Petri nets: Structural analysis

We have seen how properties of Petri nets can be proved by constructing the reachability graph and analysing it.

However, the reachability graph may become huge: exponential in the number of places (if it is finite at all).

Structural analysis enables us to prove some properties *without* constructing the reachability graph. The main techniques are:

Place invariants

Traps



3

Incidence Matrix

Let $N = \langle P, T, F, W, M_0 \rangle$ be a P/T net. The corresponding incidence matrix *C*: $P \times T \rightarrow \mathbb{Z}$ is the matrix whose rows correspond to places and whose columns correspond to transitions. Column $t \in T$ denotes how the firing of *t* affects the marking of the net: C(t, p) = W(t, p) - W(p, t).

The incidence matrix of Example 1:

$$\begin{pmatrix} t_1 & t_2 & t_3 & t_4 & t_5 & t_6 \\ -1 & 0 & 1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & -1 & 1 \\ 0 & 0 & 0 & -1 & 0 & 1 \\ \rho_5 \\ 0 & 0 & 0 & 1 & -1 & 0 \\ \rho_6 \\ \rho_7 \end{pmatrix}$$

Let us now write markings as column vectors. E.g., the initial marking in Example 1 is $M_0 = (1 \ 0 \ 0 \ 1 \ 1 \ 0 \ 0)^T$.

Likewise, we can write firing counts as column vectors with one entry for each transition. E.g., if each of the transitions t_1 , t_2 , and t_4 fires once, we can express this with $u = (1\ 1\ 0\ 1\ 0\ 0)^T$.

Then, the result of firing these transitions can be computed as $M_0 + C \cdot u$.

$$\begin{pmatrix} 1\\0\\0\\1\\-1&-1&0&0&0&0\\0&1&-1&0&0&0&0\\0&-1&1&0&-1&1\\0&0&0&-1&0&1\\0&0&0&-1&0&1\\0&0&0&0&1&-1&0\\0&0&0&0&1&-1 \end{pmatrix} \cdot \begin{pmatrix} 1\\1\\0\\1\\0\\0\\0 \end{pmatrix} = \begin{pmatrix} 0\\0\\1\\0\\0\\1\\0\\0 \end{pmatrix}$$

Let N be a P/T net with indicence matrix C, and let M, M' be two markings of N. The following implication holds:

If $M' \in reach(M)$, then there exists a vector u such that $M' = M + C \cdot u$ such that all entries in u are natural numbers.

Notice that the reverse implication does not hold in general!

E.g., bi-directional arcs (an arc from a place to a transition and back) cancel each other out in the matrix. For instance, if Example 1 contained a bi-directional arc between p_1 and t_3 , the matrix would remain the same, but the marking $\{p_3, p_6\}$ (obtained on the previous slide) would be unreachable!



Even though we have

$$\begin{pmatrix} 1\\0\\0\\0\\0 \end{pmatrix} + \begin{pmatrix} -1 & 1\\1 & -1\\-1 & 1\\0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1\\1 \end{pmatrix} = \begin{pmatrix} 1\\0\\0\\1 \end{pmatrix},$$

none of the sequences corresponding to $(1 \ 1)^T$, i.e. $t_1 t_2$ or $t_2 t_1$, can happen.

To summarize: The markings obtained by computing with the incidence matrix are an over-approximation of the actual reachable markings

However, we *can* sometimes use the matrix equations to show that a marking *M* is *unreachable*. (Compare coverability graphs...)

I.e., a corollary of the previous implication is that if M' = M + Cu has no natural solution for u, then $M' \notin reach(M)$.

Note: When we are talking about natural (integral) solutions of equations, we mean those whose components are natural (integral) numbers.

Example 3

Consider the following net and the marking $M = (1 \ 1)^T$.



$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} + \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \cdot \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

has no solution, and therefore M is not reachable.

Let *N* be a net and *C* its incidence matrix. A natural solution of the equation Cu = 0 is called a transition invariant (or: T-invariant) of *N*.

Notice that a T-invariant is a vector with one entry for each transition.

For instance, in Example 3, $u = (1 \ 1)^T$ is a T-invariant.

A T-invariant indicates a possible loop in the net, i.e. a sequence of transitions whose net effect is null, i.e. which leads back to the marking it starts in.

