Verma, Seidl, Schwentick 2005

Formulas in existential Presburger arithmetic ($\exists PA$)

are defined by

$t ::= 0 \mid 1 \mid x \mid t_1 + t_2$ $\quad\quad$ $\varphi ::= t_1 = t_2 \mid t_1 > t_2$

$\mid \varphi_1 \wedge \varphi_2 \mid \varphi_1 \vee \varphi_2 \mid \exists x \varphi.$

Theorem (Verma, Seidl, Schwentick 2005)

Given a CFG G, one can compute in linear time an existential Presburger formula φ_G so that

$$Sol(\varphi_G) = \Psi(L(G)).$$

The reduction is in two steps:

(I) We translate the given CFG G into a communication-free Petri net N_G

(II) We turn the communication-free Petri net N_G into an $\exists PA$ formula φ_G.

Petri nets:

Definition (Syntax of Petri nets)

A Petri net $N = (S, T, W)$ consists of

- S = finite set of places
- T = finite set of transitions
- $W: (S \times T) \cup (T \times S) \longrightarrow \mathbb{N}$ weight function.

A Petri net is communication-free if for each transition t there is at most one place s with weight $W(s,t) > 0$.

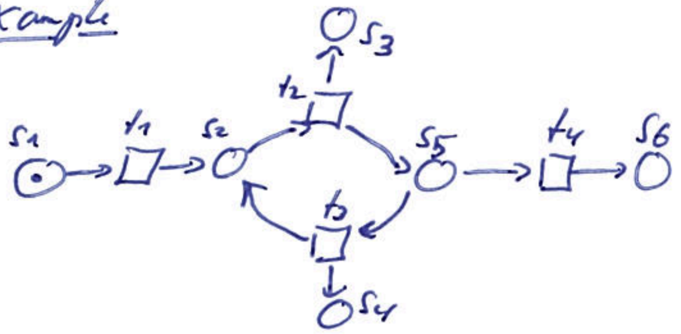
Moreover, $W(s,t) = 1$ holds.

Definition (Semantics of Petri nets):

- A marking $M: S \to \mathbb{N}$ assigns a number of tokens to each place.

- Marking M __enables__ __transition t__, if
  $M(s) \geq W(s,t)$ for all $s \in S$.
- A transition t that is enabled in M can __fire__,
  $M \xrightarrow{t} M'$ with $M'(s) = M(s) - W(s,t) + W(t,s)$ for all $s \in S$.
- Marking M' is __reachable from M__, if there are transitions $t_1, \dots, t_n$ so that
  $$M \xrightarrow{t_1} M_1 \xrightarrow{t_2} \dots \xrightarrow{t_n} M'.$$

__Example__



$$\begin{pmatrix}1\\0\\0\\0\\0\\0\end{pmatrix} \xrightarrow{t_1} \begin{pmatrix}0\\1\\0\\0\\0\\0\end{pmatrix} \xrightarrow{t_2} \begin{pmatrix}0\\0\\1\\0\\1\\0\end{pmatrix} \xrightarrow{t_4} \begin{pmatrix}0\\0\\1\\0\\0\\1\end{pmatrix}$$

__Theorem__ (Main Results in Petri Net Theory, __Mayr__ '81, __Lipton__ '76, __Rackoff__ '78, __Esparza__ '98)

- Reachability in Petri nets is decidable (Mayr '81) and (at least) EXPSPACE-hard (Lipton '76).
- Coverability ($M_1 \rightarrow M \geq M_2$?) is EXPSPACE-complete (EXPSPACE-hard Lipton '76, in EXPSPACE Rackoff '78).
- Reachability in communication-free Petri nets is NP-complete (Esparza '98)

Given a Petri net $N = (S, T, W)$, the __connectivity matrix__ $\mathbb{C} \in \mathbb{Z}^{|S| \times |T|}$ summarises the effect of firing transitions:

$$\mathbb{C}(s,t) := W(t,s) - W(s,t)$$

__Lemma (Marking Equation)__

If $M \xrightarrow{\sigma} M'$ with $\sigma \in T^*$, then $M' = M + \mathbb{C} \cdot \Psi(\sigma)$.

For communication-free Petri nets,
a weak converse of the marking equation holds.

