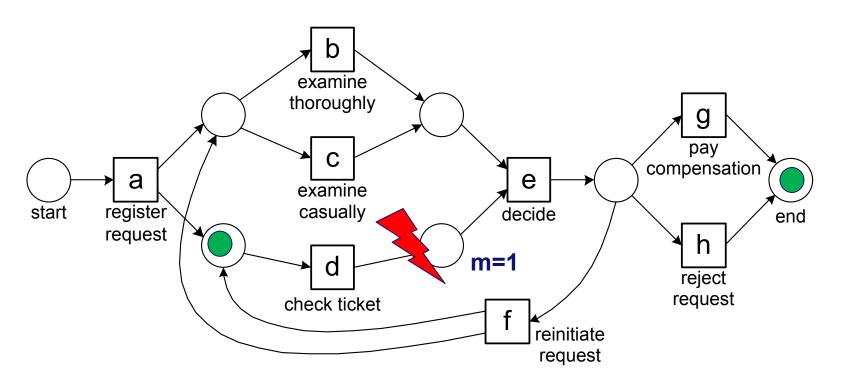




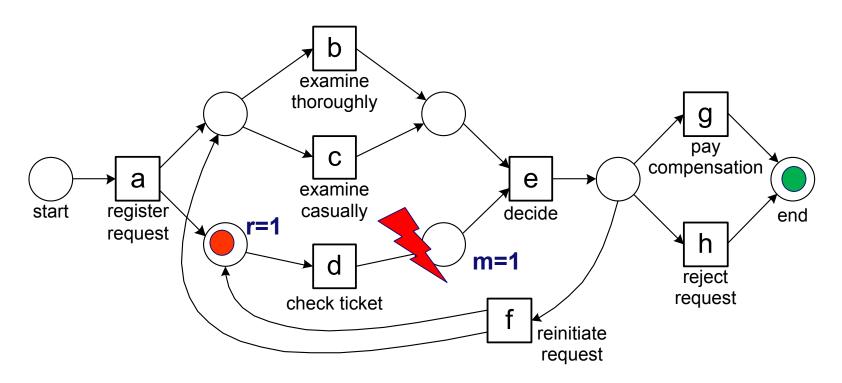


















Quantifying fitness at the trace level

$$fitness(\sigma, N) = \frac{1}{2} \left(1 - \frac{m}{c} \right) + \frac{1}{2} \left(1 - \frac{r}{p} \right)$$



Quantifying fitness at the trace level

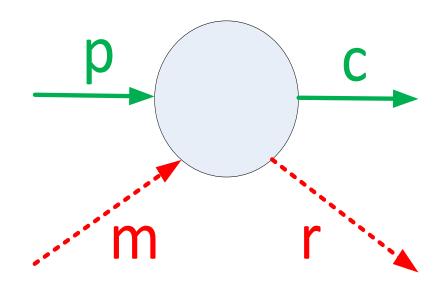
$$fitness(\sigma, N) = \frac{1}{2} \left(1 - \frac{1}{6} \right) + \frac{1}{2} \left(1 - \frac{1}{6} \right) = 0.83333$$



Approach (1/3)

Use four counters:

- p = produced tokens
- c = consumed tokens
- m = missing tokens (consumed while not there)
- r = remaining tokens
 (produced but not consumed)



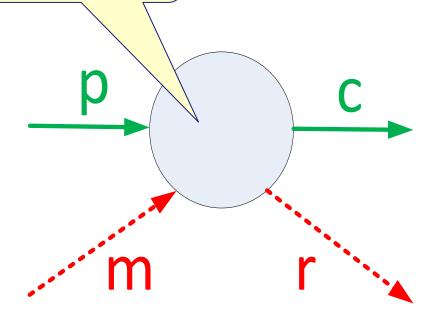


Approach (1/3)

while running p+m-c tokens

Use four counters:

- p = produced tokens
- c = consumed tokens
- m = missing tokens (consumed while not there)
- r = remaining tokens
 (produced but not consumed)



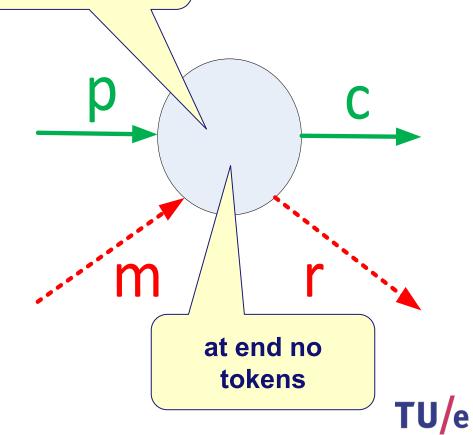


Approach (1/3)

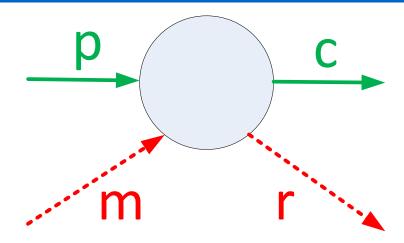
while running p+m-c tokens

Use four counters:

- p = produced tokens
- c = consumed tokens
- m = missing tokens (consumed while not there)
- r = remaining tokens
 (produced but not consumed)



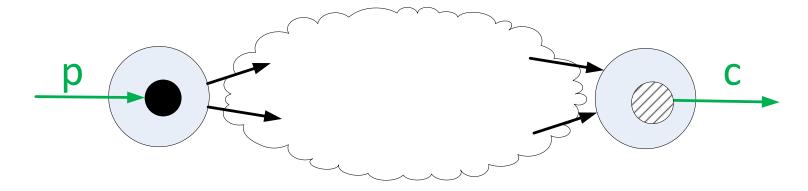
Approach (2/3)



- Invariants
 - -At any time: p+m ≥ c ≥ m (also per place)
 - -At the end: r = p + m c (also per place)



Approach (3/3)

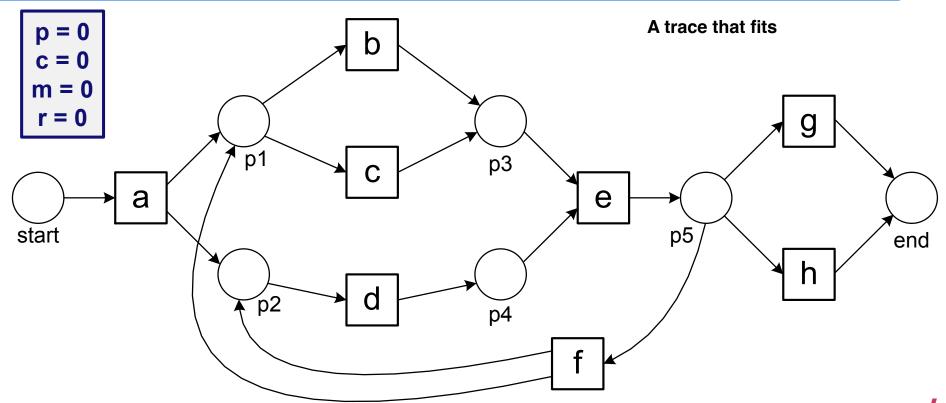


Initialization and finalization:

- In the beginning a token is produced for the source place: p = 1.
- At the end a token is consumed from the sink place (also if not there): c' = c + 1.

