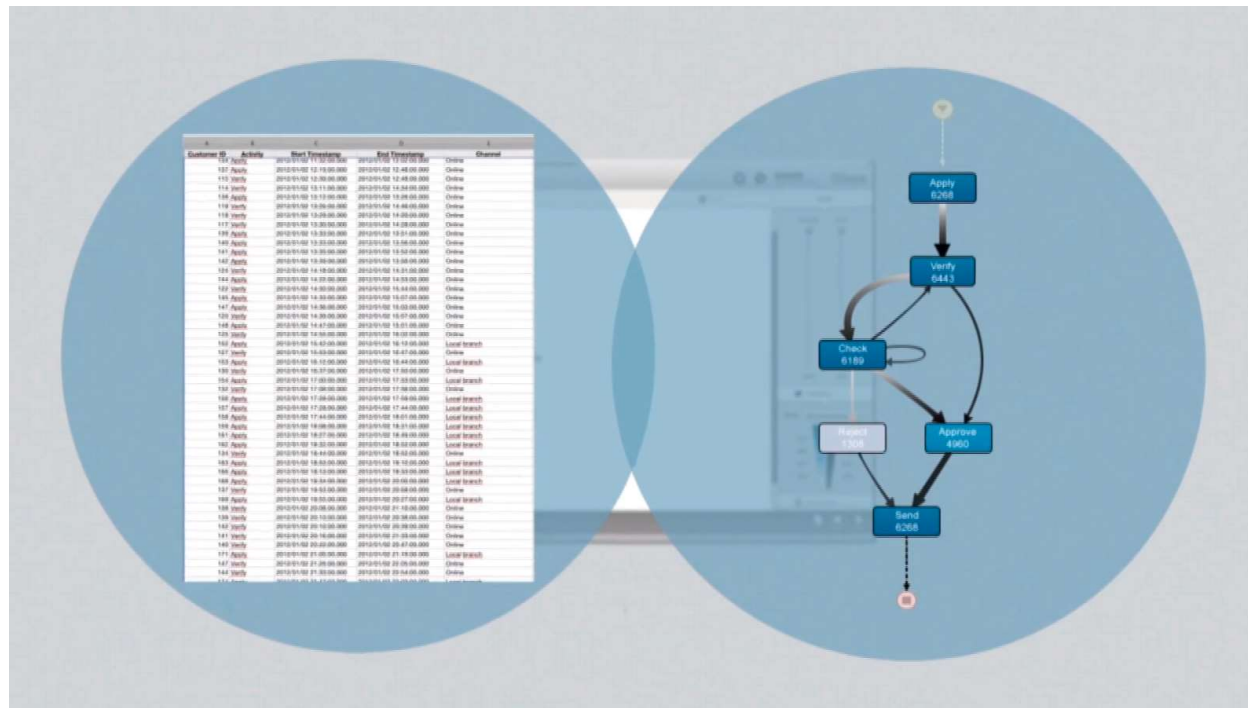


Folyamatbányászat 2

Konformancia elemzés – megfelelés elemzés



A vizsgálat szükségessége

- Ipari rendszerekben a munkafolyamatok lefutása sok esetben eltérhet a tervezett (ütemezett) lefutástól.
- Ezen eltérések okainak felkutatásának első lépése, hogy megtaláljuk, hogy a valós lefutás hol, milyen mértékben tér el a tervezettől.
- Egy ipari folyamat modellje tartalmazza az adott műveletsor lépéseit, átmeneteit, erőforrás (emberi, anyagi) felhasználását, időigényeit, viszont nem tartalmazhat minden apró részletet a folyamattal és pl. az emberekkel kapcsolatban.
- Ezért van szükség a konformancia vizsgálatra, mely segítségével megkereshetjük a modell és a tényleges lefutások közötti különbségeket.
- Mivel a modell az optimális lefutást hivatott leírni, ezért bármilyen eltérés a modelltől negatív hatással van a lefutásra és eltávolítja azt az optimálistól.

Példák

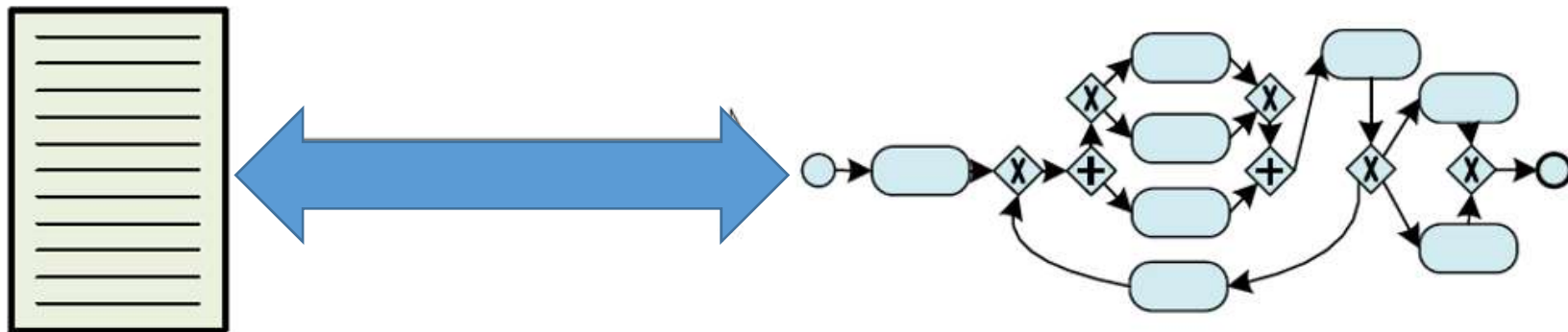
- **Első példa**, amikor **egy gépi erőforrás meghibásodik** és azt javítani vagy másikkal helyettesíteni kell. Ebben az esetben **a valós lefutás jelentősen el fog térni a modelltől**, mivel jelentős időbeli és erőforrásbeli eltéréseket mutathat.
- **Másik példa**, nagyon sok ipari folyamat során szükség van emberi erőforrásra, vagyis munkásokra, legyen ez szalag melletti összeszerelés vagy éppen könyvelés vagy más irodai munkák.

A gépesített munkavégzés egyik előnye a kézi munkavégzéssel szemben a gyorsaság és a pontosság, ellenben **amikor valamilyen hiba merül fel, akkor az ember képes azt felismerni és a legtöbb esetben gyorsan korrigálni**, míg egy meghibásodott gép sokkal drasztikusabb kiesést eredményez a termelésben.

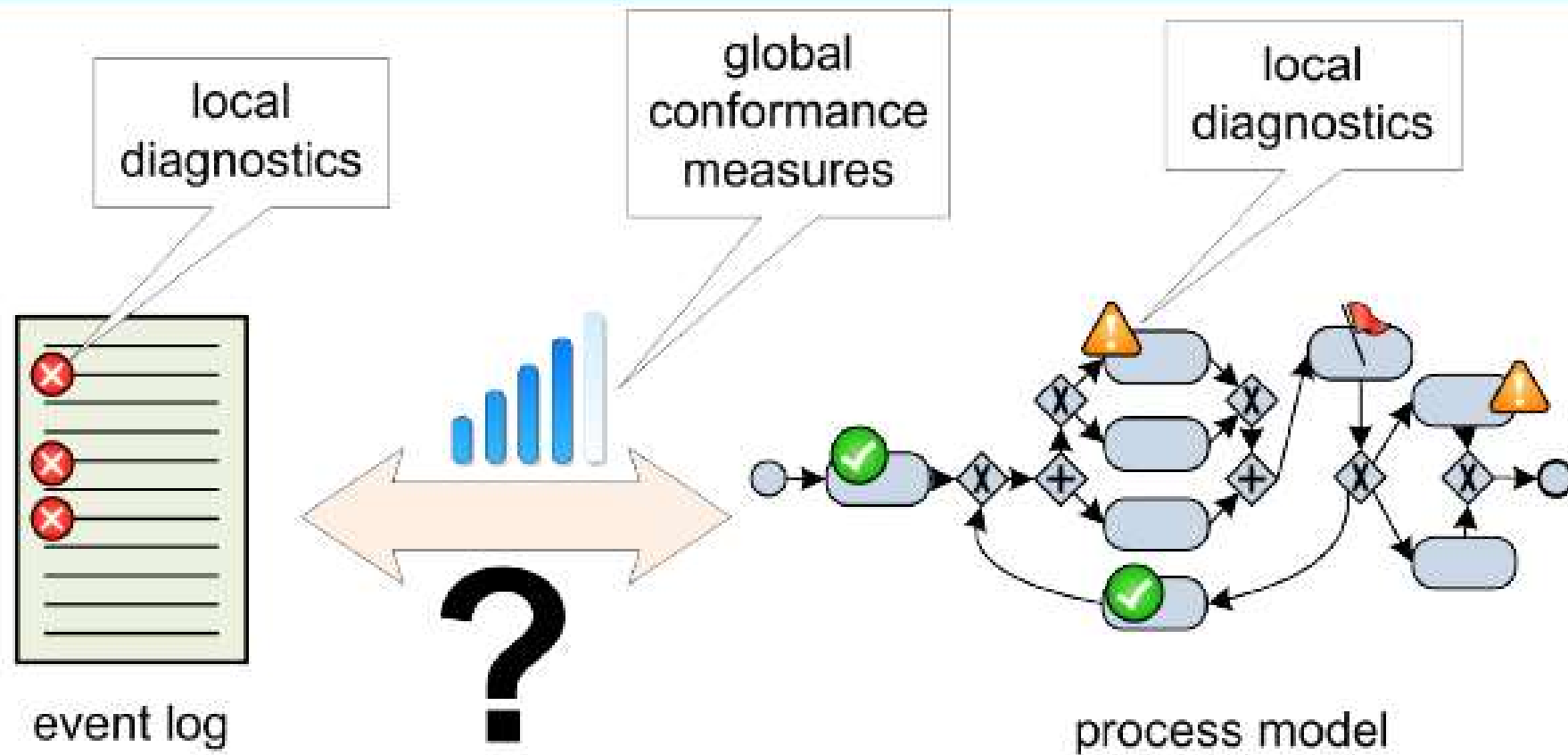
Még ezek tudatában is sokkal **nehezebb modellt készíteni egy emberek által elvégzett munkára**, mert az egyes emberek közötti különbségeken felül az idő és az aktuális hangulat is nagyon befolyásolja a munkavégzést.

Konformancia analízis (Conformance checking)

- Akkor használatos, ha van előzetes modell.
- A létező modellt hasonlítjuk össze a folyamat eseménynaplójával.
- Analizálja az eltéréseket az események és a modell között.
- Segítségével meghatározzuk az egyezés mértékét a modell által leírt működés és a tényleges esemény sorok között.



