3. Petri-Netze

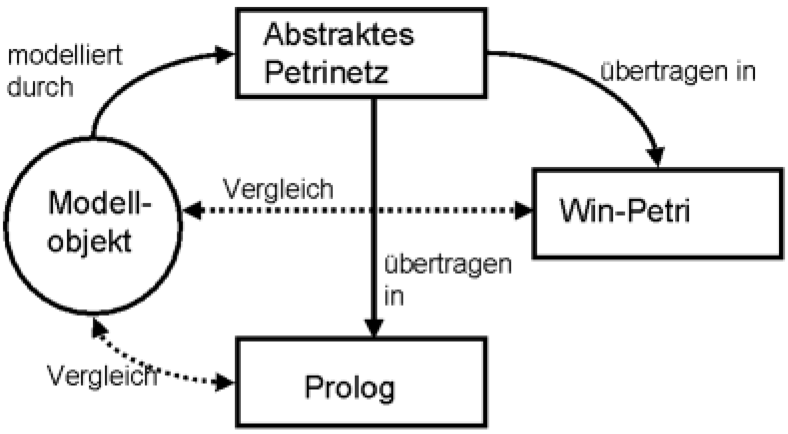
Bei Petri-Netzen handelt es sich um formale Konstrukte, die graphisch ausgestaltet sind und sich für die *Modellierung* und *Analyse* von Systemen und Prozessen eignen. Besonders gut eignen sich Petri-Netze zur Modellierung von diskreten, verteilten Systemen sowie zur Modellierung von Nebenläufigkeit und Parallelität.

Nebenläufigkeit (concurrency) von Prozessen bedeutet, daß die Prozesse unabhängig voneinander ablaufen, sich also nicht gegenseitig beeinflussen. Nichtdeterminismus bedeutet, daß es zu einem Zustand mehrere Folgezustände geben kann. Dieser Nichtdeterminismus kann wie folgt realisiert sein:

1. Eine Folgetransition wird zufällig ausgewählt (z.B. in einer Simulationsumgebung für Petri-Netze)
2. Der Benutzer wählt die Folgetransition aus.

Die Theorie der Petri-Netze geht auf die Dissertation von C.A. Petri ("Kommunikation mit Automaten", Bonn 1962) zurück. Seitdem haben Petri-Netze weltweit Verbreitung gefunden und werden in vielen Anwendungsgebieten eingesetzt, z.B. bei Ablaufbeschreibungen in ingenieurwissenschaftlichen Anwendungen, zur Modellierung von Büroabläufen (work flow) , Produktionsprozessen oder zur Prozessmodellierung bei Betriebssystemen.

Modellierungsaspekte bei Petri-Netzen:



Ein Modellobjekt, beispielsweise eine Ampelschaltung, wird zunächst durch ein abstraktes Petri-Netz modelliert. Abstrakt bedeutet, daß Petri-Netze formal abstrakt beschrieben werden müssen. Auf den entsprechenden Formalismus gehen wir später ein. Das abstrakte Petri-Netz kann anschließend mit einem Petri-Netz Werkzeug wie Winpetri visualisiert und analysiert werden. Ein Nachteil bei diesem Vorgehen ist, daß sich Simulation und Analyse des Petri-Netzes an der fest vorgegebenen Semantik des Petri-Netz Werkzeuges orientieren müssen. Ein flexiblerer Ansatz besteht darin, das abstrakte Petri-Netz in Prolog zu übertragen. Hier ist die Semantik frei definierbar.

Nach Übertragung eines abstrakten Petri-Netzes in Winpetri bzw. Prolog kann jederzeit ein Vergleich mit dem Modellobjekt erfolgen, um beispielsweise Stärken und Schwächen des erstellten Modells ermitteln zu können.

3.1 Netzgraphen (Struktur)

Ein Petri-Netz stellt aus graphentheoretischer Sicht einen bipartiten Graphen dar, der aus zwei verschiedenen Sorten von Knoten besteht: Stellen und Transitionen. Eine Stelle entspricht einer Zwischenablage für Daten bzw. Informationen und wird durch einen Kreis symbolisiert. Eine Transition beschreibt die Verarbeitung von Daten bzw. Informationen und wird durch ein Rechteck oder einen Balken symbolisiert. Ferner existieren Kanten, die jeweils nur von einer Knotensorte zur anderen führen dürfen.

Mathematisch läßt sich die Struktur eines Petri-Netzes wie folgt definieren:

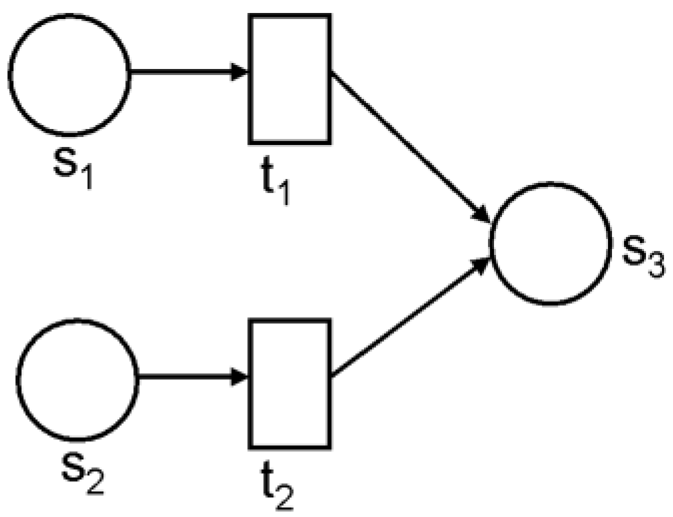
Definition **"Netzgraph"**

Ein Netzgraph N besteht aus den Komponenten S (Menge der Stellen), T (Menge der Transitionen) und F (Flußrelation), kurz: N=(S, T, F) wobei gilt:

i. $S \cap T = \emptyset$
ii. $F \subseteq (S \times T) \cup (T \times S)$

Stellen und Transitionen werden als Knoten bezeichnet. Die Elemente der Flußrelation nennt man Kanten.

Beispiel:



Das Netz N = (S, T, F) läßt sich wie folgt formalisieren:

S = {s_1, s_2, s_3)

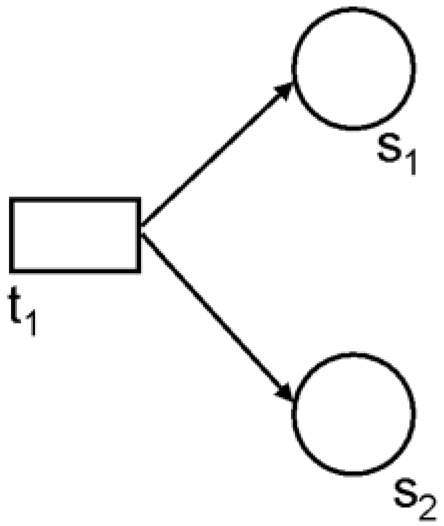
T = {t_1, t_2}

F = {(s_1, t_1), (s_2, t_2), (t_1, s_3), (t_2, s_3)}

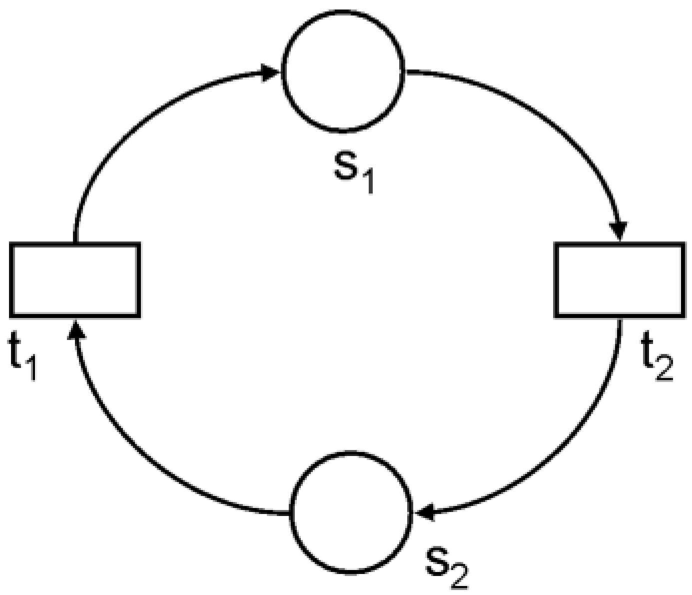
Bezeichnungen:

- **Für einen Knoten x∈ (S ∪ T) heißt • x = {y | (y, x) ∈ F} der Vorbereich von x. Wenn x eine Transition ist, dann bezeichnet • x alle Eingangs- oder Inputstellen.**
- **Für einen Knoten x∈ (S ∪ T) heißt x• = {y | (x, y) ∈ F} der Nachbereich von x. Wenn x eine Transition ist, dann bezeichnet x• alle Ausgangs- oder Outputstellen.**
- **Verzweigung:**
  x heißt vorwärtsverzweigt, falls |x• |>1 und rückwärtsverzweigt, falls |• x|>1

Beispiel für ein vorwärtsverzweigtes Netz:



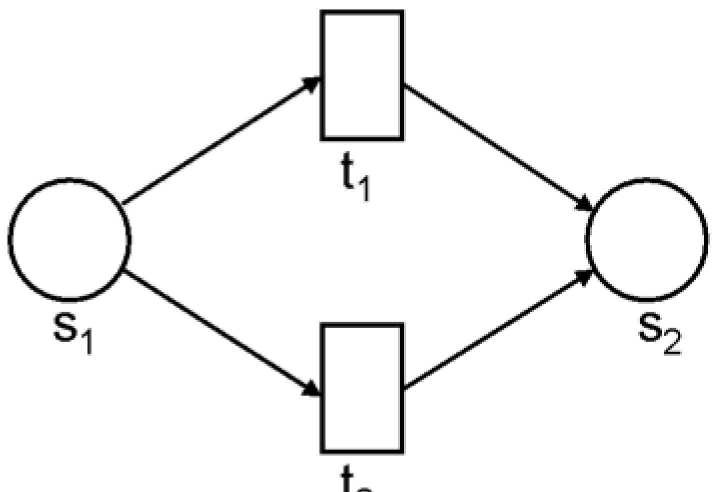
Beispiel für ein unverzweigtes Netz:



Definition **"schlicht"**

Ein Netz heißt schlicht, falls $\forall$ x,y: (($\bullet$ x = $\bullet$ y $\wedge$ x$\bullet$ = y$\bullet$) $\rightarrow$ x = y), wobei x und y Knoten sind. Mit anderen Worten: Es kann nicht zwei verschiedene Knoten x und y geben, die denselben Vor- und Nachbereich haben.

