

UNIVERSITÀ DEGLI STUDI DI NAPOLI FEDERICO II
FACOLTÀ DI INGEGNERIA
DIPARTIMENTO DI INFORMATICA E SISTEMISTICA

Corso di Ingegneria e Tecnologie dei sistemi di controllo
Anno accademico 2001-02

Dispense sulle reti di Petri
a cura di

Francesco BASILE

Maggio 2002

Indice

Introduzione	1
1. Le reti di Petri	3
1.1. Le reti di Petri: definizioni	3
1.2. Interpretazioni di una rete di Petri	7
1.3. Confronto tra automi e reti di Petri	12
1.4. Il grafo di copertura	14
1.5. Classi fondamentali di PN	18
1.6. Caratterizzazioni lineari di una PN	20
1.7. Invarianti di una PN	23
2. Le proprietà fondamentali di un sistema PN	30
2.1. Le proprietà comportamentali	30
2.1.1. Raggiungibilità	30
2.1.2. Boundedness	32
2.1.3. Liveness	34
2.1.4. Reversibilità ed home-state	41
2.2. Le proprietà strutturali	41
3. Il controllo supervisivo di una rete di Petri	45
3.1. Il problema del controllo supervisivo	45
3.2. Il metodo della programmazione lineare intera	48
3.3. Il metodo dei posti monitor	50
3.3.1. Assenza di transizioni incontrollabili	51
3.3.2. Presenza di transizioni incontrollabili	55
3.4. Considerazioni conclusive	60
A. Richiami dalla teoria dei grafi	62

Bibliografia**66**

Introduzione

Un processo manifatturiero può essere visto come una sequenza di eventi che indicano l'inizio e la terminazione di attività necessarie per l'esecuzione del processo con le seguenti caratteristiche:

- tempificazione asincrona, ovvero l'istante di occorrenza di tali eventi non presenta caratteristiche di regolarità;
- concorrenza, ovvero più attività possono essere eseguite simultaneamente;
- sincronizzazione, ovvero più attività concorrenti devono essere completate prima di poter procedere;
- conflitto, alcune attività possono richiedere l'utilizzo di risorse comuni.

Pertanto un sistema manifatturiero è un *sistema ad eventi discreti*.

Gli eventi provenienti dal campo risultano suddivisi in controllabili ed incontrollabili a seconda che si possa o meno interdire la loro occorrenza. Risulta incontrollabile il guasto di una macchina, mentre risulta essere controllabile l'avvio di un processo nel senso che è possibile interdire che dal campo possa essere iniziato un certo processo (disabilitazione di un pulsante di avvio finché non si verifica una certa condizione). In seguito al verificarsi di un evento il sistema cambia il suo stato di funzionamento. In ciascuno stato vengono eseguite determinate azioni. Un problema di controllo in tali sistemi è quello di abilitare attraverso gli eventi controllabili solo le sequenze di eventi desiderate.

Esistono fondamentalmente due classi di sistemi ad eventi discreti: i sistemi ad eventi discreti *logici* e quelli *temporizzati*. La differenza riguarda la presenza, nei sistemi temporizzati, di una tempificazione delle attività associate agli eventi: si suppone che a ciascun attività sia associata una durata di tempo, che può essere *deterministica* o *stocastica*. Pertanto tale tempificazione non va confusa con l'istante di occorrenza di un certo evento, che è rigorosamente asincrona sia nei sistemi ad

eventi logici che in quelli temporizzati. Nel corso delle presenti dispense l'attenzione sarà posta solo sui primi.

Esistono differenti formalismi per modellare i sistemi ad eventi discreti. Un sistema ad eventi discreti logico può essere modellato come un automa o come una rete di Petri (PN). Un sistema ad eventi discreti temporizzato può essere modellato come un automa temporizzato, una rete di Petri e una rete di code. Nel corso delle presenti dispense sarà presentato l'approccio basato sulle reti di Petri. Senza dilungarci troppo in questa sede diciamo che tale scelta è basata sulle seguenti motivazioni:

- le PN sono un formalismo sia grafico che matematico e pertanto risultano facilmente comprensibili anche a persone poco esperte ed allo stesso tempo offrono una vasta gamma di tecniche di analisi;
- consentono la *rappresentazione locale dello stato* di un certo sottosistema che facilita la sintesi modulare di un modello composto da più sottosistemi;
- permettono di rappresentare la concorrenza in modo esplicito;
- è facile ottenere una PN temporizzata partendo da una PN non temporizzata, evitando così di cambiare completamente modello nel caso in cui le specifiche di controllo per un certo sistema sono sia logiche che prestazionali, caso peraltro frequente nei sistemi di automazione.

Per quanto riguarda l'implementazione dei controllori sintetizzati attraverso le tecniche esposte, che non è oggetto delle presenti dispense, dobbiamo dire che i controllori presentati saranno essenzialmente un insieme di predicati logici o a loro volta delle reti di Petri, entrambi implementabili attraverso la tecnologia di controllo PC-based che si sta ampiamente diffondendo in ambito industriale.

Le presenti dispense si propongono di presentare i concetti di base per il controllo di sistemi manifatturieri flessibili (FMS) attraverso tecniche basate sulle reti di Petri.

Nel capitolo 1 vengono presentati gli elementi di base delle reti di Petri e le relative proprietà di maggior interesse nell'automazione sono esposte nel capitolo 2. Nel capitolo 3 saranno presentate due tecniche di controllo per l'imposizione dei vincoli di mutua esclusione generalizzati su una PN controllata. Oggetto dell'appendice sarà il richiamo di alcuni elementi dalla teoria dei grafi utilizzati nel corso delle presenti dispense.

1. Le reti di Petri

Questo capitolo illustra gli elementi di base delle reti di Petri ed i principali strumenti di analisi per siffatti sistemi. In particolare vengono introdotti il grafo di copertura, l'equazione di stato, i concetti di P-(T-)invariante e la suddivisione delle reti in sottoclassi basata sulla topologia del grafo della rete. Infine, vengono mostrate le maggiori potenzialità delle reti di Petri rispetto agli automi a stati finiti per la modellistica di un sistema ad eventi.

1.1. Le reti di Petri: definizioni

Una rete di Petri (PN) è un grafo bipartito orientato. I due nodi del grafo associato ad una PN sono i posti e le transizioni. Essi sono connessi da archi orientati pesati che vanno da posti a transizioni e viceversa.

Una rete di Petri posto/transizione (P/T PN) è definita come una 4-pla $N = (P, T, Pre, Post)$ dove:

- P è un'insieme di m *posti* rappresentati da cerchi;
- T è un'insieme di n *transizioni* rappresentate da linee;
- $P \cap T = \emptyset$ (cioè posti e transizioni sono insiemi disgiunti), $P \cup T \neq \emptyset$ (la rete è costituita da almeno un posto o da una transizione);
- $Pre : P \times T \rightarrow \mathbb{N}$ è la funzione di ingresso o di pre-incidenza;
- $Post : T \times P \rightarrow \mathbb{N}$ è la funzione di uscita o di post-incidenza.

Esiste un'arco che va dal posto p_i alla transizione t_j se e solo se $Pre(p_i, t_j) \neq 0$. Parimenti esiste un'arco che va dalla transizione t_j al posto p_i se e solo se $Post(t_j, p_i) \neq 0$. Una PN si dice *ordinaria* se il peso di ciascun arco è pari ad uno. Un modo pratico di rappresentare la struttura della rete è l'introduzione della matrici di pre e post

incidenza. Le funzioni di pre e post incidenza possono essere rappresentate attraverso le matrici di pre e post incidenza, entrambe di dimensione $m \times n$ ($|P| \times |T|$) e così definite: $\mathbf{Pre}(p, t) = Pre(p, t)$ e $\mathbf{Post}(p, t) = Post(t, p)$. In tal caso una rete di Petri posto/transizione (P/T PN) è definita come una 4-pla $N = (P, T, \mathbf{Pre}, \mathbf{Post})$.

Una coppia formata da un posto p ed una transizione t è chiamata *self-loop* se p è sia un posto di ingresso che di uscita di t . Una PN si dice *pura* se non contiene self-loop. Una PN pura è completamente caratterizzata dalla *matrice di incidenza*

$$\mathbf{C} = \mathbf{Post} - \mathbf{Pre}.$$

Ad una PN è associato il vettore delle marcature $\mathbf{m} : P \rightarrow \mathbb{N}$ di dimensione m , che associa ad ogni posto della rete un numero intero non negativo di *gettoni*, rappresentati da cerchietti neri o esplicitamente da un numero intero all'interno del cerchio rappresentante il posto.

Gli insiemi *preset* e *postset* di un posto sono rispettivamente:

$$\begin{aligned} \bullet p &= \{t \in T \mid Post(t, p) > 0\}, \text{ insieme delle transizioni in ingresso a } p \\ p^\bullet &= \{t \in T \mid Pre(p, t) > 0\}, \text{ insieme delle transizioni in uscita da } p \end{aligned}$$

Gli insiemi *preset* e *postset* di una transizione sono rispettivamente:

$$\begin{aligned} \bullet t &= \{p \in P \mid Pre(p, t) > 0\}, \text{ insieme dei posti in ingresso a } t \\ t^\bullet &= \{p \in P \mid Post(t, p) > 0\}, \text{ insieme dei posti in uscita da } t \end{aligned}$$

Una transizione t è abilitata nella marcatura \mathbf{m} se e solo se

$$\forall p \in \bullet t, \quad \mathbf{m}(p) \geq Pre(p, t) \Leftrightarrow \mathbf{m} \geq \mathbf{Pre}(\cdot, t) \quad (1.1)$$

Una transizione abilitata può scattare (si noti che l'istante di scatto è determinato dall'istante in cui si ha l'occorrenza dell'evento ad essa associato). Lo scatto della transizione determina il passaggio della rete nella marcatura \mathbf{m}' così definita:

$$\mathbf{m}' = \begin{cases} \mathbf{m}(p) - Pre(p, t) & \text{se } p \in \bullet t \\ \mathbf{m}(p) + Post(t, p) & \text{se } p \in t^\bullet \\ \mathbf{m} & \text{altrimenti} \end{cases}$$

La notazione $\mathbf{m}[t > \mathbf{m}']$ denota il cambio di marcatura della rete in seguito allo scatto della transizione t . Una *sequenza di scatto* da \mathbf{m}_0 è la sequenza di transizioni $\sigma = t_1..t_k$ tali che $\mathbf{m}_0[t_1 > \mathbf{m}_1[t_2 > \mathbf{m}_2..[t_k > \mathbf{m}_k$.

Un *sistema PN* è l'insieme di una PN e della sua marcatura iniziale e si indica con la notazione $\mathcal{S} = \langle N, \mathbf{m}_0 \rangle$. Una marcatura \mathbf{m} si dice *raggiungibile* nel sistema \mathcal{S} se e solo se esiste una sequenza di scatto σ tale che $\mathbf{m}_0[\sigma > \mathbf{m}$. Dato un sistema PN \mathcal{S} l'insieme delle marcature raggiungibili (chiamato *insieme di raggiungibilità*) si indica con la notazione $R(\mathcal{S})$.

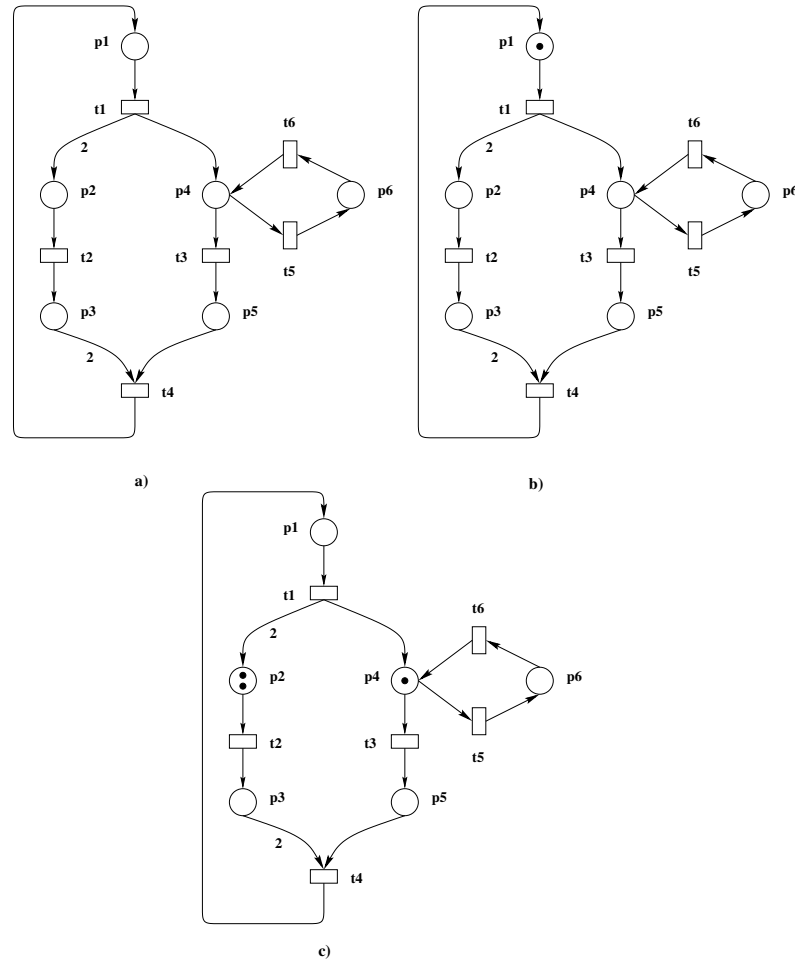


Figura 1.1.: a) una PN; b) un sistema PN; c) sistema PN in b) dopo lo scatto di t_1 .

Data una sequenza σ si introduce il vettore $\sigma : T \rightarrow \mathbb{N}$ detto *vettore di conteggio degli scatti*, ma denotato nel seguito più semplicemente come *vettore di scatto*: $\sigma(t)$ rappresenta il numero di occorrenze di t nella sequenza σ . Nota che uno stesso vettore

di scatto può essere rappresentativo di più sequenze; nel caso di una rete con tre transizioni il vettore di scatto $\sigma = [1 \ 1 \ 1]^T$ rappresenta ciascuna delle sequenze di scatto $t_1 t_2 t_3, t_2 t_1 t_3, \dots, t_3 t_2 t_1$. Pertanto il passaggio alla rappresentazione vettoriale comporta una perdita di informazione: si perde l'ordine di scatto all'interno di una sequenza.

La marcatura risultante in seguito allo scatto di una transizione t può essere rappresentata in maniera compatta dalla seguente equazione matriciale:

$$m = m_0 + C e_t$$

dove $e_t : T \rightarrow \{0, 1\}$ è il vettore caratteristico associato alla transizione t : $e_t(x) = 1$ se $x = t$, altrimenti $e_t(x) = 0$. Se invece consideriamo una sequenza σ tale che $m_0[\sigma > m]$ allora si ha:

$$m = m_0 + C \sigma \quad (1.2)$$

La (1.2) si chiama *equazione di stato* della rete.

Esempio 1 *Al fine di illustrare le definizioni su esposte si consideri la rete in fig. 1.1 (a). Dalle definizioni di matrici di incidenza, di pre-incidenza e post-incidenza si ha che:*

$$Pre = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad Post = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

$$C = \begin{bmatrix} -1 & 0 & 0 & 1 & 0 & 0 \\ 2 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -2 & 0 & 0 \\ 1 & 0 & -1 & 0 & -1 & 1 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix}$$

Gli insieme preset e postset di t_4 e p_1 sono:

$$\begin{aligned} \bullet t_4 &= \{p_3, p_5\}, & t_4 \bullet &= \{p_1\} \\ \bullet p_1 &= \{t_4\}, & p_1 \bullet &= \{t_1\} \end{aligned}$$

La rete è pura dal momento che non vi sono self-loop, mentre non è ordinaria perché esistono archi con peso diverso da uno come l'arco da t_1 a p_2 . Si consideri ora il sistema illustrato in fig. 1.1 (b) formato dalla rete in fig. 1.1 (a) sotto la marcatura iniziale $\mathbf{m}_0 = [1 \ 0 \ 0 \ 0 \ 0 \ 0]^T$. La transizione t_1 risulta abilitata dal momento che la (1.1) risulta verificata, infatti

$$\mathbf{m}_0 \geq \mathbf{Pre}(\cdot, t_1) \rightarrow \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \geq \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Pertanto in seguito allo scatto di t_1 la rete si porterà in un nuovo stato \mathbf{m}' , $\mathbf{m}[t_1 > \mathbf{m}']$. Applicando la (1.2) si ha:

$$\mathbf{m}' = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} -1 & 0 & 0 & 1 & 0 & 0 \\ 2 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -2 & 0 & 0 \\ 1 & 0 & -1 & 0 & -1 & 1 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

Nota che il vettore di scatto $\sigma = [1 \ 0 \ 0 \ 0 \ 0 \ 0]^T$ rappresenta la sequenza $\sigma = t_1$. Il nuovo stato del sistema PN considerato è illustrato in fig. 1.1 (c).

1.2. Interpretazioni di una rete di Petri

In questo paragrafo verranno mostrati due esempi di applicazioni di reti di Petri alla modellistica di sistemi ad eventi discreti: una cella robotizzata ed un programma in PASCAL concorrente. Non esiste una convenzione esplicita sul significato da dare ai posti, alle transizioni ed ai gettoni di una PN. Infatti in un sistema di automazione un posto può rappresentare un buffer o un particolare stato di una macchina, i gettoni i pezzi in lavorazione, le transizioni eventi di fine o avvio di una certa operazione. Mentre in ambito informatico i posti possono rappresentare registri, i gettoni dati e le transizioni possono essere associate all'esecuzione di un compito di elaborazione. Esiste poi il problema di stabilire se ed in che modo l'evoluzione della rete può essere associata all'esecuzione di azioni sull'ambiente, ovvero se vi è un gettone in un certo

posto viene eseguita una certa azione o se invece tale azione viene eseguita quando una certa transizione scatta. Si parla allora di *estensioni interpretate* di una PN. I due esempi che seguono prevedono tale possibilità e nel sistema manifatturiero le azioni sono associate ai posti, mentre per il programma le azioni sono associate alle transizioni.

Esempio 2 (Una cella robotizzata) *Si consideri la cella robotizzata di fig. 1.2. Essa è composta da una macchina M1, una macchina M2, un nastro trasportatore per il trasporto di pezzi grezzi all'interno della cella, un nastro trasportatore per il trasporto dei pezzi finiti all'esterno della cella, un buffer temporaneo di capacità sette ed un robot R per la movimentazione dei pezzi all'interno della cella.*

La cella ha il seguente funzionamento:

- *pezzi grezzi arrivano dal rispettivo nastro che si ferma quando la fotocellula $f1$ segnala la presenza ($f1 = 1$) di un pezzo a fine nastro;*
- *se la macchina M1 è libera ed il robot è libero, allora il robot preleva il pezzo grezzo dal nastro e lo carica sulla macchina M1;*
- *la macchina esegue l'operazione 1 sul pezzo;*
- *ad operazione terminata la macchina M1 resta in attesa per il deposito del pezzo semilavorato nel buffer temporaneo;*
- *se il robot è libero ed il buffer non è pieno il robot preleva il pezzo e lo deposita nel buffer;*
- *i pezzi depositati nel buffer vengono prelevati dal robot se esso e la macchina M2 sono liberi;*
- *la macchina M2 esegue l'operazione 2 sul pezzo semilavorato;*
- *ad operazione terminata la macchina M2 resta in attesa per il deposito del pezzo lavorato sul nastro di uscita;*
- *se il robot è libero e sul nastro di uscita in corrispondenza della posizione di deposito nessun pezzo è già presente ($f2 = 1$), allora il robot preleva il pezzo e lo deposita sul nastro di uscita;*

- una volta depositato sul nastro di uscita il pezzo viene condotto all'esterno della cella.