Konformancia, diagnosztika és mértékek



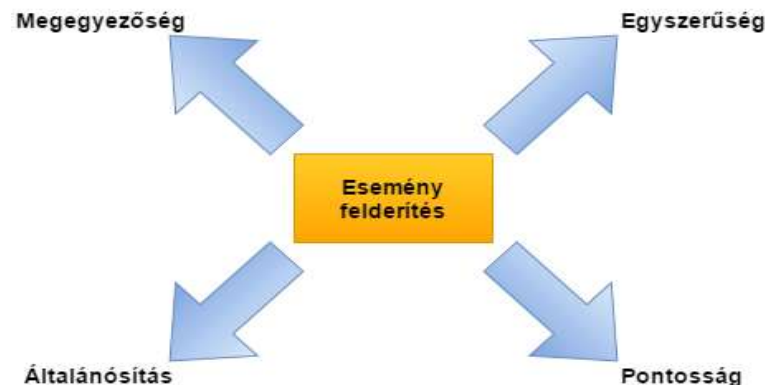
Megfelelően elkészített modell

A modell négy fő tulajdonsággal rendelkezik:

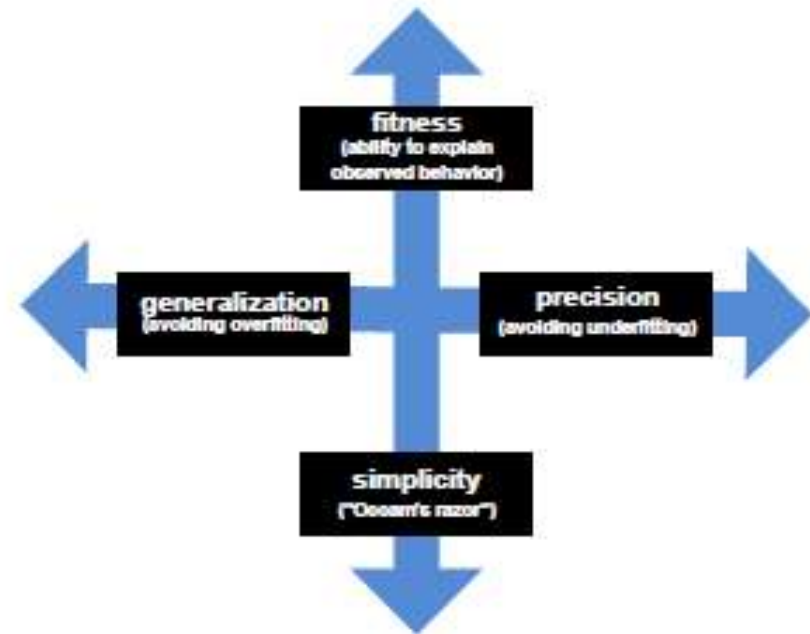
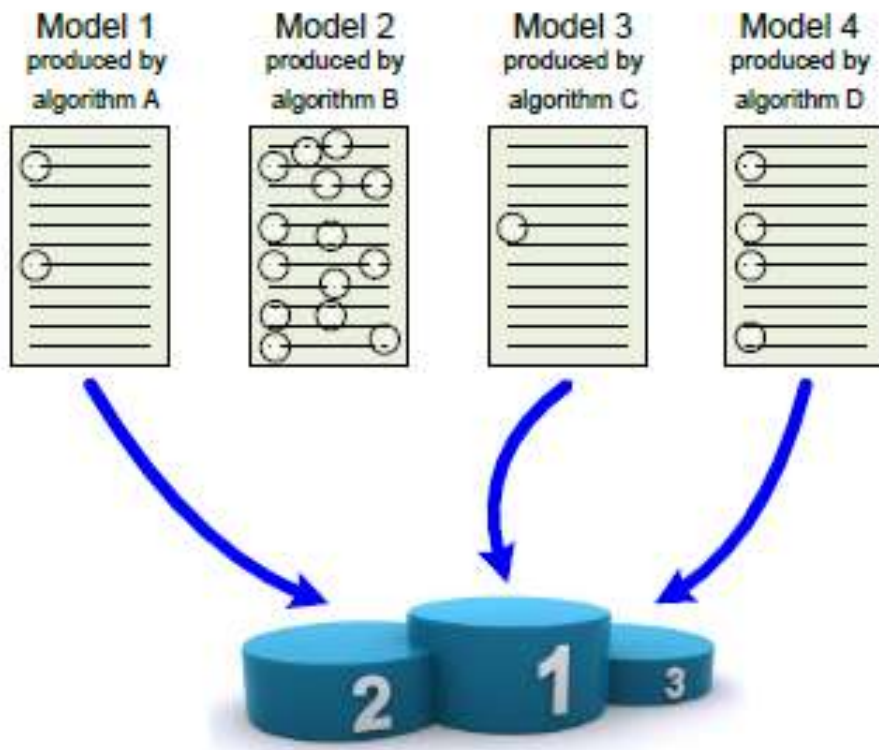
- **megegyezőség** (fitness): az eseménynaplóban felsorolt események mennyire jól illeszkednek, azaz az általuk leírt útvonallal milyen mértékig lehet végig menni a modellen,
- **egyszerűség** (simplicity): a lehetséges legegyszerűbb modell mely leírja az események viselkedését a legjobb,

(Megj.: A megegyezőség és az egyszerűség önmagukban nem elegendőek a modell minőségének megállapításához.)

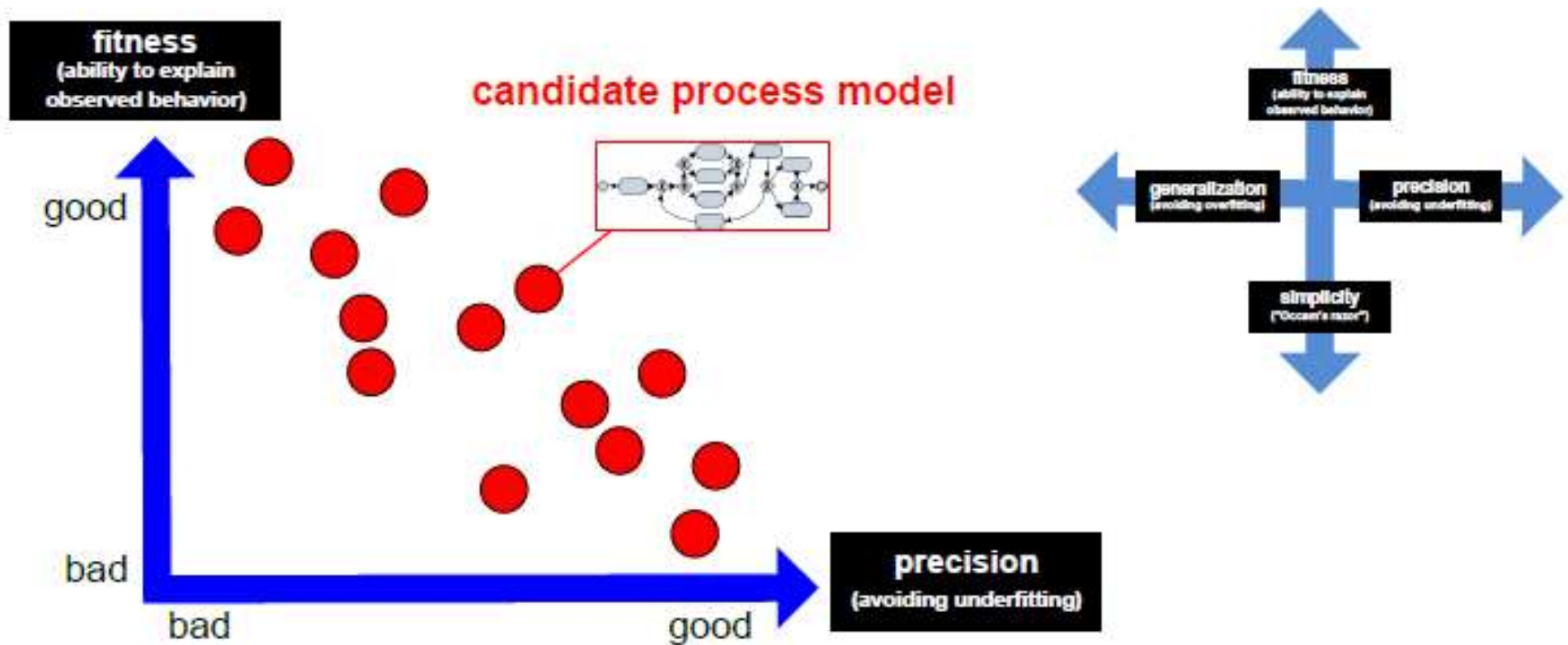
- **általánosítás** (generalization): törekedni kell minél általánosabb modell készítésére, azaz a modell minél több viselkedést ír le annál általánosabb,
- **pontosság** (precision): arra törekszünk, hogy a lehető legkevesebb viselkedést írjuk le (Megj.: Ha túlságosan specifikusan állítjuk össze a modellt, akkor az eseménynaplóban leírt események csak egy kisebb csoportját írja le jól, a többit hiányosan vagy egyáltalán nem.)



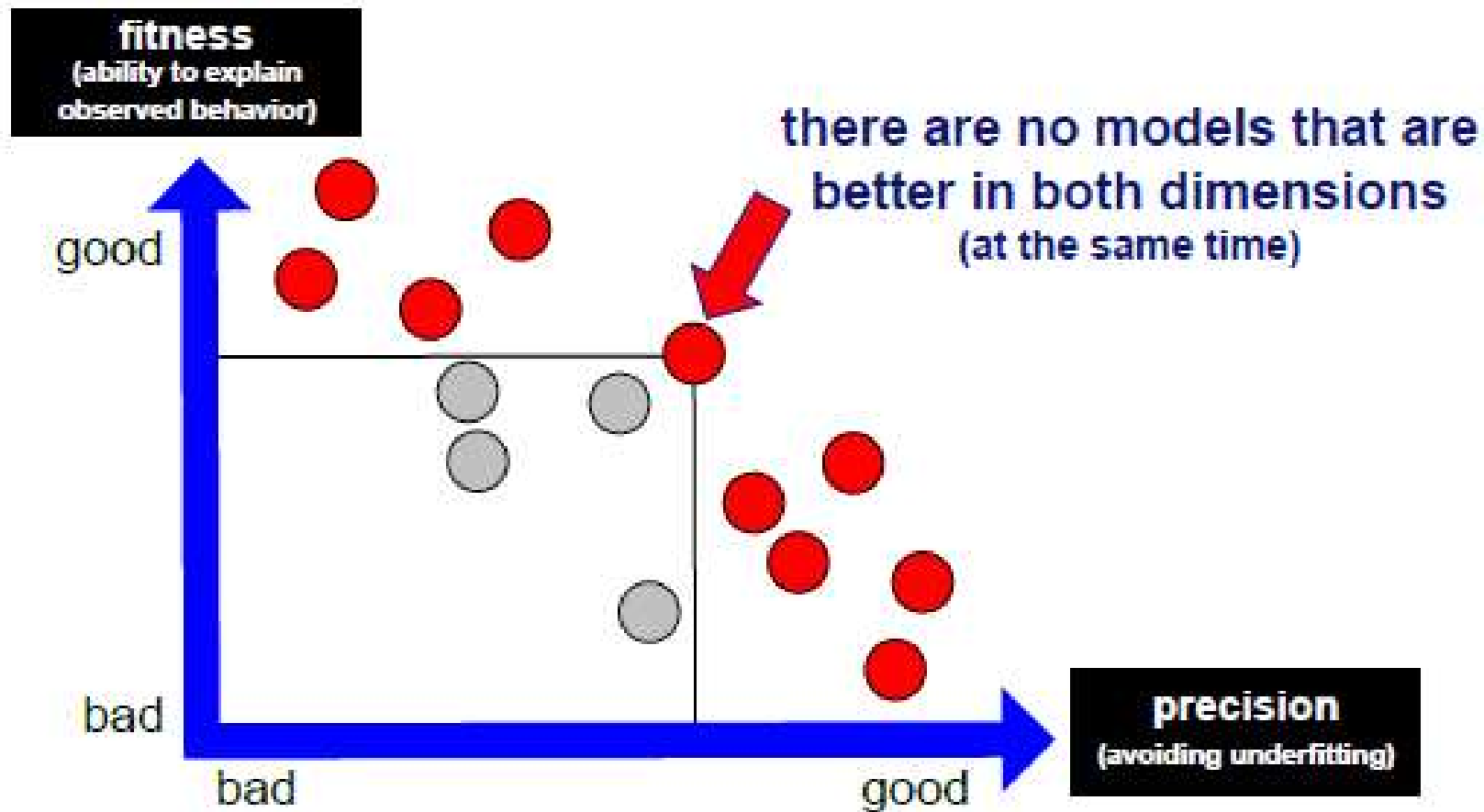
Folyamat feltáró algoritmusok vizsgálata