<u>Theorem (Esparza '98):</u>

Let (S, T, W, M_0) be a communication-free Petri net.

Consider $X : T \to \mathbb{N}$.

Then there is $\sigma \in T^*$ with $\Psi(\sigma) = X$ and $M_0 \xrightarrow{\sigma}$

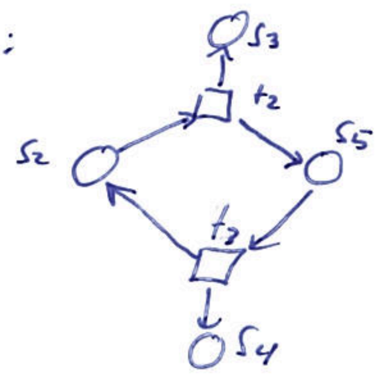
<u>iff</u> (a) $M_0 + CX \geq 0$

(b) in the subgraph induced by X, every place is reachable (in the graph-theoretic sense) from some $s \in S$ with $M_0(s) > 0$.

<u>On the example:</u>

(1) $X = (0\ 1\ 1\ 0)$

Subnet:



Property (a) holds:

$$\begin{pmatrix}1\\0\\0\\0\\0\\0\end{pmatrix} + \begin{pmatrix}-1 & 0 & 0 & 0\\ 1 & -1 & 1 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 1 & -1 & -1\\ 0 & 0 & 0 & 1\end{pmatrix}\begin{pmatrix}0\\1\\1\\0\end{pmatrix} \geq 0$$

Property (b) fails.

$\Rightarrow$ No transition sequence.

(2) $X = (1\ 0\ 0\ 1)$

Subnet

Property (a) fails:

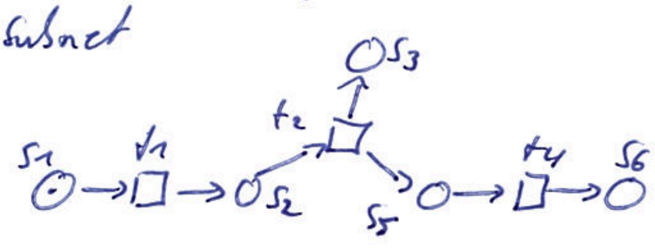
-1 on s_5.

Property (b) fails.

$\Rightarrow$ No transition sequence.

(3) $X = (1\ 1\ 0\ 1)$

Subnet



Property (a) holds.

Property (b) holds.

$\Rightarrow$ $t_1.t_2.t_4$ has required Parikh image.

Construction for Part (I):

Consider the CFG $G = (\underbrace{V}_{\text{non-terminals}}, \underbrace{\mathcal{U}}_{\text{terminals}}, P, A_0)$

Define $N_G := (V \cup \mathcal{U}, \underbrace{P}_{\text{Transitions}}, W, M_0)$

with initial marking $M_0(A_0) := 1$ and $M_0(B) := 0$ otherwise.

For the weights, consider

$$p = A \to \alpha \in P.$$

Then

$$W(A, p) := 1$$

$$W(p, B) := (\Psi(\alpha))(B) \quad \text{with } B \in V \cup \mathcal{U}.$$

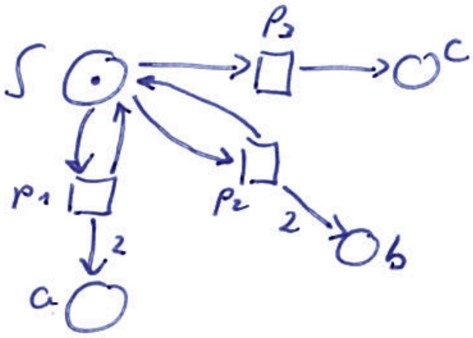
Example:

$G = (\{S\}, \{a, b, c\}, \{p_1, p_2, p_3\}, S)$

with

$p_1 = S \to aSa \qquad p_2 = S \to bSb \qquad p_3 = S \to c.$

Yields



Lemma:

Consider CFG $G = (V, \mathcal{U}, P, A_0)$ and $X : P \to \mathbb{N}$.

There is a derivation of G that uses $p \in P$ exactly $X(p)$ times

iff there is $\sigma \in T^*$ in N_G with $\Psi(\sigma) = X$ and $M_0 \xrightarrow{\sigma}$.