I singoli sottosistemi $M1$, $M2$ ed R sono supposti localmente controllati e al livello dell'architettura di controllo destinato alla coordinazione delle macchine (il livello in cui vengono di solito utilizzate le PN) sono disponibili i seguenti segnali:

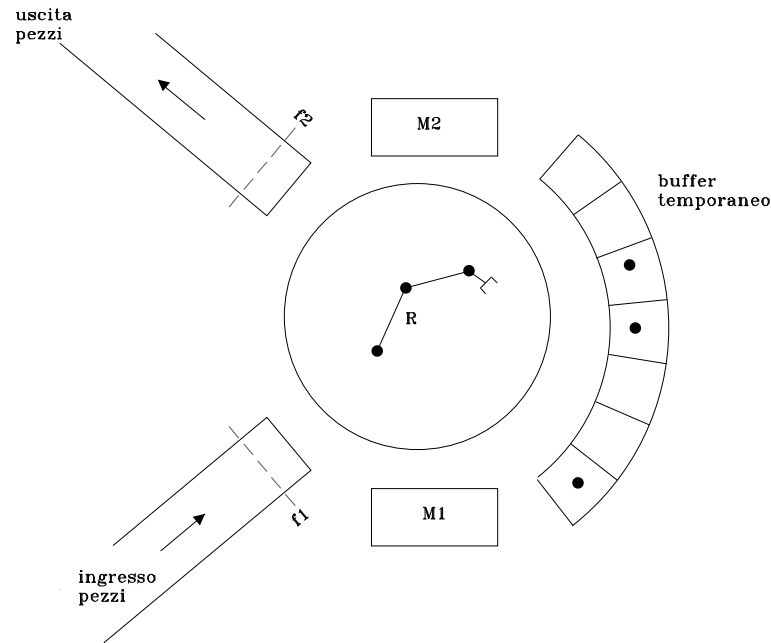


Figura 1.2.: Layout impianto FMS descritto in esempio 2

segnali di comando (azioni) per il livello locale di controllo

op1 esegui l'operazione 1

op2 esegui l'operazione 2

caricaM1 carica in M1 il pezzo dal nastro di ingresso

caricaM2 carica in M2 un pezzo dal buffer

deposito1 deposito nel buffer un pezzo lavorato da M1

deposito2 deposito su nastro di uscita un pezzo lavorato da M2

segnali di ingresso (eventi) dal livello locale di controllo

endop1 fine op1

```

    endop2 fine op2
    endc1 fine caricaM1
    endc2 fine caricaM2
    endd1 fine deposito1
    endd2 fine deposito2
segnali di ingresso (eventi) dal livello di campo
    f1 presenza pezzo da fotocellula1
    f2 presenza pezzo da fotocellula2

```

Dalle specifiche comportamentali risulta evidente che il livello di coordinamento delle macchine dovrà garantire la mutua esclusione nell'utilizzo del robot da parte delle fasi di lavorazione che lo richiedono, di evitare situazioni di overflow del buffer e di avvio di una qualunque operazione di prelievo/deposito di un pezzo da una locazione se esso non risulta presente o se rispettivamente la locazione di deposito non risulta libera.

Il modello risultante del modello PN è mostrato in fig. 1.3.

Esempio 3 (Un programma in PASCAL concorrente) Si prenda in esame il seguente programma scritto in un linguaggio di programmazione concorrente pascal-like.

```

loop
    A0;
    while C1 do (* p2 *)
        if C2 then A1 (* p3 *)
        else A2
        endif (* p4 *)
        parbegin A3, A4 (* p5, p7 *)
        parend (* p6, p8 *)
    endwhile (* p2 *)
endloop

```

Lo stato di controllo del programma ovvero l'indicazione dello stato in cui si trova in un certo istante il programma possono essere associati alla marcatura di una



PN. Le azioni da compiere in un certo stato e le condizioni associate al passaggio da uno stato all'altro possono essere associate alle transizioni di una PN. In base alla dinamica di una PN precedentemente introdotta, il comportamento della rete in fig. 1.4 descrive il comportamento del precedente programma concorrente. La cosa è interessante sia per capire come può essere implementata una PN e sia perché esse sono ampiamente utilizzate nella verifica del software.

1.3. Confronto tra automi e reti di Petri

Nel settore dell'automazione industriale uno strumento a lungo alternativo alle PN, ma tuttora utilizzato, è costituito dagli automi a stati finiti. Una PN il cui spazio di stato abbia cardinalità finita, è equivalente ad un automa a stati finiti. Un'automa è definito come la 5-pla $(Q, \Sigma, \delta, q_0, Q_m)$ dove:

- Q insieme finito di stati;
- Σ insieme finito di eventi (ingressi del sistema);
- $\delta: Q \times \Sigma \rightarrow Q$ funzione di transizione del sistema;
- $q_0 \in Q$ stato iniziale;
- $Q_m \subseteq Q$ sottoinsieme degli stati finali.

Un automa a stati finiti viene rappresentato da un grafo orientato i cui nodi rappresentano gli stati del sistema ed i cui archi sono associati agli eventi in seguito all'occorrenza dei quali si ha il passaggio da uno stato all'altro. Se si vuole tener conto delle uscite in un sistema siffatto, vanno aggiunti alla 5-pla precedente:

- Γ insieme finito delle uscite;
- $\lambda: Q \times \Sigma \rightarrow \Gamma$ (macchina di Mealy) o $\lambda: Q \rightarrow \Gamma$ (macchina di Moore) funzione di uscita del sistema.

In fig. 1.5 (a) è mostrato un sistema costituito da due macchine $M1, M2$ ed un buffer B . La macchina $M1$ ha due stati I_1 (libero) e W_1 (esecuzione op_1) ed in seguito all'occorrenza degli eventi s_1 (inizio op_1) e f_1 (fine op_1) si hanno i passaggi di stato indicati dal rispettivo automa in fig. 1.5 (b). La macchina $M2$ ha due stati I_2 (libera)

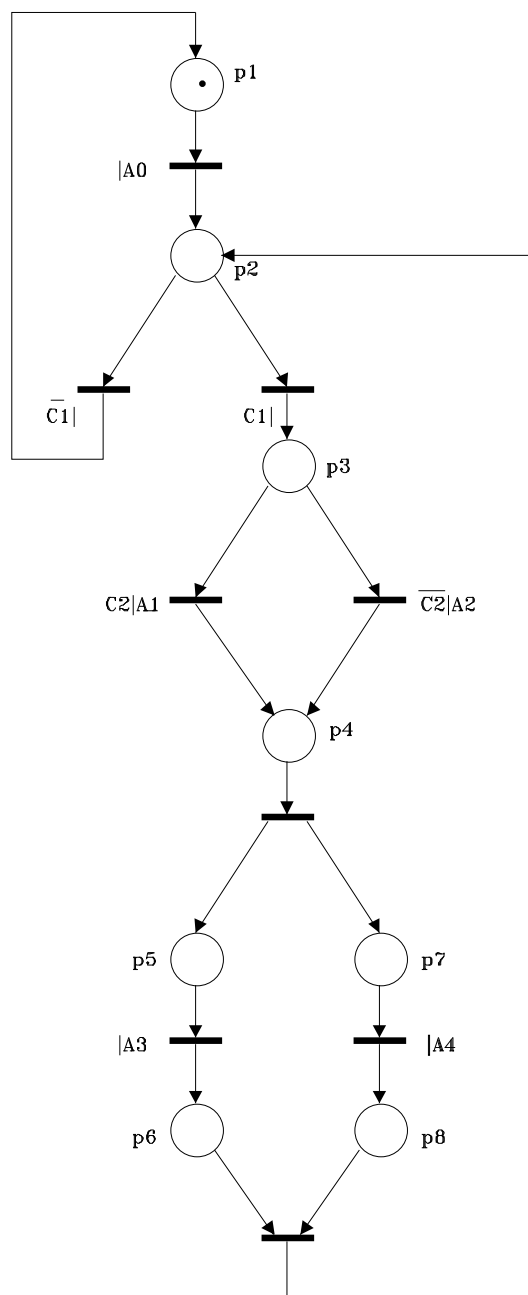


Figura 1.4.: Modello a rete di Petri del programma dell'esempio 3

e W_2 (esecuzione op_2) ed in seguito all'occorrenza degli eventi s_2 (inizio op_2) e f_2 (fine op_2) si hanno i passaggi di stato indicati dal rispettivo automa in fig. 1.5 (d). Il buffer B , che ha capacità pari ad uno, ha due stati E (vuoto) e F (pieno) ed in seguito all'occorrenza degli eventi s_2 ed f_1 si hanno i passaggi di stato indicati dal rispettivo automa in fig. 1.5 (c). Il comportamento del sistema complessivo, dato dall'esecuzione concorrente dei tre sottosistemi insieme al vincolo che $M1$ possa depositare un pezzo nel buffer B solo se esso è vuoto e che la macchina $M2$ può cominciare la lavorazione solo dopo che un pezzo è stato depositato nel buffer, è rappresentata in fig. 1.5 (e) dall'automa con otto stati.

Si noti come i modelli a PN equivalenti dei singoli sottosistemi e del sistema complessivo, mostrati nelle fig. 1.5 (f-i), evidenziano come le PN mantengono la modularità e come il numero di stati aumenta linearmente e non esponenzialmente rispetto al numero dei sottosistemi componenti. Il numero di stati di ciascun sottosistema è pari a due nel caso degli automi e ciascuna PN equivalente ha due posti, ma il sistema complessivo ha otto stati nel caso degli automi e sei posti nel caso delle PN. Questo è dovuto alla capacità di rappresentazione locale dello stato di un posto di una PN, ovvero la marcatura di un posto rappresenta lo stato di un preciso sottosistema, mentre lo stato di un automa è riferito sempre allo stato del sistema complessivo.

1.4. Il grafo di copertura

Nel seguito è descritto l'algoritmo per la costruzione del grafo di raggiungibilità. Se lo spazio di stato di una PN è finito ogni nodo di tale grafo orientato rappresenta una delle marcature raggiungibili a partire dalla marcatura iniziale della rete. Gli archi di tale grafo indicano il passaggio da una marcatura all'altra e la transizione che determina tale passaggio di stato. Nel caso che lo spazio di stato della rete non sia finito, il grafo viene chiamato grafo di copertura ed in esso compare l'elemento w (pseudo infinito) tale che $\forall a \in \mathbb{N}$:

$$a \leq w$$

$$w \leq w$$

$$w + a = w$$

$$w - a = w$$

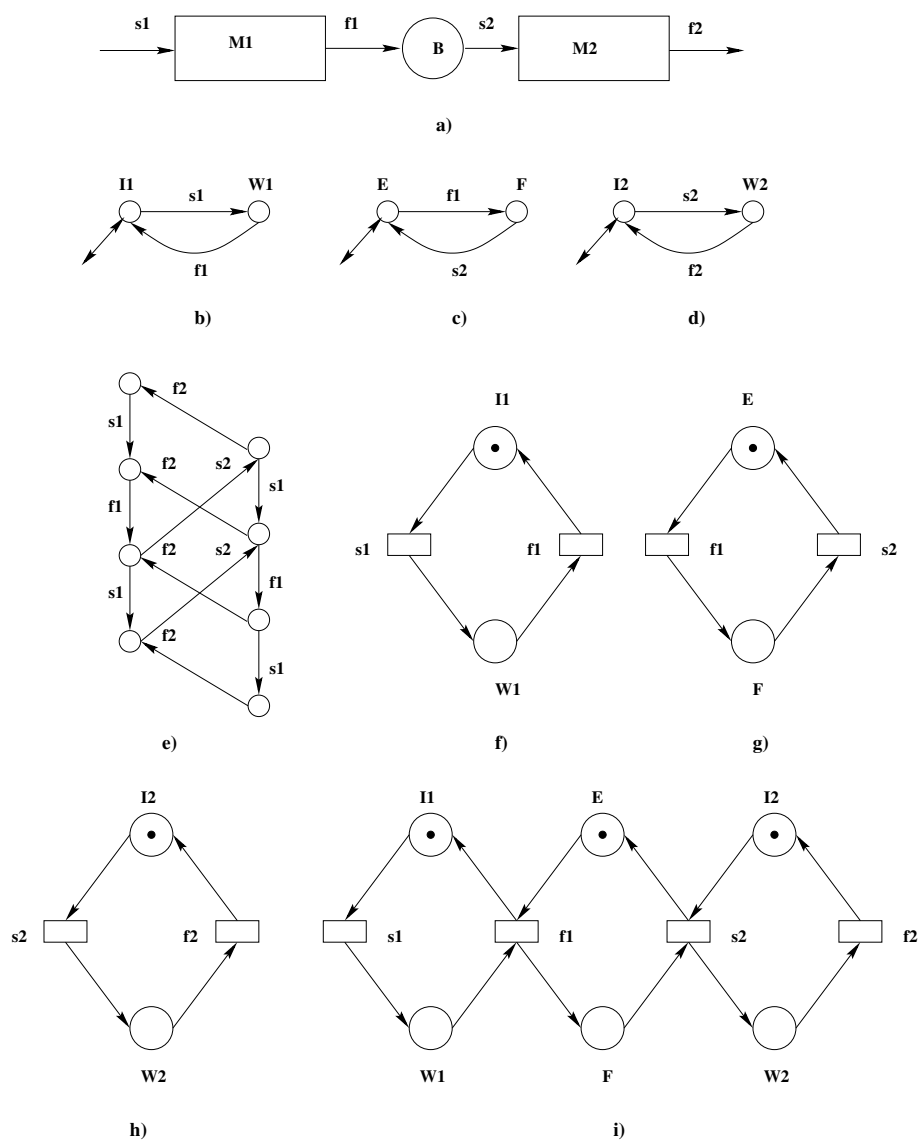


Figura 1.5.: Confronto tra automi e reti di Petri

L'introduzione di w consente di rappresentare un numero di gettoni arbitrariamente grande e quindi di compattare in un nodo del grafo un numero infinito di stati.

Si introduce inoltre la seguente relazione di copertura tra due marcature di una rete e si dice che \mathbf{x} copre \mathbf{y} , ($\mathbf{x} >_c \mathbf{y}$), se le seguenti condizioni sono verificate:

- $\mathbf{x}(p_i) \geq \mathbf{y}(p_i), \forall i \in \{1, \dots, |P|\}$;
- $\mathbf{x}(p_i) > \mathbf{y}(p_i)$ almeno per un indice $i \in \{1, \dots, |P|\}$.

Algoritmo 1 *Costruzione dell'albero di copertura.*

Step 1 *Etichetta $\langle new \rangle$ la marcatura iniziale \mathbf{m}_0 che costituisce la radice dell'albero di copertura*

Step 2 **while** *una marcatura con etichetta $\langle new \rangle$ esiste* **do**

begin

Step 2.1 *Seleziona una marcatura \mathbf{m} con etichetta $\langle new \rangle$*

Step 2.2 *Se nessuna transizione risulta abilitata nel sistema $\langle N, \mathbf{m} \rangle$ etichetta $\langle deadend \rangle$ la marcatura \mathbf{m}*

Step 2.3 *Se \mathbf{m} risulta uguale ad una marcatura \mathbf{m}' sul percorso da \mathbf{m}_0 a \mathbf{m} etichetta $\langle old \rangle$ \mathbf{m} e passa ad una nuova marcatura con etichetta $\langle new \rangle$*

Step 2.4 *Per ogni transizione abilitata nella marcatura \mathbf{m}* **do**

begin

Step 2.4.1 *ottieni una nuova marcatura \mathbf{m}'*

Step 2.4.2 *se esiste una marcatura \mathbf{m}'' sul percorso da \mathbf{m}_0 a \mathbf{m} che risulta coperta da \mathbf{m}' (ovvero, $\mathbf{m}' >_c \mathbf{m}''$) allora per ogni i tale che $\mathbf{m}'(p_i) > \mathbf{m}''(p_i)$ poni $\mathbf{m}'(p_i) = w$*

Step 2.4.3 *aggiungi all'albero il nodo \mathbf{m}' etichettando l'arco che congiunge \mathbf{m} a \mathbf{m}' con la rispettiva transizione*

Step 2.4.4 *denomina $\langle new \rangle$ \mathbf{m}'*

end

end

In fig. 1.6 è mostrata una PN ed il relativo albero di copertura.

Si noti come l'introduzione del simbolo w comporta una perdita di informazione, come mostrato in fig. 1.7, dove due diversi sistemi PN hanno lo stesso grafo di copertura.

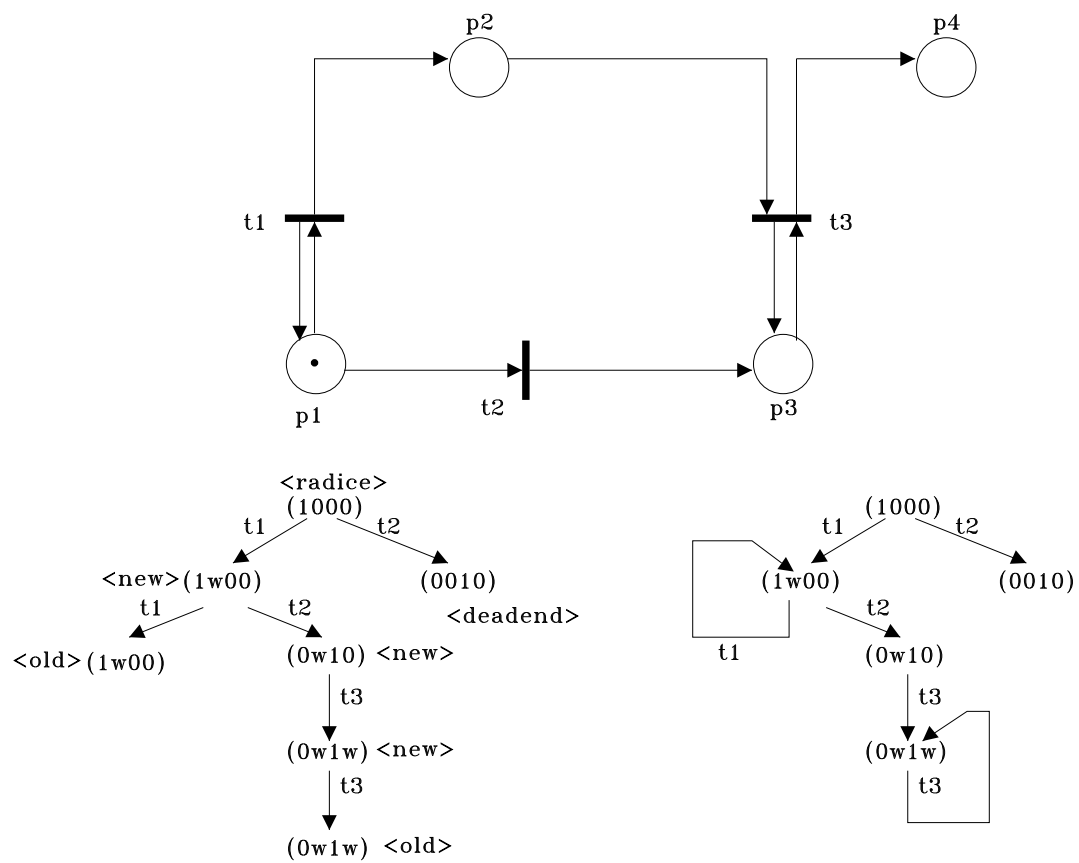


Figura 1.6.: Un sistema PN e relativo grafo di copertura.

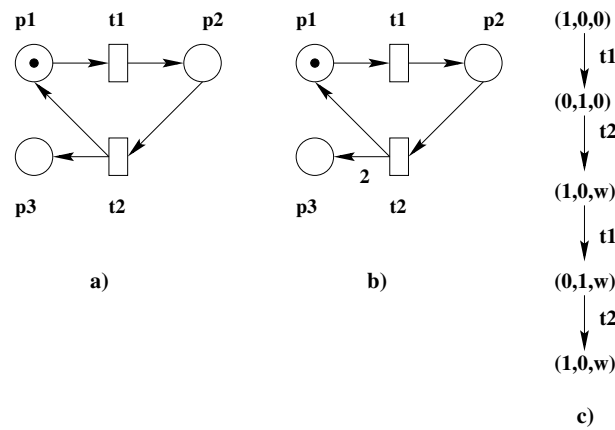


Figura 1.7.: I due sistemi PN in a) e b) hanno lo stesso grafo di copertura; nota inoltre che il simbolo w per il sistema in a) rappresenta l'insieme $\{1, 2, 3, \dots\}$, mentre per il sistema in b) l'insieme $\{2, 4, \dots\}$.

1.5. Classi fondamentali di PN

Si introducono le seguenti classi fondamentali di PN, riassunte in fig. 1.8.

Si dice *state machine (SM)* una rete di Petri ordinaria tale che:

$$\forall t \in T, |\bullet t| = 1 \quad e \quad |t\bullet| = 1$$

ovvero ogni transizione ha esattamente un posto in ingresso ed un posto in uscita.

Si dice *marked graph (MG)* una rete di Petri ordinaria tale che:

$$\forall p \in P, |\bullet p| = 1 \quad e \quad |p\bullet| = 1$$

ovvero ogni posto ha esattamente una transizione in ingresso ed una transizione in uscita.

Si dice *free-choice net (FCN)* una rete di Petri ordinaria tale che:

$$\forall p \in P, |p\bullet| \leq 1 \vee \bullet(p\bullet) = \{p\}$$

ovvero un arco in uscita da un posto p verso una transizione t o è l'unico arco in uscita da p o è l'unico arco in ingresso a t .