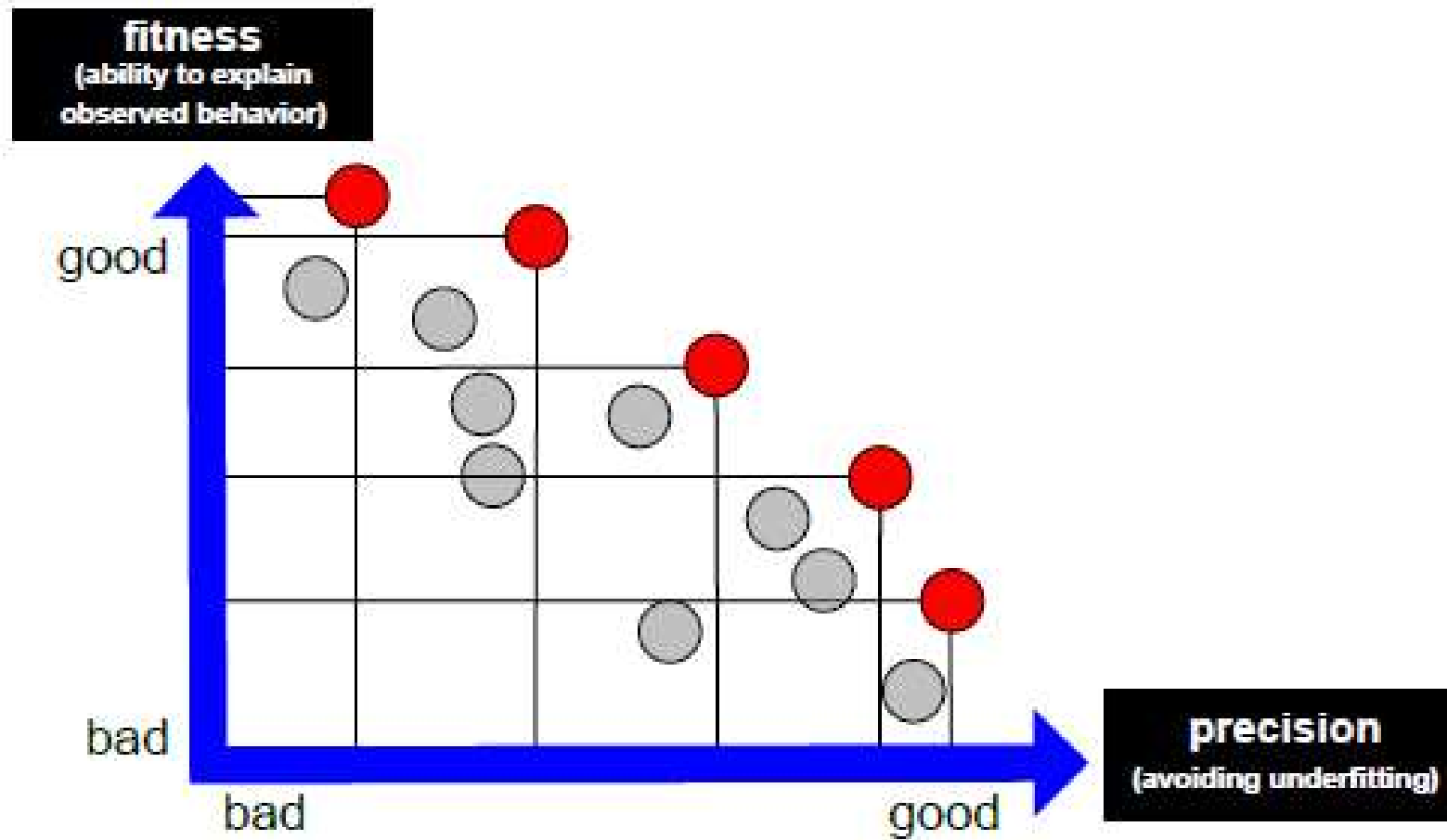
Pareto front 2 dimenziót figyelembe véve



Pareto front 2 dimenziót figyelembe véve



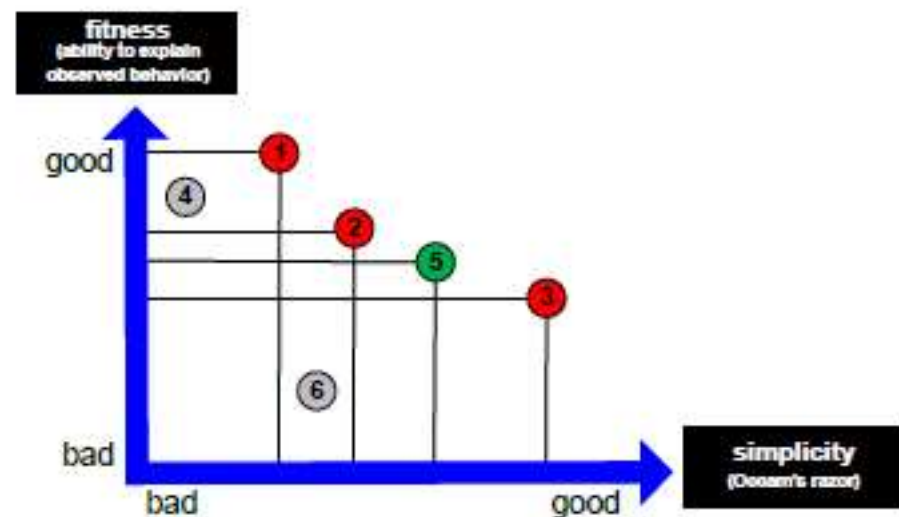
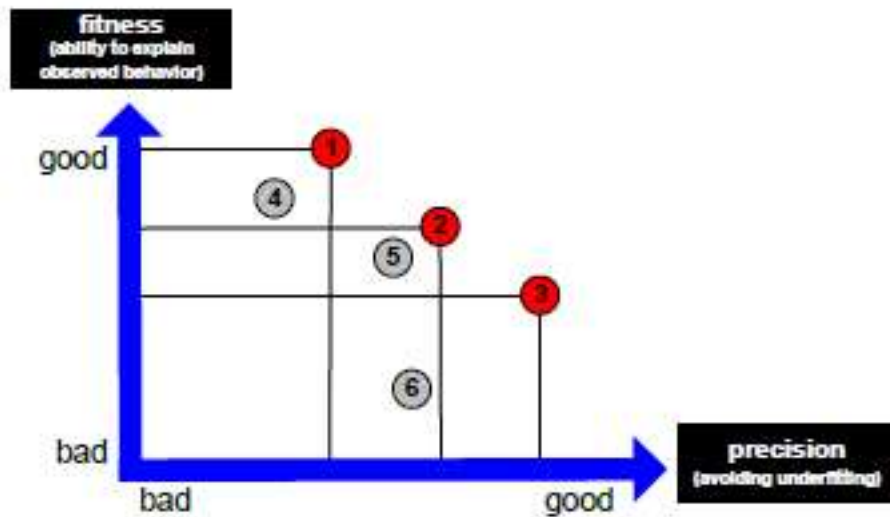
Pareto front 2 dimenziót figyelembe véve



Egy további dimenziót is megvizsgálunk

Nincs olyan modell, amely egy időben a legjobb fitness, pontosság és egyszerűség értékkel rendelkezik.

Nehéz a feltáró modelleket összehasonlítani!

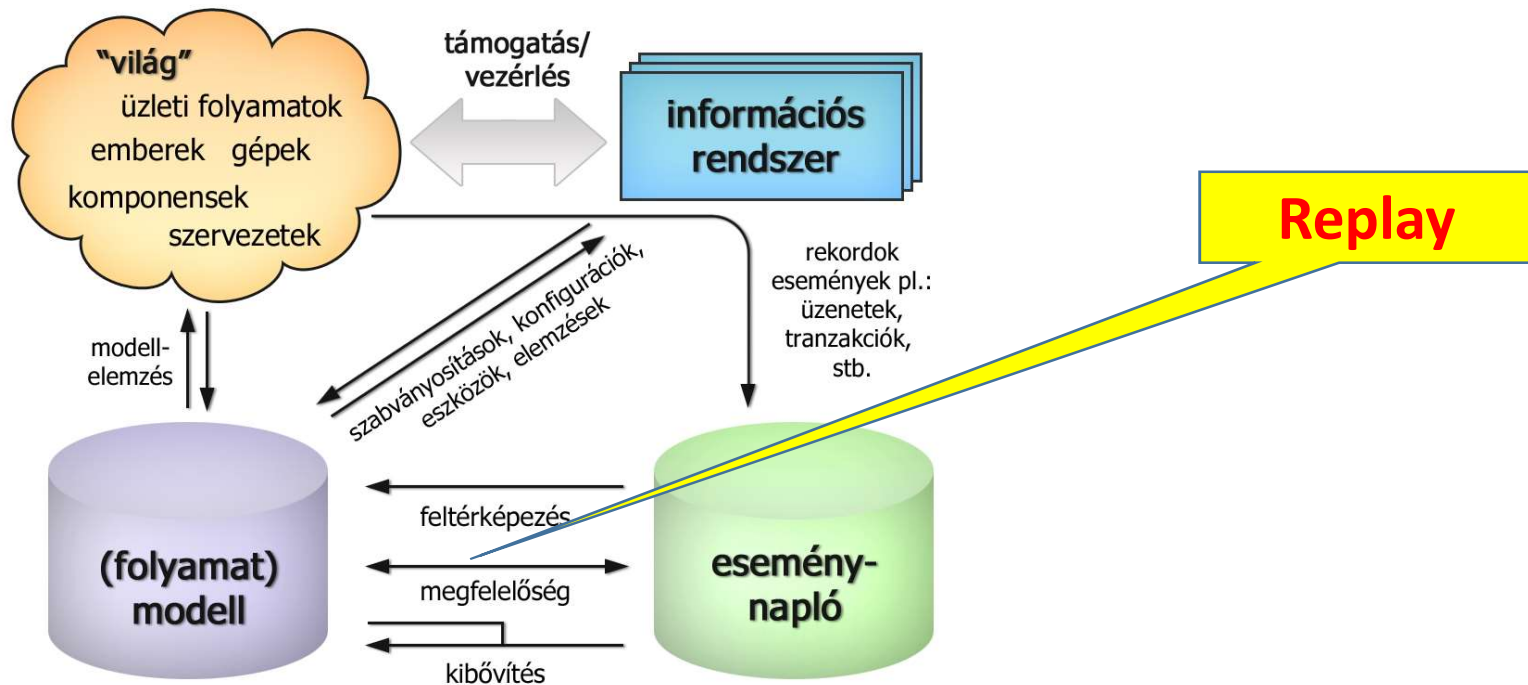


Miért nehéz a folyamat pontos feltárása?

