



ВЫСШАЯ ШКОЛА ЭКОНОМИКИ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ

Международная научно-учебная лаборатория
процессно-ориентированных
информационных систем



Обобщенный подход к декомпозиции сетей Петри для задач Process Mining

Сергей Шершаков

Москва, 2013

План

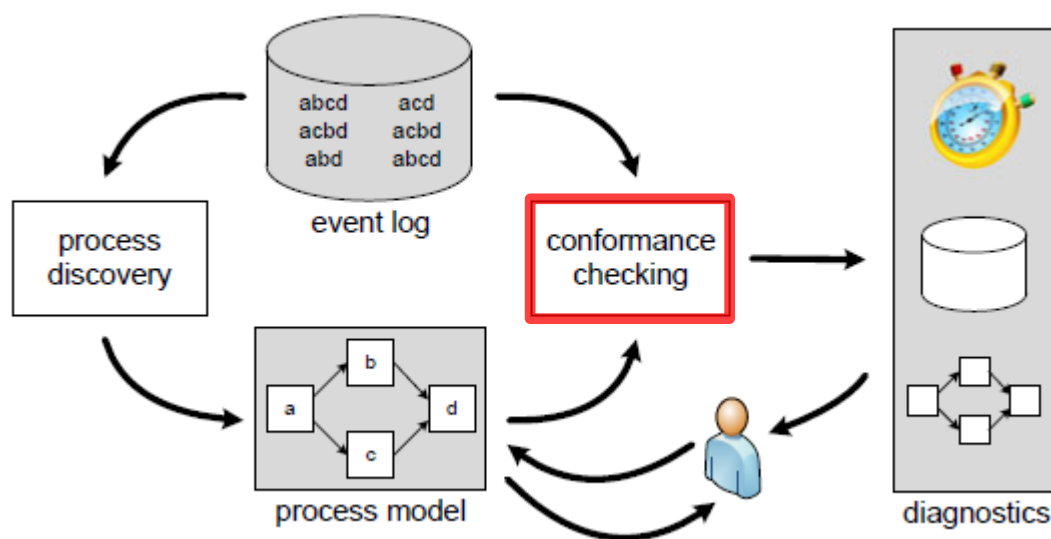
- Цели и задачи
- Лог событий и модель данных
- Проверка соответствия лога и модели
- Задача декомпозиции
- Заключение

Цель

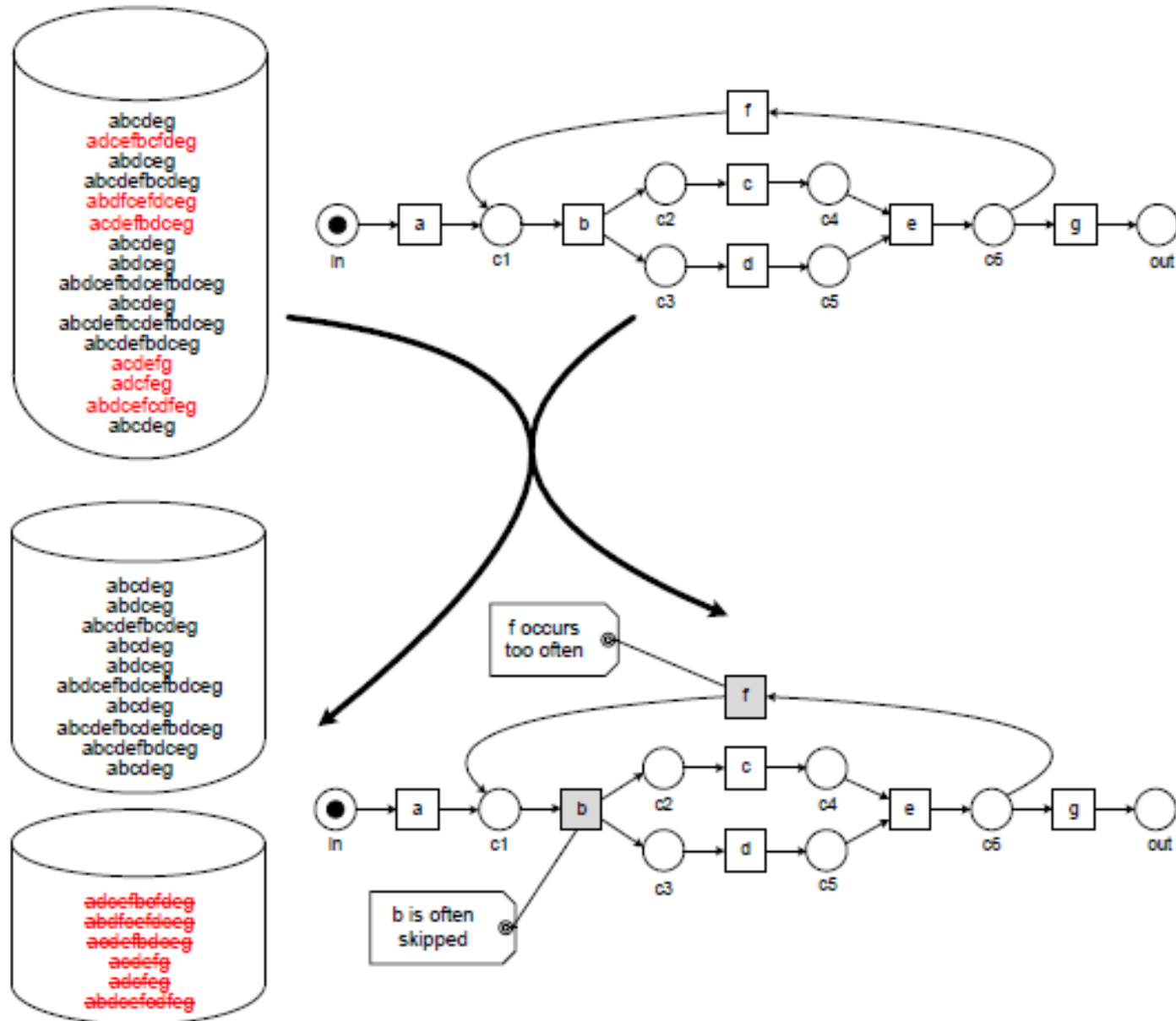
- Декомпозиция задачи Conformance Checking
 - Распределенные вычисления
 - Уменьшение времени выполнения отдельных подзадачи

Задачи Process Mining

- Process Discovery — построение модели по имеющемуся логу
- **Conformance Checking** — проверка согласованности лога и модели



Задача Conformance Checking (CC)



Критерии соответствия лога и модели

- *fitness* — пригодность, соответствие
- *simplicity* — простота
- *precision* — точность
- *generalization* — обобщенность

Способы декомпозиции задач РМ

- *replication* — репликация
- *vertical distribution* — вертикальное разделение
- ***horizontal distribution*** — горизонтальное разделение

ЛОГ СОБЫТИЙ И МОДЕЛЬ ДАННЫХ

Мультимножества

- Множество:
 - $A = \{a, b, c\}$
- Мультимножество:
 - $b_1 = []$, $b_2 = [a, a, b]$, $b_3 = [a, b, c]$,
 - $b_4 = [a, a, b, a, b, c]$, $b_5 = [a^3, b^2, c]$
- Множество всех множеств над A :
 - $\mathcal{B}(A)$

Последовательность

- Последовательность:

$$\sigma = \langle a_1, a_2, \dots, a_n \rangle \in A^*$$

- Пустая последовательность:

$$\langle \rangle$$

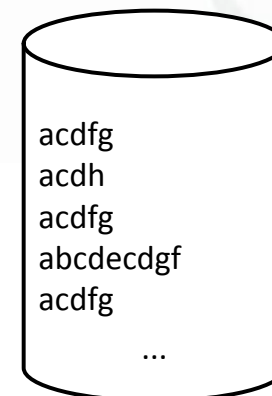
- Конкатенация двух последовательностей:

$$\sigma \cdot \sigma$$

Лог данных

- Множество активностей:
 - $A \subseteq \mathcal{U}_A$
- *Трасса* — последовательность активностей:
 - $\sigma \in A^*$
- *Лог событий* — мультимножество трасс
 - $L \in \mathcal{B}(A^*)$

$$L_1 = [\langle a, c, d, f, g \rangle^{10}, \\ \langle a, c, d, h \rangle^5, \\ \langle a, b, c, d, e, c, d, g, f \rangle^5]$$



Декартово произведение

- Декартово произведение — множество пар:
 - $R \subset X \times Y$
- Область определения:
 - $dom(R) = \{x \mid (x, y) \in R\}$
- Область значения:
 - $rgn(R) = \{y \mid (x, y) \in R\}$
- Элементы
 - $\omega(R) = dom(R) \cup rgn(R)$

$X \backslash Y$	b	c
	a	b
a	ab	ac
b	bb	bc

Функция

- Частичная функция
 - $f \in X \rightharpoonup Y$
 - $\text{dom}(f) \subseteq X$
 - $\text{rng}(f) = \{f(x) \mid x \in X\}$
- Тотальная функция
 - $f \in X \rightarrow Y$
 - $\text{dom}(f) = X$

Применение функции к последовательности

- Частичная функция

- $f \in X \rightarrow Y$

может быть применена к последовательности:

$$f(\langle \rangle) = \langle \rangle$$

$$f(\langle x \rangle \cdot \sigma) = \begin{cases} f(\sigma), & x \notin \text{dom}(f) \\ \langle f(x) \rangle \cdot f(s) & x \in \text{dom}(f) \end{cases}$$

- Пример:

- $f(\langle a, b, c \rangle) = \langle f(a), f(b), f(c) \rangle$

Проекция функции

- Проекция функции $f \in X \rightarrow Y$ на множество $Q \subseteq X$:
 - $f \upharpoonright_Q$
 - $\text{dom}(f \upharpoonright_Q) = \text{dom}(f) \cap Q$
 - $f \upharpoonright_Q(x) = f(x), \forall x \in \text{dom}(f \upharpoonright_Q)$
- Пример:
 - $[x^3, y, z^2] \upharpoonright_{\{x, y\}} = [x^3, y]$

Проекция последовательности

- Для подмножества $Q \subseteq X$ множества X можно определить функцию

- $\upharpoonright_Q \in X^* \rightarrow Q^*$

проекции последовательности из X в последовательность из Q :

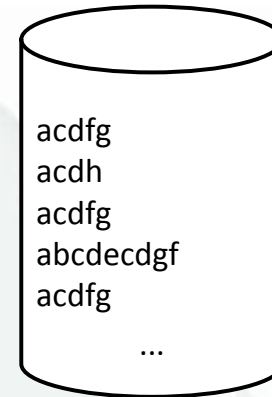
$$\langle \rangle \upharpoonright_Q = \langle \rangle$$

$$(\langle x \rangle \cdot \sigma) \upharpoonright_Q = \begin{cases} \sigma \upharpoonright_Q, & x \notin Q \\ \langle x \rangle \cdot \sigma \upharpoonright_Q & x \in Q \end{cases}$$

- Пример: $\langle y, z, y \rangle \upharpoonright_{\{x, y\}} = \langle y, y \rangle$

Лог данных: горизонтальное разделение

$$L_1 = [\langle a, c, d, f, g \rangle^{10}, \\ \langle a, c, d, h \rangle^5, \\ \langle a, b, c, d, e, c, d, g, f \rangle^5]$$