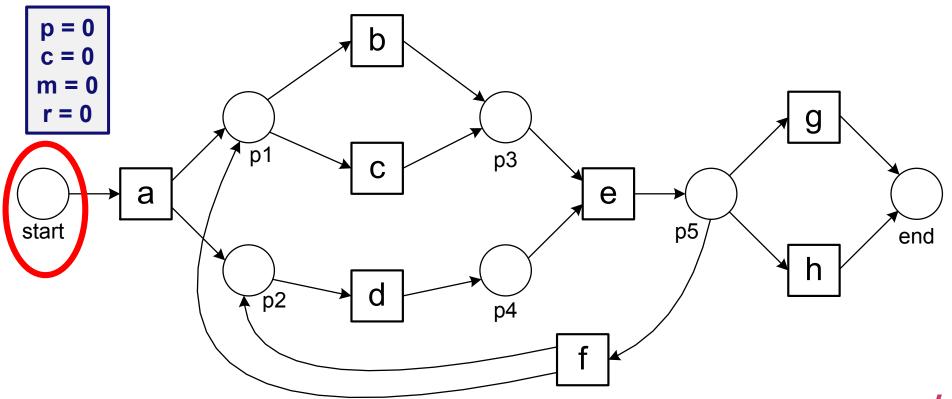


Replaying
$$\sigma_1 = \langle a, c, d, e, h \rangle$$



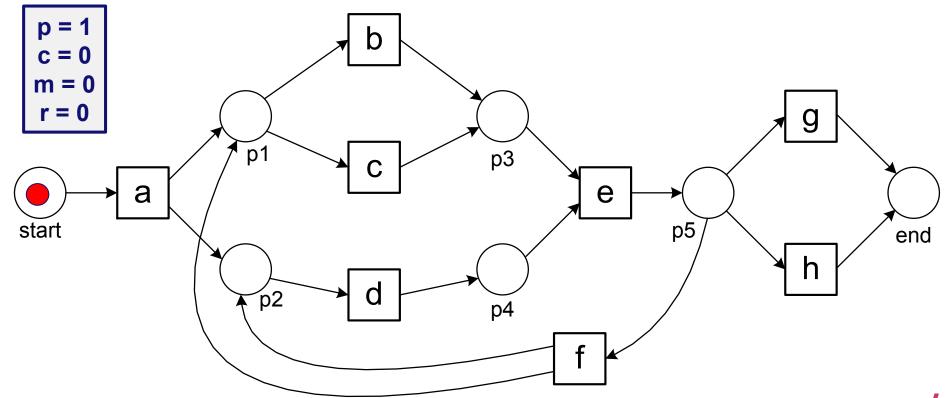


Replaying
$$\sigma_1 = \langle a, c, d, e, h \rangle$$





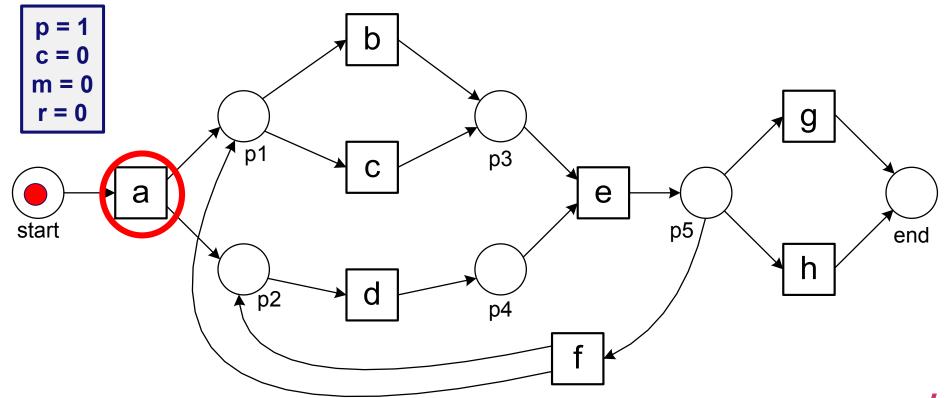
Replaying
$$\sigma_1 = \langle a, c, d, e, h \rangle$$