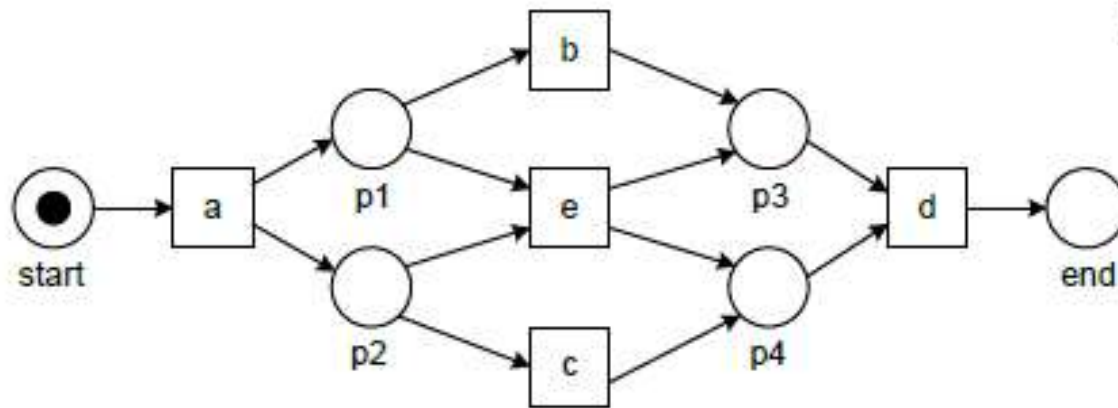
- **Nincsenek negatív példák** (azaz a logból láthatjuk, hogy mi történt, de azt nem láthatjuk, hogy mi nem történt)
- A párhuzamosság, a ciklusok és választások a keresési térnek egy komplex struktúráját adják meg, de **a logok jellemzően csak a lehetséges viselkedések egy töredék részét tartalmazzák.**
- **Nincs egyértelmű kapcsolat a modell mérete és viselkedése között.**
- Nagyon fontos a reprezentációból adódó **torzítás mértékének** meghatározása, figyelembevétele.
- A **modellező nyelvek adta lehetőségektől** is függ, hogy mit tudunk megjeleníteni, mi fér bele a modell leírásba. (UML, BPMN, Petri háló, stb.)

A konformancia ellenőrzés módszerei

1. Konformancia ellenőrzés ok-okozati lábnyomokkal,
2. Konformancia ellenőrzés token alapú visszajátszással,
3. Illesztés alapú megfelelés ellenőrzés.



Konformancia ellenőrzés ok-okozati lábnyomokkal



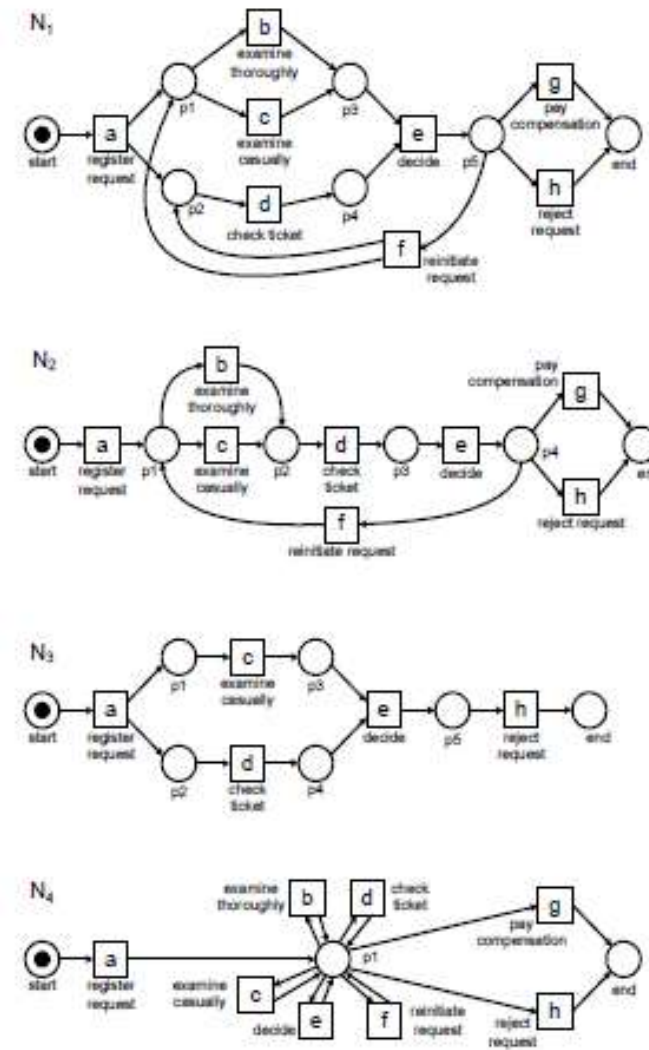
$$L_1 = [\langle a, b, c, d \rangle^3, \langle a, c, b, d \rangle^2, \langle a, e, d \rangle]$$

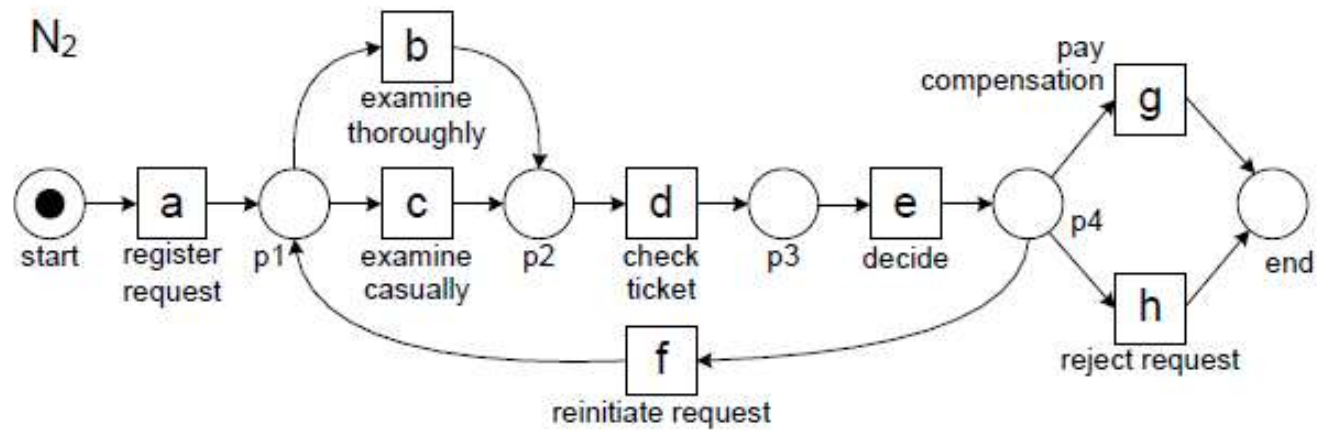
	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>
<i>a</i>	# L_1	\rightarrow_{L_1}	\rightarrow_{L_1}	# L_1	\rightarrow_{L_1}
<i>b</i>	\leftarrow_{L_1}	# L_1	\parallel_{L_1}	\rightarrow_{L_1}	# L_1
<i>c</i>	\leftarrow_{L_1}	\parallel_{L_1}	# L_1	\rightarrow_{L_1}	# L_1
<i>d</i>	# L_1	\leftarrow_{L_1}	\leftarrow_{L_1}	# L_1	\leftarrow_{L_1}
<i>e</i>	\leftarrow_{L_1}	# L_1	# L_1	\rightarrow_{L_1}	# L_1

lábnyom alapú megfelelés = 1 (1 = tökéletes egyezés, 0 = nincs egyezés)

A lábnyomok megadják a log és a modell egybevágóságát.

#	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	adcefbdefbdeh
1	adbefbdefdbeg
1	adcefbdefcdefdbeg
1391	





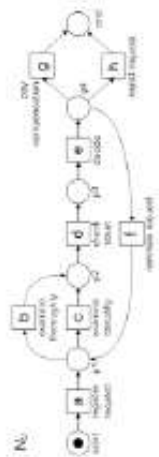
	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>	<i>g</i>	<i>h</i>
<i>a</i>	#	→	→	#	#	#	#	#
<i>b</i>	←	#	#	→	#	←	#	#
<i>c</i>	←	#	#	→	#	←	#	#
<i>d</i>	#	←	←	#	→	#	#	#
<i>e</i>	#	#	#	←	#	→	→	→
<i>f</i>	#	→	→	#	←	#	#	#
<i>g</i>	#	#	#	#	←	#	#	#
<i>h</i>	#	#	#	#	←	#	#	#

L_{full}

#	base
455	acda
191	abdeg
177	adceh
144	abdef
111	acdeg
82	adceg
58	acdeh
47	acdefgh
38	addeg
33	acdefgh
14	acdefceeg
11	acdefg
9	adcefdh
8	acdefgh
5	adcefhg
3	acdefgcbeg
2	acdefg
2	acdefgcbdeg
1	acdefghcbefh
1	acdefghcbefhg
1	acdefghcbefg

	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>	<i>g</i>	<i>h</i>
<i>a</i>	#	→	→	→	#	#	#	#
<i>b</i>	←	#	#		→	←	#	#
<i>c</i>	←	#	#		→	←	#	#
<i>d</i>	←			#	→	←	#	#
<i>e</i>	#	←	←	←	#	→	→	→
<i>f</i>	#	→	→	→	←	#	#	#
<i>g</i>	#	#	#	#	←	#	#	#
<i>h</i>	#	#	#	#	←	#	#	#

N_2



	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>	<i>g</i>	<i>h</i>
<i>a</i>	#	→	→	#	#	#	#	#
<i>b</i>	←	#	#	→	#	←	#	#
<i>c</i>	←	#	#	→	#	←	#	#
<i>d</i>	#	←	←	#	→	#	#	#
<i>e</i>	#	#	#	←	#	→	→	→
<i>f</i>	#	→	→	#	←	#	#	#
<i>g</i>	#	#	#	#	←	#	#	#
<i>h</i>	#	#	#	#	←	#	#	#

A különbségek számszerűsítése

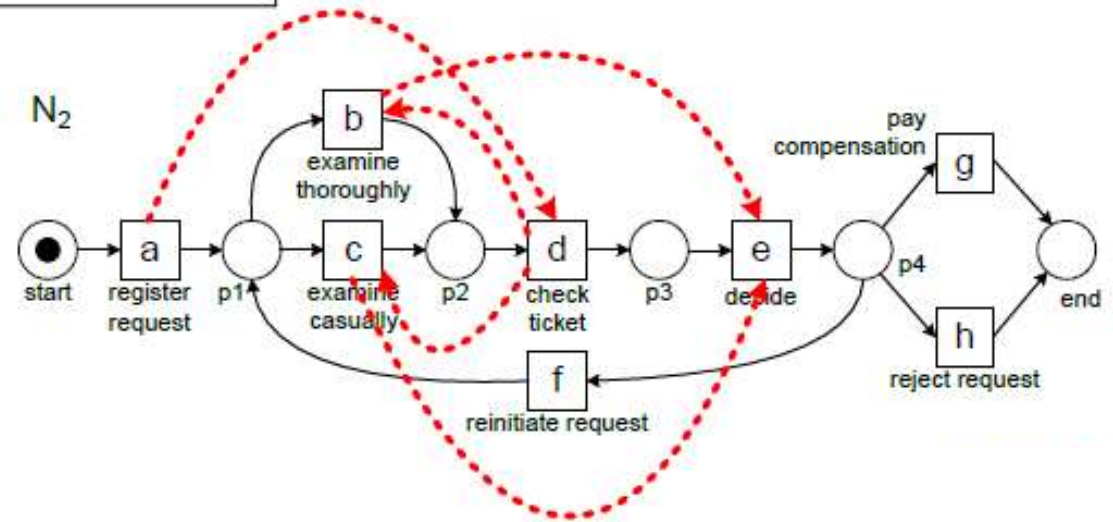
1 - x/y, ahol x a log-beli eltérések száma a modellbeli bejegyzésekhez képest, y az összes lábnyom száma a táblázatban