Beispiel:



Das Netz ist nicht schlicht, weil z.B. t_1 und t_2 denselben Vor- und Nachbereich haben, aber voneinander verschiedene Knoten sind.

Definition **"Teilnetz"**

Sei N = (S, T, F) ein Netz. Dann heißt N' = (S', T', F') Teilnetz von N g.d.w.

i. $S' \subseteq S$
ii. $T' \subseteq T$
iii. $F' \subseteq F \cap ((S' \times T') \cup (T' \times S'))$

Im Buch von Baumgarten wird zwischen einem Teilnetz und einem Netzteil unterschieden. Ein Netzteil nach Baumgarten entspricht unserer Definition eines Teilnetzes. Bei einem Teilnetz nach Baumgarten würde in Unterpunkt iii. unserer Definition anstelle von $\subseteq$ ein = stehen. Wir wollen hier nicht zwischen den Begriffen Teilnetz und Netzteil differenzieren, sondern beide Begriffe unter dem Begriff Teilnetz subsummieren.

Definition **"Rand"** eines Teilnetzes N' bzgl. N

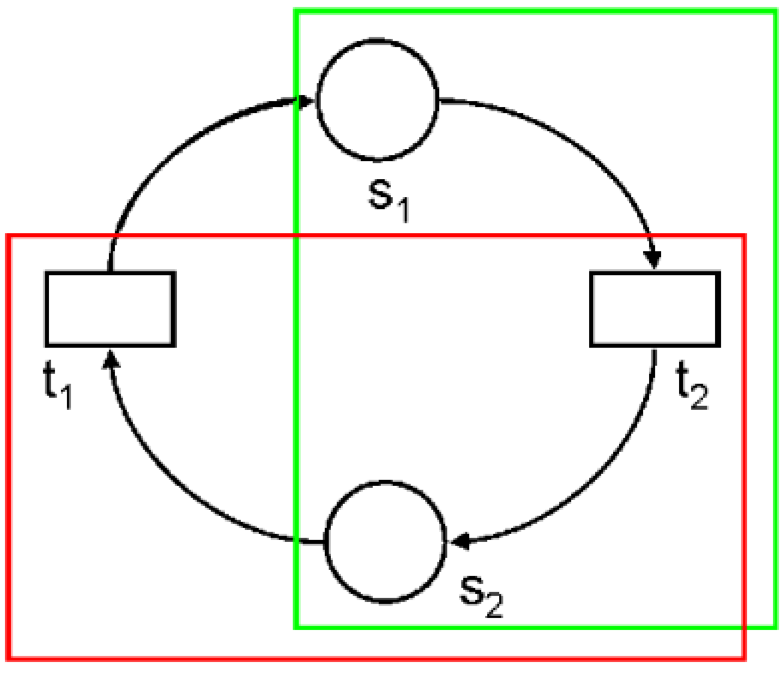
$\text{Rand}(N', N) = \{x \in S' \cup T' \,/\, (\bullet x \cup x\bullet) \setminus (S' \cup T') \neq \emptyset\}$

Beim Rand eines Teilnetzes N' bzgl. N handelt es sich also um alle Knoten des Teilnetzes, die über Kanten mit dem das Teilnetz umgebenden Restnetz verbunden sind.

Man unterscheidet in diesem Zusammenhang auch die Begriffe **"stellenberandet"** und **"transitionsberandet"**:

- Stellenberandet heißt das Teilnetz, wenn im Rand nur Stellen vorkommen, d.h. wenn Rand(N', N) $\subseteq$ S'.
- Transitionsberandet heißt das Teilnetz, wenn im Rand nur Transitionen vorkommen, d.h. wenn Rand(N', N) $\subseteq$ T'.

Beispiel:

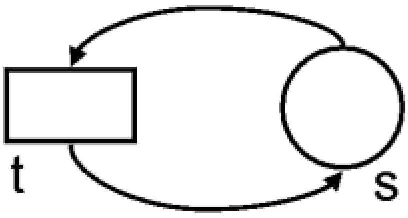


Die grüne Umrandung umschließt ein stellenberandetes, die rote Umrandung ein transitionsberandetes Teilnetz. Ein stellenberandetes Teilnetz läßt sich durch eine Stelle, ein transitionsberandetes Teilnetz durch eine Transition ersetzen.

Definition **"Schlinge"**

Ein Teilnetz bestehend aus einer Stelle s und einer Transition t, die doppelt verbunden sind (d.h. $\{(s, t), (t, s)\} \subseteq F$), heißt Schlinge.

Beispiel:

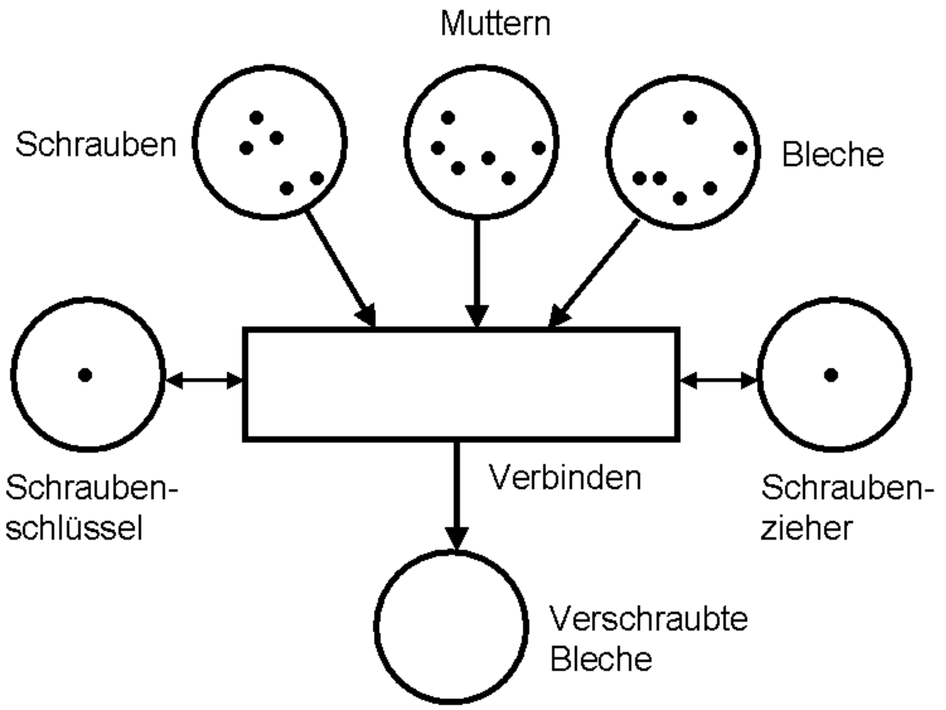


Häufig werden die beiden Pfeile auch zu einem Doppelpfeil zusammengefaßt.