Si dice *simple net (SN)* una rete di Petri ordinaria tale che:

$$\forall t \in T, |\{p \in \bullet t, \text{ tale che } |p\bullet| > 1\}| \leq 1$$

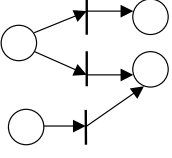
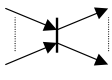
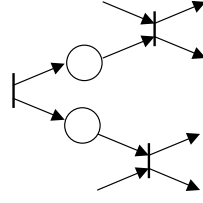
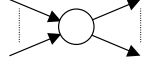
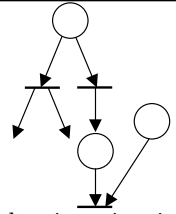
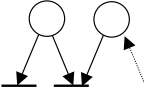
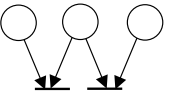
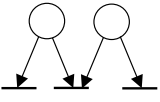
	SI	NO
SM State Machine	 <p>scelta</p>	 <p>sincronizzazione concorrenza</p>
MG Marked Graphs	 <p>sincronizzazione concorrenza</p>	 <p>scelta</p>
FCN Free Choice Net	 <p>scelta, sincronizzazione, concorrenza (la stessa transizione non può essere insieme di scelta e sincronizzazione)</p>	 <p>semaforo</p>
SN Simple Net	 <p>scelta, sincronizzazione, concorrenza, semaforo</p>	 <p>le scelte sono libere ma devono essere risolte localmente (non è possibile far dipendere una scelta da un'altra – confusione)</p>

Figura 1.8.: Principali classi di reti di Petri

Le relazioni tra il diverso potere espressivo delle classi sopra elencate è riassunto in fig. 1.9.

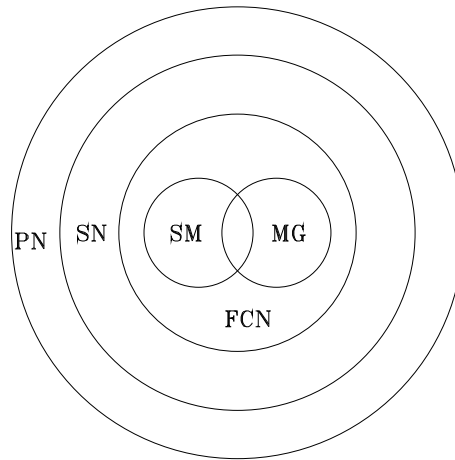


Figura 1.9.: Confronto tra il potere espressivo delle differenti classi di PN ordinarie

1.6. Caratterizzazioni lineari di una PN

L'equazione di stato non include la condizione di abilitazione di una sequenza di transizioni σ , ma solo la marcatura risultante dallo scatto di una delle sequenze il cui vettore di scatto risulta essere σ (si tenga presente infatti che il vettore di scatto $\sigma = [1 \ 2 \ 0 \ 1]$ può essere relativo ad entrambe le sequenze $\sigma_1 = t_2, t_1, t_2, t_4$ o $\sigma_2 = t_2, t_2, t_1, t_4$).

Se pertanto si considera l'insieme delle marcature appartenenti all'insieme

$$LRS^{SE}(\mathcal{S}) = \{\mathbf{m} \in \mathbb{N}^{|P|} \mid \exists \sigma \in \mathbb{N}^{|T|} : \mathbf{m} = \mathbf{m}_0 + \mathbf{C} \sigma\} \quad (1.3)$$

detto *insieme linearizzato di raggiungibilità* si troveranno alcune soluzioni *spurie*. Infatti si consideri la rete in fig. 1.10. Tutti i nodi tratteggiati del grafo non appartengono all'insieme di raggiungibilità della rete ma soddisfano l'equazione di stato. Alcuni di essi sono addirittura soluzioni il cui vettore di scatto non corrisponde ad alcuna possibile sequenza di transizioni. Ad esempio il nodo $2p_4$ (la notazione indicata è la seguente: $2p_1$ sta per $[2 \ 0 \ 0 \ 0 \ 0]$, $p_2 + p_3$ denota $[0 \ 1 \ 1 \ 0 \ 0]$, etc.) è soluzione dell'equazione di stato con vettore di scatto

pari a $\sigma^* = [3 \ 2 \ 0 \ 0 \ 1 \ 1]$. Dal momento che le uniche transizioni in uscita da p_4 sono t_3 e t_4 e sono entrambe assenti nel vettore di scatto σ^* , si comprende che la marcatura non solo è spuria ma non risulta in alcun modo raggiungibile dalla marcatura iniziale, ovvero non ha alcun nodo antecedente.

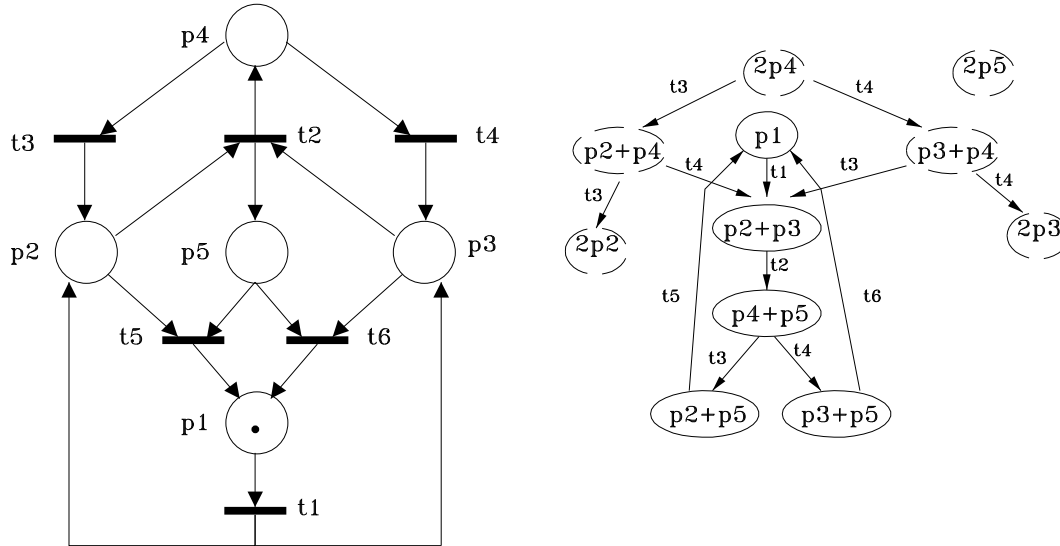


Figura 1.10.: Un sistema PN e relativo LRS^{SE}

Esempio 4 La matrice di incidenza della rete in fig. 1.11 è

$$C = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}$$

Si può verificare che la marcatura iniziale $\mathbf{m}_{p_0} = [0 \ 0]$ e il vettore di scatto $\sigma = [1 \ 1]$ verificano l'equazione di stato

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} + \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

ovvero la marcatura iniziale risulta essere raggiungibile da se stessa attraverso la sequenza $\sigma = t_1, t_2$, che è una delle possibili per la rete in questione, ma è evidente che a partire da \mathbf{m}_0 nessuna sequenza è abilitata. Tale comportamento è determinato dal fatto che l'equazione di stato è di fatto un bilancio del flusso di gettoni attraverso la rete senza tenere conto delle condizioni di abilitazione di una transizione. Quindi

ogni volta che si ha un ciclo come nel caso in esame, essa risulta verificata da ogni sequenza che includa tutte le transizioni del ciclo, indipendentemente dal fatto che esse siano o meno abilitate.

Se si volesse descrivere in forma di un sistema di equazioni tutte le marcature raggiungibili di una PN a partire dalla condizione iniziale, nel caso che la rete fosse *bounded* e fosse possibile calcolare un intero k tale che ogni stato è raggiungibile in al più k passi, le soluzioni $(\mathbf{m}_i \in \mathbb{N}^{|P|}, \boldsymbol{\sigma}_i \in \mathbb{N}^{|T|})$ del sistema

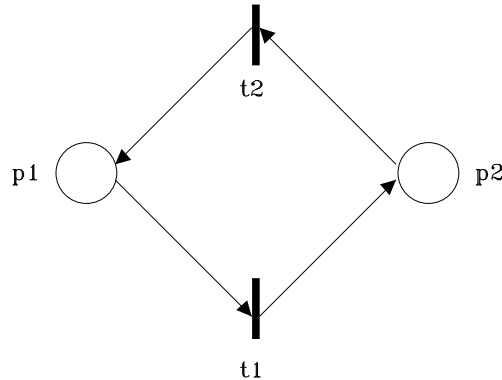


Figura 1.11.: PN in esempio 4

$$\begin{aligned}
 \mathbf{m}_0 &\geq \mathbf{Pre} \boldsymbol{\sigma}_0 \\
 \mathbf{m}_1 &= \mathbf{m}_0 + \mathbf{C} \boldsymbol{\sigma}_0 \\
 &\dots\dots\dots \\
 \mathbf{m}_{k-1} &\geq \mathbf{Pre} \boldsymbol{\sigma}_{k-1} \\
 \mathbf{m}_k &= \mathbf{m}_{k-1} + \mathbf{C} \boldsymbol{\sigma}_{k-1}
 \end{aligned}$$

sarebbero tutte e solo le marcature raggiungibili dallo stato iniziale.

Si comprende che il sistema precedente risulta tutt'altro che semplice e compatto, inoltre è di tipo non lineare e richiede un'analisi precedente per determinare in quanti passi al più ogni stato è raggiungibile. Questo è il motivo per cui l'insieme 1.3 è considerato una prima approssimazione di $R(\mathcal{S})$. In alcuni casi, data la minore complessità computazionale si prende in esame anche una forma rilassata di tale

insieme

$$LRS^{SER}(\mathcal{S}) = \{\mathbf{m} \in \mathbb{N}^{|P|} \mid \exists \boldsymbol{\sigma} \geq \mathbf{0} : \mathbf{m} = \mathbf{m}_0 + \mathbf{C} \boldsymbol{\sigma}\} \quad (1.4)$$

ottenuta rimuovendo il vincolo che $\boldsymbol{\sigma}$ sia un vettore di interi.

In fig. 1.12 sono mostrati vari esempi di linearizzazione dell'insieme di raggiungibilità. In particolare è considerato anche il caso in cui sia la marcatura che il vettore di scatto possano essere vettori di reali: in tal caso l'insieme di raggiungibilità si indica con la notazione $CRS(N, \mathbf{m}_0)$. In tal caso una transizione t è abilitata nella marcatura \mathbf{m} di una frazione $\lambda > 0$ se $\mathbf{m} \geq \lambda \mathbf{Pre}(\cdot, t)$ e in caso di scatto la rete si porta nella marcatura $\mathbf{m}' = \mathbf{m} + \lambda \mathbf{C}(\cdot, t)$. La marcatura p_2 appartiene all'insieme LRS^{SE} con riferimento al sistema PN in alto a sinistra in fig. 1.12 con vettore di scatto $\boldsymbol{\sigma} = [1 \ 1 \ 1]$. La marcatura $2p_2$ appartiene all'insieme LRS^{SER} con riferimento al sistema PN in basso a sinistra in fig. 1.12 con vettore di scatto $\boldsymbol{\sigma} = [0.5 \ 1 \ 1]$. La marcatura $2p_2$ appartiene all'insieme CRS con riferimento al sistema PN in alto a destra in fig. 1.12 con vettore di scatto $\boldsymbol{\sigma} = [0.5 \ 1]$. La marcatura p_2 appartiene all'insieme $CRS \cap LRS^{SE}$ con riferimento al sistema PN in basso a destra in fig. 1.12 con vettore di scatto $\boldsymbol{\sigma} = [1 \ 1]$ o $\boldsymbol{\sigma} = [0 \ 0.5]$.

1.7. Invarianti di una PN

Un vettore di interi $\mathbf{x} \in \mathbb{N}^{|P|}$ e tale che $\mathbf{x} \succeq \mathbf{0}$ (avente elementi interi non negativi ed almeno una componente diversa da zero) si dice

- *P-invariante* se $\mathbf{x}^T \mathbf{C} = \mathbf{0}$;
- *P-vettore crescente* se $\mathbf{x}^T \mathbf{C} \geq \mathbf{0}$;
- *P-vettore decrescente* se $\mathbf{x}^T \mathbf{C} \leq \mathbf{0}$.

In particolare si noti che

$$\mathbf{x}^T \mathbf{C} = \mathbf{0} \rightarrow \mathbf{x}^T \mathbf{m} = \mathbf{x}^T \cdot \mathbf{m}_0 + \mathbf{x}^T \cdot \mathbf{C} \boldsymbol{\sigma} = \mathbf{x}^T \cdot \mathbf{m}_0 = k, \forall \mathbf{m} \in \mathbf{R}(\mathcal{S}). \quad (1.5)$$

Pertanto gli elementi di un P-invariante rappresentano i pesi di una somma pesata delle marcature dei posti della rete che si mantiene costante nello spazio di stato della PN ed il valore di tale costante dipende dalla marcatura iniziale.

Per quanto riguarda la relazione inversa si può provare che, se tutte le transizioni possono scattare almeno una volta, allora, se vale che $\mathbf{x}^T \cdot \mathbf{m} = \mathbf{x}^T \cdot \mathbf{m}_0, \forall \mathbf{m} \in \mathbf{R}(\mathcal{S})$, si ha che $\mathbf{x}^T \cdot \mathbf{C} \cdot \boldsymbol{\sigma} = \mathbf{0}$ e quindi $\mathbf{x}^T \cdot \mathbf{C} = \mathbf{0}$.

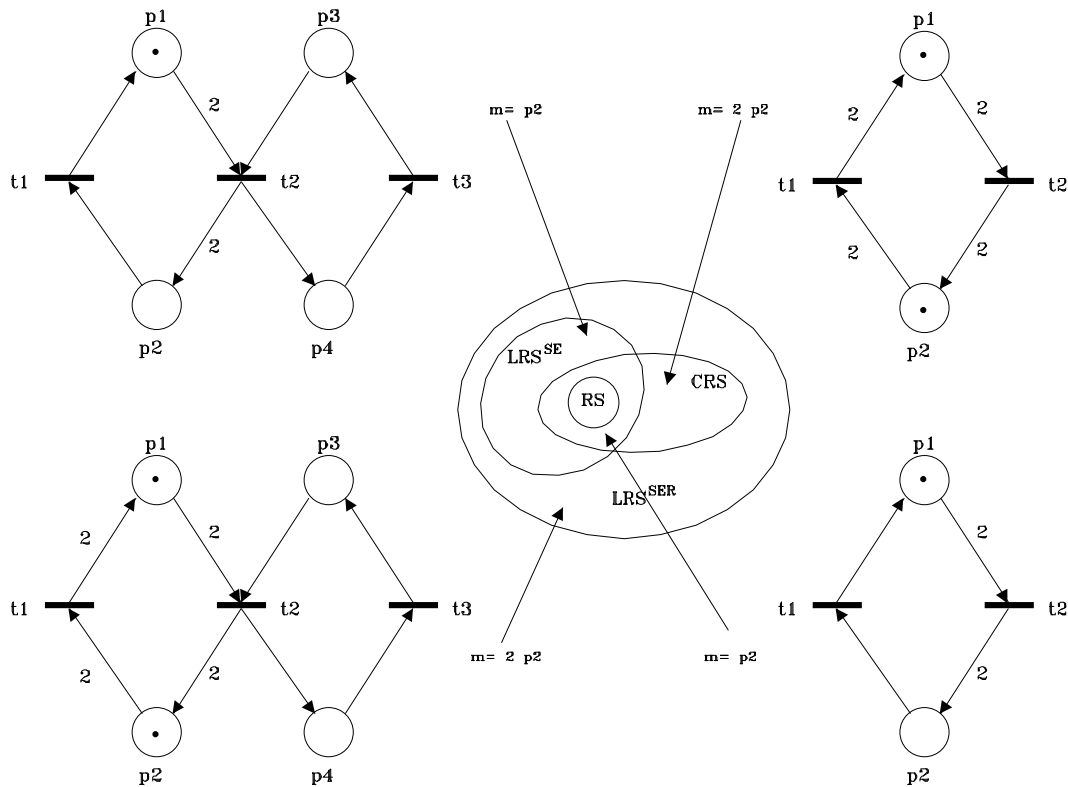


Figura 1.12.: Confronto tra differenti linearizzazioni dell'insieme di raggiungibilità

L'insieme

$$\| \mathbf{x} \| = \{p \in P \mid \mathbf{x}(p) > 0\},$$

è detto supporto del P-invariante \mathbf{x} .

Inoltre a ciascun P-invariante può essere associata la sottorete costituita dai posti in $\| \mathbf{x} \|$: in tal modo si ottiene, come illustrato in fig. 1.13, una decomposizione strutturale della rete. La rete in fig. 1.13 ha i 2 P-invarianti $\mathbf{x}_1^T = [1 \ 1 \ 0 \ 1 \ 0]$ ed $\mathbf{x}_2^T = [1 \ 0 \ 1 \ 0 \ 1]$ a cui sono associati i supporti $\| \mathbf{x}_1 \| = \{p_1, p_2, p_4\}$ e $\| \mathbf{x}_2 \| = \{p_1, p_3, p_5\}$. L'insieme della sottorete P-invariante e della marcatura iniziale dei suoi posti prende il nome di sottosistema P-invariante.

Un P-invariante \mathbf{x} è detto

- *minimale* se non esiste un P-invariante \mathbf{x}' tale che $\mathbf{x}' \leq \mathbf{x}$;

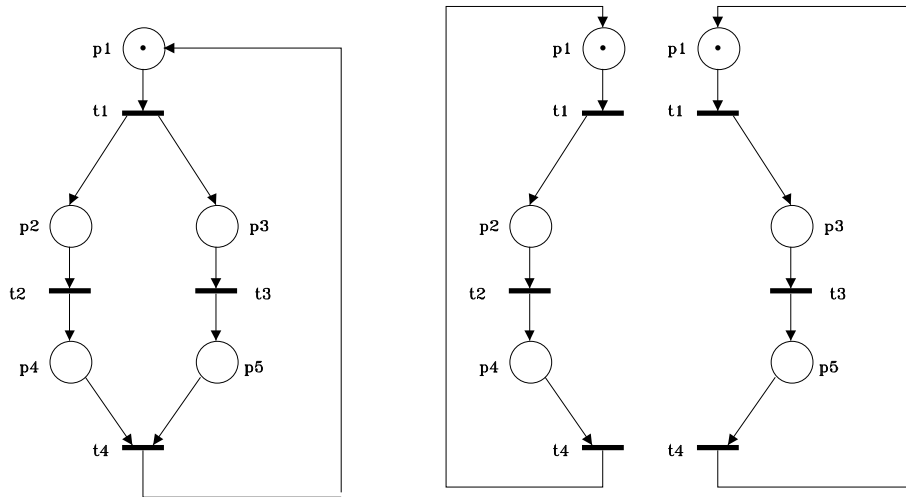


Figura 1.13.: Un sistema PN e la relativa decomposizione in termini di sottosistemi P-invarianti.

- *di supporto minimo* se il suo supporto non contiene quello di nessun altro P-invariante.

L'insieme dei P-invarianti minimali e di supporto minimo di una rete è unico ed è detto *insieme di P-invarianti fondamentali*.

Ogni P-invariante può essere ottenuto come combinazione lineare a coefficienti interi positivi di P-invarianti fondamentali.

Detta Φ la matrice avente come colonne i P-invarianti fondamentali, è possibile ottenere la seguente forma linearizzata dello spazio di raggiungibilità:

$$LRS^{Pinv}(\mathcal{S}) = \{\mathbf{m} \in \mathbf{N}^{|P|} \mid \Phi^T \cdot \mathbf{m} = \Phi^T \cdot \mathbf{m}_0\} \quad (1.6)$$

Si dimostra che:

$$\mathbf{R}(\mathcal{S}) \subseteq LRS^{SE}(\mathcal{S}) \subseteq LRS^{SER}(\mathcal{S}) \subseteq LRS^{Pinv}(\mathcal{S}). \quad (1.7)$$

Il grafo in fig. 1.10 è rappresentativo oltre che di LRS^{SE} anche di LRS^{Pinv} dal momento che si dimostra che per questa rete $LRS^{Pinv} = LRS^{SE}$. Avendo tale rete l'unico P-invariante $\mathbf{x} = [2 \ 1 \ 1 \ 1 \ 1]$ ne consegue che $LRS^{SE} = LRS^{Pinv}(\mathcal{S}) = \{\mathbf{m} \in \mathbf{N}^{|P|} \mid 2m(p_1) + m(p_2) + m(p_3) + m(p_4) + m(p_5) = 2\}$.