	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>	<i>g</i>	<i>h</i>
<i>a</i>				→: #				
<i>b</i>				:→	→: #			
<i>c</i>				:→	→: #			
<i>d</i>	←: #	:←	:←				←: #	
<i>e</i>		←: #	←: #					
<i>f</i>				→: #				
<i>g</i>								
<i>h</i>								

$$1 - \frac{12}{64} = 0.8125$$

A kapott eredmény felhasználása diagnosztikához

	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>	<i>g</i>	<i>h</i>
<i>a</i>				→: #				
<i>b</i>				:→	→: #			
<i>c</i>				:→	→: #			
<i>d</i>	←: #	:←	:←			←: #		
<i>e</i>		←: #	←: #					
<i>f</i>				→: #				
<i>g</i>								
<i>h</i>								



Konformancia token alapú egyezőség-vizsgálata

- Ehhez szükség van a **folyamat modell**jére, pl. Petri háló formájában
- A **napló fájlok** rengeteg információt tartalmaznak, melyek sok esetben nem feltétlen szükségesek a vizsgálat elvégzéséhez. Ezért az adatoknak előzetes átalakításon kell átesniük, mely folyamán csak **azon esemény tulajdonságokat tartjuk meg, melyek számunkra szükségesek.**
- Bevezetünk négy különböző mérőszámot:
 - 1. p** létrehozott token (produced token),
 - 2. c** elhasznált token (consumed token),
 - 3. m** hiányzó token (missing token),
 - 4. r** megmaradt token (remaining token).

Kiszámolhatjuk a fitnessz értékét trace szinten

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

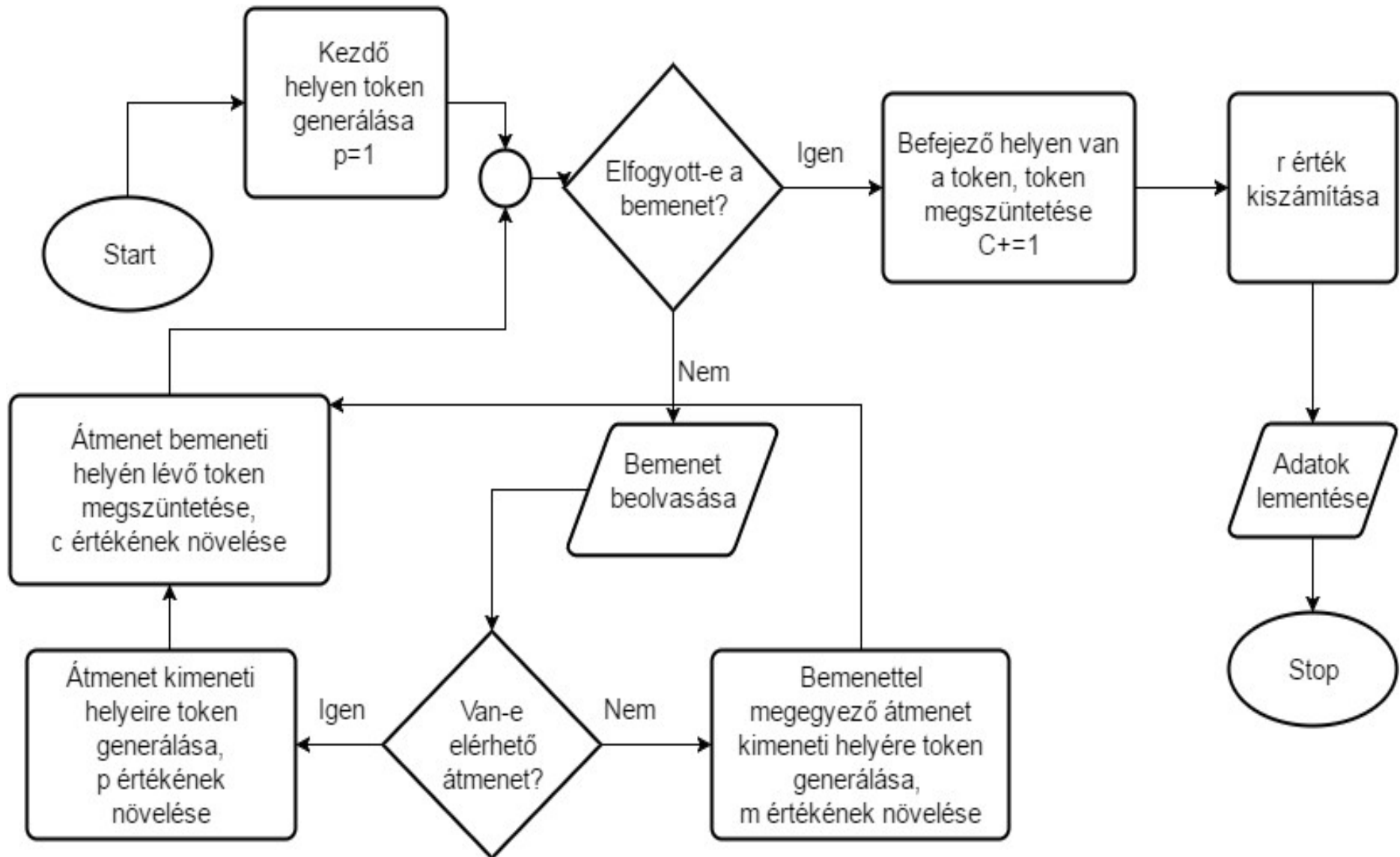
p = produced tokens = 6

c = consumed tokens = 6

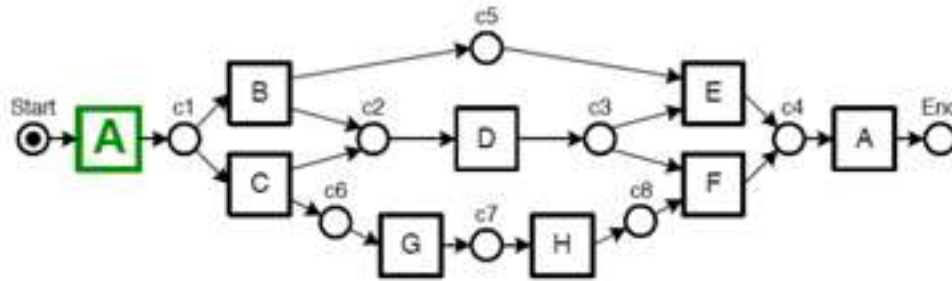
m = missing tokens = 1

r = remaining tokens = 1

Konformancia token alapú egyezőségvizsgálatának folyamatábrája

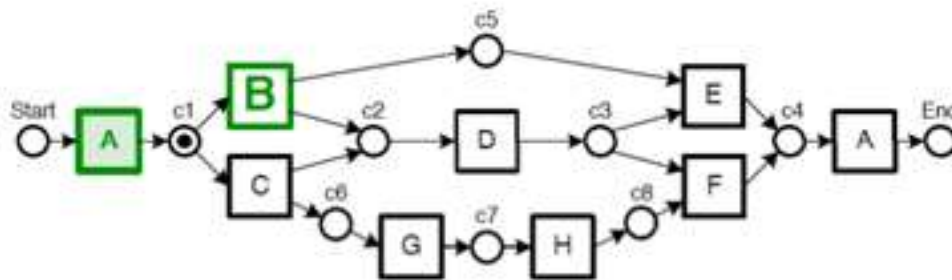


Fitnesz érték figyelése az események végig követésével egy trace-ben



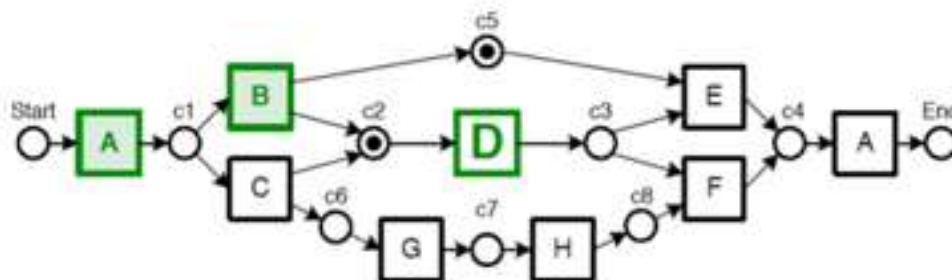
$m = 0$
 $r = 0$
 $c = 0$
 $p = 1$

No. of Instances	Log Traces
1207	→ A BDEA
145	ACDGHFA
56	ACGDHFA
23	ACHDFA
28	ACDHFA



$m = 0$
 $r = 0$
 $c = 1$
 $p = 2$

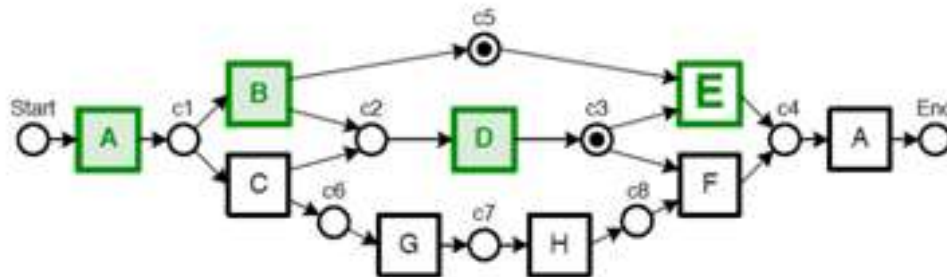
No. of Instances	Log Traces
1207	→ A B DEA
145	ACDGHFA
56	ACGDHFA
23	ACHDFA
28	ACDHFA



$m = 0$
 $r = 0$
 $c = 2$
 $p = 4$

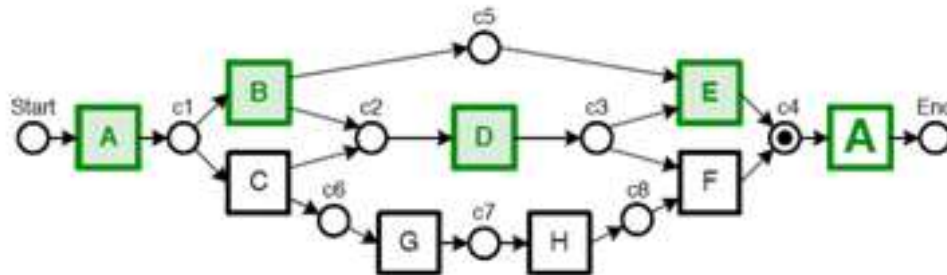
No. of Instances	Log Traces
1207	→ A B D EA
145	ACDGHFA
56	ACGDHFA
23	ACHDFA
28	ACDHFA

Ha nincs probléma ($m=0, r=0$):