We know how to check the latter using Esparza's theorem.

Construction for Part (II):

- Encode constraints (a) and (b) in Esparza's theorem using $\exists PA$.
- Let $G = (V, \mathcal{U}, P, A_0)$ be the CFG and $N_G = (V \cup \mathcal{U}, P, W, M_0)$ the associated Petri net.
- For each terminal $a \in \mathcal{U}$, let x_a be a variable.

  // These are the variables that will be free in φ_G.

  For each production $p \in P$, let y_p be a variable.

There are <u>3 kinds</u> of subformulas:

(1) For each non-terminal $A \in V$, there is an equation for requirement (a) in Esparza's theorem:

$$M_0(A) - \sum_{i=1}^{k} y_{p_i} + \sum_{p \in P} A(p) \cdot y_p = 0.$$

↳ Here, $p_1, \dots, p_k$ are the productions with lhs A.

↳ Moreover, for each production $p \in P$, $A(p)$ is the number of occurrences of A on the rhs.

↳ Equality with 0 ensures the derivation is <u>complete</u>, i.e., all non-terminals are consumed.

(There is no need to have such formulas for terminals $a \in \mathcal{U}$ as a only occurs on the rhs of productions).

(2) Ensure x_a is consistent with y_p

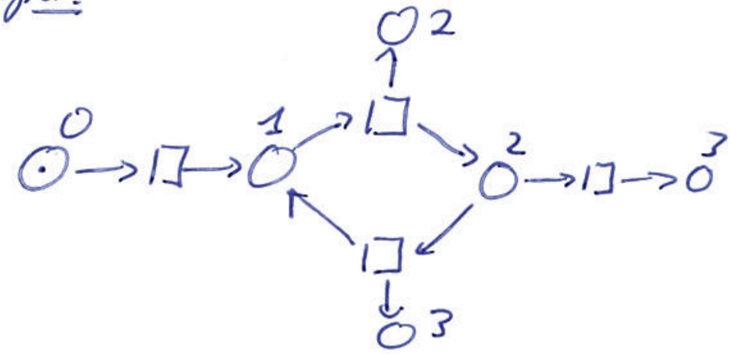
$$x_a = \sum_{p \in P} a(p) \cdot y_p.$$

(3) Express requirement (b) in Esparza's theorem.

Introduce variables z_A for each $A \in \mathcal{U} \cup V$.

Idea: z_A = distance of A from A_0
in a spanning tree on the subgraph of N_G
induced by the productions = transitions p with $y_p > 0$.

On the example:



Construction:

- For each terminal $a \in \mathcal{U}$, we have the implication

$$x_a = 0 \vee z_a > 0.$$

// To mark place a, assign z_a a value in the spanning tree.

- Consider $A \in V \cup \mathcal{U}$.

Let $p_1, \dots, p_h \in P$ be productions with
↳ A on the rhs ↳ $B_1, \dots, B_h \in V$ on the lhs

We have the implication

$$z_A = 0 \vee \bigvee_{i=1}^{h} \left(z_A = z_{B_i} + 1 \wedge y_{p_i} > 0 \wedge z_{B_i} > 0 \right)$$

If one of the B_i is the start symbol A_0,
the corresponding disjunct becomes

$$z_A = 1 \wedge y_{p_i} > 0.$$

Define:

$$\varphi_G := \exists y_1, \dots, y_{|P|} \; \exists z_1, \dots, z_{|\mathcal{U}|+|V|} : (1) \wedge (2) \wedge (3).$$

Theorem (Verma, Seidl, Schwentick):

Given a CFG G, the $\exists$PA formula φ_G

- can be computed in linear time
- and satisfies $Sol(\varphi_G) = \Psi(L(G))$.

Proof:

- Equality is by Esparza's theorem.
- ℓ_G can be computed in time linear in $|P|$,

the size of productions (rhs is part of the size of a production).

To see this, note that also

$$\mathcal{M}_0(\mathcal{A}) - \sum_{i=1}^{k} y_{p_i} + \sum_{p \in P} \mathcal{A}(p)\, y_p = 0$$

can be constructed in linear time driven by the productions.

$\square$