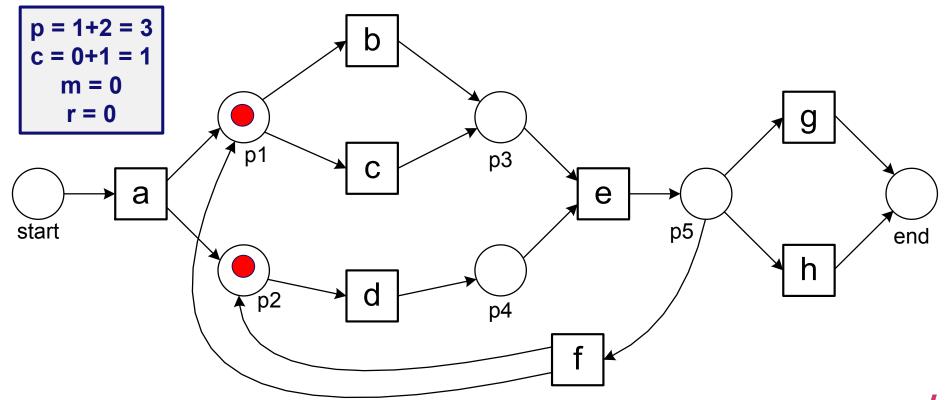
Replaying

$$\sigma_1 = \langle a | c, d, e, h \rangle$$



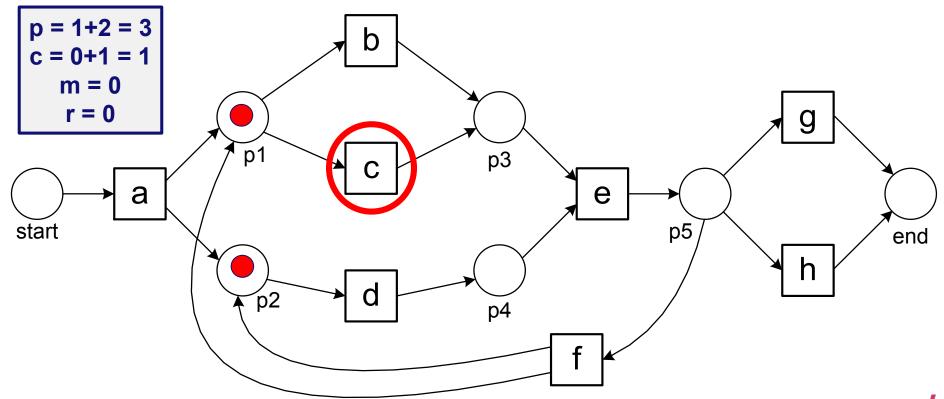


Replaying
$$\sigma_1 = \langle a, c, d, e, h \rangle$$



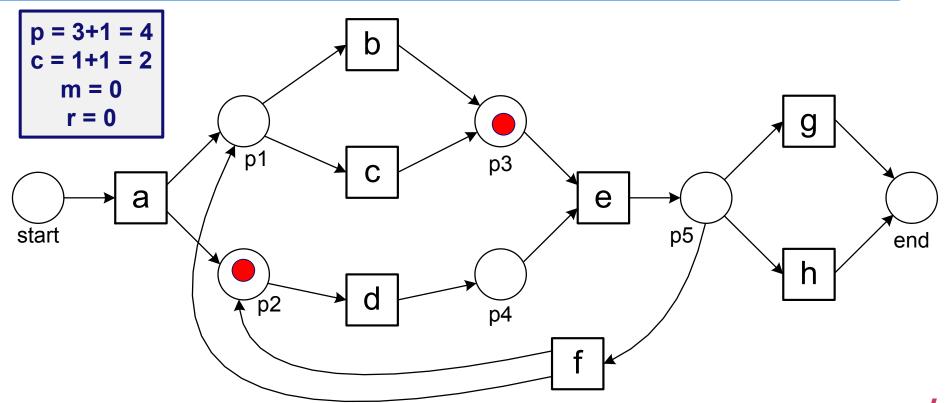


Replaying
$$\sigma_1 = \langle a(c)d, e, h \rangle$$



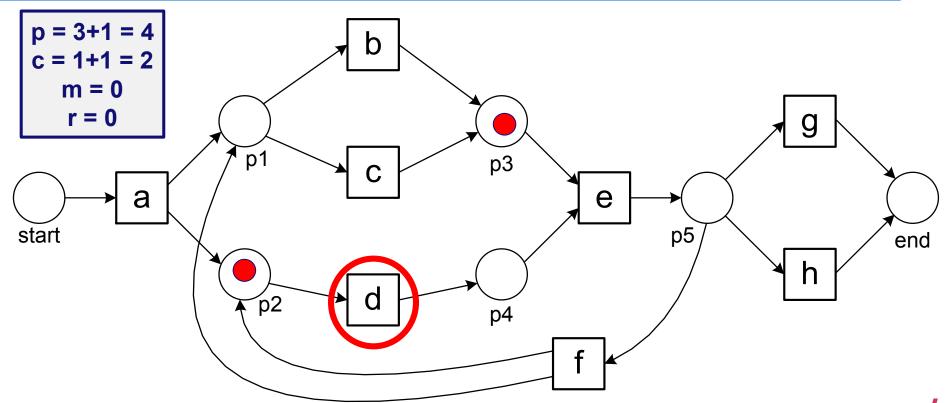


Replaying
$$\sigma_1 = \langle a, c, d, e, h \rangle$$



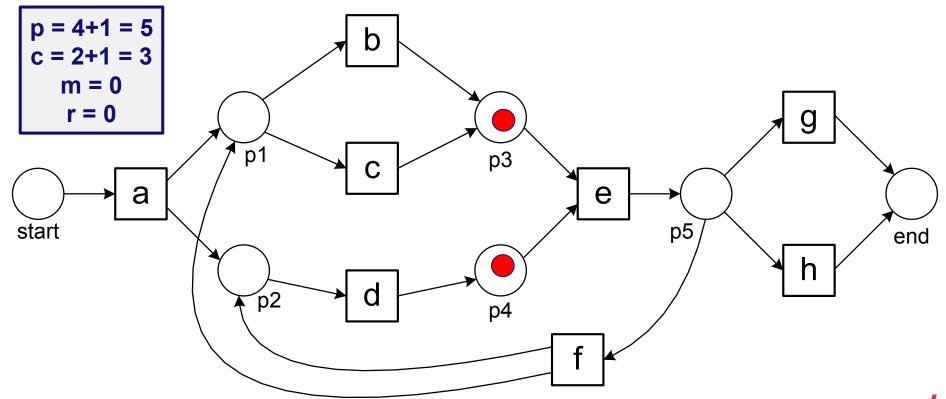


Replaying
$$\sigma_1 = \langle a, c(d)e, h \rangle$$



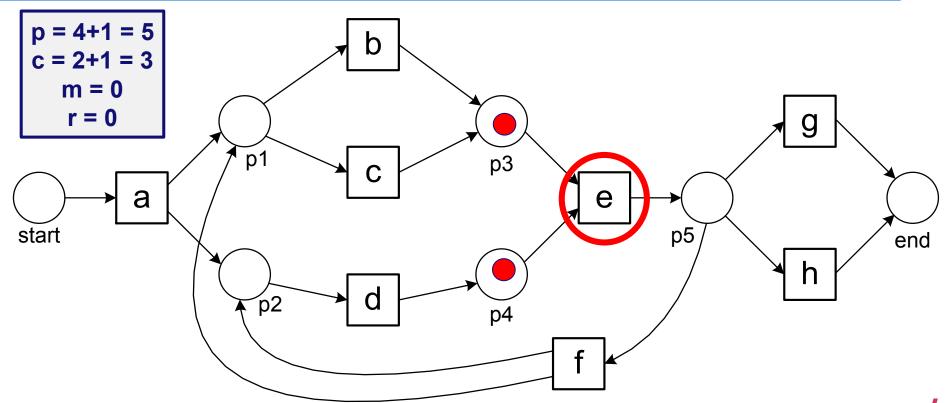


Replaying
$$\sigma_1 = \langle a, c, d, e, h \rangle$$



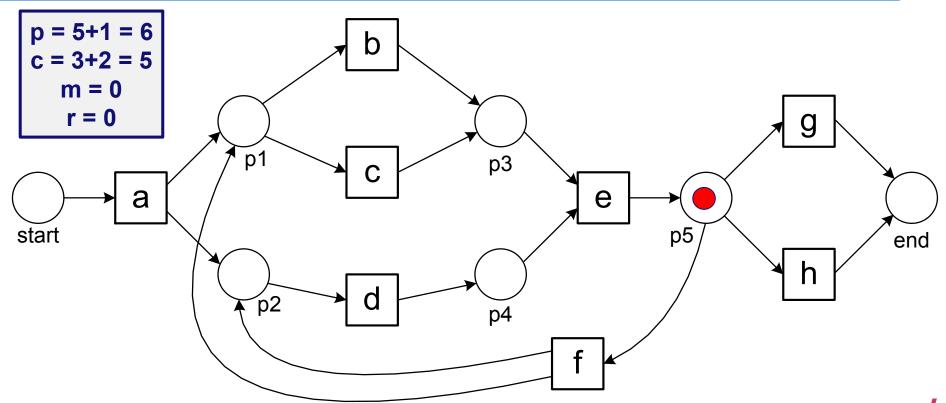


Replaying
$$\sigma_1 = \langle a, c, d(e, h) \rangle$$



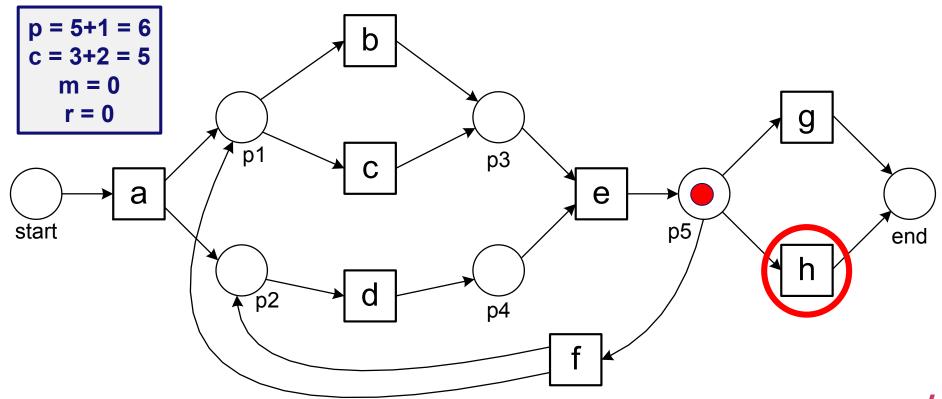


Replaying
$$\sigma_1 = \langle a, c, d, e, h \rangle$$



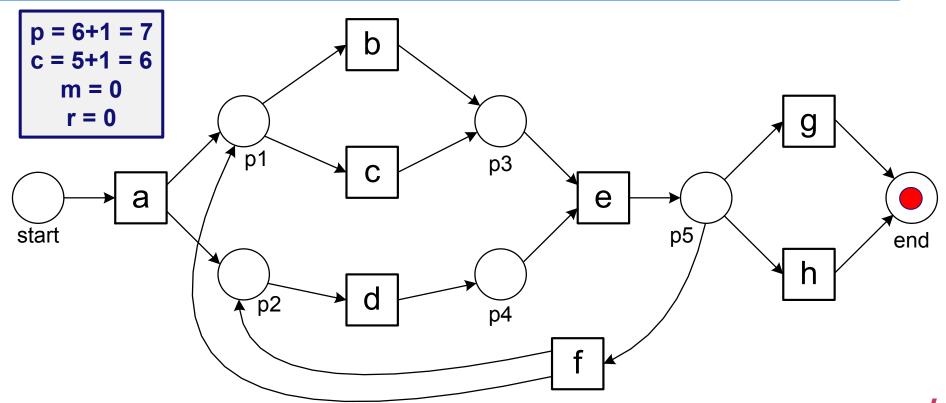