Un concetto duale ai P-invarianti è costituito dai *T-invarianti*. Un vettore di interi $\mathbf{y} \in \mathbf{N}^{|T|}$ e tale che $\mathbf{y} \succeq \mathbf{0}$ (avente elementi interi non negativi ed almeno una componente diversa da zero) si dice

- *T-invariante* se $\mathbf{C}\mathbf{y} = \mathbf{0}$;
- *T-vettore crescente* se $\mathbf{C}\mathbf{y} \geq \mathbf{0}$;
- *T-vettore decrescente* se $\mathbf{C}\mathbf{y} \leq \mathbf{0}$.

Si ha che se \mathbf{y} è un *T-invariante*

$$\mathbf{m} = \mathbf{m}_0 + \mathbf{C} \cdot \mathbf{y} = \mathbf{m}_0$$

Pertanto, un T-invariante rappresenta un vettore di scatto relativo a sequenze che riconducono la rete allo stato iniziale: la rete può presentare un comportamento ciclico. Non è però detto, infatti, che tali sequenze siano scattabili sotto una qualunque marcatura iniziale.

L'insieme

$$\|\mathbf{y}\| = \{t \in T \mid \mathbf{y}(t) > 0\},$$

è detto supporto del T-invariante \mathbf{y} .

Nel seguito viene fornito un semplice algoritmo per la determinazione dell'insieme di P-invarianti fondamentali.

Algoritmo 2 *Determinazione dell'insieme di P-invarianti fondamentali.*

1. Si costruisce la tabella $\mathbf{A} := [\mathbf{C} \mid \mathbf{I}_{m \times n}]$, dove $\mathbf{I}_{m \times m}$ è la matrice identità di dimensione m .
2. Per $j := 1, \dots, n$ (indice di colonna associato alle transizioni):
 - (a) sia $I_+ := \{i \mid A(i, j) > 0\}$ l'insieme degli indici di riga che corrispondono ad elementi positivi della colonna j ;
 - (b) sia $I_- := \{i \mid A(i, j) < 0\}$ l'insieme degli indici di riga che corrispondono ad elementi negativi della colonna j ;
 - (c) per ogni coppia $(i_+, i_-) \in I_+ \times I_-$:
 - i. sia $d := \text{mcm}\{A(i_+, j), -A(i_-, j)\}$ il minimo comune multiplo dei due elementi $A(i_+, j)$ e $-A(i_-, j)$;

- ii. sia $d_+ := d/A(i_+, j)$ e $d_- := d/A(i_-, j)$;
 - iii. si aggiunge alla tabella la nuova riga: $d_+A(i_+, \cdot) + d_-A(i_-, \cdot)$, combinazione lineare delle righe di indice i_+ e i_- (la nuova riga ha elemento j -mo pari a zero per costruzione);
 - (d) si cancellano da A le righe di indice $I_+ \cup I_-$, corrispondenti a elementi diversi da zero lungo la j -ma colonna.
3. La tabella A risultante è della forma $[\mathbf{0}_{r \times m} \mid \mathbf{X}^T]$, dove $\mathbf{0}_{r \times m}$ una matrice di zeri $r \times m$, mentre X è una matrice con m righe e r colonne. Ogni colonna di X è un P -invariante a supporto minimo. In realtà tali colonne sono anche P -invarianti minimali a meno di una eventuale divisione degli elementi di ciascuna colonna di X per il loro massimo comun divisore.

Nel precedente algoritmo si noti che se uno qualunque dei due insiemi I_+ o I_- è vuoto, al passo 2(c) non viene aggiunta alcuna riga, mentre al passo 2(d) vengono comunque eliminate le righe corrispondenti a elementi diversi da zero lungo la j -ma colonna se $I_+ \cup I_- \neq \emptyset$. Inoltre la tabella risultante sarà vuota (cioè $r = 0$) qualora la rete N non abbia alcun P -invariante.

Per il calcolo dei T -invarianti si può utilizzare lo stesso algoritmo avendo cura di inizializzare la tabella A al passo 1 ponendo $A := [C^T \mid \mathbf{I}_{m \times n}]$.

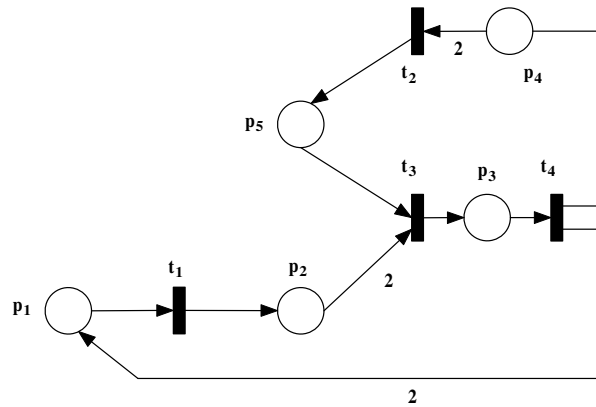


Figura 1.14.: Rete utilizzata nell'esempio 5.

Esempio 5 *In questo esempio viene utilizzato l'algoritmo per determinare l'insieme di P-invarianti fondamentali della rete in fig. 1.14. Si costruisca inizialmente la tabella*

$$\left[\begin{array}{cccc|ccccc} -1 & 0 & 0 & 2 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & -2 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 1 & 0 & 0 \\ 0 & -2 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 1 \end{array} \right] \begin{array}{l} p_1 \\ p_2 \\ p_3 \\ p_4 \\ p_5 \end{array}$$

Al passo $j = 1$ al fine di rendere nulli gli elementi della prima colonna della tabella si esegue la somma delle righe p_1 e p_2 , che vengono successivamente cancellate; viene quindi inserita la riga $p_1 + p_2$ ottenendo

$$\left[\begin{array}{cccc|ccccc} 0 & 0 & 1 & -1 & 0 & 0 & 1 & 0 & 0 \\ 0 & -2 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & -2 & 2 & 1 & 1 & 0 & 0 & 0 \end{array} \right] \begin{array}{l} p_3 \\ p_4 \\ p_5 \\ p_1 + p_2 \end{array}$$

Al passo $j = 2$ al fine di rendere nulli gli elementi della seconda colonna della tabella si esegue la somma delle righe p_4 e p_5 moltiplicata per 2, che vengono successivamente cancellate; viene quindi inserita la riga $p_4 + 2p_5$ ottenendo

$$\left[\begin{array}{cccc|ccccc} 0 & 0 & 1 & -1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & -2 & 2 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & -2 & 1 & 0 & 0 & 0 & 1 & 2 \end{array} \right] \begin{array}{l} p_3 \\ p_1 + p_2 \\ p_4 + 2p_5 \end{array}$$

Al passo $j = 3$ al fine di rendere nulli gli elementi della terza colonna della tabella si esegue la somma delle righe p_3 moltiplicata per 2 e della riga $p_1 + p_2$ oppure la somma delle righe p_3 moltiplicata per 2 e della riga $p_4 + 2p_5$; procedendo poi con le cancellazioni si perviene a

$$\left[\begin{array}{cccc|ccccc} 0 & 0 & 0 & 0 & 1 & 1 & 2 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 2 & 1 & 2 \end{array} \right] \begin{array}{l} p_1 + p_2 + 2p_3 \\ 2p_3 + p_4 + 2p_5 \end{array}$$

Al passo $j = 4$ non risulta possibile rendere nulla la quarta colonna a meno della cancellazione della riga $2p_3 + p_4 + 2p_5$, per l'algoritmo termina restituendo la tabella

$$\left[\begin{array}{cccc|ccccc} 0 & 0 & 0 & 0 & 1 & 1 & 2 & 0 & 0 \end{array} \right] \begin{array}{l} p_1 + p_2 + 2p_3 \end{array}$$

Si deduce che la rete ha l'unico invariante $\mathbf{x} = [1 \ 1 \ 2 \ 0 \ 0]^T$.

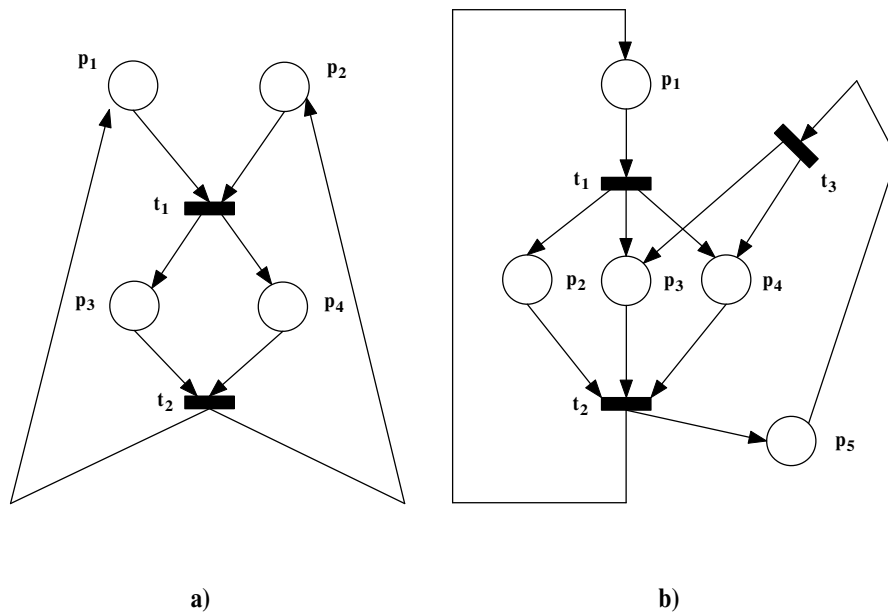


Figura 1.15.: Rete utilizzata nell'esempio 6.

Si noti come la cardinalità dell'insieme dei P-invarianti fondamentali non è legata al rango della matrice di incidenza dalla relazione $m - \text{rango}(\mathbf{C})$ come farebbe pensare il fatto che l'insieme dei P-invarianti fondamentali costituisce una base del nullo della matrice \mathbf{C}^T . Si dimostra che la cardinalità dell'insieme dei P-invarianti fondamentali è limitata dalla quantità $\binom{m}{\lfloor \frac{m}{2} \rfloor}$, dove $m = |P|$ e $\lfloor x \rfloor$ indica la parte intera di x . Tale risultato è dovuto alla proprietà che l'insieme dei P-invarianti fondamentali genera ogni altro P-invariante semplicemente mediante somme e moltiplicazioni per interi positivi.

Esempio 6 La rete in figura 1.15 (a) ha una matrice di incidenza con rango 1 ma l'insieme dei P-invarianti fondamentali ha cardinalità pari a $4 \neq m - 1 = 3$ ed è costituito dai vettori $[1 \ 0 \ 0 \ 1]^T$, $[1 \ 0 \ 1 \ 0]^T$, $[0 \ 1 \ 0 \ 1]^T$, $[0 \ 1 \ 1 \ 0]^T$.

La rete in figura 1.15 (b) ha una matrice di incidenza con rango 3 ma l'insieme dei P-invarianti fondamentali ha cardinalità pari a $1 \neq m - 3 = 2$ ed è costituito dal vettore $[1 \ 1 \ 0 \ 0 \ 0]^T$.

2. Le proprietà fondamentali di un sistema PN

Questo capitolo illustra le principali proprietà di un sistema PN; esse si dividono in proprietà *comportamentali*, se dipendono dallo stato iniziale, e *strutturali*, se non dipendono dallo stato iniziale. La verifica di tali proprietà viene condotta avvalendosi degli strumenti di analisi introdotti nel capitolo precedente.

2.1. Le proprietà comportamentali

Una proprietà di un sistema PN si dice comportamentale se dipende dallo stato iniziale.

2.1.1. Raggiungibilità

Il problema della raggiungibilità è quello di verificare se, dato un sistema \mathcal{S} ed una marcatura \mathbf{m} , esiste una sequenza σ tale che $\mathbf{m}_0[\sigma > \mathbf{m}$, ovvero se $\mathbf{m} \in \mathbf{R}(\mathcal{S})$.

Per la verifica della raggiungibilità si hanno fondamentalmente due strumenti, il grafo di copertura e l'equazione di stato.

Se la rete ha uno spazio di stato finito il grafo di copertura, non contenendo il simbolo w , ci consente di dire che una certa marcatura è raggiungibile se e solo se essa è un nodo di tale grafo. In caso contrario si può solo verificare se tale marcatura è coperta da un nodo di tale grafo.

La verifica dell'equazione di stato rappresenta in generale solo una condizione necessaria per la risoluzione del problema della raggiungibilità. Solo nel caso di reti acicliche, dei marked graph vivi (la definizione di vivezza sarà fornito nel seguito) e delle state machine possiamo dire che $\mathbf{m} \in \mathbf{R}(\mathcal{S})$ se e solo se esiste un vettore di scatto che verifica la (1.2).

In generale, per risolvere il problema della raggiungibilità va risolto il sistema $\Delta\mathbf{m} = \mathbf{C}\boldsymbol{\sigma}$, con $\Delta\mathbf{m} = \mathbf{m} - \mathbf{m}_0$ e $\boldsymbol{\sigma} \in \mathbf{N}^n$. Si può dimostrare che tale problema

avrà soluzione se $\Delta \mathbf{m}$ risulterà ortogonale alle soluzioni \mathbf{x} del sistema omogeneo $\mathbf{C}^T \mathbf{x} = \mathbf{0}$.

Data la matrice di incidenza di una PN \mathbf{C} si supponga di partizionarne la trasposta \mathbf{C}^T come segue:

$$\mathbf{C}^T = \begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{bmatrix} \quad (2.1)$$

con \mathbf{C}_{12} matrice quadrata ed invertibile di dimensione $r \times r$ con r pari al rango di \mathbf{C} .

Si dimostra che, posto

$$\mathbf{B}_f = [\mathbf{I} \quad -\mathbf{C}_{11}^T (\mathbf{C}_{12}^T)^{-1}]$$

con \mathbf{I} matrice identità di ordine $m - r$, una base per lo spazio delle soluzioni del sistema $\mathbf{C}^T \mathbf{x} = \mathbf{0}$ è dato dalle colonne di \mathbf{B}_f . Imponendo l'ortogonalità tra tali soluzioni e $\Delta \mathbf{m}$, si ottiene

$$\mathbf{m} \in \mathbf{R}(\mathcal{S}) \Rightarrow \mathbf{B}_f \cdot \Delta \mathbf{m} = \mathbf{0}. \quad (2.2)$$

In particolare nel caso dei marked graph si ha che

$$\mathbf{m} \in \mathbf{R}(\mathcal{S}) \Leftrightarrow \mathbf{B}_p^T \cdot \Delta \mathbf{m} = \mathbf{0} \quad (2.3)$$

dove \mathbf{B}_p è una matrice le cui colonne costituiscono l'insieme dei P-invarianti fondamentali, ovvero lo spazio di raggiungibilità di un marked graph è linearizzabile. Va sottolineato che \mathbf{B}_f può essere diversa da \mathbf{B}_p^T dal momento che può contenere elementi negativi o razionali.

Esempio 7 Si consideri il sistema in fig. 2.1. A partire dallo stato iniziale $\mathbf{m}_0 = [2 \ 0 \ 1 \ 0]^T$ in seguito allo scatto di t_3 , che risulta abilitata, la rete si porta nello stato $\mathbf{m}_1 = [3 \ 0 \ 0 \ 2]^T$ come risulta dall'equazione di stato.

$$\begin{bmatrix} 3 \\ 0 \\ 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ 1 \\ 0 \end{bmatrix} + \begin{bmatrix} -2 & 1 & 1 \\ 1 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & -2 & 2 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

La matrice di incidenza della rete \mathbf{C} ha rango 2 e la sua trasposta può essere partizionata in base alla (2.1) ottenendo:

$$\mathbf{C}_{11} = \begin{bmatrix} -2 & 1 \\ 1 & -1 \end{bmatrix}, \quad \mathbf{C}_{12} = \begin{bmatrix} 1 & 0 \\ 0 & -2 \end{bmatrix}$$

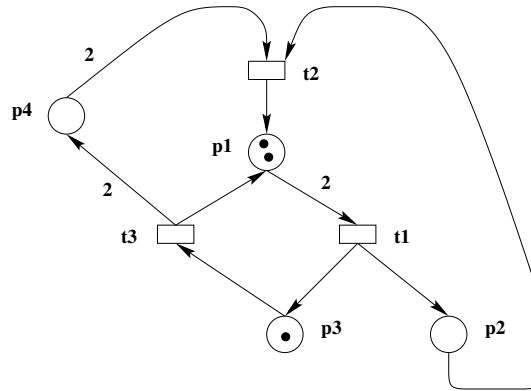


Figura 2.1.: Sistema PN dell'esempio 7

Da cui:

$$B_f = \begin{bmatrix} 1 & 0 & 2 & 1/2 \\ 0 & 1 & -1 & -1/2 \end{bmatrix}$$

Si può verificare che $B_f \cdot \Delta \mathbf{m} = \mathbf{0}$, con $\Delta \mathbf{m} = \mathbf{m}_1 - \mathbf{m}_0$.

2.1.2. Boundedness

Una sistema PN si dice *k-bounded* (k-limitato) se il numero di gettoni in ogni posto della rete non eccede un numero finito k nel corso dell'evoluzione della rete a partire dalla sua marcatura iniziale \mathbf{m}_0 , ovvero

$$\forall \mathbf{m} \in R(\mathcal{S}), \forall p \in P, \mathbf{m}(p) \leq k.$$

Se k è pari ad uno allora la rete si dice *safe*. La boundedness è di fondamentale importanza nei sistemi di automazione, sia per un motivo di sicurezza, sia per motivi gestionali quali ad esempio il dimensionamento dei buffer.

Per verificare la boundedness di un sistema PN si può utilizzare il grafo di copertura. Se nel corso del calcolo di tale grafo appare il simbolo w la rete non è limitata, altrimenti la rete si dice *k-bounded* dove k il massimo valore delle componenti di tutti i nodi del grafo che in tal caso un grafo di raggiungibilità.

Nel caso in cui la rete è non limitata si può parlare di *k-limitatezza* di un posto. Per definire tale proprietà è necessario introdurre il concetto di *w-copertura*. Un nodo del grafo di copertura in cui compare l'elemento w si denota \mathbf{m}_w . Data una marcatura $\mathbf{m} \in \mathbf{N}^m$, si dice che essa *w-coperta* da \mathbf{m}_w se $\mathbf{m}_w(p) = \mathbf{m}(p)$ per

ogni componente p tale che $\mathbf{m}_w(p) \neq w$; tale relazione si denota $\mathbf{m}_w \geq_w \mathbf{m}$. Un posto p della rete si dice k -limitato se per ogni nodo del grafo di copertura vale $\mathbf{m}_w(p) \leq k \neq w$.

Una condizione sufficiente per la k -limitatezza è rappresentata dall'equazione di stato, ovvero una rete è limitata se:

$$\forall \sigma, \mathbf{m} \geq \mathbf{0}, \quad \max \mathbf{m} \leq \mathbf{k}, \quad \mathbf{k} \in \mathbf{N}^m, \quad \text{con } \mathbf{m} = \mathbf{m}_0 + \mathbf{C} \sigma \quad (2.4)$$

La (2.4) può essere verificata attraverso la programmazione lineare come di seguito esposto.

$$\begin{aligned} & \max_{\mathbf{m}, \sigma} \mathbf{m} \\ \text{s.t. } & \begin{cases} \mathbf{m} = \mathbf{m}_0 + \mathbf{C} \sigma \\ \sigma \in \mathbf{N}^n \\ \mathbf{m} \in \mathbf{N}^m \end{cases} \end{aligned} \quad (2.5)$$

da cui viene che

$$\begin{aligned} & \max_{\mathbf{m}, \sigma} \begin{bmatrix} \mathbf{0}_{m \times n} & \mathbf{I}_{m \times m} \end{bmatrix} \begin{bmatrix} \sigma \\ \mathbf{m} \end{bmatrix} \\ \text{s.t. } & \begin{cases} \begin{bmatrix} \mathbf{C} & -\mathbf{I} \end{bmatrix} \begin{bmatrix} \sigma \\ \mathbf{m} \end{bmatrix} = -\mathbf{m}_0 \\ \begin{bmatrix} \sigma \\ \mathbf{m} \end{bmatrix} \in \mathbf{N}^{n+m} \end{cases} \end{aligned} \quad (2.6)$$

La (2.6) presenta una funzione obiettivo di tipo vettoriale ma può essere ricondotta ad m problemi del tipo

$$\begin{aligned} & \max_{\mathbf{m}, \sigma} \begin{bmatrix} \mathbf{0}_{1 \times n} & \mathbf{e}_p \end{bmatrix} \begin{bmatrix} \sigma \\ \mathbf{m} \end{bmatrix} \\ \text{s.t. } & \begin{cases} \begin{bmatrix} \mathbf{C} & -\mathbf{I} \end{bmatrix} \begin{bmatrix} \sigma \\ \mathbf{m} \end{bmatrix} = -\mathbf{m}_0 \\ \begin{bmatrix} \sigma \\ \mathbf{m} \end{bmatrix} \in \mathbf{N}^{n+m} \end{cases} \end{aligned} \quad (2.7)$$

dove $\mathbf{e}_p : P \rightarrow \{0, 1\}$ è il vettore caratteristico associato al posto p : $\mathbf{e}_p(x) = 1$ se $x = p$, altrimenti $\mathbf{e}_p(x) = 0$. Ogni problema del tipo (2.7) rappresenta un sistema di programmazione lineare intera nella forma standard

$$\begin{aligned} & \max_{\mathbf{x}} \mathbf{c}^T \mathbf{x} \\ \text{s.t. } & \begin{cases} \mathbf{A}\mathbf{x} = \mathbf{b} \\ \mathbf{x} \in \mathbf{N}^n \end{cases} \end{aligned} \quad (2.8)$$

Il vincolo di interezza di σ e \mathbf{m} può essere rimosso, tenendo però presente che, trattandosi di una condizione sufficiente, essa risulterebbe ancor meno restrittiva.