Let *N* be a net and *C* its incidence matrix. A natural solution of the equation $C^T x = 0$ such that $x \neq 0$ is called a place invariant (or: P-invariant) of *N*.

Notice that a P-invariant is a vector with one entry for each place.

For instance, in Example 1, $x_1 = (1 \ 1 \ 1 \ 0 \ 0 \ 0)^T$, $x_2 = (0 \ 0 \ 1 \ 1 \ 0 \ 0 \ 1)^T$, and $x_3 = (0 \ 0 \ 0 \ 1 \ 1 \ 1)^T$ are all P-invariants.

A P-invariant indicates that the number of tokens in all reachable markings satisfies some linear invariant (see next slide).

Let *M* be marking reachable with a transition sequence whose firing count is expressed by *u*, i.e. $M = M_0 + Cu$. Let *x* be a P-invariant. Then, the following holds:

$$M^{T}x = (M_{0} + Cu)^{T}x = M_{0}^{T}x + (Cu)^{T}x = M_{0}^{T}x + u^{T}C^{T}x = M_{0}^{T}x$$

For instance, invariant x_2 means that all reachable markings *M* satisfy (switching to the functional notation for markings):

$$M(p_3) + M(p_4) + M(p_7) = M_0(p_3) + M_0(p_4) + M_0(p_7) = 1$$
(1)

As a special case, a P-invariant in which all entries are either 0 or 1 indicates a set of places in which the number of tokens remains unchanged in all reachable markings.

Note that linear combinations of P-invariants (i.e. multiplying an invariant by a constant or component-wise addition of two invariants) will again yield a P-invariant.

We can use P-invariants to prove mutual exclusion properties.

Example: According to equation 1, in every reachable marking of Example 1 exactly one of the places p_3 , p_4 , and p_7 is marked. In particular, p_3 and p_7 cannot be marked concurrently!

P-invariants can also be useful as a *pre-processing step* for reachability analysis.

Suppose that when computing the reachability graph, the marking of a place is normally represented with *n* bits of storage. E.g. the places p_3 , p_4 , and p_7 together would require 3n bits.

However, as we have discovered invariant x_2 , we know that exactly one of the three places is marked in each reachable marking.

Thus, we just need to store in each marking *which* of the three is marked, which required just two bits.

To compute some P-invariants, one can use the algorithm due to J. Farkas (1902).

Unfortunately there are P/T-nets with an exponential number of linearly independent P-invariants (in the number of places of the net). Thus the Farkas algorithm may take exponential time in the worst case.

Input: the incidence matrix *C* with *n* rows (places), and *m* columns (transitions).

Output: A set of place invariants.

Notation: $(C \mid E_n)$ denotes the juxtaposition of C by E_n , the $n \times n$ identity matrix.

 $D_0 := (C \mid E_n);$

for i := 1 to m do

for d_1 , d_2 rows in D_{i-1} such that $d_1(i)$ and $d_2(i)$ have opposite signs do $d := |d_2(i)| \cdot d_1 + |d_1(i)| \cdot d_2;$ (* d(i) = 0 *) $d' := d/\gcd(d(1), d(2), \dots, d(m+n));$ augment D_{i-1} with d' as last row;

endfor;

delete all rows of the (augmented) matrix D_{i-1} whose *i*-th component

is different from 0, the result is D_i ;

endfor;

delete the first m columns of D_m

An example

Let us assume the following incidence matrix:

$$C = \begin{pmatrix} -1 & 1 & 1 & -1 \\ 1 & -1 & -1 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & -1 \\ -1 & 0 & 0 & 1 \end{pmatrix}$$
$$D_0 = (C \mid E_5) = \begin{pmatrix} -1 & 1 & 1 & -1 & | \ 1 & 0 & 0 & 0 & 0 \\ 1 & -1 & -1 & 1 & | \ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & -1 & | \ 0 & 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 & | \ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Addition of the rows 1 and 2, 1 and 4, 2 and 5, 4 and 5:

$$D_1 = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & -2 & 1 & 0 & 0 & 1 & 0 \\ 0 & -1 & -1 & 2 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix}$$

Addition of rows 3 und 4:

$$D_3 = D_4 = \begin{pmatrix} 0 & 0 & 0 & 0 & | 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & | 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & | 1 & 1 & 0 & 1 & 1 \end{pmatrix}$$

Minimal P-invariants are (1, 1, 0, 0, 0) and (0, 0, 0, 1, 1).