Replaying
$$\sigma_1 = \langle a, c, d, e \rangle$$



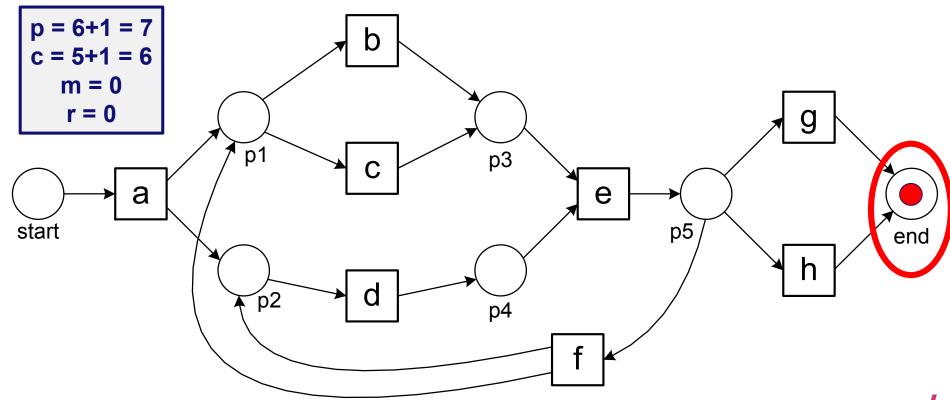


Replaying
$$\sigma_1 = \langle a, c, d, e, h \rangle$$



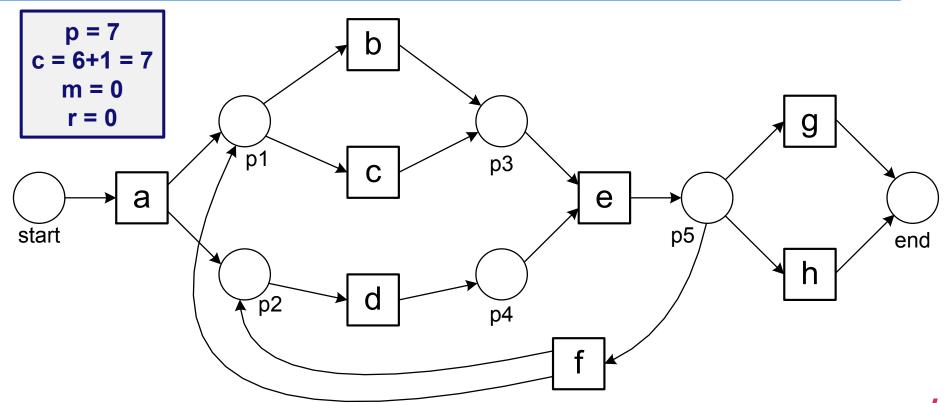


Replaying
$$\sigma_1 = \langle a, c, d, e, h \rangle$$





Replaying
$$\sigma_1 = \langle a, c, d, e, h \rangle$$





Quantifying fitness at the trace level

$$\sigma_1 = \langle a, c, d, e, h \rangle$$

$$fitness(\sigma, N) = \frac{1}{2} \left(1 - \frac{m}{c} \right) + \frac{1}{2} \left(1 - \frac{r}{p} \right)$$



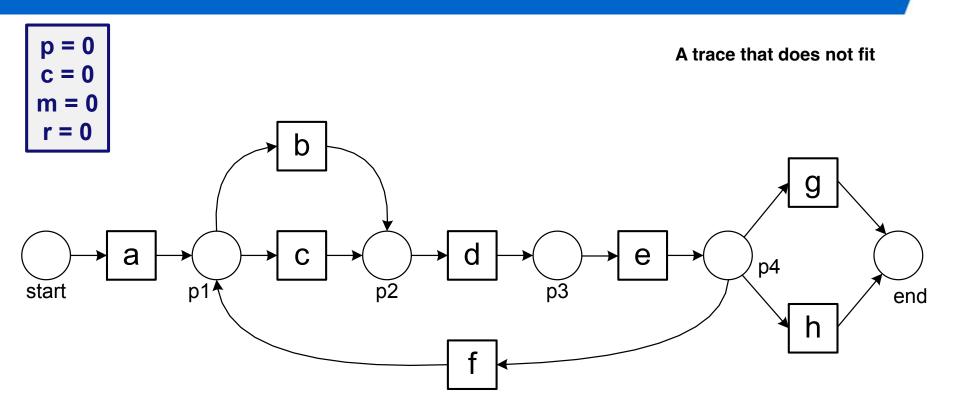
Quantifying fitness at the trace level

$$\sigma_1 = \langle a, c, d, e, h \rangle$$

$$fitness(\sigma, N) = \frac{1}{2} \left(1 - \frac{\mathbf{0}}{\mathbf{7}} \right) + \frac{1}{2} \left(1 - \frac{\mathbf{0}}{\mathbf{7}} \right) = \mathbf{1}$$