Infine si può dimostrare, a partire dalla (1.5), che un sistema PN è limitato se ciascun posto è coperto da almeno un P-invariante (ovvero se ciascun posto compare nel supporto di almeno un P-invariante); il contrario non è sempre vero.

2.1.3. Liveness

Una sistema PN si dice *deadlock free* se almeno una transizione può sempre scattare. Una transizione si dice *live* (viva) se a partire da qualsiasi marcatura $\mathbf{m} \in R(\mathcal{S})$ esiste una sequenza σ attraverso cui è possibile raggiungere una marcatura \mathbf{m}' , ovvero $\mathbf{m}[\sigma > \mathbf{m}'$, tale che in tale marcatura essa risulta abilitata. Un sistema PN si dice *live* se tutte le transizioni sono *live*. Spesso dato che tale proprietà risulta assai difficile da ottenersi si introducono differenti gradi di *liveness* di una transizione t , precisamente una transizione t si dice:

- *dead* (*L0-live*) se in nessuna marcatura $\mathbf{m} \in R(\mathcal{S})$ risulta abilitata;
- *L1-live* se una transizione può scattare almeno una volta in una marcatura $\mathbf{m} \in R(\mathcal{S})$;
- *L2-live* se, dato un numero intero k , può scattare almeno k volte in alcune sequenze a partire da \mathbf{m}_0 ;
- *L3-live* se è contenuta infinite volte in qualche sequenza a partire da \mathbf{m}_0 ;
- *L4-live* or *live* se una transizione è *L1-live* da qualunque marcatura $\mathbf{m} \in R(\mathcal{S})$.

Nei sistemi di automazione tale proprietà è di importanza vitale dal momento che un blocco del sistema risulta quasi sempre connesso con il blocco di parti nel processo, e pertanto esso comporta una perdita economica oltre alle difficoltà della reinizializzazione del sistema.

Un primo strumento per la verifica della liveness di una rete è rappresentato dal grafo di copertura. Dato un sistema PN limitato ed il suo grafo di raggiungibilità:

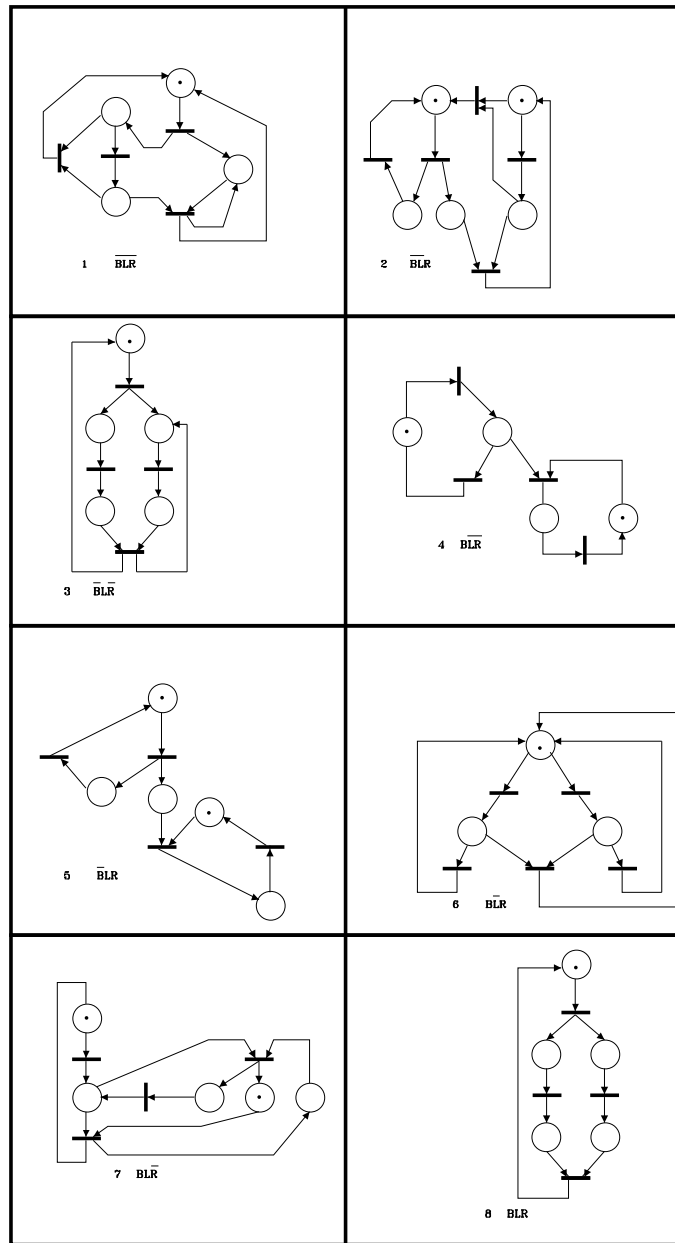


Figura 2.2.: Boundedness (B), liveness (L) e reversibilità (R) sono proprietà indipendenti

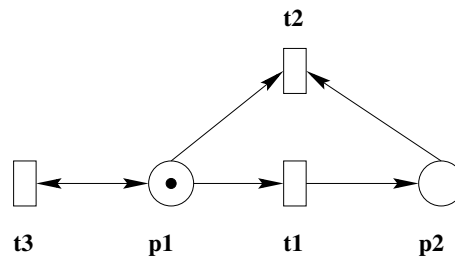


Figura 2.3.: La transizione t_2 risulta dead, perché non può mai scattare; la transizione t_1 è L1-live poiché può scattare solo una volta; la transizione t_3 è L3-live poiché può scattare infinite volte, ma a condizione che non scatti t_1 , motivo per cui non è L4-live.

- la transizione t è dead $\Leftrightarrow t$ non compare nel grafo;
- la transizione t è L1-live $\Leftrightarrow t$ compare nel grafo;
- la transizione t è L4-live \Leftrightarrow un arco etichettato t compare in ogni componente assorbente del grafo.

Dato un sistema PN ed il suo grafo di copertura:

- la transizione t è dead $\Leftrightarrow t$ non compare nel grafo;
- la transizione t è L1-live $\Leftrightarrow t$ compare nel grafo;
- la transizione t è L4-live \Rightarrow un arco etichettato t compare in ogni componente assorbente del grafo;
- la marcatura \mathbf{m} è dead \Leftarrow esiste nel grafo un nodo $\mathbf{m}_w \geq_w \mathbf{m}$ da cui non escono archi.

Alcuni risultati basati sulla classificazione delle PN insieme ai concetti di siphon e trap sono di aiuto per la verifica della liveness di un sistema. Un insieme non vuoto di posti S di una PN si dice *siphon* se $\bullet S \subseteq S^\bullet$, ovvero se una transizione ha un arco di uscita verso un posto dell'insieme ne ha anche sicuramente uno di ingresso, come mostrato in fig. 2.4 (a). Se un siphon non è marcato tale rimane nel corso dell'evoluzione della rete. Un insieme non vuoto di posti S di una PN si dice *trap* se $S^\bullet \subseteq \bullet S$, ovvero se una transizione ha un arco in ingresso da un posto dell'insieme, allora ha sicuramente un arco di uscita verso uno dei posti dell'insieme, come

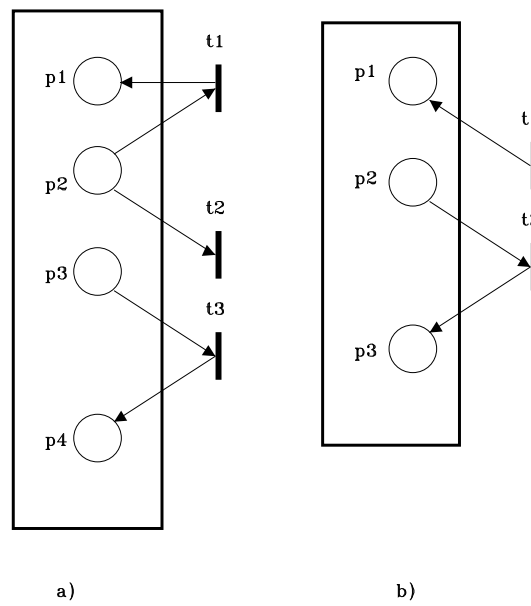


Figura 2.4.: Illustrazione di un siphon (a) ed un trap (b)

mostrato in fig. 2.4 (b). Se un trap è marcato tale rimane nel corso dell'evoluzione della rete, ovvero non sarà possibile rimuovere dai posti di cui è costituito tutti i gettoni. Si noti che il supporto $\| \mathbf{x} \|$ di un P-invariante \mathbf{x} è sia un trap che un siphon. I siphon sono di particolare interesse per l'analisi della vivezza di una rete: se un siphon si svuota dei gettoni in esso contenuto, resterà definitivamente vuoto; pertanto, tutte le transizioni che hanno in ingresso archi provenienti dai posti del siphon saranno dead.

Di seguito sono riportati alcuni dei risultati presenti in letteratura per la verifica della liveness di un sistema, basati sui concetti sopra esposti.

- una state machine (SM) è live se e solo se è strettamente connessa ed \mathbf{m}_0 ha almeno un gettone;
- un marked graph (MG) connesso è live se e solo se \mathbf{m}_0 assegna almeno un gettone ad ogni circuito diretto della rete;
- una free-choice net (FCN) è live se e solo se ogni siphon contiene un trap marcato;

- una simple net (SN) è live se ogni siphon contiene un trap marcato;
- se N è una rete ordinaria, una marcatura \mathbf{m} è di blocco se l'insieme $S = \{p | \mathbf{m}(p) = 0\}$ è un siphon e $\forall t \in T, \bullet t \cap S \neq \emptyset$.

Proposizione 1 *Data una rete di Petri avente matrice di incidenza \mathbf{C} si ha che*

$$\mathbf{x} \geq \mathbf{0}, \mathbf{x}^T \cdot \mathbf{C} \leq \mathbf{0} \Rightarrow S = \{p | \mathbf{x}(p) > 0\} \text{ è un siphon.}$$

Per la determinazione dei siphon di una rete è di fondamentale importanza il risultato espresso nella proposizione seguente.

Proposizione 2 *Data una rete di Petri avente matrice di incidenza $\mathbf{C} = \mathbf{Post} - \mathbf{Pre}$ si costruisca la matrice $\mathbf{C}' = \mathbf{Post}' - \mathbf{Pre}'$ dove*

$$\mathbf{Post}' = \mathbf{Post}$$

e

$$\forall t \in T, \forall p \in \bullet t, \mathbf{Pre}'(p, t) = k \mathbf{Pre}(p, t),$$

con $k = \sum_{p' \in \bullet t} \mathbf{Post}(p', t)$.

Si dimostra che

$\mathbf{x} \geq \mathbf{0}, \mathbf{x}^T \cdot \mathbf{C}' \leq \mathbf{0} \Leftrightarrow S = \{p | \mathbf{x}(p) > 0\}$ è un siphon della rete con matrice di incidenza \mathbf{C} .

L'algoritmo 2 può essere modificato per determinare i siphon:

- al passo 1 bisogna porre $\mathbf{A} := [\mathbf{C}' \mid \mathbf{I}_{m \times n}]$;
- al passo (2d) invece di eliminare tutte le righe di $I_+ \cup I_-$, vanno eliminate solo quelle di indice I_+ dal momento che sono ammessi elementi negativi nel prodotto $\mathbf{x}^T \cdot \mathbf{C}'$.

In tal caso l'algoritmo fornirà tutti i siphon della rete e non solo quelli minimali. Un siphon si dice minimale se non esiste un altro siphon non vuoto contenuto in esso.

Esempio 8 *Si consideri la rete in fig. 2.5. In tale esempio verrà mostrato il calcolo dei siphon di tale rete.*

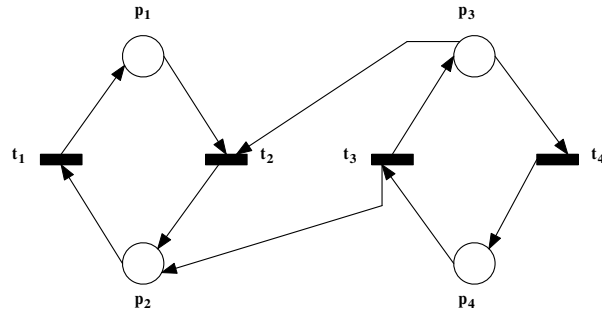


Figura 2.5.: Rete utilizzata nell'esempio 8.

$$Pre = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

$$Post = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$Post' = Post$$

$$Pre' = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 2 & 0 \end{bmatrix}$$

$$C' = \begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 1 & 1 & 0 \\ 0 & -1 & 1 & -1 \\ 0 & 0 & -2 & 1 \end{bmatrix}$$

Si costruisca la tabella $A := [C' \mid I_{4 \times 4}]$

$$\left[\begin{array}{cccc|cccc} 1 & -1 & 0 & 0 & 1 & 0 & 0 & 0 \\ -1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & -1 & 1 & -1 & 0 & 0 & 1 & 0 \\ 0 & 0 & -2 & 1 & 0 & 0 & 0 & 1 \end{array} \right] \begin{array}{l} r_1 \\ r_2 \\ r_3 \\ r_4 \end{array}$$

Al passo $j = 1$ al fine di rendere nulli gli elementi della prima colonna della tabella si esegue la somma delle righe r_1 e r_2 e viene successivamente cancellata la sola riga r_1 dal momento che la riga r_2 ha un elemento negativo in prima colonna; viene quindi inserita la riga $r_1 + r_2$ ottenendo

$$\left[\begin{array}{cccc|cccc} -1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & -1 & 1 & -1 & 0 & 0 & 1 & 0 \\ 0 & 0 & -2 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 \end{array} \right] \begin{array}{l} r_2 \\ r_3 \\ r_4 \\ r_1 + r_2 \end{array}$$

Al passo $j = 2$ al fine di rendere nulli gli elementi della seconda colonna della tabella si esegue la somma delle righe r_2 e r_3 e viene successivamente cancellata la sola riga r_2 dal momento che la riga r_3 ha un elemento negativo in seconda colonna; viene quindi inserita la riga $r_2 + r_3$ ottenendo

$$\left[\begin{array}{cccc|cccc} 0 & -1 & 1 & -1 & 0 & 0 & 1 & 0 \\ 0 & 0 & -2 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 \\ -1 & 0 & 2 & -1 & 0 & 1 & 1 & 0 \end{array} \right] \begin{array}{l} r_3 \\ r_4 \\ r_1 + r_2 \\ r_2 + r_3 \end{array}$$

Al passo $j = 3$ al fine di rendere nulli gli elementi della terza colonna della tabella si esegue la somma delle righe $2r_3$ ed r_4 , r_4 ed $2(r_1 + r_2)$, r_4 ed $r_2 + r_3$ e vengono successivamente cancellate le righe r_3 , $r_1 + r_2$, $r_2 + r_3$ ottenendo

$$\left[\begin{array}{cccc|cccc} 0 & 0 & -2 & 1 & 0 & 0 & 0 & 1 \\ 0 & -2 & 0 & -1 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 1 & 2 & 2 & 0 & 1 \\ -1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{array} \right] \begin{array}{l} r_4 \\ 2r_3 + r_4 \\ r_4 + 2(r_1 + r_2) \\ r_4 + (r_2 + r_3) \end{array}$$

Al passo $j = 4$ al fine di rendere nulli gli elementi della quarta colonna della tabella si esegue la somma delle righe r_4 ed $2r_3 + r_4$, $2r_3 + r_4$ ed $r_4 + 2(r_1 + r_2)$ e vengono successivamente cancellate le righe r_4 , $r_4 + 2(r_1 + r_2)$ ottenendo

$$\left[\begin{array}{cccc|cccc} 0 & -2 & 0 & -1 & 0 & 0 & 2 & 1 \\ -1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & -2 & -2 & 0 & 0 & 0 & 2 & 2 \\ 0 & -2 & 0 & 0 & 2 & 2 & 2 & 2 \end{array} \right] \begin{array}{l} 2r_3 + r_4 \\ r_4 + (r_2 + r_3) \\ r_4 + (2r_3 + r_4) \\ (2r_3 + r_4) + (r_4 + 2(r_1 + r_2)) \end{array}$$

Da cui ne viene che i siphon della rete sono dati dai supporti dei vettori

$[0 \ 0 \ 2 \ 1]^T$, $[0 \ 1 \ 1 \ 1]^T$, $[0 \ 0 \ 2 \ 2]^T$, $[2 \ 2 \ 2 \ 2]^T$. Poichè due di essi hanno lo stesso supporto i siphon della rete sono $S_1 = \{p_3, p_4\}$, $S_2 = \{p_2, p_3, p_4\}$, $S_3 = \{p_1, p_2, p_3, p_4\}$. Pertanto S_1 risulta essere il siphon minimale.

Infine, va detto che esistono delle condizioni sufficienti basate sull'equazione di stato anche per la verifica della liveness, che saranno tralasciate in questa sede.

2.1.4. Reversibilità ed home-state

Una rete si dice *reversibile* se

$$\forall \mathbf{m} \in R(\mathcal{S}), \mathbf{m}_0 \in R(N, \mathbf{m}).$$

A volte è importante che valga una versione rilassata di tale proprietà: l'esistenza di un *home state*. Uno stato \mathbf{m}' raggiungibile da \mathbf{m}_0 si dice home-state se

$$\forall \mathbf{m} \in R(\mathcal{S}), \mathbf{m}' \in R(N, \mathbf{m}).$$

Se una rete è reversibile la marcatura iniziale è un home-state.

Dato un sistema PN limitato ed il suo grafo di raggiungibilità:

il sistema è reversibile \Leftrightarrow il grafo è strettamente connesso.

Dato un sistema PN ed il suo grafo di copertura:

il sistema è reversibile \Rightarrow ogni componente assorbente del grafo contiene un nodo $\mathbf{m}_w \geq_w \mathbf{m}_0$.

La rete in fig. 2.6, che si può dimostrare essere reversibile, presenta una sola componente assorbente (costituita dai nodi $[0 \ 0 \ 1 \ 0 \ w \ 1]^T, [1 \ 0 \ 0 \ 0 \ w \ 1]^T, [0 \ 1 \ 0 \ 0 \ w \ 1]^T, [0 \ 0 \ 0 \ 1 \ w \ 1]^T$) che contiene la marcatura $[0 \ 0 \ 1 \ 0 \ w \ 1]^T \geq_w [0 \ 0 \ 1 \ 0 \ 0 \ 1]^T = \mathbf{m}_0$.

Nel caso generale il problema della reversibilità si riconduce a quello della raggiungibilità.

2.2. Le proprietà strutturali

Boundedness (B), liveness (L) e reversibilità (R) sono proprietà indipendenti come risulta dalla fig. 2.2. Tali proprietà di una rete sopra elencate si dicono comportamentali perché dipendono dallo stato iniziale. Della liveness e della boundedness esiste anche la versione strutturale. Precisamente se un sistema PN è limitato indipendentemente dal suo stato iniziale si parla di *boundedness strutturale* della PN ad essa associata. Inoltre, una sistema PN è *strutturalmente live* se esiste una marcatura iniziale tale che esso risulta live.