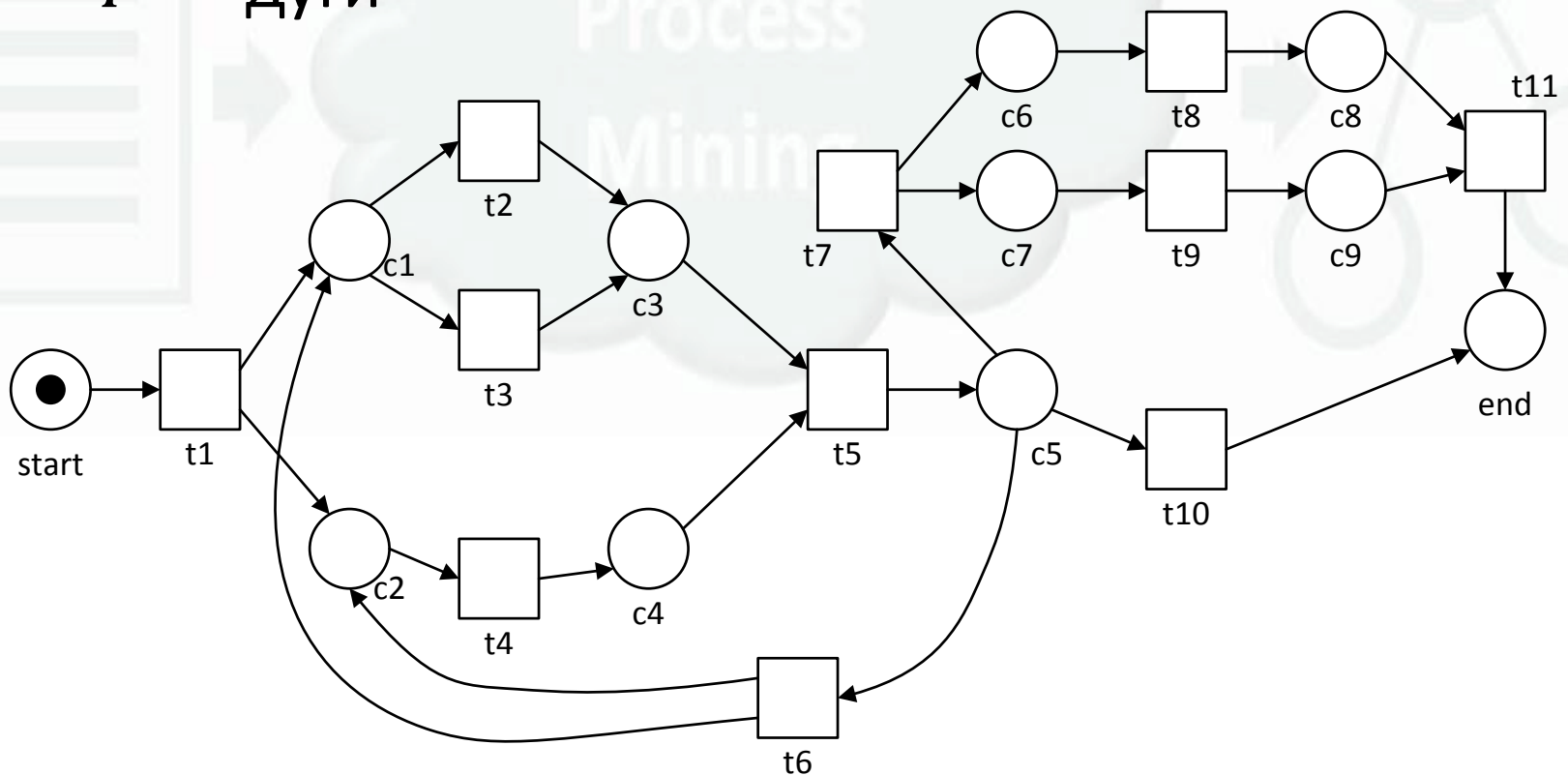
$$L_1^1 = L_1 \upharpoonright_{\{a\}} = [\langle a \rangle^{20}]$$

$$L_1^2 = L_1 \upharpoonright_{\{a, b, d, e\}} = [\langle a, d \rangle^{15}, \langle a, b, d, e, d \rangle^5]$$

$$L_1^3 = L_1 \upharpoonright_{\{a, c, e\}} = [\langle a, c \rangle^{15}, \langle a, c, e, c \rangle^5]$$

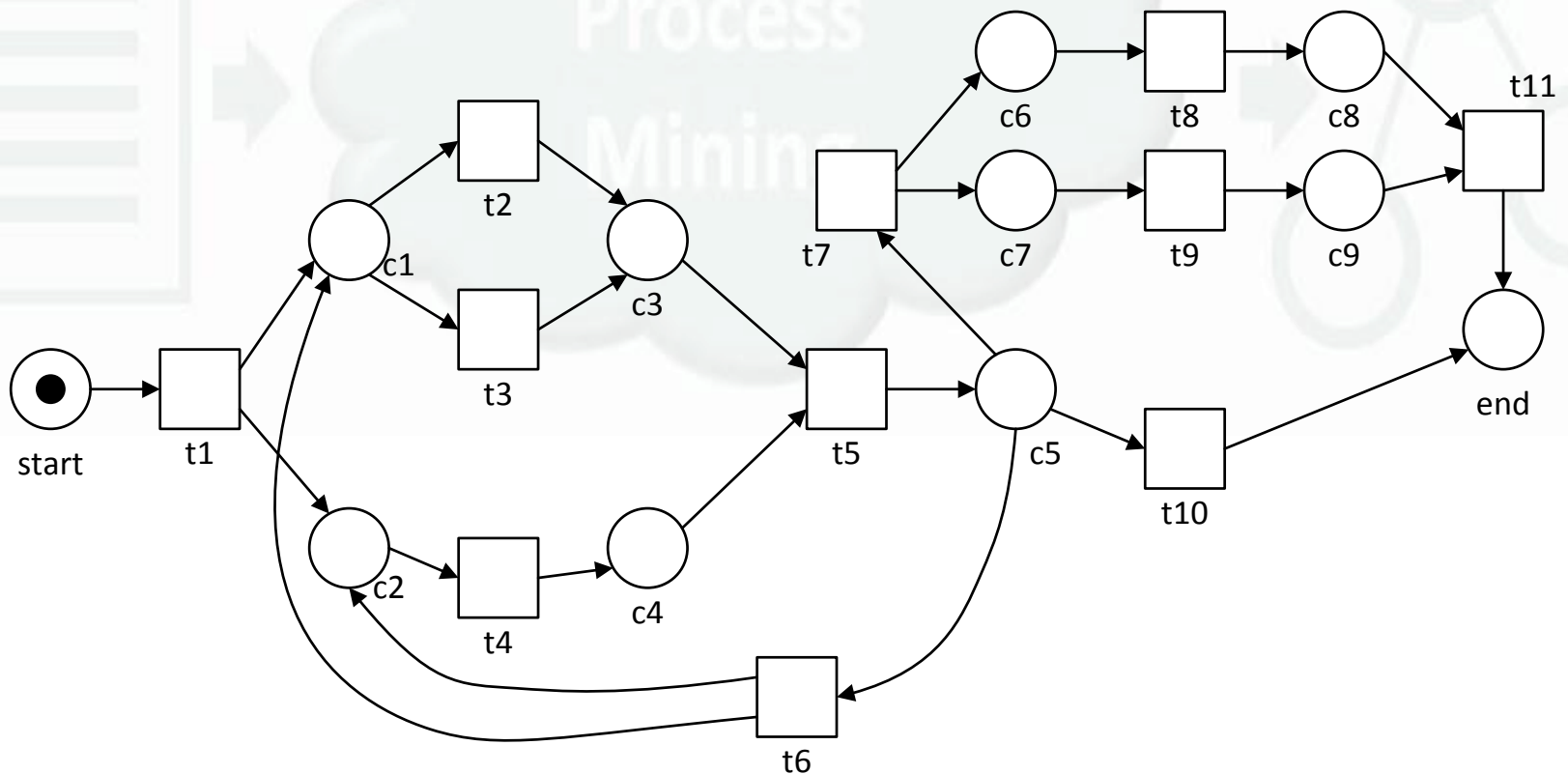
Сеть Петри

- Сеть Петри это тройка $N = (P, T, F)$:
 - P — узлы
 - T — переходы
 - F — дуги
 - $P \cap T = \emptyset$
 - $F \subseteq (P \times T) \cup (T \times P)$



