

8.4 Das Bankiersproblem: Sichere Zustände

Die Problematik von Verklemmungen bei der Betriebsmittelvergabe wurde von Dijkstra als Problem des Bankiers dargestellt [Dij68]. Hier wurde auch der Begriff des sicheren Zustands eingeführt.

Ein Bankier besitzt ein Kapital g . Seine n Kunden erhalten wechselnden Kredit. Jeder Kunde muss seinen maximalen Kreditwunsch von vornherein bekanntgeben und wird nur als Kunde akzeptiert, wenn dieser das Kapital nicht übersteigt. Kredite werden nicht entzogen. Dafür muss aber jeder Kunde versprechen, den Maximalkredit auf einmal nach endlicher Zeit zurückzuzahlen. Der Bankier verspricht, jede Bitte um Kredit in endlicher Zeit zu erfüllen. Für den Bankier besteht natürlich das Problem der Verklemmung: es kann sein, dass mehrere Kunden noch nicht ihren Maximalkredit erhalten haben, das Restkapital des Bankiers aber zu klein ist, mindestens einen Kunden total zu befriedigen, um dann nach einer Frist wieder neues Kapital zu haben.

Eine Instanz $\iota = (n, f, g)$ des Bankiersproblems besteht aus einer Zahl $n \in \mathbb{N}$, einem n -Tupel $f = (f_1, \dots, f_n)$ und einer Zahl $g \in \mathbb{N}$. Ein Zustand einer solchen Instanz ist ein n -tupel $r = (r_1, \dots, r_n)$ das die Restforderung der Kunden beschreibt. Anfangs gilt $r = f$. Ein Zustand heißt *sicher*, wenn er nicht notwendig zu einer Verklemmung führt.

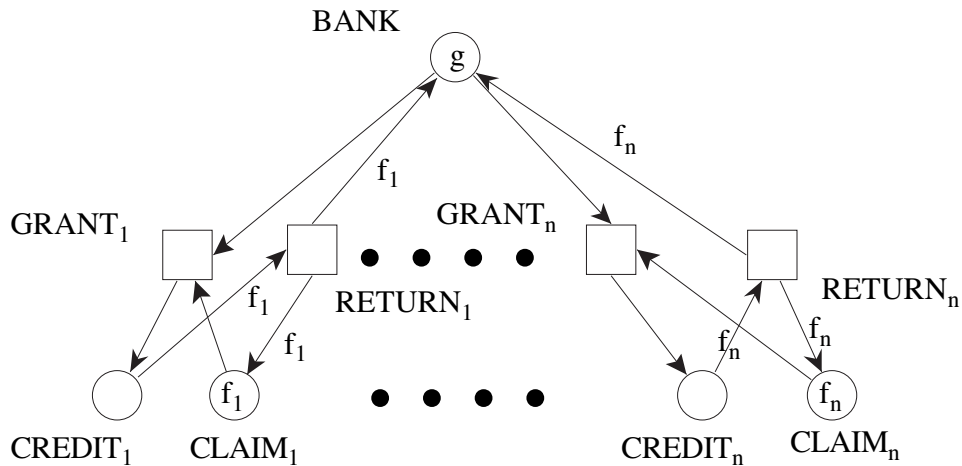


Abbildung 8.24: Das Bankiersproblem mit n Kunden

Das P/T-Netz in Abb. 8.24 modelliert das beschriebene Bankiersproblem. Der Platz *BANK* enthält so viele Marken wie das Kapital des Bankiers in Geldeinheiten umfasst. $CREDIT_i$ und $CLAIM_i$ stehen für Kredit und Restforderung. Durch die Transition $GRANT_i$ erhält der Kunde i so viele Geldeinheiten, wie sie schaltet. $RETURN_i$ transferiert das Geld zurück. Diese Transition kann nur schalten, wenn der Bankier die Maximalforderung f_i des Kunden erfüllt hat. Gleichzeitig wird die Ausgangsforderung wieder hergestellt.

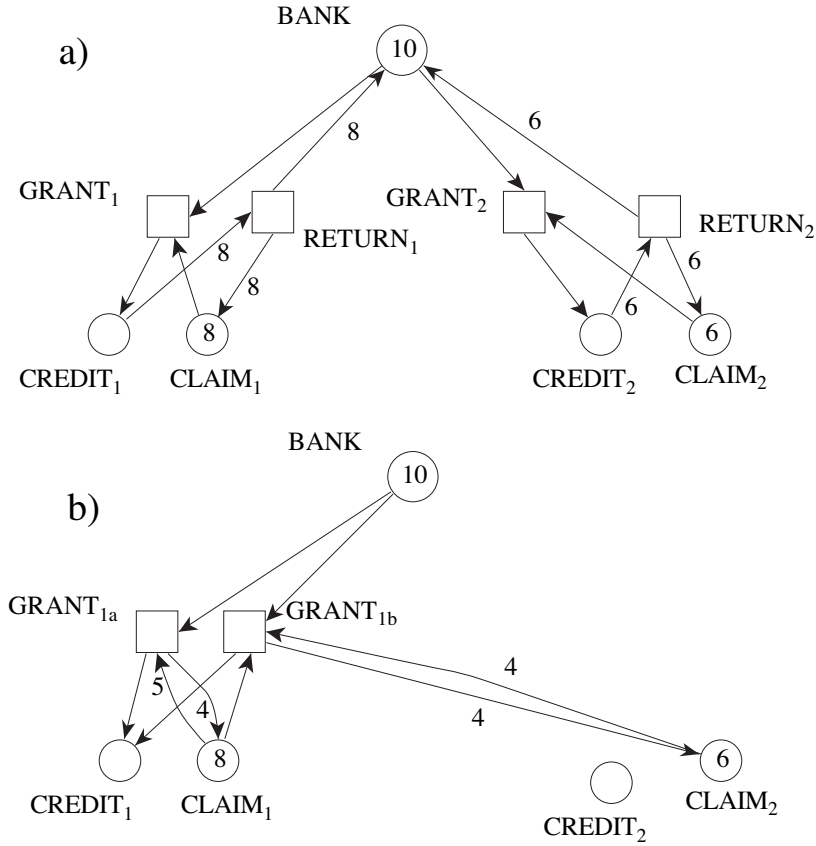


Abbildung 8.25: Eine Bankiersprobleminstanz mit 2 Kunden und Modifikation zur Vermeidung von Verklemmungen

Betrachten wir zwei spezielle Instanzen, nämlich $\iota_1 = (2, (8, 6), 10)$ (Abb. 8.25) und $\iota_2 = (3, (8, 3, 9), 10)$ (Abb. 8.27). An ihnen werden Erreichbarkeitsgraph und P-Invarianten erläutert.