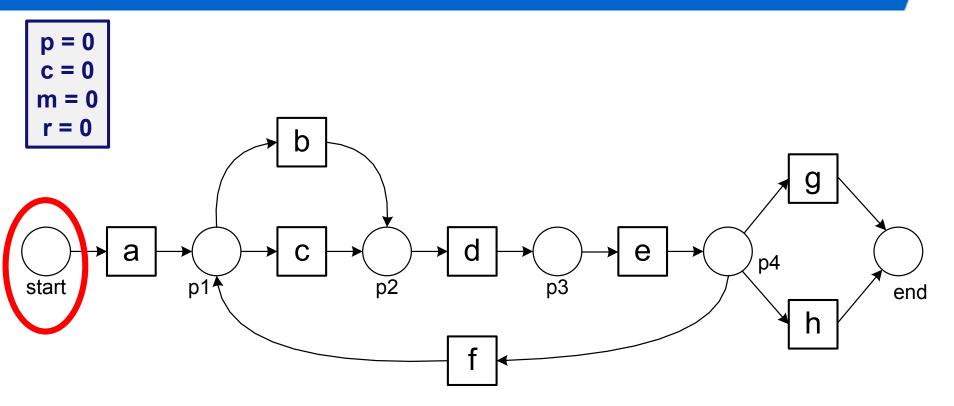


$$\sigma_3 = \langle a, d, c, e, h \rangle$$



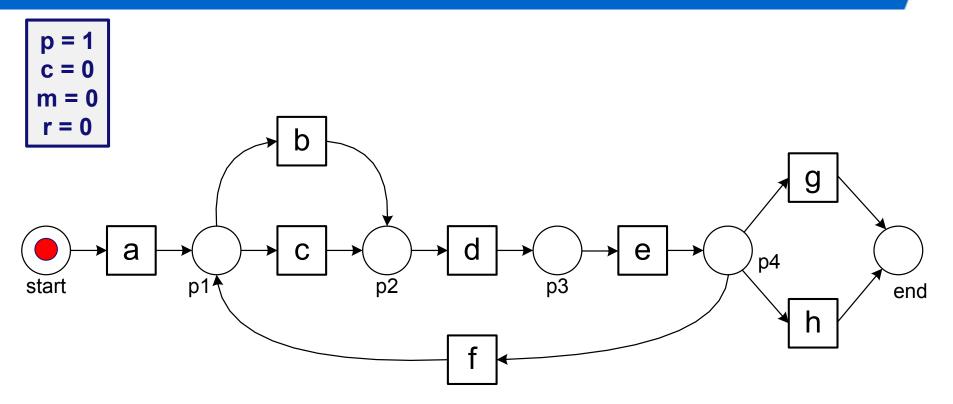


$$\sigma_3 = \langle a, d, c, e, h \rangle$$



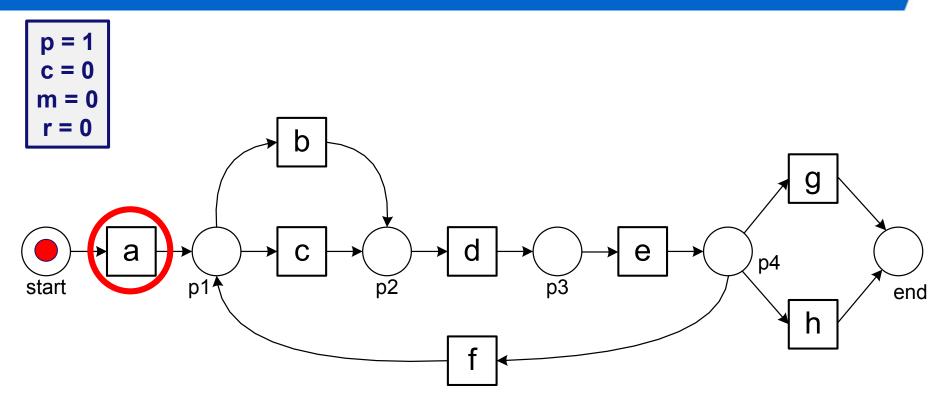


$$\sigma_3 = \langle a, d, c, e, h \rangle$$



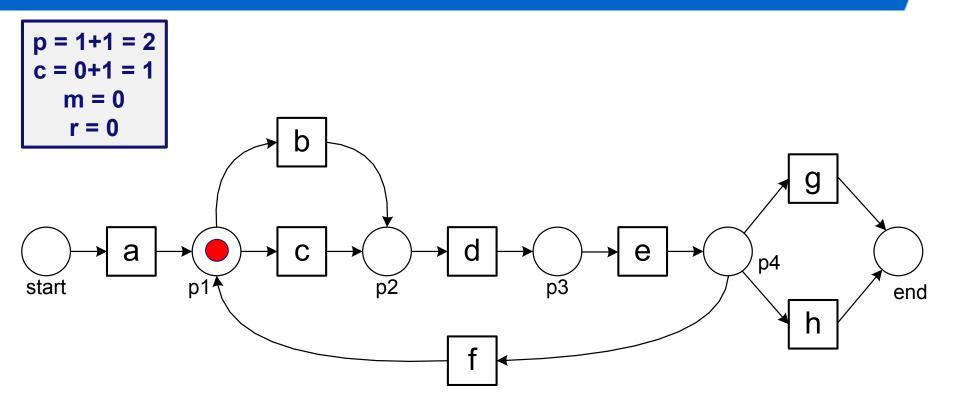


$$\sigma_3 = \langle a | d, c, e, h \rangle$$



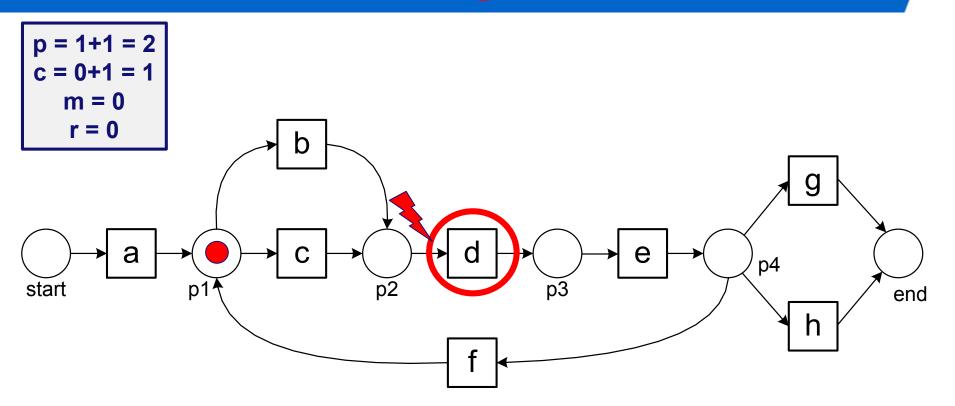


$$\sigma_3 = \langle a, d, c, e, h \rangle$$



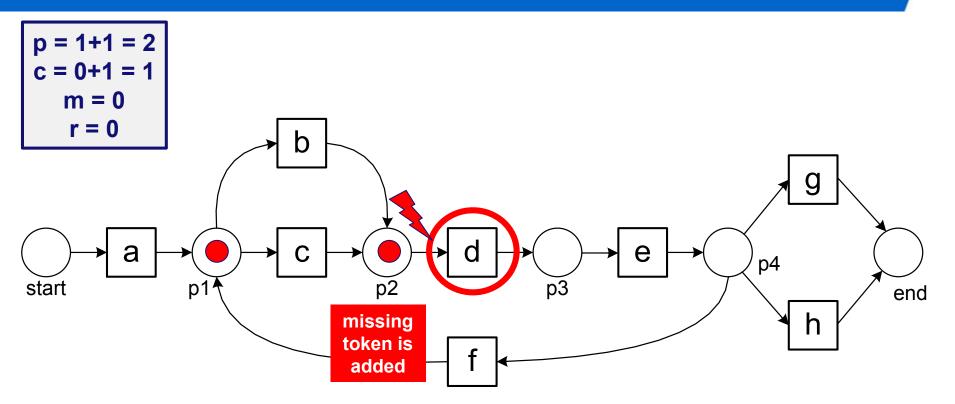


$$\sigma_3 = \langle a, d, c, e, h \rangle$$



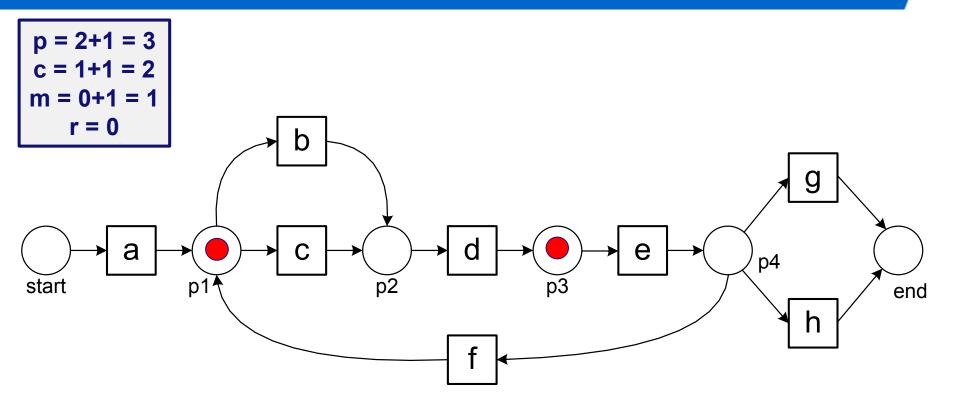


$$\sigma_3 = \langle a, d, c, e, h \rangle$$



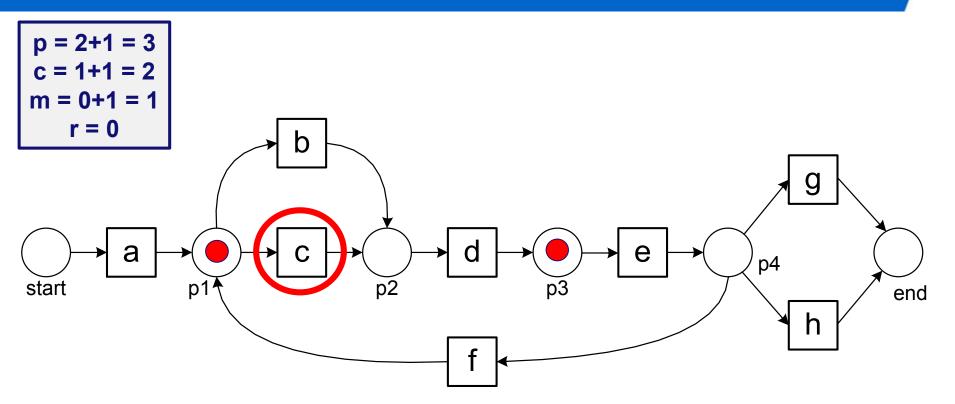


$$\sigma_3 = \langle a, d, c, e, h \rangle$$



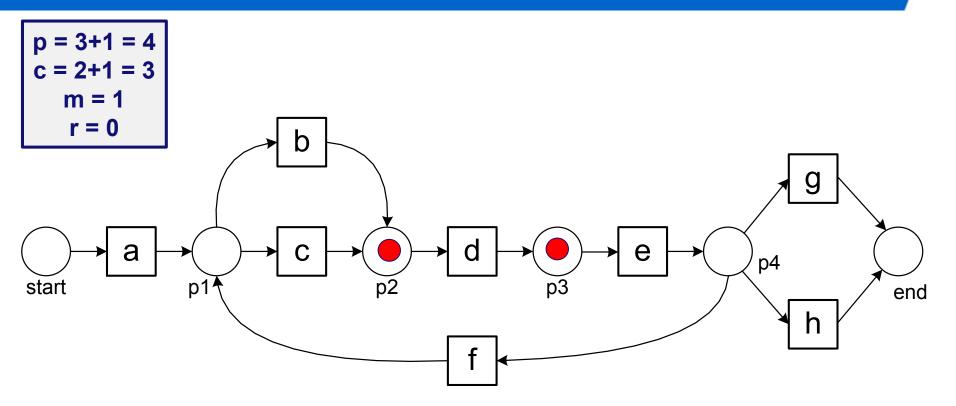


$$\sigma_3 = \langle a, d(c, e, h) \rangle$$



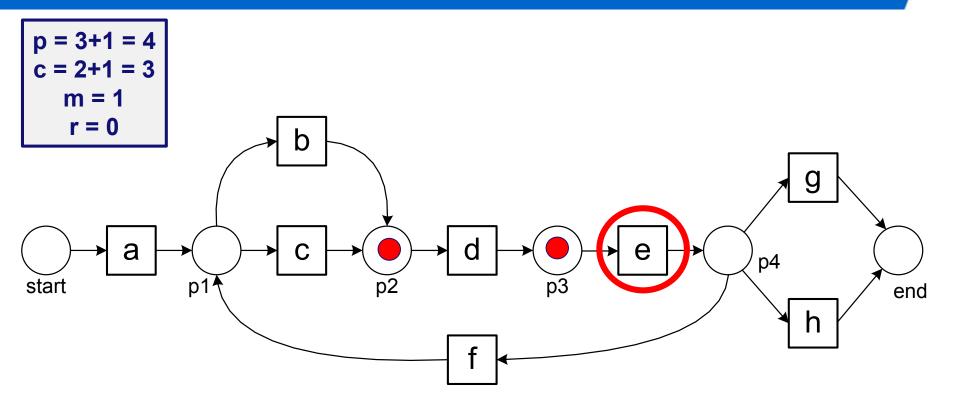


$$\sigma_3 = \langle a, d, c, e, h \rangle$$



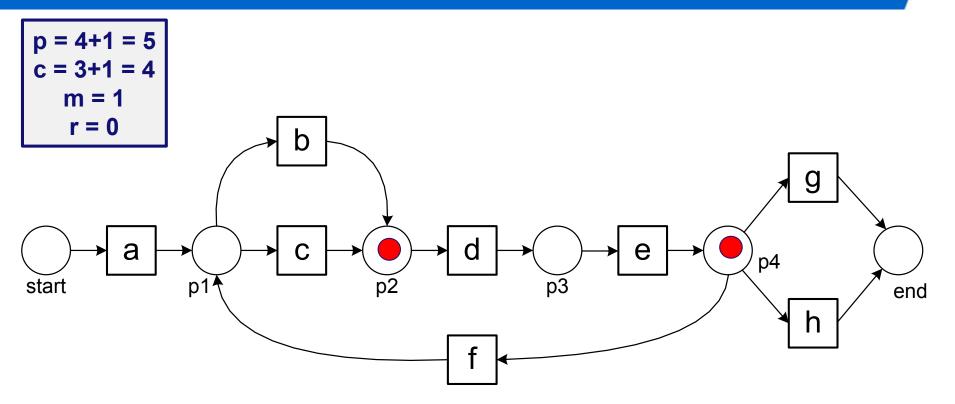


$$\sigma_3 = \langle a, d, c(e, h) \rangle$$



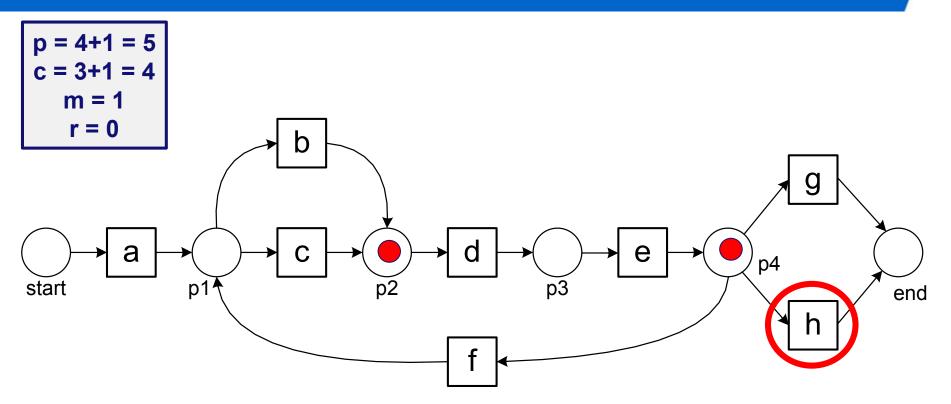


$$\sigma_3 = \langle a, d, c, e, h \rangle$$



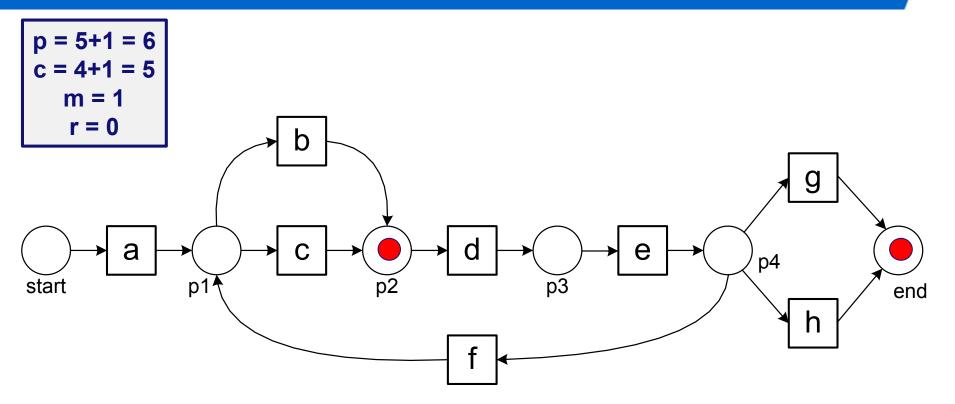


$$\sigma_3 = \langle a, d, c, e, h \rangle$$



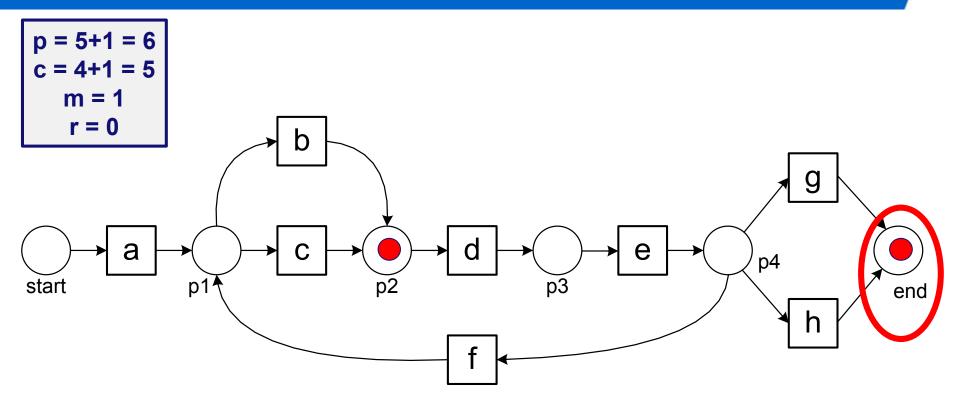


$$\sigma_3 = \langle a, d, c, e, h \rangle$$



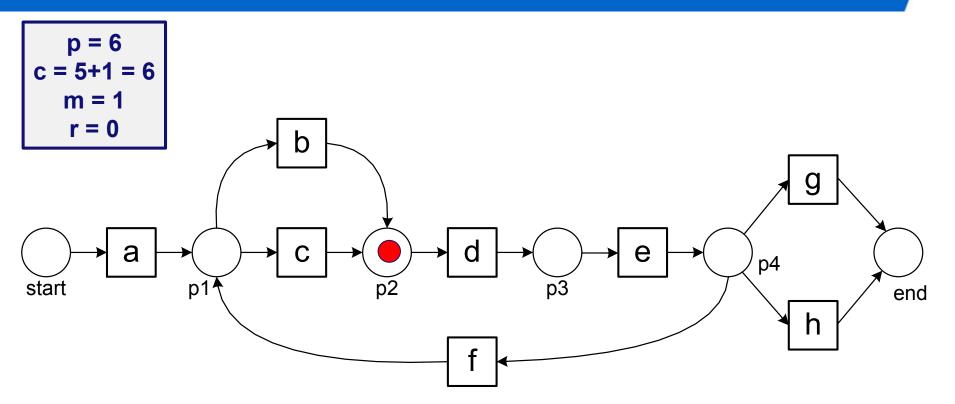


$$\sigma_3 = \langle a, d, c, e, h \rangle$$



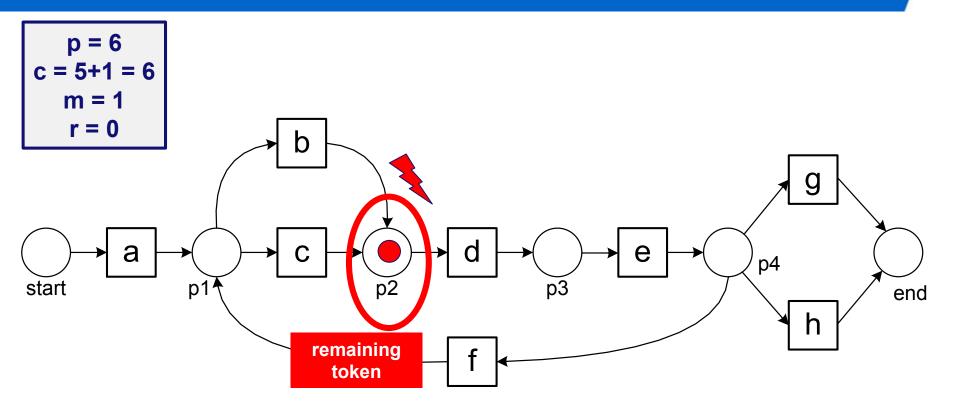


$$\sigma_3 = \langle a, d, c, e, h \rangle$$



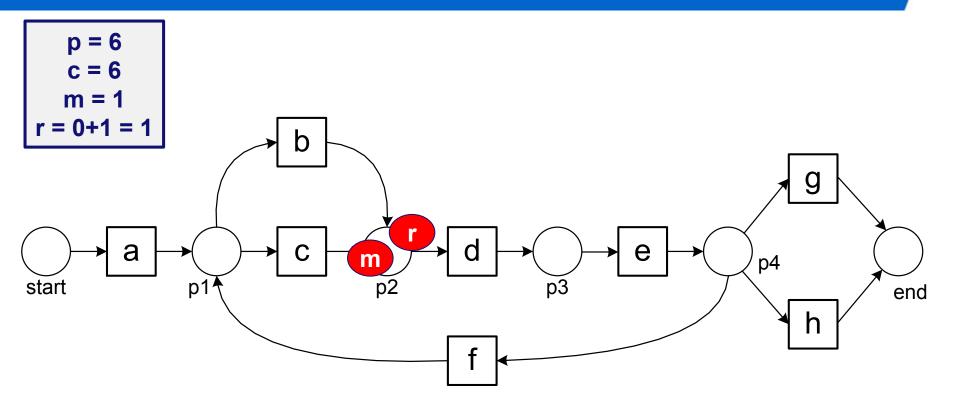


$$\sigma_3 = \langle a, d, c, e, h \rangle$$





$$\sigma_3 = \langle a, d, c, e, h \rangle$$





Quantifying fitness at the trace level

$$\sigma_3 = \langle a, d, c, e, h \rangle$$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, h \rangle$
 $\sigma_3 = \langle a, h, h$

$$fitness(\sigma, N) = \frac{1}{2} \left(1 - \frac{m}{c} \right) + \frac{1}{2} \left(1 - \frac{r}{p} \right)$$



Quantifying fitness at the trace level

$$\sigma_3 = \langle a, d, c, e, h \rangle$$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, c, e, h \rangle$
 $\sigma_3 = \langle a, d, h \rangle$
 $\sigma_3 = \langle a, h \rangle$
 σ

$$fitness(\sigma, N) = \frac{1}{2} \left(1 - \frac{1}{6} \right) + \frac{1}{2} \left(1 - \frac{1}{6} \right) = 0.8333$$