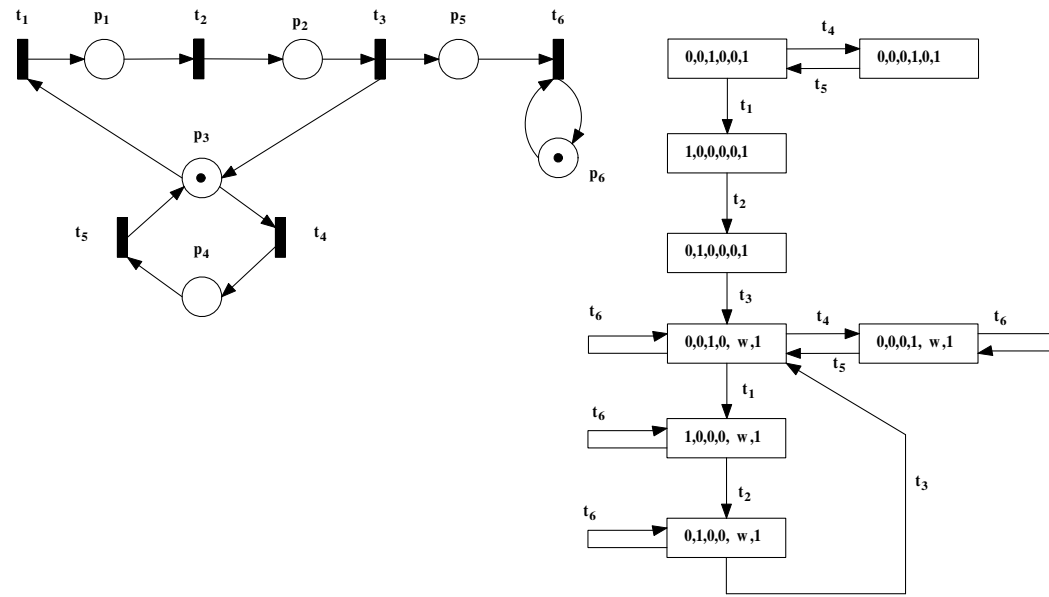


Figura 2.6.: Sistema PN e relativo grafo di copertura.

Si introducono nel seguito le definizioni delle principali proprietà strutturali di una PN, ovvero di quelle proprietà che a differenza della liveness, boundedness e reversibilità non dipendono per definizione dallo stato iniziale.

Una PN si dice (parzialmente) *conservativa* se e solo se esiste un vettore di interi positivi (non negativi) \mathbf{x} tale che

$$\forall \mathbf{m}_0 \in \mathbb{N}^{|P|}, \forall \mathbf{m} \in R(\mathcal{S}), \mathbf{x}^T \mathbf{m} = \mathbf{x}^T \mathbf{m}_0 = \text{costante}.$$

Una PN si dice (parzialmente) *ripetitiva* se esiste una marcatura iniziale \mathbf{m}_0 ed una sequenza σ da \mathbf{m}_0 tale che (qualche) ogni transizione occorre infinite volte in σ .

Una PN si dice (parzialmente) *consistente* se esiste una marcatura iniziale \mathbf{m}_0 ed una sequenza σ da \mathbf{m}_0 a \mathbf{m}_0 tale che (qualche) ogni transizione occorre almeno una volta in σ .

In fig. 2.7 ed in tab. 2.2 sono riassunte vari esempi di reti che godono di differenti proprietà strutturali, mentre in tab. 2.1 sono elencate le condizioni a cui una rete deve soddisfare per godere di tali proprietà.

Proprietà	Condizione necessaria e sufficiente
Strutturalmente bounded	$\exists \mathbf{x} > \mathbf{0}, \mathbf{x}^T \mathbf{C} \leq \mathbf{0}$
Conservativa	$\exists \mathbf{x} > \mathbf{0}, \mathbf{x}^T \mathbf{C} = \mathbf{0}$
Parzialmente conservativa	$\exists \mathbf{x} \geq \mathbf{0}, \mathbf{x}^T \mathbf{C} = \mathbf{0}$
Ripetitiva	$\exists \mathbf{y} > \mathbf{0}, \mathbf{C} \mathbf{y} \geq \mathbf{0}$
Parzialmente ripetitiva	$\exists \mathbf{y} \geq \mathbf{0}, \mathbf{C} \mathbf{y} \geq \mathbf{0}$
Consistente	$\exists \mathbf{y} > \mathbf{0}, \mathbf{C} \mathbf{y} = \mathbf{0}$
Parzialmente consistente	$\exists \mathbf{y} \geq \mathbf{0}, \mathbf{C} \mathbf{y} = \mathbf{0}$

Tabella 2.1.: Proprietà strutturali di una PN e relative condizioni necessarie e sufficienti

	a	b	c	d	e	f	g	h	i	j
Strutturalmente bounded	+	-	-	+	-	+	-	+	+	-
Conservativa	-	-	-	-	-	+	-	-	-	-
Parzialmente conservativa	-	-	-	-	-	+	+	+	-	-
Ripetitiva	-	+	+	-	-	+	+	-	-	+
Parzialmente ripetitiva	-	+	+	-	+	+	+	-	+	+
Consistente	-	-	+	-	-	+	-	-	-	-
Parzialmente consistente	-	-	+	-	-	+	-	-	+	+
Strutturalmente live	-	+	+	-	-	+	+	-	-	+

Tabella 2.2.: Proprietà strutturali delle reti in fig. 2.7

Esistono delle connessioni tra proprietà strutturali e comportamentali, alcune di seguito riportate: si dimostra che se una rete è sia strutturalmente limitata che strutturalmente live allora è sia conservativa che consistente; inoltre, una rete live e limitata è consistente.

La proprietà più difficile da trattare in termini strutturali è la liveness. In particolare si ha che un MG è sempre strutturalmente live e che una FCN è strutturalmente live se e solo se ogni siphon contiene un trap.

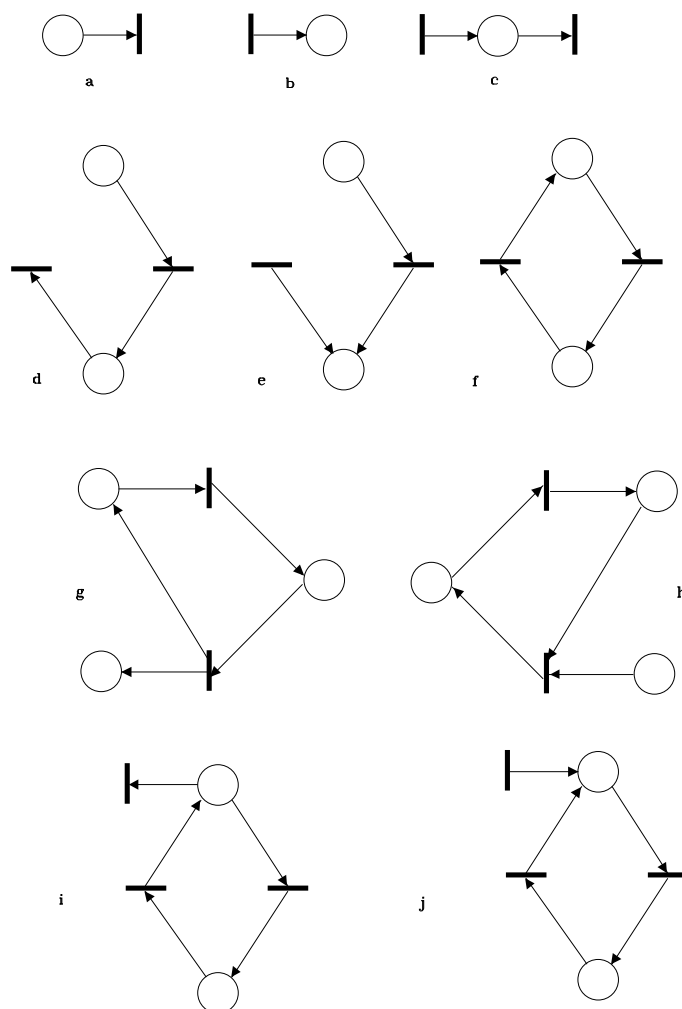


Figura 2.7.: Reti con differenti proprietà strutturali

3. Il controllo supervisivo di una rete di Petri

Oggetto di questo capitolo è il controllo di un FMS modellato come una rete di Petri controllata, ovvero in cui alcune transizioni possono essere disabilitate da un agente esterno, detto *supervisore*, al fine di interdire il raggiungimento di un insieme specificato di stati. Questa problematica di controllo prende il nome di *controllo supervisivo*.

Va precisato che il controllo supervisivo include anche il caso in cui la specifica di controllo venga assegnata come linguaggio che non sarà oggetto di queste dispense. In tal caso si assume che ad ogni transizione venga assegnata una lettera di un certo alfabeto e che pertanto le possibili evoluzioni della rete siano rappresentabili come un linguaggio: tale approccio è frequente soprattutto nell'ambito dei protocolli di comunicazione.

Il supervisore potrà o meno essere rappresentato da una rete di Petri. Il tipo di specifica preso in considerazione è quello dei vincoli di mutua esclusione generalizzata: ovvero l'insieme di stati proibiti viene specificato per mezzo di una disuguaglianza lineare sulle marcature dei posti della rete. Viene inoltre introdotto il concetto di controllore massimamente permissivo. Tra le tecniche di controllo supervisivo proposte di recente in letteratura, vengono illustrate il metodo della programmazione lineare intera ed il metodo dei posti monitor.

3.1. Il problema del controllo supervisivo

Si suppone che l'insieme delle transizioni T sia partizionato in due insiemi: T_c insieme delle transizioni *controllabili*, disegnate con rettangoli vuoti e T_{uc} insieme delle transizioni *incontrollabili*, disegnate con rettangoli pieni. Le transizioni controllabili possono essere disabilitate, ovvero si suppone che ad esse sia associato un ingresso

binario ($\mathbf{u} : T_c \rightarrow \{0, 1\}$) dal cui valore dipende se la transizione può scattare, indipendentemente dal fatto che tale transizione sia abilitata sotto una certa marcatura \mathbf{m} , ovvero che la (1.1) sia verificata. Formalmente si ha:

$$\begin{aligned} &\text{a ciascuna transizione } t \in T_c \text{ è associato un ingresso } \mathbf{u}(t) \text{ ed essa è abilitata} \\ &\text{se e solo se vale la (1.1) e } \mathbf{u}(t) = 1. \end{aligned} \quad (3.1)$$

Compito del controllore è quello di determinare tale ingresso per ciascuna transizione controllabile in maniera che, se in un certo stato una transizione controllabile risulta abilitata da tale ingresso, la rete non possa raggiungere uno stato proibito in seguito al suo scatto. Le transizioni incontrollabili non possono essere disabilitate. Una rete può o meno avere transizioni incontrollabili.

Dato un insieme di marcature $\mathcal{L} \subseteq \mathbf{N}^m$ si vuole progettare il controllore tale che l'insieme di raggiungibilità del sistema a ciclo chiuso sia $\mathcal{L} \cap R(\mathcal{S})$. L'insieme \mathcal{L} è detto insieme delle marcature *legali*.

Di particolare interesse per i sistemi di automazione risulta essere il caso in cui \mathcal{L} è assegnato nella forma di n_c disuguaglianze lineari sulle marcature dei posti della rete, dette vincoli di mutua esclusione generalizzati (GMEC). In tal caso si ha che

$$\mathcal{L} = \mathcal{M}(\mathbf{L}, \mathbf{k}) \equiv \{\mathbf{m} \in \mathbf{N}^m \mid \mathbf{L}\mathbf{m} \leq \mathbf{k}\}, \quad (3.2)$$

dove $\mathbf{L} \in \mathbf{Z}^{n_c \times m}$, $\mathbf{k} \in \mathbf{Z}^{n_c}$. L'interesse nasce dal fatto che attraverso tale specifica si riescono a modellare molte specifiche di controllo per tali sistemi tra cui la mutua esclusione, la limitazione della marcatura di un certo insieme di posti, implicazioni di tipo logico ed in taluni casi anche condizioni di liveness (come nel caso dell'esempio 13).

In presenza di transizioni incontrollabili, con riferimento ad un sistema \mathcal{S} il controllore dovrà garantire che

$$\bigcup_{\mathbf{m} \in \mathcal{L} \cap R(\mathcal{S})} R(N_{uc}, \mathbf{m}) \subseteq \mathcal{L} \quad (3.3)$$

se tale relazione è verificata \mathcal{L} si dice *controllabile*.

Si indichi con T_{uc}^* l'insieme di tutte le possibili sequenze di sole transizioni incontrollabili. In tal caso bisognerà interdire oltre agli stati che non appartengono all'insieme assegnato $\mathcal{M}(\mathbf{L}, \mathbf{k})$ anche l'insieme di stati

$$\mathcal{M}_{fuc}(\mathbf{L}, \mathbf{k}) = \{\mathbf{m} \in \mathbf{N}^m \mid \mathbf{m}[\sigma > \mathbf{m}', \mathbf{m}' \notin \mathcal{M}(\mathbf{L}, \mathbf{k}), \sigma \in T_{uc}^*]\}$$

cioè gli stati legali da cui si può raggiungere in seguito allo scatto di sole transizioni non controllabili uno stato non legale. Formalmente la cosa si esprime dicendo che l'insieme delle marcature ammissibili sotto controllo è

$$\mathcal{M}_c(\mathbf{L}, \mathbf{k}) = (\mathcal{M}(\mathbf{L}, \mathbf{k}) \cap R(N, \mathbf{m}_0)) \setminus \mathcal{M}_{fuc}(\mathbf{L}, \mathbf{k})$$

e si vede chiaramente che $|\mathcal{M}_c(\mathbf{L}, \mathbf{k})| \leq |\mathcal{M}(\mathbf{L}, \mathbf{k})|$ ovvero l'insieme delle marcature legali sotto controllo risulta avere una cardinalità minore rispetto all'insieme specificato delle marcature legali.

Se \mathcal{L} non è controllabile, si considera la classe dei sottoinsiemi controllabili di \mathcal{L} , ovvero, la classe $\Omega(\mathcal{L}) = \{\mathcal{K} \subseteq \mathcal{L} \mid \mathcal{K} \text{ è controllabile}\}$. Per ciascun sottoinsieme \mathcal{K} in $\Omega(\mathcal{L})$ si può costruire un supervisore tale che $\mathcal{K} \cap R(\mathcal{S}) \subset \mathcal{L} \cap R(\mathcal{S})$. Si può dimostrare che poichè tale classe è non vuota e chiusa sotto l'operatore di intersezione esiste un estremo superiore di tale insieme. Esso è denotato con $\mathcal{L}^\uparrow = \sup \Omega(\mathcal{L})$ e costituisce la soluzione ottima al nostro problema.

Nel caso generale il controllore è costituito da predicati logici dal cui valore dipende se ad un certo istante una transizione controllabile è abilitata o meno. Tali controllori si dicono interpretati e non consentono di ottenere un modello a ciclo chiuso dell'impianto controllato. Tuttavia essi sono spesso massimamente permissivi, nel senso che non vengono interdetto marcature al di fuori di quelle che effettivamente violano il vincolo o marcature da cui a seguito dello scatto di sole transizioni incontrollabili il vincolo può essere violato.

Se si vuole invece ottenere un controllore che sia anch'esso una PN, in presenza di transizioni non controllabili bisogna sacrificare la massima permissività del vincolo come illustrato nel seguito. In tal caso saranno interdetto anche marcature legali e non esiste un solo controllore ma una famiglia di controllori. Tale tipo di controllori è detto compilato ed è descritto dallo stesso formalismo con cui si descrive il processo: è possibile l'analisi a ciclo chiuso. In tal caso in luogo di un ingresso $\mathbf{u}(t)$ ci saranno archi dai posti di controllo alle transizioni controllabili della rete, come indicato in fig. 3.1 (b). La condizione di abilitazione per una transizione controllabile non sarà la (3.1), ma la (1.1) riscritta per la rete a ciclo chiuso.

Nel seguito nel corso degli esempi useremo spesso la notazione $m_i = \mathbf{m}(p_i)$.

Esempio 9 *Si consideri il sistema in fig. 3.1 (a), supponiamo di voler imporre il vincolo $m_1 \leq 1$. Dalla specifica di controllo si evince che la marcatura $\mathbf{m}_1 = [2 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0]$ è raggiungibile se x, y, z risultano maggiori o*

uguali a due. Si ha che $T_c = \{t_1, t_6, t_7, t_8\}$, $T_{uc} = \{t_2, t_3, t_4, t_5\}$. Oltre a tutte le marcature che presentano due gettoni in p_1 , a causa della transizioni incontrollabili bisognerà impedire il raggiungimento anche di quelle marcature come $\mathbf{m}_2 = [1 \ 0 \ 0 \ 0 \ 1 \ 1 \ 1 \ 0 \ 0 \ 0]$. Infatti pur essendo \mathbf{m}_2 una marcatura legale, a partire da \mathbf{m}_2 risulta non possibile impedire lo scatto della sequenza $t_3 t_4 t_5 t_2$ che conduce il sistema nello stato \mathbf{m}_1 , che non è uno stato legale. Da cui si evince che in presenza di transizioni non controllabili, l'insieme delle marcature legali sotto controllo è minore di quello delle marcature legali assegnate.

3.2. Il metodo della programmazione lineare intera

Si consideri la sottorete non controllabile N_{uc} ottenuta considerando le sole transizioni non controllabili della rete dell'impianto, si ha che l'insieme delle marcature potenzialmente raggiungibili a partire da una marcatura $\mathbf{m} \in R(\mathcal{S})$ ed in seguito allo scatto di sole transizioni incontrollabili è dato da

$$PR_{uc}(N, \mathbf{m}) = \{\mathbf{m}' \in \mathbf{N}^m \mid \mathbf{m}' = \mathbf{m} + \mathbf{C}_{uc} \boldsymbol{\sigma}_{uc}, \boldsymbol{\sigma}_{uc} \geq \mathbf{0}\} \quad (3.4)$$

dove $\mathbf{C}_{uc} \in \mathbf{Z}^{m \times n_{uc}}$ è la matrice di incidenza della rete N_{uc} ed n_{uc} è il numero delle transizioni non controllabili.

Il termine potenzialmente raggiungibile è dovuto al fatto che il soddisfacimento dell'equazione di stato come esposto nel capitolo 2 non costituisce una condizione sufficiente per la raggiungibilità. Nel caso che N_{uc} sia aciclica (ipotesi che faremo sempre nel seguito), allora l'insieme di sopra rappresenta effettivamente gli stati raggiungibili in seguito allo scatto di sole transizioni incontrollabili e lo denotiamo come $R_{uc}(N, \mathbf{m})$.

L'insieme delle marcature non legali dovute allo scatto delle transizioni non controllabili risulterà dunque dato da:

$$\mathcal{F}_{uc}(\mathbf{L}, \mathbf{k}) = \{\mathbf{m} \in \mathbf{N}^m \mid \exists \boldsymbol{\sigma}_{uc} \geq \mathbf{0}, \mathbf{L}\mathbf{m} + \mathbf{L}\mathbf{C}_{uc} \boldsymbol{\sigma}_{uc} > \mathbf{k}\} \quad (3.5)$$

Pertanto il controllore dovrà garantire che sotto una certa marcatura \mathbf{m}' siano abilitate solo le transizioni controllabili che consentono il raggiungimento di marcature della rete \mathbf{m}'' dell'impianto tali che sia verificato il predicato logico:

$$\mathbf{L}\mathbf{m}'' + \mathbf{L}\mathbf{C}_{uc} \boldsymbol{\sigma}_{uc}^* \leq \mathbf{k} \quad (3.6)$$

dove σ_{uc}^* è la soluzione del ILP $\forall \mathbf{m} \geq \mathbf{0}$

$$\begin{aligned} & \max_{\sigma_{uc}} \mathbf{L} \mathbf{C}_{uc} \sigma_{uc} \\ \text{s.t. } & \begin{cases} \sigma_{uc} \geq \mathbf{0} \\ \mathbf{C}_{uc} \sigma_{uc} \geq -\mathbf{m}'' \end{cases} \end{aligned} \quad (3.7)$$

Poichè la soluzione del problema di programmazione lineare intera precedente σ_{uc}^* è in generale una funzione non lineare di \mathbf{m} , il controllore non sempre è implementabile come una PN.