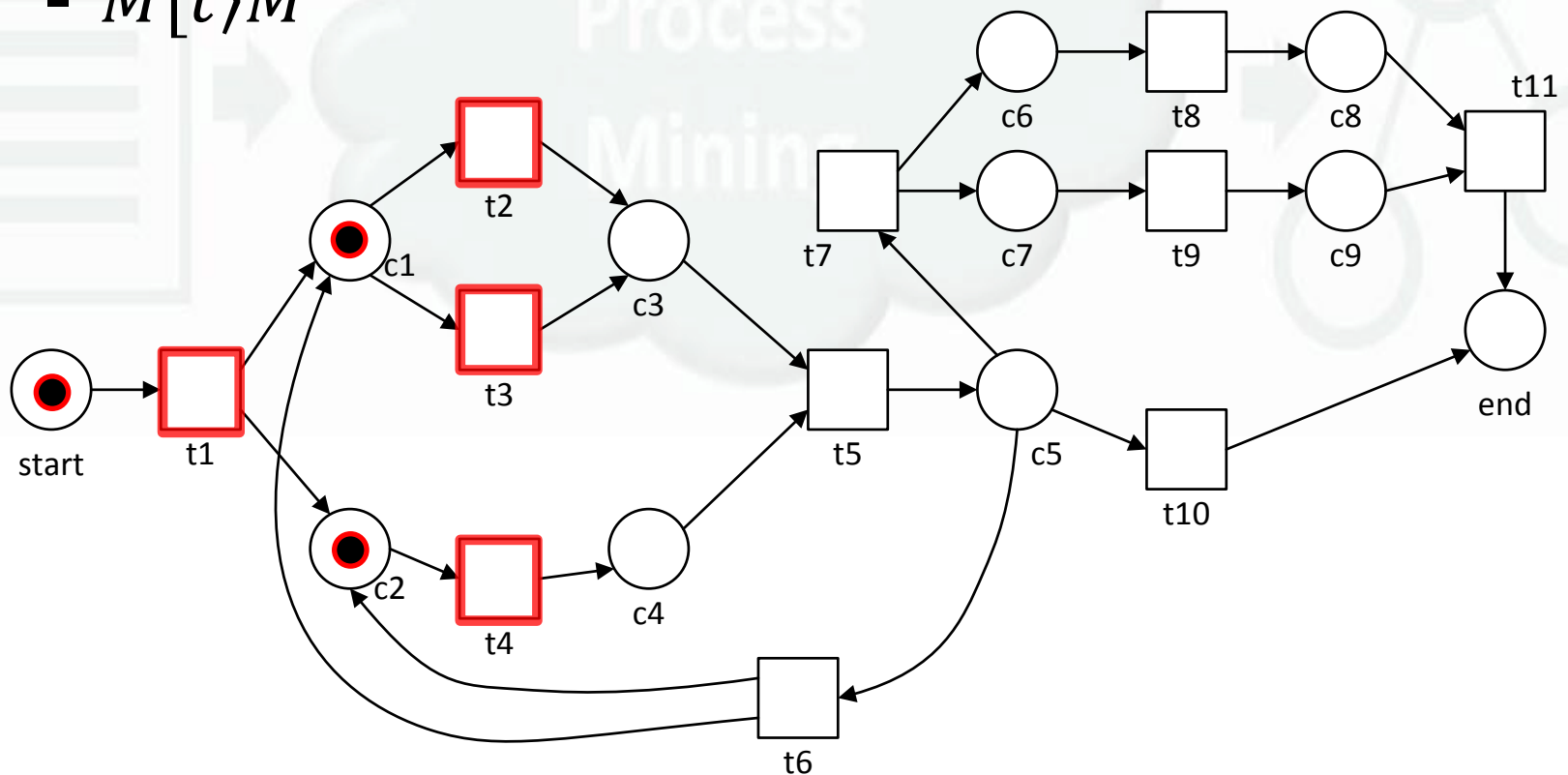
Сеть Петри: разметка

- Разметка — мультимножество позиций:
 - $M \in \mathcal{B}(A)$



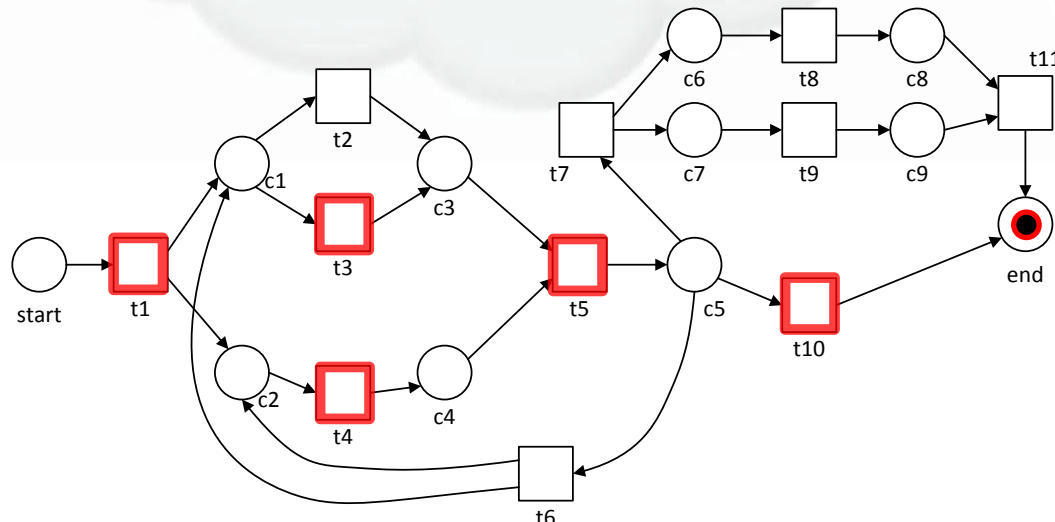
Сеть Петри: разрешенный переход

- Переход $t \in T$ разрешен для разметки M :
 - $M[t\rangle$
- Переход «срабатывает»
 - $M[t\rangle M'$



Сеть Петри: последовательность переходов

- Пусть $\sigma = \langle t_1, t_2, \dots, t_n \rangle \in T^*$ — последовательность переходов. Тогда
- $M[\sigma]M'$ означает:
 - $\exists(M_0, M_1, \dots, M_n)$:
 - $M_0 = M, M_n = M'$
 - $\forall i: 0 \leq i < n: M_i[t_{i+1}]M'_{i+1}$



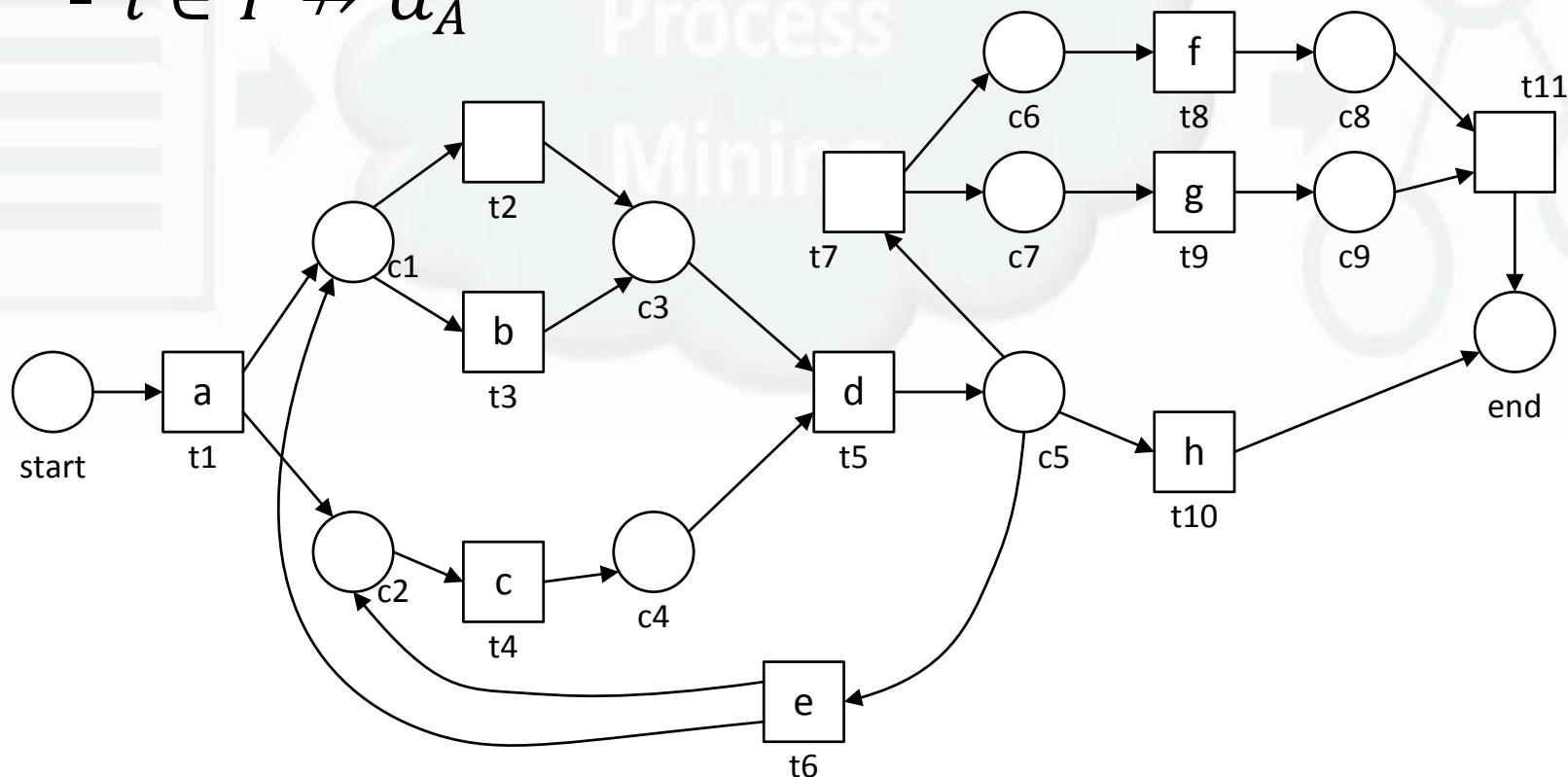
Помеченная сеть Петри (labeled Petri net)

- Помеченная сеть

$$N = (P, T, F, l):$$

- $l \in T \rightarrow \mathcal{U}_A$

a = регистрация запроса
b = проверка файла
c = проверка квитанции
d = принятие решения
e = перезапрос
f = послать подтв. письмо
g = выплата компенсации
h = послать отказ. письмо

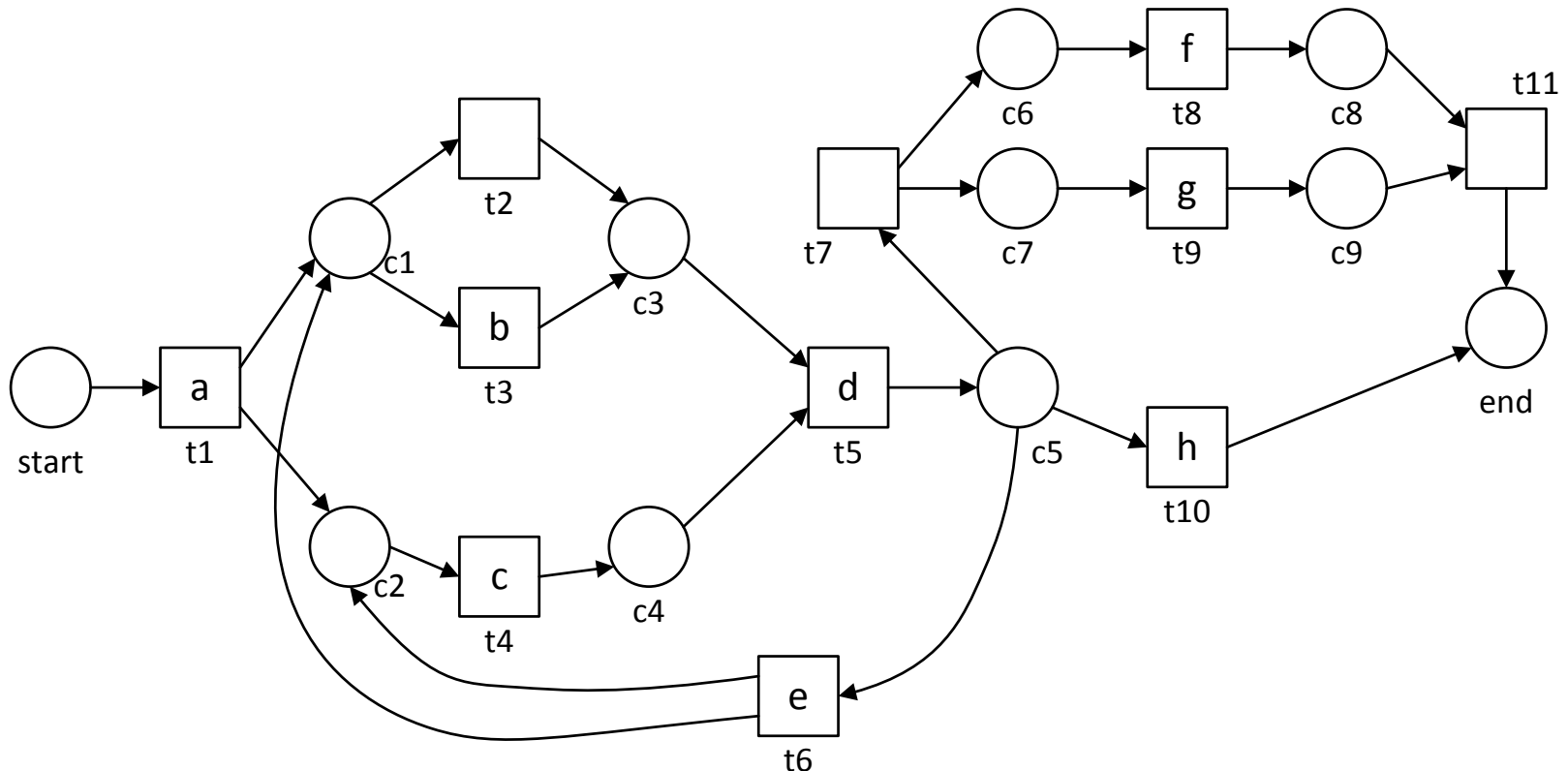


Маршрут из активностей

- Пусть $N = (P, T, F, l)$ — помеченная сеть,
- $\sigma_v = \langle a_1, a_2, \dots, a_n \rangle \in \mathcal{U}_A$ — последовательность активностей, тогда
- $M[\sigma_v \triangleright M'$ — последовательность переходов по активностям, если:
- $\exists \sigma \in T^*: M[\sigma \rangle M', l(\sigma) = \sigma_v$