Für die Instanz $\iota_1 = (2, (8, 6), 10)$ in Abb. 8.25a wird jeder Zustand durch eine Markierung dargestellt, d.h. durch einen 5-dimensionalen Vektor. Für alle erreichbaren Markierungen \mathbf{m} gelten die folgenden drei Gleichungen:

- $\mathbf{m}[BANK] + \mathbf{m}[CREDIT_1] + \mathbf{m}[CREDIT_2] = 10$
(Das Geld ist bei der Bank oder bei den Kunden und zwar genau 10 Einheiten insgesamt.)
- $\mathbf{m}[CLAIM_1] + \mathbf{m}[CREDIT_1] = 8$
(Kredit und Restforderung des Kunden 1 beträgt zusammen immer genau 8 Einheiten.)
- $\mathbf{m}[CLAIM_2] + \mathbf{m}[CREDIT_2] = 6$
(Kredit und Restforderung des Kunden 2 beträgt zusammen immer genau 6 Einheiten.)

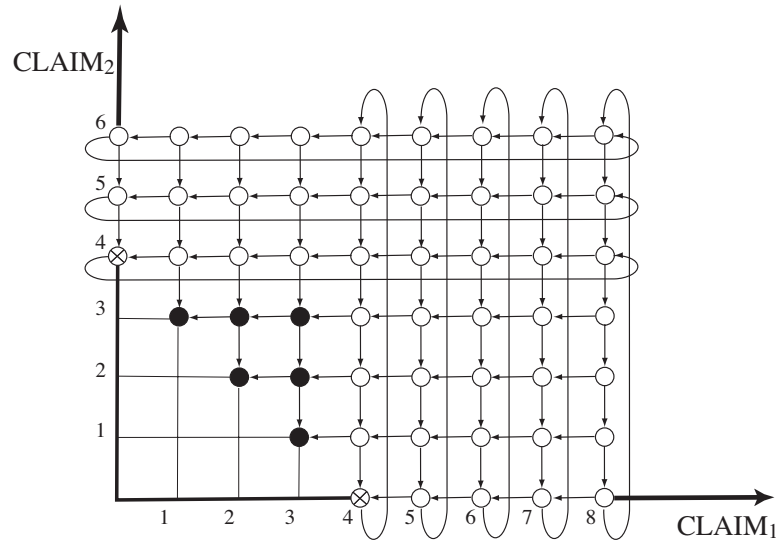


Abbildung 8.26: Erreichbarkeitsgraph des Netzes 8.25

(Sie heißen P-Invarianten-Gleichungen und werden im nächsten Abschnitt 7.5 systematisch behandelt.) Also ist jede erreichbare Markierung schon durch 2 ihrer Komponenten festgelegt, nämlich $(CLAIM_1, CLAIM_2)$. Die anderen 3 Komponenten, $BANK$, $CREDIT_1$ und $CREDIT_2$ können mit den Gleichungen daraus berechnet werden. Dadurch kann der Erreichbarkeitsgraph in einem ebenen Gitter dargestellt werden (Abb. 8.26).

Die Anfangsmarkierung $\mathbf{m}_0 = (10, 0, 8, 0, 6)$ (wobei die Plätze folgendermaßen geordnet seien: $(BANK, CREDIT_1, CLAIM_1, CREDIT_2, CLAIM_2)$) wird auf das Paar $(\mathbf{m}_0(CLAIM_1), \mathbf{m}_0(CLAIM_2)) = (8, 6)$ reduziert. Es entspricht dem Knoten rechts oben im Graphen von Abb. 8.26.

Alle Pfade, die in diesem Knoten anfangen, entsprechen Schaltfolgen. Kanten nach links, rechts, unten und oben entsprechen jeweils dem Schalten der Transition $GRANT_1$, $RETURN_1$, $GRANT_2$ und $RETURN_2$. Werden die Kunden strikt nacheinander bedient, so gibt es keine Probleme. Falls sich jedoch die Ausleihschritte überlappen, kann einer der drei Verklemmungen $(1,3)$, $(2,2)$ und $(3,1)$ kommen.

Dijkstra hat darauf hingewiesen, dass es noch andere kritische Zustände gibt, nämlich diejenigen, die unvermeidlich zu einer Verklemmung führen, wie z.B. $(3,3)$. Er nannte sie *unsicher*. Sie sind als schwarze Knoten dargestellt. In [HV87] und [VJ85] wurde gezeigt, dass *sichere Zustände* (weiße Knoten) durch ihre *minimalen Elemente* darstellbar sind: $(0,4)$ und $(4,0)$, (mit Kreuz).

Wie kann man unsichere Zustände vermeiden? Der Algorithmus von Dijkstra berechnet vor jedem Ausleihschritt, ob von dem dann erreichten Nachfolgezustand der Anfangszustand noch erreichbar ist.