Incidence matrix for a net with 2*n* places:

$$C^{T} = \begin{pmatrix} -1 & -1 & 1 & 1 & 0 & 0 & \cdots & 0 & 0 \\ -1 & -1 & 0 & 0 & 1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots \\ -1 & -1 & 0 & 0 & 0 & 0 & \cdots & 1 & 1 \end{pmatrix}$$

 $(y_1, 1 - y_1, y_2, 1 - y_2, \dots, y_n, 1 - y_n)$ is an invariant for every $y_1, y_2, \dots, y_n \in \{0, 1\}$, and so there are 2^n linearly independent P-invariants.

Let $\langle P, T, F, W, M_0 \rangle$ be a P/T net. A trap is a set of places $S \subseteq P$ such that $S^{\bullet} \subseteq {}^{\bullet}S$.

In other words, each transition which removes tokens from a trap must also put at least one token back to the trap.

A trap *S* is called marked in marking *M* iff for at least one place $s \in S$ it holds that $M(s) \ge 1$.

Note: If a trap S is marked initially (i.e. in M_0), then it is also marked in all reachable markings.

In Example 4 (see next slide), $S_1 = \{nc_1, nc_2\}$ is a trap.

The only transitions that remove tokens from this set are t_2 and t_5 . However, both also add new tokens to S_1 .

 S_1 is marked initially, and therefore in all reachable markings *M* the following inequality holds: $M(nc_1) + M(nc_2) \ge 1$

Traps can be useful in combination with place invariants to recapture information lost in the incidence matrix due to the cancellation of self-loop arcs.

Consider the following attempt at a mutual exlusion algorithm for cr_1 and cr_2 :



The idea is to achieve mutual exclusion by entering the critical section only if the other process is not already there.

Proving mutual exclusion properties using traps

In Example 4, we want to prove that in all reachable markings M, cr_1 and cr_2 cannot be marked at the same time. This can be expressed by the following inequality:

 $M(cr_1) + M(cr_2) \leq 1$

The P-invariants we can derive in the net yield these equalities:

$$M(q_1) + M(pend_1) + M(cr_1) = 1$$

$$M(q_2) + M(pend_2) + M(cr_2) = 1$$

$$M(cr_1) + M(nc_1) = 1$$

$$M(cr_2) + M(nc_2) = 1$$
(2)
(3)
(3)
(4)
(5)

However, these equalities are insufficient to prove the desired property!

Recall that $S_1 = \{nc_1, nc_2\}$ is a trap.

 S_1 is marked initially and therefore in all reachable markings *M*. Thus:

$$M(nc_1) + M(nc_2) \ge 1 \tag{6}$$

Now, adding (4) and (5) and subtracting (6) yields $M(cr_1) + M(cr_2) \le 1$, which proves the mutual exclusion property.

Petri nets: Unfoldings

Unfoldings are a data structure that represents the reachable markings of a Petri net.

They are used for *bounded* nets! In the following, we assume 1-boundedness; the technique can be extended to arbitrary bounds.

Unfoldings represent a trade-off in terms of time/space requirements; their size is in between that of a net and its reachability graph, and checking whether a marking is reachable becomes easier than for the net, but more difficult than from the reachability graph.

Unfoldings exploit the inherent concurrency of a Petri net.

Unfoldings for finite transition systems

Let T be a finite transition system with initial state X. One can define the acyclic unfolding U_T (which is used for CTL model checking):



Remark: \mathcal{U}_T can be viewed as a structure in which every state is labelled by a state from \mathcal{T} . We denote this labelling by the function B.

 $\mathcal{U}_{\mathcal{T}}$ contains the same behaviours as \mathcal{T} (and the same reachable states). Additionally, $\mathcal{U}_{\mathcal{T}}$ has a simpler structure (acyclic, in fact, a tree). However, in general, $\mathcal{U}_{\mathcal{T}}$ is *infinite*. \mathcal{P} is called a prefix of $\mathcal{U}_{\mathcal{T}}$ if \mathcal{P} is obtained by "pruning" arbitrary branches of $\mathcal{U}_{\mathcal{T}}$. Example:



Observation: One can always find a *finite* prefix containing the same reachable states as the infinite unfolding (by unrolling loops exactly once). We shall call such a prefix complete.