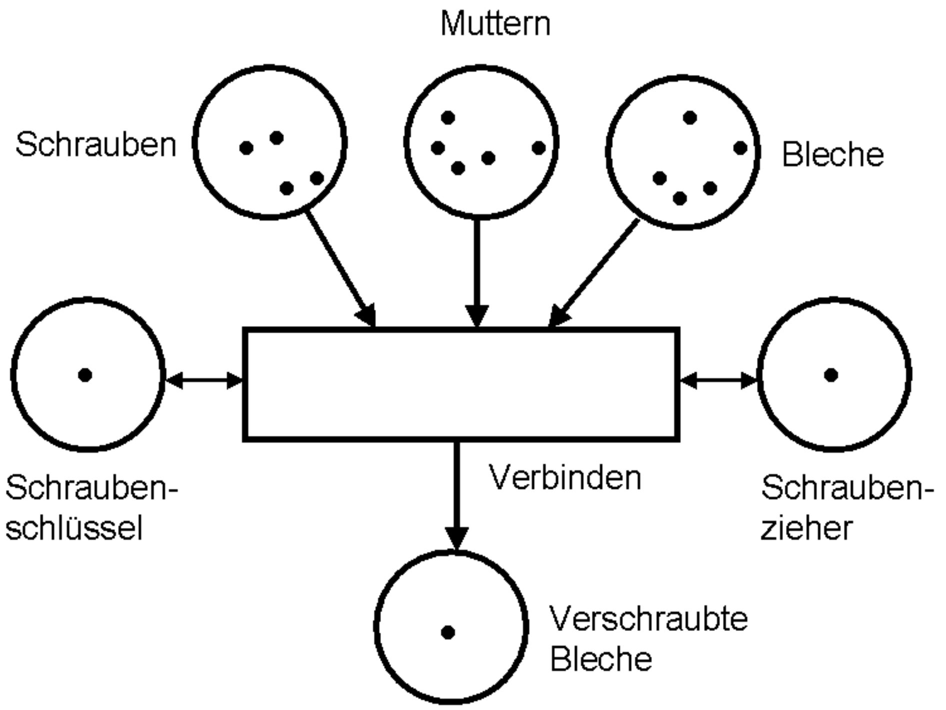
3.2. Dynamik von Netzen (Stellen-Transitions-Systeme, kurz: S/T-Systeme)

S/T-Systeme sind prozessorientierte Interpretationen von Netzgraphen, die durch Hinzunahme nicht unterscheidbarer (anonymer) Marken entstehen. Um dynamische Vorgänge zu beschreiben, werden die Stellen mit Marken "belegt", die entweder durch schwarze Punkte oder Zahlen (Winpetri) in den Stellen symbolisiert werden können. Beim Schalten einer Transition werden Marken an Inputstellen verbraucht und an Outputstellen erzeugt, wobei sich ihre Gesamtzahl meistens ändert.

Beispiel: Bleche und Schrauben



Nach dem Schalten der Transition "Verbinden" sind Bleche, Schrauben und Muttern miteinander verbunden. Der Schraubenschlüssel wird durch das Schalten der Transition nicht verbraucht und steht daher nach dem Schalten der Transition wieder zur Verfügung (Schlinge). Die Gesamtzahl der Marken hat sich nach einmaligem Schalten der Transition geändert:



Nachfolgend werden wir den Begriff S/T-System sowie die Dynamik von S/T-Systemen formalisieren:

Def.: **"S/T-System"**

Eine Struktur $Y=(S, T, F, K, W, M_0)$ heißt S/T-System g.d.w.

i.) (S, T, F)	ein Netzgraph
ii.) $K: S \rightarrow IN \cup \{\infty\}$	"Kapazität für Stellen
iii.) $W: F \rightarrow IN$	"Kantengewichte"
iv.) $M_0: S \rightarrow IN_0$	Anfangsmarkierung mit $M_0(s) \leq K(s)$ für alle s

Hinweis: Standardwerte ("Default") sind $K=\infty$ und $W=1$.

Allgemein heißt jede Zuordnung $M: S \rightarrow IN_0$ mit $M(s) \leq K(s)$ für alle s eine **Markierung** oder **Belegung.**

Def.: **"aktiviert"**

Man bezeichnet eine Transition $t \in T$ als aktiviert unter einer Markierung M g.d.w.

" $s \in \bullet t$: $M(s) \geq W(s, t)$ und

" $s \in t\bullet : M(s) + W(t, s) \leq K(s)$

Abkürzend schreibt man für die Aktiviertheit einer Transition auch M[t>.

Def.: **"Schalten"**

Man sagt t schaltet von M nach M’ falls M[t> gilt und

$$M'(s) = \begin{cases} M(s) - W(s, t) \text{ für } s \in \bullet t \setminus t\bullet \\ M(s) + W(t, s) \text{ für } s \in t\bullet \setminus \bullet t \\ M(s) - W(s, t) + W(t, s) \text{ für } s \in \bullet t \cap t\bullet \\ M(s) \text{ sonst } (s \notin \bullet t \cup t\bullet) \end{cases}$$

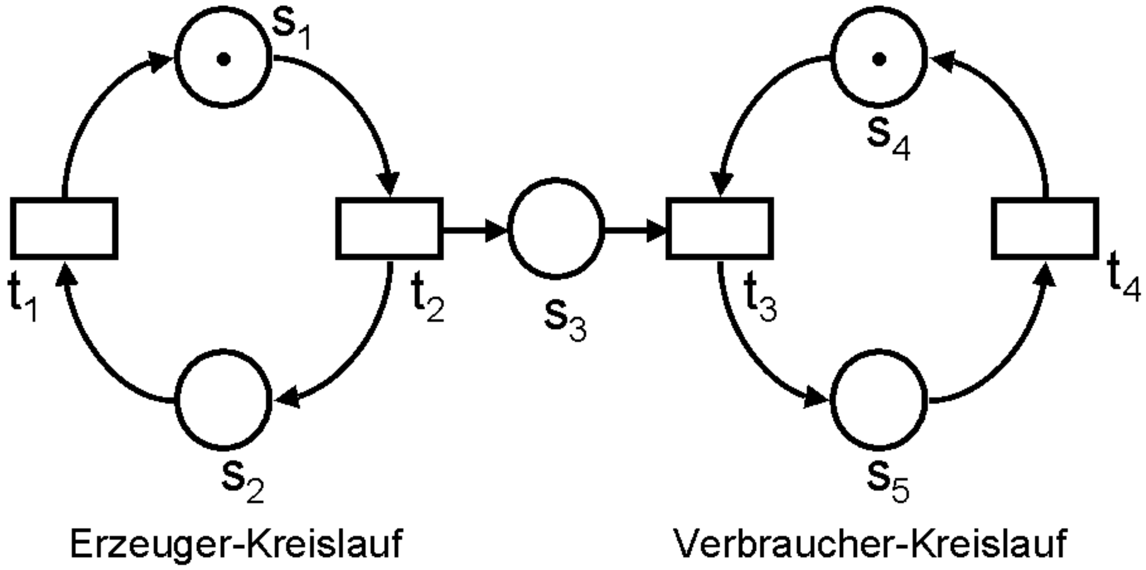
Formal wird der Schaltvorgang als M[t>M' geschrieben, wobei M' als die Folgemarkierung von M bezeichnet wird.

Werden in obiger Fallbeschreibung die Kantengewichte von Kanten, die nicht vorkommen, auf 0 gesetzt, so genügt als allgemeiner Fall die dritte Fallzeile:

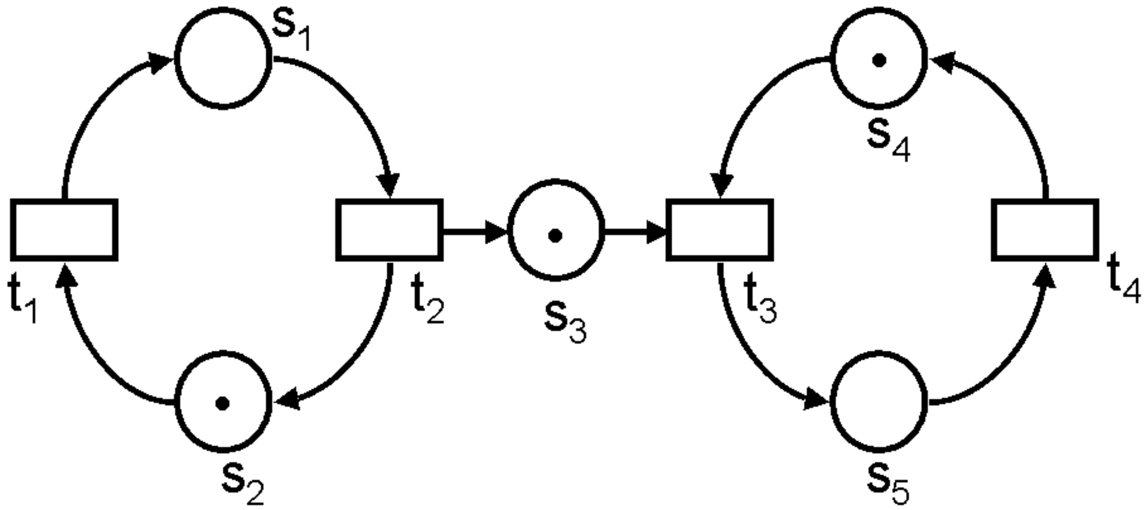
$M'(s) = M(s) - W(s, t) + W(t, s)$ für $s \in \bullet t \cap t\bullet$

Beispiel:

Nachfolgend ist ein Erzeuger-Verbraucher-System mit der Anfangsmarkierung $M_0=\{(s_1, 1), (s_2, 0), (s_3, 0), (s_4, 1), (s_5, 0)\}$ gezeigt.



Schaltet t_2, so ergibt sich die Folgemarkierung M_1={(s_1, 0), (s_2, 1), (s_3, 1), (s_4, 1), (s_5, 0)}:



Bei dem Schaltvorgang hat sich die Gesamtzahl der Marken erhöht.

Um sich einen Überblick über das Schaltverhalten eines S/T-Systems zu verschaffen, kann man für alle beliebigen Schaltfolgen die Menge der erreichbaren Folgemarkierungen eines S/T-Systems ermitteln. Diese Menge bezeichnet man als Erreichbarkeitsmenge und den Vorgang des Ermittelns dieser Menge als Erreichbarkeitsanalyse. Zunächst zur Definition der Begriffe Schaltfolge und Erreichbarkeitsmenge:

Def. **"Schaltfolge"**

Eine Schaltfolge $M_0[t_1...t_n>$ steht für $\exists M_i: M_0[t_1>M_1[t_2>...M_{n-1}[t_n>M_n$.