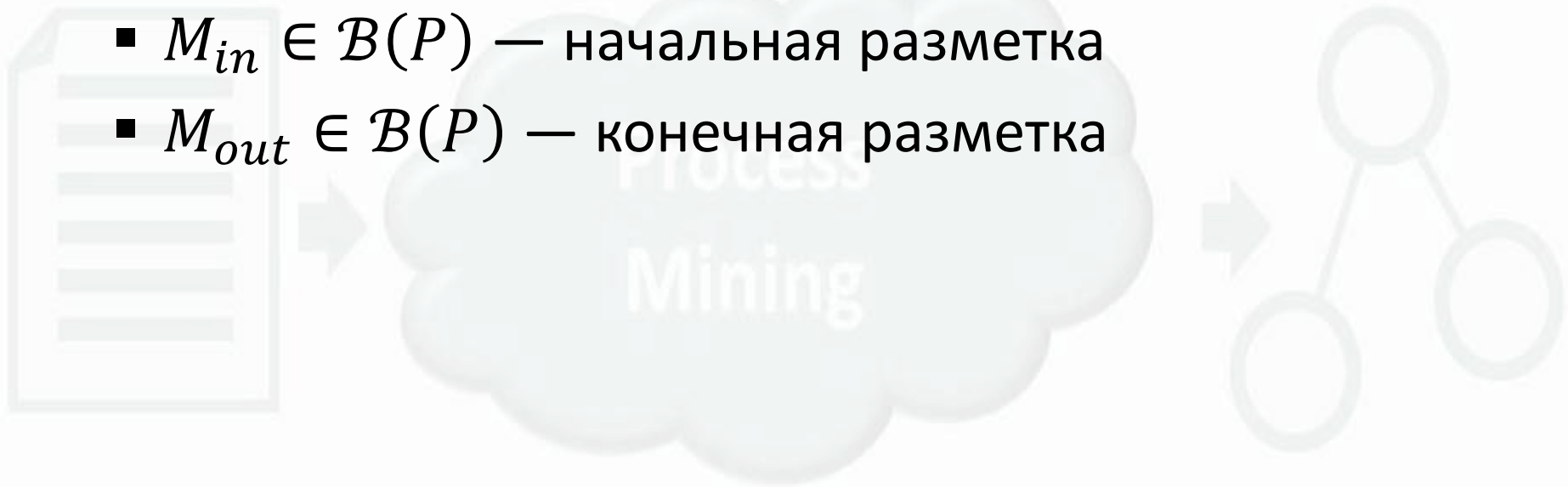
Маршрут из активностей: пример

- $\sigma_v = \langle a, c, d, f, g \rangle$
- $\sigma = \langle t1, t2, t4, t5, t7, t8, t9, t11 \rangle$
- $([start])(\sigma_v \triangleright ([end]))$



Сеть Петри: системная сеть

- Системная сеть — тройка $SN = (N, M_{in}, M_{out})$:
 - $N = (P, T, F, l)$ — сеть Петри
 - $M_{in} \in \mathcal{B}(P)$ — начальная разметка
 - $M_{out} \in \mathcal{B}(P)$ — конечная разметка

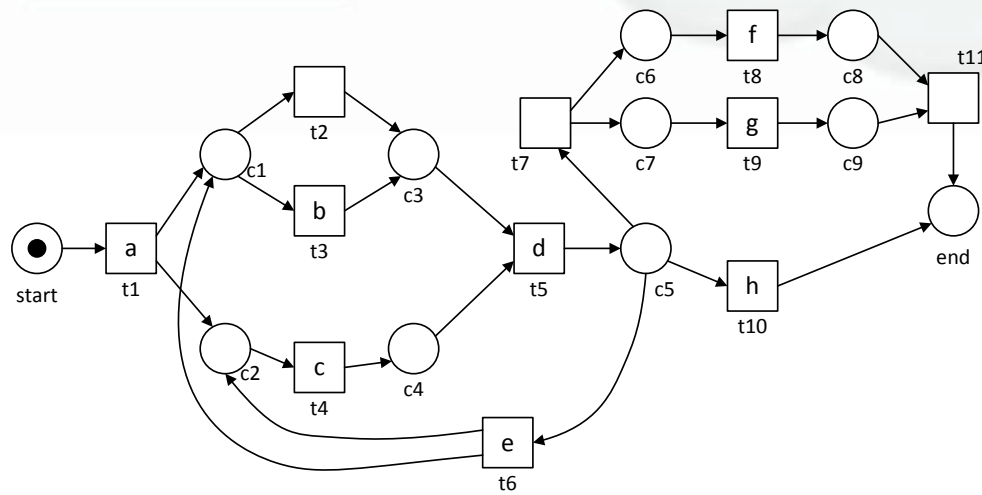


Системная сеть: элементы

- $SN = (N, M_{in}, M_{out})$, тогда:
 - $T_v(SN) = dom(l)$ —мн-во видимых переходов;
 - $A_v(SN) = rgn(l)$ —мн-во соответствующих наблюдаемых активностей;
 - $T_v^u(SN) = \left\{ t \in T_v(SN) \mid \forall t' \in T_v(SN): \begin{array}{l} l(t) = l(t') \Rightarrow t = t' \end{array} \right\}$ — мн-во уникальных видимых переходов;
 - $A_v^u(SN) = \{l(t) \mid t \in T_v^u(SN)\}$ — мн-во соответствующих уникальных наблюдаемых активностей

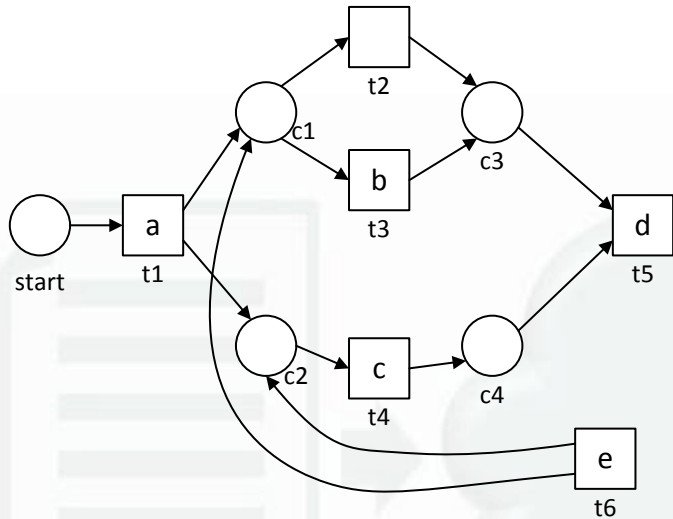
Сеть Петри: трассы

- Пусть $SN = (N, M_{in}, M_{out}) \in \mathcal{U}(SN)$ — системная сеть, тогда:
 - $\phi(SN) = \{\sigma_v | (M_{in})[\sigma_v \triangleright (M_{out})\}$ — мн-во всех видимых трасс;
 - $\phi_f(SN) = \{\sigma | (M_{in})[\sigma \rangle (M_{out})\}$ — мн-во соответствующих полных последовательностей срабатывания



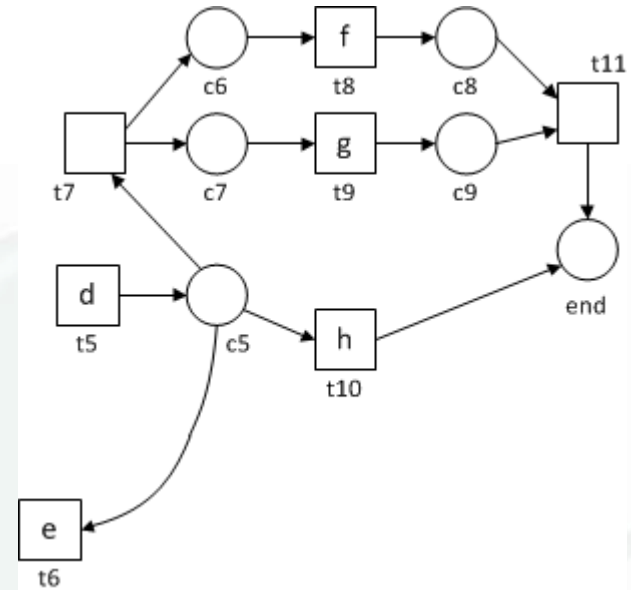
$\phi(SN) = \{$	$\phi_f(SN) = \{$
$\langle a, c, d, f, g \rangle,$	$\langle t1, t2, t4, t5, t7, t8, t9, t11 \rangle,$
$\langle a, c, b, d, f, g \rangle,$	$\langle t1, t3, t4, t5, t10 \rangle, \dots\}$
$\langle a, c, d, h \rangle,$	
$\langle a, b, c, d, e, c, d, h \rangle, \dots\}$	

Объединение сетей Петри



$$SN^1 = (N^1, M_{in}^1, M_{out}^1) \in \mathcal{U}(SN)$$

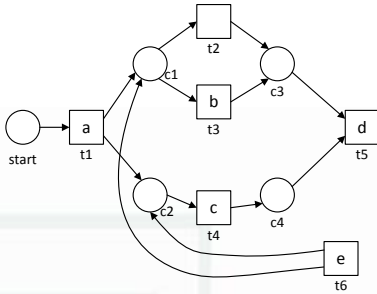
$$N^1 = (P^1, T^1, F^1, l^1)$$



$$SN^2 = (N^2, M_{in}^2, M_{out}^2) \in \mathcal{U}(SN)$$

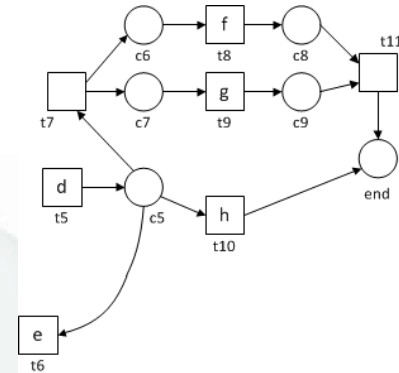
$$N^2 = (P^2, T^2, F^2, l^2)$$

Объединение сетей Петри



$$SN^1 = (N^1, M_{in}^1, M_{out}^1) \in \mathcal{U}(SN)$$

$$N^1 = (P^1, T^1, F^1, l^1)$$



$$SN^2 = (N^2, M_{in}^2, M_{out}^2) \in \mathcal{U}(SN)$$

$$N^2 = (P^2, T^2, F^2, l^2)$$

- $l^3 \in (T^1 \cup T^2) \rightarrow \mathcal{U}_A$ — объединение ф. разметки:

- $dom(l^3) = dom(l^1) \cup dom(l^2)$
- $l^3(t) = l^1(t) \Rightarrow t \in dom(l^1)$
- $l^3(t) = l^2(t) \Rightarrow t \in dom(l^2) \setminus dom(l^1)$

- $N^1 \cup N^2 = (P^1 \cup P^2, T^1 \cup T^2, F^1 \cup F^2, l^3)$

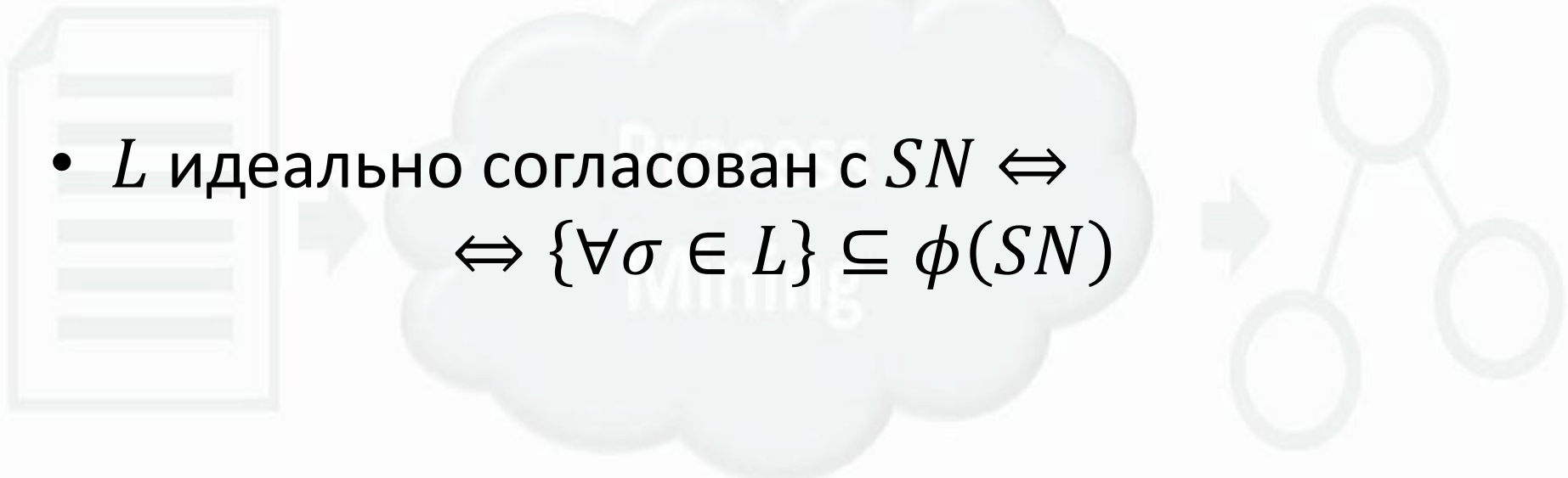
- $SN^1 \cup SN^2 = (N^1 \cup N^2, M_{in}^1 \uplus M_{in}^2, M_{out}^1 \uplus M_{out}^2)$

||

ПРОВЕРКА СООТВЕТСТВИЯ ЛОГА И МОДЕЛИ

Идеально (при)годный лог

- $L \in \mathcal{B}(A^*)$ — лог
- $SN = (N, M_{in}, M_{out}) \in \mathcal{U}_{SN}$ — системная сеть
- L идеально согласован с $SN \Leftrightarrow$
 $\Leftrightarrow \{\forall \sigma \in L\} \subseteq \phi(SN)$