Let us discuss an algorithm to obtain a complete prefix of $\mathcal{U}_{\mathcal{T}}$.

The algorithm maintains a set \mathcal{E} , the set of states observed so far.

Some arcs in the prefix will be called cutoffs, we shall mark them red.

1. Initially, the prefix contains only the root, labelled by X. We set $\mathcal{E} := \{X\}$.

2. Select a node *n* on the prefix that is not the target of a cutoff edge. Let B(n) = Y be the label of the node, and let *Z* be a state with $Y \rightarrow Z$ such that the prefix does not contain any edge from *n* to a *Z*-labelled node.

2a. If no such pair n, Z exists, we are done.

2b. Otherwise, add a new, Z-labelled node to the prefix and add an edge from n to it.

2c. If $Z \in \mathcal{E}$, then the new edge is a cutoff. Otherwise, set $\mathcal{E} := \mathcal{E} \cup \{Z\}$.

3. Continue at step 2.

Step-by-step construction of the prefix in the previous example:



Observation (1): A complete prefix contains as many transitions as T.

Observation (2): The shape of the prefix depends on the *order* in which edges are added!

We generalize unfoldings for Petri nets, as follows:

The unfolding of a Petri net \mathcal{P} (or, a prefix of the same) is an *acyclic* Petri net \mathcal{Q} .

Assumption: Suppose that \mathcal{P} is 1-safe.

Remark: In the following, we call the places of Q conditions, the transitions of Q events. This merely serves to better distinguish the elements of P and Q, functionally they are the same!

```
Every condition of Q is labelled by a place of P, every event of Q by a transition of P.
```

Every event t' is of the form (P', t), where P' is the preset of t' and t the label of t'.

Let P' be a set of conditions. B(P') denotes the set of places labelling the elements of P'.

Every condition has exactly one incoming arc.

Some events in a complete prefix are labelled as cutoffs.

1. Let M_0 be the initial marking of \mathcal{P} . Then \mathcal{Q} initially contains one condition for each place in M_0 . The initial marking of \mathcal{Q} contains exactly these conditions. We set $\mathcal{E} := \{M_0\}$.

2. Let *t* be a transition of \mathcal{P} and P' a set of conditions none of which is the output place of a cutoff transition. Moreover, let P' be coverable in \mathcal{Q} (i.e., part of a reachable marking), let $B(P') = {}^{\bullet}t$, and suppose that (P', t) is not yet contained in \mathcal{Q} .

2a. If no such pair (P', t) exists, we are done.

2b. Add the event t' := (P', t) to the prefix (with P' as preset and label t). Moreover, extend the prefix by one condition for every output place of t and make it an output place of t'.

2c. We associate with t' a marking M'_t (which is reachable in \mathcal{P}) (see below). If $M_{t'} \in \mathcal{E}$, then t' is a cutoff. Otherwise $\mathcal{E} := \mathcal{E} \cup \{M_{t'}\}$.

Remark: If we omit step 2c (no cutoffs), then we obtain the full unfolding of \mathcal{P} .

The shape of Q again depends on the order in which events are added. (More on this in a moment!)

Example 1: Petri net...



... and a possible prefix of the unfolding



When adding t' = (P', t) to the prefix, $M_{t'}$ is determined as follows:

Idea: $M_{t'}$ is the marking obtained by making the "minimal" effort to fire t'.

Let *x*, *y* be two nodes (conditions *or* events) in Q. Let < be the smallest partial order where *x* < *y* if there is an edge from *x* to *y*.

Let x be a node of Q. We define $\lfloor x \rfloor := \{ y \mid y \le x \}$.

Let $M_{t'}$ be the marking obtained by firing the transitions of $\lfloor t' \rfloor$ (in any order). Note: Such a firing sequence exists since P' is coverable.



The construction of Q terminates since $M_{t'}$ is reachable in \mathcal{P} and since there are only finitely many reachable markings in \mathcal{P} .

In most cases, the prefix is

bigger than \mathcal{P} ;

smaller than its reachability graph.