Fitness at the log level

$$fitness(L,N) = \frac{1}{2} \left(1 - \frac{\sum_{\sigma \in L} L(\sigma) \times m_{N,\sigma}}{\sum_{\sigma \in L} L(\sigma) \times c_{N,\sigma}} \right) +$$

Looks scary, but one just needs to take the sums of p, c, m, and r over the multiset of traces in de event log ...

$$\frac{1}{2} \left(1 - \frac{\sum_{\sigma \in L} L(\sigma) \times r_{N,\sigma}}{\sum_{\sigma \in L} L(\sigma) \times p_{N,\sigma}} \right)$$

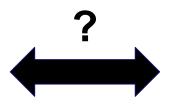


Fitness at the log level

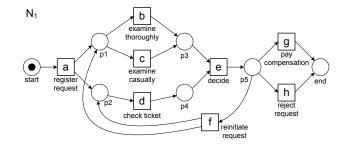
$$fitness(L,N) = \frac{1}{2} \left(1 - \frac{\sum_{\sigma \in L} L(\sigma) \times m_{N,\sigma}}{\sum_{\sigma \in L} L(\sigma) \times c_{N,\sigma}}\right) + \frac{1}{2} \left(1 - \frac{\sum_{\sigma \in L} L(\sigma) \times r_{N,\sigma}}{\sum_{\sigma \in L} L(\sigma) \times r_{N,\sigma}}\right) + \frac{1}{2} \left(1 - \frac{\sum_{\sigma \in L} L(\sigma) \times r_{N,\sigma}}{\sum_{\sigma \in L} L(\sigma) \times r_{N,\sigma}}\right) + \frac{1}{2} \left(1 - \frac{\sum_{\sigma \in L} L(\sigma) \times r_{N,\sigma}}{\sum_{\sigma \in L} L(\sigma) \times r_{N,\sigma}}\right) + \frac{1}{2} \left(1 - \frac{\sum_{\sigma \in L} L(\sigma) \times r_{N,\sigma}}{\sum_{\sigma \in L} L(\sigma) \times r_{N,\sigma}}\right) + \frac{1}{2} \left(1 - \frac{\sum_{\sigma \in L} L(\sigma) \times r_{N,\sigma}}{\sum_{\sigma \in L} L(\sigma) \times r_{N,\sigma}}\right) + \frac{1}{2} \left(1 - \frac{\sum_{\sigma \in L} L(\sigma) \times r_{N,\sigma}}{\sum_{\sigma \in L} L(\sigma) \times r_{N,\sigma}}\right) + \frac{1}{2} \left(1 - \frac{\sum_{\sigma \in L} L(\sigma) \times r_{N,\sigma}}{\sum_{\sigma \in L} L(\sigma) \times r_{N,\sigma}}\right) + \frac{1}{2} \left(1 - \frac{\sum_{\sigma \in L} L(\sigma) \times r_{N,\sigma}}{\sum_{\sigma \in L} L(\sigma) \times r_{N,\sigma}}\right) + \frac{1}{2} \left(1 - \frac{\sum_{\sigma \in L} L(\sigma) \times r_{N,\sigma}}{\sum_{\sigma \in L} L(\sigma) \times r_{N,\sigma}}\right) + \frac{1}{2} \left(1 - \frac{\sum_{\sigma \in L} L(\sigma) \times r_{N,\sigma}}{\sum_{\sigma \in L} L(\sigma) \times r_{N,\sigma}}\right) + \frac{1}{2} \left(1 - \frac{\sum_{\sigma \in L} L(\sigma) \times r_{N,\sigma}}{\sum_{\sigma \in L} L(\sigma) \times r_{N,\sigma}}\right) + \frac{1}{2} \left(1 - \frac{\sum_{\sigma \in L} L(\sigma) \times r_{N,\sigma}}{\sum_{\sigma \in L} L(\sigma) \times r_{N,\sigma}}\right) + \frac{1}{2} \left(1 - \frac{\sum_{\sigma \in L} L(\sigma) \times r_{N,\sigma}}{\sum_{\sigma \in L} L(\sigma) \times r_{N,\sigma}}\right) + \frac{1}{2} \left(1 - \frac{\sum_{\sigma \in L} L(\sigma) \times r_{N,\sigma}}{\sum_{\sigma \in L} L(\sigma) \times r_{N,\sigma}}\right) + \frac{1}{2} \left(1 - \frac{\sum_{\sigma \in L} L(\sigma) \times r_{N,\sigma}}{\sum_{\sigma \in L} L(\sigma) \times r_{N,\sigma}}\right) + \frac{1}{2} \left(1 - \frac{\sum_{\sigma \in L} L(\sigma) \times r_{N,\sigma}}{\sum_{\sigma \in L} L(\sigma) \times r_{N,\sigma}}\right) + \frac{1}{2} \left(1 - \frac{\sum_{\sigma \in L} L(\sigma) \times r_{N,\sigma}}{\sum_{\sigma \in L} L(\sigma) \times r_{N,\sigma}}\right) + \frac{1}{2} \left(1 - \frac{\sum_{\sigma \in L} L(\sigma) \times r_{N,\sigma}}{\sum_{\sigma \in L} L(\sigma) \times r_{N,\sigma}}\right) + \frac{1}{2} \left(1 - \frac{\sum_{\sigma \in L} L(\sigma) \times r_{N,\sigma}}{\sum_{\sigma \in L} L(\sigma) \times r_{N,\sigma}}\right) + \frac{1}{2} \left(1 - \frac{\sum_{\sigma \in L} L(\sigma) \times r_{N,\sigma}}{\sum_{\sigma \in L} L(\sigma) \times r_{N,\sigma}}\right) + \frac{1}{2} \left(1 - \frac{\sum_{\sigma \in L} L(\sigma) \times r_{N,\sigma}}{\sum_{\sigma \in L} L(\sigma) \times r_{N,\sigma}}\right) + \frac{1}{2} \left(1 - \frac{\sum_{\sigma \in L} L(\sigma) \times r_{N,\sigma}}{\sum_{\sigma \in L} L(\sigma) \times r_{N,\sigma}}\right) + \frac{1}{2} \left(1 - \frac{\sum_{\sigma \in L} L(\sigma) \times r_{N,\sigma}}{\sum_{\sigma \in L} L(\sigma) \times r_{N,\sigma}}\right) + \frac{1}{2} \left(1 - \frac{\sum_{\sigma \in L} L(\sigma) \times r_{N,\sigma}}{\sum_{\sigma \in L} L(\sigma) \times r_{N,\sigma}}\right) + \frac{1}{2} \left(1 - \frac{\sum_{\sigma \in L} L(\sigma) \times r_{N,\sigma}}{\sum_{\sigma \in L} L(\sigma) \times r_{N,\sigma}}\right) + \frac{1}{2} \left(1 - \frac{\sum_{\sigma \in L} L(\sigma) \times r_{N,\sigma}}{\sum_{\sigma \in L} L(\sigma) \times r_{N,\sigma}}\right) + \frac{1}{2} \left(1 - \frac{\sum_{\sigma \in L} L(\sigma) \times r_{N,\sigma}}{\sum_{\sigma \in L} L(\sigma) \times r_{N,\sigma}}\right) + \frac{1}{2} \left(1 -$$

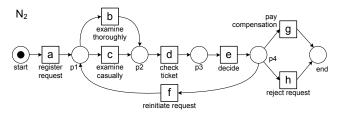


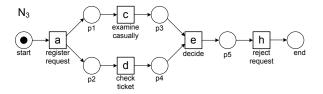
#	trace
455	acdeh
191	abdeg
177	adceh
144	abdeh
111	acdeg
82	adceg
56	adbeh
47	acdefdbeh
38	adbeg
33	acdefbdeh
14	acdefbdeg
11	acdefdbeg
9	adcefcdeh
8	adcefdbeh
5	adcefbdeg
3	acdefbdefdbeg
2	adcefdbeg
2	adcefbdefbdeg
1	adcefdbefbdeh
1	adbefbdefdbeg
1	adcefdbefcdefdbe
1301	

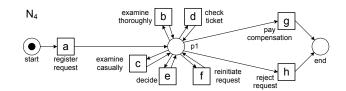


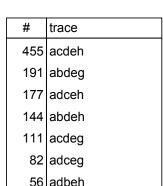
$$fitness(L,N) = \frac{1}{2} \left(1 - \frac{\sum_{\sigma \in L} L(\sigma) \times m_{N,\sigma}}{\sum_{\sigma \in L} L(\sigma) \times c_{N,\sigma}} \right) + \frac{1}{2} \left(1 - \frac{\sum_{\sigma \in L} L(\sigma) \times r_{N,\sigma}}{\sum_{\sigma \in L} L(\sigma) \times p_{N,\sigma}} \right)$$







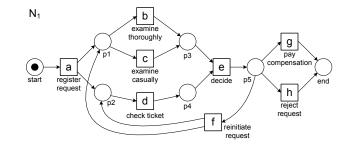


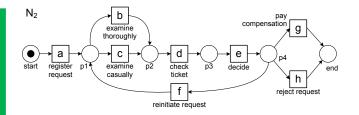


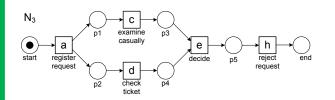
47 acdefdbeh

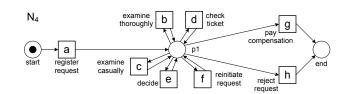
1391

$$fitness(L,N) = \frac{1}{2} \left(1 - \frac{\sum_{\sigma \in L} L(\sigma) \times m_{N,\sigma}}{\sum_{\sigma \in L} L(\sigma) \times c_{N,\sigma}} \right) + \frac{1}{2} \left(1 - \frac{\sum_{\sigma \in L} L(\sigma) \times r_{N,\sigma}}{\sum_{\sigma \in L} L(\sigma) \times p_{N,\sigma}} \right)$$







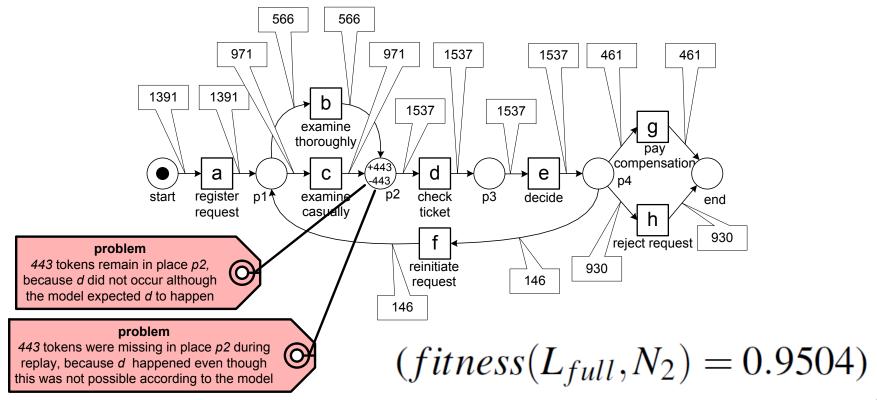


$fitness(L_{full}, N_1) = 1$ $fitness(L_{full}, N_2) = 0.9504$ $fitness(L_{full}, N_3) = 0.8797$ $fitness(L_{full}, N_4) = 1$

38 adbeg 33 acdefbdeh 14 acdefbdeg 11 acdefdbeg 9 adcefcdeh 8 adcefdbeh 5 adcefbdeg 3 acdefbdefdbeg 2 adcefdbeg 2 adcefbdefbdeg 1 adcefdbefbdeh 1 adbefbdefdbeg 1 adcefdbefcdefdbeg

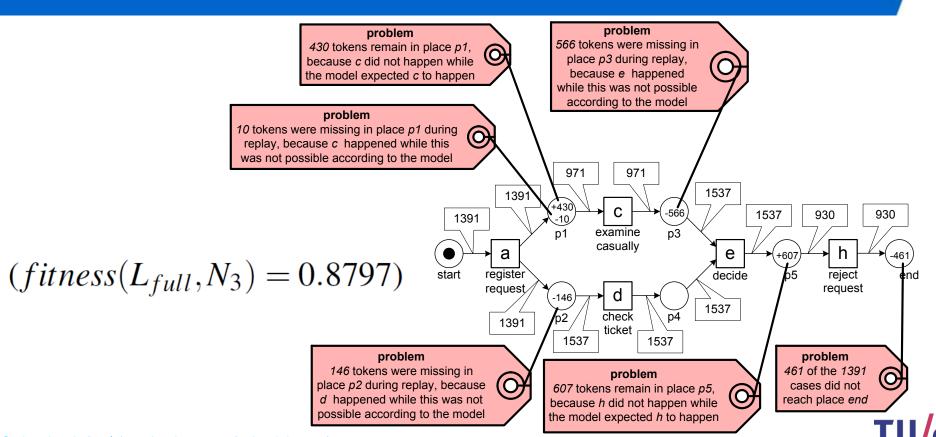
do properly evaluate the flower model we shall used precision

Diagnostics





Diagnostics



Part I: Introduction

Chapter 1 Data Science in Action

Chapter 2 Process Mining: The Missing Link

Part II: Preliminaries

Chapter 3 **Process Modeling** and Analysis

Chapter 4 **Data Mining**

Part III: From Event Logs to Process Models

Chapter 5 Getting the Data

Chapter 6 Process Discovery: An Introduction

Chapter 7

Advanced Process Discovery Techniques

Part IV: Beyond Process Discovery

Chapter 8 Conformance Checking

Chapter 9 Mining Additional Perspectives

Chapter 10 **Operational Support**

Chapter 14

Processes"

Analyzing "Spaghetti

Part V: Putting Process Mining to

Chapter 11

Process Mining Software

Chapter 12

Process Mining in the Large

apter 13

Analyzing "Lasagna Processes"

Process Mining

Wil van der Aalst

Data Science in Action Second Edition



Part VI: Reflection

Chapter 15 Cartography and

Navigation

Chapter 16 **Epilogue**