Voraussetzung für das Schalten der Schaltfolge ist natürlich, daß zu jeder Zwischenmarkierung M_i die entsprechende Transition t_{i+1} auch wirklich aktiviert ist.

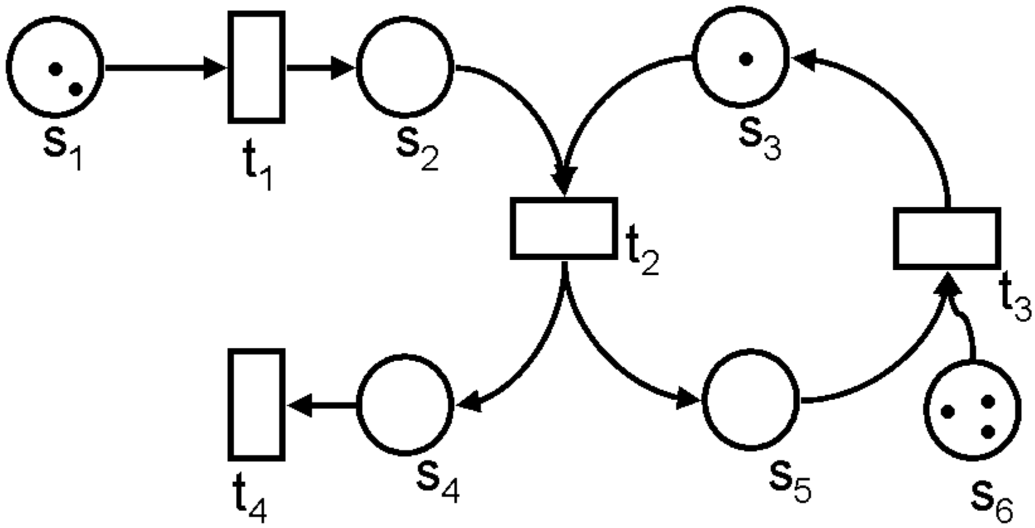
Def.: **"Erreichbarkeitsmenge"**

$[M_0>:=\{M' \ / \ M[w>M', w\in T^*\}$

Hierbei ist w ein Wort aus Transitionen, welches Element der Menge aller beliebigen Wörter aus Transitionen T* (incl. des leeren Wortes) ist.

Erreichbarkeitsanalyse

Die Erreichbarkeitsanalyse erfolgt durch systematisches Abprüfen aller Schaltmöglichkeiten eines S/T-Systems, wobei die erreichbaren Markierungen im allgemeinen tabellarisch aufgelistet werden. Dieses Vorgehen wird an nachfolgendem S/T-System X_{bsp} = (S, T, F, K, W, M_0) mit S={s_1, s_2, s_3, s_4, s_5, s_6}, T={t_1, t_2, t_3, t_4}, F={(s_1, t_1), (t_1, s_2), (s_2, t_2), (t_2, s_4), (s_4, t_4), (s_3, t_2), (t_2, s_5), (s_5, t_3), (t_3, s_3), (s_6, t_3)} und M_0={(s_1, 2), (s_2, 0), (s_3, 1), (s_4, 0), (s_5, 0), (s_6, 3)}erläutert. Die Kanten haben das Kantengewicht 1, die Stellen die Kapazität oo.



Erreichbarkeitstabelle:

	s_1	s_2	s_3	s_4	s_5	s_6
M_0	2	0	1	0	0	3
M_1	1	1	1	0	0	3
M_2	0	2	1	0	0	3
M_3	1	0	0	1	1	3
M_4	0	1	0	1	1	3
M_5	1	0	1	1	0	2
M_6	1	0	0	0	1	3
M_7	0	1	1	1	0	2
M_8	0	1	0	0	1	3
M_9	1	0	1	0	0	2
M_{10}	0	0	0	2	1	2
M_{11}	0	1	1	0	0	2
M_{12}	0	0	1	2	0	1
M_{13}	0	0	0	1	1	2
M_{14}	0	0	1	1	0	1
M_{15}	0	0	0	0	1	2
M_{16}	0	0	1	0	0	1

Die Vorgehensweise zum Erstellen der obigen Erreichbarkeitstabelle läßt sich wie folgt beschreiben:

a. Zunächst wird die Anfangsmarkierung M_0 in die Tabelle eingetragen. M_0 ist jetzt die aktuelle Markierung M_c. Für M_c wird der Reihe nach jede Transition t darauf untersucht, ob sie aktiviert ist.
    - Wenn nein, sind wir momentan mit dieser Transition fertig

- Wenn ja, wird die Folgemarkierung zu M_c ermittelt.

  - Ist die Folgemarkierung neu, wird sie als neue Markierung in die Erreichbarkeitstabelle eingetragen.
  - Ist sie bereits in der Erreichbarkeitstabelle vorhanden, so erfolgt kein neuer Eintrag.

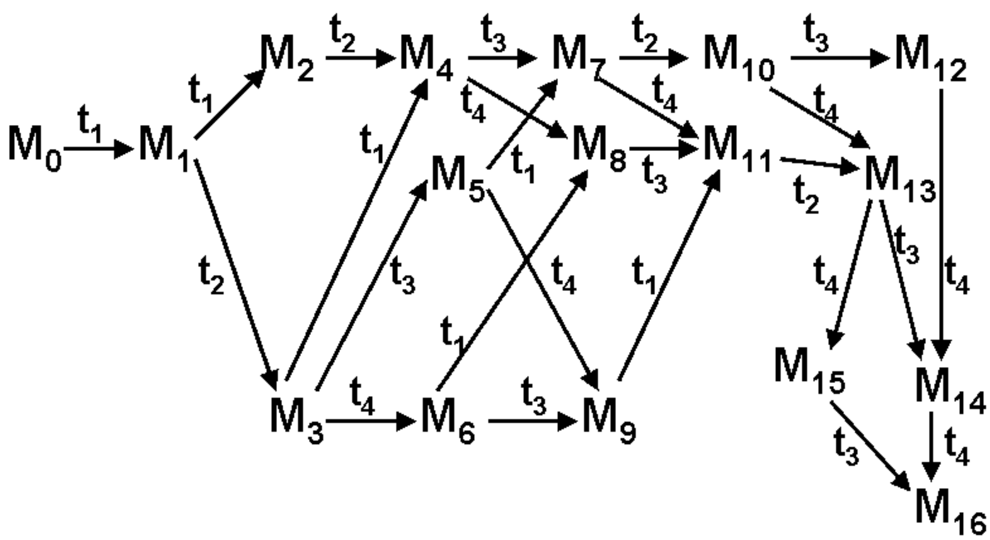
b. Für die in a. ermittelten Folgemarkierungen wird wie in a. beschrieben verfahren, d.h. die jeweilige Folgemarkierung ist die neue aktuelle Markierung M_c.

c. Sind alle in die Tabelle eingetragenen Markierungen nach dem in a. beschriebenen Verfahren untersucht, so ist man mit der Erreichbarkeitsanalyse fertig.

Erreichbarkeitsgraph:

Das Ergebnis einer Erreichbarkeitsanalyse kann auch mit Hilfe eines Erreichbarkeitsgraphen veranschaulicht werden. Dies sei an obigem S/T-System X_{bsp} veranschaulicht:

Der zu diesem Petri-Netz gehörige endliche Erreichbarkeitsgraph sieht wie folgt aus:

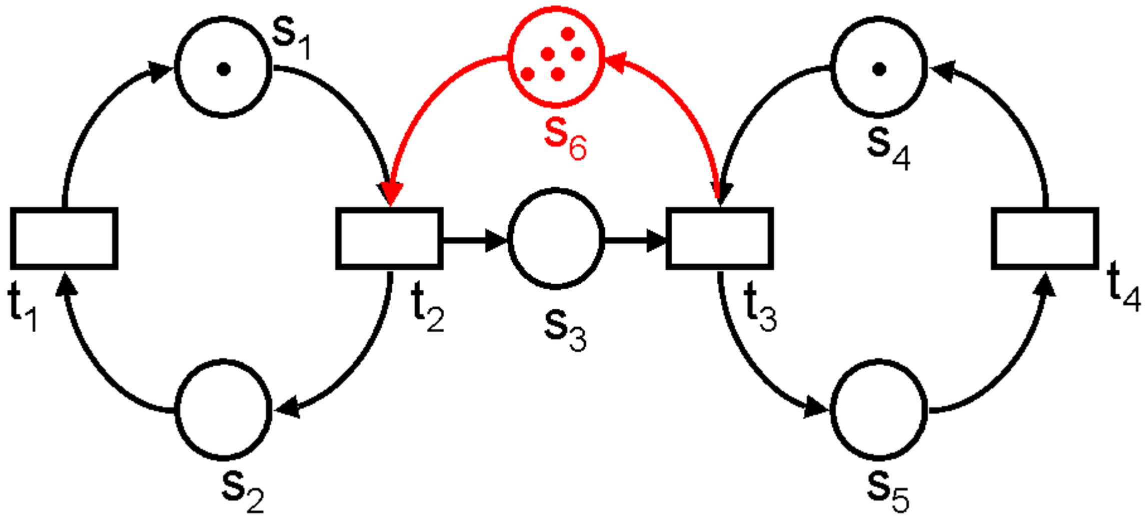


Erreichbarkeitsanalysen können auch unendlich sein. Ein Beispiel für eine unendliche Erreichbarkeitsanalyse liefert das oben dargestellte Erzeuger-Verbraucher-System:

	s_1	**s_2**	**s_3**	**s_4**	**s_5**	**Aktivierte Transitionen**
M_0	1	0	0	1	0	$M_0[t_2>M_1$
M_1	0	1	1	1	0	$M_1[t_1>M_2$ $M_1[t_3>M_3$
M_2	1	0	1	1	0	$M_2[t_3>M_4$ $M_2[t_2>M_5$
M_3	0	1	0	0	1	$M_3[t_1>M_4$ $M_3[t_4>M_6$
M_4	1	0	0	0	1	$M_4[t_4>M_0$ $M_4[t_2>M_7$
M_5	0	1	2	1	0	...
M_6	0	1	0	1	0	...
M_7	0	1	1	0	1	...