From the structure of the unfolding we can derive statements about the mutual relationships of conditions:

Let p, q be two (different) conditions of Q.

p, *q* are called causally dependent if p < q or q < p. (I.e., in every firing sequence containing both conditions, one condition must be consumed to generate the other.)

p, *q* are in conflict if there are events *t*, *u* (where $t \neq u$), $t \in \lfloor p \rfloor$, $u \in \lfloor q \rfloor$, and $\bullet t \cap \bullet u \neq \emptyset$. (I.e., *p*, *q* can *never* occur in a reachable marking!)

p, *q* are called concurrent if they are neither causally dependent nor in conflict with one another (I.e., *p*, *q can* occur together in some reachable marking!)

Remark: A set of conditions P' in Q is coverable iff all pairs $p, q \in P'$ are concurrent.

The concurrency relation C (where p C q iff p, q are concurrent) can be computed efficiently while generating the unfolding.

If P' is reachable (resp. coverable) in Q, then so is B(P') in P. Question: Does the reverse hold? For finite-state systems \mathcal{T} we have:

A state Y is reachable in T iff a Y-labelled node is reachable in any prefix of T constructed according to our algorithm.

This holds independently of the order in which events are added to the prefix.

For Petri nets \mathcal{P} (and a prefix \mathcal{Q}) we would like to have the following completeness property:

A marking *M* is reachable in \mathcal{P} *iff* a marking *M'* with B(M') = M is reachable in any prefix \mathcal{Q} constructed according to our algorithm.

Unfortunately, this does not holds for all prefixes. For Petri net unfoldings, whether Q is complete, does depend on the order in which events are generated!

Example 2

Consider the following Petri net:



In Example 2 the marking $\{p\}$ is reachable, e.g. by firing ABT.

The net can also reach the marking $\{e, f\}$ by firing either A C or B D, and then return by firing E F to the initial marking.

We shall see that a prefix generated according to depth-first order will "overlook" the transition T.

Depth-first order generated the prefix shown below (order and cutoffs indicated in red):



Let Q^* be the (usually infinite) unfolding of P obtained from constructing the prefix without cutoff conditions.

Let \prec be a total order of the events that refines < (i.e. t < t' implies $t \prec t'$).

Intuition: \prec is a possible order in which the events of Q^* can be generated.

Let \mathcal{Q}_{\prec} be the (unique!) prefix of \mathcal{Q}^* , where the events are added in the order given by \prec .

We call \prec adequate iff \mathcal{Q}_{\prec} is complete.

Let *M* be a reachable marking in Q^* . Then we call $C_M := \bigcup_{p \in M} \lfloor p \rfloor$ a configuration.

When given a configuration C, its maximal (w.r.t. <) conditions constitute the corresponding reachable marking, denoted M_C .

Remark: For every event *t*, the set $\lfloor t \rfloor \cup t^{\bullet} \cup M_0 =: C_t$ is a configuration. We have $M_{C_t} = M_t$.

We call *E* an extension of *C* iff $C \cap E = \emptyset$ and $C \cup E$ is a configuration. In this case, we write $C \oplus E$ to denote the configuration $C \cup E$.

Let M, M' be two markings of Q^* such that B(M) = B(M'). If E is an extension of C_M , then there is an extension E' of $C_{M'}$ that is isomorphic to E.

The following condition guarantees that \prec is adequate:

Let t, t' be two events with $t \prec t'$ and $M_t = M_{t'}$, and E an extension of C_t and E' the extension of $C_{t'}$ isomorphic to E. Then $u \prec u'$ must hold, where $C_u = C_t \cup E$ and $C_{u'} = C_{t'} \cup E'$.

E.g., this is implied by taking a total order satisfying $t \prec t'$ if $|C_t| < |C_{t'}|$.

Proof: Let \prec be an order satisfying the above constraint. We show that Q_{\prec} is complete. So let *M* be a marking reachable in \mathcal{P} . Then there is a marking *M'* in Q^* with B(M') = M. Either $C_{M'}$ is contained in Q_{\prec} , or $C_{M'} = C_t \oplus E$ for some cutoff event *t*. But then there is another event *t'* with $M_{t'} = M_t$ and $t' \prec t$ and therefore a configuration $C' := C_{t'} \oplus E'$, where *E'* is isomorphic to *E*, and we have $B(M_{C'}) = B(M') = M$. Either *C'* is contained in Q_{\prec} , or one repeats the argument, but only finitely often since \prec is well-founded.