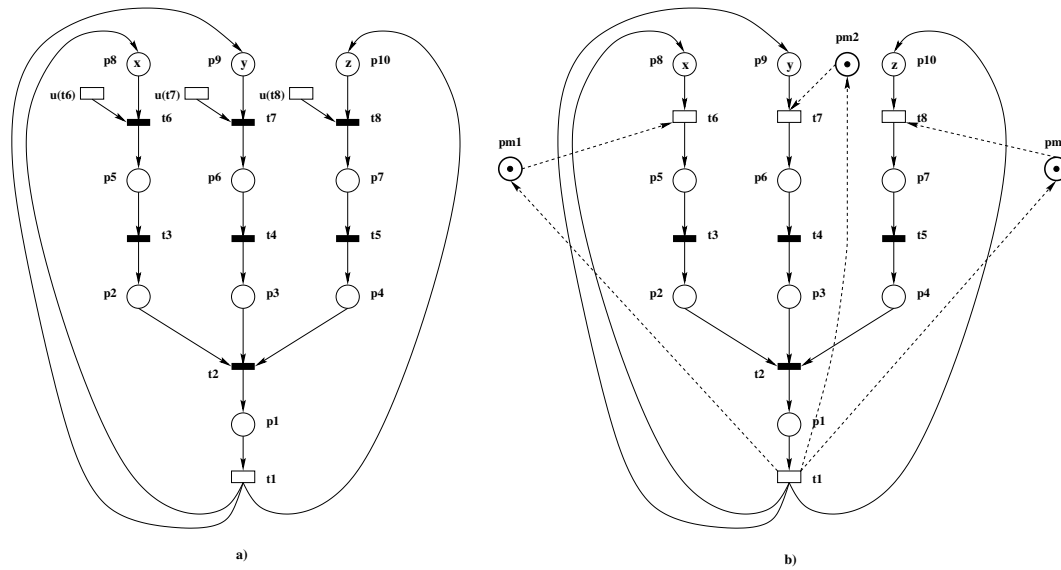


Figura 3.1.: Sistema PN dell'esempio 10

Esempio 10 Data la rete in fig. 3.1 (a), supponiamo di voler imporre il vincolo $m_1 \leq b$. La matrice di incidenza della parte incontrollabile della rete è

$$\mathbf{C}_{uc} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Dalla specifica di controllo segue che $\mathbf{L} = [1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0]$, $\mathbf{k} = [b]$.

Con riferimento al sistema 3.7 si ha

$$\begin{aligned} & \max_{\sigma_2, \sigma_3, \sigma_4, \sigma_5} \sigma_2 \\ \text{s.t.} \quad & \begin{cases} \sigma_2, \sigma_3, \sigma_4, \sigma_5 \geq 0 \\ \sigma_2 \geq -m_1 \\ -\sigma_2 + \sigma_3 \geq -m_2 \\ -\sigma_2 + \sigma_4 \geq -m_3 \\ -\sigma_2 + \sigma_5 \geq -m_4 \\ -\sigma_3 \geq -m_5 \\ -\sigma_4 \geq -m_6 \\ -\sigma_5 \geq -m_7 \end{cases} \end{aligned}$$

che dopo alcune manipolazioni diventa

$$\begin{aligned} & \max_{\sigma_2, \sigma_3, \sigma_4, \sigma_5} \sigma_2 \\ \text{s.t.} \quad & \begin{cases} \sigma_2, \sigma_3, \sigma_4, \sigma_5 \geq 0 \\ \sigma_2 \geq -m_1 \\ \sigma_2 \leq m_2 + \sigma_3 \\ \sigma_2 \leq m_3 + \sigma_4 \\ \sigma_2 \leq m_4 + \sigma_5 \\ \sigma_3 \leq m_5 \\ \sigma_4 \leq m_6 \\ \sigma_5 \leq m_7 \end{cases} \end{aligned}$$

e quindi si che $\sigma_2^* = \min(m_2 + m_5, m_3 + m_6, m_4 + m_7)$.

In base alla 3.6 si ha che sotto una certa marcatura \mathbf{m}' , $\forall t \in T_c$, $\mathbf{u}(t) = 1$ se

$$m_1'' + \min(m_2'' + m_5'', m_3'' + m_6'', m_4'' + m_7'') \leq b. \quad (3.8)$$

con $\mathbf{m}'[t] > \mathbf{m}''$.

3.3. Il metodo dei posti monitor

Ipotizziamo che $n_c = 1$, pertanto la specifica di controllo risulta essere:

$$\sum_{i=1}^m l_i m_i \leq b. \quad (3.9)$$

Tale disequaglianza può essere trasformata in una eguaglianza se si introduce una variabile slack non negativa tale che:

$$\sum_{i=1}^m l_i m_i + m_c = b. \quad (3.10)$$

Essendo m_i la marcatura del posto p_i , si pensa di associare alla variabile slack m_c il significato di marcatura di un posto di controllo p_c . L'idea di tale approccio è di garantire la (3.9) attraverso l'introduzione di un posto di controllo che contenga sotto ciascun stato della rete un certo numero di gettoni che normalmente la dinamica delle rete lascerebbe fluire nei suoi posti, violando la disequaglianza. La (3.10) può essere riscritta come

$$\mathbf{x}^T \widehat{\mathbf{m}} = \mathbf{x}^T \cdot \widehat{\mathbf{m}}_0 \quad (3.11)$$

con $\mathbf{x} = [\mathbf{l} \ 1]$, $\widehat{\mathbf{m}} = [\mathbf{m} \ m_c]^T$ e $\mathbf{x}^T \cdot \widehat{\mathbf{m}}_0 = b$. La (3.11) esprime il fatto che una somma pesata delle marcature dei posti di un rete che denotiamo con \widehat{N} , pari a quella originaria con l'aggiunta di un posto di controllo p_c , deve rimanere costante nel corso dell'evoluzione delle rete a partire dalla marcatura iniziale. Sappiamo che, se \mathbf{x} fosse un P-invariante della rete \widehat{N} , la (3.11) risulterebbe verificata $\forall \widehat{\mathbf{m}} \in R(\widehat{N}, \widehat{\mathbf{m}}_0)$. Il metodo esposto nel seguente paragrafo si basa sulla determinazione degli elementi della matrice di incidenza del posto p_c e della sua marcatura iniziale in maniera da verificare la (3.11). Il posto p_c prende il nome di posto *monitor*.

3.3.1. Assenza di transizioni incontrollabili

Supponiamo che tutte le transizioni di un sistema \mathcal{S} siano controllabili e che la specifica di controllo sia al solito

$$\mathbf{L}\mathbf{m} \leq \mathbf{k}, \quad (3.12)$$

con $\mathbf{L} \in \mathbf{Z}^{n_c \times m}$, $\mathbf{k} \in \mathbf{Z}^{n_c \times 1}$. Introducendo per ciascuno degli n_c vincoli una variabile slack, tale che:

$$\mathbf{L}\mathbf{m} + \mathbf{m}_c = \mathbf{k}, \quad (3.13)$$

con $\mathbf{m}_c \in \mathbf{Z}^{n_c \times 1}$ vettore delle marcature dei posti della rete di controllo da determinare. La (3.11) è garantita da tutti i P-invarianti così determinati:

$$\begin{aligned} \mathbf{X}^T \widehat{\mathbf{C}} &= [\mathbf{L} \ \mathbf{I}] \begin{bmatrix} \mathbf{C} \\ \mathbf{C}_c \end{bmatrix} = \mathbf{0}, \\ \mathbf{L}\mathbf{C} + \mathbf{C}_c &= \mathbf{0}, \end{aligned} \quad (3.14)$$

con $\widehat{C} \in \mathbf{Z}^{(m+n_c) \times n}$ matrice di incidenza della rete a ciclo chiuso ottenuta aggiungendo alla rete originaria un posto di controllo per ciascuno degli n_c vincoli (per cui essa ha n_c righe in più rispetto a C), I matrice identità di dimensione $n_c \times n_c$. Il blocco incognito C_c della matrice \widehat{C} contiene i pesi degli archi che connettono i posti di controllo alle transizioni della rete originaria. Dalla (3.14) segue che essa risulta completamente determinata dalla:

$$C_c = -LC. \quad (3.15)$$

Si noti che C_c è di fatto la matrice di incidenza della rete di controllo. Tale metodo caratterizza tale rete di controllo completamente in termini della matrice di incidenza (non è possibile basandosi semplicemente su una proprietà di conservazione ottenere elementi delle matrici di pre e post incidenza): pertanto tale rete non conterrà self-loop. Inoltre, essa è determinata esclusivamente sulla base della matrice di incidenza della rete originaria C , senza tenere pertanto conto del fatto che essa possa contenere dei self-loop. Pertanto, tale metodo risulta applicabile solo in assenza di self-loop nella rete che modella il processo da controllare.

Per quanto riguarda la marcatura iniziale del controllore, essa può essere ricavata imponendo che la (3.13) sia verificata anche sotto la marcatura iniziale:

$$Lm_0 + m_{c0} = k,$$

da cui.

$$m_{c0} = k - Lm_0. \quad (3.16)$$

Va precisato che m_{c0} deve essere un vettore di interi positivi. Per cui un certo vincolo (L, k) può essere imposto su un certo sistema S se e solo se

$$k - Lm_0 \geq 0. \quad (3.17)$$

La (3.17) è detta *condizione di ammissibilità* del vincolo.

Esempio 11 Si consideri il sistema in fig. 3.2 con matrice di incidenza

$$C = \begin{bmatrix} -1 & 0 & 0 & 1 & 0 \\ 1 & 1 & -1 & 0 & 1 \\ 0 & -1 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 \end{bmatrix}$$

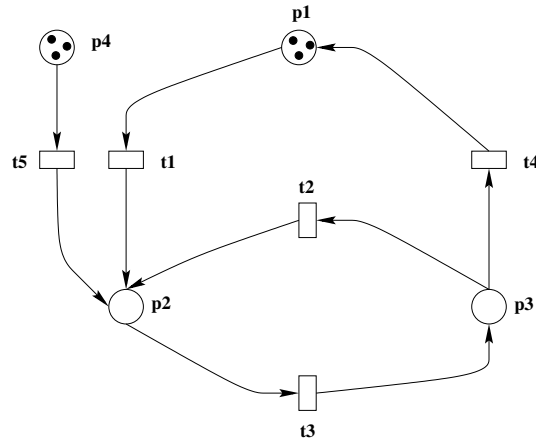


Figura 3.2.: Sistema PN dell'esempio 11

e stato iniziale è $\mathbf{m}_0 = [3 \ 0 \ 0 \ 3]^T$. La matrice \mathbf{C} ha un unico P -invariante uguale a $\mathbf{x}_1 = [1 \ 1 \ 1 \ 1]^T$.

La specifica di controllo è l'imposizione del GMEC

$$m_2 + m_3 \leq 1$$

da cui

$$\mathbf{L} = [0 \ 1 \ 1 \ 0], \mathbf{k} = [1].$$

Chiaramente tale specifica non risulta rispettata dalla rete in assenza di controllo dal momento che \mathbf{L} non risulta multiplo di \mathbf{x}_1 . A tal fine va aggiunto un posto monitor la cui riga nella matrice di incidenza del sistema a ciclo chiuso $\hat{\mathbf{C}}$ risulta data da

$$\mathbf{C}_c = -\mathbf{L}\mathbf{C} = [-1 \ 0 \ 0 \ 1 \ -1]^T.$$

La marcatura iniziale del posto monitor risulta invece pari a

$$\mathbf{m}_{c0} = 1 - \mathbf{L}\mathbf{m}_0 = 1.$$

Il sistema a ciclo chiuso risultante è mostrato in fig. 3.3, dove gli archi del posto monitor sono disegnati in linea tratteggiata.

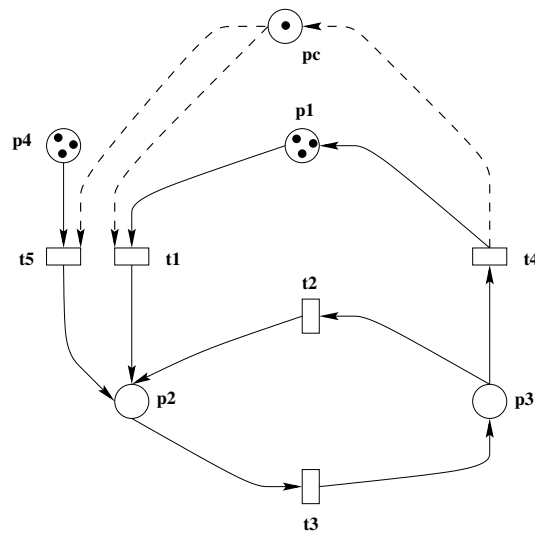


Figura 3.3.: Sistema PN di fig. 3.2 su cui è forzato il vincolo $m_1 + m_3 \leq 1$ per mezzo del posto monitor p_c .

Proposizione 3 Dato il sistema $\langle N, \mathbf{m}_0 \rangle$, con matrice di incidenza \mathbf{C} , ed un GMEC (\mathbf{l}, k) , sia $\widehat{\mathcal{S}} \langle \widehat{N}, \widehat{\mathbf{m}}_0 \rangle$ il sistema a ciclo chiuso con matrice di incidenza

$$\widehat{\mathbf{C}} = \begin{bmatrix} \mathbf{C} \\ -\mathbf{l}\mathbf{C} \end{bmatrix}$$

e marcatura iniziale

$$\widehat{\mathbf{m}}_0 = \begin{bmatrix} \mathbf{m}_0 \\ k - \mathbf{l}\mathbf{m}_0 \end{bmatrix}$$

ottenuto aggiungendo alla rete a ciclo aperto un posto monitor p_c per garantire il rispetto del vincolo (\mathbf{l}, k) .

Data una marcatura $\widehat{\mathbf{m}} = [\mathbf{m} \quad m(p_c)]^T \in R(\widehat{\mathcal{S}})$ raggiungibile a ciclo chiuso vale:

- $\mathbf{l}\mathbf{m} \leq k$, cioè la marcatura della rete del processo rispetta il vincolo;
- data una qualunque transizione t , se $\mathbf{m}[t > \mathbf{m}'$ e $m(p_c) < \widehat{Pre}(p_c, t)$ allora $\mathbf{l}\mathbf{m}' > k$, cioè se una transizione t risulta abilitata dalla marcatura \mathbf{m} ma disabilitata dal monitor il suo scatto porterebbe ad una marcatura non legale.

Dim. Per il primo punto si osservi che il vettore $\mathbf{x} = [\mathbf{l} \ 1]^T$ è un P-invariante del sistema a ciclo chiuso dal momento che

$$\mathbf{x}^T \widehat{\mathbf{C}} = [\mathbf{l} \ 1] \begin{bmatrix} \mathbf{C} \\ \mathbf{c}_c \end{bmatrix} = [\mathbf{l} \ 1] \begin{bmatrix} \mathbf{C} \\ -\mathbf{l}\mathbf{C} \end{bmatrix} = \mathbf{l}\mathbf{C} - \mathbf{l}\mathbf{C} = \mathbf{0}$$

da cui discende che deve valere $\mathbf{x}^T \widehat{\mathbf{m}} = \mathbf{x}^T \widehat{\mathbf{m}}_0$ ovvero

$$\mathbf{x}^T \mathbf{m} + m(p_c) = \mathbf{x}^T \widehat{\mathbf{m}} = \mathbf{x}^T \widehat{\mathbf{m}}_0 = \mathbf{x}^T \mathbf{m}_0 + k - \mathbf{x}^T \mathbf{m}_0 = k$$

per cui data la marcatura $\widehat{\mathbf{m}} \in \widehat{\mathcal{S}}$, la marcatura \mathbf{m} sarà legale.

Per il secondo punto, si consideri una transizione t tale che $\mathbf{m}' = \mathbf{m} + \mathbf{C}(\cdot, t)$. Per definizione di posto monitor si avrà un arco da p_c a t se e solo se $\mathbf{c}_c(t) < 0$ ed in tal caso $\mathbf{c}_c(t) = \widehat{Pre}(p_c, t)$. Per cui se la transizione t risulta disabilitata dal monitor vale $m(p_c) < \widehat{Pre}(p_c, t) = \mathbf{l}\mathbf{C}(\cdot, t)$ da cui

$$\mathbf{l}\mathbf{m}' = \mathbf{l}[\mathbf{m} + \mathbf{C}(\cdot, t)] = \mathbf{l}\mathbf{m} + \mathbf{l}\mathbf{C}(\cdot, t) > \mathbf{l}\mathbf{m} + m(p_c) > k$$

da cui discende che la marcatura \mathbf{m}' risulterebbe essere non legale.

Risulta provato in tal modo che il monitor è in assenza di transizioni controllabili un controllore massimamente permissivo. La generalizzazione della dimostrazione al caso di $n_c > 1$ è immediata.

3.3.2. Presenza di transizioni incontrollabili

In presenza di transizioni non controllabili va imposto il vincolo aggiuntivo:

$$\mathbf{L}\mathbf{C}_{uc} \leq \mathbf{0} \tag{3.18}$$

Dalla (3.15), se la (3.18) non fosse soddisfatta, gli elementi della matrice di incidenza della PN del controllore \mathbf{C}_c presenterebbero archi in uscita verso le transizioni incontrollabili che invece non possono essere disabilitate.

L'idea è quella di mantenere la semplicità del controllore del caso in cui tutte le transizioni sono controllabili. Dato l'insieme \mathcal{L} espresso da n_c disequazioni lineari si definisce la classe

$$\Omega_{n_c}(\mathcal{L}) = \{\mathcal{K} \subseteq \mathcal{L} \mid \mathcal{K} \text{ è controllabile, } \exists \mathbf{L}' \in \mathbf{Z}^{n_c \times m}, \mathbf{k}' \in \mathbf{Z}^{n_c} : \mathcal{K} = \mathcal{M}(\mathbf{L}', \mathbf{k}')\}$$

dei sottoinsiemi di \mathcal{L} *controllabili ed espressi da n_c disequazioni lineari*. Si dimostra che tale classe non ammette un estremo superiore perchè non è chiusa sotto

l'operatore di intersezione. Per tale motivo non esiste un monitor ottimo nel caso in cui sono presenti transizioni incontrollabili. Tuttavia è possibile calcolare tutti gli elementi di $\Omega_{n_c}(\mathcal{L})$: un criterio diverso da quello di inclusione deciderà la soluzione ottima.

È possibile trasformare (\mathbf{L}, \mathbf{k}) al fine di includere (3.18), preservando il fatto che la matrice di incidenza del controllore possa essere ottenuta semplicemente dalla moltiplicazione della matrice del vincolo per quella del processo.

Proposizione 4 *Date le matrici \mathbf{R}_1 e \mathbf{R}_2 tali che*

$$\mathbf{R}_1 \in \mathbf{Z}^{n_c \times m}, \mathbf{R}_1 \mathbf{m} \geq \mathbf{0}, \forall \mathbf{m} \quad (3.19)$$

$$\mathbf{R}_2 \in \mathbf{N}^{n_c \times n_c}, \text{ definita positiva e diagonale} \quad (3.20)$$

e

$$[\mathbf{R}_1 \quad \mathbf{R}_2] \begin{bmatrix} \mathbf{C}_{uc} & \mathbf{m}_0 \\ \mathbf{L}\mathbf{C}_{uc} & \mathbf{L}\mathbf{m}_0 - \mathbf{k} - \mathbf{1} \end{bmatrix} \quad (3.21)$$

$$\leq [\mathbf{O}_{n_c \times n_{uc}} \quad -\mathbf{1}]$$

con $\mathbf{1}$ vettore $n_c \times 1$ di 1, il controllore calcolato come

$$\mathbf{C}_c = -\mathbf{L}'\mathbf{C} \quad (3.22)$$

$$\mathbf{m}_{c0} = \mathbf{k}' - \mathbf{L}'\mathbf{m}_0 \quad (3.23)$$

dove

$$\mathbf{L}' = \mathbf{R}_1 + \mathbf{R}_2\mathbf{L} \quad (3.24)$$

$$\mathbf{k}' = \mathbf{R}_2(\mathbf{k} + \mathbf{1}) - \mathbf{1}. \quad (3.25)$$

assicura che per il sistema retroazionato si abbia

$$\mathbf{L}\mathbf{m} \leq \mathbf{k}$$

$$\mathbf{L}'\mathbf{C}_{uc} \leq \mathbf{0}$$

$$\mathbf{0} \leq (\mathbf{R}_1\mathbf{m}_0) \leq \mathbf{R}_2(\mathbf{k} + \mathbf{1} - \mathbf{L}\mathbf{m}_0) - \mathbf{1} \quad (3.26)$$

La (3.26) equivale alla condizione di ammissibilità ($L'm_0 \leq k'$) per il vincolo trasformato.

La proposizione 4 è basata sull'implicazione $L'm \leq k' \Rightarrow Lm \leq k$. Essa può essere dimostrata come segue:

$$L'm \leq k' \Leftrightarrow (R_1 + R_2 L)m \leq R_2(k + 1) - 1$$

da cui trattandosi di una disuguaglianza tra interi

$$(R_1 + R_2 L)m < R_2(k + 1)$$

ed essendo R_2 invertibile

$$R_2^{-1}R_1 m + Lm < k + 1$$

si ha che per la (3.19) e la (3.20) $R_2^{-1}R_1 m \geq 0$ e quindi $Lm \leq k$.