Man erkennt, daß der Zwischenspeicher s_3 beliebig "aufgepumpt" werden kann. Die Erreichbarkeitsanalyse ist somit unendlich. Die unendliche Erreichbarkeitsmenge wird formal als $|[M_0>|$ = oo geschrieben.

Durch Zufügen einer weiteren Stelle s_6 kann erreicht werden, daß nur noch endlich viele Markierungen erreichbar sind:



Die Summe der Marken in s_3 und s_6 ist immer gleich 5. Die Vorratsproduktion auf s_3 und damit die unendliche Erreichbarkeitsmenge können so vermieden werden.

Vektoradditionssysteme:

Die Grundideen bei Vektoradditionssystemen sind die folgenden:

- Markierungen werden als Tupel (Vektoren) dargestellt
- Eine sog. Inzidenzmatrix codiert alle Transitionen

Der Begriff "Inzidenzmatrix" wird nachfolgend definiert:

Def.: **"Inzidenzmatrix"**

Sei N = (S, T, F) das Netz eines S/T-Systems mit endlicher Stellen- und Transitionenmenge:

$S=\{s_1, ..., s_m\}$

$T=\{t_1, ..., t_n\}$

Die Inzidenzmatrix $C=(C_{ab})$ mit $a\in \{1..m\}$ und $b\in \{1..n\}$ ist dann wie folgt gegeben:

$$C_{ab} = \begin{cases} W(t_b, s_a) \text{ falls } (t_b, s_a) \in F \setminus F^{-1} \\ -W(s_a, t_b) \text{ falls } (s_a, t_b) \in F \setminus F^{-1} \\ W(t_b, s_a) - W(s_a, t_b) \text{ falls } (t_b, s_a) \in F \cap F^{-1} \\ 0 \text{ sonst} \end{cases}$$

Hinweis: $F \setminus F^{-1}$ bedeutet, daß Transition und Stelle zwar miteinander verbunden sind, aber keine Schlinge vorliegt. F^{-1} ist die Umkehrrelation der Flußrelation.

Auch hier gilt wiederum $C_{ab} = W(t_b, s_a) - W(s_a, t_b)$ falls wir $W(t_b, s_a)$ bzw. $W(s_a, t_b)=0$ für $W(t_b, s_a)$ bzw. $W(s_a, t_b) \notin F$ annehmen.

Allgemein läßt sich die Inzidenzmatrix wie folgt darstellen:

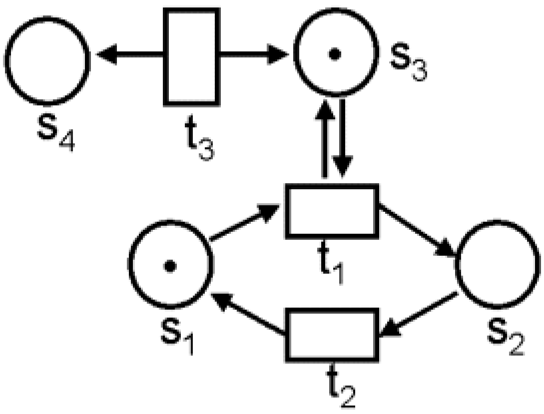
$$C = \begin{pmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & & \vdots \\ c_{m1} & \cdots & c_{mn} \end{pmatrix}$$

Für das bereits vorgestellte Erzeuger-Verbraucher-System ergibt sich die folgende Inzidenzmatrix:

$$C = \begin{array}{cc} \begin{pmatrix} 1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 1 & -1 \end{pmatrix} & \begin{array}{c} s1 \\ s2 \\ s3 \\ s4 \\ s5 \end{array} \\ \begin{array}{cccc} t1 & t2 & t3 & t4 \end{array} & \end{array}$$

Jede Spalte entspricht einem Vektor, der wiederum den Beitrag der jeweiligen Transition zu den Stellen angibt. Beispielsweise wird durch Schalten von t_1 eine Marke zu s_1 hinzugefügt, während gleichzeitig von s_2 eine Marke abgezogen wird.

Die Inzidenzmatrix spiegelt alle wesentlichen Eigenschaften eines Netzes wider - bis auf die Schlingen. Dies soll an nachfolgendem Beispiel gezeigt werden:



Die zugehörige Inzidenzmatrix ist die folgende:

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

Man sieht hier in der Inzidenzmatrix in Spalte 1, Zeile 3 der Inzidenzmatrix lediglich, daß nach Schalten von t_1 sich die Markenzahl von s_3 um 0 Marken verändert hat. Dies erfolgte durch Abziehen einer Marke von s_3 und anschließendes Hinzufügen einer Marke zu s_3. Letzteres kann man an der Inzidenzmatrix allein aber nicht erkennen, sondern muß das Netz zur Hilfe nehmen.

Wir erhalten den zu einer Transition gehörigen **"Schaltvektor"** durch Multiplikation der Inzidenzmatrix mit dem charakteristischen Vektor der Transition. Der zur Transition t_1 gehörende charakteristische Vektor ist der folgende:

$$\vec{\delta}(t_1) = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \text{ bei } t_1, \ldots, t_3$$

Der obige charakteristische Vektor "extrahiert" nun die erste Spalte aus der Inzidenzmatrix und liefert damit den Schaltvektor:

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

Allgemein formuliert wird durch den charakteristischen Vektor die j-te Spalte aus einer Inzidenzmatrix "extrahiert", wobei j die Stelle im charakteristischen Vektor bezeichnet, an welcher eine 1 vorkommt. Es gilt also:

$\vec{c}(t_j) = C \bullet \vec{\delta}(t_j)$

$\vec{c}(t_j)$: Schaltvektor (j-te Spalte)

$\vec{\delta}(t_j)$: charakteristischer Vektor

Die Folgemarkierung nach Schalten einer Transition t_j ist hiermit so darstellbar:

$M[t_j> M'$ mit

$$M' = M + C \bullet \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} \leftarrow \text{j-te Spalte}$$

Die Wirkung einer **Schaltfolge** läßt sich entsprechend durch den **Häufigkeitsvektor ("Parikh-Vektor")** darstellen:

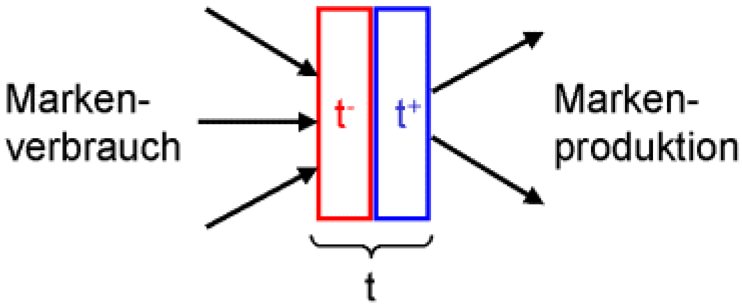
Sei $T=\{t_1, t_3, t_1, t_4\}$ und sei unter einer Markierung M die Folge $\upsilon=t_1t_3t_1t_4$ aktiviert.

Mit $\vec{\upsilon} = \begin{bmatrix} 2 \\ 0 \\ 1 \\ 1 \end{bmatrix}$ (entsprechend auftretender Häufigkeit)

gilt dann für $M[\upsilon>M'$: $M' = M+C \bullet \vec{\upsilon}$

Bemerkungen:

a. Aus der obigen Darstellung ergibt sich die Invarianz der Folgemarkierung unter (aktivierten!) Permutationen einer Schaltfolge.
b. Die Aktiviertheit bedarf einer genaueren Prüfung. Dieser Prüfung liegt die folgende Idee zugrunde:



Wir vereinbaren, daß zuerst t^- und anschließend t^+ ausgeführt wird. Dabei ist:

$$\vec{c}(t_j^-) = \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix} \text{ mit } x_i = \begin{cases} -W(s_i, t_j) \text{ falls } s_i \in \bullet t_j \\ 0 \text{ sonst} \end{cases}$$

$$\vec{c}(t_j^+) = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} \text{ mit } y_i = \begin{cases} W(t_j\ s_i) \text{ falls } s_i \in t_j \bullet \\ 0 \text{ sonst} \end{cases}$$

Es gilt natürlich:

$$\vec{c}(t_j) = \vec{c}(t_j^-) + c(t_j^+)$$

Damit folgt nun für die Aktiviertheit einer Transition:

Eine Transition ist ausführbar (d.h. aktiviert) unter einer Belegung M g.d.w. $M[t_j^->M'$ sowie $M'[t_j^+>$.