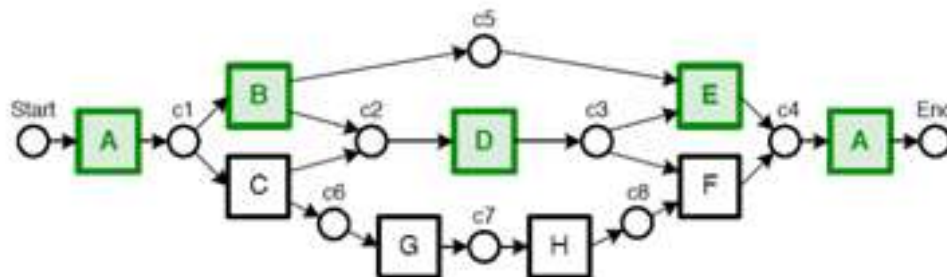
$m = 0$
 $r = 0$
 $c = 3$
 $p = 5$

No. of Instances	Log Traces
1207	→ ABDEA
145	ACDGHFA
56	ACGDHFA
23	ACHDFA
28	ACDHFA



$m = 0$
 $r = 0$
 $c = 5$
 $p = 6$

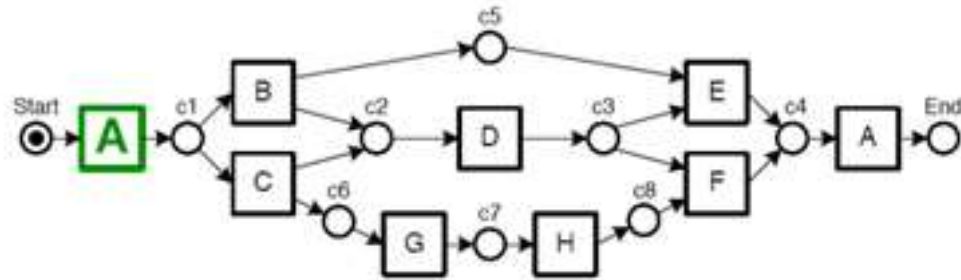
No. of Instances	Log Traces
1207	→ ABDEA
145	ACDGHFA
56	ACGDHFA
23	ACHDFA
28	ACDHFA



$m = 0$
 $r = 0$
 $c = 7$
 $p = 7$

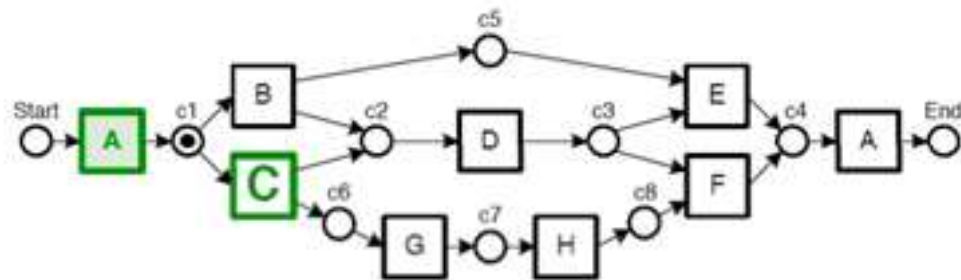
No. of Instances	Log Traces
1207	→ ABDEA
145	ACDGHFA
56	ACGDHFA
23	ACHDFA
28	ACDHFA

Egy másik (megvalósíthatatlan) trace



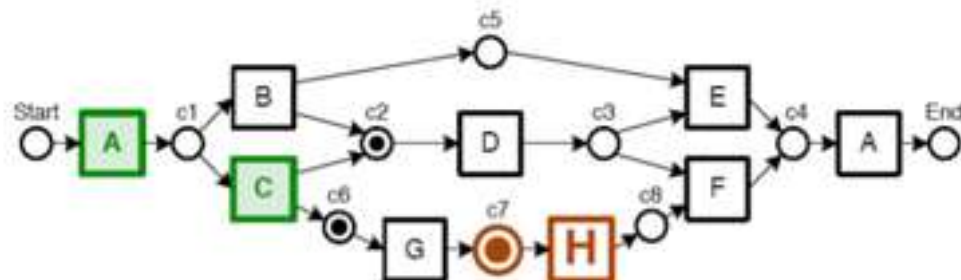
$m = 0$
 $r = 0$
 $c = 0$
 $p = 1$

No. of Instances	Log Traces
1207	ABDEA
145	ACDGHFA
56	ACGDHFA
23	→ACHDFA
28	ACDHFA



$m = 0$
 $r = 0$
 $c = 1$
 $p = 2$

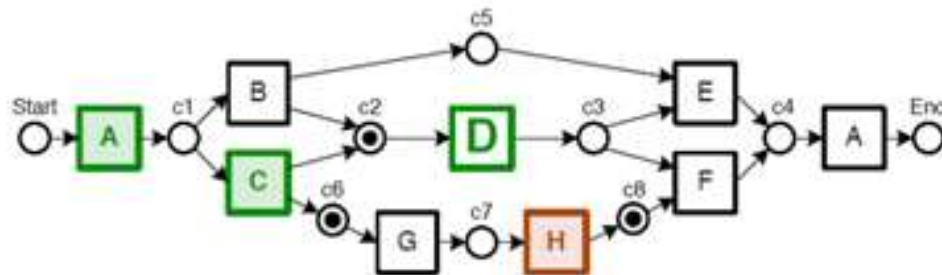
No. of Instances	Log Traces
1207	ABDEA
145	ACDGHFA
56	ACGDHFA
23	→ACHDFA
28	ACDHFA



$m = 1$
 $r = 0$
 $c = 2$
 $p = 4$

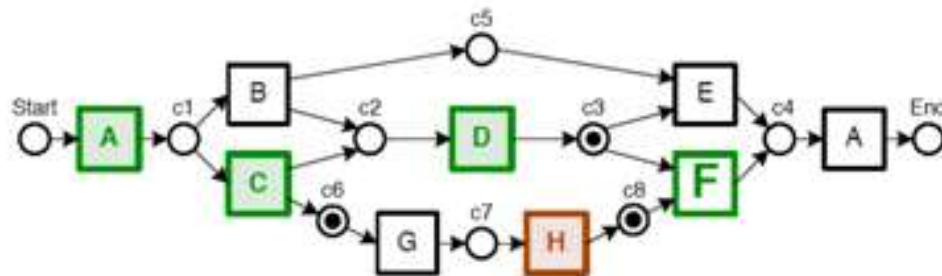
No. of Instances	Log Traces
1207	ABDEA
145	ACDGHFA
56	ACGDHFA
23	→ACHDFA
28	ACDHFA

Folytatás



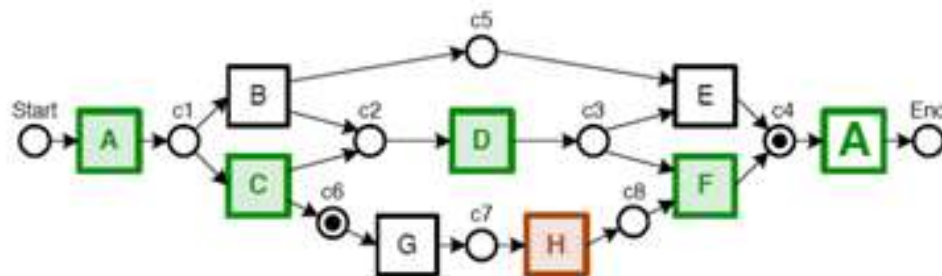
$m = 1$
 $r = 0$
 $c = 3$
 $p = 5$

No. of Instances	Log Traces
1207	ABDEA
145	ACDGHFA
56	ACGDHFA
23	→ ACHDFA
28	ACDHFA



$m = 1$
 $r = 0$
 $c = 4$
 $p = 6$

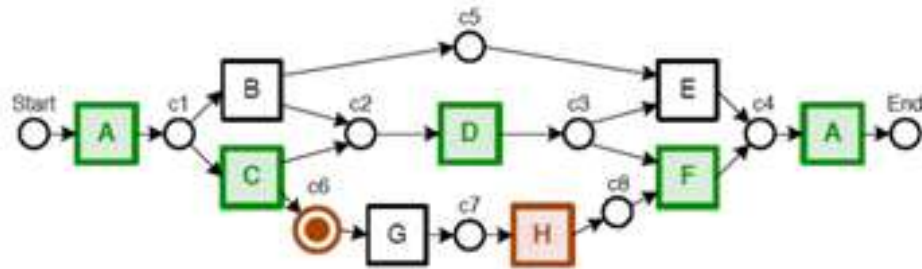
No. of Instances	Log Traces
1207	ABDEA
145	ACDGHFA
56	ACGDHFA
23	→ ACHDFFA
28	ACDHFA



$m = 1$
 $r = 0$
 $c = 6$
 $p = 7$

No. of Instances	Log Traces
1207	ABDEA
145	ACDGHFA
56	ACGDHFA
23	→ ACHDFFA
28	ACDHFA

Fitnesz érték számolása



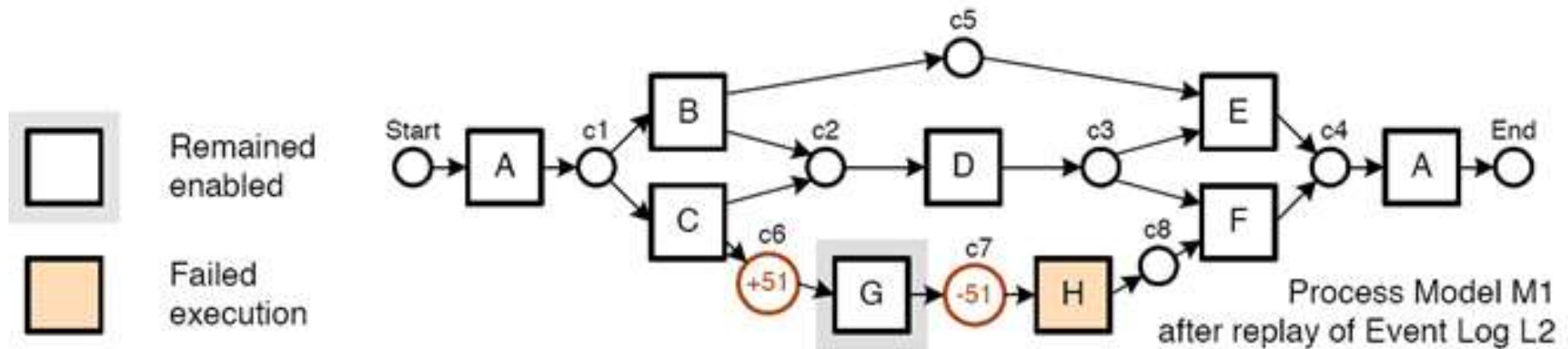
$m = 1$
 $r = 1$
 $c = 8$
 $p = 8$

No. of Instances	Log Traces
1207	ABDEA
145	ACDGHFA
56	ACGDHFA
23	→ ACHDFA
28	ACDHFA

$$f = \frac{1}{2} \left(1 - \frac{\sum_{i=1}^k n_i m_i}{\sum_{i=1}^k n_i c_i} \right) + \frac{1}{2} \left(1 - \frac{\sum_{i=1}^k n_i r_i}{\sum_{i=1}^k n_i p_i} \right)$$

$$f(M1, L2) = \frac{1}{2} \left(1 - \frac{51}{10666} \right) + \frac{1}{2} \left(1 - \frac{51}{10666} \right) \approx 0.995$$