Nell'ipotesi (sempre supposta in questa sede) che N_{uc} sia aciclica, è possibile calcolare le matrici R_1 and R_2 con semplici operazioni di moltiplicazioni di riga per interi positivi ed addizioni tra righe della tabella di interi

$$\left[\begin{array}{c|c|c} C_{uc} & I & O \\ \hline N & R_1 & R_2 \end{array} \right] \quad (3.27)$$

in modo da annullare gli eventuali elementi positivi del blocco N , che al passo iniziale dell'algoritmo si suppone pari a LC_{uc} . Va tenuto presente che per il rispetto della (3.19) e della (3.20) non si possono sottrarre due righe nè moltiplicare una riga per un intero negativo. Ad ogni passo dell'algoritmo si sceglie un elemento positivo del blocco N , sia esso $N(i, j)$. Si procede addizionando la i -esima riga del blocco $[N \ R_1 \ R_2]$ con una riga del blocco C_{uc} avente un elemento negativo nella colonna j , sia l l'indice di tale riga. Prima di eseguire l'addizione tra le due righe, la i -esima riga del blocco $[N \ R_1 \ R_2]$ va moltiplicata per $C_{uc}(l, j)/N(i, j)$. Se ci sono più elementi negativi nella colonna $C_{uc}(\cdot, j)$, ciascuno di essi porta ad una soluzione diversa. L'algoritmo si arresta quando tutti gli elementi di N sono negativi o nulli o se la colonna $C_{uc}(\cdot, j)$ non contiene elementi negativi, ovvero se la trasformazione del vincolo è impossibile. Va tenuto presente che nel corso dell'algoritmo il blocco $[C_{uc} \ I \ O]$ non viene modificato. Al passo iniziale $R_1 = [O_{n_c \times m}]$ e $R_2 = [I_{n_c \times n_c}]$.

Esempio 12 La matrice di incidenza della rete in fig. 3.4 è

$$\mathbf{C} = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -2 & 1 & 0 \\ 0 & -1 & 0 & 1 \\ 2 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \end{bmatrix}$$

La marcatura iniziale è $\mathbf{m}_0 = [0 \ 0 \ 0 \ x \ y]$. Se la specifica di controllo assegnata è $m_1 \leq b$, ovvero $\mathbf{L} = [1 \ 0 \ 0 \ 0 \ 0]$, $k = b$ si ha che

$$\left[\begin{array}{c|c|c} \mathbf{C}_{uc} & \mathbf{I} & \mathbf{O} \\ \hline \mathbf{N} & \mathbf{R}_1 & \mathbf{R}_2 \end{array} \right] =$$

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

Bisogna azzerare $\mathbf{N}(1,1)$ e scegliamo come elemento pivot $\mathbf{C}_{uc}(2,1)$, da cui moltiplicando per 2 l'ultima riga ed addizionandola alla seconda si ottiene:

$$\left[\begin{array}{c|c|c} \mathbf{C}_{uc} & \mathbf{I} & \mathbf{O} \\ \hline \mathbf{N} & \mathbf{R}_1 & \mathbf{R}_2 \end{array} \right] =$$

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

e quindi $\mathbf{R}_1 = [0 \ 1 \ 0 \ 0 \ 0]$, $\mathbf{R}_2 = [2]$. In tal caso il vincolo risulterà trasformato come segue:

$$\mathbf{L}'_1 = [2 \ 1 \ 0 \ 0 \ 0], \quad k'_1 = 2(b+1) - 1,$$

ed il posto monitor sarà determinato da

$$\mathbf{C}_{c1} = -\mathbf{L}'_1 \mathbf{C} = [2 \ 0 \ -1 \ 0]$$

$$m_{c01} = k' - \mathbf{L}'_1 \mathbf{m}_0 = 2(b+1) - 1 - b = b+1$$

Scegliendo come elemento pivot $C_{uc}(3,1)$, si ottiene invece $\mathbf{R}_1 = [0 \ 0 \ 1 \ 0 \ 0]$, $\mathbf{R}_2 = [1]$, da cui

$$\mathbf{L}'_2 = [1 \ 0 \ 1 \ 0 \ 0], \ k'_2 = b,$$

quindi

$$\mathbf{C}_{c2} = -\mathbf{L}'_2 \mathbf{C} = [1 \ 0 \ 0 \ -1]$$

$$m_{c02} = k' - \mathbf{L}'_2 \mathbf{m}_0 = b$$

Se avessimo adoperato il metodo della programmazione lineare avremmo ottenuto che sotto una certa marcatura \mathbf{m}' , $\forall t \in T_c$, $\mathbf{u}(t) = 1$ se

$$\mathbf{u}(t) = 1, \forall t \in T_c \text{ se } m''_1 + \min(m''_3, [\frac{m''_2}{2}]) \leq b$$

con $\mathbf{m}'[t] > \mathbf{m}''$.

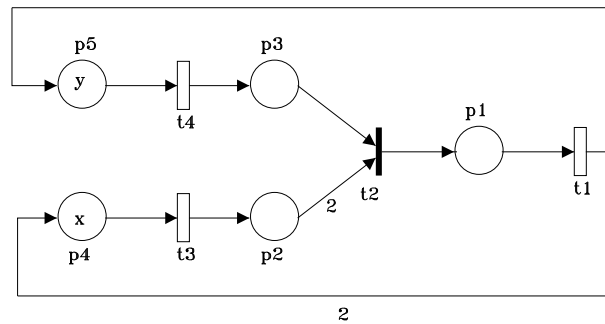


Figura 3.4.: Sistema PN nell'esempio 12

Esempio 13 Si noti che la rete in fig. 3.5 presenta tre siphon: $S_1 = \{p_1, p_2\}$, $S_2 = \{p_1, p_3\}$, $S_3 = \{p_1, p_2, p_3\}$. Dal momento che S_1 ed S_2 possono essere svuotati se una delle transizioni t_1 o t_2 scatta due volte prima che scatti l'altra, la rete risulta non viva. Per renderla viva va imposto che la differenza tra il numero di scatti delle due transizioni sia sempre inferiore o uguale ad uno, che equivale ad imporre che i due sifoni minimali S_1 ed S_2 non possano essere svuotati. In termini di vincoli sui posti della rete la specifica si traduce nei due vincoli: $m_2 \leq 1$, $m_3 \leq 1$. Per ciascuno di essi può essere ricavato un posto di controllo, ottenendo la PN controllata mostrata in fig. 3.5.

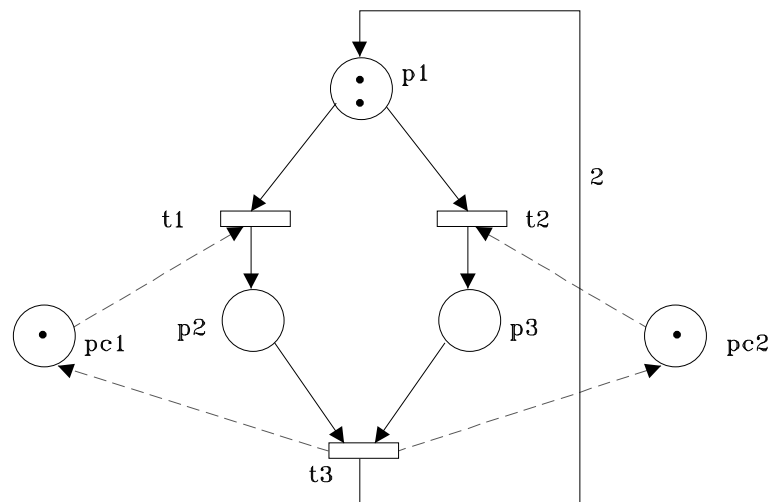


Figura 3.5.: Sistema PN dell'esempio 13

3.4. Considerazioni conclusive

Al fine di effettuare un confronto tra i metodi presentati si prenda in esame il sistema in fig. 3.1 (a), che nel corso dell'esempio 10 è stato controllato attraverso il metodo della programmazione lineare intera, con la specifica $m_1 \leq b$. In fig. 3.1 (b) è disegnata la stessa rete con la soluzione in termini di posti monitor, nel caso di $b = 1$: ciascuno dei posti p_{mi} costituisce una possibile soluzione. Nel caso del metodo dei monitor dunque in generale la soluzione non è unica, essa è inoltre non massimamente permissiva. Si consideri, infatti, il caso in cui $x = y = z = 3$ e sia stato scelto il monitor p_{m1} . La marcatura $\mathbf{m} = [0 \ 0 \ 0 \ 0 \ 2 \ 0 \ 0 \ 1 \ 3 \ 3]$ risulta proibita, mentre essa è legale per la (3.5) ed infatti ammessa nella soluzione ottenuta con il metodo della programmazione lineare intera come si può verificare attraverso la (3.8).

Il metodo dei monitor consente però nel caso in esame di valutare se la rete a ciclo chiuso è viva o reversibile dal momento che essa risulta un MG qualunque sia il posto monitor scelto. Se la specifica richiedesse l'ottimizzazione del tempo ciclo, nel caso in cui le transizioni fossero temporizzate, si potrebbe scegliere in base alle durate di tali transizioni il monitor che ottimizza il tempo ciclo contemporaneamente all'imposizione del GMEC. Ed in tal caso quindi si otterrebbe con il metodo dei

monitor una soluzione ottima e non subottima come nel caso della sola specifica di tipo logico. Il motivo risiede nel fatto che i monitor rappresentano un controllore di tipo compilato, ovvero sono essi stessi una PN, e pertanto consentono di analizzare le proprietà della rete a ciclo chiuso con le stesse tecniche usate per l'analisi della rete del processo. L'analisi a ciclo chiuso sarebbe possibile nel caso della programmazione lineare intera solo attraverso tecniche simulative o esaustive.

A. Richiami dalla teoria dei grafi

Definizione 1 Un grafo è una coppia $G = (V, A)$, dove V è l'insieme dei vertici (o nodi), e $A \subseteq V \times V$ è l'insieme degli archi.

Ogni arco incide fra due nodi, cioè collega due nodi, detti adiacenti, fra loro. In fig. A.1 (a), è rappresentato il grafo con insieme dei nodi $V = \{v_1, v_2, v_3\}$ e insieme degli archi $A = \{a_1, a_2\}$. I nodi sono rappresentati da punti o cerchi, mentre gli archi da tratti di linea che congiungono i due nodi su cui l'arco incide.

Definizione 2 Un grafo è detto orientato se gli archi hanno un verso di percorrenza che si indica mediante una freccia.

Definizione 3 Un grafo $G = (V, A)$ è detto bipartito se è possibile partizionare l'insieme dei nodi in due sottoinsiemi disgiunti $V = V_1 \cup V_2$, tali che $A \subseteq (V_1 \times V_2) \cup (V_2 \times V_1)$, cioè ogni arco incide fra nodi appartenenti a sottoinsiemi diversi.

In fig. A.1, i grafi (a,b) sono bipartiti se si considera $V_1 = \{v_1, v_3\}$, $V_2 = \{v_2\}$.

Definizione 4 Un grafo $G = (V, A)$ è detto pesato se ad esso è associata una funzione peso $f : A \rightarrow \mathbf{R}$; per ogni $a \in A$, $f(a)$ è detto il peso dell'arco a .

Ogni rete di Petri è un grafo bipartito orientato e pesato: infatti l'insieme dei nodi è costituito da due insieme disgiunti rappresentati da P e T ed i pesi di tale grafo sono dati dagli elementi delle matrici **Pre** e **Post**.

Definizione 5 Un cammino in un grafo è una sequenza $\gamma = v_{j_0} a_{j_1} v_{j_1} a_{j_2} \dots a_{j_k} v_{j_k}$ composta alternativamente di nodi $v_j \in V$ e di archi $a_j \in A$ e dove $a_j = \{v_{j_{i-1}}, v_{j_i}\}$, cioè ogni nodo $v_{j_{i-1}}$ è adiacente tramite l'arco a_{j_i} al nodo successivo v_{j_i} . Si dice anche che tale cammino porta da v_{j_0} a v_{j_k} ed ha lunghezza k (la lunghezza di un cammino è il pari al numero di archi che lo compongono). Un cammino $\gamma = v_{j_0} a_{j_1} v_{j_1} a_{j_2} \dots a_{j_k} v_{j_k}$ è detto orientato se $a_j = \{v_{j_{i-1}}, v_{j_i}\}$, cioè ogni arco a_{j_i} ha un verso di percorrenza che va dal nodo $v_{j_{i-1}}$ al nodo v_{j_i} .

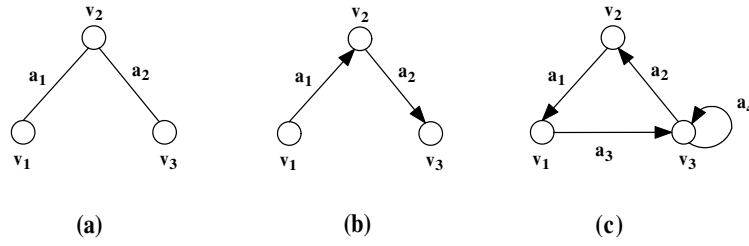


Figura A.1.: (a) un grafo connesso non orientato, (b) un grafo connesso orientato, (c) un grafo fortemente connesso ed orientato.

Definizione 6 *Un ciclo è un cammino γ in cui $v_{j_0} = v_{j_k}$, cioè il nodo iniziale e quello finale coincidono. Un ciclo è detto orientato se ogni arco a_j ha un verso di percorrenza che va dal nodo $v_{j_{i-1}}$ al nodo v_{j_i} . Un ciclo è detto elementare se non passa più di una volta per lo stesso nodo, escluso ovviamente quello iniziale e finale.*

Definizione 7 *Un grafo è connesso se per ogni coppia di nodi $v, v' \in V$ esiste un cammino che porta da v a v' . Un grafo orientato è fortemente connesso se per ogni coppia ordinata di nodi (v, v') esiste un cammino orientato che porta da v a v' .*

In fig. A.1 il grafo (b) non è fortemente connesso dal momento che non esiste nessun cammino orientato che porta da v_2 a v_1 mentre esiste un cammino orientato che porta da v_1 a v_2 .

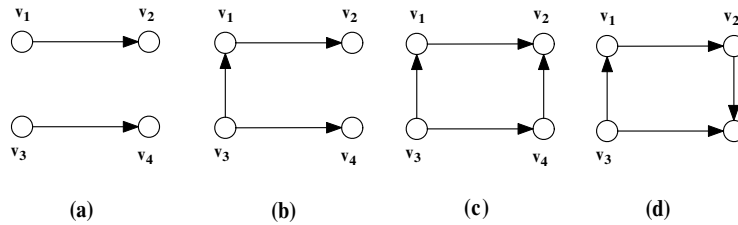


Figura A.2.: (a) un grafo non connesso, (b) un albero, (c) un grafo aciclico (d) un grafo ciclico.

Definizione 8 *Un albero è un grafo connesso in cui non esistono cicli .*

Definizione 9 *Un grafo orientato è detto aciclico se non contiene cicli orientati.*

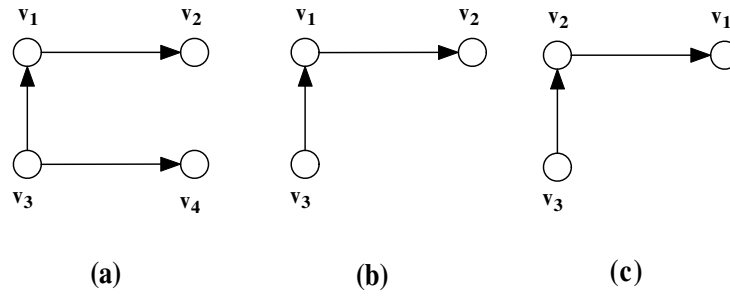


Figura A.3.: (a) un grafo, (b) un suo sottografo, (c) un altro sottografo ad esso isomorfo.

Definizione 10 Dato un grafo $G = (V, A)$, si dice che il grafo $G' = (V', A')$ è un suo sottografo se vale $V' \subseteq V$ e $A' \subseteq A \cap (V' \times V')$. Per denotare che G' è un sottografo di G si scrive $G' \subseteq G$. In particolare, se $A' = A \cap (V' \times V')$, cioè se G' contiene tutti gli archi di G che incidono fra due nodi in V' , si dice che G' è il sottografo indotto da V' .

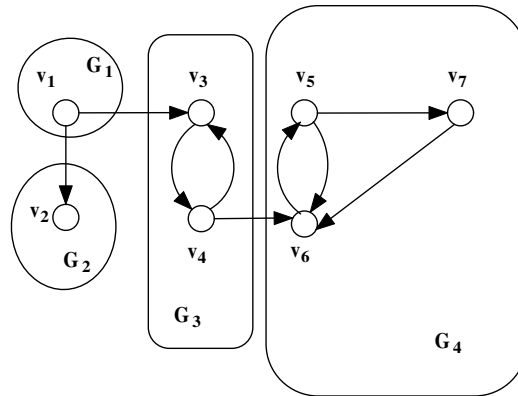


Figura A.4.: un grafo orientato e le sue componenti fortemente connesse.

Definizione 11 Dato un grafo $G = (V, A)$, si definisce componente di G un sottografo $G' \subseteq G$ che sia connesso e massimale, cioè non deve esistere un altro sottografo connesso G'' tale che $G' \subset G'' \subseteq G$.

Definizione 12 Dato un grafo orientato $G = (V, A)$, si definisce componente fortemente connessa di G un sottografo $G' \subseteq G$ che sia fortemente connesso e massimale, cioè non deve esistere un altro sottografo fortemente connesso G'' tale che $G' \subset G'' \subseteq G$.

Proposizione 5 Se un grafo (ovvero, grafo orientato) $G = (V, A)$ ha r componenti (ovvero, componenti fortemente connesse), l'insieme dei nodi V può essere partizionato in r sottoinsiemi $V = V_1 \cup V_2 \cup \dots \cup V_r$ tali che ogni sottografo indotto da una generica V_i è una componente (ovvero, componente fortemente connessa) di G .

Definizione 13 Dato un grafo orientato $G = (V, A)$, una sua componente fortemente connessa $G' = (V', A')$ è detta assorbente se $A \cap (V' \times (V \setminus V')) = \emptyset$, cioè se non esiste alcun arco in G che esce da un nodo in V' e va verso un nodo non in V' .

In fig. A.4 le componenti G_2 e G_4 sono assorbenti.

Bibliografia

- [1] C. G. Cassandras and S. Lafortune, *Introduction to Discrete Event Systems*. Kluwer Academic Publisher, 1999.
- [2] J. Proth and X. Xie, *Petri Nets: A Tool for Design and Management of Manufacturing Systems*. John Wiley & Sons, 1996.
- [3] P.J.Ramadge and W.M. Wohnam, “The control of discrete event systems,” *Proceedings of IEEE*, vol. 77, pp. 637–659, January 1989.
- [4] R. David and H. Alla, “Petri nets for modelling of dynamical systems: A survey,” *Automatica*, vol. 30, no. 2, pp. 175–202, 1994.
- [5] T. Murata, “Petri nets: Properties, analysis and applications,” *Proceedings of IEEE*, vol. 77, pp. 541–580, April 1989.
- [6] L.E. Holloway, B.H. Krogh, and A. Giua, “A survey of petri nets methods for controlled discrete event systems,” *Discrete Event Dynamic Systems: Theory and Applications*, vol. 7, no. 7, pp. 151–190, 1997.
- [7] M. Silva and E. Teruel, “Petri nets for the design and operation of manufacturing systems,” *European Journal of Control*, vol. 3, no. 3, pp. 182–199, 1997.
- [8] Bruce H. Krogh and Lawrence E. Holloway, “Synthesis of feedback control logic for discrete manufacturing systems,” *Automatica*, vol. 27, pp. 641–651, July 1991.
- [9] K. Yalimandou, J. Moody, M. Lemmon, and P. Antsaklis, “Feedback control of petri nets based on place invariants,” *Automatica*, vol. 32, pp. 15–28, January 1996.

-
- [10] Y. Li and W. Wonham, “Control of vector discrete-event systems I - the base model,” *IEEE Trans. on Automatic Control*, vol. 38, pp. 1214–1227, August 1993.
 - [11] Y. Li and W. Wonham, “Control of vector discrete-event systems II - controller synthesis,” *IEEE Trans. on Automatic Control*, vol. 39, pp. 512–531, March 1994.
 - [12] A. D. Febbraro and A. Giua, *Sistemi ad eventi discreti*. Milano: McGraw-Hill, 2002.