3.3 Grundbegriffe - Grundsituationen

Definition **"Lebendigkeit"**, **"Deadlock"**

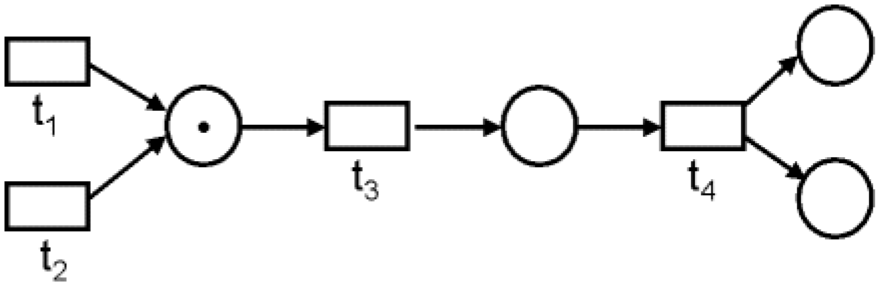
a. Ein S/T-System heißt "tot" unter einer Belegung M, falls keine Transition unter M aktiviert ist. Mit anderen Worten gilt: $\neg$ M[t> für alle $t \in T$.
b. Ein S/T-System mit der Anfangsbelegung M_0 heißt (schwach) lebendig oder "Deadlock-frei", wenn jede erreichbare Markierung wenigstens eine Transition aktiviert. Mit anderen Worten gilt: $\forall M' \in [M_0> \exists t \in T$: M[t>

Definition **"Kausalität"**

In einem S/T-System mit Anfangsmarkierung M_0 ist die Transition $t_1 \in T$ notwendige

Bedingung für das Schalten von $t_2 \in T$ g.d.w. $\forall w \in T^*: M_0[wt_2> \Rightarrow$ w enthält t_1.

Beispiel:



Das Schalten von t_3 ist eine notwendige Bedingung für das Schalten von t_4.

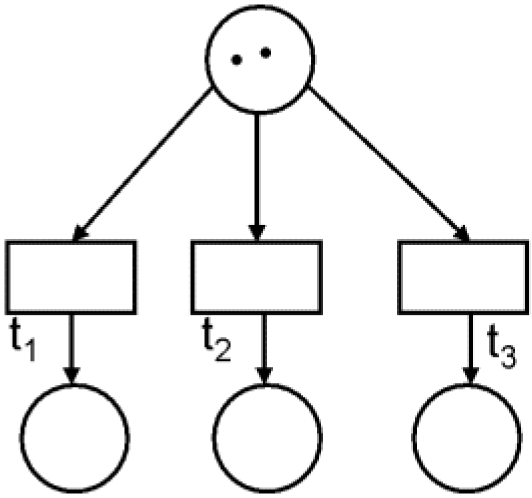
Bemerkungen:

a. Die "notwendige Bedingung" ist eine transitive Relation.
b. Die Kausalität ist besonders ausgeprägt in sog. "Kausalnetzen" (kreisfreie Netze)

"Nebenläufigkeit"

Dieser Begriff bezeichnet die Unabhängigkeit von Transitionen (Ereignissen, Ereignisfolgen).

Beispiel:



In diesem Beispiel können t_1, t_2 und t_3 paarweise unabhängig voneinander schalten, jedoch nicht alle drei Transitionen. Die paarweise Unabhängigkeit der Transitionen kann man überprüfen, indem man für alle möglichen Zweierkombinationen von Transitionen überprüft, ob die aus der jeweiligen Zweierkombination gebildete Transitionenfolge als Schaltfolge schaltbereit ist.

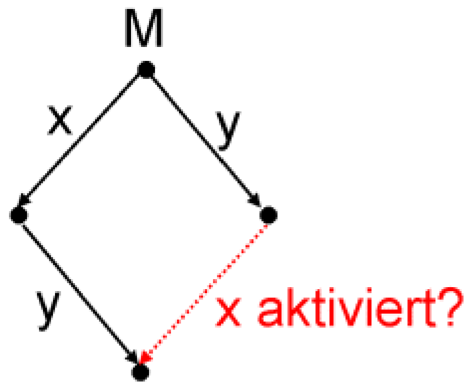
Definition **"nebenläufige Aktivierung"**

Eine Liste von Transitionen (d.h. eine Zusammenfassung, ggf. mit Wiederholungen) $T=[t_1, ..., t_n]$ heißt "nebenläufig aktiviert" unter einer Markierung M, falls alle Permutationen von T unter M als Schaltfolgen aktiviert sind. Aufgrund der Ergebnisinvarianz bei Umordnung können wir schreiben:

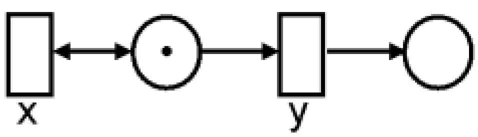
$M[[t_1, ..., t_n]>M'$ mit eindeutig bestimmtem M' ($M' = Mt_1, ..., t_n$).

Anmerkungen:

a. Falls T nur aus unterschiedlichen Elementen besteht, können wir auch $T=\{t_1, ..., t_n\}$, $M[\{t_1, ..., t_n\}>$ schreiben.
b. Baumgarten verwendet die [ ]-Schreibweise für "starke Nebenläufigkeit" im Sinne der Nebenläufigkeit von $[t_1^-, t_1^+, t_2^-, ..., t_n^-, t_n^+]$
c. Das Beispiel zeigt, dass aus paarweiser Nebenläufigkeit (von $t_1, .., t_3$) nicht auf Nebenläufigkeit von $[t_1, t_2, t_3]$ geschlossen werden kann.
   Unter gewissen Voraussetzungen kann jedoch auf Nebenläufigkeit geschlossen werden:
   Satz:
   Betrachte ein schlingenfreies S/T-System mit K(s)= oo für alle $s \in S$. Dann gilt für beliebige Markierungen M und je zwei Transitionen $x,y \in T$:
   $M[y> \wedge M[xy> \Rightarrow M[yx>$.
   D.h. es gilt auch M[[x,y]> im Sinne der Nebenläufigkeit. Dies läßt sich mit Hilfe des folgenden Schemas veranschaulichen:



   Die Schlingenfreiheit ist eine notwendige Voraussetzung, d.h. die Behauptung gilt nicht für Netze mit Schlingen. Dies kann man sich an folgendem Beispiel klar machen:



   Beweis des Quadratschlusses:
   Wir führen zunächst den Begriff der **"Schwellenmarkierung"** ein ($s \in S$):

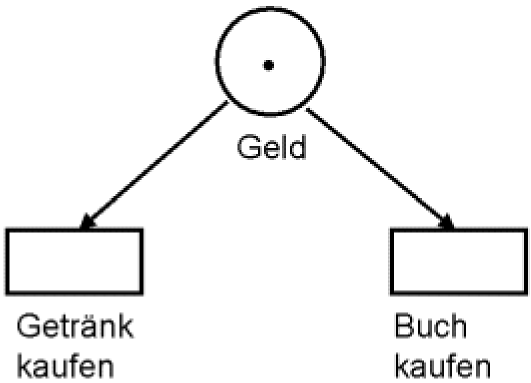
$$M_t^-(s) := \begin{cases} W(s, t) \text{ falls } s \in \bullet t \\ 0 \text{ sonst} \end{cases}$$

$$M_t^+(s) := \begin{cases} W(t, s) \text{ falls } s \in t \bullet \\ 0 \text{ sonst} \end{cases}$$

Sei nun $s \in S$ eine beliebige Stelle aus $\bullet x \cap \bullet y$, also aus dem gemeinsamen Vorbereich von x und y.
Zu zeigen: Nach Schalten von y wird die Schwellenmarkierung von x nicht unterschritten, so daß anschließend x noch schalten kann. Mit anderen Worten muß gezeigt werden:
$M(s) - M_y^-(s) + M_y^+(s) \geq M_x^-(s)$
In diesem Zusammenhang interessiert uns nur die Stelle, die sowohl im Vorbereich von x als auch im Vorbereich von y liegt, also $s \in \bullet x \cap \bullet y$. Aufgrund der Hypothese des Satzes ist:
i. $M(s) \geq M_y^-(s)$
ii. $M(s) \geq M_x^-(s)$
iii. $M(s) - M_x^-(s) + M_x^+(s) \geq M_y^-(s)$
Wegen der vorausgesetzten Schlingenfreiheit und da wir nur die Stelle $s \in \bullet x \cap \bullet y$ betrachten, gilt: $M_y^+(s)=0$ und $M_x^+(s)=0$.
Damit folgt:
i. $M(s) - M_x^-(s) + M_x^+(s) \geq M_y^-(s)$
ii. $M(s) - M_x^-(s) + M_y^+(s) \geq M_y^-(s)$ (wegen $M_y^+(s)= M_x^+(s)=0$)
Durch Umformen der Gleichung ii. ergibt sich:
$M(s) - M_y^-(s) + M_y^+(s) \geq M_x^-(s)$
Also gilt M[yx>. Dies war zu zeigen.