Wie aus Abb. 8.26 ersichtlich kann dies auch dadurch geschehen, dass Transition $GRANT_1$ nur in Markierungen aktivierbar ist, die (in mindestens einer Komponente) größer als $(4, 0)$ oder $(0, 4)$ sind. Dies wird erreicht, wenn man die Transition $GRANT_1$ durch zwei modifizierte Kopien $GRANT_{1a}$ und $GRANT_{1b}$ (Abb. 8.25b)⁴ ersetzt. Diese beiden Transitionen haben den gleichen Schalteffekt wie die ursprüngliche, aber einen höheren Aktivierungs-„Schwellwert“. Die entsprechende Konstruktion ist bei $GRANT_2$ anzuwenden. Der allgemeine Algorithmus ist in [VJ85] und teilweise in [JV87] S.216 ff. zu finden.

Die zweite Instanz $\iota_2 = (3, (8, 3, 9), 10)$ ist ein Beispiel aus dem Buch [BH73] über Betriebssysteme. Seine Netzdarstellung befindet sich in Abbildung 8.27. Es enthält sieben

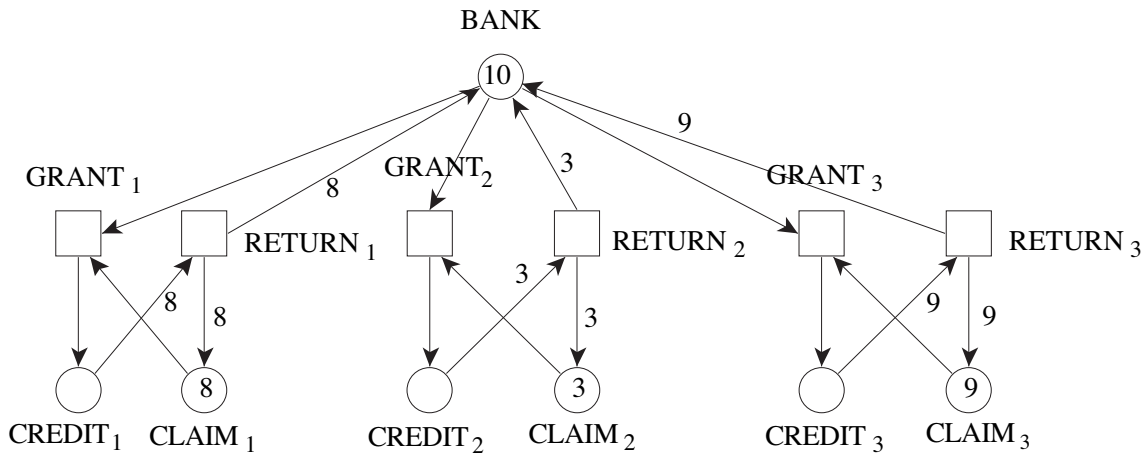


Abbildung 8.27: Eine Instanz des Bankiersproblems mit drei Kunden

Plätze. Wegen der nun vier Gleichungen kann der Erreichbarkeitsgraph in drei Dimensionen dargestellt werden (siehe dazu auch Abb. 8.28). Bezeichnend ist das Anwachsen seiner Größe. Er enthält 195 erreichbare Markierungen. Die Teilmenge von 137 sicheren Markierungen (weiße Knoten) wird von 10 minimalen Elementen (dem **Residuum**: weiße Knoten mit Kreuz)) abgegrenzt. Das Residuum als Charakterisierung der kleinsten Menge der noch gerade sicheren Zustände liefert ein wichtiges Kriterium für die sinnvollerweise erlaubten Handlungen in einem System. Eine allgemeine Methode zu ihrer Berechnung wird in [HV87] gegeben. Verlässt man diesen Bereich, kommt man zwangsweise zu im Allgemeinen unerwünschten (partiellen oder totalen) Verklemmungen: den *unsicheren Markierungen*. Die 58 unsicheren Markierungen sind wieder schwarz dargestellt.

Die zweite und größere Instanz hat aber auch Eigenschaften, die in der ersten nicht zu beobachten waren: sie enthält Markierungen, von denen aus Verklemmungen vermieden werden können, die aber nicht sicher sind, wenn sicher heißt, dass alle Kunden ihre Transaktionen beenden können. Beispielsweise kann von der Markierung $(4, 3, 6)$ ausge-

⁴Diese Abbildung zeigt nur, wie $GRANT_1$ durch zwei Transitionen zu ersetzen ist. Entsprechendes muss auch auf $GRANT_2$ angewandt werden, nicht jedoch auf $RETURN_1$ und $RETURN_2$.

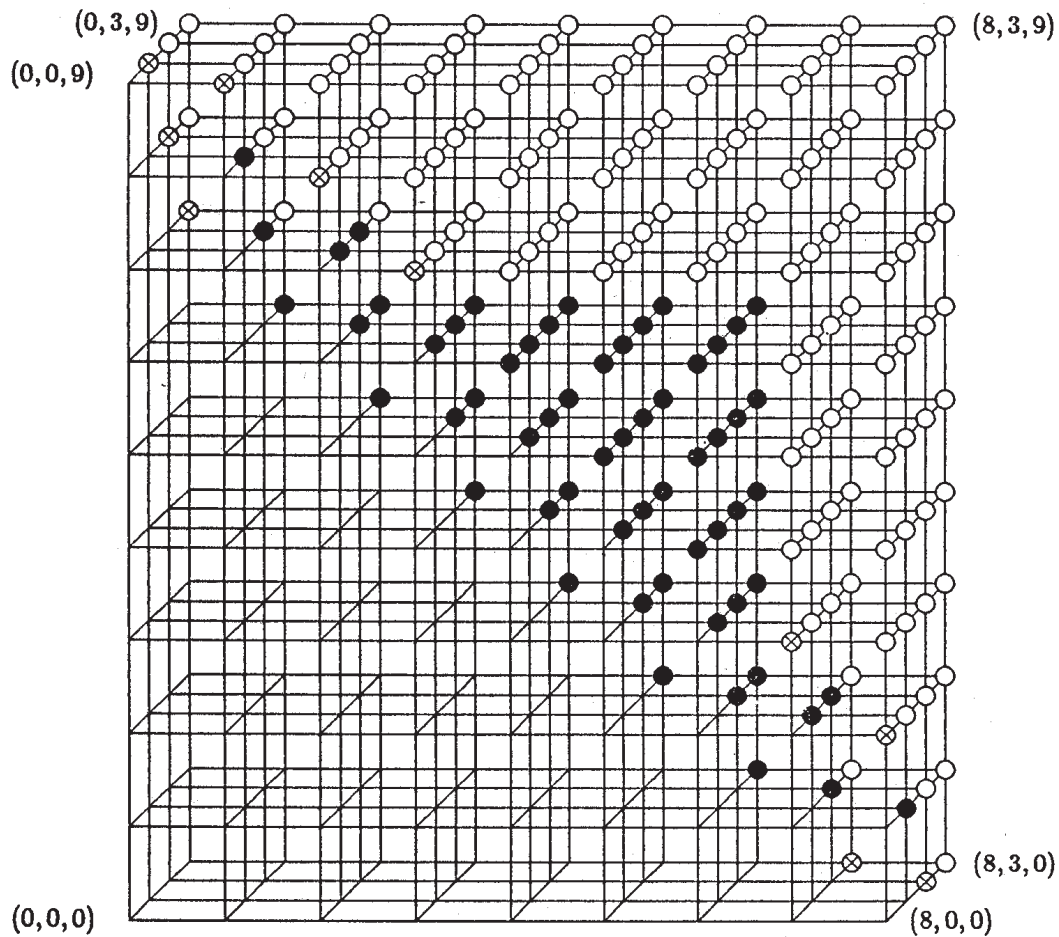


Abbildung 8.28: Erreichbarkeitsgraph des Netzes 8.27

hend, der zweite Kunde beliebig viele Transaktionen ausführen, während die Kunden 1 und 3 nicht einmal einen Ausleihvorgang vollständig zu Ende bringen können. Solche Situationen heißen *partielle Verklemmung*. Ein Netz ohne partielle Verklemmungen heißt *lebendig*. Lebendigkeit wird im Abschnitt 7.1.3 behandelt.

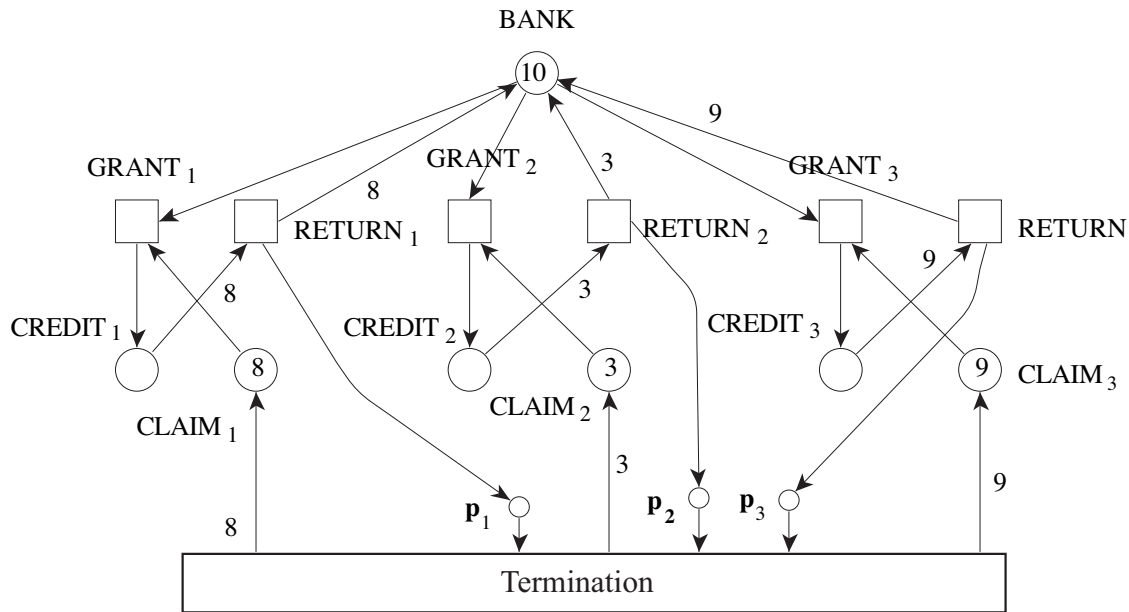


Abbildung 8.29: Bankiersnetz mit Terminationstransition

Um die ursprüngliche Definition von Dijkstra beizubehalten, kann man wie in Abb. 8.29 eine Transition *Termination* einführen, die dafür sorgt, dass alle Kunden einen Ausleihvorgang beendet haben, bevor eine neue Runde started. Die Einschränkungen der möglichen Abläufe ergibt sich aus der expliziten Synchronisation. Um den Begriff der sicheren Markierung sinnvoll auch in dem Netz 8.27 zu benutzen, könnte man folgende Definitionen unter c) benutzen, die eine sinnvolle Übertragungen des Begriffs der Sicherheit auf Netze liefern:

- Eine Markierung \mathbf{m} heißt sicher, wenn eine spezifizierte Endmarkierung (hier die Anfangsmarkierung) wieder erreichbar ist.
- Eine Markierung \mathbf{m} heißt sicher, wenn eine bestimmte Transition (z.B. *Termination*) zum Schalten gebracht werden kann.
- Eine Markierung \mathbf{m} heißt sicher, wenn von ihr aus eine unendliche Schaltfolge möglich ist, die alle Transitionen des Netzes unendlich oft enthält.

Ist in a) die Anfangsmarkierung gemeint, dann wird diese Eigenschaft in der Petrinetzliteratur auch „home property“ genannt. Die Eigenschaft c) hat mit dem „fairen Verhalten“ eines Netzes zu tun. Fairness und Lebendigkeit stehen in komplexen Beziehungen zueinander.

8.5 Das Renew-Werkzeug

Das am Arbeitsbereich TGI in Hamburg entwickelte Werkzeug RENEW [KWD] ermöglicht die graphische Darstellung der hier behandelten Netzmodelle. Unter <http://www.Renew.de> findet sich neben den Installationsdateien und -hinweisen sowie dem vollständigen Quell-Code der Software auch ein Handbuch, das die folgenden Beschreibungen ausführlich ergänzt. Im Folgenden werden anhand einiger Bildschirmdarstellungen zentrale Konzepte und die Nutzung von RENEW erläutert.

Die RENEW-Werkzeugleiste ist in der Abb. 8.30 dargestellt.



Abbildung 8.30: Werkzeugleiste des RENEW-Werkzeugs

RENEW kann auch für die Ausführung der erzeugten Netze benutzt werden. Als Beschriftungssprache wurden Elemente der Programmiersprache Java [GJS97] gewählt, die in vielen ihrer Eigenschaften Petrinetze sinnvoll ergänzt. Das Werkzeug ist selbst in Java geschrieben und durch Transitionen können Java-Klassen ausgeführt werden. *Ausführung* wird in der Petrinetzliteratur auch als *Simulation* bezeichnet. Dies bedeutet die Simulation des formalen Modells und nicht eines realen Weltausschnittes. Letzteres gilt natürlich auch, wenn das Petrinetz den realen Weltausschnitt hinreichend genau darstellt. Dazu können auch Zeitschranken für das Schalten benutzt werden. Eine Übersicht zu Petrinetz-Werkzeugen gibt die Internetseite *The Petri Nets World* [Pet]. Über sie sind auch Informationen zu Literatur zu Petrinetzen, Forschungsgruppen, -aktivitäten, -projekte und Anwendungen von Petrinetzen zugänglich.

Die Abbildung 8.31 zeigt ein P/T-Netz in RENEW (nach [Kum01]). Seine drei Marken

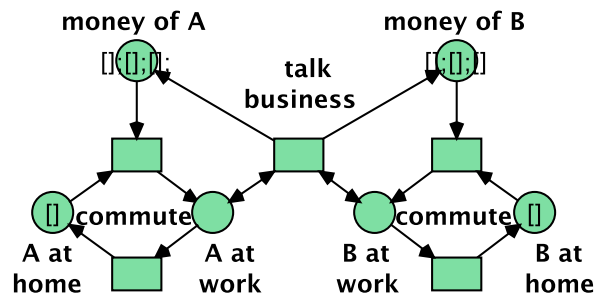


Abbildung 8.31: Das Leben zweier Geschäftsleute

im Platz *money of A* werden als $[]; []; []$ dargestellt. Dieses Netz kann folgendermaßen interpretiert werden: zwei Geschäftsleute A und B können jeweils von zu Hause (*at home*) zum Arbeitsplatz (*at work*) wechseln. Die Fahrt kostet Geld. Dieses (und mehr) kann jedoch beim gemeinsamen Handel (*talk business*) wieder verdient werden. Abstrakt gesehen handelt sich hier also um zwei Betriebsmittel verbrauchende und erzeugende Funktionseinheiten.

Durch Faltung erhält man das gefärbte Netz in Abbildung 8.32. Geschäftsleute werden

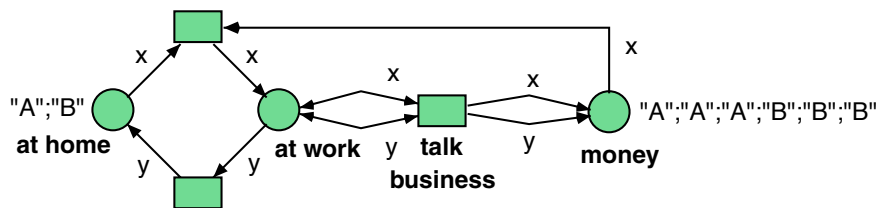


Abbildung 8.32: Faltung von Netz 8.31 zu einem gefärbten Netz

durch *individuelle Marken* "A" und "B" dargestellt, ebenso wie ihr jeweiliges Guthaben im Platz *money*.

Wird eine Darstellung mit Bezeichner und Wert gewünscht, so kann man wie in Abb. 8.33 vorgehen. Die Kantenbeschriftungen sind hier 2-Tupel (in JAVA-Notation), deren erste

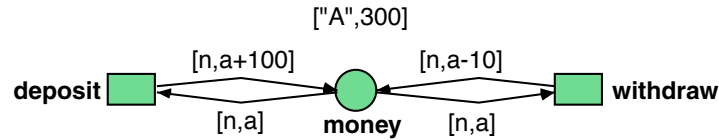


Abbildung 8.33: Ein nicht ganz so einfaches Bankkonto

Komponente den Kontoinhaber und deren zweite Komponente den Wert oder den zu verändernden Wert darstellen.

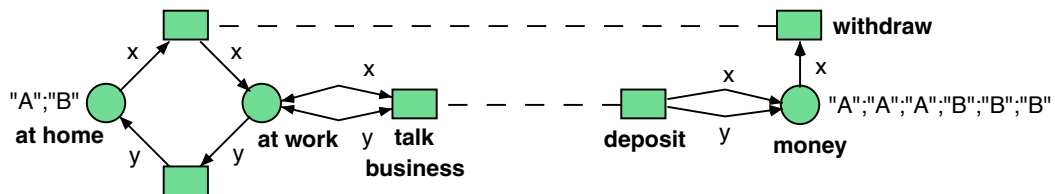


Abbildung 8.34: Personen und Konten werden geteilt

Ein ernst zu nehmender Einwand gegen das gefärbte Netz aus Abb. 8.32 ist, dass die Bewegung von Geld und die Bewegung der Personen zu stark miteinander verwoben sind. Auf den ersten Blick ist es nicht klar, welche Teile des Netzes welchen Aspekt beschreiben.

In Abb. 8.34 bereiten wir eine Teilung des Systems aus Abb. 8.32 vor. Zwei der Transitionen werden doppelt in verschiedenen Teilen des Netzdiagramms aufgezeichnet. Um die Schaltsemantik des Netzes korrekt beizubehalten, müssten die Transitionen wieder verschmolzen werden.

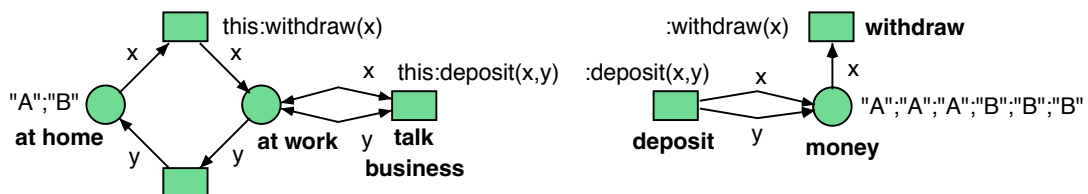


Abbildung 8.35: Textuelle Anschriften kennzeichnen Synchronisation

Wir gehen einen Schritt weiter und deuten die beabsichtigte Bedeutung des Netzes durch eine textuelle Anschrift an. In Abb. 8.35 bedeutet die Anschrift `this:withdraw(x)`, dass in diesem Netz (Englisch *this net*) eine andere Transition vorhanden sein sollte, die die

Auszahlung von Geld vom Konto von x übernehmen kann. Wir geben dabei die Variable am Ende der Anschrift an, um klarzumachen, welche Information umhergereicht werden muss.

Solche Anschriften werden als synchrone Kanäle bezeichnet, weil sie die Transitionen zwingen, synchron zu schalten und weil sie den Fluss von Informationen kanalisieren. Die Autoren von [CDH92] haben dieses Konzept für höhere Petrinetze eingeführt. Gegenüber gewöhnlichen Petrinetzen fügen synchrone Kanäle die Fähigkeit hinzu, über gleichzeitige, nicht nur über aufeinanderfolgende oder über nebenläufige Handlungen zu sprechen (Rendez-Vous-Synchronisation).

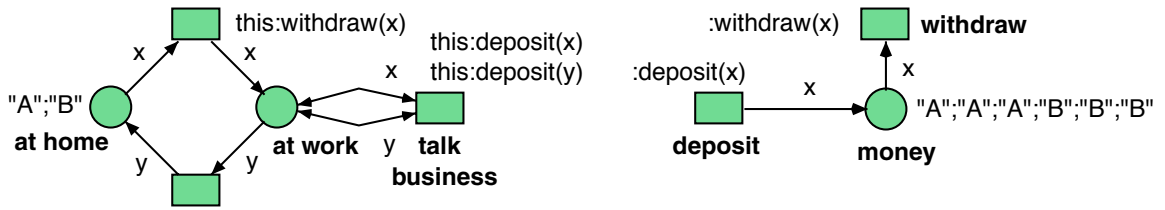


Abbildung 8.36: Wiederverwendung eines Kanals

Weil wir textuelle Anschriften für die Kanäle verwendet haben, können wir mehrere Synchronisationen gleichzeitig anfordern, ohne dass das Diagramm seine Klarheit verliert. In Abb. 8.36 wurde das Netz aus Abb. 8.35 so umstrukturiert, dass von der Transition **talk business** zwei Aufrufe des Kanals **deposit** getätigt werden. Dies veranlasst die Transition **deposit**, zweimal zu schalten, was das gewünschte Verhalten in unserem Beispiel ist.

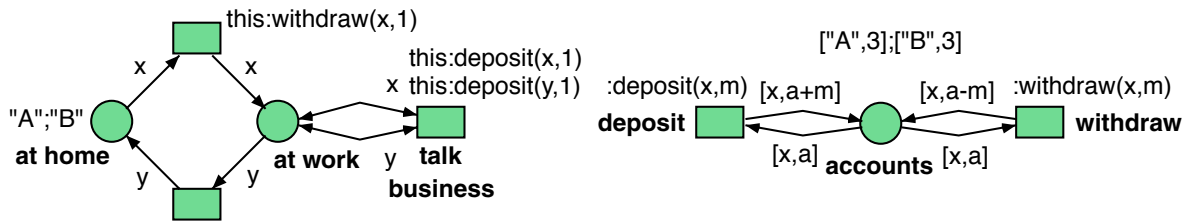


Abbildung 8.37: Buchführung mit Algebra

Jetzt können wir die Lösung aus Abb. 8.36 mit dem gefärbten Netz zur Modellierung eines Bankkontos aus Abb. 8.33 kombinieren und jedes Konto als Wertepaar repräsentieren. Abb. 8.37 zeigt das Resultat. Beachtenswert ist, wie die Anzahl der Geldeinheiten, die ein- oder auszuzahlen ist, als zweiter Parameter über den Kanal übergeben wird.

Üblicherweise werden die beiden Netze in zwei verschiedenen Fenstern wie in Abbildung 8.38 dargestellt.

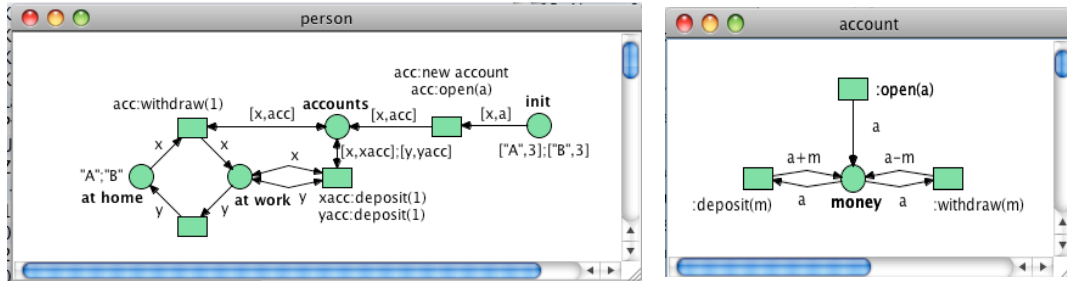


Abbildung 8.38: RENEW: Buchführung mit Algebra in zwei Fenstern

Die Abbildung 8.39 zeigt das Netz **person** nach einem Simulationsschritt.

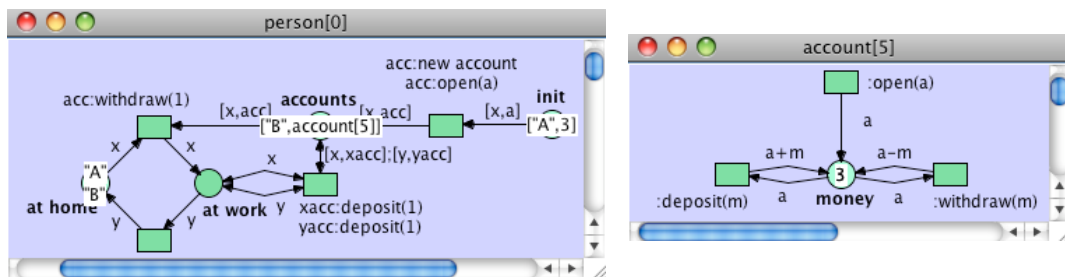


Abbildung 8.39: RENEW (Simulationsmodus): Buchführung mit Algebra in zwei Fenstern

Beim Schalten der Transition rechts oben wurde mit der Bindung $[x = "B", a = 3]$ ein Exemplar **account[5]** des Musters **account** erzeugt und an die Variable **acc** gebunden. Dadurch wird die Bindung zu $[x = "B", a = 3, acc = account[5]]$ erweitert. Im Platz **accounts** verfügt nun der Geschäftsmann "B" über eine Referenz auf sein Konto **account[5]** mit Inhalt 3. Beim nochmaligen Schalten dieser Transition würde entsprechendes für den Geschäftsmann "A" gelten, natürlich mit der Referenz auf ein eigenes Exemplar von **account**. Diese Fähigkeit von RENEW Exemplare von Mustern zu bilden und die Referenzen wie Marken zu behandeln ist eine über übliche gefärbte Netze hinausführende Eigenschaft, die aber nicht Gegenstand dieser Vorlesung ist.

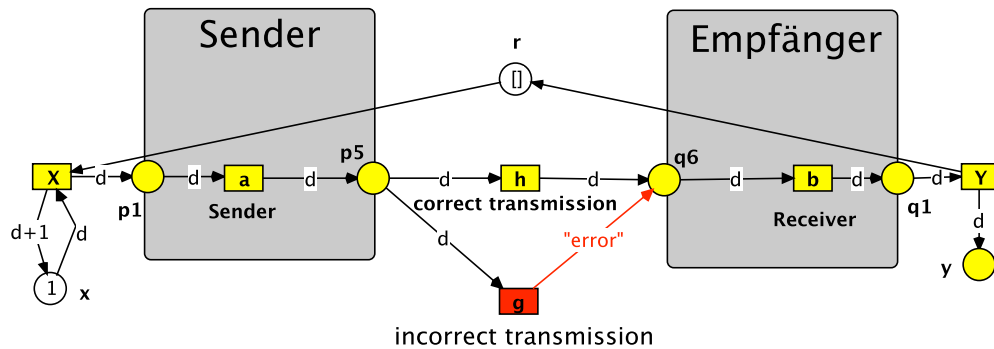


Abbildung 8.41: Übermittlung mit Fehlern: Netz abp-2

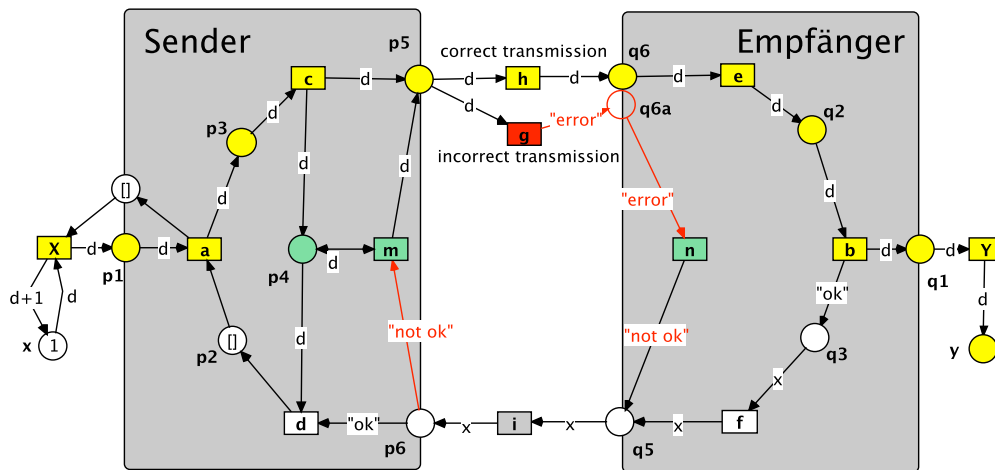


Abbildung 8.42: Übermittlung mit Quittung: Netz abp-3

nicht betrachtet wird. Das Netz abp-1 kann also als Spezifikation der Aufgabenstellung angesehen werden.

Im Netz abp-2 (Abb. 8.41) kann der Kanal Daten fehlerhaft übermitteln. Dies wird durch das alternative Schalten der Transition g modelliert, die die Fehlermeldung "error" erzeugt. Dies muss in Anwendungen durch unterliegende Protokollschichten implementiert werden, bzw. durch ein *timeout* im Fall von Datenverlust. Das Netz abp-2 hat ein von der Spezifikation abp-1 verschiedenes Verhalten.

Ein der Spezifikation entsprechendes Verhalten wird durch das Netz abp-3 durch die Rücksendung einer positiven ("ok") oder negativen ("not ok") Quittung erreicht. Bei einer negativen Quittung wird das im Platz $p4$ gespeicherte Datum solange wieder versandt, bis eine positive Quittung erfolgt. Das Netz erfüllt jedoch nicht die Spezifikation, wenn auch die Rücksendung über die Transition i fehlerhaft ist.

Zur Vorbereitung der endgültigen Lösung wird mit abp-4 ein zu abp-3 verhaltensäquivalentes Protokoll eingeführt. Dem Datum d wird durch die Transition a ein

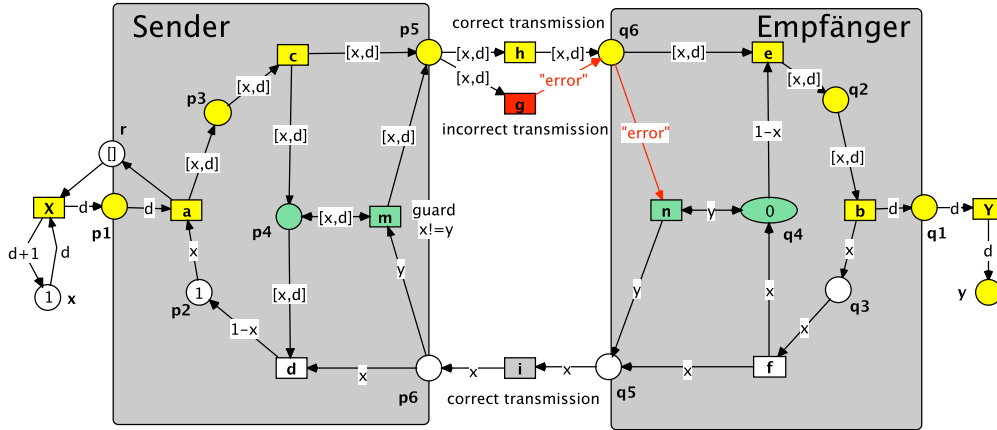


Abbildung 8.43: Übermittlung mit Quittung durch Alternierbit: Netz abp-4

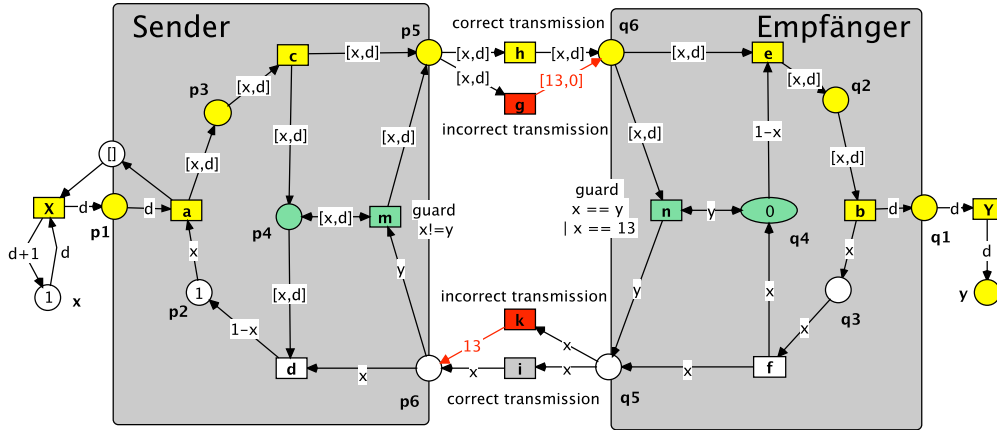


Abbildung 8.44: Das Alternierbitprotokoll abp-5

Bit x beigefügt, das anfangs den Wert 1 hat. Bei fehlerfreier Übermittlung schaltet die Transition e , da $q4$ den komplementären Wert $1 - x$ (anfangs 0) enthält. Danach wird durch die Transition b das Bit wieder gelöscht und das Datum d wie vorher abgeliefert. Die Quittung erfolgt durch das (nicht geänderte) Bit x . Im Fehlerfall wird jedoch durch die Transition n mit dem komplementären Bit (anfangs 0) quittiert, wodurch die erneute Sendung eingeleitet wird. Dies erkennt der Sender dadurch, dass das Quittungs-Bit y von dem gespeicherten Bit x verschieden ist (**guard** $x \neq y$). Um die nächste Datenübermittlung einzuleiten, wird das Ausgangsbit komplementiert nach $p2$ gelegt. Der Vorgang wiederholt sich mit dem jeweils komplementären Bitwert.

Nun ist es nur noch ein kleiner Schritt zur endgültigen Lösung. Im Netz abp-5 (Abb. 8.44) ist nun auch im Quittungs-Kanal eine Störung möglich. Um den Integer-Test **guard** $x \neq y$ beibehalten zu können, wurde statt der Fehlermeldung "error" die (Unglücks-)Zahl 13 gewählt, d.h. auch bei einem Fehler in diesem Kanal wird das Datum d erneut

verschickt. Nun tritt aber folgendes Problem auf. Falls das Datum vorher störungsfrei übermittelt wurde, würde in diesem Fall eine verdopplete Ankunft beim Empfänger erfolgen. Dies kann der Empfänger jedoch daran erkennen, dass das Bit nicht gewechselt hat. Die verdopplte Sendung wird durch den Guard $x==y \mid x==13$ (d.h. $x = y \vee x = 13$) in der Transition n gelöscht.

Als Besonderheit kommt dieses Protokoll zur Quittierung mit nur einem Bit x aus ([BS69]) - daher der Name *Alternierbitprotokoll*. Ein Beweis dafür, dass die Lösung korrekt ist, kann z.B. dadurch erfolgen, dass das Netz $abp-5$ als verhaltensäquivalent zu seiner Spezifikation $abp-1$ bewiesen wird. Verhaltensäquivalenz kann z.B. durch Bisimulation formalisiert werden. Dies wird im Rahmen der Prozessalgebra durchgeführt. Eine alternative Verifikationsmethode ist das *Model-Checking*. Hierbei wird die Spezifikation anhand des Erreichbarkeitsgraphen untersucht, der im vorliegenden Fall etwa 70 Zustände hat. Es wurden auf diese Weise bereits Systeme mit 10^{50} Zuständen verifiziert.