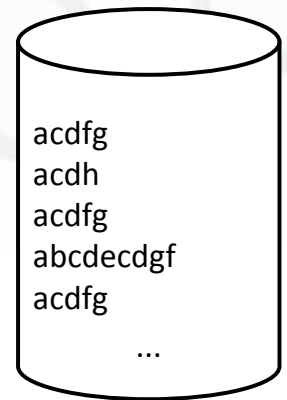
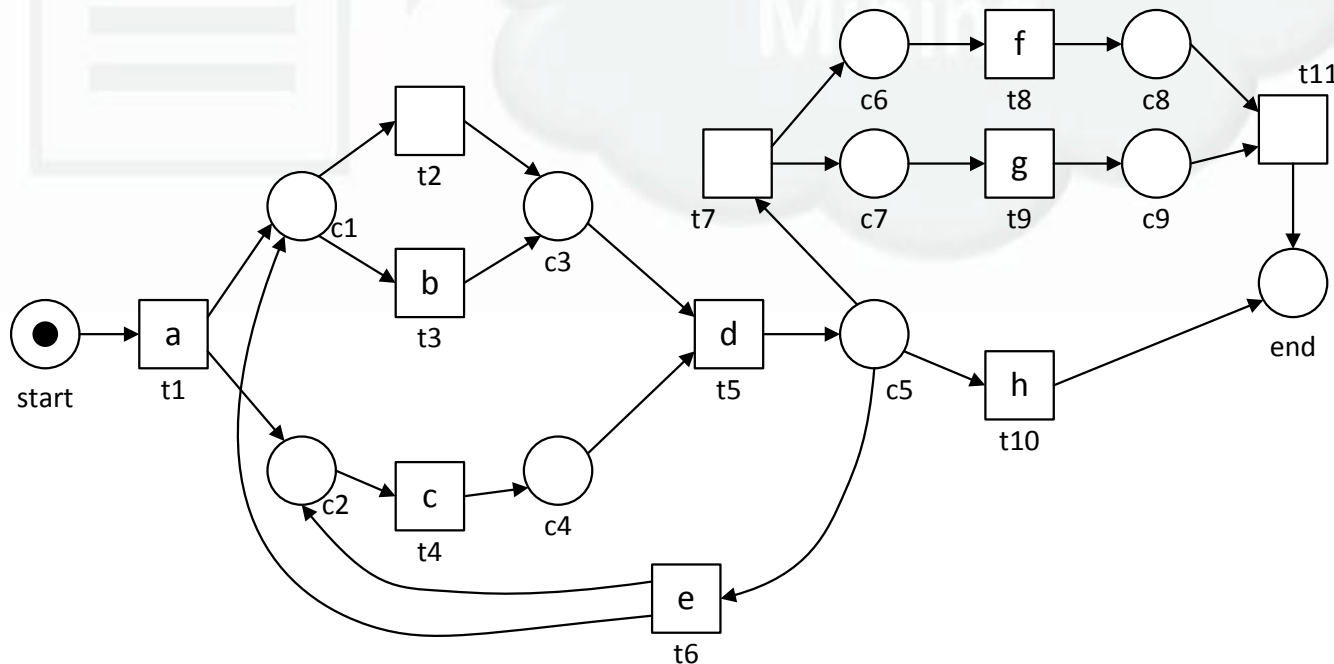
Как измерить годность? – Выравнивание

$$L_1 = \left[\langle a, c, d, f, g \rangle^{10}, \langle a, c, d, h \rangle^5, \langle a, b, c, d, e, c, d, g, f \rangle^5 \right]$$



$$\gamma_1 = \begin{array}{|c|c|c|c|c|c|c|c|} \hline a & c & \gg & d & \gg & f & g & \gg \\ \hline a & c & \tau & d & \tau & f & g & \tau \\ \hline t1 & t4 & t2 & t5 & t7 & t8 & t9 & t11 \\ \hline \end{array}$$

$$\gamma_2 = \begin{array}{|c|c|c|c|c|} \hline a & c & \gg & d & h \\ \hline a & c & \tau & d & h \\ \hline t1 & t4 & t2 & t5 & t10 \\ \hline \end{array}$$

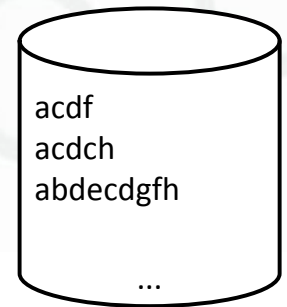
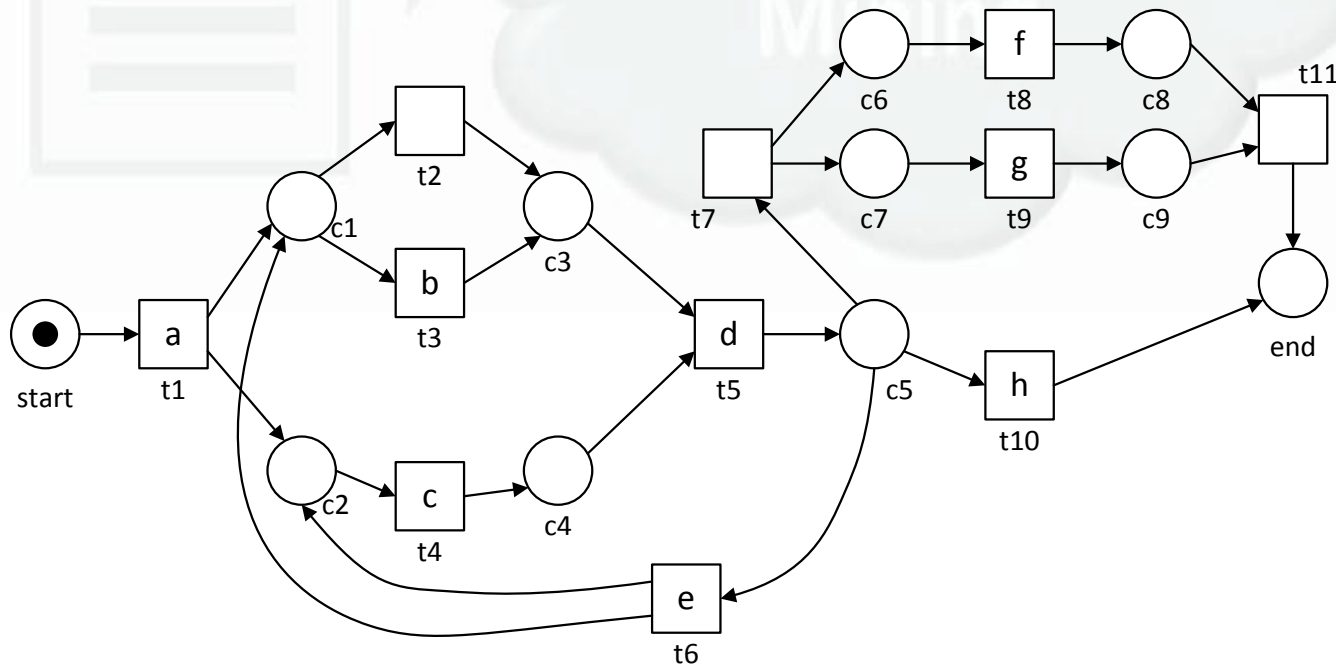
$$\gamma_3 = \begin{array}{|c|c|c|c|c|c|c|c|c|c|c|} \hline a & b & c & d & e & c & \gg & d & \gg & g & f & \gg \\ \hline a & b & c & d & e & c & \tau & d & \tau & g & f & \tau \\ \hline t1 & t3 & t4 & t5 & t6 & t4 & t2 & t5 & t7 & t9 & t8 & t11 \\ \hline \end{array}$$


Выравнивание 2

$$L_2 = \left[\langle a, c, d, f \rangle^{10}, \langle a, c, d, c, h \rangle^5, \langle a, b, d, e, c, d, g, f, h \rangle^5 \right]$$



$\gamma_4 =$	a	c	$\gg$	d	$\gg$	f	$\gg$	$\gg$	$\gamma_5 =$	a	c	$\gg$	d	c	h	$\gamma_6 =$	a	b	$\gg$	d	e	c	$\gg$	d	$\gg$	g	f	$\gg$	h
	a	c	τ	d	τ	f	g	τ		a	c	τ	d	$\gg$	h		a	b	c	d	e	c	τ	d	τ	g	f	τ	$\gg$
	t_1	t_4	t_2	t_5	t_7	t_8	t_9	t_{11}		t_1	t_4	t_2	t_5		t_{10}		t_1	t_3	t_4	t_5	t_6	t_4	t_2	t_5	t_7	t_9	t_8	t_{11}	



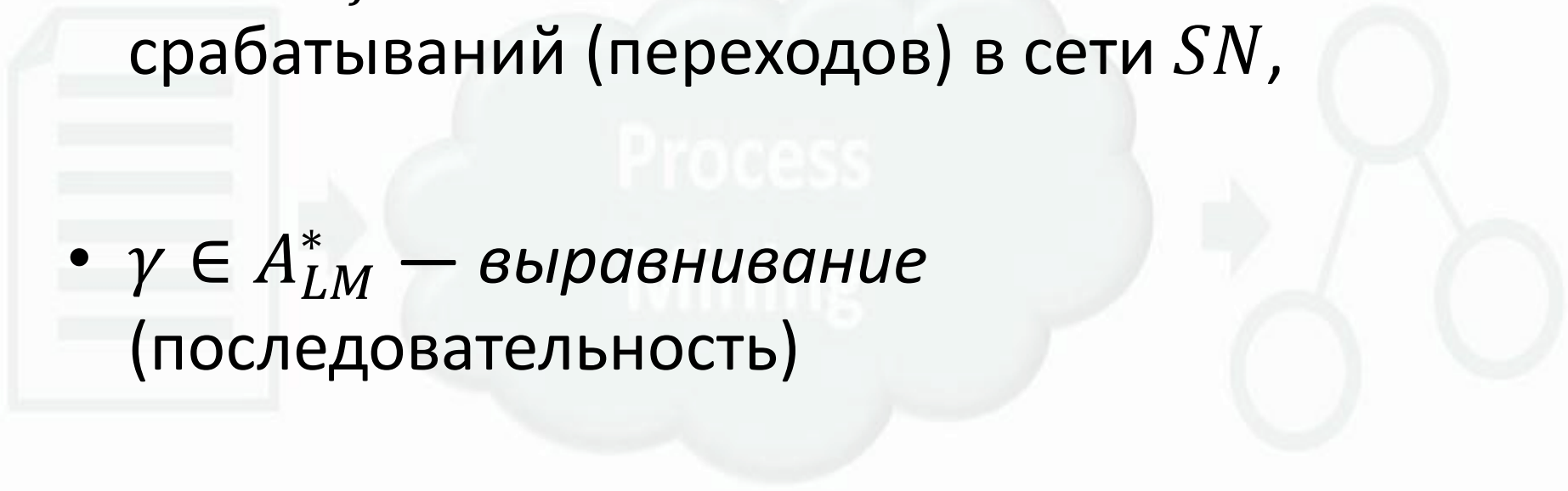
Допустимые ходы

- Ход — пара $(x, (y, t))$,
 - x — эл-т лога
 - (y, t) — эл-т модели
- $L \in \mathcal{B}(A^*)$ — лог,
- $SN = (N, M_{in}, M_{out}) \in \mathcal{U}_{SN}$ — системная сеть,
- множество допустимых ходов:

$$\begin{aligned} A_{LM} = & \left\{ (x, (x, t)) \mid x \in A, t \in T, l(t) = x \right\} \cup \\ & \cup \left\{ (\gg, (x, t)) \mid t \in T, l(t) = x \right\} \cup \\ & \cup \left\{ (x, \gg) \mid x \in A \right\} \end{aligned}$$

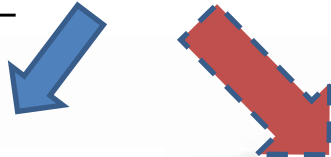
Выравнивание

- $\sigma_L \in L$ — трасса в логе,
- $\sigma_M \in \phi_f(SN)$ — полная последовательность срабатываний (переходов) в сети SN ,
- $\gamma \in A_{LM}^*$ — *выравнивание* (последовательность)

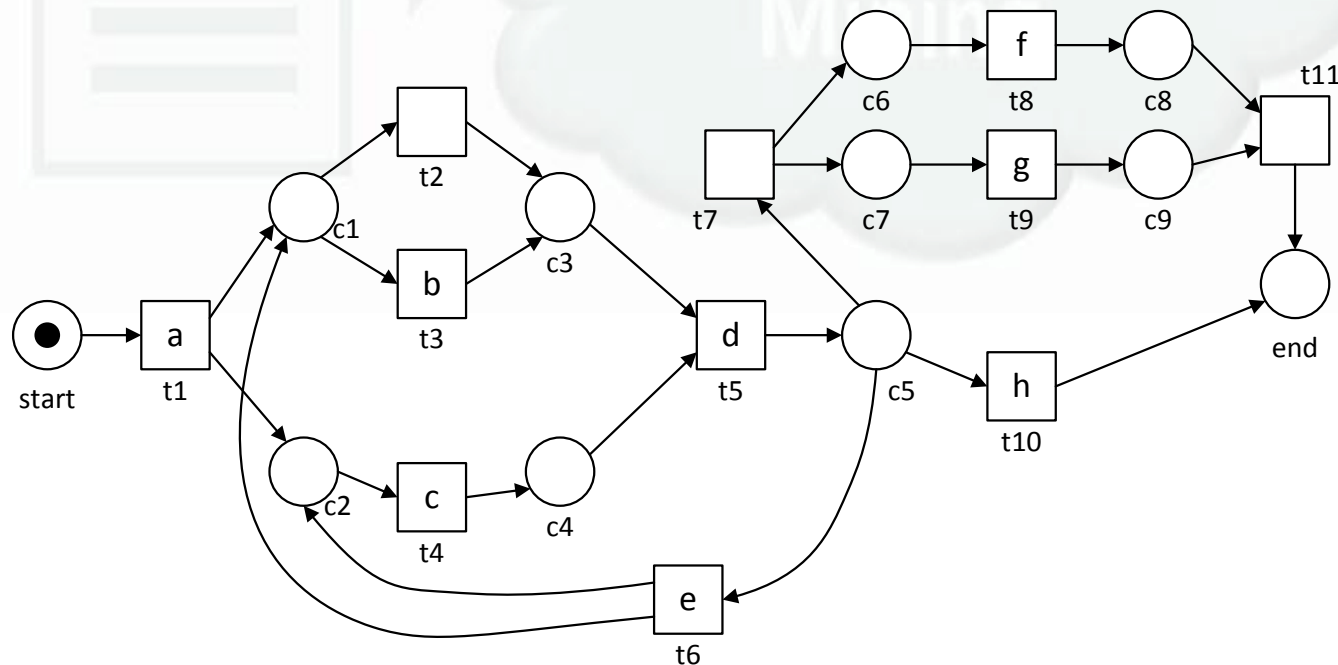


О множественности выравниваний

$$L_2 = \left[\langle a, c, d, f \rangle^{10}, \langle a, c, d, c, h \rangle^5, \langle a, b, d, e, c, d, g, f, h \rangle^5 \right]$$



$$\gamma_4 = \begin{array}{|c|c|c|c|c|c|c|c|} \hline a & c & \gg & d & \gg & f & \gg & \gg \\ \hline a & c & \tau & d & \tau & f & g & \tau \\ \hline t1 & t4 & t2 & t5 & t7 & t8 & t9 & t11 \\ \hline \end{array}$$

$$\gamma'_4 = \begin{array}{|c|c|c|c|c|c|c|} \hline a & c & \gg & d & \gg & f & \gg \\ \hline a & c & b & d & \tau & \gg & h \\ \hline t1 & t4 & t3 & t5 & t7 & & t10 \\ \hline \end{array}$$


Стоимостная функция

- $\delta \in A_{LM} \rightarrow \mathbb{N}$ — стоимостная функция;
- «стоимость» выравнивания $\gamma \in A_{LM}^*$:

$$\delta(\gamma) = \sum_{(x,y) \in \gamma} \delta(x,y)$$

$$\gamma_4 = \begin{array}{c|c|c|c|c|c|c|c|c|} a & c & \gg & d & \gg & f & \gg & \gg & \\ \hline a & c & \tau & d & \tau & f & g & \tau & \\ \hline t1 & t4 & t2 & t5 & t7 & t8 & t9 & t11 & \end{array}$$

$$\delta_s(\gamma_4) = 1$$

$$\gamma'_4 = \begin{array}{c|c|c|c|c|c|c|c|c|} a & c & \gg & d & \gg & f & \gg & & \\ \hline a & c & b & d & \tau & \gg & h & & \\ \hline t1 & t4 & t3 & t5 & t7 & & t10 & & \end{array}$$

$$\delta_s(\gamma'_4) = 3$$

Оптимальное выравнивание

- $L \in \mathcal{B}(A^*)$ — лог, $A \subseteq \mathcal{U}_A$
- $SN = (N, M_{in}, M_{out}) \in \mathcal{U}_{SN}$ — сеть: $\phi(SN) \neq \emptyset$
- для $\sigma_L \in L$ определяем
$$\Gamma_{\sigma_L, SN} = \left\{ \gamma \in A_{LM}^* \mid \exists \sigma_M \in \phi_f(SN): \gamma = align(\sigma_L, \sigma_M) \right\};$$
- выравнивание $\gamma \in \Gamma_{\sigma_L, SN}$ оптимально для трассы $\sigma_L \in L$ и SN
 $\Leftrightarrow \gamma' \in \Gamma_{\sigma_L, M}: \delta(\gamma') \geq \delta(\gamma);$
- $\lambda_{SN} \in A^* \rightarrow A_{LM}^*$ сопоставляет каждой трассе σ_L оптимальное выравнивание: $\lambda_{SN}(\sigma_L) \in \Gamma_{\sigma_L, SN}$ — оптимально;
- $costs(L, SN, \delta) = \sum_{\sigma_L \in L} \delta(\lambda_{SN}(\sigma_L))$ — стоимость рассогласования для всего лога

Идеально согласованный лог

- $L \in \mathcal{B}(A^*)$ — лог, $A \subseteq \mathcal{U}_A$
- $SN = (N, M_{in}, M_{out}) \in \mathcal{U}_{SN}$ — сеть:
 $\phi(SN) \neq \emptyset$
- лог L идеально согласован с сетью SN
 $\Leftrightarrow$
- $costs(L, SN, \delta) = 0$

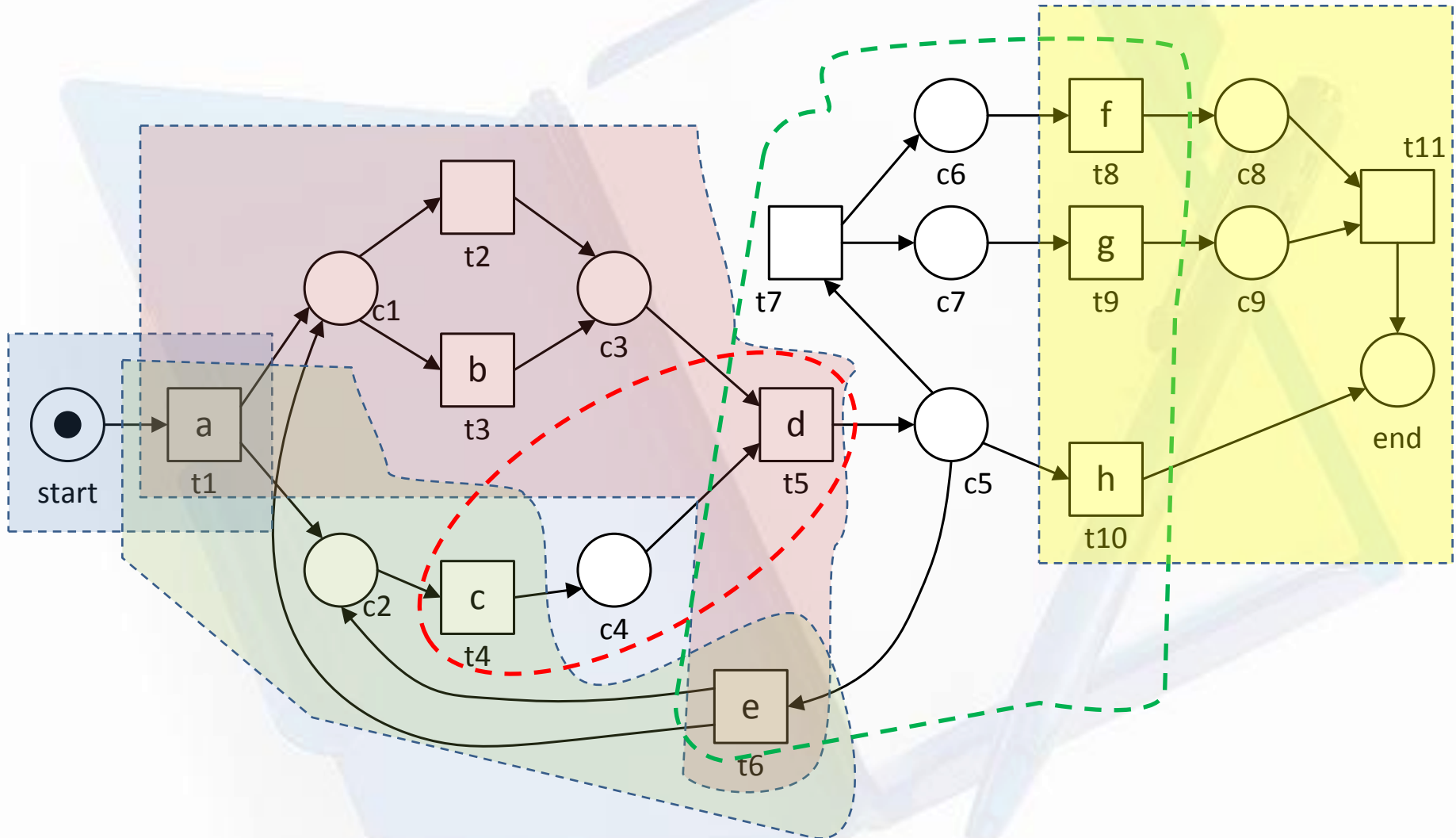
III

ЗАДАЧА ДЕКОМПОЗИЦИИ

Истинная декомпозиция

- Пусть $SN \in \mathcal{U}_{SN}$ — сеть с функцией разметки l ,
- $D = \{SN^1, SN^2, \dots, SN^n\} \subseteq \mathcal{U}_{SN}$ — истинная декомпозиция $\Leftrightarrow$
 - $SN^i = (N^i, M_{in}^i, M_{out}^i)$ — с.сеть где $N^i = (P^i, T^i, F^i, l^i) \forall (1 \leq i \leq n)$
 - $l^i = l \upharpoonright_{T^i} \forall (1 \leq i \leq n)$ — частичная ф-ция разм.
 - $P^i \cap P^j = \emptyset \forall (1 \leq i < j \leq n)$
 - $T^i \cap T^j \subseteq T_v^u(SN) \emptyset \forall (1 \leq i < j \leq n)$
- $SN = \bigcup_{1 \leq i \leq n} SN^i$
- $\mathcal{D}(SN)$ — мн-во всех истинных декомпозиций

Максимальная декомпозиция



Максимальная декомпозиция

• Пусть $SN \in \mathcal{U}_{SN}$ — сеть,

• $D_{\max}(SN)$ — максимальная декомпозиция:

1) $\forall (x, y) \in F: \llbracket (x, y) \rrbracket \subseteq F$ — эл. сеть:

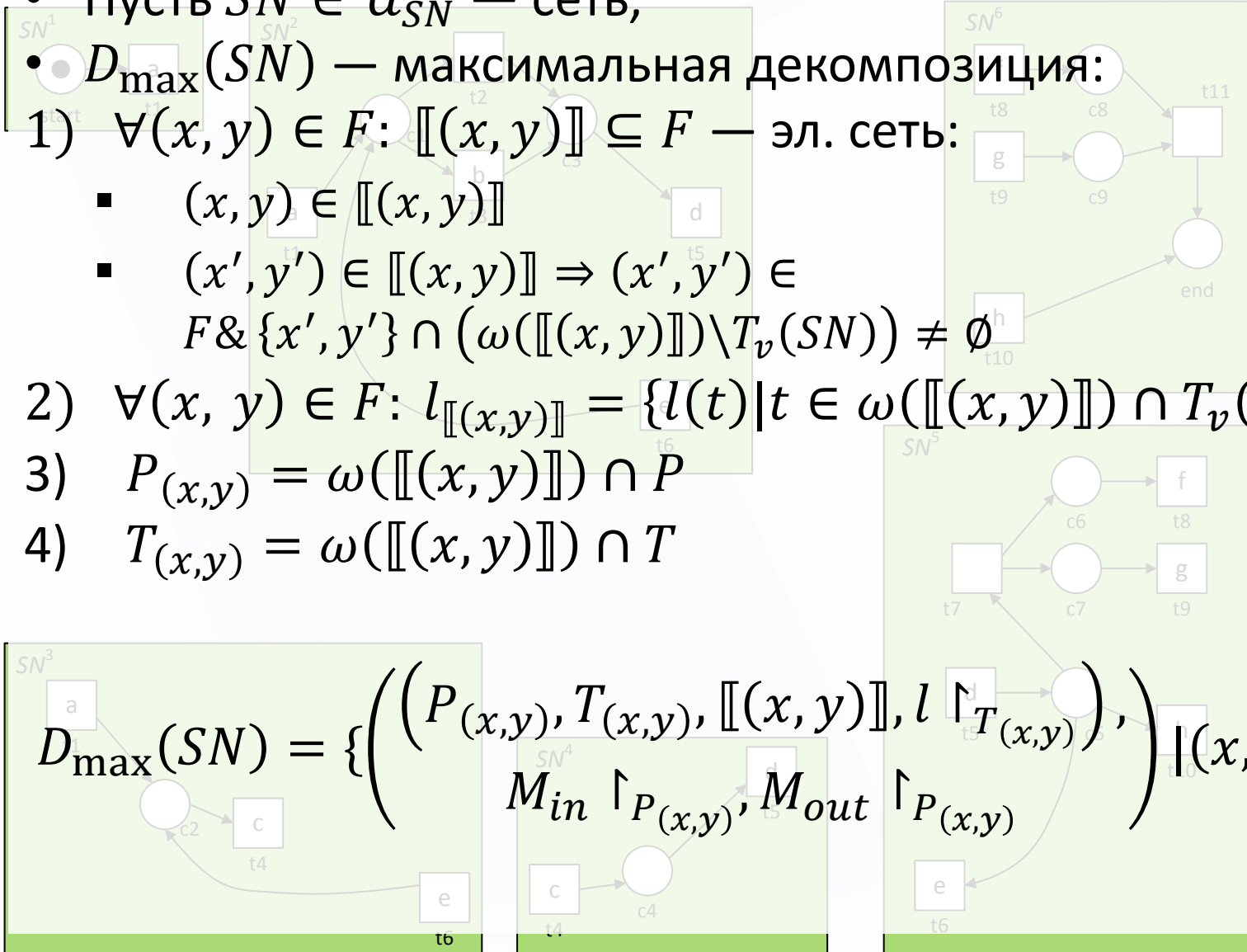
- $(x, y) \in \llbracket (x, y) \rrbracket$
- $(x', y') \in \llbracket (x, y) \rrbracket \Rightarrow (x', y') \in F \& \{x', y'\} \cap (\omega(\llbracket (x, y) \rrbracket) \setminus T_v(SN)) \neq \emptyset$

2) $\forall (x, y) \in F: l_{\llbracket (x, y) \rrbracket} = \{l(t) | t \in \omega(\llbracket (x, y) \rrbracket) \cap T_v(SN)\}$

3) $P_{(x,y)} = \omega(\llbracket (x, y) \rrbracket) \cap P$

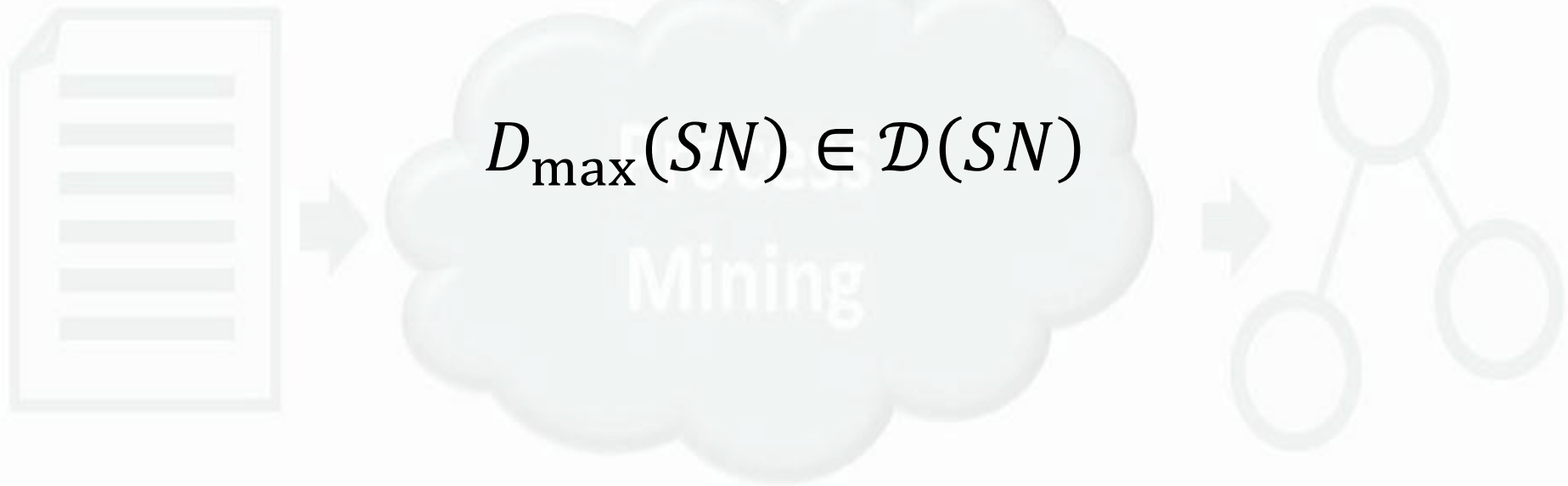
4) $T_{(x,y)} = \omega(\llbracket (x, y) \rrbracket) \cap T$

$$D_{\max}(SN) = \left\{ \left(\left(P_{(x,y)}, T_{(x,y)}, \llbracket (x, y) \rrbracket, l \upharpoonright_{T_{(x,y)}} \right), \left(M_{in} \upharpoonright_{P_{(x,y)}}, M_{out} \upharpoonright_{P_{(x,y)}} \right) \right) \mid (x, y) \in F \right\}$$

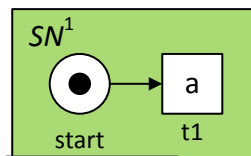


Максимальная декомпозиция истинна

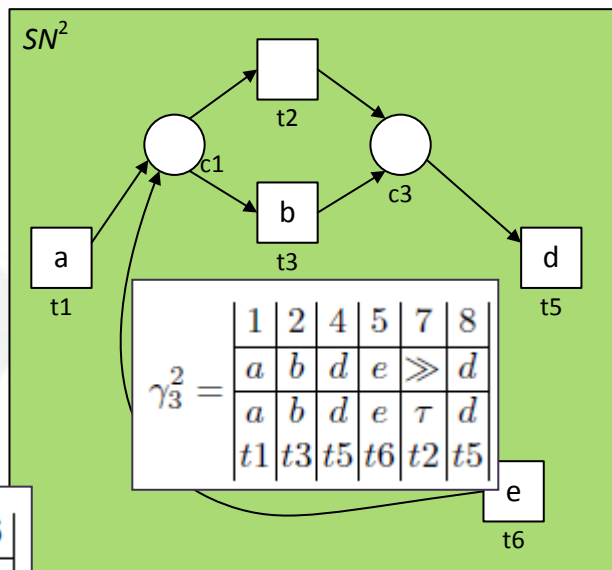
- $SN \in \mathcal{U}_{SN}$ — сеть,
- $D_{\max}(SN)$ — максимальная декомпозиция

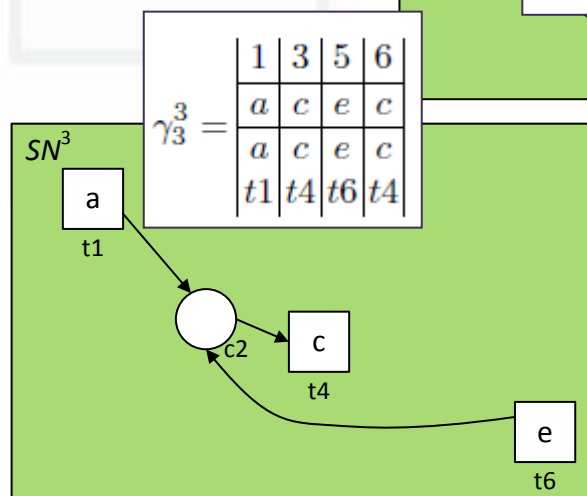


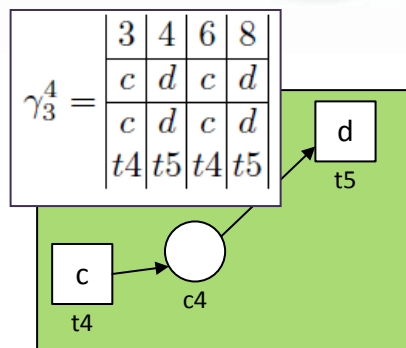
Декомпозиция выравнивания

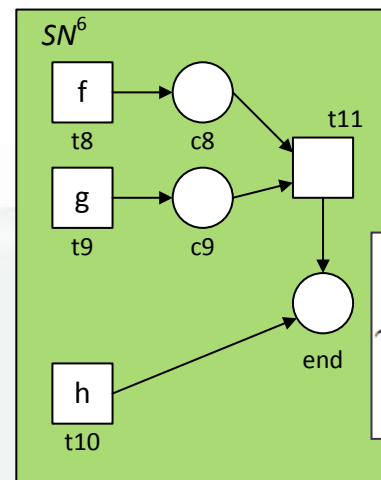


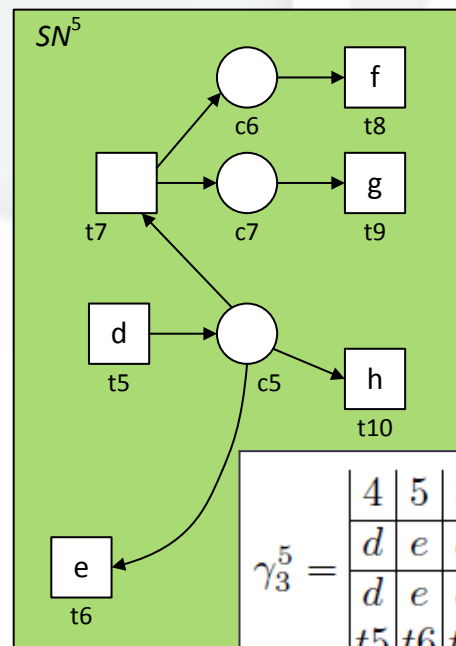
$$\gamma_3^1 = \begin{array}{|c|} \hline 1 \\ \hline a \\ \hline a \\ \hline t1 \\ \hline \end{array}$$

$$\gamma_3 = \begin{array}{|c|c|c|c|c|c|c|c|c|c|c|c|} \hline 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\ \hline a & b & c & d & e & c & \gg & d & \gg & g & f & \gg \\ \hline a & b & c & d & e & c & \tau & d & \tau & g & f & \tau \\ \hline t1 & t3 & t4 & t5 & t6 & t4 & t2 & t5 & t7 & t9 & t8 & t11 \\ \hline \end{array}$$


$$\gamma_3^2 = \begin{array}{|c|c|c|c|c|c|} \hline 1 & 2 & 4 & 5 & 7 & 8 \\ \hline a & b & d & e & \gg & d \\ \hline a & b & d & e & \tau & d \\ \hline t1 & t3 & t5 & t6 & t2 & t5 \\ \hline \end{array}$$


$$\gamma_3^3 = \begin{array}{|c|c|c|c|} \hline 1 & 3 & 5 & 6 \\ \hline a & c & e & c \\ \hline a & c & e & c \\ \hline t1 & t4 & t6 & t4 \\ \hline \end{array}$$


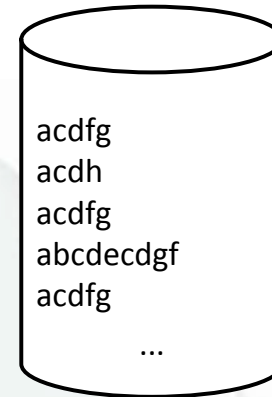
$$\gamma_3^4 = \begin{array}{|c|c|c|c|} \hline 3 & 4 & 6 & 8 \\ \hline c & d & c & d \\ \hline c & d & c & d \\ \hline t4 & t5 & t4 & t5 \\ \hline \end{array}$$


$$\gamma_3^6 = \begin{array}{|c|c|c|} \hline 10 & 11 & 12 \\ \hline g & f & \gg \\ \hline g & f & \tau \\ \hline t9 & t8 & t11 \\ \hline \end{array}$$


$$\gamma_3^5 = \begin{array}{|c|c|c|c|c|c|} \hline 4 & 5 & 8 & 9 & 10 & 11 \\ \hline d & e & d & \gg & g & f \\ \hline d & e & d & \tau & g & f \\ \hline t5 & t6 & t5 & t7 & t9 & t8 \\ \hline \end{array}$$

Выделение сублогов

$$L_1 = [\langle a, c, d, f, g \rangle^{10}, \\ \langle a, c, d, h \rangle^5, \\ \langle a, b, c, d, e, c, d, g, f \rangle^5]$$



$$L_1^1 = L_1 \upharpoonright_{\{a\}} = [\langle a \rangle^{20}]$$

$$L_1^2 = L_1 \upharpoonright_{\{a, b, d, e\}} = [\langle a, d \rangle^{15}, \langle a, b, d, e, d \rangle^5]$$

$$L_1^3 = L_1 \upharpoonright_{\{a, c, e\}} = [\langle a, c \rangle^{15}, \langle a, c, e, c \rangle^5]$$

Утверждение о возможности декомпозиции проверки согласованности

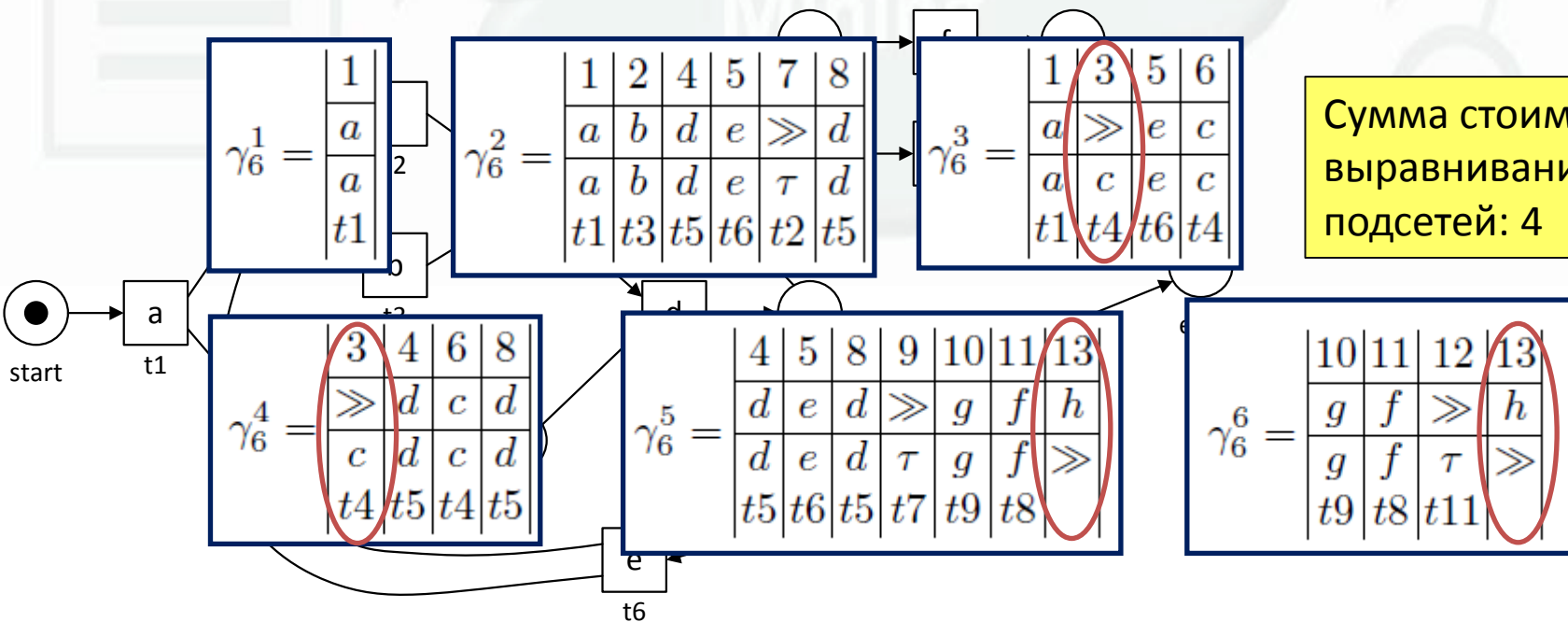
- $L \in \mathcal{B}(A^*)$ — лог, где $A \subseteq \mathcal{U}_A$,
 - $SN \in \mathcal{U}_{SN}$ — сеть.
-
- Если $D = \{SN^1, SN^2, \dots, SN^n\} \subseteq \mathcal{D}(SN)$ — истинная декомпозиция, тогда
лог L идеально соответствует сети SN
 $\Leftrightarrow$
 $\forall (1 \leq i \leq n) L \upharpoonright_{A_v(SN^i)}$
идеально соответствует SN^i

Проблема подсчета стоимости

- Трасса $\langle a, b, d, e, c, d, g, f, h \rangle$
- Оптимальное выравнивание

$$\gamma_6 = \begin{array}{c|cccccccccccccc} & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 \\ \hline a & b & \gg & d & e & c & \gg & d & \gg & g & f & \gg & h \\ \hline a & b & c & d & e & c & \tau & d & \tau & g & f & \tau & \gg \\ \hline t1 & t3 & t4 & t5 & t6 & t4 & t2 & t5 & t7 & t9 & t8 & t11 & \end{array}$$

Стоимость
оптимального
выравнивания: 2



Сумма стоимостей
выравнивания
подсетей: 4

Адаптированная функция стоимости

- $D = \{SN^1, SN^2, \dots, SN^n\} \subseteq \mathcal{D}(SN)$ — истинная декомпозиция сети SN ,
- $\delta \in A_{LM} \rightarrow \mathbb{N}$ — функция стоимости:
- функция $c_Q(a, (a, t)) = c_Q(>>, (a, t)) = c_Q(a, >>) = |\{1 \leq i \leq n | a \in A_i\}|$ подсчитывает подсетей, содержащих наблюдаемую активность a ;
- $\delta_Q(x, y) = \frac{\delta(x, y)}{c_Q(x, y)}$ — адаптированная функция стоимости

О нижней границе стоимости рассогласования

- $L \in \mathcal{B}(A^*)$ — лог, где $A \subseteq \mathcal{U}_A$,
 - $SN \in \mathcal{U}_{SN}$ — сеть,
 - δ — функция стоимости.
- Если $D = \{SN^1, SN^2, \dots, SN^n\} \subseteq \mathcal{D}(SN)$ — истинная декомпозиция, тогда

$$\text{costs}(L, SN, \delta) \geq \sum_{1 \leq i \leq n} \text{costs}(L \upharpoonright_{A_v(SN^i)}, SN^i, \delta_Q)$$

О доле идеально пригодных трасс

- $L \in \mathcal{B}(A^*)$ — лог, где $A \subseteq \mathcal{U}_A$,
- $SN \in \mathcal{U}_{SN}$ — сеть.

- Если $D = \{SN^1, SN^2, \dots, SN^n\} \subseteq \mathcal{D}(SN)$ — истинная декомпозиция, тогда

$$\begin{aligned} \frac{|[\sigma \in L | \sigma \in \phi(SN)]|}{|L|} &= \\ &= \frac{|[\sigma \in L | \forall (1 \leq i \leq n) \sigma \upharpoonright_{A_v(SN^i)} \in \phi(SN^i)]|}{|L|} \end{aligned}$$

Заключение

- Горизонтальное разделение логов
- Задача проверки согласованности
- Расчет коэффициента пригодности
- Другое использование выравниваний
- Основная цель: методика декомпозиции задачи Conformance Checking получена и доказана формально!

Связанные работы

- Декомпозиция Conformance Checking с использованием нотации «пассажей»
- Декомпозиция Process Discovery — построение модели по логу

Литература

- W.M.P. van der Aalst. Process Mining: Discovery, Conformance and Enhancement of Business Processes. Springer-Verlag, Berlin, 2011.
- W.M.P. van der Aalst. Decomposing Process Mining Problems Using Passages. In S. Haddad and L. Pomello, editors, Applications and Theory of Petri Nets 2012, volume 7347 of Lecture Notes in Computer Science, pages 72{91. Springer-Verlag, Berlin, 2012.
- W.M.P. van der Aalst. Decomposing Petri Nets for Process Mining. A Generic Approach. — 2012.

Спасибо за внимание!