Konflikt
Als Konflikt wird eine nicht-nebenläufige Aktivierung mehrerer Transitionen bezeichnet.

Beispiel:



Definition **"Konflikt"**

Zwei verschiedene Transitionen t, t'$\in$ T stehen unter Belegung M und K(s)=oo für alle s "in Konflikt" miteinander g.d.w.

i. M[t>
ii. M[t'>
iii. $\neg$ M[[t, t']>

Es gilt außerdem der folgende **Satz**:

a. Zwei in Konflikt stehende Transitionen t_1, t_2 haben mindestens eine gemeinsame markentragende Inputstelle $s \in \bullet t_1 \cap \bullet t_2$.
b. Sind außerdem alle Kantengewichte W=1, so kann s für zwei in Konflikt stehende Transitionen t_1 und t_2 so gewählt werden, dass $s \in \bullet t_1 \cap \bullet t_2 \setminus t_1\bullet \cap t_2\bullet$. Dann gilt auch: M(s)=1.

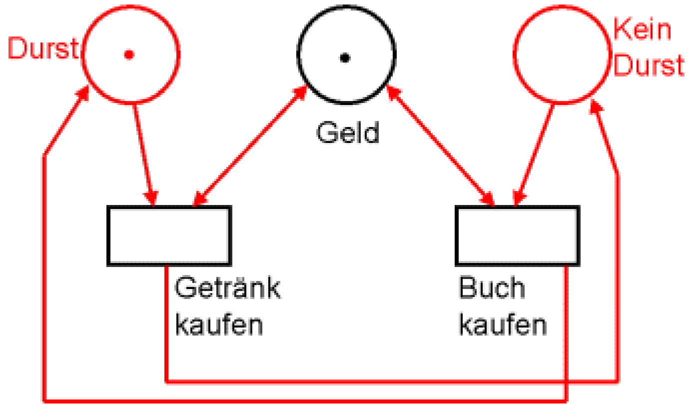
Beweis zu b.:

Sei W=1. Annahme: Es existiert für zwei in Konflikt stehende Transitionen t_1 und t_2 kein $s \in S$ mit der Eigenschaft $s \in \bullet t_1 \cap \bullet t_2 \setminus t_1\bullet \cap t_2\bullet$.

D.h.: Alle gemeinsamen Input-Stellen sind Schlingen bzgl. t_1 und t_2. $\Rightarrow$ $Mt_i(s)$ = M(s)-1+1=M(s) für i=1,2. $\Rightarrow$ Durch s kann kein Konflikt eingeführt werden. Dies ist ein Widerspruch.

Konfliktbeseitigung:

Eine Konfliktbeseitigung ist durch Einführung zusätzlicher Eingabestellen möglich. Dies ist an obigem Beispiel gezeigt:

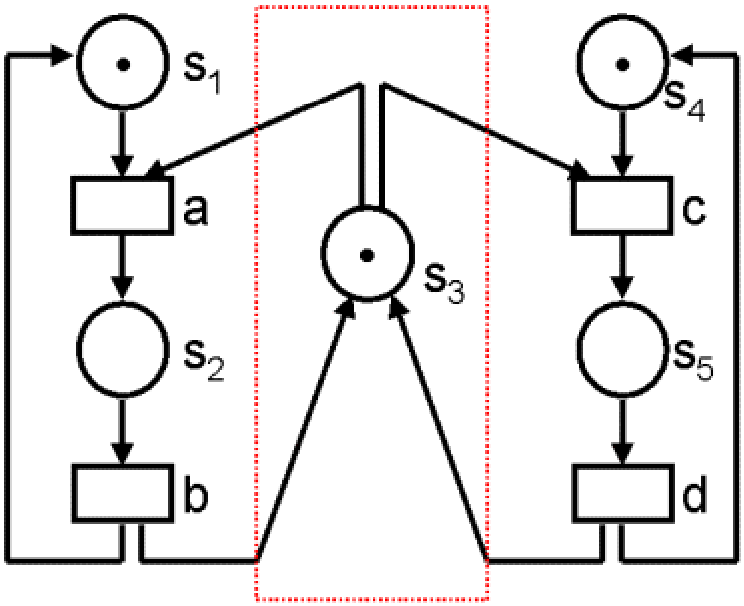


Die zusätzlich eingeführten Stellen und Kanten (rot markiert) beseitigen den Konflikt zwischen den Transitionen "Getränk kaufen" und "Buch kaufen", denn die Transition "buch kaufen" ist nicht mehr schaltbereit.

"Synchronisation"

Unter Synchronisation versteht man im allgemeinen das Einführen von Abhängigkeiten zwischen Transitionsfolgen. Es kommt also zu einer Wegnahme von Nebenläufigkeit.

Dies ist an folgendem Beispiel erläutert:



Hier wird durch die Einführung eines Synchronisationsbereiches eine Abhängigkeit zwischen den Transitionenfolgen [a,b] und [c,d] eingeführt. In der Praxis könnte das beispielsweise folgendes bedeuten: Man will verhindern, daß zwei Prozesse gleichzeitig auf einen Drucker zugreifen. Daher wird ein Synchronisationsbereich eingeführt, der dafür sorgt, daß der Drucker (repräsentiert durch die Stelle s_3) nur von jeweils einem Prozess genutzt werden kann.

Ein Maß für den Grad der Synchronisation liefert der **"Synchronie-Abstand"**.

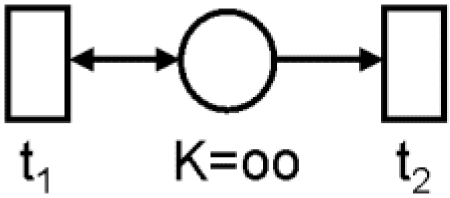
Def.: **"Synchronie-Abstand"**

Seien $T_1, T_2 \subset T$ Transitionenmengen eines S/T-Systems mit Anfangsmarkierung M_0.

a. Der "verankerte Synchronie-Abstand" von T_1, T_2 ist wie folgt definiert:
   $\sigma_a(T_1, T_2) := \sup\{|\#(T_1, w) - \#(T_2, w)| \mid M_0[w>\}$
   wobei #(T, w) die Häufigkeit des Auftretens von Elementen aus T in w bezeichnet.
   Beispiel: #({r,o,t}, 'risotto')=5
   Anschaulich gesprochen bezeichnet σ_a das maximale anfängliche Vorauseilen einer Transitionenmenge T_1 vor einer Transitionenmenge T_2.
   Für unser Synchronisations-Beispiel ergibt sich für den verankerten Synchronie-Abstand $\sigma_a(\{a, b\}, \{c, d\})$=oo.
b. Der "freie Synchronie-Abstand" von T_1, T_2 ist wie folgt definiert:
   $\sigma_f(T_1, T_2) := \sup\{|\#(T_1, w) - \#(T_2, w)| \mid M \in [M_0>, M[w>\}$
   Im Unterschied zum verankerten Synchronieabstand geht man hier von einer frei gewählten Markierung aus, die innerhalb der Erreichbarkeitsmenge liegt.

Bemerkungen/Eigenschaften:

a. σ_a (T, T) = σ_f (T,T) = 0
b. Seien $t_1 \in T_1$ und $t_2 \in T_2$. Dann gilt für



σ_a ({t_1}, {t_2}) = σ_f ({t_1}, {t_2}) = oo
c. $\sigma_a (T_1, T_2) \leq \sigma_f (T_1, T_2)$
d. Unter bestimmten Voraussetzungen (starke Lebendigkeit des S/T-Systems) gilt eine Metrik-Eigenschaft:
i. $\sigma(T_1, T_2)=0 \Leftrightarrow T_1 = T_2$
ii. "Symmetrie": $\sigma(T_1, T_2) = \sigma(T_2, T_1)$
iii. "Dreiecksungleichung": $\sigma(T_1, T_3) \leq \sigma(T_1, T_2) + \sigma(T_2, T_3)$

Eine Variante stellt der sog. **"Fairness-Abstand"** dar. Dieser bezeichnet das maximale anfängliche bzw. gelegentliche Vorkommen der einen Transitionenmenge ohne die andere. Seien $T_1, T_2 \subset T$ Transitionenmengen eines S/T-Systems mit Anfangsmarkierung M_0. Der Fairness-Abstand wird dann wie folgt definiert:

a. Der "verankerte Fairness-Abstand" von T_1, T_2 ist wie folgt definiert:
$\varphi_a (T_1, T_2)$:=max (sup{#(T_1, w) / M_0[w> $\wedge$ #(T_2, w)=0}, sup{#(T_2, w) / M_0[w> $\wedge$ #(T_1, w)=0})
b. Der "freie Fairness-Abstand" von T_1, T_2 ist wie folgt definiert:
$\varphi_f (T_1, T_2)$:=max (sup{#(T_1, w) / M$\in$ [M_0>, M[w> $\wedge$ #(T_2, w)=0}, sup{#(T_2, w) / M$\in$ [M_0>, M[w> $\wedge$ #(T_1, w)=0}

Für unser Synchronisations-Beispiel ergibt sich für den verankerter Fairness-Abstand φ_a({a, b}, {c, d})=oo.