Hogyan használhatjuk a módszert diagnosztikához

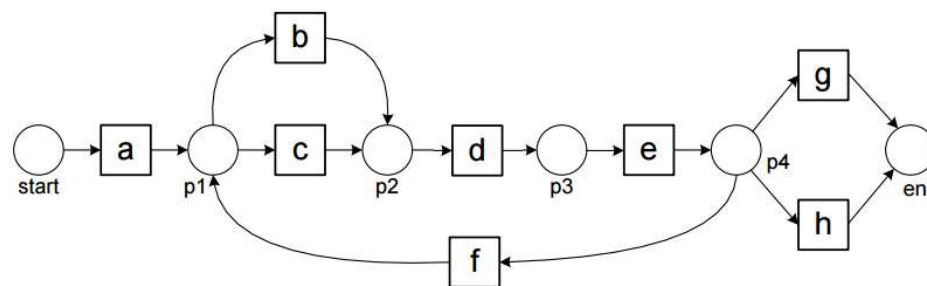


Másik példa

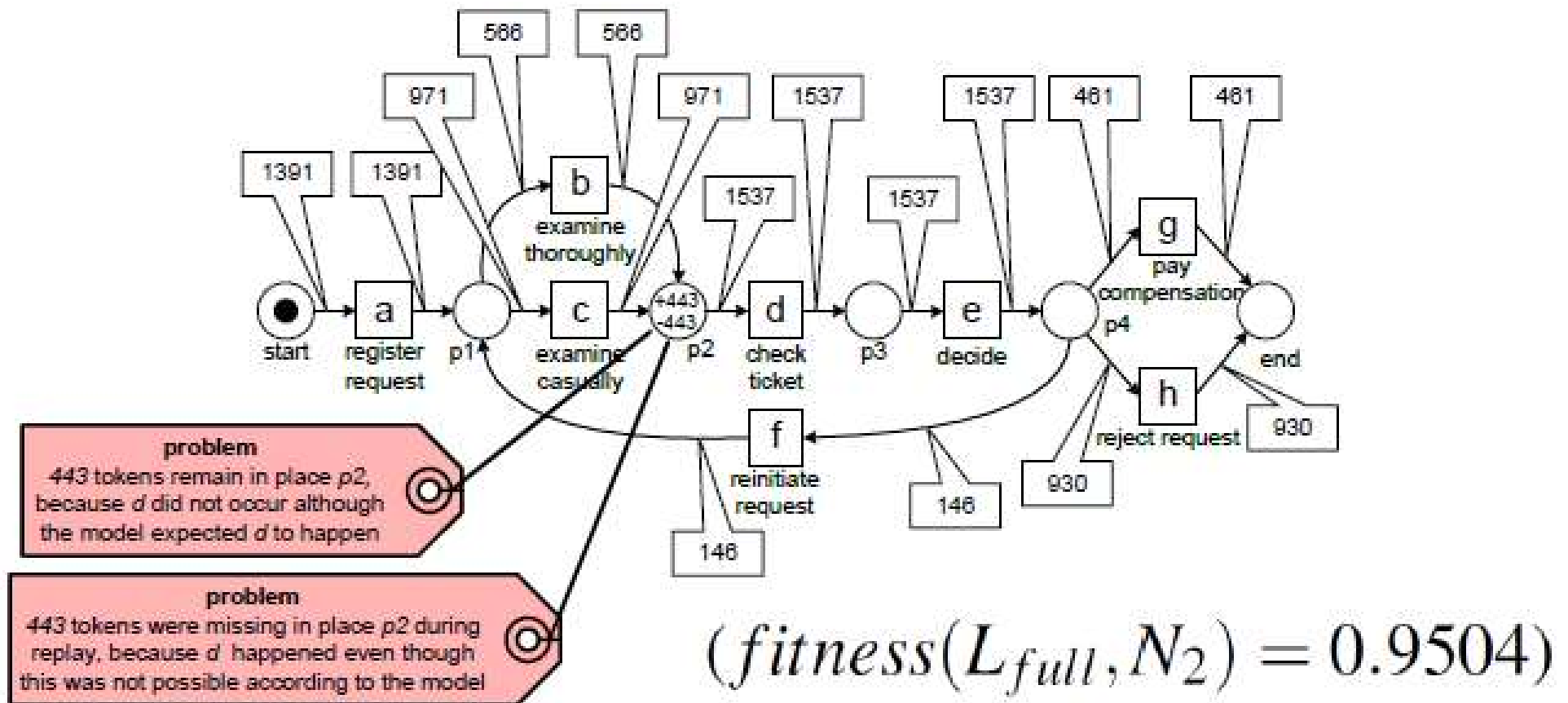
- A napló fájl átalakítása után csak azon esemény tulajdonságokat tartjuk meg melyek számunkra szükségesek.
- Ezután előfordulási gyakoriság szerint összerendezzük őket.
- A vizsgálat elvégzéséhez szükséges adatok maradnak meg: az azonosító, a gyakoriság és az útvonal

455	σ_1	$\langle a, c, d, e, h \rangle$
191	σ_2	$\langle a, b, d, e, g \rangle$
177	σ_3	$\langle a, d, c, e, h \rangle$
144	σ_4	$\langle a, b, d, e, h \rangle$
111	σ_5	$\langle a, c, d, e, g \rangle$
82	σ_6	$\langle a, d, c, e, g \rangle$
56	σ_7	$\langle a, d, b, e, h \rangle$
47	σ_8	$\langle a, c, d, e, f, d, b, e, h \rangle$
38	σ_9	$\langle a, d, b, e, g \rangle$
33	σ_{10}	$\langle a, c, d, e, f, b, d, e, h \rangle$
14	σ_{11}	$\langle a, c, d, e, f, b, d, e, g \rangle$
11	σ_{12}	$\langle a, c, d, e, f, d, b, e, g \rangle$
9	σ_{13}	$\langle a, d, c, e, f, c, d, e, h \rangle$

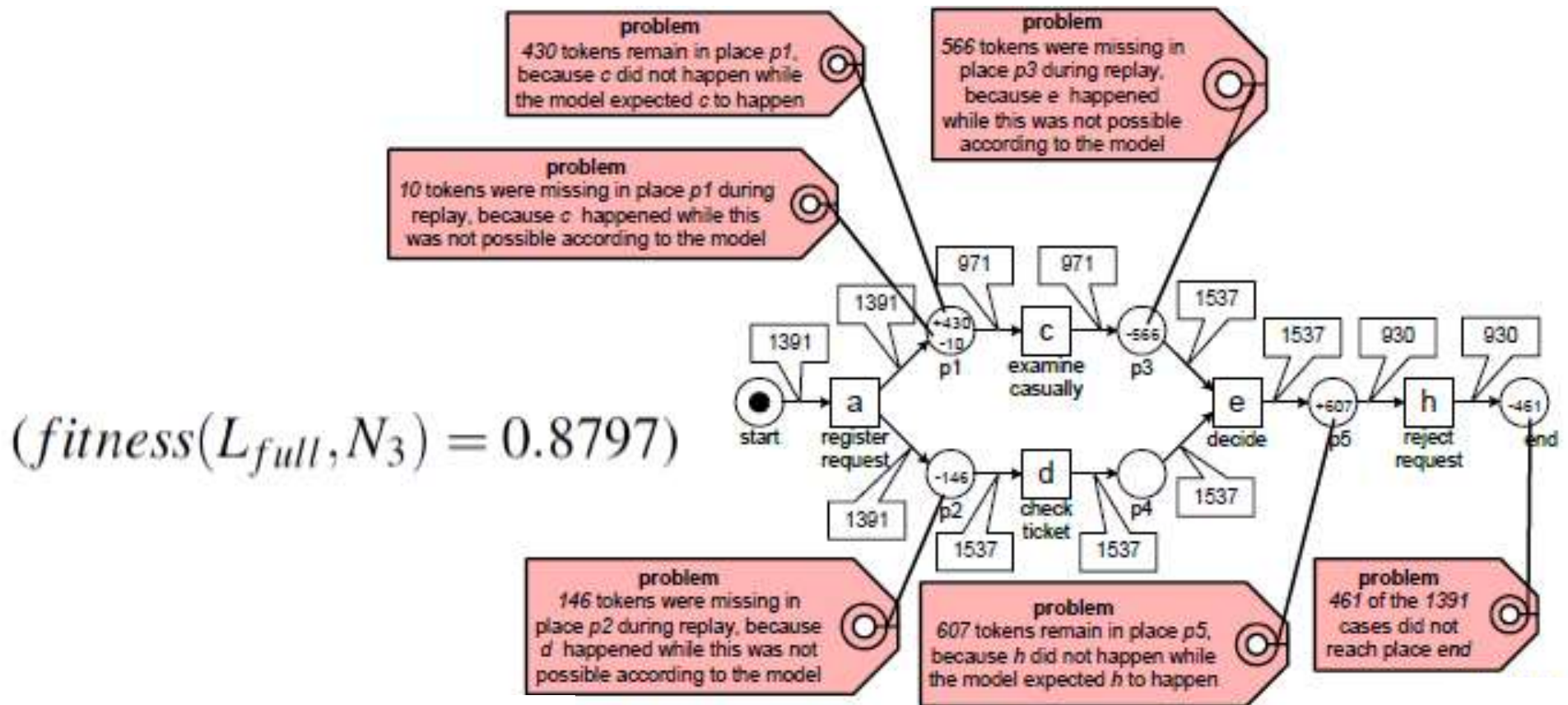
Elemezzük a megjelölt eseménysorozatot a rendelkezésre álló modell alapján!



A módszer alkalmazása modell diagnosztikához

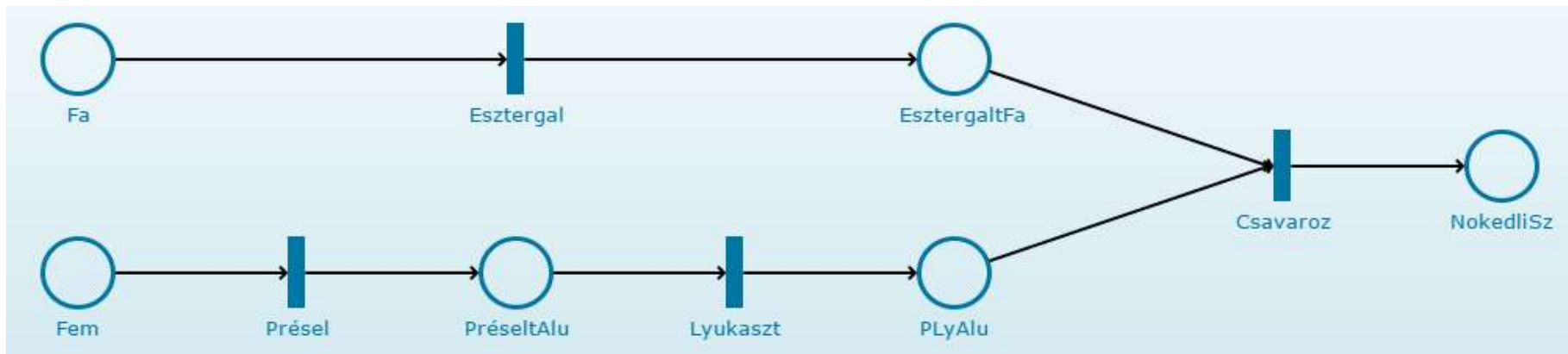


A módszer alkalmazása modell diagnosztikához



A konformancia elemzés alternatív megközelítései

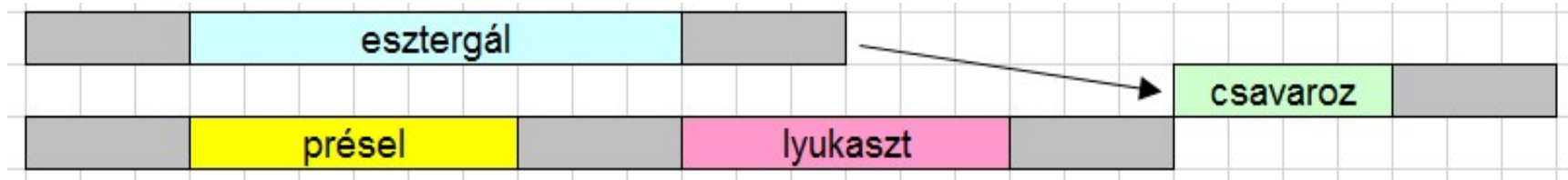
- A log fájlok feladata az esemény logok tárolása. Ezt nagyon sok formátumban megtehetjük (XML, CSV, JSON, stb.). Vegyük pl. JSON formátumot (JavaScript Object Notation).
- Tekintsük az alábbi gyártási folyamatot:



- A folyamat lényege, hogy mindkét részfolyamatnak el kell készülnie mielőtt összecsavarozásra kerülne a termék.

Másfajta szemléltetése a folyamatnak

- A Gantt-diagramon a szürke részek az egyes műveletek közötti szállítást jelölik, ez a Petri-hálóba nem jelenik meg.



Petri háló kiértékelése

Az algoritmus a készterméktől indulva számolja ki, hogy az egyes félkész nyersanyagoknak a kezdéstől számítva mennyi idő alatt kell kész lenniük.

helyek = petri-háló helyei kimenő és bejövő átmenetek listájával

eljárás számoldó(hely):

idő = 0;

átmenet = hely bejövő átmenete;

maxldő = 0;

ciklus p = átmenet bejövő helyei:

templdő = számoldó(p);

ha templdő nagyobb maxldő:

maxldő = templdő;

feltétel vége

ciklus vége

idő = maxldő + átmenet.idő;

hely.idő = idő;

visszatér idő;

eljárás vége

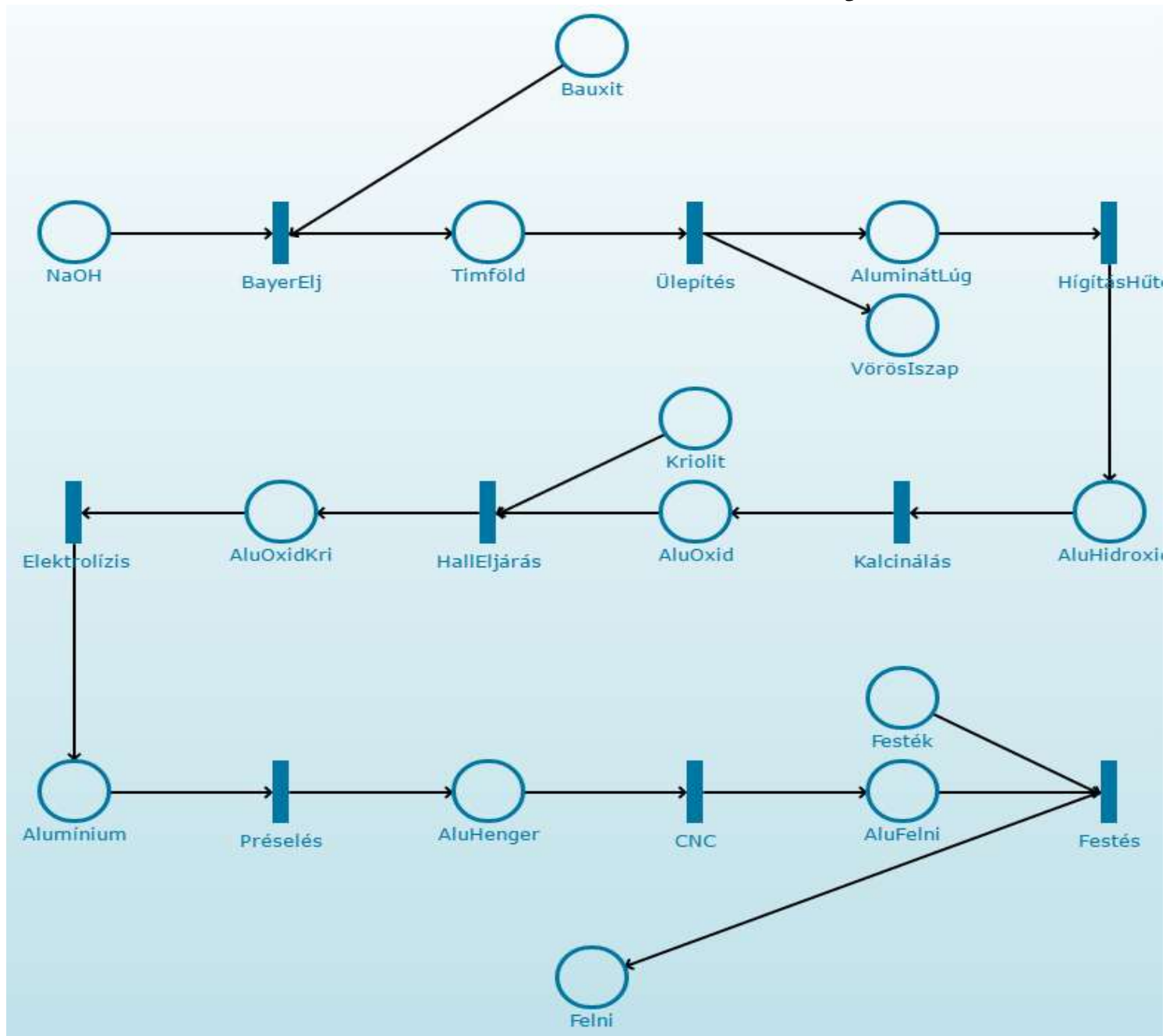
A célállapotból kiindulva rekurzívan kiszámoljuk az addig eltelt időt.

Amennyiben egy átmenetbe több helyről jön él (pl.: csavaroz), akkor a bejövő élekből a legtöbb időt igénylőt választjuk, mert nem lehet elindítani a műveletet, amíg a legtöbb időt igénylő előfeltétel nem készült el.

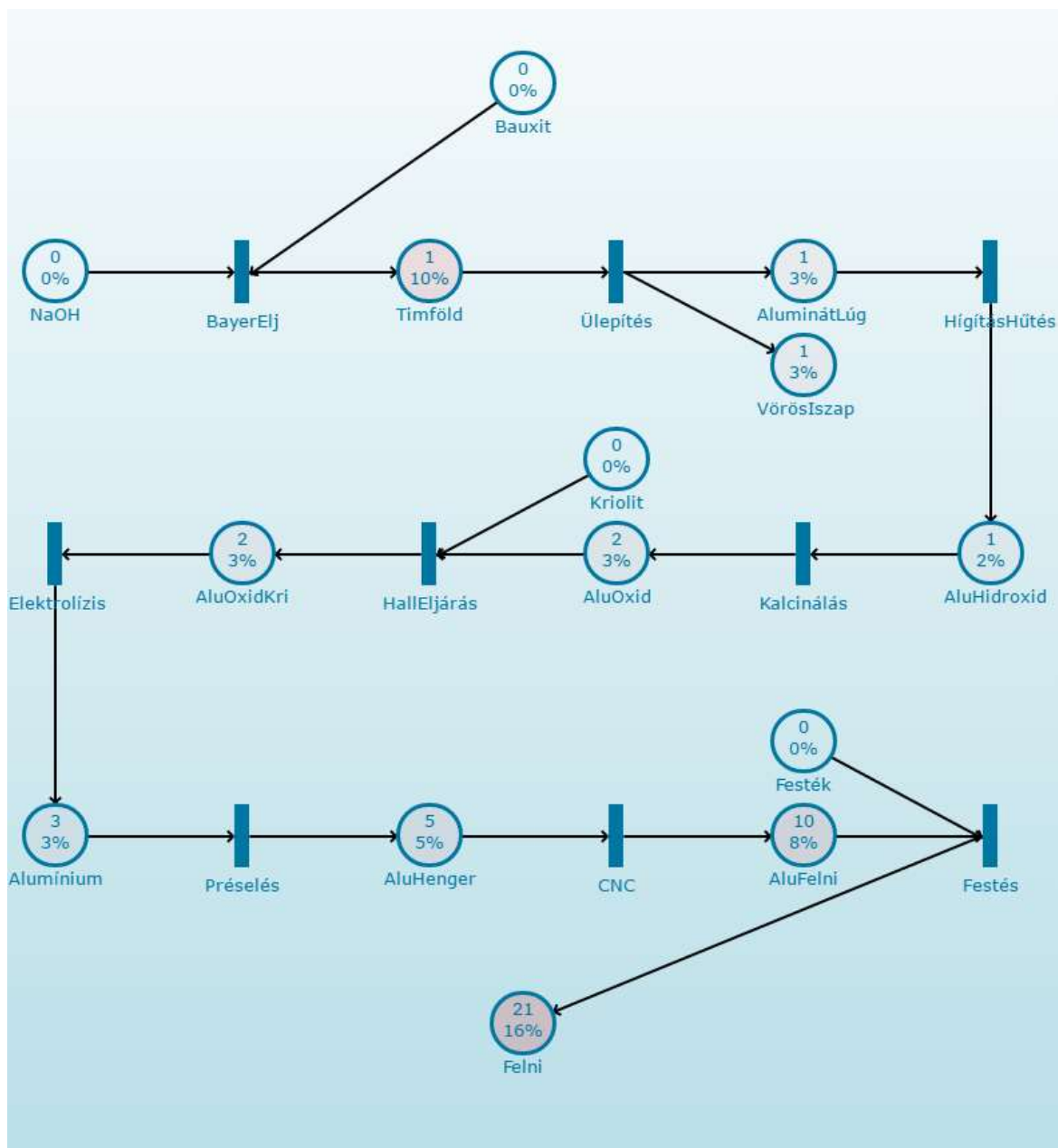
Eltérések keresése

- Időbeli eltéréseket keresünk jelen állapotban (vizsgálható lehet még a költség vagy az erőforrások).
- Az eltérések kereséséhez egy log fájlt kell feldolgoznunk. A fájl az egyes műveletek kezdeti és befejezési idejét tartalmazza.
- Ezeket az időket megfeleltetjük a hálóban az átmenet előtti és az azt követő hely idejével, majd letranszformáljuk úgy, hogy a folyamat nullával kezdődjön.
- Például ha a log fájlban az esztergálás 2017-01-13 11:39:32-től 2017-01-13 11:39:56-ig tart, akkor a transzformálás után a Fa hely 0 időértéket, az EsztergáltFa hely pedig 24 értéket kap.
- Ha ezt az egész háló minden helyére elvégeztük, akkor hasonló adatszerkezetet kapunk mint a Petri-háló feldolgozásakor kaptunk. Ezen két adatszerkezetet összehasonlítva megkapjuk, hogy a folyamat egyes állomásai időben hogy viszonyulnak a modellhez.
- A rendszer ezeket az eltéréseket színek segítségével jelölheti a Petri-hálón, ezzel vizualizálva, hogy hol vannak komolyabb időbeli eltérések, illetve hol sikerült lefaragni a modell szerinti időből.

Alumínium felni készítési modellje



Az elemzés eredménye



- Az egyes helyeken megjelennek az időbeli eltérések értékei és százalékos arányuk.
- Bizonyos erősségű vörös háttérrel vannak a helyek a Petri-hálóban jelölve.