3.4 Weitere Netztypen:

Bislang haben wir uns mit dem Netztyp "S/T-Systeme" beschäftigt. Es gibt aber noch viele weitere Netztypen bei Petri-Netzen. Zwei wichtige werden im folgenden kurz vorgestellt.

Netztyp **B/E-Systeme**:

Definition: Ein Bedingungs-Ereignis-System, kurz B/E-System ist ein S/T-System mit schlingenfreiem und schlichtem Netz. Außerdem müssen alle Stellenkapazitäten und alle Kantengewichte gleich 1 sein.

Anmerkungen:

1. Sei N=(S, T, F) das Netz eines B/E-Systems. Dann bezeichnen wir S als die Menge der Bedingungen (B) und T als die Menge der Ereignisse (E). Markierungen können als Teilmengen von S dargestellt werden: M$\in$ $\mathcal{P}$(s), wobei $\mathcal{P}$

die Potenzmenge bezeichnet.

2. Schalten einer Transition $t \in T$: $M[t>M'$.
   Damit t überhaupt schalten kann, müssen zunächst die folgenden Voraussetzungen im Hinblick auf $M[t>$ erfüllt sein:

   I. $\bullet t \subset M$

   II. $t\bullet \subset \overline{M} = S \setminus M$

   Damit gilt für die Folgemarkierung M': $M'=(M \setminus \bullet t) \cup t\bullet$

3. Zwei Transitionen $t_1, t_2 \in T$, $t_1 \neq t_2$ sind unter M nebenläufig aktiviert g.d.w.: $M[t_1>$, $M[t_2>$ sowie
   i. $\bullet t_1 \cap \bullet t_2 = \emptyset$
   ii. $t_1\bullet \cap t_2\bullet = \emptyset$

Definition **"Konflikt / Kontakt"**

Unter einem Kontakt versteht man die wechselseitige ausgabeseitige Behinderung von Transitionen. Ein Konflikt hingegen beschreibt die wechselseitige eingabeseitige Behinderung von Transitionen.

$M[t_1>$	$M[t_2>$	$\neg M[\{t_1, t_2\}>$	
$\bullet t_1 \in M$	$\bullet t_2 \in M$	$\bullet t_1 \cap \bullet t_2 \neq \emptyset \vee$	Konflikt
$t_1\bullet \in \overline{M}$	$t_2\bullet \in \overline{M}$	$t_1\bullet \cap t_2\bullet \neq \emptyset$	Kontakt

Definition **"Kontaktfreiheit"**

Ein B/E-System heißt kontaktfrei falls $\forall t \in T$ und $\forall M \in [M_0>$ gilt:

$$\bullet t \subset M \Rightarrow t\bullet \subset \overline{M}$$
$$\wedge\, t\bullet \subset M \Rightarrow \bullet t \subset \overline{M}$$

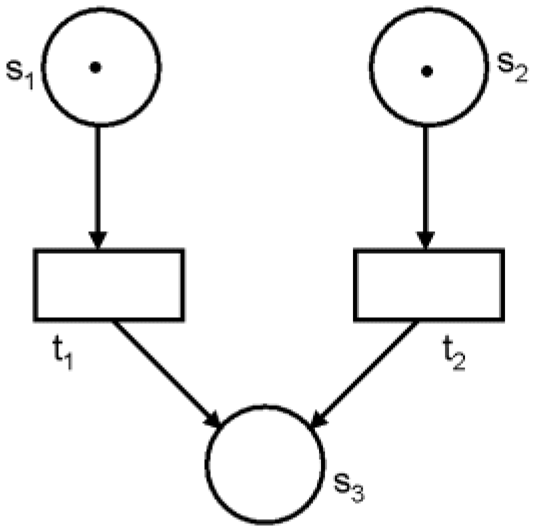
Mit anderen Worten falls t eingabeseitig aktiviert ist, so ist t auch ausgabeseitig frei. Ist t hingegen ausgabeseitig blockiert, so ist diese Transition auch nicht eingabeseitig aktiviert.

Jedes S/T-System lässt sich kontaktfrei machen, indem man "Komplementärstellen" bildet.

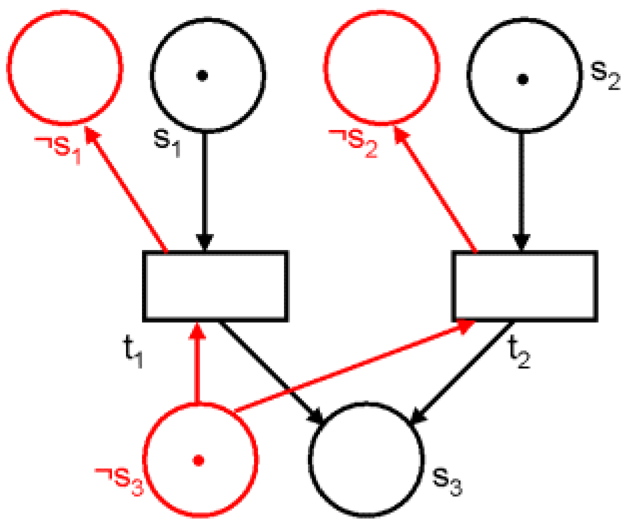
Definition **"Komplementärstelle"**

In einem B/E-System nennt man eine Stelle $s_k \in S$ Komplementärstelle der Stelle $s \in S$ g.d.w. $\bullet s_k = s\bullet$ sowie $\bullet s = s_k\bullet$ und wenn von jedem Paar (s, s_k) bzw. (s_k, s) entweder s oder s_k eine Marke trägt.

Dies ist am Beispiel des folgenden B/E-Systems gezeigt:



Bei diesem B/E-System besteht ein Kontakt zwischen den Transitionen t_1 und t_2. Bildet man nun zu allen Stellen die jeweiligen Komplementärstellen, so ergibt sich das folgende kontaktfreie B/E-System:



Allgemein gilt: Jedes B/E-System läßt sich kontaktfrei machen indem man zu allen Stellen die jeweiligen Komplementärstellen bildet.

Netze mit individuellen Marken

Netze mit individuellen Marken sind Erweiterungen von S/T-Systemen. Sie werden auch als Prädikat-Transitions-Netze bzw. als "coloured petri nets" bezeichnet. In diesen Systemen werden den Marken Werte zugeordnet, wodurch man sie unterscheiden und mit ihnen rechnen kann.

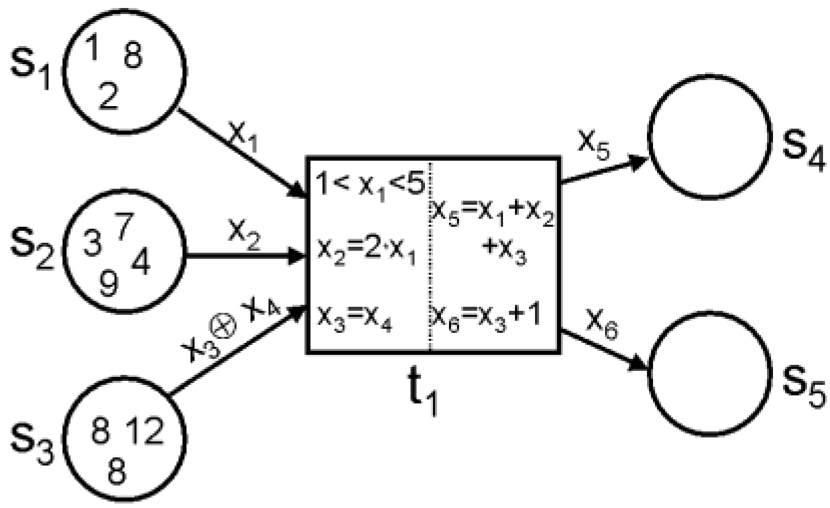
Eine informelle Definition für Prädikat-Transitions-Netze lautet wie folgt:

- Grundlage ist wie bisher ein Netz N=(S, T, F).
- Eine Markierung M(s) stellt bei Prädikat-Transitions-Netzen eine Multimenge von Worten aus einem Wertebereich D dar. Jeder Stelle kann in diesem Zusammenhang ein bestimmter Datentyp zugeordnet sein, z.B. die Menge der reellen oder natürlichen

Zahlen.

- Das Schalten einer Transition kann bei Netzen mit individuellen Marken von mehr als nur dem Vorliegen gewisser Anzahlen von Marken auf den Inputstellen abhängig gemacht werden. Zusätzliche Bedingungen (Prädikate) können qualitative Forderungen bzgl. Der durch eine Transition verbrauchten bzw. erzeugten Marken ausdrücken. Die Schaltregel läßt sich bei Prädikat-Transitions-Netzen nun wie folgt definieren:
  i. Es kommt durch das Schalten einer Transition zu einem Abzug jeweils eines Wertes $x_i \in D(s_i)$ aus jeder Eingangsstelle s_i
  ii. Der Abzug der Marken aus einer Eingangsstelle und die Ablage der Marke in eine Ausgangsstelle ist an Bedingungen (Prädikate) für die Werte x_1, x_2, ... geknüpft. Damit eine Transition schalten kann, müssen alle ihr zugeordneten Prädikate wahr sein.
  iii. Es kommt durch das Schalten einer Transition zu einer Ablage spezifizierter Ausgangswerte in den Nachbereichsstellen.

Beispiel:



Nach Schalten von t_1 ergibt sich als Folgemarkierung: