

Uniwersytet Warszawski
Wydział Matematyki, Informatyki i Mechaniki

Procesy nieskończone systemów współbieżnych

rozprawa doktorska

Łukasz Mikulski
Wydział Matematyki i Informatyki
Uniwersytet Mikołaja Kopernika w Toruniu

Promotor

dr hab. Edward Ochmański, prof. UMK

Zakład Lingwistyki Matematycznej i Teorii Współbieżności
Wydział Matematyki i Informatyki
Uniwersytet Mikołaja Kopernika w Toruniu

Lipiec 2011

Oświadczenie autora rozprawy:
oświadczam, że niniejsza rozprawa została napisana przeze mnie samodzielnie.

14 lipca 2011
data

.....
Łukasz Mikulski

Oświadczenie promotora rozprawy:
niniejsza rozprawa jest gotowa do oceny przez recenzentów.

14 lipca 2011
data

.....
dr hab. Edward Ochmański, prof. UMK

Procesy nieskończone systemów współbieżnych*

Słowa kluczowe:

Sieci Petriego, ślady Mazurkiewicza, procesy nieskończone, współbieżność, konflikt

AMS Mathematical Subject Classification 2000:

68Q85 Models and methods for concurrent and distributed computing

Streszczenie

W rozprawie, przez pryzmat zjawiska konfliktu, badane są procesy nieskończone systemów współbieżnych. Podstawowymi narzędziami są sieci Petriego oraz ślady Mazurkiewicza.

Ślady procesów nieskończonych opisywane są za pomocą skończonych zbiorów nieskończonych rzutów, które są podklasą klasy zbiorów rzutowych. Wyodrębnienie zbiorów wewnątrznie bezkonfliktowych, będących dokładnie reprezentacjami rzutowymi śladów, następuje poprzez definicję strukturalną. Podejście to wpisuje się, upraszczając jedno z podejść, w dotychczasowe studia nad nieskończonymi śladami Mazurkiewicza.

Analiza procesów trwałych p/t-sieci prowadzi do pojęcia sieci pokojowych, a w dalszej dyskusji do jego uogólnień, czyli sieci k-pokojowych i super-pokojowych. Pojęcia te unifikują i uogólniają zjawiska konfliktu oraz konfuzji. Przedyskutowane zostały podstawowe problemy decyzyjne związane ze zbiorami markingów osiągalnych w tych nowych typach sieci, co doprowadziło do zaklasyfikowania sieci k-pokojowych jako uniwersalnych modeli obliczeń. Przedstawiono też efektywne narzędzia do dalszych badań nad tymi klasami sieci oraz zbadano pewne ich istotne podklasy. Nieco trudniejsza okazała się dyskusja nad sieciami super-pokojowymi. Przedstawione wyniki dotyczące tej klasy sieci nie wyczerpują tematu; pozostawiono kilka problemów otwartych wymagających dalszych badań.

*Badania wspierane przez Ministerstwo Nauki i Szkolnictwa Wyższego, grant N N206 258035

Infinite processes of concurrent systems*

Keywords:

Petri nets, Mazurkiewicz traces, infinite processes, concurrency, conflict

AMS Mathematical Subject Classification 2000:

68Q85 Models and methods for concurrent and distributed computing

Abstract

This thesis deals, through the prism of conflict, with infinite processes of concurrent systems. The basic tools used in the thesis are Petri nets and Mazurkiewicz traces.

We propose to describe the infinite traces of processes by the finite sets of infinite projections. Such sets are treated as a subclass of general projection sets. Conflict-free projection sets, which are exactly the projection representations of the traces, are introduced by the structural definition. We show, how this approach fits, simplifying one of the previous approaches, in the studies on the infinite Mazurkiewicz traces.

Analysis of the persistent processes of p/t-nets leads to the concept of nonviolent nets. Further discussion gives their generalizations - k-nonviolent nets and super-nonviolent nets. We show how these concepts unify and generalize the phenomenon of conflict and confusion. We discuss the decision problems associated with sets of reachable markings in these new types of nets, which leads to the classification of k-nonviolent nets as a universal model of computation. We also give effective tools for further studies on nonviolent nets and study some of their relevant sub-classes. Slightly more difficult was the discussion on super-nonviolent nets. We present some partial results, posing several open problems.

*The research supported by Ministry of Science and Higher Education of Poland, grant N N206 258035

Spis treści

Wstęp	3
1 Modele systemów współbieżnych	7
1.1 Podstawowe notacje	7
1.2 Sieci Petriego	10
1.2.1 Elementarne sieci Petriego	10
1.2.2 Sieci markowane (p/t-sieci)	13
1.2.3 Rozszerzenia p/t-sieci	19
1.2.4 Problemy decyzyjne w sieciach Petriego	24
1.3 Ślady Mazurkiewicza	41
1.3.1 Ślady skończone	41
1.3.2 Ślady nieskończone	44
1.4 Związki sieci elementarnych i śladów	46
2 Zbiory rzutowe i rekonstrukcje	49
2.1 Własności rzutowe śladów	49
2.1.1 Wektory Parikha	49
2.1.2 Reprezentacja rzutowa śladu	50
2.2 Zbiory rzutowe	50
2.3 Wewnętrznie bezkonfliktowe zbiory rzutowe	52
2.4 Rekonstrukcja śladów	54
2.5 Konkatenacja śladów nieskończonych	57
3 Procesy bezkonfliktowe	61
3.1 Konflikty w sieciach Petriego	61
3.2 Sieci pokojowe	66
3.3 Problemy decyzyjne w sieciach pokojowych	69
3.3.1 Nierozstrzygalność problemu osiągalności pustego markingu	72
3.3.2 Równoważność markingowa sieci pokojowych i inhibitorowych	76
3.3.3 Pokojowe sieci wolnego wyboru	78
3.4 Uogólnienia sieci pokojowych	80
3.4.1 Hierarchia sieci pokojowych	84
3.4.2 Lematy o propagacji i genezie konfliktów	86
3.4.3 Sieci ω -pokojowe	89
Podsumowanie	95
Literatura	97

*Serdecznie dziękuje mojej żonie Sylwii, za wsparcie moralne
oraz Edwardowi, za poświęcony czas i wsparcie merytoryczne.*

Wstęp

Systemy reakcyjne, takie jak systemy operacyjne czy programy kontrolujące linie produkcyjne, można modelować jako ciągi następujących po sobie akcji. Jako że systemy takie powinny teoretycznie działać bez przestojów, modele przyjmują, że ciągi te są nieskończone. Z powodów czysto praktycznych, takich jak kończące się możliwości zwiększania efektywności procesorów działających sekwencyjnie, jak i złożoności rozważanych systemów innego typu, często stosuje się modele współbieżne. Kolejność wykonania niektórych akcji jest w nich nieistotna.

Sieci Petriego [40] oraz ślady Mazurkiewicza [34] to dwa popularne modele związane ze współbieżnością. O ile sieci Petriego dają intuicyjny model systemów współbieżnych, to ślady koncentrują się bardziej na opisie procesów takich systemów. Samą nieskończoność procesów można rozumieć w kontekście tych modeli wielorako. Mogą to być pewne własności uniwersalne sieci Petriego, rozszerzenie modelu przez wprowadzenie nieskończonej liczby żetonów do sieci, czy w końcu rozważanie nieskończonych obliczeń, jakie mogą odbywać się w sieciach Petriego i opisywanie ich przy pomocy śladów Mazurkiewicza jako procesów nieskończonych.

Rozprawa składa się z trzech rozdziałów, opatrzonych wprowadzeniem i podsumowaniem. W rozdziale pierwszym przedstawione zostały podstawowe pojęcia i fakty wykorzystywane w pracy. Przy istotnych elementach przeprowadzona została dokładniejsza dyskusja (ze szczególnym uwzględnieniem problemów decyzyjnych związanych z markingami sieci Petriego, grafów pokrywalności sieci Petriego oraz modeli nieskończonych śladów Mazurkiewicza). Zasadniczą część rozprawy stanowią rozdziały drugi i trzeci. W rozdziale drugim ślady nieskończone zanurzone zostały w ogólniejszej klasie zbiorów rzutowych. Przedyskutowane zostały płynące z tego korzyści. Rozdział trzeci koncentruje się na pojęciu konfliktu w sieciach Petriego oraz związanych z nim własności sieci i procesów - bezkonfliktowości i trwałości.

Zbiory rzutowe Za punkt wyjścia dla rozdziału drugiego przyjęto wektory Parikha, opisujące w algebraiczny sposób ślady obliczeń w najprostszych w analizie typie alfabetów współbieżnych - alfabetach całkowicie przemennych. Pozwala to w naturalny sposób rozszerzyć wektory Parikha do reprezentacji rzutowych śladów, zaś te do zbiorów rzutowych. Takie podejście do śladów nieskończonych, w przeciwieństwie do podejścia przepłotowego, pozwoliło na redukcję jednego z nieskończonych wymiarów. Zamiast nieskończonych zbiorów utożsamionych ze sobą nieskończonych obliczeń, ślady stają się skończonymi zbiorami nieskończonych rzutów.

W klasie zbiorów rzutowych wyodrębniono przy pomocy indukcyjnej konstrukcji (jako

najmniejsza podklasa spełniająca pewne warunki), podklasę zbiorów wewnętrznie bezkonfliktowych (zwaną też, za Gastinem, rekonstruowalnymi). Są one reprezentacjami rzutowymi śladów. Kilka technicznych lematów doprowadziło do sprecyzowania procedury rekonstrukcji śladu ze zbioru rzutowego. Zdefiniowany został zbiór wszystkich rekonstruowalnych prefiksów zbioru rzutowego. Wykorzystano to do równoważnego sformułowania pokazanej wcześniej operacji rekonstrukcji. Dla zbioru rzutowego operacja rekonstrukcji daje w wyniku ślad, którego reprezentacja rzutowa jest supremum zbioru wszystkich jego wewnętrznie bezkonfliktowych prefiksów.

Zaproponowana rekonstrukcja pozwoliła znacznie uprościć jedno z klasycznych podejść do konkatenacji śladów nieskończonych. Definiując konkatenację w szerszej dziedzinie zbiorów rzutowych i stosując operację rekonstrukcji, uzyskano operację równoważną tej zdefiniowanej przez Kwiatkowską [29]:

Procesy bezkonfliktowe Rozdział trzeci koncentruje się na pojęciu konfliktu. Rozpoczyna się on dyskusją na temat różnych kontekstów w jakich to pojęcie pojawia się w literaturze. Prowadzi to do rozróżnienia dwóch typów konfliktów - statycznych, będących topologiczną własnością grafów sieci Petriego, oraz dynamicznych, które w istotny sposób odwołują się obliczeń wykonalnych w danej sieci i zależą od jej markingu początkowego. W dalszej części pokazane zostało, że każdy konflikt dynamiczny związany jest z pewnym konfliktem statycznym. Nie jest natomiast prawdziwa własność odwrotna. Przytoczony jest przykład, w którym realizacja w postaci dynamicznej konfliktu statycznego zależy od przyjętego markingu początkowego.

Następnie zacytowane są pojęcia trwałości, bezkonfliktowości i konfuzji. Badania nad własnościami trwałych obliczeń dowolnych sieci doprowadziły do zdefiniowania nowego typu sieci Petriego - sieci pokojowych. W nowej klasie sieci przedyskutowane zostały problemy decyzyjne związane z markingami i udowodniona ich decyzyjna równoważność. Wykorzystane tu zostały wyniki dotyczące p/t-sieci, którym poświęcono dużo uwagi i miejsca w rozdziale pierwszym. Dyskusję uzupełnia dowód nierozstrzygalności problemu Pustości Stanu w sieciach pokojowych.

Stosując podobną do wykorzystanej w tym dowodzie technikę, wykazano też równoważność, w sensie obliczeniowym, sieci pokojowych z sieciami inhibitorowymi. Dzięki wynikom Agerwali i Minskiego, pozwoliło to wysnuć wniosek o uniwersalności modelu obliczeń dostarczonego przez sieci pokojowe.

Opisana została słabsza w sensie mocy obliczeniowej podklasa sieci pokojowych - pokojowe sieci wolnego wyboru. Wskazano konstrukcję pozwalającą symulować działanie pokojowych sieci wolnego wyboru przez p/t-sieci (nie jest to możliwe w przypadku dowolnych sieci pokojowych).

W dalszej części rozdziału trzeciego, przez iterację rozumowania prowadzącego do pojęcia sieci pokojowych, zdefiniowana została nieskończona hierarchia sieci pokojowych, a przez rozumowanie graniczne - sieci super-pokojowe.

Pozwoliło to opisać w języku wykonalności obliczeń w kolejnych klasach sieci pokojowych zarówno zjawisko konfliktu jak i konfuzji, unifikując i rozszerzając w sposób naturalny te dwa pojęcia. Sformułowane i udowodnione zostały dwa techniczne lematy (o propagacji oraz genezie konfliktów). Pozwalają one w efektywny sposób badać różne poziomy pokojowości, a w konsekwencji - przerwania wykonalności na różnych poziomach. Lematy te dają pełen opis konfliktów uogólnionych i wskazują ich pochodzenie.

Podsumowaniem rozdziału trzeciego i całej rozprawy jest dyskusja na temat mocy obliczeniowej sieci pokojowych. Rozpoczyna ją wskazanie zależności między zbiorami osiągalnych markingów oraz językami wykonalnych obliczeń na różnych poziomach pokojowości. Przy ograniczeniu się do konkretnej sieci statycznej, tworzą one ciągi zstępujące. W przypadku języków wykonalnych w tych sieciach obliczeń, mogą to być nawet ciągi ściśle zstępujące.

Nieco trudniejsza okazała się dyskusja dotycząca sieci super-pokojowych. Jedynymi sieciami, w których każde obliczenie jest ω -trwałe, są klasyczne sieci trwałe. Sformułowana i udowodniona została wersja twierdzenia Kellera dla sieci ω -pokojowych. Przytoczony przykład pokazuje, że zachowanie niektórych sieci ω -pokojowych symulować można za pomocą trwałych p/t-sieci. Pozwoliło to wysnuć kończące całą rozprawę przypuszczenie, że prawdziwa jest

Hipoteza Sieci ω -pokojowe są w sensie mocy obliczeniowej słabsze od maszyn Turinga, a nawet od p/t-sieci.

Wiele wyników niniejszej rozprawy opublikowane zostało w pracach [4, 5, 25, 36, 37] i przedstawione na lokalnych oraz międzynarodowych konferencjach.

Rozdział 1

Modele systemów współbieżnych

1.1 Podstawowe notacje

Teoria języków formalnych. Niech Σ oznacza alfabet (zbiór niepusty) skończony, litery (elementy) alfabetu Σ oznaczać będziemy małymi literami z początku alfabetu łacińskiego (np. a, b, c).

Zbiór wszystkich słów nad alfabetem Σ oznaczamy przez $\Sigma^\infty = \Sigma^* \cup \Sigma^\omega$, gdzie Σ^* i Σ^ω to odpowiednio zbiory słów skończonych i nieskończonych. Słowa zapisywać będziemy małymi literami z końca alfabetu greckiego (np. σ, τ, ν). Słowo puste oznaczane będzie przez ϵ . Przez $\sigma_{(n)}$ rozumiemy prefiks długości n słowa σ , zaś przez $\sigma[n]$ n -tą literę tego słowa (litery liczymy począwszy od jedynek).

Podalfabet generowany przez słowo σ to $Alph(\sigma) = \{a \mid \exists_n a = \sigma[n]\}$. Zamiast $a \in Alph(\sigma)$ używać będziemy skróconego zapisu $a \in \sigma$. Zbiór wszystkich prefiksów słowa σ oznaczać będziemy przez $\downarrow \sigma$, zaś zbiór słów, których prefiksem jest σ przez $\uparrow \sigma$. Długość słowa σ oznaczać będziemy przez $|\sigma|$, dla $\sigma \in \Sigma^\omega$ przyjmujemy $|\sigma| = \omega$, podobne oznaczenie używane będzie dla mocy zbioru; $|A| = \omega$ oznacza, że moc zbioru A wynosi \aleph_0 . Przez $|\sigma|_a$ oznaczać będziemy liczbę wystąpień litery a w słowie σ . Zgodnie z przyjętymi oznaczeniami, $|\sigma_{(i)}|_{\sigma_{[i]}}$ to liczba wystąpień i -tej litery słowa σ w jego prefiksie długości i .

Standardowym, nieprzemiennym działaniem na słowach jest konkatencja. Zazwyczaj będziemy pomijać znak \cdot symbolizujący to działanie. Przyjmiemy też, że dla słowa nieskończonego σ i dowolnego słowa τ mamy $\sigma\tau = \sigma$.

Definicja 1.1.1 Przez $\Pi_{a,b} : \Sigma^\infty \rightarrow \{a, b\}^\infty$ oznaczać będziemy funkcję rzutowania słowa na alfabet $\{a, b\}$, czyli funkcję spełniającą warunki:

$$\begin{aligned} \Pi_{a,b}(\epsilon) &= \epsilon \\ \Pi_{a,b}(c) &= c && \text{dla } c \in \{a, b\} \\ \Pi_{a,b}(c) &= \epsilon && \text{dla } c \notin \{a, b\} \\ \Pi_{a,b}(c\sigma) &= \Pi_{a,b}(c)\Pi_{a,b}(\sigma) && \text{dla } c \in \Sigma \text{ oraz } \sigma \in \Sigma^\infty \end{aligned}$$

W szczególności $\Pi_{a,a}$ będzie rzutem na alfabet jednoliterowy, oznaczanym również Π_a .

Teoria mnogości. Dla zbioru A przez funkcję charakterystyczną jego podzbioru $B \subseteq A$ rozumiana będzie funkcja $\aleph_B : A \rightarrow \mathbb{N}$ dana wzorem

$$\aleph_B(a) = \begin{cases} 1 & \text{gdy } a \in B \\ 0 & \text{gdy } a \notin B \end{cases}.$$

Jeśli zbiór A jest przeliczalny (o mocy co najwyżej \aleph_0) i dobrze uporządkowany (zadany jest na nim dobry porządek pozwalający jednoznacznie określić bijektywną funkcję $O : N \rightarrow A$, gdzie $N \subseteq \mathbb{N}$ jest zbiorem $|A|$ początkowych elementów zbioru liczb naturalnych), to podzbiory zbioru A utożsamiane będą z wektorami należącymi do $\{0, 1\}^{|A|}$. Wielozbiory (zwane też multizbiorami) elementów pochodzących z A , czyli funkcje $F : A \rightarrow \mathbb{N}$ utożsamiane będą z wektorami należącymi do $v \in \mathbb{N}^{|A|}$. Formalnie powiemy, że $v(i) = F(O(i))$ dla $i \in N$.

Wektorem charakterystycznym zbioru $B \subseteq A$ nazywany będzie wektor (oznaczany tak samo jak funkcja charakterystyczna) $\aleph_B \in \{0, 1\}^{|A|}$ utożsamiany z funkcją charakterystyczną zbioru B (jako podzbioru zbioru A).

Relacja na zbiorze X jest relacją częściowego porządku, jeśli jest zwrotna, przechodnia i antysymetryczna. Zbiór częściowo uporządkowany to para (X, \leq) , gdzie X jest zbiorem, a \leq relacją częściowego porządku na X .

Podzbiór DS zbioru częściowo uporządkowanego (X, \leq) nazywany jest skierowanym, jeśli

$$DS \neq \emptyset \wedge \forall_{x, y \in DS} \exists_{z \in DS} (x \leq z \wedge y \leq z).$$

Przez ograniczenie górne zbioru A , będącego podzbiorem zbioru częściowo uporządkowanego (X, \leq) rozumiemy dowolny element zbioru X większy od każdego elementu ze zbioru A , bardziej formalnie $x \in X$ jest ograniczeniem górnym zbioru A wtedy i tylko wtedy, gdy

$$\forall_{a \in A} a \leq x.$$

Przez $\sup(A)$, gdzie $A \subseteq X$ rozumiemy (o ile istnieje) supremum, czyli najmniejsze ograniczenie górne zbioru A , bardziej formalnie

$$\sup(A) = x \Leftrightarrow x \text{ - ograniczenie górne zbioru } A \text{ oraz } \forall_{z \text{ - ograniczenie górne } A} x \leq z.$$

Zbiór częściowo uporządkowany nazywamy kierunkowo-zupełnym (ang. *dcpo*) jeśli każdy jego podzbiór skierowany posiada kres górny. Jeśli ponadto istnieje element najmniejszy, to zbiór częściowo uporządkowany nazywany jest zupełnym (ang. *cpo*).

Element $x \in X$ zbioru częściowo uporządkowanego (X, \leq) nazywamy skończonym, jeśli dla dowolnego zbioru skierowanego DS zachodzi:

$$x \leq \sup DS \Rightarrow \exists_{d \in DS} x \leq d,$$

czyli jeśli element skończony jest mniejszy od supremum dowolnego zbioru skierowanego, to jest też mniejszy od jednego z elementów tego zbioru. Zbiór częściowo uporządkowany kierunkowo-zupełny (X, \leq) jest zbiorem częściowo uporządkowanym algebraicznym, jeśli każdy element należący do zbioru X jest supremum pewnego zbioru skierowanego zawierającego tylko elementy skończone.

Jeśli dodatkowo zbiór elementów skończonych w X (nazywany bazą zbioru częściowo

uporządkowanego (X, \leq) jest przeliczalny, to ten zbiór częściowo uporządkowany nazywamy ω -algebraicznym.

Niech \sqsubseteq będzie relacją na Σ^∞ , taką że

$$\forall_{\sigma, \tau \in \Sigma^\infty} \sigma \sqsubseteq \tau \Leftrightarrow \exists_{v \in \Sigma^\infty} \tau = \sigma v$$

Relacja \sqsubseteq jest relacją częściowego porządku; będziemy ją nazywać prefiksową relacją częściowego porządku dla słów. Zbiór częściowo uporządkowany $(\Sigma^\infty, \sqsubseteq)$ jest zbiorem częściowo uporządkowanym zupełnym, którego elementy są skończone dokładnie wtedy, gdy są słowami skończonymi (o skończonej długości). W związku z tym jest on ω -algebraiczny.

Korzystając z prefiksowej relacji częściowego porządku wprowadzimy częściową operację przeciwną do konkatencji - różnicę prefiksową $\setminus_{\sqsubseteq} : \Sigma^\infty \times \Sigma^\infty \rightarrow \Sigma^\infty$. Różnica prefiksowa określona jest dla par słów będących w prefiksowej relacji częściowego porządku, jeśli $\sigma \sqsubseteq \tau$, wówczas $\tau \setminus_{\sqsubseteq} \sigma$ to unikalne słowo v takie, że $\sigma v = \tau$. Natomiast w przypadku, gdy $\tau \sqsubseteq \sigma$, to $\tau \setminus_{\sqsubseteq} \sigma = \epsilon$.

Algebra i teoria grafów. Przez \mathbb{N} oznaczany będzie zbiór liczb naturalnych z zerem. Przez \mathbb{N}^k oznaczany będzie zbiór k -wymiarowych wektorów liczb naturalnych. Zbiór \mathbb{N}^k z działaniem dodawania po współrzędnych tworzy monoid. Element neutralny dla tego działania oznaczany będzie $\mathbb{0} = (0, 0, \dots, 0)$, zaś wektor złożony z jedynek - przez $\mathbb{1} = (1, 1, \dots, 1)$.

Dodawanie można rozszerzyć na dowolne podzbiory $X, Y \subseteq \mathbb{N}^k$ określając $X + Y = \{x + y \mid x \in X \wedge y \in Y\}$. Ponadto na podzbiórach monoidu \mathbb{N}^k określona będzie operacja gwiazdki - $X^* = \bigcup_{i \in \mathbb{N}} X_i$, przy czym $X_0 = \mathbb{0}$, zaś $X_{i+1} = X_i + X$. Częściowy porządek $\leq \subseteq \mathbb{N}^k \times \mathbb{N}^k$ określony jest po współrzędnych (tj. $x \leq y$ wtedy i tylko wtedy, gdy każda ze współrzędnych wektora x jest niemniejsza od odpowiedniej współrzędnej wektora y).

Podobnie, przez \mathbb{N}^P rozumiany będzie zbiór funkcji działających ze zbioru P do liczb naturalnych. Dla takich funkcji (o wspólnej dziedzinie) określić można działanie dodawania po współrzędnych oraz stałe funkcje zerową - $\mathbb{0}$ i jednostkową - $\mathbb{1}$.

Podobnie jak w przypadku wektorów działanie dodawania można rozszerzyć na zbiory funkcji $X, Y \subseteq \mathbb{N}^P$ określając $X + Y = \{x + y \mid x \in X \wedge y \in Y\}$ oraz zdefiniować operację gwiazdki - $X^* = \bigcup_{i \in \mathbb{N}} X_i$, przy czym $X_0 = \mathbb{0}$, zaś $X_{i+1} = X_i + X$. Częściowy porządek $\leq \subseteq \mathbb{N}^P \times \mathbb{N}^P$ określony jest po współrzędnych (tj. $x \leq y$ wtedy i tylko wtedy, gdy wartość dla każdego z argumentów funkcji x jest niemniejsza od wartości dla odpowiadającego mu argumentu funkcji y).

Grafem skierowanym nazywać będziemy parę (V, E) , gdzie V jest dowolnym zbiorem przeliczalnym, zaś $E \subseteq V \times V$ jest relacją definiującą łuki w tym grafie. *Ścieżka* w grafie to skończony ciąg wierzchołków v_1, \dots, v_n , spełniający warunek

$$\forall_{i < n} (v_i, v_{i+1}) \in E.$$

Jeśli $v_1 = v_n$, to ścieżkę taką nazywamy *cyklem*.

1.2 Sieci Petriego

Sieci Petriego są intuicyjnym narzędziem pozwalającym modelować systemy i procesy współbieżne. Koncepcję łudząco podobną do sformalizowanej dwie dekady później teorii sieci Petriego, Carl Adam Petri wykorzystał już w roku 1939, w wieku 13 lat. Użył jej wtedy do opisu procesów chemicznych [42]. Oficjalna publikacja [40] pochodzi z roku 1962, tłumaczenie angielskie [41] z 1966.

Teoria sieci Petriego jest wnikliwie badana od lat sześćdziesiątych XX wieku. Wynikiem tych badań było zdefiniowanie wielu typów sieci i bardzo bogata literatura opisująca ich własności. Mimo to, nie ma zbyt wielu opracowań ogólnych w języku polskim, a dwa najpopularniejsze, będące tłumaczeniami monografii Reisiga [44] i Starkego [47] pochodzą z lat osiemdziesiątych XX wieku. W związku z tym poświęcimy dość dużo miejsca na precyzyjne wprowadzenie wykorzystywanych pojęć i zilustrujemy je wymownymi przykładami, proponując czasem inne niż w przytoczonych monografiach nazewnictwo.

Skoncentrujemy się głównie na trzech podstawowych typach sieci, w kolejności rosnącej siły ekspresji będą to sieci elementarne, markowane oraz inhibitorowe. W późniejszych rozdziałach wprowadzimy też nowy typ sieci - sieci pokojowe. Nieco bliżej przyjrzymy się językom obliczeń, ze szczególnym uwzględnieniem obliczeń nieskończonych, jakie są wykonalne w omawianych typach sieci oraz zbiorom markingów osiągalnych i pokrywalnych w tych sieciach, ze szczególnym uwzględnieniem rozszerzonych markingów dopuszczających nieskończenie wypełnione miejsca w sieci. Przedyskutujemy też problemy decyzyjne związane z markingami, powołując się w wielu miejscach na pracę Michaela Hacka [23].

1.2.1 Elementarne sieci Petriego

Definicja 1.2.1 *Elementarną siecią Petriego* nazywamy czwórkę $S = (P, T, F, M_0)$, gdzie P i T są dowolnymi, rozłącznymi zbiorami skończonymi, $F \subseteq P \times T \cup T \times P$ jest *relacją przepływu* zaś $M_0 : P \rightarrow \{0, 1\}$ funkcją określającą *stan (marking) początkowy*. Elementy zbioru P nazywać będziemy *miejscami*, zaś elementy zbioru T - *akcjami*. Przyjmować będziemy, że zbiór P jest uporządkowany w sposób liniowy. *Markingami* nazywane będą funkcje $M : P \rightarrow \{0, 1\}$, lub (stosowane głównie do zadania konkretnej wartości) równoważne im wektory $M \in \{0, 1\}^{|P|}$ których kolejne współrzędne odpowiadają kolejnym miejscom (jeśli definicja sieci nie wskazuje inaczej, dla miejsc przyjmujemy porządek leksykograficzny ich nazw).

Relacja umożliwienia $\mathcal{U} \subseteq \{0, 1\}^P \times T$, zdeterminowana przez relację przepływu i określająca które akcje w danym markingu mają wystarczającą liczbę zasobów aby się wykonać, zdefiniowana jest następująco:

$$M \mathcal{U} a \Leftrightarrow \forall_{p \in P} ((pFa \Rightarrow M(p) = 1) \wedge (aFp \Rightarrow M(p) = 0)).$$

Relacja wykonalności $\mathcal{W} \subseteq \mathcal{U}$, określająca które akcje mogą zostać wykonane w danym markingu, jest w przypadku sieci elementarnych równa relacji umożliwienia.

Krokiem nazywamy trójkę $(M, a, M') \in \{0, 1\}^P \times T \times \{0, 1\}^P$ taką, że $M \mathcal{W} a$ oraz

$$M'(p) = \begin{cases} 1 - M(p) & \text{gdy } aFp \vee pFa \\ M(p) & \text{w pozostałych przypadkach} \end{cases} .$$

Zapis Ma oznaczać będzie marking M' uzyskany z markingu M po wykonaniu akcji a .
Wykonalny ciąg kroków w sieci elementarnej to ciąg kroków (skończony lub nie)

$$u = ((M_1, a_1, M'_1), (M_2, a_2, M'_2), \dots,$$

przy czym dla każdego $i \in \{1, 2, \dots\}$ mamy $M'_i = M_{i+1}$. Ciąg kroków u zapisywany będzie w skrócie następująco $u = M_1 a_1 a_2 \dots$, zaś ciąg akcji $w = a_1 a_2 \dots$ nazywany będzie obliczeniem. Zapis $M' = Mw$ oznaczać będzie, że MwM' jest skończonym i wykonalnym ciągiem kroków, zaś $M\mathcal{W}w$ - że w markingu M można wykonać ciąg akcji w (czyli Mw jest wykonalnym ciągiem kroków). Język wszystkich słów u będących obliczeniami wykonalnymi w markingu początkowym M_0 sieci S to *język definiowany przez sieć S* (oznaczany $L(S)$).

Zbiór wszystkich markingów osiągalnych w sieci S z danego markingu M to

$$[M >= \{M' \in \{0, 1\}^{|P|} \mid \exists_{w \in T^*} MwM' \text{ jest wykonalnym w sieci } S \text{ ciągiem kroków}\}.$$

Sieci statyczne. Pierwsze trzy elementy sieci elementarnej S , czyli zbiory miejsc i akcji oraz relacja przepływu, są statycznymi elementami sieci. Ze względu na nieco odmienną naturę, pozbawioną elementów dynamicznych (przepływu żetonów), wyróżnia się je często [44, 47] jako osobny obiekt - *sieć statyczną* $N = (P, T, F)$ (zwaną też czasem *grafem sieci Petriego* [39]).

Dla każdej akcji a , z tak zdefiniowanej sieci statycznej, oznaczamy:

$$\begin{aligned} \bullet a &= \{p \in P \mid (p, a) \in F\} - \text{zbiór wejść do akcji } a \\ a \bullet &= \{p \in P \mid (a, p) \in F\} - \text{zbiór wyjść z akcji } a. \end{aligned}$$

Symboli $\bullet a$ oraz $a \bullet$ używać będziemy również w stosunku do odpowiadającym tym zbiorom ich funkcji charakterystycznych oraz, podobnie jak w przypadku markingów, $|P|$ -elementowych wektorów o wartościach ze zbioru $\{0, 1\}$.

Graficznie sieci statyczne prezentowane są przez grafy skierowane $G = (P \cup T, F)$. Zbiór wierzchołków złożony jest z dwóch typów obiektów: oznaczanych owalami miejsc oraz przedstawianych przy pomocy prostokątów akcji. Zbiór skierowanych krawędzi koduje relację przepływu F . Zgodnie z definicją relacji przepływu (Definicja 1.2.1), krawędzie muszą łączyć obiekty dwóch różnych typów, czyli graf sieci jest grafem dwudzielnym.

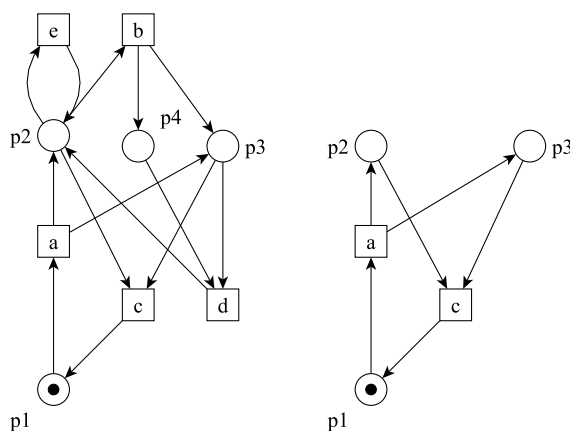
Wprowadzone oznaczenie pozwala sformułować bardziej intuicyjnie relację umożliwienia oraz krok sieci elementarnej. Akcja $a \in T$ w sieci S jest umożliwiona w markingu M (czyli zachodzi $M \mathcal{U} a$) wtedy i tylko wtedy, gdy:

$$\forall_{p \in \bullet a} M(p) = 1 \wedge \forall_{p \in a \bullet} M(p) = 0,$$

czyli wszystkie wejścia do akcji a są pełne, zaś wszystkie wyjścia z akcji a są puste. W każdym z kolejnych kroków wykonywana jest jedna z wykonalnych akcji; wybór akcji odbywa się w sposób niedeterministyczny. Wykonanie polega na zmianie markingu poprzez dekrementację wszystkich wejść do wykonanej akcji oraz inkrementację wszystkich wyjść z tej akcji. Formalnie, wartość wynikowego markingu można zapisać również $M' = M - \bullet a + a \bullet$, gdzie zbiory wejść oraz wyjść z akcji a rozumiane są jako funkcje charakterystyczne tych zbiorów, zaś odejmowanie i dodawanie odbywa się po współrzędnych.

Marking początkowy (i markingi uzyskane w czasie działania sieci) reprezentowane będą graficznie przez umieszczane wewnątrz miejsc czarne, wypełnione owale, zwane żetonami lub tokenami. W sieciach elementarnych w każdym miejscu może znajdować się co najwyżej jeden żeton. Obecność żetonu równoważna jest wartości 1 na współrzędnej odpowiadającej temu miejscu i miejsce takie nazywamy zamarkowanym. W przypadku braku żetonu - marking w danym miejscu wynosi zero.

Uwaga 1.2.2 *Definicja sieci elementarnych dopuszcza, aby miejsce p było zarówno wejściem do jak i wyjściem z akcji a . Sytuację taką nazywamy pętelką (lub ciasną pętlą [44]) i w przypadku sieci elementarnych, jej wystąpienie implikuje permanentne uniemożliwienie akcji a . Kiedy miejsce p będzie zamarkowane, to jako wyjście z akcji a będzie ono uniemożliwiało wykonanie tej akcji. W przeciwnym wypadku, kiedy miejsce p nie będzie zamarkowane, akcja a będzie uniemożliwiona ze względu na brak żetonu w jednym ze swoich wejść. Naturalne jest więc rozważanie sieci, w których nie występują pętelki, zwanych sieciami bezpętelkowymi (w literaturze można je spotkać pod nazwą sieci czystych [13]).*



Rysunek 1.1: Po lewej, sieć elementarna S z markingiem początkowym $(1,0,0,0)$. Po prawej, sieć elementarna S' powstała z sieci S przez usunięcie niewykorzystywanych akcji i miejsc

Przykład 1.2.3 Na Rysunku 1.1 zaprezentowano dwie sieci elementarne. Po lewej stronie znajduje się sieć S , składająca się z czterech miejsc (p_1, p_2, p_3 i p_4) oraz z pięciu akcji (a, b, c, d i e). W markingu początkowym M_0 , żeton znajduje się jedynie w miejscu p_1 , markingowi takiemu odpowiada wektor $(1,0,0,0)$. Na mocy Uwagi 1.2.2, akcja e z pewnością nie będzie nigdy umożliwiona - jest ona połączona pętelką z miejscem p_2 . Ponadto, przy bieżącym markingu początkowym, również akcje b i d nie będą mogły się nigdy wykonać - po wykonaniu jedynej umożliwionej akcji a , otrzymamy marking $(0,1,1,0)$. Wówczas jedyną możliwą akcją jest c , po wykonaniu której, powracamy do markingu początkowego M_0 . Wynika stąd, że zarówno akcja b , jak i d nie będą nigdy umożliwione, czyli ich obecność nie wnosi niczego do działania sieci. Poza tym, stan miejsca p_4 nigdy nie ulega zmianie, a co za tym idzie, to miejsce również można usunąć. Po usunięciu zbędnych akcji i miejsc otrzymujemy identycznie zachowującą się (definiującą identyczne języki obliczeń wykonalnych) sieć elementarną S' , przedstawioną po prawej stronie. Warto zauważyć, że sytuacja uległaby diametralnej zmianie, gdyby marking początkowy M_0 wynosił $(0,1,0,0)$. Wówczas akcje b oraz d wykonywałyby się periodycznie, zmieniając przy tym stan miejsc p_2, p_3 i p_4 , natomiast bezużyteczne byłyby akcje a i c oraz miejsce p_1 .

Jak widać w Przykładzie 1.2.3, trwałe uniemożliwienie akcji może wynikać ze struktury sieci (włącznie z markingiem początkowym) w sposób bardziej skomplikowany niż przedstawiono w Uwadze 1.2.2. Akcje takie uznać można za bezużyteczne i usunąć z sieci bez żadnej straty. Podobnie, w sieci elementarnej mogą istnieć miejsca, których marking nie będzie się zmieniał w żadnym możliwym do wykonania obliczeniu danej sieci. Takie miejsca również można usunąć. Warto w tym miejscu zauważyć, że usuwanie zbędnych miejsc powinno nastąpić po usunięciu zbędnych akcji. W przeciwnym przypadku pewne stałe uniemożliwione akcje mogłyby stać się umożliwiające z powodu usunięcia zawsze pustego wejścia.

W sposób naturalny nasuwa się pytanie: *Czy potrafimy w sposób automatyczny wyłapać wszystkie takie sytuacje na podstawie struktury sieci?* Jest to pytanie o wykonalność akcji oraz pustość lub niepustość miejsca. Przykład 1.2.3 pokazuje, że sama sieć statyczna nie daje wystarczających informacji. Podobne pytania pojawią się w przypadku innych typów sieci Petriego. Pytania te, jak również kilka innych związanych z markingiem problemów decyzyjnych, zostaną sformułowane formalnie i niezależnie od typu sieci w Sekcji 1.2.4. Przedyskutowane też będą zagadnienia związane z rozstrzygalnością tych problemów.

Uwaga 1.2.4 *Chociaż niektóre z wyżej opisanych problemów, ściśle związane z dynamiką sieci, mogą być nierozstrzygalne w niektórych klasach sieci, to część sytuacji patologicznych można wykryć na podstawie budowy sieci statycznej. Od tej chwili przyjęte zostanie założenie, że w żadnej z rozpatrywanych sieci Petriego nie występują miejsca ani akcje izolowane (w sensie izolowanych wierzchołków w teorii grafów). Ponadto, termin sieci elementarne stosowany będzie wyłącznie do bezpętelkowych sieci elementarnych.*

1.2.2 Sieci markowane (p/t-sieci)

Definicja 1.2.5 *Markowaną siecią Petriego (p/t-siecią) nazywać będziemy czwórkę $S = (P, T, W, M_0)$, gdzie P i T są, podobnie jak w przypadku sieci elementarnych, dowolnymi, rozłącznymi zbiorami skończonymi, $W : P \times T \cup T \times P \rightarrow \mathbb{N}$ jest funkcją wagową, odpowiadającą relacji przepływu w sieciach elementarnych, zaś $M_0 : P \rightarrow \mathbb{N}$ funkcją określającą marking początkowy. Podobnie jak w przypadku sieci elementarnych, markingami nazywać będziemy nie tylko funkcje $M : P \rightarrow \mathbb{N}$, ale również równoważne im wektory $M \in \mathbb{N}^{|P|}$, przy czym miejsca porządkujemy zgodnie z porządkiem leksykograficznym (chyba, że definicja sieci precyzuje inaczej).*

Symbolem W oznaczane będą też funkcje $W : 2^P \times T \cup T \times 2^P \rightarrow \mathbb{N}^P$, określone dla $p \in X$ jako $W(X, a)(p) = W(p, a)$ i $W(a, X)(p) = W(a, p)$.

Relacja umożliwienia w p/t-sieci $\mathcal{U} \subseteq \mathbb{N}^P \times T$ zdefiniowana jest następująco:

$$M \mathcal{U} a \Leftrightarrow W(P, a) < M.$$

Relacja wykonalności $\mathcal{W} \subseteq \mathcal{U}$, podobnie jak w przypadku sieci elementarnych, jest równa relacji umożliwienia.

Krokiem nazywamy trójkę $(M, a, M') \in \mathbb{N}^P \times T \times \mathbb{N}^P$ taką, że $M \mathcal{W} a$ oraz $M' = Ma = M - W(P, a) + W(a, P)$. Obliczenie oraz język definiowany przez sieć S rozumiane są identycznie jak w przypadku sieci elementarnych.

Zbiór wszystkich markingów osiągalnych w sieci S z danego markingu M to

$$[M \succ = \{M' : P \rightarrow \mathbb{N} \mid \exists_{w \in T^*} M w M' \text{ jest wykonalnym w sieci } S \text{ ciągiem kroków}\}.$$

Sieć statyczna. W sieci statycznej $N = (P, T, W)$ wchodzącej w skład p/t-sieci, miejsce relacji przepływu zajmuje funkcja wagowa. Jest ona jednak w sposób bezpośredni związana z występującą w przypadku sieci elementarnych relacją. Para (x, y) należy do relacji przepływu F dokładnie wtedy, gdy $W(x, y) > 0$, natomiast w sytuacji, gdy $W(x, y) = 0$ para (x, y) nie należy do F . Znajduje to odzwierciedlenie w reprezentacji graficznej p/t-sieci - podobnie jak w przypadku relacji przepływu, niezerowe wartości funkcji wagowej reprezentowane są przez krawędzie skierowane, etykietowane tym razem wagą połączenia (brak etykiety wagowej przy łuku (x, y) oznacza, że $W(x, y) = 1$).

Wiele pozycji (np. [13, 44, 47]) definiuje p/t-sieci jako piątki uporządkowane, dodając jawnie relację przepływu i używając jej jako dziedziny dodatkowo określonej funkcji wagowej. Zastosowanie zaproponowanej w [23] definicji pomijającej jawną relację przejścia pozwala uprościć pewne notacje, tam gdzie to wygodniejsze odwołując się jednak do zdefiniowanej na podstawie funkcji wagowej W relacji przejścia F . Ponadto, niektórzy autorzy ([13]) odróżniają p/t-sieć (rozumianą jako sieć statyczna) od p/t-systemu (nazywanego tu p/t-siecią). Czasem mogą być dodane ograniczenia na maksymalną pojemność miejsc. Można jednak pokazać ([13]), że takie ograniczenia nie wnoszą niczego do siły ekspresji p/t-sieci, w związku z czym, w dalszych rozważaniach przyjęto, że pojemność wszystkich miejsc jest nieograniczona.

Ponadto, odwołując się do pośredniej definicji relacji przejścia F w p/t-sieciach, oznaczymy zdefiniowane dla sieci elementarnych zbiory wejść do danej akcji oraz wyjść z tej akcji:

$$\begin{aligned} \bullet a &= \{p \in P \mid W(p, a) > 0\} - \text{zbiór wejść do akcji } a \\ a^\bullet &= \{p \in P \mid W(a, p) > 0\} - \text{zbiór wyjść z akcji } a. \end{aligned}$$

Uwaga 1.2.6 *W przeciwieństwie do sieci elementarnych, w p/t-sieciach nie interesuje nas zawartość miejsc będących wyjściami z akcji a . Założenie o nieograniczoności miejsc pozwala umieścić w nich dowolnie dużo żetonów. Istotne natomiast są natomiast wagi połączeń (binarne w sieciach elementarnych).*

Sieci bezpętelkowe. W przeciwieństwie do sieci elementarnych, występowanie pęteli w p/t-sieciach nie powoduje negatywnych skutków. Połączenie akcji z miejscem przy pomocy pętliki nie uniemożliwia wykonania tej akcji. Można jednakże rozpatrywać takie p/t-sieci, w których nie występują pętliki, zwane *sieciami bezpętelkowymi* (lub *czystymi*).

Definicja 1.2.7 P/t-sieć $S = (P, T, W, M_0)$ nazywamy siecią bezpętelkową wtedy i tylko wtedy, gdy

$$\forall_{(x,y) \in P \times T \cup T \times P} W(x, y) > 0 \Rightarrow W(y, x) = 0.$$

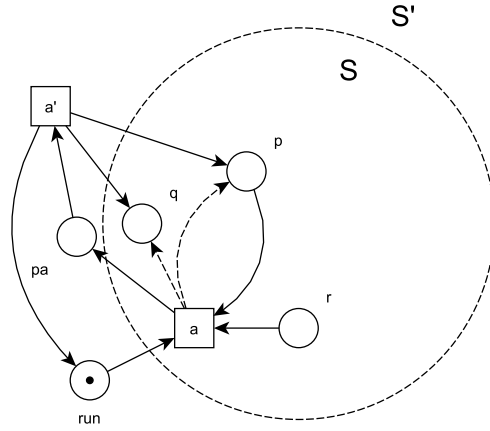
Oznacza to, że żadne miejsce nie może jednocześnie należeć do wejść oraz wyjść z tej samej akcji.

Bezpętelkowe p/t-sieci mają taką samą siłę wyrazu, jak p/t-sieci zawierające pętliki. Rozumiane jest przez to, że dla każdej p/t-sieci S można skonstruować bezpętelkową p/t-sieć S' tak, aby istniała odpowiednia zależność między markingami osiągalnymi w tych sieciach oraz wykonalnymi w nich obliczeniami (patrz Lemat 1.2.10). W jednej z

możliwych konstrukcji wykorzystuje się miejsce sterujące siecią (zwane przez Hacka *run* [23]). Takie dodatkowe miejsce, połączone ze wszystkimi akcjami pierwotnej sieci, jest bardzo wygodnym i często stosowanym narzędziem.

Konstrukcja 1.2.8 Niech $S = (P, T, W, M_0)$ oznacza p/t-sieć. Konstrukcję sieci $S' = (P', T', W', M'_0)$ rozpoczniemy od skopiowania z sieci S zbiorów miejsc i akcji oraz markingu początkowego. Następnie, dla każdej akcji a , dodamy do sieci S' jej kopię a' oraz dodatkowe, puste w markingu początkowym, miejsce p_a . Miejsce p_a jest wyjściem z akcji a oraz wejściem do akcji a' . Dalej, rozdzielona pomiędzy akcją oryginalną i primowaną, zostaje rola akcji w sieci S - wszystkie wyjścia z akcji a zostają przekazane akcji a' . Formalnie oznacza to, że

$$\forall p \in P \quad W'(p, a) = W(p, a) \wedge W'(p, a') = W'(a, p) = 0 \wedge W'(a', p) = W(a, p).$$



Rysunek 1.2: Idea konstrukcji czystej p/t-sieci S' , na podstawie p/t-sieci S

Ponadto, dodajemy wspomniane wyżej globalne miejsce *run*, łącząc je ze wszystkimi akcjami w sieci S' tak, aby było wejściem do akcji oryginalnych i wyjściem z akcji primowanych. W markingu początkowym M'_0 , w miejscu *run* ustawiamy wartość jeden. Formalnie oznacza to, że

$$\begin{aligned} \forall a \in P \quad W'(run, a) = 1 \wedge W'(a, run) = 0, \\ \forall a' \in P' \setminus P \quad W'(run, a') = 0 \wedge W'(a', run) = 1 \end{aligned}$$

oraz

$$M'_0(run) = 1.$$

Na tak zadanym zbiorze miejsc P' zadajemy porządek ustawiając w kolejności leksykograficznej najpierw miejsca ze zbioru P , a następnie nowe miejsca począwszy od primowanych a skończywszy na miejscu *run*, które jest tym samym elementem maksymalnym. W podobny sposób definiować będziemy porządek dla nowego, rozszerzanego zbioru miejsc w pozostałych konstrukcjach tego typu.

Warto zauważyć, że każdemu wykonaniu akcji a w sieci S odpowiada wykonanie sekwencji akcji aa' w sieci S' , a poza pierwotną siecią S krąży cały czas dokładnie jeden żeton, początkowo znajdujący się w miejscu *run*.

Uwaga 1.2.9 W opisywanym przypadku p/t -sieci, powyższa konstrukcja byłaby poprawna również bez dodatkowego miejsca *run*. Dzięki jego dodaniu staje się ona bardziej przejrzysta i uniwersalna (łatwiej jest przenieść ją na inne typy sieci, między innymi na sieci inhibitorowe).

Dla markingu M w sieci S określimy odpowiadający mu marking $10M$ w sieci S' następująco:

$$10M(p) = \begin{cases} M(p) & \text{dla } p \in P \\ 1 & \text{dla } p = \textit{run} \\ 0 & \text{w pozostałych miejscach dodanych.} \end{cases}$$

Oznacza to między innymi, że markingowi początkowemu M_0 w sieci S odpowiada marking początkowy $10M_0 = M'_0$ w sieci S' . Ponadto

Lemat 1.2.10 *Marking M jest osiągalny w sieci S wtedy i tylko wtedy, gdy marking $10M$ jest osiągalny w sieci S' .*

Dowód. Przeanalizujmy na początek znaczenie miejsca *run* dla dynamiki sieci. Zauważmy, że poza oryginalną siecią S krąży cały czas dokładnie jeden żeton (znajdujący się w stanie początkowym w miejscu *run*), zaś akcja a jest umożliwiona w markingu M sieci S dokładnie wtedy, gdy jest umożliwiona w markingu $10M$ w sieci S' . Ponadto, w markingach postaci $10M$ umożliwiaone mogą być tylko akcje ze zbioru T (miejsce *run* jest pełne, jako jedyne z nowych miejsc), podczas gdy w pozostałych osiągalnych markingach sieci S' z pewnością nie są one umożliwiaone (brakuje wtedy żetonu w miejscu *run*). Prócz tego, jedyną akcją umożliwiaoną po wykonaniu akcji a jest a' oraz wykonalność kroku (M, a, M') w sieci S oznacza, że $10Maa'10M'$ jest ciągiem kroków w sieci S' .

(\Rightarrow): Weźmy teraz marking M osiągalny w sieci S i przeprowadźmy proste rozumowanie indukcyjne ze względu na długość minimalnego obliczenia prowadzącego do danego markingu. Jeśli marking M jest osiągalny przez obliczenie o długości zero, czyli ϵ , to jest to marking początkowy M_0 . Marking $10M_0$ jest osiągalny w sieci S' w sposób trywialny, jako marking początkowy w tej sieci. Załóżmy teraz, że marking M jest osiągalny w sieci S przez pewne obliczenie $v = ua$ długości n , zaś dla każdego markingu M' osiągalnego w sieci S przez krótsze niż n obliczenia marking $10M'$ jest osiągalny w sieci S' . Rozważmy marking $M' = M_0u$, osiągalny w sieci S przez obliczenie długości $n - 1$. Na mocy założenia indukcyjnego, odpowiadający mu marking $10M'$ jest osiągalny w sieci S' . Na mocy poczynionych uwag jest w nim umożliwiaona akcja a , po wykonaniu której umożliwiaona jest jedynie akcja a' , oraz $10M'aa'10M$ jest ciągiem kroków w sieci S' .

(\Leftarrow): Z drugiej strony, niech $10M$ będzie markingiem osiągalnym w sieci S' . Podobnie jak poprzednio, przeprowadźmy proste rozumowanie indukcyjne ze względu na długość minimalnego obliczenia prowadzącego do danego markingu. Jeśli marking M jest osiągalny przez obliczenie o długości zero, to jest to marking początkowy $10M_0$, a marking M_0 jest w sposób trywialny osiągalny w sieci S . Na mocy ostatniej z uwag wstępnych, w sieci S' na zmianę wykonywane są akcje ze zbioru T oraz z jego dopełnienia do zbioru T' . W związku z tym, jeśli marking $10M$ jest osiągalny w sieci S przez obliczenie dłuższe niż ϵ , to ma ono postać uaa' , zaś marking osiągnięty przez obliczenie u jest postaci $10M'$. Na mocy założenia indukcyjnego, marking M' jest osiągalny w sieci S . Ponadto (M', a, M) jest krokiem w sieci S , co kończy dowód. \square

Sieci zwykłe (bez wag). Kolejną cechą odróżniającą p/t-sieci od sieci elementarnych jest występowanie funkcji wagowej. Także w tym aspekcie, można wyodrębnić podklasę sieci upodabniających się do sieci elementarnych, w których funkcja wagowa przyjmuje tylko dwie wartości - 0 oraz 1. Takie p/t-sieci nazywamy sieciami zwykłymi.

Definicja 1.2.11 P/t-sieć $S = (P, T, W, M_0)$ nazywamy siecią zwykłą wtedy i tylko wtedy, gdy

$$\forall_{(x,y) \in P \times T \cup T \times P} W(x, y) \leq 1.$$

Podklasa sieci zwykłych ma także taką samą siłę wyrazu jak cała klasa p/t-sieci. Analogicznie do przypadku podklasy sieci bezpętelkowych, dla dowolnej p/t-sieci S można skonstruować sieć zwykłą S' tak, aby istniała odpowiednia zależność między markingami osiągalnymi w tych sieciach. Precyzyjnie istnieje będzie funkcja $f : P \rightarrow 2^{P'}$ definiująca podział zbioru S' w taki sposób, że marking M jest osiągalny w sieci S dokładnie wtedy, gdy istnieje osiągalny w sieci S' marking M' taki że $M(p) = \sum_{q \in f(p)} M'(q)$ (patrz Lemat 1.2.13).

Wynikiem działania poniższej konstrukcji [47] będzie nie tylko sieć zwykła, ale również bezpętelkowa. Nie daje się ona w tak szerokim zakresie rozszerzyć na inne klasy sieci (miedzy innymi dla sieci inhibitorowych potrzebne będzie złożenie obu prezentowanych konstrukcji).

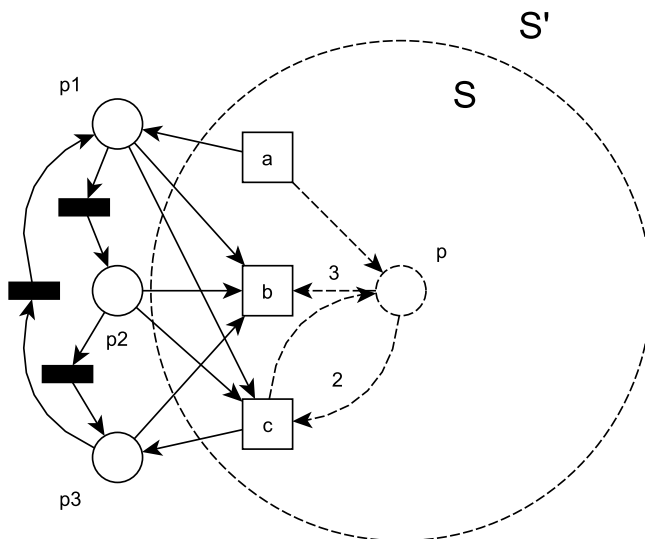
Konstrukcja 1.2.12 Niech $S = (P, T, W, M_0)$ oznacza p/t-sieć. Skonstruujemy na jej podstawie p/t-sieć $S' = (P', T', W', M'_0)$, w której akcje pozostaną niezmienione, natomiast dla każdego miejsca $p \in P$ utworzymy zbiór miejsc $\{p_1, p_2, \dots, p_k\}$, połączonych nowymi akcjami $t_1^p, t_2^p, \dots, t_k^p$ (tworząc tym samym cykl w grafie sieci). Rozmiar takiego zbioru jest zależny od funkcji wagowej W i może być różny dla różnych miejsc.

W celu ustalenia jego wielkości, zdefiniujemy nową funkcję $R : P \rightarrow \mathbb{N}$, określającą rząd każdego z miejsc. Funkcja ta przyporządkowuje każdemu z miejsc maksymalną sumę wag, z jakimi dane miejsce połączone jest z którąś z akcji; formalnie $R(p) = \max\{W(a, p) + W(p, a) \mid a \in T\}$. Ważnym, szczególnym przypadkiem, są miejsca rzędu jeden. Tworzymy dla nich zbiór składający się wyłącznie z miejsca p_1 - dodanie akcji t_1^p nie wnosi nic do działania takiej sieci (wejściem do tej akcji i wyjściem z niej byłoby miejsce p_1) i wprowadza zbędną pętelkę.

Ostatecznie sieć $S' = (P', T', W', M'_0)$ definiujemy następująco:

- $P' = \bigcup_{p \in P} \{p_k \mid k = 1, \dots, R(p)\};$
- $T' = T \cup \bigcup_{p \in P, R(p) > 1} \{t_i^p \mid i = 1, \dots, R(p)\};$
- $$F' = \bigcup_{(p,a) \in F} \{(p_i, a) \mid i = 1, \dots, W(p, a)\} \\ \cup \bigcup_{(a,p) \in F} \{(a, p_i) \mid i = W(p, a) + 1, \dots, W(p, a) + W(a, p)\} \\ \cup \bigcup_{p \in P, R(p) > 1} (\{(p_i, t_i^p) \mid i = 1, \dots, R(p)\} \\ \cup \{(t_i^p, p_i + 1) \mid i = 1, \dots, R(p) - 1\} \\ \cup (t_R^p(p), p_1));$$
- $$W'(x, y) = \begin{cases} 1 & \text{dla } (x, y) \in F' \\ 0 & \text{dla } (x, y) \notin F' \end{cases};$$

$$\bullet M'_0(p_i) = \begin{cases} M_0(p) & \text{dla } i = 1 \\ 0 & \text{dla } i \neq 1 \end{cases}.$$



Rysunek 1.3: Idea konstrukcji zwykłej i bezpętelkowej p/t-sieci S' , na podstawie p/t-sieci S . Miejsce p zastąpione jest przez zbiór złożony z trzech miejsc - p_1 , p_2 oraz p_3 oraz akcji t_1^p , t_2^p i t_3^p symbolizowanych przez czarne prostokąty. Przerwane strzałki symbolizują relację przepływu w sieci S (wraz z wagami w indeksie), zaś ciągłe - relację przepływu w sieci S'

Zdefiniujemy teraz relację \equiv na zbiorze $\mathbb{N}^{P'}$, zawierającym wszystkie osiągalne w sieci S' markingi. Weźmy dwa markingi $M_1, M_2 \in \mathbb{N}^{P'}$. Powiemy, że

$$M_1 \equiv M_2 \text{ wtw } \forall_{p \in P} \sum_{i=1, \dots, R(p)} M_1(p_i) = \sum_{i=1, \dots, R(p)} M_2(p_i),$$

czyli markingi te mają takie same liczby żetonów w każdym z nowo powstałych cykli. Łatwo sprawdzić, że tak zdefiniowana relacja \equiv jest relacją równoważności. Zbiór ilorazowy powstały z podzielenia zbioru $\mathbb{N}^{P'}$ przez relację \equiv jest natomiast izomorficzny ze zbiorem \mathbb{N}^P . Ponadto, dla markingu M osiągalnego w p/t-sieci S określimy odpowiadający mu marking $10M$ w bezpętelkowej i zwykłej sieci S' :

$$10M(p_i) = \begin{cases} M(p) & \text{dla } i = 1 \\ 0 & \text{dla } i \neq 1. \end{cases}$$

Tak więc $M'_0 = 10M_0$.

Poniższy lemat określa zależności między markingami osiągalnymi w pierwotnej sieci S oraz powstałej w wyniku konstrukcji sieci S' .

Lemat 1.2.13 *Niech $S = (P, T, W, M_0)$ będzie p/t-siecią, $S' = (P', T', W', M'_0)$ bezpętelkową i zwykłą p/t-siecią powstałą z sieci S w wyniku Konstrukcji 1.2.12, zaś $\equiv \subseteq \mathbb{N}^{P'} \times \mathbb{N}^{P'}$ relacją równoważności markinguów opisaną w tej konstrukcji. Wówczas osiągalność markingu w sieci S' jest niezmiennikiem relacji \equiv , czyli*

$$(M'_1 \equiv M'_2) \Rightarrow (M'_1 \in [M'_0 > \Leftrightarrow M'_2 \in [M'_0 >).$$

Ponadto, marking M jest osiągalny w sieci S dokładnie wtedy, gdy marking $10M$ jest osiągalny w sieci S' .

Dowód. Na początek zauważmy, że dzięki akcjom t_i^p w sieci S' żetony mogą w swobodny sposób przepływać w obrębie powstałych w konstrukcji cykli. Oznacza to, że dla markingów M'_1 i równoważnego mu markingów M'_2 istnieje ciąg akcji t_1, \dots, t_k taki, że $M'_1 t_1 \dots t_k M'_2$ jest wykonalnym ciągiem kroków, co uzasadnia pierwszą część tezy.

Weźmy teraz marking $M : P \rightarrow \mathbb{N}$ i akcję $a \in T$ umożliwiającą w tym markingu. Uzasadnimy, że wśród markingów równoważnych markingowi $10M$ istnieje taki, w którym akcja a jest umożliwiona. W tym celu rozważmy marking M' zadany w następujący sposób:

$$M'(p_i) = \begin{cases} M(p) - W(p, t) + 1 & \text{dla } i = 1 \\ 1 & \text{dla } 1 < i \leq W(p, t) \\ 0 & \text{dla } i > W(p, t). \end{cases}$$

Ponieważ akcja a jest umożliwiona markingu M to $M(p) - W(p, t) + 1 > 0$. W sposób oczywisty $M' \equiv 10M$. Ponadto, zgodnie z definicją relacji F , akcja a jest umożliwiona w markingu M' .

Z drugiej strony, założmy że akcja $a \in T$ jest umożliwiona w markingu $M' : P \rightarrow \mathbb{N}$. Wówczas jest ona też umożliwiona w markingu M'_n , w którym w każdym niepustym miejscu markingu M' znajduje się dokładnie jeden żeton. W markingu M_n odpowiadającym $10M_n \equiv M'_n$, zgodnie z określeniem relacji F , w każdym miejscu znajduje się przynajmniej $W(p, a)$ żetonów. W związku z tym akcja a jest umożliwiona w markingu M_n , a zatem również w większym markingu M odpowiadającym $10M \equiv M'$.

Pokazaliśmy, że każda akcja $a \in T$ jest umożliwiona w markingu $M : P \rightarrow \mathbb{N}$ dokładnie wtedy, gdy jest ona umożliwiona w przynajmniej jednym z markingów równoważnych odpowiadającemu mu markingowi $10M$. Ponieważ $M'_0 = 10M_0$, proste rozumowanie indukcyjne daje drugą część tezy. \square

Uwaga 1.2.14 Powyższy lemat ustala równoważność w sensie osiągalnych markingów między p/t -siecią S a powstałą w wyniku Konstrukcji 1.2.12 jej zwykłą i bezpętelkową odpowiedniczką. Warto jednak zwrócić uwagę na dwie płaszczyzny, w których odbywają się obliczenia w sieci S' . Z jednej strony wykonywane są w niej akcje przekopiowane bezpośrednio z sieci S , które można określić jako akcje istotne z punktu widzenia zachowania systemu. Z drugiej strony, wykonywane są również dodatkowe akcje t_i^p , umożliwiające swobodny przepływ żetonów wewnątrz cykli, ale nie zmieniające w istotny sposób stanu sieci. Występowanie tego drugiego typu akcji może zdominować obliczenia w sieci, a ta, po transformacji, może stracić pozytywne z punktu widzenia etyki obliczeń własności.

1.2.3 Rozszerzenia p/t -sieci

Sieci inhibitorowe.

Definicja 1.2.15 Siecią inhibitorową [3, 23] nazywamy piątkę $S = (P, T, W, I, M_0)$, gdzie (P, T, W, M_0) jest p/t -siecią, natomiast $I \subseteq P \times T$ relacją inhibicji. Relacja umożliwienia $\mathcal{U} \subseteq \mathbb{N}^P \times T$ określona jest następująco:

$$M \mathcal{U} a \Leftrightarrow [\forall_{p \in \bullet a} W(p, a) \leq M(p) \wedge \forall_{p \in \circ a} M(p) = 0,$$

gdzie

${}^{\circ}a = \{p \in P \mid (p, a) \in I\}$ - zbiór *inhibitorów* dla akcji a .

Podobnie jak w przypadku sieci elementarnych i p/t-sieci, relacja wykonalności \mathcal{W} jest równa relacji umożliwienia. Krok, obliczenie i zbiór wszystkich markingów osiągalnych definiujemy i oznaczamy tak jak w przypadku p/t-sieci.

W grafie reprezentującym sieć, relacja inhibicji znajduje odzwierciedlenie jako nowy typ połączenia węzłów. W przeciwieństwie do zwykłych grotów kończących krawędzie oznaczające relację przepływu, krawędzie inhibitorowe zakończone są pustymi okręgami (w tej pracy rombami).

Podobnie jak w przypadku p/t-sieci, na podstawie funkcji wagowej W zdefiniować można relację przepływu F (dodawaną czasem jawnie w definicji sieci inhibitorowej). Analogicznie wygląda również kwestia pojemności miejsc (przyjmować będziemy, że jest ona nieograniczona) oraz zbiorów wejść do akcji i wyjść z niej.

Różnica w działaniu, w stosunku do klasycznych p/t-sieci, jest widoczna w definicji umożliwienia, a co za tym idzie również wykonalności akcji. Aby akcja była możliwa do wykonania, potrzebna jest nie tylko wystarczająca liczba żetonów w miejscach wejściowych, ale również pustość wszystkich miejsc inhibitorowych dla tej akcji. Nieistotna pozostaje liczba żetonów w wyjściach.

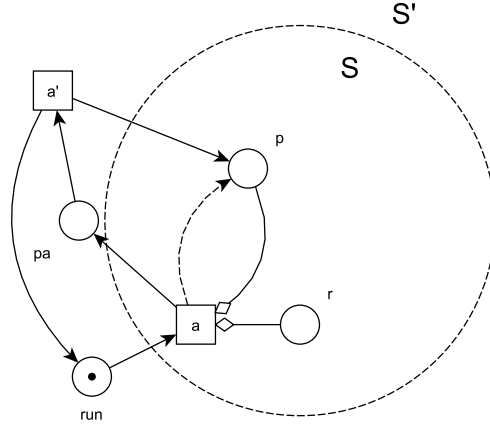
Uwaga 1.2.16 *Podobnie jak w przypadku sieci elementarnych, umożliwienie akcji w sieciach inhibitorowych może wymagać pustości niektórych miejsc. Sama definicja sieci dopuszcza sytuację, w której to samo miejsce p jest wejściem oraz inhibitorem dla tej samej akcji a . Powoduje to oczywiście permanentne uniemożliwienie akcji a , gdyż miejsce p nie może jednocześnie być puste i zawierać dodatniej liczby żetonów. Możliwe, a nawet wygodne z praktycznego punktu widzenia, są natomiast dwa typy pętelek - kiedy miejsce jest jednocześnie wejściem i wyjściem z tej samej akcji oraz wyjściem z akcji i inhibitorem dla tej samej akcji. Pierwszy przypadek może być wykorzystany do sprawdzenia jakiegoś warunku bez jego modyfikacji, drugi może zostać użyty przy modelowaniu semaforów.*

Bezpętelkowe i zwykłe sieci inhibitorowe. Ponownie, sieci bez pętelek (zarówno zwykłych jak i inhibitorowych) nazywać będziemy inhibitorowymi sieciami bezpętelkowymi (czystymi), zaś sieci z zero-jedynkową funkcją wagową - zwykłymi. Jednakże, w związku z występowaniem w sieciach inhibitorowych nowego typu zależności między miejscami i akcjami, a co za tym idzie, również nowych typów pętelek, pojawia się w sposób naturalny pytanie o aktualność Konstrukcji 1.2.8 oraz Konstrukcji 1.2.12, pozwalających usunąć kosztem rozmiaru sieci pętelki i wagi.

Niemal bez zmian pozostaje Konstrukcja 1.2.8, przekształcająca sieć inhibitorową w czystą sieć inhibitorową. Tym razem globalne miejsce *run* jest konieczne. Bez niego można by było na przykład wykonać, jedna po drugiej, dwie akcje które, przy pomocy odpowiednich inhibitorów, wykluczają się nawzajem.

Konstrukcja 1.2.17 Ponownie, dla każdej akcji a , dodajemy jej kopię a' oraz miejsce p_a . Akcja a' przejmuje rolę wyjścia i wszystkie połączenia z wyjściami z akcji a w pierwotnej sieci, oryginalnej akcji a pozostaje zaś rola sprawdzenia wykonalności i pobrania żetonów z wejść - bez zmian pozostają więc zależności z miejscami będącymi wejściami do oraz

inhibitorami dla akcji a . Ponadto pojawia się globalne miejsce run , tworzące wraz z każdą trójką (a, p_a, a') cykl.



Rysunek 1.4: Idea konstrukcji bezpiecznej sieci inhibitorowej S' , na podstawie sieci inhibitorowej S . Przedstawione zostały tylko sytuacje nie uwzględnione na Rysunku 1.2

Nieco większy problem pojawia się w przypadku drugiej konstrukcji, pozwalającej usunąć z sieci wagi. Tym razem konieczność sprawdzenia pustości danego miejsca, wymaga sprawdzenia pustości całego odpowiadającego mu zbioru miejsc. Tym samym, jakakolwiek pętla inhibitorowa występująca w sieci S , pozostanie również w sieci S' . Tak więc, dla dowolnej sieci inhibitorowej otrzymamy jedynie zwykłą sieć inhibitorową.

Konstrukcja 1.2.18 Niech $S = (P, T, W, I, M_0)$ będzie siecią inhibitorową. Skonstruujemy dla niej sieć $S' = (P', T', W', I', M'_0)$, w której funkcja wagowa będzie funkcją zero-jedynkową. Większość konstrukcji przebiega analogicznie do Konstrukcji 1.2.12. Tak więc ponownie kopiujemy akcje ze zbioru T , zaś dla każdego miejsca $p \in P$ tworzymy zbiór miejsc połączonych w cykl nowymi akcjami, którego wielkość zależy od rzędu $R(p)$ miejsca p . Identycznie określamy także relację $\equiv \subseteq \mathbb{N}^{P'} \times \mathbb{N}^{P'}$.

Jedynym rozszerzeniem konstrukcji klasycznej jest określenie nowej relacji inhibicji I' . Dla każdej pary $(p, a) \in I$ określamy zbiór par łączący cały zbiór miejsc odpowiadający miejscu p z akcją a . Formalnie:

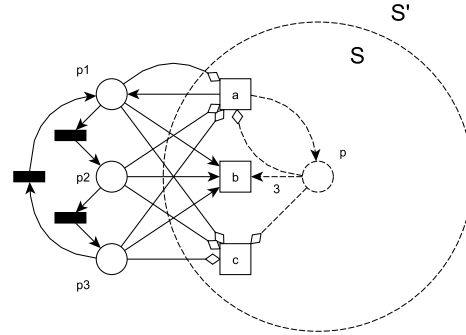
$$I' = \{(p_i, a) \mid (p, a) \in I \wedge i = 1, \dots, R(p)\}.$$

Dzięki Konstrukcji 1.2.18 uzyskujemy rezultat podobny do tegoż dla p/t-sieci:

Lemat 1.2.19 Niech $S = (P, T, W, I, M_0)$ będzie siecią inhibitorową, $S' = (P', T', W', I', M'_0)$ zwykłą siecią inhibitorową powstałą z sieci S w wyniku Konstrukcji 1.2.18, zaś $\equiv \subseteq \mathbb{N}^{P'} \times \mathbb{N}^{P'}$ relacją równoważności markingów opisaną w Konstrukcji 1.2.12. Wówczas osiągalność markingów w sieci S' jest niezmiennikiem relacji \equiv , czyli

$$(M'_1 \equiv M'_2) \Rightarrow (M'_1 \in [M'_0 > \Leftrightarrow M'_2 \in [M'_0 >).$$

Ponadto, marking M jest osiągalny w sieci S dokładnie wtedy, gdy marking $10M$ jest osiągalny w sieci S' .



Rysunek 1.5: Idea konstrukcji zwykłej sieci inhibitorowej S' , na podstawie sieci inhibitorowej S . Miejsce p zastąpione jest przez zbiór złożony z trzech miejsc - $p1$, $p2$ oraz $p3$, które wraz z akcjami t_1^p , t_2^p oraz t_3^p symbolizowanymi przez czarne prostokąty, tworzą cykl. Przerwane strzałki symbolizują relacje przepływu oraz inhibicji w sieci S (wraz z wagami w indeksie), zaś ciągle - relacje w sieci S'

Dowód. Dowód można przeprowadzić analogicznie do dowodu Lematu 1.2.13. Pierwsza część jest nawet identyczna - w obrębie nowo powstałych cykli nie występują żadne łuki inhibitorowe.

Nie stanowią one też problemu w dowodzie drugiej części tezy, gdyż aby akcja $a \in T$ połączona w sieci S łukiem inhibitorowym z miejscem p mogła być umożliwiona w markingu M' w sieci S' , to wszystkie miejsca p_i muszą być puste. \square

Zgodnie z tezą powyższego lematu oraz poczynionymi wcześniej uwagami, uzyskana w wyniku Konstrukcji 1.2.18 sieć inhibitorowa S' jest siecią zwykłą, ale niekoniecznie czystą. Jednakże, w połączeniu z Konstrukcją 1.2.17 uzyskamy pożądaný rezultat:

Fakt 1.2.20 *Niech $S = (P, T, W, I, M_0)$ będzie czystą siecią inhibitorową. Wówczas powstała w wyniku Konstrukcji 1.2.18 zwykła sieć inhibitorowa S' jest również siecią czystą.*

Inne rozszerzenia sieci Petriego. Sekcja ta zawierać będzie tylko kilka definicji o charakterze informacyjnym. Choć dalsze rozważania skupią się na trzech przedstawionych klasach sieci Petriego - sieciach elementarnych, markowanych oraz inhibitorowych, to dla pełnego obrazu niektórych problemów mogą mieć miejsce również odwołania do innych klas sieci. W szczególności dotyczy to sieci czyszczących oraz podwajających. Dodatkowo zdefiniujemy w tej sekcji sieci samomodyfikujące [48] - uogólnienie sieci markowanych pozwalające wymodelować większość uogólnień sieci Petriego. Zdefiniujemy również strategię maksymalnych kroków Burkharda [8], która oddziela umożliwienie od wykonalności. Sekcja ta jest ważna z punktu widzenia opisywanych w Rozdziale 3 sieci pokojowych, które podobnie jak maksymalne kroki Burkharda mają ścisły związek z konfliktami między akcjami.

Definicja 1.2.21 *Czyszczącą siecią Petriego [3, 10, 17] nazywamy p/t-sieć rozszerzoną o relację czyszczenia $R \subseteq T \times P$, czyli jest to piątka $S = (P, T, W, R, M_0)$, gdzie z relacją R związane są nowe łuki w tej sieci - łuki czyszczące. Relacje umożliwienia i wykonalności są tożsame i identyczne jak w przypadku p/t-sieci.*

Nieco inaczej wygląda definicja kroku, a właściwie markingu wynikowego M' . Dla miejsc p nie będących z akcją a w relacji R , identycznie jak w p/t-sieciach, żetony zdejmowane są z wejść do a oraz dodawane są do wyjść z a ($M'(p) = M(p) + W(a, p) - W(p, a)$). Czyszczone są natomiast wszystkie wyjścia czyszczące z a , czyli $aRp \Rightarrow M'(p) = 0$.

Wprost z definicji zależność aRp powoduje, że wartość jaką przyjmuje $W(a, p)$ nie ma znaczenia. Można przyjąć, że $aRp \Rightarrow W(a, p) = 0$, czyli obecność łuku czyszczącego między miejscem p a akcją a wyklucza możliwość występowania między tymi dwoma węzłami zwykłego łuku. Łuki czyszczące, podobnie jak łuki inhibitorowe, zakończone są pustymi okręgami (w tej pracy rombami).

Definicja 1.2.22 *Podwajającą siecią Petriego* [17] nazywamy p/t-sieć rozszerzoną o relację podwajania $D \subseteq T \times P$. Formalnie jest to więc piątka $S = (P, T, W, D, M_0)$, gdzie z relacją D związane są nowe łuki w tej sieci - łuki podwajające.

Relacje umożliwienia i wykonalności są tożsame i identyczne jak w przypadku p/t-sieci. Ponownie, nieco inaczej wygląda definicja kroku. Identycznie jak w p/t-sieciach, żetony zdejmowane są z wejść do a zaś dodawane do miejsc, które są zwykłymi (nie będącymi w relacji D) wyjściami z akcji a . Stan miejsc będących wyjściami podwajającymi są natomiast podwajane, czyli $aDp \Rightarrow M'(p) = 2M(p)$.

Ponadto, zależność aDp pociąga za sobą $W(a, p) = 0$, czyli łuk podwajający można traktować jako specjalny typ zwykłego łuku wyjściowego. Niemożliwe jest jednoczesne wystąpienie między miejscem a akcją podwajającego i zwykłego łuku wyjściowego. Łuki podwajające zakończone są podwójnymi grotami.

Definicja 1.2.23 *Samomodyfikującą lub uogólnioną siecią Petriego* [48, 17] nazywamy czwórkę $N = (P, T, \mathbb{W}, M_0)$, gdzie P , T oraz M_0 są podobnie jak w przypadku p/t-sieci zbiorami miejsc, akcji oraz markingu początkowym. Czwarty element - funkcja wagowa $\mathbb{W} : P \times T \cup T \times P \rightarrow \mathbb{N}[P]$ (gdzie przez $\mathbb{N}[P]$ rozumiemy zbiór wielomianów nad liczbami naturalnymi o zmiennych pochodzących ze zbioru P) pozwala opisać umożliwienie i modyfikację stanu sieci używając funkcji wielomianowych, a nie stałych jak w przypadku p/t-sieci.

Podobnie jak w przypadku p/t-sieci, akcja a jest umożliwiona, jeśli w każdym jej wejściu znajduje się wymagana liczba żetonów, które są zużywane przez tę akcję. Wykonanie akcji owocuje zwiększeniem zawartości jej wyjść, zgodnie z wartościami wielomianów wyliczonymi przed wykonaniem akcji. Można więc traktować sieć samomodyfikującą jako p/t-sieć, której funkcja wagowa zmienia się w zależności od markingu (w sposób określony przez funkcję \mathbb{W}).

Uwaga 1.2.24 *Dotychczas zdefiniowane nowe łuki możemy opisać używając sieci samomodyfikujących w następujący sposób:*

$$\begin{aligned} \text{inhibitorowe } I(p, a) &\rightarrow \mathbb{W}(p, a) = 2p \\ \text{czyszczące } R(a, p) &\rightarrow \mathbb{W}(p, a) = p \\ \text{podwajające } D(a, p) &\rightarrow \mathbb{W}(a, p) = p \end{aligned}$$

Definicja 1.2.25 Sieć markowana ze strategią *maksymalnych kroków Burkharda*, to zwykła p/t-sieć $S = (P, T, W, M_0)$, z klasycznie zdefiniowanym umożliwieniem. Sytuacja zmienia się w definicji pojedynczego kroku. Zamiast mówić o pojedynczych akcjach, pod uwagę bierzemy ich jednocześnie wykonywane zbiory - *transakcje*. W ramach jednej transakcji musimy wykonać maksymalnie dużo współbieżnych akcji. Pozostałe po pobraniu wymaganych zasoby nie mogą być wystarczające dla jeszcze jednej akcji.

Formalnie transakcja t to podzbiór zbioru akcji T . Transakcja $t \subseteq T$ jest wykonalna w markingu M wtedy i tylko wtedy gdy $\sum_{a \in t} W(P, a) \leq M$ oraz $\neg(\exists_{b \notin t} W(P, b) + \sum_{a \in t} W(P, a) \leq M)$, czyli suma wejść do akcji należących do transakcji t nie przekracza M i nie ma wystarczającej ilości zasobów na żadne rozszerzenie transakcji t .

1.2.4 Problemy decyzyjne w sieciach Petriego

W Przykładzie 1.2.3 zasygnalizowane zostały problemy z wykrywaniem trwale uniemożliwionych akcji oraz miejsc, których stan nie ulega zmianie w czasie obliczeń w sieci. Choć przykład ten dotyczy sieci elementarnych, podobne sytuacje zachodzą we wszystkich typach sieci Petriego. Daje to motywację dla sprecyzowania i rozważenia problemów związanych z osiągalnymi w sieciach Petriego markingami. Problemy te sformułujemy w postaci ogólnej (używając słowa sieć), pasującej do dowolnej klasy sieci Petriego. Oczywiście, ich status decyzyjny (rozstrzygalność bądź nierozstrzygalność) może być różny (i jest różny) w różnych klasach sieci, co będzie badane w dalszej fazie rozprawy. Ze względu na charakterystykę, problemy te podzielone zostały na dwie grupy – problemy związane z pokrywalnością oraz problemy związane z osiągalnością danego markingu.

Problemy związane z pokrywalnością.

Problem 1.2.26 P1 - Pokrywalność markingu.

Jest to pytanie, czy w danej sieci pokrywalny jest z góry zadany marking, czyli czy da się tak poprowadzić obliczenia, aby w każdym z miejsc pojawiła się co najmniej z góry zadana liczba żetonów. Bardziej formalnie, jest to pytanie, czy dla danej sieci S oraz danego markingu $M : P \rightarrow \mathbb{N}$:

$$\exists_{M' \in [M_0]} \forall_{p \in P} M'(p) \geq M(p).$$

Problem 1.2.27 P2 - Wykonalność akcji.

Jest to pytanie, czy w danej sieci konkretna akcja ma możliwość wykonania się (jest w relacji \mathcal{W} z jakimś markingiem osiągalnym), to znaczy czy istnieje jakiegokolwiek obliczenie w tej sieci zawierające tę akcję. Bardziej formalnie, jest to pytanie, czy dla danej sieci S oraz akcji $a \in T$:

$$\exists_{M \in [M_0]} M \mathcal{W} a.$$

Problem 1.2.28 P3 - Niepustość miejsca.

Jest to pytanie, czy w danej sieci osiągalny jest jakikolwiek marking, w którym w z góry zadany miejscu jest co najmniej jeden żeton. Bardziej formalnie, jest to pytanie, czy dla danej sieci S oraz miejsca $p \in P$:

$$\exists_{M \in [M_0]} M(p) > 0.$$

Problemy związane z osiągalnością.

Problem 1.2.29 O1 - Osiągalność markingu.

Jest to pytanie, czy w danej sieci osiągalny jest z góry zadany marking, czyli czy da się tak poprowadzić obliczenia, aby w każdym z miejsc pojawiła się z góry zadana liczba żetonów. Bardziej formalnie, jest to pytanie, czy dla danej sieci S o zbiorze miejsc P i markingu początkowym $M_0 : P \rightarrow \mathbb{N}$ oraz danego markingu $M : P \rightarrow \mathbb{N}$:

$$M \in [M_0 > .$$

Problem 1.2.30 O2 - Pustość stanu.

Jest to pytanie, czy w danej sieci osiągalny jest z pustym markingiem, czyli czy da się tak poprowadzić obliczenia, aby wyczyścić wszystkie miejsca jednocześnie. Bardziej formalnie, jest to pytanie, czy dla danej sieci S o markingu początkowym M_0 :

$$\emptyset \in [M_0 > .$$

Problem 1.2.31 O3 - Pustość miejsca.

Jest to pytanie, czy w danej sieci osiągalny jest jakikolwiek marking, w którym w z góry zadanym miejscu nie ma żadnego żetonu. Bardziej formalnie, jest to pytanie, czy dla danej sieci S o zbiorze miejsc P i markingu początkowym $M_0 : P \rightarrow \mathbb{N}$ oraz miejsca $p \in P$:

$$\exists_{M \in [M_0 >} M(p) = 0.$$

Problem 1.2.32 O4 - Osiągalność pokrytego podmarkingu.

Jest to pytanie, czy w danej sieci osiągalny jest jakikolwiek marking $M' : P \rightarrow \mathbb{N}$, w którym określona część miejsc ma wartości zgodne z markingiem M . Bardziej formalnie, jest to pytanie, czy dla danej sieci S o zbiorze miejsc P i markingu początkowym $M_0 : P \rightarrow \mathbb{N}$, markingu M oraz zbioru miejsc $D \subseteq P$:

$$\exists_{M' \in [M_0 >} (\forall_{p \notin D} M'(p) \geq M(p)) \wedge (\forall_{p \in D} M'(p) = M(p)).$$

Zależności między problemami związanymi z markingiem. Wymienione w poprzednim paragrafie problemy są konstruktywnie decyzyjnie równoważne wewnątrz grup – w każdej z wymienionych klas sieci problemy z grupy P1-P3 są wszystkie rozstrzygalne bądź wszystkie nierozstrzygalne (i to samo w grupie O1-O4). Trochę trudniej znaleźć zależności pomiędzy tymi grupami. I tak, o ile rozstrzygalność problemu osiągalności konstruktywnie pociąga za sobą rozstrzygalność problemu pokrywalności we wszystkich klasach sieci, to konstruktywna równoważność została udowodniona tylko w przypadku sieci inhibitorowych, a w przypadku sieci czyszczących czy podwajających [17] mamy wręcz rozstrzygalną pokrywalność przy nierozstrzygalnej osiągalności. Przez dowolną klasę sieci Petriego będziemy w dalszej części rozumieć tylko p/t-sieci i sieci inhibitorowe. Wiele z przedstawionych konstrukcji działa również dla pozostałych z wymienionych w tym rozdziale klas.

W dalszej części tego paragrafu uzasadnimy decyzyjną równoważność wewnątrz grup problemów związanych z pokrywalnością oraz osiągalnością. użytą metodą jest redukcja jednego problemu do drugiego. Zaczniemy od pierwszej grupy, czyli trzech problemów związanych pokrywalnością. Jako pierwsze dwa problemy rozważmy pokrywalność oraz niepustość miejsca.

Fakt 1.2.33 P1 rozstrzygalny \Rightarrow P3 rozstrzygalny:

Załóżmy, że w jakiejś klasie sieci Petriego rozstrzygalny jest problem pokrywalności. Wówczas jest w niej rozstrzygalny również problem niepustości miejsca, jako szczególny przypadek problemu pokrywalności - pytamy wówczas o pokrywalność markingu pustego we wszystkich miejscach oprócz p ; w tym jednym, wyróżnionym w problemie P3 miejscu, żądamy chociaż jednego żetonu.

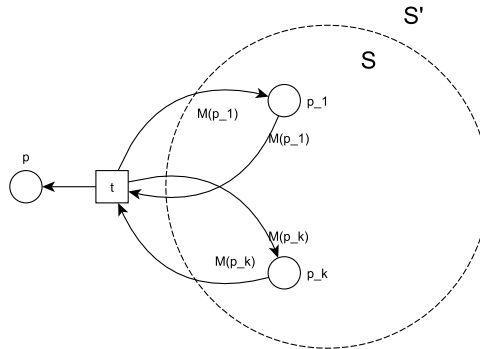
Można również pokazać, że umiejętność rozstrzygnięcia problemu niepustości miejsca jest wystarczająca przy rozstrzygnięciu pokrywalności markingu. W tym wypadku uzasadnienie jest nieco bardziej skomplikowane, posłużymy się następującą konstrukcją:

Konstrukcja 1.2.34 P3 rozstrzygalny \Rightarrow P1 rozstrzygalny:

Weźmy sieć Petriego $S = (P, T, W, M_0)$ i ustalmy marking $\bar{M} : P \rightarrow \mathbb{N}$. Na bazie sieci S skonstruujemy taką sieć $S' = (P', T', W', M'_0)$, że marking \bar{M} będzie pokrywalny w sieci S dokładnie wtedy, gdy jedno, z góry ustalone miejsce w sieci S' będzie niepuste.

Sieć S' powstaje z sieci S przez dodanie do zbioru miejsc P oraz zbioru akcji T po jednym nowym elemencie. Są to: miejsce p , którego niepustość będziemy testować oraz akcja t , która będzie wykonalna w momencie, gdy marking \bar{M} zostanie pokryty.

Aby uzyskać taki efekt, łączymy pętelką każde z miejsc $p_i \in P$ oryginalnej sieci S z akcją t , ustalając $W'(t, p_i) = W'(p_i, t) = \bar{M}(p_i)$. Miejsce p z kolei będzie jedynym wyjściem z akcji t . Osiągniemy to ustalając $W'(t, p) = 1$. Pozostałe wartości funkcji wagowej W' zostają ustalone zgodnie z wartościami funkcji W lub wynoszą 0. Ostatnim elementem sieci S' jest marking początkowy M'_0 . W miejscach należących do zbioru miejsc sieci S przyjmuje on wartości równe wartościom markingu M_0 , natomiast w miejscu p marking M'_0 przyjmuje wartość 0.



Rysunek 1.6: Konstrukcja sieci S' na podstawie sieci S . Marking M jest pokrywalny w sieci S dokładnie wtedy, gdy osiągalna jest niepustość miejsca p w sieci S'

Z każdym markingiem $M : P \rightarrow \mathbb{N}$ w sieci S , zwiążemy odpowiadający mu marking $0M : P' \rightarrow \mathbb{N}$ w sieci S' w następujący sposób:

$$0M(p) = \begin{cases} M(p) & \text{dla } p_i \in P \\ 0 & \text{dla } p. \end{cases}$$

Marking początkowy M'_0 jest wówczas tożsamy z markingiem $0M_0$.

Przejdźmy do uzasadnienia postawionej tezy. Na początek zauważmy, że każde obliczenie u wykonalne w sieci S jest również wykonalne w sieci S' . Podobnie, każde obliczenie v

nie zawierające akcji t i wykonalne w sieci S' jest również wykonalne w sieci S . Załóżmy, że marking \bar{M} jest pokrywalny w sieci S (czyli $\bar{M} \in [M_0 >)$). Wówczas istnieje obliczenie u wykonalne w sieci S takie, że $M_0 u M$ oraz $M \geq \bar{M}$. Oznacza to że w sieci S' osiągalny jest marking $0M \geq 0\bar{M}$ w którym wykonalna jest akcja t . Tak więc wykonalne jest obliczenie ut , którego wykonanie skutkuje wrzuceniem jednego żetonu do miejsca p , czyli miejsce p jest niepuste w sieci S' .

Z drugiej strony załóżmy, że miejsce p jest niepuste w sieci S' . Ponieważ w markingu początkowym nie ma w nim żadnego żetonu, musi istnieć obliczenie zawierające jedyną akcję mogącą zmienić stan miejsca p - akcję t . Rozważmy prefiks v tego obliczenia, taki że akcja t nie występuje w v oraz vt jest wykonalne. Mamy teraz wykonalne ciągi kroków $0M_0 v 0M$ oraz $0MtM'$. Obliczenie v , jako nie zawierające akcji t jest wykonalne również w sieci S i prowadzi do markingu M . Ponieważ akcja t jest wykonalna w markingu $0M$, to marking M pokrywa marking \bar{M} , co kończy uzasadnienie poprawności konstrukcji.

W Konstrukcji 1.2.34 uwagę zwraca szczególny związek akcji t oraz miejsca p . Wykorzystamy tą obserwację do uzasadnienia decyzyjnej równoważności problemu niepustości miejsca oraz wykonalności akcji:

Fakt 1.2.35 P3 rozstrzygalny \Rightarrow P2 rozstrzygalny:

Zauważmy, że problem niepustości miejsca p jest ciekawy tylko w sytuacji, gdy w markingu początkowym miejsce p jest puste. Załóżmy, że w jakiejś klasie sieci Petriego rozstrzygalny jest problem wykonalności akcji. Wówczas, dla miejsca p możemy określić czy jakakolwiek akcja, dla której miejsce p jest wyjściem jest wykonalna w tej sieci. Jako że miejsce p jest puste w markingu początkowym, jest to jedyna możliwość uzyskania markingu, w którym miejsce p byłoby niepuste.

W przypadku, gdy rozstrzygalna jest niepustość miejsca, postępujemy podobnie jak w Konstrukcji 1.2.34:

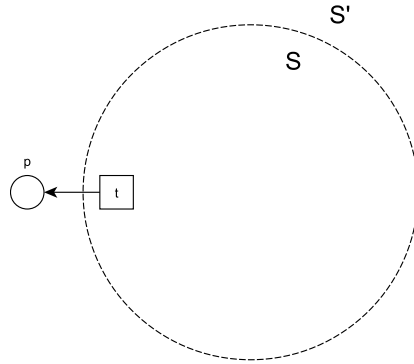
Konstrukcja 1.2.36 P2 rozstrzygalny \Rightarrow P3 rozstrzygalny:

Weźmy sieć Petriego $S = (P, T, W, M_0)$ i ustalmy akcję $t \in T$. Na bazie sieci S skonstruujemy sieć $S' = (P', T, W', M'_0)$ tak, aby akcja t była wykonalna w sieci S dokładnie wtedy, gdy jedno, z góry ustalone miejsce w sieci S' będzie niepuste.

Modyfikacja sieci S jest bardzo prosta. Dodajemy do niej jedno miejsce p . Jest ono wyjściem z akcji t , czyli $W'(t, p) = 1$ (pozostałe wartości funkcji wagowej W' są identyczne z wartościami W lub wynoszą zero). Poza tym, jest ono puste w markingu początkowym M'_0 .

Podobnie jak w Konstrukcji 1.2.34, miejsce p pozostaje puste, dopóki akcja t nie wykona się chociaż raz. Tak więc problem wykonalności akcji t w sieci S jest równoważny problemowi niepustości miejsca p w sieci S' .

Pokazaliśmy, że umiemy rozstrzygać o niepustości miejsca, zarówno dla p/t-sieci jak i dla sieci inhibitorowych, potrafimy też rozstrzygać o wykonalności akcji oraz pokrywalności markingu. Podobnie - mając narzędzia pozwalające rozstrzygnąć problem pokrywalności markingu albo wykonalności akcji potrafimy rozstrzygać niepustość miejsca. Zebrane razem, fakty te dają konstruktywną równoważność przedstawionych problemów. Identyczne rozumowanie można przeprowadzić w pozostałych klasach sieci Petriego. Problematyczny



Rysunek 1.7: Konstrukcja sieci S' na podstawie sieci S . Akcja t jest wykonalna w sieci S dokładnie wtedy, gdy osiągalna jest niepustość miejsca p w sieci S'

jest jednak Fakt 1.2.35 dla sieci samomodyfikujących (przedstawione rozumowanie nie działa w tym przypadku).

Przejdźmy teraz do drugiej grupy, czyli problemów O1, O2, O3 oraz O4. Na początek zajmiemy się problemami O1 oraz O4.

Fakt 1.2.37 O4 rozstrzygalny \Rightarrow O1 rozstrzygalny:

Załóżmy, że w jakiejś klasie sieci Petriego rozstrzygalny jest problem osiągalności pokrytego podmarkingu. Wówczas rozstrzygalny jest w niej, jako szczególny przypadek problemu O4 z $D = P$, problem osiągalności markingu.

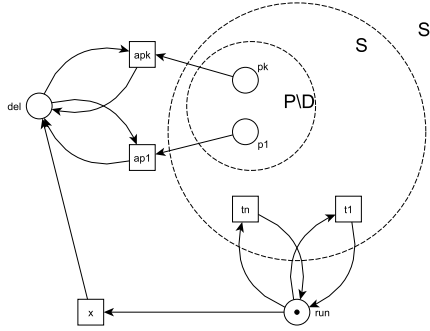
Więcej trudności sprawia uzasadnienie drugiej implikacji. Posłużymy się do tego następującą konstrukcją:

Konstrukcja 1.2.38 O1 rozstrzygalny \Rightarrow O4 rozstrzygalny:

Weźmy sieć Petriego $S = (P, T, W, M_0)$, marking $M : P \rightarrow \mathbb{N}$ oraz podzbiór miejsc $D \subset P$. Na bazie tej sieci skonstruujemy sieć $S' = (P', T', W', M'_0)$ tak, aby marking 01M odpowiadający markingowi M był osiągalny w sieci S' dokładnie wtedy, kiedy marking M jest pokrywalny w sieci S przez marking równy mu we wszystkich miejscach ze zbioru D .

Zbiór miejsc P' sieci S' to zbiór S rozszerzony o dwa miejsca - *run* oraz *del*. Natomiast zbiór akcji T' , to zbiór T rozszerzony o nowe akcje a_p , po jednej dla każdego miejsca $p \in P \setminus D$ oraz akcję x .

Miejsce *run* jest połączone pętelką z każdą akcją $t \in T$ oraz jest wejściem do akcji x . Miejsce *del* połączone jest pętelką z każdą akcją a_p oraz wyjściem z akcji x . Ponadto, każda akcja a_p połączona jest z odpowiadającym jej miejscem p pętelką o wadze wejściowej równej $M(p) + 1$ oraz wadze wyjściowej równej $M(p)$. Każdemu markingowi $M : P \rightarrow \mathbb{N}$ odpowiada marking 01M : $P' \rightarrow \mathbb{N}$, który jest mu równy we wszystkich miejscach ze zbioru P a ponadto $01M(\text{del}) = 1$ oraz $01M(\text{run}) = 0$. W końcu marking początkowy M'_0 jest równy markingowi M_0 dla wszystkich miejsc ze zbioru P . W dwóch nowych miejscach jego wartości są, zupełnie odwrotnie niż w markingu $01M_0$, równe odpowiednio 0 dla miejsca *del* oraz 1 dla miejsca *run*.



Rysunek 1.8: Konstrukcja sieci S' na podstawie sieci S . Na przedstawionym diagramie pominięto wagi krawędzie wejściowych i wyjściowych miejsca del . Marking M jest pokrywalny przez tożsamy z nim na zbiorze D marking M' w sieci S dokładnie wtedy, gdy marking $01M$ jest osiągalny w sieci S'

W tak skonstruowanej sieci S' akcja x może się wykonać co najwyżej raz i każde wykonalne obliczenie u możemy podzielić na dwie fazy - przed i po wykonaniu akcji x . Jeśli akcja x nie wykona się ani razu w czasie obliczenia u , to druga faza jest pusta. Faza pierwsza odpowiada za obliczenie wykonywane w obrębie pierwotnej sieci S . Miejsce run jest wówczas zamarkowane i każda akcja ze zbioru T jest wykonalna w markingu $01M$ w sieci S' o ile tylko jest wykonalna w markingu M w sieci S . Z drugiej strony, miejsce del jest wówczas puste i żadna akcja a_p nie jest umożliwiona. Po wykonaniu akcji x sytuacja się odwraca. Wszystkie akcje ze zbioru T są teraz uniemożliwione, natomiast potencjalnie wykonalne są akcje a_p . Pozwalają one zredukować liczbę żetonów w miejscach spoza zabronionego zbioru D , ale tak, aby nie zejść poniżej wartości markingu M w tych miejscach. Dzięki temu, jeśli marking M jest pokryty w sieci S przez osiągalny marking M' równy markingowi M we wszystkich miejscach ze zbioru D , to marking $01M$ jest osiągalny w sieci S' .

Z drugiej strony, jeśli marking $01M$ jest osiągalny w sieci S' , to wykonalne jest w niej obliczenie $u = u_1xu_2$ takie, że M'_0u01M oraz $x \notin Alph(u_1u_2)$. Zgodnie z wcześniejszą obserwacją $Alph(u_1) \subseteq T$ oraz $Alph(u_2) \cap T = \emptyset$, więc w sieci S wykonalne jest obliczenie u_1 , prowadzące do markingu M_1 . Po wykonaniu obliczenia u_1x otrzymujemy w sieci S' marking $01M_1$. W dalszej części obliczenia u usuwamy część żetonów z miejsc ze zbioru $P \setminus D$, co prowadzi do markingu $01M$. W związku z tym, marking $01M_1$ pokrywa marking $01M$ i jest z nim tożsamy na zbiorze D , czyli marking M_1 pokrywa marking M i jest z nim tożsamy na zbiorze D , co kończy dowód.

W dalszej kolejności uzasadnimy, że rozstrzygalność problemu osiągalności pociąga za sobą rozstrzygalność problemu pustości stanu.

Fakt 1.2.39 *O1 rozstrzygalny* \Rightarrow *O2 rozstrzygalny*:

Założmy, że w jakiejś klasie sieci Petriego rozstrzygalny jest problem osiągalności. Wówczas jest w niej również rozstrzygalny problem pustości stanu, jako szczególny przypadek problemu osiągalności - pytamy wówczas o osiągalność markingu, w którym każde z miejsc jest puste.

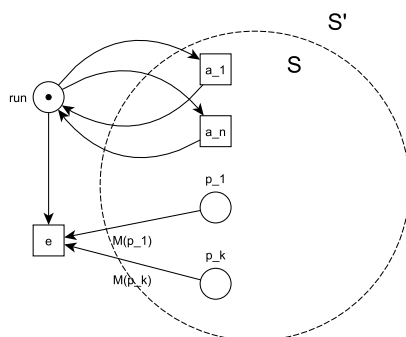
Nieco bardziej skomplikowane jest rozstrzygnięcie problemu osiągalności markingu w oparciu o odpowiedź na pytanie o osiągalność markingu pustego. Podobną do poniższej konstrukcję można znaleźć w [47].

Konstrukcja 1.2.40 O2 rozstrzygalny \Rightarrow O1 rozstrzygalny:

Weźmy sieć Petriego $S = (P, T, W, M_0)$, gdzie P oznacza zbiór miejsc, zaś T - zbiór akcji w tej sieci. Ustalmy marking $M : P \rightarrow \mathbb{N}$. Na bazie sieci S skonstruujemy taką sieć $S' = (P', T', W', M'_0)$, że marking M będzie osiągalny w sieci S dokładnie wtedy, gdy w sieci S' będzie osiągalny marking pusty.

W konstrukcji skorzystamy z opisaną wcześniej idei miejsca sterującego siecią. Pozwoli to w dowolnym momencie zatrzymać działanie sieci i wykonać pewne operacje na uchwyconym w ten sposób markingu.

Sieć S' powstaje z sieci S poprzez dodanie globalnego miejsca run zawierającego jeden żeton w markingu początkowym oraz globalnej akcji e . Miejsce run jest połączone pętelką z każdą akcją $a \in T$. Ponadto każde miejsce $p \in P$, które jest niepuste w markingu M , jest wejściem do akcji e . Wagi krawędzi łączących miejsca oryginalnej sieci S z akcją e są zdeterminowane przez marking M , którego osiągalność chcielibyśmy zbadać. W końcu, miejsce run jest również wejściem do akcji e .



Rysunek 1.9: Konstrukcja sieci S' na podstawie sieci S . Marking M jest osiągalny w sieci S dokładnie wtedy, gdy marking pusty jest osiągalny w sieci S'

Zauważmy, że zanim nie zostanie wykonana akcja e , sieć S' działa dokładnie tak samo, jak sieć S . Po wykonaniu akcji e osiągamy marking martwy; każda akcja wymaga obecności żetonu w miejscu run , które jest czyszczone przez akcję e . W związku z tym, każdemu markingowi M osiągalnemu w sieci S odpowiada osiągalny w sieci S' marking $1M$, gdzie $1M(p) = M(p)$ dla $p \in P$ oraz $1M(run) = 1$. Ponadto, jeśli w markingu M umożliwiona jest akcja $a \in T$, to jest ona umożliwiona również w markingu $1M$. Marking początkowy w sieci S' to $M'_0 = 1M_0$. Poza tym, w sieci S' osiągalne są jeszcze dodatkowe markinki z pustym miejscem run .

Załóżmy, że marking M jest osiągalny w sieci S . Wówczas istnieje obliczenie u w sieci S takie, że $M_0 u M$. Zgodnie z powyższą obserwacją, obliczenie u jest wykonalne również w sieci S' i $1M_0 u 1M$. W markingu $1M$ akcja e jest wykonalna, a jej wykonanie prowadzi do markingu pustego.

Z drugiej strony, przy założeniu, że marking pusty jest osiągalny z sieci S' , musiał on

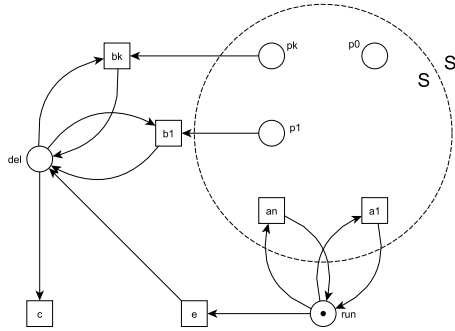
zostać osiągnięty przy obliczeniu, którego ostatnią akcją jest e . Oznacza to, że istnieje takie obliczenie v' w sieci S' , że $M'_0 v' \circledast$ oraz $v' = ve$. Akcja e może się wykonać w sieci S' co najwyżej raz, w związku z czym obliczenie v jest wykonalne w sieci S . Pozostaje jeszcze zauważyć, że $M_0 v M$.

Jako uzupełnienie równoważności w rozważanej grupie pozostało jeszcze wykazanie równoważności rozstrzygalności problemu pustości miejsca z jednym z trzech wyżej wymienionych problemów. Pokażemy, że problem O3 jest równoważny problemowi osiągalności stanu pustego. Zaczniemy od konstrukcji, która pozwala przetestować pustość miejsca, pod warunkiem, że umiemy testować pustość stanu.

Konstrukcja 1.2.41 O2 rozstrzygalny \Rightarrow O3 rozstrzygalny:

W konstrukcji tej ponownie skorzystamy z idei miejsca *run*, sterującego działaniem sieci. Podobnie jak wcześniej, pozwoli on uchwycić w dowolnym momencie stan, w jakim znalazła się sieć i wykonać pewne testy. Będzie to próba wyczyszczenia sieci S bez modyfikowania stanu miejsca p_0 .

Sieć S' powstaje z sieci S przed dodaniem do niej dwóch nowych miejsc - *run* i *del*, oraz $|P| + 1$ nowych akcji. W szczególności, dla każdego miejsca $p_i \in P$ sieci S , różnego od miejsca p_0 , dodajemy akcję b_i , dla której p_i jest jednym z wejść. Akcje te pozwalają wyczyścić miejsca sieci S . Ponadto, miejsce *del* jest jednocześnie wejściem i wyjściem z każdej akcji b_i . Dwie pozostałe akcje to akcja e - wyłączająca sieć S i rozpoczynająca procedurę usuwania oraz akcja c - usuwająca żeton z miejsca *del*. Tak więc miejsce *del* jest wejściem do akcji c oraz wyjściem z akcji e . Pozostałe miejsce *run* jest połączone pętelką z każdą akcją $t \in T$ oryginalnej sieci S oraz jest wejściem do akcji e . Marking M'_0 ustalamy na równy markingowi M_0 dla wszystkich miejsc $p \in P$ oraz $M'_0(\text{run}) = 1$ i $M'_0(\text{del}) = 0$.



Rysunek 1.10: Konstrukcja sieci S' na podstawie sieci S . Marking w którym miejsce p_0 jest puste jest osiągalny w sieci S dokładnie wtedy, gdy marking pusty jest osiągalny w sieci S'

Zauważmy, że zanim zostanie wykonana akcja e , sieć S' działa dokładnie tak samo jak sieć S . Akcja ta może wykonać się tylko raz, po tym fakcie wszystkie akcje $t \in T$ oraz akcja e będą już stale uniemożliwione, potencjalne umożliwienie zyskują natomiast wszystkie nowe akcje b_i oraz akcja c . Teraz z kolei, wykonanie akcji c prowadzi do markingu martwego. Tak więc każde obliczenie w sieci S' jest prefiksem obliczenia postaci $v_1 e v_2 c$, gdzie $v_1 \in T^*$ zaś $v_2 \in ((T' \setminus T) \setminus \{c, e\})^*$.

Dla uproszczenia zapisu, z każdym markingiem $M : P \rightarrow \mathbb{N}$ skojarzymy markingi

$M_z, M_r, M_d \in \mathbb{N}^{P'}$ w taki sposób, że $\forall_{p \in P} M(p) = M_z(p) = M_r(p) = M_d(p)$ oraz $M_r(run) = M_d(del) = 1$ i $M_z(del) = M_z(run) = M_r(del) = M_d(run) = 0$. Wówczas mamy, że ciąg kroków $M'_0 v_1 M_r e M_d v_2 M'_d c M'_z$ jest wykonalny, przy czym M' jest markingiem osiągalnym w sieci S .

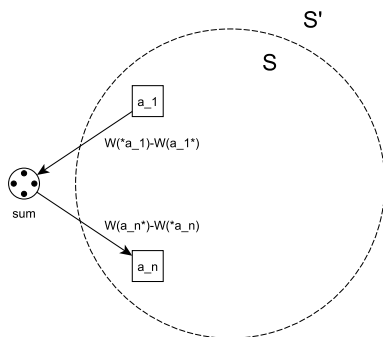
Założmy, że marking M , w którym miejsce p_0 jest puste jest osiągalny w w sieci S . Wówczas, na mocy poczynionej uwagi, w sieci S' osiągalne są stany M_r oraz M_d . W stanie M_d można wykonać akcje b_i tak długo, aż wszystkie miejsca p_i (z wyjątkiem już pustego miejsca p_0) zostaną wyczyszczone. Doprowadzi to markingu M'_d . Po wykonaniu akcji c osiągamy stan pusty, czyli marking M'_z .

Z drugiej strony, założmy że marking pusty jest osiągalny w sieci S' . Oznacza to, że musiała zostać wykonana akcja e . Jest ona jedyną akcją usuwającą żeton z miejsca run . Założmy, że w odpowiadającym mu markingu M , który jest osiągalny w sieci S , miejsce p_0 było niepuste. Wówczas po wykonaniu pozostałych akcji prowadzących do stanu pustego w sieci S' miejsce to pozostawałoby niepuste, co daje oczywistą sprzeczność. Tak więc miejsce p_0 jest puste w markingu M osiągalnym w sieci S , co kończy dowód.

Pozostała nam do przeprowadzenia konstrukcja pokazująca, że umiemy rozstrzygać o pustości miejsca, potrafimy rozstrzygać o pustości stanu.

Konstrukcja 1.2.42 O3 rozstrzygalny \Rightarrow O2 rozstrzygalny:

Tym razem idea i sama konstrukcja są dużo prostsze. Do sieci S dodamy jedno miejsce sum liczące sumę żetonów w całej sieci S . Miejsce to będzie wyjściem z każdej akcji a , która pobiera z sieci S mniej żetonów niż do niej zwraca. Waga tak powstałego połączenia będzie równa tej różnicy ($W'(a, sum) = \sum_{p \in P} (W(a, p) - W(p, a))$). Podobnie, miejsce sum będzie wejściem do każdej akcji a , która pobiera z oryginalnej sieci więcej żetonów niż do niej zwraca. Waga tak powstałego połączenia wynosi $W'(sum, a) = \sum_{p \in P} (W(p, a) - W(a, p))$. Ponadto, w markingu początkowym w miejscu sum znajduje się tyle żetonów ile we wszystkich pozostałych miejscach razem wziętych ($M'_0(sum) = \sum_{p \in P} M_0(p)$).



Rysunek 1.11: Konstrukcja sieci S' na podstawie sieci S . Marking pusty jest osiągalny w sieci S dokładnie wtedy, gdy marking, w którym miejsce sum jest puste jest osiągalny w sieci S'

Nietrudno zauważyć, że nowe miejsce nie wpływa na obliczenia umożliwiające w sieci S , a ponadto w każdym osiągalnym markingu w miejscu sum znajduje się dokładnie połowa wszystkich znajdujących się w sieci S' żetonów. Uzasadnia to, że stan, w którym miejsce

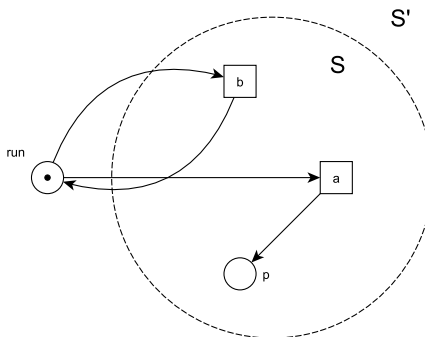
sum jest puste osiągalny jest w sieci S' dokładnie wtedy, gdy osiągalny jest w niej (a co za tym idzie również w sieci S) stan pusty.

Przedstawiona powyżej seria faktów i konstrukcji pokazują, że w obrębie zaproponowanych grup, rozstrzygalność problemów związanych z markiem jest konstruktywnie równoważna. Warto zauważyć, że Konstrukcja 1.2.42 nie działa dla sieci czyszczących i podwajających (choć uzasadniana implikacja zachodzi w tych klasach sieci), zaś w przypadku sieci samomodyfikujących wymaga drobnych modyfikacji. Pokażemy teraz, że nieco słabszy wynik dla rozważanych klas sieci Petriego możemy uzyskać dla dowolnej pary problemów pochodzących z różnych grup. Mianowicie, rozstrzygalność dowolnego problemu związanego z osiągalnością pociąga za sobą rozstrzygalność dowolnego problemu związanego z pokrywalnością. Zależność tę wykażemy przez konstrukcję pozwalającą rozstrzygnąć problem niepustości miejsca z wykorzystaniem rozstrzygalności problemu pustości miejsca.

Konstrukcja 1.2.43 O3 rozstrzygalny \Rightarrow P3 rozstrzygalny:

Problem niepustości z góry określonego miejsca $p \in P$ sieci $S = (P, T, W, M_0)$ jest nietrywialny tylko w przypadku, gdy $M_0(p) = 0$. Możemy wówczas skonstruować sieć $S' = (P', T', W', M'_0)$, która do momentu wrzucenia pierwszych żetonów do miejsca p będzie działała identycznie jak sieć S . Po wykonaniu pierwszej akcji, wrzucającej żeton do miejsca p , sieć S' osiągnie marking martwy.

Sieć S' powstaje z sieci S przed dodaniem do zbioru miejsc nieco zmodyfikowanego miejsca run . Jest ono połączone pętelkami z tymi akcjami $a \in T$ sieci S , które nie powodują zwiększenia liczby żetonów w miejscu p ($W(a, p) = 0$). Dla pozostałych akcji, czyli akcji wypełniających miejsce p , miejsce run jest tylko wejściem.



Rysunek 1.12: Konstrukcja sieci S' na podstawie sieci S . Marking, w którym miejsce p jest niepuste jest osiągalny w sieci S dokładnie wtedy, gdy marking, w którym miejsce run jest puste jest osiągalny w sieci S'

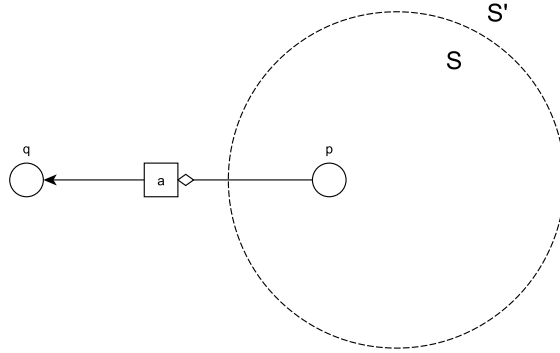
Konstrukcja ta jest poprawna dla wielu klas sieci, będących rozszerzeniami p/t-sieci, gdyż wszystkie dołączone łuki są zwykłymi łukami wejściowymi bądź wyjściowymi. Ze zdefiniowanych w tej rozprawie problem sprawiają, podobnie jak w przypadku Faktu 1.2.35, tylko sieci samomodyfikujące.

Pokazanie podobnej zależności w drugą stronę dla p/t-sieci nie jest znane. Uprościłoby ono znacznie dowód rozstrzygalności problemu osiągalności markingu w tej klasie sieci. Jest to natomiast dość proste w sieciach inhibitorowych:

Konstrukcja 1.2.44 P3 \Rightarrow O3 (dla sieci inhibitorowych):

Rozważmy sieć inhibitorową $S = (P, T, W, I, M_0)$ z wyróżnionym miejscem p . Skonstruujemy sieć inhibitorową $S' = (P', T', W', I', M'_0)$ ze wskazanym miejscem $q \in P'$ tak, aby osiągalność niepustości miejsca q w sieci S' była równoważna osiągalności pustości miejsca p w sieci S .

Sieć S' powstaje z sieci S przez dodanie nowego miejsca q , nowej akcji a oraz dwóch łuków: inhibitorowego, z p do a i zwykłego z a do q . Markingiem początkowy M'_0 w sieci S' jest identyczny markingowi początkowemu M_0 na wszystkich miejscach ze zbioru P , zaś $M'_0(q) = 0$.



Rysunek 1.13: Konstrukcja sieci S' na podstawie sieci S . Marking, w którym miejsce p jest puste jest osiągalny w sieci S dokładnie wtedy, gdy marking, w którym miejsce q jest niepuste jest osiągalny w sieci S'

Poprawność dwóch ostatnich konstrukcji jest dość prosta i pominięto jej szczegółowe uzasadnienie.

Rozstrzygalność. Po omówieniu zależności decyzyjnych między różnymi problemami związanymi z markingiem warto zastanowić się nad ich rozstrzygalnością. Jednym z głównych narzędzi ułatwiających analizę zbiorów markingów osiągalnych sieci Petriego są grafy osiągalności (zwanych też grafami przypadków [44] lub grafami konfiguracji [47]).

Definicja 1.2.45 Niech S będzie siecią Petriego o zbiorze wejść P , zbiorze akcji T oraz markingu początkowym M_0 . Grafem osiągalności nazwiemy multigraf skierowany $RG(S) = ([M_0 >, E, e)$, gdzie $[M_0 >$ jest zbiorem wierzchołków tego grafu odpowiadających wszystkim markingom osiągalnym w tej sieci, E zbiorem krawędzi, zaś $e \subseteq E \times T$ relacją etykietującą krawędzie. Krawędź (M_1, M_2) należy do zbioru E dokładnie wtedy, gdy istnieje akcja $a \in T$ taka, że (M_1, a, M_2) jest krokiem. Ponadto krawędź ta jest wtedy w relacji e z akcją a .

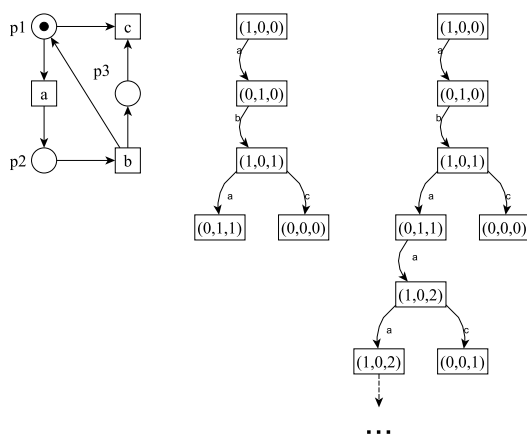
Przykład 1.2.46 Rozważmy następującą sieć elementarną $S = (P, T, F, M_0)$, gdzie $P = \{p_1, p_2, p_3\}$ zaś $T = \{a, b, c\}$. Miejsce p_1 jest wejściem do akcji a i c oraz wyjściem z akcji

b , miejsce p_2 - wejściem do akcji b oraz wyjściem z a , natomiast p_3 jest wyjściem z akcji b i wejściem do akcji c . Markingowi początkowemu odpowiada wektor $(1, 0, 0)$, czyli jedyny żeton znajduje się w miejscu p_1 .

Analiza działania tej sieci jest dość prosta - jedyną akcją umożliwiającą w markingu początkowym jest akcja a , prowadzi to do markingu $(0, 1, 0)$, w którym możemy wykonać tylko akcję b , co prowadzi do markingu $(0, 1, 1)$. Teraz z kolei mamy możliwość wykonania zarówno akcji a jak i c . Obie te akcje prowadzą do markingów martwych (w których żadna akcja nie jest umożliwiona), $(0, 1, 1)$ oraz $(0, 0, 0)$ odpowiednio. Sytuację tą obrazuje znajdujący się poniżej graf osiągalności dla tej sieci.

Dla odmiany, potraktujmy teraz sieć S jako p/t-sieć $S' = (P, T, W, M_0)$. Funkcję wagową określimy zgodnie z Rysunkiem 1.14 (tak, aby odpowiadająca jej relacja przepływu pokrywała się z F , oraz sieć S' była zwykłą): $W(x, y) = 1 \Leftrightarrow (x, y) \in F$, w przeciwnym przypadku $W(x, y) = 0$.

Wszystkie obliczenia wykonalne w sieci elementarnej S są też wykonalne w p/t-sieci S' . Różnica pojawia się przy markingu opisanym wektorem $(0, 1, 1)$. W sieci S jest to marking martwy, natomiast w sieci S' umożliwiona jest akcja b . Jej wykonanie prowadzi do markingu, któremu odpowiada wektor $(1, 0, 2)$. Dalej można na zmianę wykonywać akcje a oraz b zwiększając liczbę żetonów w miejscu p_3 w sposób nieograniczony lub w pewnym momencie wykonać akcję c , która zawsze prowadzi do martwego markingu. W związku z tym graf osiągalności sieci S' jest nieskończony.



Rysunek 1.14: Sieci S oraz S' oraz ich grafy osiągalności, w środku, gdy sieć traktujemy jako sieć elementarną (S), po prawej jako p/t-sieć (S')

Uwaga 1.2.47 Trywialność problemów związanych z markingiem w sieciach elementarnych.

Zauważmy, że w sieciach elementarnych zbiory wszystkich markingów osiągalnych są skończone (nie większe niż $2^{|P|}$). W związku z tym, grafy osiągalności dla sieci z tej klasy są również skończone, a co za tym idzie - wszystkie zdefiniowane wcześniej problemy związane z markingiem są w sposób trywialny rozstrzygalne.

Wygodnie jest też wprowadzić, często nieskończone, ukorzenione drzewa osiągalności z krawędziami etykietowanymi akcjami sieci. W drzewach takich wiele wierzchołków może być etykietowanych tym samym markingiem. Korzeń etykietowany jest markingiem początkowym, a każda gałąź drzewa odpowiada obliczeniu wykonalnemu w danej sieci. Graf

osiągalności powstaje wówczas ze sklejenia w jeden wierzchołek wszystkich wierzchołków o takiej samej etykiecie. Grafy przedstawione na Rysunku 1.14 są drzewami, każdy marking $M \in [M_0 >$ możemy osiągnąć przy pomocy unikatowego obliczenia. W związku z tym, nie występuje żadne sklejenie.

W przypadku klasycznych p/t-sieci nazwy drzewa osiągalności [26, 38] używa się czasem zamiennie z drzewami pokrywalności [39]. Wyraźnie rozróżnimy te dwa pojęcia. Drzewa pokrywalności znajdują zastosowanie w rozstrzyganiu problemu pokrywalności (Lemat 1.2.51), choć Mayr próbował użyć ich do rozstrzygania problemu osiągalności [32]. Wraz z nimi często pojawia się graf pokrywalności [44], związany z drzewem w podobny sposób, jak graf osiągalności związany jest z drzewem osiągalności. Jednak w przeciwieństwie do tamtej sytuacji, zarówno drzewo jak i graf pokrywalności są zawsze skończone (co formalnie udowodnimy jako Lemat 1.2.52). Aby wprowadzić definicję drzewa pokrywalności, potrzebujemy dodać do zbioru liczb naturalnych nowy element, który będziemy traktować jak nieskończoność, oraz rozszerzyć pojęcie markingu o możliwość zapełniania miejsc nieskończonymi liczbami żetonów:

Definicja 1.2.48 Dodajmy do zbioru liczb naturalnych dodatkowy element ω tak, aby był to element największy w nowopowstałym zbiorze $\mathbb{N} \cup \{\omega\}$. W wyniku dodawania i mnożenia przez element ω otrzymujemy ω (dodawanie i mnożenie pozostają symetryczne). Ponadto wynikiem odejmowania, gdzie odjemnikiem jest ω jest 0, zaś w przypadku, gdy ω jest odjemną, a odjemnik pochodzi ze zbioru \mathbb{N} - wynikiem jest ω . Bardziej formalnie:

Niech $n \in \mathbb{N}$ zaś $x \in \mathbb{N} \cup \{\omega\}$,

$$\begin{aligned} x + \omega &= \omega + x = \omega \\ x * \omega &= \omega * x = \omega \\ x - \omega &= 0 \\ \omega - n &= \omega \end{aligned}$$

Definicja 1.2.49 Przez marking uogólniony p/t-sieci $S = (P, T, W, M_0)$ rozumiemy funkcję $\bar{M} : P \rightarrow (\mathbb{N} \cup \{\omega\})$. Co prawda, markinki uogólnione w których choć na jednej współrzędnej występuje ω nie są osiągalne przez żadne skończone obliczenie w sieci, to jednak są bardzo wygodne w opisie markinków pokrywalnych oraz obliczeń nieskończonych.

Umożliwienie, wykonalność oraz wykonanie akcji umożliwionej w markingu uogólnionym będziemy rozumieć zgodnie z Definicją 1.2.5

Drzewa pokrywalności definiowane są za pomocą (niestety w niektórych wersjach niedeterministycznego [38]) algorytmu zaproponowanego przez Karpa i Millera [26]. W myśl tego algorytmu, budowę drzewa rozpoczynamy od korzenia, który jest etykietowany markingiem początkowym sieci. Dalej postępujemy jakbyśmy mieli zbudować drzewo osiągalności. Różnica pojawia się, gdy któryś z wierzchołków znajdujących się na tej samej gałęzi co nowo dodawany, jest etykietowany równym lub mniejszym markingiem. Niech M oznacza etykietę nowo dodawanego wierzchołka, zaś M' - znalezionej równego bądź mniejszego. W pierwszym przypadku ($M = M'$) zaprzestajemy dalszego rozbudowywania tej krawędzi, w drugim dodajemy nowy wierzchołek, lecz etykietujemy go odpowiednim markingiem uogólnionym \bar{M} zawierającym symbole ω zgodnie z zasadą:

$$\bar{M}(p) = \begin{cases} M(p) & \text{jeśli } M(p) = M'(p) \\ \omega & \text{jeśli } M(p) < M'(p) \end{cases}$$

Możliwość pojawienia się w tym momencie kilku różnych mniejszych markingów jest odpowiadająca za niedeterminizm tego algorytmu. Niedeterminizm ten nie występował w oryginalnym sformułowaniu Karpa i Millera, gdzie nowy uogólniony marking \bar{M} powstaje w opisany sposób nie z pojedynczego, większego markingu M' , ale z markingu M'' o współrzędnych $M''(p) = \max\{M'(p) \mid M' \leq M\}$. Różnicę, wynikającą z tej kosmetycznej na pozór zmiany, widać w Przykładzie 1.2.53.

Definicja 1.2.50 Niech $S = (P, T, W, M_0)$ będzie p/t-siecią. Przez drzewo pokrywalności rozumiemy graf $CT(S) = (V, E, v, e)$, gdzie $V \subseteq (\mathbb{N} \cup \{\omega\})^P$, zaś $E \subseteq V \times V$, $v : V \rightarrow P$ oraz $e : E \rightarrow T$ są funkcjami etykietującymi odpowiednio wierzchołki oraz krawędzie, który powstaje zgodnie z opisaną powyżej procedurą.

Grafem osiągalności nazywamy graf $CG(S) = (V, E, e)$, gdzie $V \subseteq (\mathbb{N} \cup \{\omega\})^P$, $E \subseteq V \times V$, zaś $e : E \rightarrow T$ jest funkcją etykietującą krawędzie, który powstaje z drzewa $CT(S)$ przez sklejenie wierzchołków o jednakowych etykietach.

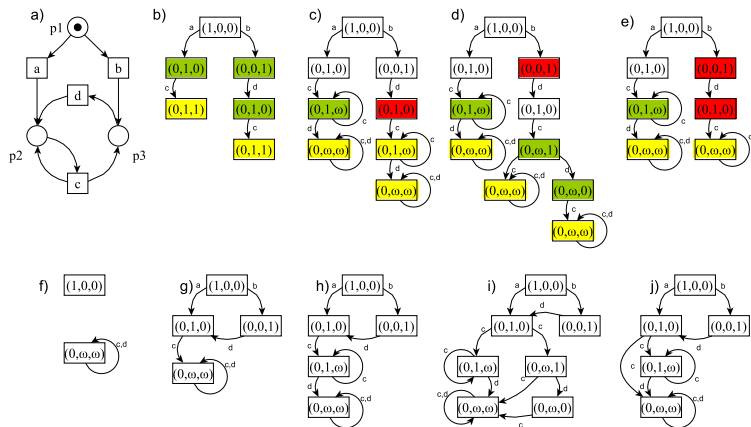
Lemat 1.2.51 Niech $S = (P, T, W, M_0)$ będzie p/t-siecią, zaś $CT(S) = (V, E, v, e)$ jej drzewem pokrywalności. Wówczas dla markingu $M : \rightarrow \mathbb{N}$ istnieje pewien marking $M' \in [M_0 >$, taki że $M \leq M'$ dokładnie wtedy, gdy w drzewie $CT(S)$ istnieje wierzchołek $w \in V$ etykietowany uogólnionym markingiem $\bar{M} = v(w)$, taki że $M \leq \bar{M}$.

Dowód. Skojarzmy markingi osiągalne w sieci S z węzłami drzewa $CT(S)$ za pomocą prowadzących do nich obliczeń. Jeśli marking M można osiągnąć z markingu początkowego przez wykonanie obliczenia u , czyli $M_0 u M$, to skojarzmy go z wierzchołkiem w drzewie $CT(S)$, do którego dojdziemy od korzenia poruszając się łukami etykietowanymi kolejnymi akcjami obliczenia u . Oczywiście może się zdarzyć, że z jednym markingiem osiągalnym skojarzmy wiele wierzchołków, jednak z każdym skojarzmy chociaż jeden. Zauważmy, że jeśli marking skojarzony jest z wierzchołkiem, to jest od niego mniejszy bądź równy (bardzo łatwo uzasadnić to indukcyjnie, indukcja po długości kojarzącego obliczenia). W sposób bezpośredni daje to tezę. \square

Lemat 1.2.52 Dla drzewa pokrywalności $CT = (V, E, v, e)$ zbiór V jest skończony.

Dowód. Załóżmy, że drzewo CT jest nieskończone. Ponieważ każdy wierzchołek ma liczbę dzieci ograniczoną rozmiarem zbioru akcji T , to na mocy lematu Königa musi istnieć ścieżka nieskończona. Zauważmy, że gdyby znajdowało się na niej nieskończenie wiele wierzchołków, to musiałyby istnieć para porównywalnych wierzchołków, z których wcześniejszy byłby niemniejszy od późniejszego. Zatem byłyby równe i procedura rozbudowy drzewa zakończyła by się w tym momencie co przeczy nieskończoności gałęzi lub pojawiłaby się przynajmniej jedna współrzędna ω . W drugim przypadku, analogiczny argument można zastosować dla całej czas nieskończonej gałęzi rozpoczynającej się od wierzchołka zawierającego ową omegę. W związku z tym, ponownie, procedura musiałaby się zakończyć lub pojawiłaby się ω na innej współrzędnej. Ponieważ współrzędnych jest tyle co miejsc w wyjściowej sieci, ostatecznie gałąź ta nie mogłaby być nieskończona, co dowodzi, że drzewo pokrywalności każdej sieci jest skończone. \square

Przykład 1.2.53 Rozważmy, pochodzącą z [44], p/t-sieć $S = (P, T, W, M_0)$ przedstawioną na Rysunku 1.15(a).



Rysunek 1.15: Sieć S (a), miejsca w którym pojawiają się różnice w procedurze generowania drzewa pokrywalności w porównaniu do procedury generowania drzewa osiągalności (b), różne możliwe drzewa pokrywalności (c-e) oraz grafy pokrywalności (f-j) tej sieci

Na Rysunku 1.15(b) przedstawiona została sytuacja, w której procedura budowy drzewa pokrywalności różni się od procedury budowy drzewa osiągalności. Na obu gałęziach budowanego drzewa pojawiają się markingi (oznaczone kolorem żółtym) mniejsze od markingów (oznaczonych kolorem zielonym) znajdujących się na zbudowanej części gałęzi. W przypadku gałęzi lewej, występuje tylko jeden taki marking. Wtedy wszystkie wersje procedury Karpa-Millera działają identycznie. Natomiast w przypadku gałęzi prawej pojawiają się dwa mniejsze markingi, które są nieporównywalne.

Zastosowanie w tym przypadku procedury opisanej przez Muratę prowadzi do drzewa 1.15(c), jeśli weźmiemy pod uwagę oznaczony na czerwono marking $(0,1,0)$, lub 1.15(d) dla markingu $(0,0,1)$. Oryginalna procedura biorąca pod uwagę oba markingi pozwala zbudować drzewo 1.15(e). W przypadku wszystkich tych drzew zaadaptowaliśmy zaproponowaną przez Muratę konwencję oznaczania drugiego wystąpienia na jednej ścieżce takiego samego markingu przez krawędzie powrotne odwołujące się do wystąpienia poprzedniego (w tych drzewach są to wyłącznie pętliki), która zdecydowanie skraca i upraszcza zapis.

Na kolejnych Rysunkach (1.15(f-j)) pojawiły się możliwe grafy pokrywalności tej sieci. Są to minimalny graf pokrywalności Finkela (f), możliwe efekty zastosowania procedury budowy grafu pokrywalności opisanej przez Reisiga w [44] (g-h) oraz grafy powstałe przez sklejenie identycznych wierzchołków drzew pokrywalności (c-e), czyli odpowiednio grafy (h-j). Warto w tym miejscu zauważyć, że jeden z możliwych do uzyskania grafów pokrywalności Reisiga (h) jest identyczny z tym powstałym w wyniku sklejenia identycznych wierzchołków w drzewie Muraty (j).

Uwaga 1.2.54 Przykład 1.2.53 pokazuje, że utworzone, zgodnie z procedurami podobnymi do tej zaproponowanej przez Karpa i Millera, drzewa oraz grafy pokrywalności nie muszą być wyznaczone jednoznacznie.

Uwaga 1.2.55 *Bazując na wprowadzonym przez Finkela [19], a rozważanym wcześniej przez Mayra [32] minimalnym pokryciu zbioru markingów osiągalnych można zdefiniować jednoznaczny, minimalny graf pokrywalności (graf f). Nie musi on jednak być możliwy do wyznaczenia przez opisaną procedurę tworzenia grafu pokrywalności (a więc nie musi być grafem pokrywalności w sensie definicji Karpa i Millera).*

Wniosek 1.2.56 *Rozstrzygalność problemu pokrywalności w p/t -sieciach.*

Niech $S = (P, T, W, M_0)$ będzie p/t -siecią, zaś $M : P \rightarrow \mathbb{N}$ markingiem. Problem pokrywalności markingu M w sieci S jest rozstrzygalny.

Powołując się na Lematy 1.2.51 oraz 1.2.52, pokrywalność danego markingu w p/t -sieci jesteśmy w stanie rozstrzygnąć na podstawie skończonego drzewa (bądź grafu) pokrywalności zbudowanego zgodnie z jedną z przedstawionych lub zacytowanych procedur. W związku z tym nasuwa się sygnalizowany wcześniej

Wniosek 1.2.57 *Problem Wykonalności Akcji jest rozstrzygalny, czyli dla danej p/t -sieci można wykryć i usunąć wszystkie akcje bezużyteczne.*

Poza tym, rozstrzygalne są również pozostałe problemy z grupy $P1 - P3$, jak również związana z tymi problemami ograniczoność sieci. Mając do dyspozycji graf pokrywalności, wystarczy poszukać w nim węzłów zawierających współrzędne równe ω . Inne natychmiastowe zastosowanie to sprawdzenie żywości akcji w danym markingu, które tłumaczy się wprost na odpowiednią instancję Problemu Pokrywalności.

Na początku lat osiemdziesiątych dwudziestego wieku Mayr [32] oraz Kosaraju [28] pokazali, że rozstrzygalny jest również problem osiągalności.

Fakt 1.2.58 *Niech $S = (P, T, W, M_0)$ będzie p/t -siecią, zaś $M : P \rightarrow \mathbb{N}$ markingiem. Problem osiągalności markingu M w sieci S jest rozstrzygalny.*

Dowód oparty o arytmetykę Presburgera jest dość zmundny, a złożoność obliczeniowa powstałej w jego wyniku procedury duża. Sam fakt pociąga za sobą wiele ciekawych wniosków, podamy jeden z nich, istotny z punktu widzenia dyskusji na temat usuwania z sieci zbędnych elementów:

Wniosek 1.2.59 *Problem Osiągalności Pokrytego Podmarkingu, a co za tym idzie problem użyteczności miejsca, jest rozstrzygalny. Znając marking początkowy, możemy wykorzystując odpowiednie instancje przytoczonego problemu sprawdzić, czy liczba żetonów w tym miejscu może ulec zmianie. Mając daną p/t -sieć bez akcji bezużytecznych można wykryć i usunąć miejsca bezużyteczne.*

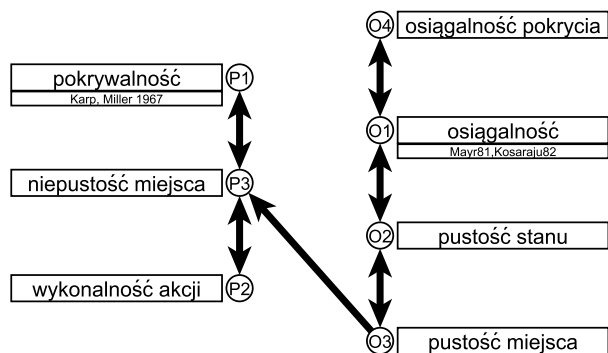
Poza tym, rozstrzygalne są również pozostałe problemy z grupy $O1 - O4$, jak również związana z nimi żywość sieci (implikująca odporność na zakleszczenia), która wymaga przetestowania osiągalności skończonej liczby potencjalnie problematycznych markingów (to czy są one faktycznie problematyczne, czyli czy jakaś akcja jest w nich martwa, można rozstrzygnąć wykorzystując odpowiednią instancję Problemu Pokrywalności).

Przejdźmy teraz do sieci inhibitorowych. Nierozstrzygalność problemu osiągalności w sieciach inhibitorowych pokazał Agerwala, w pracy w której wprowadził pojęcie takich sieci

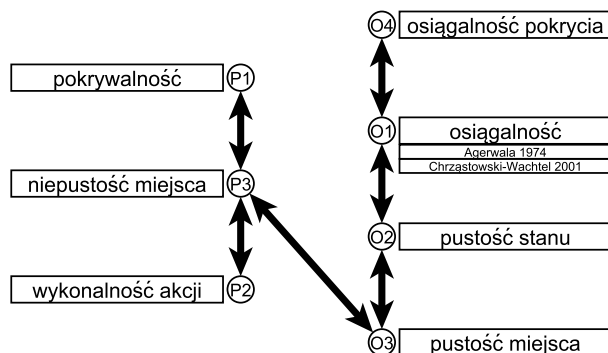
[2]. Wykorzystał w tym celu ich równoważność obliczeniową z maszynami licznikowymi, oraz, pokazaną przez Minsky'ego, równoważność tych ostatnich z maszynami Turinga. Fakt można też pokazać inaczej [10], wykorzystując nierozstrzygalność 10 problemu Hilberta (Matijasevich [31]).

Fakt 1.2.60 Niech $S = (P, T, W, I, M_0)$ będzie siecią inhibitorową, zaś $M : P \rightarrow \mathbb{N}$ markingiem. Problem osiągalności markingu M w sieci S jest nierozstrzygalny.

Ostatecznie więc, wszystkie problemy związane z markingiem są rozstrzygalne w p/t-sieciach (odpowiednie zależności pokazane w tej sekcji zebrane są na Rysunku 1.16). Natomiast w sieciach inhibitorowych problemy te są nierozstrzygalne i konstruktywnie równoważne (Rysunek 1.17).



Rysunek 1.16: Struktura problemów związanych z markingiem w p/t-sieciach, z wyszczególnieniem dowodów rozstrzygalności pokrywalności oraz osiągalności. Krawędź skierowana z wierzchołka odpowiadającego problemowi A do wierzchołka odpowiadającego problemowi B oznacza, że rozstrzygalność problemu A implikuje konstruktywnie rozstrzygalność problemu B



Rysunek 1.17: Struktura problemów związanych z markingiem w sieciach inhibitorowych, z wyszczególnieniem dowodów nierozstrzygalności osiągalności. Krawędź skierowana z wierzchołka odpowiadającego problemowi A do wierzchołka odpowiadającego problemowi B oznacza, że rozstrzygalność problemu A implikuje konstruktywnie rozstrzygalność problemu B

1.3 Ślady Mazurkiewicza

Teoria śladów obliczeń, zwanych śladami Mazurkiewicza, dostarcza matematycznego modelu który pozwala opisać i badać zachowanie systemów współbieżnych. Za pierwsze prace dotyczące tej teorii uznaje się książkę Pierre’a Cartiera i Dominique’a Foaty [9] oraz raport Antoniego Mazurkiewicza [34]. Przez ostatnie dekadę ubiegłego wieku teoria śladów skończonych przeżywała burzliwy rozwój, którego efektem była wydana w roku 1995 *The Book of Traces* [16]. Mimo rozległej bibliografii (sama *The Book of Traces* cytuje 278 pozycji), tematyka ta jest nadal aktualna, czego dowodzi choćby konferencja *Developments and New Tracks in the Trace Theory*, która odbyła się w roku 2008 w Cremonie.

Ponadto popularne stały się różne rozszerzenia i uogólnienia pierwotnej teorii. Zaliczyć do nich można ślady nieskończone [15, 21, 29, 36] modelujące obliczenia w działających bez przerwy systemach współbieżnych czy pozwalające lepiej uchwycić pewne aspekty współbieżności półślady (*ang. semitraces*) [11, 12] oraz ślady transakcji współbieżnych (*ang. comtraces*) [24].

1.3.1 Ślady skończone

Niech $I \subseteq \Sigma \times \Sigma$ będzie antyzwrotną i symetryczną relacją, zwaną dalej relacją niezależności. Dopełnienie relacji niezależności I , oznaczane D , zwane będzie relacją zależności, zaś para (Σ, D) - alfabetem współbieżnym.

Definicja 1.3.1 Niech $\equiv \subseteq \Sigma^* \times \Sigma^*$ będzie relacją równoważności indukowaną przez I w następujący sposób:

$$\sigma \equiv \tau \Leftrightarrow \forall_{aDb} \Pi_{a,b}(\sigma) = \Pi_{a,b}(\tau),$$

tzn. rzuty słów σ i τ na pary liter zależnych są równe.

Tak zdefiniowana relacja zachowuje działanie złożenia, jest więc kongruencją.

Uwaga 1.3.2 *Monoid ilorazowy tej relacji Σ^*/\equiv nazywany będzie monoidem śladów, a jego elementy śladami. Zapisywać je będziemy małymi literami z początku alfabetu greckiego (np. α, β, γ). Natomiast parę (Σ, D) , czyli alfabet wraz z relacją zależności, nazywać będziemy alfabetem współbieżnym. Symbolu relacji zależności D będziemy też używać nie tylko w kontekście zbioru wszystkich par liter zależnych ale również rodziny wszystkich jedno- i dwuelementowych podzbiorów (podalfabetów) Σ , w których relacja zależności dziedziczona po D jest relacją pełną.*

Wprost z definicji dla liter $a, b \in \Sigma$ jeśli aDb to $\Pi_{a,b} : \Sigma^*/\equiv \rightarrow \{a, b\}^\infty$, taka że $\Pi_{a,b}([\sigma]) = \Pi_{a,b}(\sigma)$ jest dobrze określona. Funkcji rzutowania będziemy używać w przypadku liter zależnych zarówno dla argumentów będących słowami, jak i śladami. Poza tym dwa słowa, które są równoważne w sensie \equiv są równej długości (bo mają taką samą liczbę wystąpień każdej z liter).

Mając relację zależności w alfabecie współbieżnym, analizę obliczenia sekwencyjnego (słowa) możemy oprzeć o graf zależności tego obliczenia [20], który eksponuje następstwo czasowe kolejnych akcji i po domknięciu przechodnim pozwala traktować obliczenie jako zbiór częściowo uporządkowany. W celu większej przejrzystości sposobu konstrukcji takiego grafu

wprowadzimy pojęcie akcji etykietowanej (niejawnie użytej przez Gastina w [20] na podejściu opisanym w tym artykule opiera się przyjęta w tej rozprawie definicja rzutowa śladów nieskończonych) rozróżniającej kolejne wystąpienia tych samych akcji w trakcie wykonywania obliczenia:

Definicja 1.3.3 Niech $\sigma \in \Sigma^*$. Funkcję $E_\sigma : \{1, \dots, |\sigma|\} \rightarrow \mathbb{N}$, określoną jako:

$$E_\sigma(i) = |\sigma_{(i)}|_{\sigma[i]}$$

nazywamy etykietowaniem akcji obliczenia σ . Akcjami etykietowanymi tego obliczenia nazywamy pary postaci $(\sigma[i], E_\sigma(i))$. Dla obliczenia σ przez σ_{Et} oznaczymy ciąg akcji etykietowanych odpowiadający obliczeniu σ ($\sigma_{Et}[i] = (\sigma[i], E_\sigma(i))$).

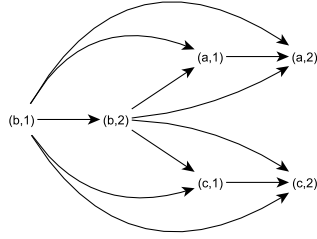
Definicja 1.3.4 Niech (Σ, D) będzie alfabetem współbieżnym. Wówczas *grafem zależności obliczenia* $\sigma \in \Sigma^*$ nazywamy etykietowany, acykliczny i skończony graf skierowany $G = (V, E, \lambda)$, gdzie funkcja $\lambda : V \rightarrow \Sigma \times \mathbb{N}$ przyporządkowuje wierzchołkom grafu akcje etykietowane ciągu σ_{Et} : dwa wierzchołki $v_1, v_2 \in V$ są w relacji E (występuje między nimi krawędź skierowana od v_1 do v_2), dokładnie wtedy gdy akcja etykietowana $\lambda(v_1) = (a_1, x_1)$ występuje w σ_{Et} przed akcją etykietowaną $\lambda(v_2) = (a_2, x_2)$ oraz pierwsze współrzędne tych akcji etykietowanych są zależne, czyli $a_1 D a_2$ (przypomnijmy, że dla każdego $a \in \Sigma$ zachodzi $a D a$).

Uwaga 1.3.5 W praktyce zamiast grafów zależności obliczenia często używa się ich skompresowanej wersji - diagramów Hassego tych grafów. Diagram Hassego powstaje z oryginalnego grafu przez usunięcie niewiele wnoszących krawędzi. Jeśli w grafie $G = (V, E, \lambda)$ istnieje ścieżka długości większej niż jeden v_1, \dots, v_n oraz krawędź (v_1, v_n) należy do E , to ścieżka ta koduje w wystarczający sposób informację o zależności związanej z częściowym porządkiem, która zachodzi między wierzchołkami v_1 oraz v_n , w związku z czym usunięcie tej krawędzi nie spowoduje utraty istotnych informacji. Zależność między grafem zależności $G = (V, E, \lambda)$ a odpowiadającym mu diagramem Hassego $H = (V, E_H, \lambda)$ można precyzyjnie zdefiniować w następujący sposób: zbiory wierzchołków są identyczne zaś relacja sąsiedztwa E_H diagramu Hassego jest najmniejszą relacją, której domknięcie zwrotno-przechodnie daje relację równą domknięciu zwrotno-przechodniemu relacji sąsiedztwa E w grafie zależności G .

Przykład 1.3.6 Rozważmy alfabet współbieżny składający się z $\Sigma = \{a, b, c, d\}$ i relacji zależności $D = Id \cup \{(a, b), (b, a), (b, c), (c, b)\}$ oraz słowo $\sigma = b b d d a c a c$. Do klasy abstrakcji (ślądu) obliczenia [bbddacac] należy 168 słów, w tym: *bbaaccdd, ddbbcca, bdbdacac*. Etykietowanie tego słowa przyporządkowuje kolejnym literom następujące etykiety:

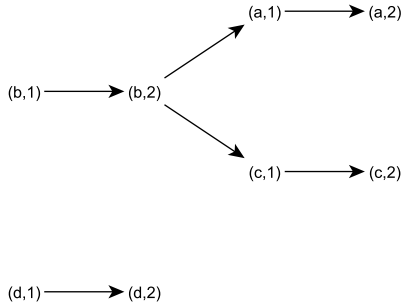
$$\{1, 2, 1, 2, 1, 1, 2, 2\}.$$

Graf zależności ślądu $[\sigma]$ to:



Rysunek 1.18: Graf zależności śladu [bbddacac]

Natomiast jego diagram Hassego ma postać:



Rysunek 1.19: Diagram Hassego śladu [bbddacac]

W literaturze [35, 20], ślady definiowane są zazwyczaj w inny, równoważny powyższemu sposób. Wykorzystywane są do tego pojęcia kongruencji lub domknięcia zwrotno-przechodniego. Równoważność standardowych definicji znaleźć można w wielu opracowaniach na temat śladów Mazurkiewicza (między innymi w [35]) pokazana jest w poniższym fakcie:

Fakt 1.3.7 Niech $\{\Sigma, D\}$ będzie alfabetem współbieżnym, zaś $\equiv \subseteq \Sigma^* \times \Sigma^*$ relacją równoważności z Definicji 1.3.1. Wówczas relacja \equiv jest

- najmniejszą kongruencją, dla której $aIb \Rightarrow ab \equiv ba$
- zwrotno-przechodnim domknięciem relacji $\{(\sigma ab\tau, \sigma ba\tau); \sigma, \tau \in \Sigma^* \wedge aIb\}$
- najmniejszą relacją utożsamiającą obliczenia dające identyczne grafy zależności.

Podobnie jak w przypadku słów, możemy zdefiniować dla śladów relację prefikсового porządku. Dwa ślady będą w tej relacji, jeśli pierwszy z nich jest prefiksem drugiego.

Definicja 1.3.8 Niech \sqsubseteq^{\equiv} będzie relacją na $\Sigma^*/_{\equiv}$, taką że

$$\forall_{\alpha, \beta \in \Sigma^*/_{\equiv}} \alpha \sqsubseteq^{\equiv} \beta \Leftrightarrow \forall_{aDb} \Pi_{a,b}(\alpha) \sqsubseteq \Pi_{a,b}(\beta).$$

Relacja prefiksowego porządku określona na słowach jest zachowana przez odpowiednie ślady (tzn. $\sigma \sqsubseteq \tau \Rightarrow [\sigma] \sqsubseteq^{\equiv} [\tau]$).

Fakt 1.3.9 *Relacja \sqsubseteq^{\equiv} jest relacją częściowego porządku.*

Ponadto, dwa ślady możemy składać biorąc dwa słowa będące ich reprezentantami, składając je i jako wynik biorąc ślad złożenia:

Definicja 1.3.10 Niech $[\sigma], [\tau] \in \Sigma^*/_{\equiv}$. Wówczas

$$[\sigma][\tau] = [\sigma\tau].$$

Z tak zdefiniowaną konkatenacją śladów, możemy uzasadnić równoważność definicji 1.3.8 (relacji prefiksowego porządku) z klasycznymi, naśladującymi definicję prefiskowego porządku dla słów, wersjami tej definicji:

Fakt 1.3.11 *Niech $\alpha, \beta \in \Sigma^{\infty}/_{\equiv}$ oraz $\alpha \sqsubseteq^{\equiv} \beta$. Wówczas*

$$\exists_{\tau} \beta = \alpha\tau.$$

1.3.2 Ślady nieskończone

W modelowaniu i analizie systemów działających bez przerw wygodnie jest używać obliczeń nieskończonych. We współbieżnych wersjach takich systemów, naturalne jest z kolei rozważanie nieskończonych śladów. Podobnie jak w przypadku definiowania śladów skończonych, jest wiele podejść do ich rozszerzania do wersji nieskończonej. Wszystkie one dopuszczają konkatenację dwóch całkowicie niezależnych śladów (również nieskończonych), prowadzą jednak do różnych modeli.

Model 1 - domknięcie w sensie Scotta. Pierwszy z modeli za podstawę bierze prefiksowy częściowy porządek na śladach skończonych. Takie podejście do śladów nieskończonych można znaleźć między innymi w [29]. W modelu tym ślady (skończone i nieskończone) utożsamiamy z ideałami, a właściwie z supremami ideałów, zbioru częściowo uporządkowanego $(\Sigma^*/_{\equiv}, \sqsubseteq^{\equiv})$. Powoduje to, że ślad nieskończony rozumiany jest jako zbiór swoich skończonych prefiksów i dalsze definicje odwołują się w sposób naturalny właśnie do zbiorów skończonych prefiksów. W [36] pokazano, że identyczny wynik daje rozszerzenie na przypadek nieskończony definicji rzutowej. Łatwo też zdefiniować niełączną konkatenację śladów nieskończonych zgodnych z definicją przedstawioną w [29], bez odwoływania się do skończonych prefiksów. Więcej na ten temat znaleźć można w Rozdziale 2.

Definicja ideałowa prowadzi ostatecznie do relacji równoważności $\equiv_S \subseteq \Sigma^{\infty} \times \Sigma^{\infty}$ na słowach, zadanej następująco:

$\sigma \equiv_S \tau$ wtedy i tylko wtedy gdy

$$\begin{aligned} (\forall_{\sigma_0 \in \Sigma^*} \sigma_0 \sqsubseteq \sigma \Rightarrow (\exists_{\tau_0 \in \Sigma^*} \tau_0 \sqsubseteq \tau \wedge [\tau_0] \sqsubseteq^{\equiv} [\sigma_0])) \wedge \\ (\forall_{\tau_0 \in \Sigma^*} \tau_0 \sqsubseteq \tau \Rightarrow (\exists_{\sigma_0 \in \Sigma^*} \sigma_0 \sqsubseteq \sigma \wedge [\sigma_0] \sqsubseteq^{\equiv} [\tau_0])), \end{aligned}$$

czyli każdy skończony prefiks jednego z nich jest prefiksem w sensie śladowym jednego z prefiksów drugiego z nich.

Model 2 - monoid Gastina. Podobnie jak w tej rozprawie, to jest korzystając z rzutów na podalfabety całkowicie zależne, do sprawy śladów nieskończonych podchodzi w [20] Gastin. Zamiast rzutów na zależne podalfabety dwu i jednoliterowe (które są szczególnym przypadkiem pokrycia klikowego) używa on rzutów na dowolne pokrycia klikowe grafu relacji zależności i relację równoważności $\equiv_G \subseteq \Sigma^\infty \times \Sigma^\infty$ na słowach definiuje bezpośrednio:

$$\sigma \equiv_G \tau \text{ wtedy i tylko wtedy gdy } \forall_{S \in D} \Pi_S(\sigma) = \Pi_S(\tau).$$

Problem nielączności pojawiający się przy próbie definicji konkatencji Gastin rozwiązuje dodając do zbioru śladów zero (oznaczane przez \perp) będące wynikiem składania niekompatybilnych śladów.

Model 3 - ślady zespolone [15, 21]. W końcu trzeci, najbardziej rozbudowany model bazuje na grafach zależności śladów. W myśl tej definicji śladem jest każdy acykliczny graf skierowany, w którym wierzchołki etykietowane są literami alfabetu współbieżnego (a nie parami litera i liczba naturalna jak w przypadku Definicji 1.3.4), zaś krawędzie występują między każdą parą wierzchołków etykietowanych literami zależnymi. Pominięty zostaje tu zakaz istnienia ścieżek nieskończonych między dwoma wierzchołkami. Pozwala to używać śladów nieskończonych, a nawet w sposób naturalny i łączny składać je ze sobą. Rozszerza jednak zbiór śladów nie tylko o grafy klas równoważności słów nieskończonych, ale także o tak zwane ślady zespolone, nie związane z żadnym elementem zbioru Σ^∞ .

Dla każdego słowa $\sigma \in \Sigma^\infty$ możemy jednak wyznaczyć graf $G(\sigma)$, w którym zbiór wierzchołków może być nieskończony, ale nie istnieją nieskończone ścieżki między wierzchołkami. I ograniczając się do takich grafów zdefiniować relację równoważności $\equiv_C \subseteq \Sigma^\infty \times \Sigma^\infty$ na słowach:

$$\sigma \equiv_C \tau \text{ wtedy i tylko wtedy gdy } G(\sigma) = G(\tau).$$

Choć zbiory obiektów zdefiniowanych w obrębie przedstawionych modeli różnią się w sposób istotny, to indukują identyczne relacje równoważności w zbiorze Σ^∞ :

Fakt 1.3.12 *Niech $\sigma, \tau \in \Sigma^\infty$ będą obliczeniami sekwencyjnymi. Wówczas następujące warunki są równoważne:*

- $\sigma \equiv_S \tau$
- $\sigma \equiv_G \tau$
- $\sigma \equiv_C \tau$

W dalszej części rozważać będziemy ślady nieskończone w sensie przedstawionym jako Model 2 (choć bez dodatkowego elementu \perp). Możemy wtedy, w naturalny sposób (powtarzając Definicję 1.3.8), przenieść na ślady nieskończone zdefiniowaną wcześniej prefiksową relację częściowego porządku (użyjemy dla niej tego samego oznaczenia):

Definicja 1.3.13 Niech \sqsubseteq^\equiv będzie relacją na Σ^*/\equiv , taką że

$$\forall_{\alpha, \beta \in \Sigma^\infty/\equiv} \alpha \sqsubseteq^\equiv \beta \Leftrightarrow \forall_{aDb} \Pi_{a,b}(\alpha) \sqsubseteq \Pi_{a,b}(\beta).$$

Tak zdefiniowany zbiór częściowo uporządkowany $(\Sigma^\infty / \equiv, \sqsubseteq^\equiv)$ jest, zgodnie z przyświecającą Modelowi I ideą, domknięty na suprema, czyli zupełny w sensie Scotta. Ponadto, elementami skończonymi tego zbioru częściowo uporządkowanego są ślady skończone, dowolne elementy można uzyskać jako suprema ciągów elementów skończonych, zaś elementem minimalnym jest ślad pusty. W związku z tym prawdziwy jest następujący

Fakt 1.3.14 *Dla dowolnego alfabetu współbieżnego (Σ, D) zbiór częściowo uporządkowany $(\Sigma^\infty / \equiv, \sqsubseteq^\equiv)$ jest ω – algebraiczny.*

1.4 Związki sieci elementarnych i śladów

Sieci Petriego i ślady Mazurkiewicza są dwoma matematycznymi formalizmami służącymi do modelowania współbieżności, na pierwszy rzut oka dość odległymi. Przecząc tej tezie, twórca teorii śladów Mazurkiewicz podaje, obok teorii języków formalnych, teorię sieci Petriego jako motywację dla powstania teorii śladów [34, 35]. W tej sekcji spróbujemy naszkicować podstawowe związki między tymi dwoma teoriami, w szczególności przytoczymy za Mazurkiewiczem definicję zależności akcji sieci elementarnych.

W elementarnych sieciach Petriego związków ze śladami Mazurkiewicza najprościej szukać poprzez języki wykonalnych obliczeń, oznaczane dla sieci Petriego S przez $L(S)$. Chcielibyśmy tak zdefiniować relację zależności, aby wraz z każdym wykonalnym obliczeniem wykonalny był też każdy jego śladowy równoważnik, a jednocześnie, aby była to relacja możliwie najmniejsza. Oczywiście kandydatem na warunek niezależności akcji [34, 35] jest całkowita rozłączność sąsiadujących z akcją miejsc:

Definicja 1.4.1 Niech $S = (P, T, F, M_0)$ będzie siecią elementarną, zaś $a, b \in T$ - dwoma akcjami tej sieci. Mówimy, że akcje a oraz b są niezależne (aIb), jeśli

$$(\bullet a \cup a \bullet) \cap (\bullet b \cup b \bullet) = \emptyset.$$

Biorąc jako relację zależności D dopełnienie tak określonej relacji niezależności możemy rozszerzyć alfabet T akcji sieci S do alfabetu współbieżnego (T, D) . Z empirycznego punktu widzenia, dwie akcje a oraz b będące w konflikcie (patrz Definicja 3.1.1) nigdy nie będą mogły się wykonać jedna po drugiej, w związku z czym nigdy nie dojdzie do sytuacji, gdy w wykonalne obliczenie będzie postaci $\sigma ab\tau$. Uprawnia to do uznania akcji skonfliktowanych jako niezależnych (co sprzyja minimalizacji relacji zależności), jest jednak sprzeczne z intuicją. Pozostaniemy w związku z tym przy klasycznej definicji zależności akcji sieci elementarnych.

Przy tak zdefiniowanej relacji zależności, wraz z każdym obliczeniem σ , do języka obliczeń wykonalnych w danej sieci Petriego należy cały ślad $[\sigma]$.

Lemat 1.4.2 *Niech $S = (P, T, F, M_0)$ będzie siecią elementarną, zaś $\sigma \in T^\omega$ obliczeniem. Wówczas*

$$\sigma \in L(S) \wedge \sigma \equiv \tau \Rightarrow \tau \in L(S).$$

Dowód. Pokażemy, że lemat ten jest prawdziwy jeśli $\sigma = uabv$, zaś $\tau = ubav$, gdzie aIb są dwoma niezależnymi akcjami, zaś u, v - ciągami akcji. Załóżmy, że M_0uM , MaM' , $M'bM''$ są wykonalnymi ciągami kroków oraz obliczenie v jest wykonalne w markingu M'' .

Z warunku niezależności akcji a oraz b , ponieważ akcja b jest umożliwiona w markingu $M' = Ma$, to musi być też umożliwiona w markingu M . Podobnie, akcja a musi być umożliwiona w markingu Mb . Ponieważ sieci elementarne spełniają twierdzenie Kellera (mają własność diamentu), to marking wynikowy nie zależy od kolejności w jakiej wykonujemy akcje, czyli $Mab = Mba = M''$, co uzasadnia prawdziwość lematu w tej sytuacji. Ponieważ dowolne dwa reprezentanty śladu można uzyskać jeden z drugiego przy pomocy skończonej liczby takich komutacji, to lemat jest również prawdziwy dla dowolnych obliczeń σ oraz τ . \square

W związku z tym, przez $L(S)$ oznaczać będziemy również język śladów obliczeń wykonalnych w sieci elementarnej S .

Uwaga 1.4.3 *W podobny sposób (przez uzależnienie akcji posiadających wspólne incydentne place) i uzyskując podobne efekty można wprowadzić relację niezależności akcji w p/t-sieciach i sieciach inhibitorowych.*

Rozdział 2

Zbiory rzutowe i rekonstrukcje

W Rozdziale 1, gdzie wprowadziliśmy definicję śladów skończonych, używaliśmy dość często funkcji rzutowania na podalfabety, w których relacja zależności jest relacją pełną. Wielu badaczy zauważyło, że traktowanie śladów jako zbiorów lub wektorów takich rzutów, szczególnie kiedy interesują nas ślady nieskończone, jest bardzo wygodne [20, 36, 43, 46]. Myśląc o śladach jako o klasach abstrakcji sekwencyjnych obliczeń, mamy do czynienia z nieskończonymi zbiorami nieskończonych słów. Zbiory rzutowe pozwalają zredukować jeden nieskończony wymiar i traktować ślady jako skończone zbiory nieskończonych słów.

2.1 Własności rzutowe śladów

2.1.1 Wektory Parikha

Najprostszymi podalfabetami, w których zwrotna relacja zależności jest relacją pełną są podalfabety jednoliterowe. Rozważając rzuty śladu na wszystkie takie podalfabety otrzymujemy informacje o liczbie (ze zbioru $\mathbb{N} \cup \{\omega\}$) wystąpień kolejnych liter w śladzie. Prowadzi to bezpośrednio do definicji wektora Parikha śladu:

Definicja 2.1.1 Niech $v \in \Sigma^\infty$ będzie słowem. Wektorem Parikha tego śladu nazwiemy funkcję $Par_v : \Sigma \rightarrow \mathbb{N} \cup \{\omega\}$ daną wzorem:

$$Par_v(a) = |v|_a = |\Pi_a(v)|.$$

Dla uproszczenia notacji, zamiast $Par_\alpha(a)$ będziemy często pisać $|\alpha|_a$.

Uwaga 2.1.2 *Wektory Parikha różnych słów mogą być identyczne, co w naturalny sposób generuje relację równoważności Parikha na słowach. W języku teorii śladów jest to po prostu relacja równoważności śladowej \equiv indukowanej przez maksymalną antyzwrotną relację niezależności $I = \Sigma \times \Sigma \setminus Id$, a zbiór wszystkich wektorów Parikha nad danym alfabetem Σ jest tożsamy z językiem wszystkich śladów Σ^∞ / \equiv nad całkowicie przemiennym alfabetem współbieżnym (Σ, D) .*

Uwaga 2.1.3 *Wobec oczywistego faktu równości wektorów Parikha dla słów równoważnych ($u \equiv v \Rightarrow Par_u = Par_v$) możemy mówić o wektorach Parikha śladów:*

$$Par_\alpha = Par_w, \text{ gdzie } w \in \alpha.$$

2.1.2 Reprezentacja rzutowa śladu

O ile tylko alfabet współbieżny (Σ, D) nie jest alfabetem całkowicie przemiennym, dwa różne ślady mogą mieć takie same wektory Parikha. W celu ich rozróżnienia, możemy użyć rzutów na dwuliterowe podalfabety alfabetu Σ , a wystarczające okazują się tylko te, w których relacja zależności dziedziczona z (Σ, D) jest relacją pełną. Uzyskujemy w ten sposób reprezentacje rzutowe śladów:

Definicja 2.1.4 Niech (Σ, D) będzie alfabetem współbieżnym, zaś $\alpha \in \Sigma^\infty / \equiv$ śladem nad tym alfabetem. Zbiorem rzutowym śladu α nazywać będziemy funkcję $\Pi_D(\alpha) : D \rightarrow \Sigma^\omega$ (przy czym D rozumiane jest w sensie Uwagi 2.1.2 - jako rodzina zbiorów) daną wzorem:

$$\Pi_D(\alpha)(\{a, b\}) = \Pi_{a,b}(\alpha).$$

Uwaga 2.1.5 *Nietrudno zauważyć, że reprezentacja rzutowa śladu jest tylko wyeksponowaniem rzutowego charakteru Definicji 1.3.1. W związku z tym, wprost z definicji, funkcja Π_D jest funkcją różnowartościową.*

2.2 Zbiory rzutowe

W czasie rozważań rzutów na podalfabety o mocy większej niż jeden, informacja o liczbie elementów wchodzących w skład tych rzutów przestała być wystarczająca. W miejsce wektorów Parikha, w analogiczny sposób, wprowadzimy uogólnienie reprezentacji rzutowych śladów - zbiory rzutowe [37], czyli elementy iloczynu kartezjańskiego języków sekwencyjnych nad jedno i dwuliterowymi podalfabetami z pełną relacją zależności (zbiór tych podalfabetów jest szczególnym przypadkiem rozbicia klikowego grafu relacji D):

Definicja 2.2.1 Niech (Σ, D) będzie alfabetem współbieżnym. Wówczas zbiorem rzutowym nazywać będziemy funkcję $P : D \rightarrow \Sigma^\omega$ (przy czym D rozumiane jest w sensie Uwagi 2.1.2), która spełnia warunek:

$$P(\{a, b\}) \in \{a, b\}^\infty.$$

Rodzinę wszystkich zbiorów rzutowych oznaczać będziemy przez $\mathbb{P}\mathbb{S}$.

Przykład 2.2.2 Niech $(\Sigma = \{a, b, c\}, D = \Sigma \times \Sigma)$ będzie alfabetem współbieżnym (całkowicie zależnym). Rozważmy następujące przykłady zbiorów rzutowych nad tym alfabetem:

$$P_1 = \left\{ \begin{array}{l} (\{a, a\}, aa) \\ (\{a, b\}, abab) \\ (\{a, c\}, aca) \\ (\{b, b\}, bb) \\ (\{b, c\}, cbb) \\ (\{c, c\}, c) \end{array} \right\} P_2 = \left\{ \begin{array}{l} (\{a, a\}, aa) \\ (\{a, b\}, aba) \\ (\{a, c\}, aca) \\ (\{b, b\}, bb) \\ (\{b, c\}, cbb) \\ (\{c, c\}, c) \end{array} \right\} P_3 = \left\{ \begin{array}{l} (\{a, a\}, aa) \\ (\{a, b\}, abab) \\ (\{a, c\}, aca) \\ (\{b, b\}, bb) \\ (\{b, c\}, bbc) \\ (\{c, c\}, c) \end{array} \right\}.$$

Spośród przedstawionych zbiorów rzutowych, tylko P_1 jest reprezentacją rzutową śladu (precyzyjnie rzecz ujmując - śladu $[acbab]$).

W rodzinie zbiorów rzutowych, podobnie jak w przypadku języków słów i śladów, możemy wprowadzić prefiksową relację częściowego porządku:

Definicja 2.2.3 Niech (Σ, D) będzie alfabetem współbieżnym. Wówczas prefiksową relacją częściowego porządku w rodzinie wszystkich zbiorów rzutowych \mathbb{PS} nazwiemy relację $\sqsubseteq^\Pi \subseteq \mathbb{PS} \times \mathbb{PS}$ daną formułą:

$$P_1 \sqsubseteq^\Pi P_2 \Leftrightarrow \forall_{S \in D} P_1(S) \sqsubseteq P_2(S).$$

Zacytujemy teraz definicję oraz ogólne twierdzenie z teorii dziedzin[1]:

Definicja 2.2.4 Produktem kartezjańskim dwóch zbiorów częściowo uporządkowanych kierunkowo-zupełnych (D, \leq) oraz (E, \leq) nazywamy zbiór częściowo uporządkowany $(D \times E, \leq')$, gdzie

- $D \times E = \{(x, y) \mid x \in D \wedge y \in E\}$,
- $(x, y) \leq' (x', y')$ wtedy i tylko wtedy gdy $x \leq x' \wedge y \leq y'$

Fakt 2.2.5 *Produkt kartezjański $(D \times E, \leq')$ dwóch zbiorów częściowo uporządkowanych kierunkowo-zupełnych (D, \leq) oraz (E, \leq) jest zbiorem częściowo uporządkowanym kierunkowo-zupełnym. Suprema oraz infima liczone są po współrzędnych.*

Konstrukcja zbioru częściowo uporządkowanego zbiorów rzutowych spełnia założenia tego twierdzenia uogólnionego na produkt skończonej liczby zbiorów częściowo uporządkowanych kierunkowo-zupełnych, w związku z czym:

Wniosek 2.2.6 *Relacja \sqsubseteq^Π jest częściowym porządkiem. Zbiór częściowo uporządkowany $(\mathbb{PS}, \sqsubseteq^\Pi)$ jest zupełny w sensie Scotta. Suprema oraz infima liczone są po współrzędnych.*

Na koniec zdefiniujemy jeszcze operację konkatenacji dwóch zbiorów rzutowych. Zgodnie z dotychczas przyjętą zasadą oprzemy ją o produktowy charakter tych obiektów i zdefiniujemy po współrzędnych.

Definicja 2.2.7 Niech $P_1, P_2 \in \mathbb{PS}$ będą dwoma zbiorami rzutowymi. Wówczas przez $P_1 P_2$ rozumieć będziemy zbiór rzutowy P_3 taki, że

$$\forall_{S \in D} P_3(S) = P_1(S)P_2(S).$$

Ponieważ reprezentacja rzutowa śladu jest szczególnym przypadkiem zbioru rzutowego, a tak sformułowana definicja konkatenacji jest równoważna definicji składania śladów skończonych rozumianych jako zbiory rzutowe, możnaby pokusić się o próbę wykorzystania jej do zdefiniowania konkatenacji śladów. Poniższy przykład pokazuje dobitnie, że nie da się tego zrobić bezpośrednio.

Przykład 2.2.8 Rozważmy następujący alfabet współbieżny $(\Sigma = \{a, b, c\}, I = \{(a, c)\})$ i zbiory rzutowe, będące reprezentacjami rzutowymi śladów $\alpha = [a^\omega]$, $\beta = [b]$ oraz $\gamma = [c]$:

$$\Pi_D(\alpha) = \left\{ \begin{array}{l} (\{a, a\}, a^\omega) \\ (\{a, b\}, a^\omega) \\ (\{b, b\}, \epsilon) \\ (\{b, c\}, \epsilon) \\ (\{c, c\}, \epsilon) \end{array} \right\}, \quad \Pi_D(\beta) = \left\{ \begin{array}{l} (\{a, a\}, \epsilon) \\ (\{a, b\}, b) \\ (\{b, b\}, b) \\ (\{b, c\}, b) \\ (\{c, c\}, \epsilon) \end{array} \right\}, \quad \Pi_D(\gamma) = \left\{ \begin{array}{l} (\{a, a\}, \epsilon) \\ (\{a, b\}, \epsilon) \\ (\{b, b\}, \epsilon) \\ (\{b, c\}, c) \\ (\{c, c\}, c) \end{array} \right\}.$$

Wynikiem konkatenacji tych trzech reprezentacji rzutowych jest zbiór rzutowy

$$P = \left\{ \begin{array}{l} (\{a, a\}, a^\omega) \\ (\{a, b\}, a^\omega) \\ (\{b, b\}, b) \\ (\{b, c\}, bc) \\ (\{c, c\}, c) \end{array} \right\},$$

który w sposób oczywisty (rzut na $\{b, b\}$ zawiera jedno wystąpienie b , zaś rzut na $\{a, b\}$ nie zawiera żadnego) nie jest reprezentacją rzutową żadnego śladu.

Efekt opisany w powyższym przykładzie powoduje, że określona na zbiorach rzutowych konkatenacja wyprowadza poza rodzinę zbiorów rzutowych będących reprezentacjami śladów. Kolejne sekcje poświęcimy bliższemu przyjrzeniu się tej rodzinie oraz takiemu poprawieniu konkatenacji, aby poza nią nie wyprowadzała.

2.3 Wewnętrznie bezkonfliktowe zbiory rzutowe

W swojej wczesnej pracy na temat śladów nieskończonych [20], Gastin używał obiektów podobnych do przedstawionych w Definicji 2.2.1 zbiorów rzutowych. Te z nich, z których da się zbudować grafy śladów, a więc są reprezentacjami rzutowymi śladów - nazwał *rekonstruowalnymi*.

W Przykładzie 2.2.2 pokazaliśmy trzy, bardzo podobne, zbiory rzutowe. Pierwszy z nich jest rekonstruowalny, dwa pozostałe nie. W przypadku drugiego nie zgadzają się liczby liter występujących w różnych rzutach (precyzyjnie - w drugim rzucie litera b występuje tylko raz, a nie dwa razy, jak w rzutach czwartym i piątym). Problem z trzecim przykładem jest bardziej subtelny - wystąpiło tam klasyczne zjawisko zakleszczenia; nie da się tak ułożyć liter występujących w tym obiekcie (co prowadziłoby do odtworzenia jednego z reprezentantów śladu) aby występowały one w zaproponowanej w rzutach drugim, trzecim i piątym kolejności. Oba te przykłady pokazują pewne typy wewnętrznych konfliktów występujących w zbiorach rzutowych, które powodują, że nie są one rekonstruowalne.

Przedstawimy teraz klasę zbiorów rzutowych wolnych od tego typu konfliktów, nazwanych wewnętrznie bezkonfliktowymi zbiorami rzutowymi [37]:

Definicja 2.3.1 Rodziną wewnętrznie bezkonfliktowych zbiorów rzutowych (w skrócie *wbzt*) nazywamy najmniejszą rodzinę F spełniającą warunki:

- $(\forall_{S \in D} P(S) = \epsilon) \Rightarrow P \in F$
- $P \in F \wedge P$ jest skończony $\Rightarrow \forall_{c \in \Sigma} P' = P \Pi_D([c]) \in F$
- $\forall_i (P_i \text{ jest skończony} \wedge P_i \sqsubseteq^\Pi P_{i+1} \wedge P_i \in F) \Rightarrow \sup P_i \in F$.

Uwaga 2.3.2 Drugi punkt powyższej definicji, polegający na dopisaniu do skończonego zbioru rzutowego pojedynczej litery we wszystkich rzutach w których ona występuje można, zgodnie z oryginalnym sformułowaniem [37], równoważnie zapisać:

$$P \in F \wedge P \text{ jest skończony} \wedge (\exists_{c \in \Sigma} \forall_{S \in D} P'(S) = P(S) \Pi_S(c)) \Rightarrow P' \in F$$

Jedną z użytecznych obserwacji dotyczących wewnętrznie bezkonfliktowych zbiorów rzutowych jest, gwarantowana przez trzeci punkt Definicji 2.3.1, domkniętość jej skierowanych podrodzin na suprema. Drobnym i wyłącznie technicznym problemem są tutaj mogące się pojawić w danej podrodzinie skierowanej nieskończone zbiory rzutowe. Przydatna będzie do tego określona po współrzędnych operacja przeciwna do konkatenacji, którą podobnie jak w przypadku słów, nazwiemy różnicą prefiksową:

Definicja 2.3.3 Niech $P_1, P_2 \in \mathbb{P}\mathbb{S}$ będą dwoma kompatybilnymi, czyli posiadającymi wspólne ograniczenie górne zbiorami rzutowymi. Wówczas $P_3 = P_2 \setminus_{\sqsubseteq^{\Pi}} P_1$ oznacza unikatowy zbiór rzutowy taki, że $\forall_{x \in D} P_3(x) = P_2(x) \setminus_{\sqsubseteq} P_1(x)$. Naturalnie, jeśli $P_1 \sqsubseteq^{\Pi} P_2$, to $P_1 P_3 = P_2$.

Lemat 2.3.4 Niech $S \subseteq F$ będzie skierowaną podrodziną wbrz nad alfabetem współbieżnym (Σ, D) . Wówczas $\sup S \in F$, czyli ograniczenie górne tej podrodziny istnieje i jest wewnętrznie bezkonfliktowe.

Dowód. Skorzystamy z Wniosku 2.2.6, dotyczącego zbioru częściowo uporządkowanego $(\mathbb{P}\mathbb{S}, \sqsubseteq^{\Pi})$. Ponieważ rodzina $\mathbb{P}\mathbb{S}$ wszystkich zbiorów rzutowych jest dziedziną Scotta, $\sup S$ jako ograniczenie górne jej podrodziny skierowanej istnieje i należy do $\mathbb{P}\mathbb{S}$. Pozostaje pokazać, że musi ono należeć również do podrodziny wewnętrznie bezkonfliktowych zbiorów rzutowych.

Weźmy wstępujący ciąg u zbiorów z rodziny S , który w granicy daje $\sup S$. Od pewnego miejsca zbiory te mogą być nieskończone. Pokażemy, jak zastąpić je zbiorami skończonymi i skonstruować ciąg v , aby supremum ciągu nie zmieniło się. Na początek wszystkie skończone elementy ciągu u przepisujemy do ciągu v . Weźmy teraz parę (v_i, u_{i+1}) taką, że v_i -skończony, zaś u_{i+1} -nieskończony. Weźmy różnicę prefiksową tych elementów $x = u_{i+1} \setminus_{\sqsubseteq^{\Pi}} v_i$ a następnie wyznaczmy zbiór A wszystkich akcji znajdujących się w pierwszej warstwie x , czyli znajdujących się na pierwszych pozycjach w każdym z rzutów na podalfabety je zawierające. Jako v_{i+1} przyjmiemy v_i skonkatenowane (w sensie drugiego punktu Definicji 2.3.1) w dowolnej kolejności ze wszystkimi elementami zbioru A .

Operację tę powtarzamy uzyskując ciąg złożony z samych skończonych zbiorów rzutowych. Pozostaje nam pokazać, że supremum tego ciągu jest identyczne z supremum wyjściowego ciągu u . Do tego z kolei wystarczy wziąć dowolny element u_i ciągu u i pokazać, że każdy skończony prefiks w tego elementu jest prefiksem $\sup v$. Weźmy $w' = w \setminus_{\sqsubseteq^{\Pi}} v_i$. Nietrudno zauważyć, że kolejne warstwy (być może rozszerzony o pewne niezależne elementy) zbioru rzutowego w' były dodawane do v_i w kolejnych krokach opisanej procedury. A zatem, ze skończoności zbioru rzutowego w , istnieje takie j , że $w \sqsubseteq^{\Pi} v_j$. Z dowolności elementów u_i oraz w otrzymujemy, że $\sup v = \sup u$, czyli na mocy trzeciego punktu Definicji 2.3.1 $\sup u$ jest wewnętrznie bezkonfliktowym zbiorem rzutowym, co kończy dowód. \square

Wniosek 2.3.5 Lemat 2.3.4 wraz z Wnioskiem 2.2.6 pozwalają stwierdzić, że rodzina wewnętrznie bezkonfliktowych zbiorów rzutowych wraz z prefiksową relacją częściowego porządku \sqsubseteq^{Π} jest zbiorem częściowo uporządkowanym zupełnym w sensie Scotta.

Innym ciekawym faktem jest możliwość rozszerzenia na nieskończone wewnętrznie bezkonfliktowe zbiory rzutowe operacji dopisywania zbiorów jednoliterowych opisanej w punkcie

drugim Definicji 2.3.1. Jest ona możliwa również w przypadku gdy niektóre z rzutów (zupełnie niezwiązanych z dodawaną literą) są nieskończone.

Lemat 2.3.6 *Niech $P \in \mathbb{P}\mathbb{S}$ będzie wewnętrznym bezkonfliktowym zbiorem rzutowym nad alfabetem współbieżnym (Σ, D) , zaś $a \in \Sigma$ literą tego alfabetu. Jeśli wszystkie współrzędne zbioru P związane z tą literą są skończone, to zbiór rzutowy $P \cdot \Pi_D([a])$ jest wewnętrznym bezkonfliktowym. Bardziej formalnie:*

$$(\forall_{S \in D} a \in S \Rightarrow |P(S)| < \omega) \Rightarrow P \cdot \Pi_D([a]) \in \text{wbzr.}$$

Dowód. W sytuacji, gdy zbiór rzutowy P jest skończony, teza jest prawdziwa na mocy Definicji 2.3.1. W przeciwnym przypadku, istnieje ciąg P_i skończonych, wewnętrznym bezkonfliktowych zbiorów rzutowych dający w granicy P ($\sup_n P_n = P$). Na mocy Definicji 2.3.1, zbiory $P'_i = P_i \cdot \Pi_D([a])$ są wewnętrznymi bezkonfliktowymi zbiorami rzutowymi, natomiast założenie lematu gwarantuje, że istnieje takie $k \in \mathbb{N}$ począwszy od którego współrzędne na których występuje litera a (podzbiory alfabetu zawierające tę literę) w zbiorach rzutowych P_i są stałe (równe odpowiednim współrzędnym w zbiorze P_k i w całym P). Zauważmy, że począwszy od tego momentu zbiory P'_i tworzą zbiór skierowany, a więc istnieje dla niego najmniejsze ograniczenie z góry P' , które wprost z definicji jest wewnętrznym bezkonfliktowym. Pozostaje uzasadnić, że $P \cdot \Pi_D([a]) = P'$. Ponieważ w dziedzinach Scotta granice liczy się po współrzędnych, wystarczy aby $\forall_{S \in D} P \cdot \Pi_D([a])(S) = P'(S)$. Rozważmy dwa przypadki, kiedy $a \in S$ oraz kiedy $a \notin S$. W przypadku pierwszym, zachodzi opisana wyżej sytuacja, w której począwszy od $i = k$ mamy $P(S)a = P_i(S)a = P'_i(S) = P'(S)$. W przypadku pozostałych współrzędnych, zbiory rzutowe P_i oraz P'_i są identyczne, czyli $P(S) = \sup_i P_i(S) = \sup_i P'_i(S) = P'(S)$, co kończy dowód. \square

Uwaga 2.3.7 *Zauważmy, że powyższy lemat jest prawdziwy również gdy wszystkie współrzędne zawierające literę a są nieskończone (co pociąga za sobą, że litera a występuje w nich nieskończenie często). Natomiast, gdy istnieje zarówno skończona jak i nieskończona współrzędna na której występuje a , teza lematu jest fałszywa (litera a zostanie dopisana do współrzędnej skończonej, zaś pominięta przy współrzędnej nieskończonej).*

Wniosek 2.3.8 *Niech $P, R \in \mathbb{P}\mathbb{S}$ będą dwoma wewnętrznymi bezkonfliktowymi zbiorami rzutowymi spełniającymi warunek $|P(S)| = \omega \Rightarrow R(S) = \epsilon$. Wówczas PR jest również wewnętrznym bezkonfliktowym zbiorem rzutowym.*

Sekcję tę podsumujemy, nasuwającym się w świetle przedstawionych własności struktury śladów nieskończonych, wnioskiem.

Wniosek 2.3.9 *Zbiór rzutowy $P \in \mathbb{P}\mathbb{S}$ jest rekonstruowalny dokładnie wtedy, gdy jest wewnętrznym bezkonfliktowym.*

2.4 Rekonstrukcja śladów

W podsekcji 2.1.2 zdefiniowaliśmy funkcję rzutowania przyporządkowującą każdemu śladowi zbiór rzutowy. W kolejnych sekcjach wskazaliśmy, że funkcja rzutowania nie działa

na rodzinę zbiorów rzutowych (nie jest surjekcją). Przyjrzyjmy się teraz nieco bliżej, wykorzystanej w dowodzie Lematu 2.3.4, idei warstw zbioru rzutowego. Przy ich pomocy zdefiniujemy funkcję działającą z rodziny wszystkich zbiorów rzutowych do zbioru śladów, która zbiorom rzutowym przyporządkuje ślady odpowiadające największej, mniejszej lub równej od danego zbioru rzutowego reprezentacji rzutowej. W szczególności, ograniczona do reprezentacji rzutowych śladów, będzie to funkcja odwrotna do rzutowania.

W celu zapewnienia uczciwości poniższej konstrukcji, operacja rekonstrukcji opisana w [36] wykonywana była wzdłuż specjalnego kanonicznego wzorca, w którym każda litera z alfabetu występuje nieskończenie często. Zabezpieczało to przed pominięciem akcji niezależnych od nieskończenie wielu innych, występujących w zbiorze rzutowym akcji. Zadanie to można, z podobnym skutkiem, zrealizować przy pomocy wprowadzonych w dowodzie Lematu 2.3.4 warstw, co zastosowano niżej.

Konstrukcja 2.4.1 Weźmy zbiór rzutowy P . Skonstruujemy wstępujący ciąg słów v_n , będących kolejnymi przybliżeniami jednego z reprezentantów śladu, związanego w opisany wyżej sposób ze zbiorem rzutowym P . Równocześnie konstruować będziemy ciąg zbiorów rzutowych P_n , będących sufiksami wyjściowego zbioru P . Za v_0 przyjmijmy słowo puste ϵ , zaś za P_0 - wyjściowy zbiór rzutowy P .

Przypomnijmy, że przez pierwszą warstwę zbioru rzutowego rozumiemy zbiór liter złożony z tych, które występują jako pierwsze we wszystkich rzutach na podalfabety zawierające tę literę. W kolejnych krokach wyznaczać będziemy pierwszą warstwę A i -tego sufiksu P_i i tworzyć $i + 1$ - szą sufiks przez jej odjęcie (traktując tę warstwę jako zbiór rzutowy) oraz kolejne przybliżenie v_{i+1} przez dodanie do niego liter z warstwy A (dla ustalenia uwagi - w kolejności leksykograficznej, choć nie ma to znaczenie, gdyż litery występujące w jednej warstwie są parami niezależne). Bardziej formalnie:

$$\begin{aligned} A &= \{a \in \Sigma \mid \forall_{S \in D} a \in S \Rightarrow P_i(S)[1] = a\}, \\ P_{i+1} &= P_i \setminus \sqsubseteq^{\Pi} \Pi_D([\text{lex}(A)]), \\ v_{i+1} &= v_i \text{lex}(A), \end{aligned}$$

gdzie przez $\text{lex}(A)$ rozumiemy leksykograficzne uszeregowanie zbioru A .

Wynikiem operacji rekonstrukcji $R_D : \mathbb{P}\mathbb{S} \rightarrow \Sigma^\infty / \equiv$ jest ślad obliczenia będącego granicą ciągu v_i , czyli

$$R_D(P) = [\text{sup}_n v_n].$$

Przykład 2.4.2 (Rekonstrukcja) Niech $\Sigma = \{a, b, c, d\}$, zaś I - symetryczne uzupełnienie relacji $\{(a, d), (a, c), (b, d), (c, d)\}$. Przeanalizujmy konstrukcję ciągów (v_n) oraz (P_n) dla zbioru rzutowego

$$P = \left\{ \begin{array}{l} (\{a, a\}, a^\omega) \\ (\{a, b\}, bba^\omega) \\ (\{b, b\}, bb) \\ (\{b, c\}, bbc^\omega) \\ (\{c, c\}, c^\omega) \\ (\{d, d\}, dd) \end{array} \right\},$$

będącego reprezentacją rzutową śladu $[bbdd(ac)^\omega]$. Wyznamy pierwszą warstwę A dla zbioru rzutowego $P_0 = P$. Należą do niej dwie litery spełniające zawarty w definicji warunek. Są to b oraz d . Litery a oraz c występują co prawda na pierwszych miejscach w

rzutach $P_0(\{a, a\})$ oraz $P_0(\{c, c\})$ odpowiednio, ale w rzutach $P_0(\{a, b\})$ oraz $P_0(\{b, c\})$ na pierwszych pozycjach występuje litera b . W związku z tym $v_1 = \text{lex}(\{b, d\}) = bd$, zaś

$$P_1 = \left\{ \begin{array}{l} (\{a, a\}, a^\omega) \\ (\{a, b\}, ba^\omega) \\ (\{b, b\}, b) \\ (\{b, c\}, bc^\omega) \\ (\{c, c\}, c^\omega) \\ (\{d, d\}, d) \end{array} \right\}.$$

W drugim kroku, pierwsza warstwa rozważanego zbioru rzutowego (P_1) to ponownie dwuelementowy zbiór złożony z liter b oraz d . W związku z tym $v_2 = bd \cdot bd = bdbd$, zaś

$$P_2 = \left\{ \begin{array}{l} (\{a, a\}, a^\omega) \\ (\{a, b\}, a^\omega) \\ (\{b, b\}, \epsilon) \\ (\{b, c\}, c^\omega) \\ (\{c, c\}, c^\omega) \\ (\{d, d\}, \epsilon) \end{array} \right\}.$$

W kolejnych krokach pierwszymi warstwami są dwuelementowe zbiory $A = \{a, c\}$, zaś począwszy od $i = 3$

$$P_i = \left\{ \begin{array}{l} (\{a, a\}, a^\omega) \\ (\{a, b\}, a^\omega) \\ (\{b, b\}, \epsilon) \\ (\{b, c\}, c^\omega) \\ (\{c, c\}, c^\omega) \\ (\{d, d\}, \epsilon) \end{array} \right\},$$

podczas gdy $v_{i+2} = bdbd(ac)^i$. Ostatecznie, w wyniku rekonstrukcji otrzymujemy ciąg v_i dający w granicy słowo nieskończone $bdbd(ac)^\omega$, czyli ślad $[bdbd(ac)^\omega] = [bddd(ac)^\omega]$. Warto zauważyć, że biorąc pod uwagę ciąg P_n , w granicy nie uzyskaliśmy pustego zbioru rzutowego $\Pi_D([\epsilon])$, mimo że wyjściowy zbiór rzutowy P jest rekonstruowalny.

Uwaga 2.4.3 *W rekonstrukcji opisanej w [36] w kolejnych krokach zamiast całych warstw brano pojedyncze litery. Takie postępowanie daje poprawne wyniki w przypadku skończonych zbiorów rzutowych, jednak w ogólności kryje w sobie dość poważne zagrożenie. Gdybyśmy w poprzednim przykładzie zamiast warstw brali najmniejsze leksykograficznie litery będące pierwszymi literami w odpowiednich rzutach, wynikiem byłoby $[bba^\omega]$. W cytowanej pracy efekt ten wyeliminowano dzięki kontroli uczciwości wyboru liter dołączanych do kolejnych elementów ciągu v_i .*

Wykorzystamy teraz informację o wyniku operacji R_D i opiszemy ją w sposób bardziej formalny. Użyjemy do tego celu zbiory wszystkich wewnętrznie bezkonfliktowych prefiksów zbiorów rzutowych:

Definicja 2.4.4 Niech $P \in \mathbb{PS}$ będzie zbiorem rzutowym, zaś $F \subset \mathbb{PS}$ oznacza rodzinę bezkonfliktowych zbiorów rzutowych. Przez $WBP(P) \subseteq \mathbb{PS}$ oznaczmy rodzinę wewnętrznie bezkonfliktowych zbiorów rzutowych mniejszych lub równych P . Bardziej formalnie:

$$WBP(P) = \{P' \in F \mid P' \sqsubseteq^\Pi P\}.$$

Zbiory WBP posiadają kilka ciekawych własności. Mianowicie, są to zawsze zbiory niepuste (do każdego takiego zbioru należy $\Pi_D([\epsilon])$). Poza tym są one zbiorami skierowanymi, czyli

Lemat 2.4.5 *Niech $P \in \mathbb{PS}$ będzie zbiorem rzutowym, zaś $P_0, P_1 \in WBP(P)$ jego wewnętrznie bezkonfliktowymi prefiksami. Wówczas w rodzinie $WBP(P)$ istnieje wspólne ograniczenie z góry dla zbiorów P_0 oraz P_1 , czyli taki zbiór $P_2 \in WBP(P)$, że $P_0 \sqsubseteq^\Pi P_2 \wedge P_1 \sqsubseteq^\Pi P_2$.*

Dowód. Na początku zauważmy, że na każdej współrzędnej $S \in D$ zbiory rzutowe P_0 oraz P_1 są porównywalne. To znaczy, $P_0(S)$ jest prefiksem $P_1(S)$ lub na odwrót, $P_1(S)$ jest prefiksem $P_0(S)$. W przeciwnym przypadku zbiory te nie mogłyby być jednocześnie prefiksami P .

Rozważmy zbiór rzutowy $P' = P_1 \setminus \sqsubseteq^\Pi P_0$. Zauważmy, że na mocy poczynionej na wstępie uwagi, współrzędna $P'(S)$ może być niepusta tylko wtedy, gdy odpowiadająca jej współrzędna $P_0(S)$ jest skończona (w przeciwnym przypadku $P_0(S)$ nie mogłoby być właściwym prefiksem $P_1(S)$). Stąd P_0P' musi być, na mocy Wniosku 2.3.8 wewnętrznie bezkonfliktowy. Oczywiście P_0 jest prefiksem P_0P' , ale również P_1 jest prefiksem P_0P' . Aby to uzasadnić rozważmy dowolną współrzędną $S \in D$ zbioru P_0P' . Jeśli $P_1(S) \sqsubseteq P_0(S)$, to także $P_1(S) \sqsubseteq P_0P'(S)$. Jeśli natomiast $P_0(S) \sqsubseteq P_1(S)$, to $P_0P'(S) = P_0(S)(P_1(S) \setminus \sqsubseteq P_0(S)) = P_1(S)$, czyli $P_1(S) = P_0P'(S)$. Do wykazania, że P_0P' należy do rodziny $WBP(P)$, pozostaje uzasadnić, że $P_0P' \sqsubseteq^\Pi P$. Wprost z określenia P' widać, że każda współrzędna $S \in D$ tego zbioru jest odpowiednią współrzędną zbioru P_0 lub zbioru P_1 . Daje to, wprost z definicji częściowego porządku dla zbiorów rzutowych, tezę i kończy dowód. \square

W związku z tym zbiory te posiadają najmniejsze ograniczenia z góry (pewnym ograniczeniem, niekoniecznie najmniejszym, jest sam zbiór P). Ponadto, na mocy Lematu 2.3.4 zbiory WBP są domknięte w sensie Scotta, czyli zawierają swoje suprema. Bardziej formalnie:

Wniosek 2.4.6 *Niech $P \in \mathbb{PS}$ będzie zbiorem rzutowym. Wówczas*

$$\sup(WBP(P)) \in WBP(P).$$

Korzystając z tych własności i mając w pamięci nieformalny opis wyniku operacji R_D możemy sformułować

Wniosek 2.4.7 [37] *Dla zbioru rzutowego $P \in \mathbb{PS}$ operacja rekonstrukcji R_D daje w wyniku ślad, którego reprezentacja rzutowa jest supremum zbioru $WBP(P)$, czyli*

$$\Pi_D(R_D(P)) = \sup(WBP(P)).$$

2.5 Konkatenacja śladów nieskończonych

Sekcję tę rozpoczniemy od przeglądu różnych definicji konkatenacji śladów nieskończonych, jakie można znaleźć w literaturze. Odpowiadać one będą trzem przedstawionym w

Rozdziale 1 modelom rozszerzania definicji śladów na przypadek nieskończony.

Na początek przyjrzymy się modelowi trzeciemu - śladom zespolonym, czyli odpowiednim acyklicznym grafom skierowanym. Przypomnijmy, że w tym modelu funkcja etykietująca ma inne ($\lambda : V \rightarrow \Sigma$) niż zawarte w Definicji 1.3.4 sformułowanie ($\lambda : V \rightarrow \Sigma \times \mathbb{N}$). Przyjęcie tutaj etykietowania zgodnego z tą definicją jest problematyczne, gdyż porządek wierzchołków etykietowanych tą samą literą zgodny z kolejnością występowania tych wierzchołków może nie być dobrym porządkiem.

W tym modelu nie mamy większych problemów ze zdefiniowaniem konkatenacji. Świetnie działa definicja przeniesiona ze świata skończonego. Grafy składa się przez dodanie nowych krawędzi dla wszystkich par liter zależnych, z których jedna należy do pierwszego argumentu, zaś druga - do drugiego, zachowując przy tym porządek argumentów. Bardziej formalnie, niech $G_1 = (V_1, E_1, \lambda_1)$ oraz $G_2 = (V_2, E_2, \lambda_2)$ będą dwoma grafami śladów zespolonych o rozłącznych zbiorach wierzchołków. Wówczas $G_3 = G_1.G_2$ jest grafem zespolonym $G_3 = (V_3, E_3, \lambda_3)$, gdzie

- $V_3 = V_1 \cup V_2$
- $E_3 = E_1 \cup E_2 \cup \{(v_1, v_2) \mid v_1 \in V_1 \wedge v_2 \in V_2 \wedge \lambda_1(v_1)D\lambda_2(v_2)\}$
- $\lambda_3(v) = \begin{cases} \lambda_1(v) & \text{gdy } v \in V_1 \\ \lambda_2(v) & \text{gdy } v \in V_2. \end{cases}$

Przejdźmy teraz do modelu drugiego - śladów Gastina. Przypomnijmy, że w tym modelu śladami są elementy zbioru $\Sigma^\infty / \equiv \cup \{\perp\}$. Konkatenację definiujemy analogicznie jak w przypadku rzeczywistych śladów zespolonych, pod warunkiem, że wynik pozostaje śladem rzeczywistym (w grafie nie ma ścieżek między dwoma wierzchołkami, których długość jest nieskończona - zdefiniowanie funkcji etykietującej $\lambda : V \rightarrow \Sigma \times \mathbb{N}$ przestaje być wówczas problemem). W przeciwnym wypadku wynikiem konkatenacji jest \perp . Natomiast \perp składany z dowolnym elementem, zarówno z lewej jak i z prawej strony, daje w wyniku \perp .

Z kolei w modelu pierwszym koncentrujemy się na części rzeczywistej konkatenacji śladów zespolonych. Za Kwiatkowską [29] możemy ją zdefiniować, zachowując oryginalne oznaczenia, w następujący sposób:

$$\alpha\beta = \sup\{\alpha_1\beta_1 \mid \alpha_1 \sqsubseteq_{fin}^{\equiv} \alpha \wedge \beta_1 \sqsubseteq_{fin}^{\equiv} \beta \wedge \beta_1 \text{ jest ogonowo niezależna z } \alpha \text{ po } \alpha_1\},$$

przy czym relacja $\sqsubseteq_{fin}^{\equiv}$ oraz niezależność ogonowa zdefiniowane są w następujący sposób:

- Mówimy, że $\alpha \sqsubseteq_{fin}^{\equiv} \beta$ jeśli dla każdego zbioru skierowanego DS takiego, że $\beta \sqsubseteq^{\equiv} \sup DS$ istnieje $\gamma \in DS$ taka, że $\alpha \sqsubseteq^{\equiv} \gamma$.
- Ślad β jest ogonowo niezależny od śladu α po jego prefiksie α_1 jeśli

$$\forall \alpha_2 \in \Sigma^\infty / \equiv \alpha_1\alpha_2 \sqsubseteq_{fin}^{\equiv} \alpha \Rightarrow \alpha_2 I \beta_1$$

Wszystkie trzy definicje działają w obrębie wspólnej części dziedziny (Σ^∞ / \equiv) identycznie w sytuacjach, gdy pierwszy argument jest skończony lub niezależny od pewnego miejsca od drugiego. Różnice uwidaczniają się jedynie w przypadku, gdy pierwszy argument jest nieskończony i występują w nim nieskończenie często akcje zależne od przynajmniej jednej akcji występującej w drugim argumencie. W sytuacji śladów zespolonych wychodzimy

wtedy do szerszego świata, ślady Gastina zachowują informację o takim zjawisku w postaci elementu specjalnego \perp , natomiast oparte o zbiory częściowo uporządkowane ślady Kwiatkowskiej pozwalają wpleść maksymalnie duży niezależny prefiks drugiego argumentu w pierwszy. Z punktu widzenia modelowania systemów współbieżnych to podejście wydaje się najciekawsze. Pokażemy teraz, jak wykorzystać zbudowaną w poprzedniej sekcji teorię rekonstrukcji zbiorów rzutowych do zdefiniowania takiej konkatenacji w równoważny sposób:

Definicja 2.5.1 Niech $\alpha, \beta \in \Sigma^\infty / \equiv$ będą śladami. Wówczas ich konkatenacja $\cdot : \Sigma^\infty / \equiv \times \Sigma^\infty / \equiv \rightarrow \Sigma^\infty / \equiv$ określona jest następująco:

$$\alpha \cdot \beta = R_D(\Pi_D(\alpha)\Pi_D(\beta)).$$

Fakt 2.5.2 [36] *Wprowadzona Definicja 2.5.1 jest równoważna definicji Kwiatkowskiej [29], bardziej formalnie:*

$$\begin{aligned} \sup\{\alpha_1\beta_1 \mid \alpha_1 \sqsubseteq_{fin}^{\equiv} \alpha \wedge \beta_1 \sqsubseteq_{fin}^{\equiv} \beta \wedge \beta_1 \text{ jest ogonowo niezależna z } \alpha \text{ po } \alpha_1\} \\ = R_D(\Pi_D(\alpha)\Pi_D(\beta)). \end{aligned}$$

Jedną z największych wad takiej konkatenacji jest jej pokazany w Przykładzie 2.5.3 brak łączności, co między innymi powoduje, że Σ^∞ / \equiv wraz z działaniem \cdot nie tworzą monoidu.

Przykład 2.5.3 Rozważmy (podobnie jak w Przykładzie 2.2.8 następujący alfabet współbieżny ($\Sigma = \{a, b, c\}, I = \{(a, c)\}$) i zbiory rzutowe, będące reprezentacjami rzutowymi śladów $\alpha = [a^\omega]$, $\beta = [b]$ oraz $\gamma = [c]$:

$$\Pi_D(\alpha) = \left\{ \begin{array}{l} (\{a, a\}, a^\omega) \\ (\{a, b\}, a^\omega) \\ (\{b, b\}, \epsilon) \\ (\{b, c\}, \epsilon) \\ (\{c, c\}, \epsilon) \end{array} \right\}, \quad \Pi_D(\beta) = \left\{ \begin{array}{l} (\{a, a\}, \epsilon) \\ (\{a, b\}, b) \\ (\{b, b\}, b) \\ (\{b, c\}, b) \\ (\{c, c\}, \epsilon) \end{array} \right\}, \quad \Pi_D(\gamma) = \left\{ \begin{array}{l} (\{a, a\}, \epsilon) \\ (\{a, b\}, \epsilon) \\ (\{b, b\}, \epsilon) \\ (\{b, c\}, c) \\ (\{c, c\}, c) \end{array} \right\}.$$

Wynikiem konkatenacji tych trzech zbiorów rzutowych, według Definicji 2.2.7, jest zbiór rzutowy

$$P = \left\{ \begin{array}{l} (\{a, a\}, a^\omega) \\ (\{a, b\}, a^\omega) \\ (\{b, b\}, b) \\ (\{b, c\}, bc) \\ (\{c, c\}, c) \end{array} \right\},$$

który nie jest reprezentacją rzutową żadnego śladu. Ponadto, składając te zbiory według Definicji 2.5.1, zależnie od kolejności, otrzymamy różne wyniki

$$\Pi_D(\alpha)(\Pi_D(\beta)\Pi_D(\gamma)) = \left\{ \begin{array}{l} (\{a, a\}, a^\omega) \\ (\{a, b\}, a^\omega) \\ (\{b, b\}, \epsilon) \\ (\{b, c\}, c) \\ (\{c, c\}, c) \end{array} \right\},$$

oraz

$$(\Pi_D(\alpha)\Pi_D(\beta))\Pi_D(\gamma) = \left\{ \begin{array}{l} (\{a, a\}, a^\omega) \\ (\{a, b\}, a^\omega) \\ (\{b, b\}, \epsilon) \\ (\{b, c\}, \epsilon) \\ (\{c, c\}, \epsilon) \end{array} \right\},$$

zatem ta operacja nie jest łączna.

Rozdział 3

Procesy bezkonfliktowe

3.1 Konflikty w sieciach Petriego

W praktycznych zastosowaniach sieci Petriego często mamy do czynienia z sytuacjami niepożądanymi. Jednym z nich jest konflikt [30] między dwoma akcjami sieci. Konflikt taki występuje, kiedy dwie jednocześnie wykonalne akcje nie mogą wykonać się współbieżnie. Wykonanie jednej z nich powoduje uniemożliwienie drugiej.

Konflikty dynamiczne i statyczne w różnych typach sieci. Precyzowanie możliwości i uniemożliwienia pojawiających się w ogólnym opisie sytuacji konfliktowej rozpoczniemy od sieci elementarnych. Przypomnijmy, że uniemożliwienie wykonalnej akcji a może w nich nastąpić zarówno w przypadku zabrania żetonu z jednego z wejść do akcji a jak i poprzez zapełnienie jednego z wyjść z akcji a . Reisig [44] mówi o konflikcie między akcjami posiadającymi wspólne miejsca sąsiadujące, podczas kiedy Starke [47] kładzie duży nacisk na dynamiczne aspekty konfliktu, czyli uniemożliwienie wcześniej umożliwionej akcji. W kontekście Przykładu 3.1.9 wydaje się dość naturalne rozróżnić te dwa typy konfliktów.

Definicja 3.1.1 Niech $N = (P, T, F, M_0)$ będzie elementarną siecią Petriego, $a_1, a_2 \in T$ dwoma różnymi akcjami w tej sieci. Mówimy, że akcje a_1 oraz a_2 są w konflikcie *strukturalnym* (statycznym), jeśli

$$(\bullet a_1 \cup a_1 \bullet) \cap (\bullet a_2 \cup a_2 \bullet) \neq \emptyset.$$

Definicja 3.1.2 Niech N będzie dowolną siecią Petriego o zbiorze akcji T , zaś $a_1, a_2 \in T$ dwoma różnymi akcjami w tej sieci. Mówimy, że akcje a_1 oraz a_2 są w konflikcie *dynamicznym* w osiągalnym markingu M , jeśli

$$M \mathcal{W} a_1 \wedge M \mathcal{W} a_2 \wedge (\neg(M a_1) \mathcal{W} a_2 \vee \neg(M a_2) \mathcal{W} a_1),$$

czyli w markingu M można wykonać zarówno akcję a_1 jak i a_2 , ale niemożliwe jest ich wykonanie jedna po drugiej.

Definicja konfliktu dynamicznego zależy tylko od wykonalności akcji w danym typie sieci i jest wspólna dla wszystkich typów sieci Petriego, stąd słowo dowolna w tej definicji. Nieco inaczej przedstawia się sytuacja konfliktów statycznych. Współdzielenie wyjścia przez dwie akcje przestaje być problemem w p/t-sieciach i ich uogólnieniach. W konflikcie

statycznym w p/t-sieciach znajdują się tylko te akcje, które współdzielą wejście. W sieciach inhibitorowych dochodzą jeszcze sytuacje, gdy wyjście jednej z akcji jest jednocześnie miejscem inhibitorowym dla innej:

Definicja 3.1.3 Niech $N = (P, T, F, M_0)$ będzie p/t-siecią, $a_1, a_2 \in T$ dwoma różnymi akcjami w tej sieci. Mówimy, że akcje a_1 oraz a_2 są w konflikcie strukturalnym, jeśli

$$\bullet a_1 \cap \bullet a_2 \neq \emptyset.$$

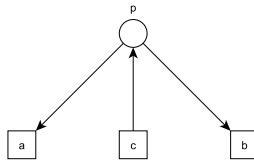
Definicja 3.1.4 Niech $N = (P, T, F, M_0)$ będzie siecią inhibitorową, $a_1, a_2 \in T$ dwoma różnymi akcjami w tej sieci. Mówimy, że akcje a_1 oraz a_2 są w konflikcie strukturalnym, jeśli

$$(\bullet a_1 \cap \bullet a_2) \cup (\bullet a_1 \cap \circ a_2) \cup (\circ a_2 \cap \circ a_1) \neq \emptyset.$$

Uwaga 3.1.5 Zbiory konfliktów strukturalnych we wszystkich trzech typach sieci są skończone i można je konstruktywnie wyznaczyć.

Ponadto w sieciach elementarnych oraz czystych i zwykłych p/t-sieciach konflikty dynamiczne są symetryczne. Jeśli w markingu M akcja a jest w konflikcie z akcją b , to akcja b jest w konflikcie z akcją a . Nie jest to prawdą dla dowolnych p/t-sieci (ze względu na możliwe wagi oraz pętelki) oraz w sieciach inhibitorowych (nawet zwykłych i czystych).

Przykład 3.1.6 W przedstawionej sieci mamy do czynienia z konfliktem strukturalnym między akcjami a oraz b . Niezależnie od typu sieci, może się wykonać akcja c , co owocuje powstaniem konfliktu dynamicznego. Jeśli sieć tę potraktujemy jako p/t-sieć (lub jej rozszerzenie) możemy pozbyć się tego konfliktu wykonując raz jeszcze akcję c , co prowadzi do umieszczenia w miejscu p drugiego żetonu.



Rysunek 3.1: Klasyczny przykład konfliktu

Trwałość i wykrywanie konfliktów. Ścisłe związane z konfliktami są pojęcia sieci bezkonfliktowych oraz trwałych [26, 27, 30]. Pojęcie konfliktów, a właściwie ich braku, przeniesione zostaje w szerszą perspektywę całych obliczeń. Trwałym nazywa się obliczenie, w którym żadna akcja wykonalna nie traci możliwości wykonania aż do momentu faktycznego wykonania. W przypadku sieci elementarnych, oznacza to brak rozwiązywania jakichkolwiek konfliktów. Z kolei trwałą nazywamy sieć, w której konflikty dynamiczne nie występują w ogóle. Wszystkie obliczenia takiej sieci są obliczeniami trwałymi. Nie musi to jednak oznaczać braku występowania konfliktów strukturalnych (patrz Przykład 3.1.9).

Definicja 3.1.7 Niech $S = (P, T, F, M_0)$ będzie dowolną siecią Petriego, zaś $w \in L(S)$ oznacza dowolne obliczenie wykonalne w tej sieci.

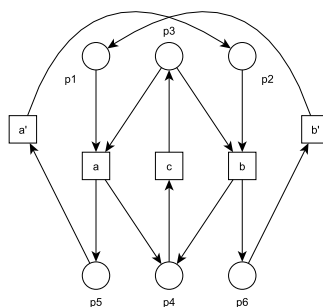
Mówimy, że obliczenie $w = a_1 a_2 \dots$, o kolejnych markingach pośrednich M_1, M_2, \dots , jest trwałe, o ile $\forall_i \forall_{a \in P \setminus \{a_1\}} M_{i-1} \mathcal{W} a_i \wedge M_{i-1} \mathcal{W} a \Rightarrow (M_{i-1} a_i) \mathcal{W} a \wedge (M_{i-1} a) \mathcal{W} a_1$.

Mówimy, że marking M jest *markingiem trwałym*, o ile wykonanie dowolnej akcji umożliwionej w tym markingu jest trwałe, zaś sieć S jest *siecią trwałą*, jeśli każdy marking osiągalny w tej sieci jest trwały [26].

Braku jakichkolwiek konfliktów strukturalnych oczekujemy od sieci bezkonfliktowych. Przy definicji tej klasy sieci, Landweber i Robertson zawężają pojęcie konfliktu strukturalnego w zwykłych p/t-sieciach [30]. Definiują oni występowanie konfliktu strukturalnego w miejscu p będącym wejściem do dwóch różnych akcji tylko w przypadku, gdy choć jedna z tych akcji nie jest połączona z miejscem p pętelką. Uzyskują dzięki temu charakteryzację sieci bezkonfliktowych przy użyciu pojęcia trwałości. My przyjmiemy jednak definicję korzystającą z wprowadzonego wcześniej pojęcia konfliktu strukturalnego:

Definicja 3.1.8 Niech S będzie dowolną siecią Petriego. Sieć tę nazwiemy bezkonfliktową, o ile nie występują w niej żadne konflikty strukturalne (w myśl Definicji 3.1.1, 3.1.3 oraz 3.1.4). W dalszej części mówiąc o sieciach bezkonfliktowych będziemy mieli na myśli bezkonfliktowe p/t-sieci.

Przykład 3.1.9 Rozważmy sieć S składającą się z pięciu akcji oraz sześciu miejsc przedstawioną na Rysunku 3.2. Występuje w niej oczywisty konflikt strukturalny między akcjami a oraz b (w każdym z trzech typów opisanych sieci). Współdzielą one miejsce p_3 . Potraktujmy tę sieć statyczną jako sieć elementarną z różnymi markingami początkowymi.



Rysunek 3.2: Sieć elementarna, której bezkonfliktowość zależy od markingu początkowego

Konflikt strukturalny zamienia się w konflikt dynamiczny w markingu $M_0 = (1, 1, 1, 0, 0, 0)$. Ponieważ już w markingu początkowym mamy do czynienia z konfliktem pomiędzy jedynymi dwoma wykonalnymi akcjami - żadne obliczenie w tej sieci nie będzie bezkonfliktowe.

Rozważmy dalej marking początkowy $M_0 = (1, 0, 1, 0, 0, 0)$. Wówczas jedyną możliwą akcją jest akcja a , po niej współbieżnie wykonać się mogą akcje a' oraz c prowadząc do markingu $M_1 = (0, 1, 1, 0, 0, 0)$. Dalej wymuszona jest akcja b i, symetrycznie, współbieżnie akcje b' oraz c , po czym następuje powrót do markingu początkowego.

Przykład ten pokazuje, że sieć statyczna S w której za marking początkowy przyjęto $M_0 = (1, 0, 1, 0, 0, 0)$ jest siecią trwałą.

W końcu rozważmy marking początkowy $M_0 = (1, 0, 1, 0, 0, 1)$. Wówczas akcje a oraz b nie znajdują się w konflikcie dynamicznym, możemy wykonać je w dowolnej kolejności. Przy przyjęciu zasady, że akcje a, b oraz c są wykonywane w pierwszej kolejności, jeśli tylko są wykonalne, uzyskamy obliczenia bezkonfliktowe w tej sieci. Jednocześnie, wykonanie akcji a' prowadzi do markingu $M = (1, 1, 1, 0, 0, 0)$, w którym występuje znany nam konflikt dynamiczny między akcjami a oraz b .

Reasumując, nawet w przypadku najprostszej klasy sieci Petriego, sieć o identycznej strukturze może, w zależności od markingu początkowego, być trwała, pozwalając na wykonywanie obliczeń trwałych lub zmuszać do rozwiązywania konfliktów w każdym obliczeniu. Podobne wnioski można wysnuć w przypadku pozostałych typów sieci Petriego.

Powyższy przykład pokazuje, że struktura sieci niewiele mówi o konfliktach dynamicznych. Kluczowy staje się tu marking początkowy i dynamika sieci. Precyzując poczynione wcześniej uwagi, możemy jednak sformułować warunek konieczny występowania konfliktów dynamicznych - istnienie w sieci konfliktów strukturalnych.

Lemat 3.1.10 *Niech $S = (P, T, F, M_0)$ będzie elementarną siecią Petriego, $a_1, a_2 \in T$ dwoma różnymi akcjami w tej sieci znajdującymi się w konflikcie dynamicznym w osiągalnym w sieci S markingu M . Wówczas akcje a_1 oraz a_2 znajdują się w konflikcie strukturalnym w sieci S .*

Dowód. Załóżmy, że między akcjami a_1 oraz a_2 nie występuje konflikt statyczny, czyli

$$(\bullet a_1 \cup a_1^\bullet) \cap (\bullet a_2 \cup a_2^\bullet) = \emptyset,$$

choć są one jednocześnie umożliwiające w markingu M .

Wówczas $\bullet a_1 \cap a_2^\bullet = \emptyset$ oraz $\bullet a_2 \cap a_1^\bullet = \emptyset$, czyli $\bullet a_2 \leq M a_1$ oraz żaden nowy żeton nie trafił do wyjścia z a_2 przy wykonywaniu w markingu M akcji a_1 , a zatem $M(a_1) \mathcal{W} a_2$, co przeczy konfliktowi dynamicznemu tych akcji i kończy dowód. \square

Podobne lematy są prawdziwe w przypadku p/t-sieci i sieci inhibitorowych.

Dalsza analiza problemu występowania w sieci konfliktów dynamicznych prowadzi do problemu osiągalności podmarkingu. Konflikt strukturalny między akcjami a oraz b zamienia się w konflikt dynamiczny, kiedy osiągalny jest jeden z podmarkingów pokrywający marking, w którym wykonalne są obie te akcje z ustaloną w jednym z ich wspólnych wejść liczbą tokenów uniemożliwiająca ich jednoczesne wykonanie. Pamiętając o obserwacjach poczynionych przez Landwebera i Robertsona (ważnym elementem statycznym w analizie obliczeń trwałych są pętliki, potrafiące uniemożliwić przekształcenie się konfliktu statycznego w dynamiczny) możemy sformułować formalnie warunki w jakich konflikty statyczne zmieniają się w dynamiczne:

Lemat 3.1.11 (dla sieci elementarnych)

Niech $S = (P, T, F, M_0)$ będzie elementarną siecią Petriego, zaś $a_1, a_2 \in T$ dwiema różnymi akcjami będącymi w konflikcie strukturalnym. Konflikt ten zamieni się w konflikt dynamiczny, jeśli tylko obie te akcje będą kiedykolwiek jednocześnie umożliwiające, czyli

$$\exists M \in [M_0 > (p \in \bullet a_1 \cup \bullet a_2 \Rightarrow M(p) = 1) \wedge (p \in a_1^\bullet \cup a_2^\bullet \Rightarrow M(p) = 0).$$

Lemat 3.1.12 (dla p/t-sieci)

Niech $S = (P, T, W, M_0)$ będzie p/t-siecią, zaś $a_1, a_2 \in T$ dwiema różnymi akcjami będącymi w konflikcie strukturalnym. Konflikt ten zamieni się w konflikt dynamiczny, jeśli osiągalny będzie jeden z markingów M spełniający warunki:

1. $M \leq \max\{W(P, a_1), W(P, a_2)\}$;
2. $\exists_{p \in \bullet_{a_1} \cap \bullet_{a_2}} M(p) < W(a_1, p) + W(a_2, p) - \max\{W(p, a_1), W(p, a_2)\}$.

Wniosek 3.1.13 Ponieważ osiągalność jednego z markingów opisanych w Lemacie 3.1.12 sprowadza się do rozstrzygnięcia kilku instancji problemu osiągalności pokrytego podmarkingu (O4), który jest rozstrzygalny w p/t-sieciach, problem zamiany konfliktu strukturalnego na dynamiczny jest w nich również rozstrzygalny.

Wniosek 3.1.14 Trwałość p/t-sieci jest rozstrzygalna.

Inne dowody tego faktu można znaleźć w [6, 22, 33]. O ile w pierwszym z cytowanych źródeł pomysł zawarty w dowodzie jest podobny do przedstawionego rozumowania, to w pozostałych przedstawiono dowody, które nie wykorzystują rozstrzygalności problemu osiągalności markingów (O1) w p/t-sieciach.

W dalszej części tej sekcji przyjrzymy się kilku klasycznym własnościom bezkonfliktowych p/t-sieci. W cytowanym już artykule [30] Landweber i Robertson udowodnili, że zbiór markingów osiągalnych w sieciach bezkonfliktowych (o szerszej niż nasza definicji) jest zawsze zbiorem półliniowym, czyli

Fakt 3.1.15 Niech $S = (P, T, W, M_0)$ będzie bezkonfliktową p/t-siecią. Wówczas zbiór markingów osiągalnych w tej sieci można zapisać jako sumę zbiorów liniowych postaci:

$$\left\{ X_0 + \sum_{i=1}^n a_i X_i \mid a_i \geq 0 \wedge X_i \in \mathbb{N}^P \right\},$$

czyli jest on zbiorem półliniowym.

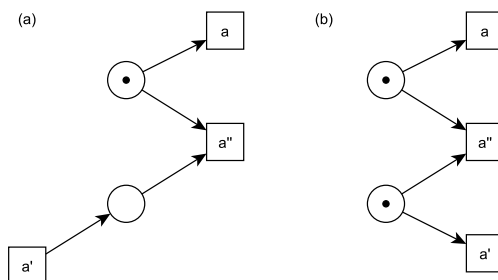
Ponadto, uzasadniają oni, że zbiory te są efektywnie wyznaczalne, a to z kolei daje rozstrzygalność problemu równości tych zbiorów dla dwóch sieci bezkonfliktowych [18].

Na koniec klasycznych rozważań na temat konfliktów w sieciach Petriego podamy jeszcze definicje ściśle związanej z konfliktami dynamicznymi *konfuzji* oraz sieci wolnego wyboru, w których konflikty strukturalne dotyczą całych zbiorów wejść a nie tylko pojedynczych akcji.

Pojęcie konfuzji bardzo często występuje wraz z dyskusją na temat współbieżności i konfliktów. Zazwyczaj konfuzja opisywana jest jako zjawisko na granicy tych dwóch pojęć. Wielu autorów, w tym Reisig [44] i Starke [47], ogranicza się jednak do wskazania przykładów konfuzji bez podawania precyzyjnej definicji. My oprzemy się na definicji jaką można znaleźć u Rosenberga i Thiagarajana [45].

Definicja 3.1.16 Niech S będzie siecią Petriego, T zbiorem akcji tej sieci, M ustalonym markingiem osiągalnym w tej sieci zaś $a, b \in T$ dwoma współbieżnymi (niezależnymi w sensie opisanym w Rozdziale 1) akcjami umożliwiającymi w tym markingu. Mówimy, że

akcja a znajdują się w markingu M w konfuzji z akcją b o ile zbiory skonfliktowanych z nią akcji w markingach M oraz Mb są różne.



Rysunek 3.3: Na przedstawionych przykładach konfuzji [47] mamy do czynienia z sytuacją (a) w której może pojawić się nowy konflikt oraz (b), gdzie konflikt może zniknąć

Przykład 3.1.17 Na Rysunku 3.3 przedstawiono sieci w których akcje a oraz a' są niezależne i umożliwiające, czyli mogą się wykonać współbieżnie. W obu przypadkach akcja a jest w konflikcie statycznym z akcją a'' . W sieci (a) wykonanie akcji a' przed akcją a powoduje, że statyczny konflikt między akcjami a oraz a'' zamienia się w dynamiczny, natomiast w drugim przypadku wykonanie akcji a' w pierwszej kolejności usuwa konflikt dynamiczny między akcjami a i a'' . Jednakże akcja a' przed wykonaniem sama była w konflikcie z akcją a'' . Zauważmy, że w p/t-sieci z Rysunku 3.1 mamy do czynienia z kolejnym typem konfuzji usuwającej konflikt opisany w Przykładzie 3.1.6.

Pojęcie sieci wolnego wyboru, którego zdefiniowaniem zakończymy tę sekcję, ewoluowało w czasie. My, za Desalem i Esparzą [14], posłużymy się definicją z roku 1995. Wcześniej tego typu sieci nazywano rozszerzonymi sieciami wolnego wyboru (*ang. extended free-choice nets*).

Definicja 3.1.18 Niech S będzie zwykłą siecią Petriego, zaś T - zbiorem akcji tej sieci. Sieć S nazywać będziemy *siecią wolnego wyboru*, jeśli dla dowolnych dwóch akcji $a, b \in T$ zbiory ich wejść są rozłączne lub pokrywają się. Bardziej formalnie:

$$\forall_{a,b \in T} (\bullet a \cap \bullet b = \emptyset) \vee (\bullet a = \bullet b).$$

3.2 Sieci pokojowe

W kontekście unikania sytuacji konfliktowych, w literaturze [18, 23, 30, 38] rozpatrywane są zazwyczaj sieci trwałe lub bezkonfliktowe, czyli takie, w których każde obliczenie jest trwałe. W tej sekcji zajmiemy się trwałymi obliczeniami w dowolnych p/t-sieciach. Już Przykład 3.1.9 pokazuje, że problem ten nie jest trywialny i mamy do czynienia z wieloma możliwymi przypadkami, nawet w sieciach o identycznej strukturze statycznej.

Umożliwienie a wykonalność Modyfikacje klasycznych p/t-sieci odbywają się zazwyczaj przez umożliwiające dodatkowe akcje rozszerzenia (sieci inhibitorowe, czyszczące, podwajające czy samomodyfikujące). Dodawane są wtedy nowe obiekty (relacje), dla których zmieniona zostaje zasada umożliwienia pewnych akcji. Wszystkie umożliwione akcje są w tych przypadkach wykonalne. W znikomych sytuacjach modyfikacja taka odbywa się przez redukcję możliwości p/t-sieci i zmianę reguły wykonalności akcji. Przykładem są maksymalne kroki współbieżne Burkharda [7, 47]. Tutaj, spośród zbioru wszystkich umożliwionych akcji, wybieramy jeden z maksymalnych współbieżnych podzbiorów i wykonujemy go w całości. Uzyskujemy w ten sposób dość zaskakujący wynik - ograniczenie zbioru wszystkich osiągalnych markingów na poziomie pojedynczych sieci powoduje zwiększenie mocy obliczeniowej modelu. Sieci Burkharda są w sensie obliczeniowym równoważne maszynom Turinga. Wydaje się jednak, że tracą wiele w kontekście modelowania współbieżności. Wykonalność akcji przestaje mieć charakter lokalny - trzeba znać stan całej sieci.

Zaproponujemy teraz inną modyfikację p/t-sieci, która nie zmienia jej struktury ale wyłącznie zasadę wykonalności, pozostawiając przy tym lokalny charakter tej zasady. Koncepcja tej modyfikacji ma źródła w modelowaniu trwałych obliczeń w p/t-sieciach. Umożliwienie traktować będziemy analogicznie jak w przypadku klasycznych p/t-sieci, wyznaczać je będzie odpowiednia liczba żetonów w wejściach do akcji. Zmianie natomiast ulegnie, utożsamiana zazwyczaj z umożliwieniem, wykonalność akcji. Wykonalne będą tylko trwałe akcje, to jest takie, które swoim wykonaniem nie uniemożliwią żadnej innej. Ze względu na bezkonfliktowe zasady działania, sieci takie nazwiemy *pokojowymi* [4].

Definicja 3.2.1 *Pokojową siecią Petriego* nazywamy sieć o strukturze p/t-sieci (czyli czwórkę $S = (P, T, W, M_0)$), gdzie podobnie jak w przypadku klasycznym, marking M jest w relacji \mathcal{U} z akcją a (mówimy, że akcja a jest umożliwiona w markingu M) wtedy i tylko wtedy, gdy

$$W(P, a) \leq M,$$

czyli we wszystkich wejściach do akcji a jest wystarczająco dużo żetonów.

Mówimy natomiast, że marking M jest w relacji wykonalności \mathcal{W} z akcją a wtedy i tylko wtedy, gdy

$$M \mathcal{U} a \wedge \forall_{b \neq a} M \mathcal{U} b \Rightarrow (Ma) \mathcal{U} b,$$

czyli każda różna od a , ale każda umożliwiona w markingu M akcja pozostaje umożliwiona po wykonaniu akcji a w p/t-sieci o identycznej jak sieć S strukturze. Analogicznie do p/t-sieci zdefiniowany jest krok oraz obliczenie w sieci pokojowej, a także zbiór markingów osiągalnych pokojowo z markingu M (oznaczany przez $\underline{[M >]}$).

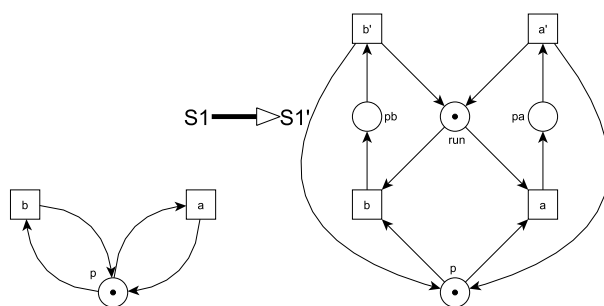
Tak rozumiany warunek wykonalności może wydawać się równie globalny jak w przypadku sieci Burkharda. Widzimy jednak, że w sieciach pokojowych, w odróżnieniu od sieci Burkharda, wykonalność akcji zależy wyłącznie od akcji będących z nimi w konflikcie strukturalnym, a więc topologicznych sąsiadów. Stąd lokalność zasady wykonalności w sieciach pokojowych.

Zauważmy, że wprost z definicji, zbiory markingów osiągalnych w sieciach pokojowych są podzbiorem markingów osiągalnych w p/t-sieciach o identycznej strukturze (czyli $\underline{[M_0 >]} \subseteq [M_0 >]$). Nasuwa się naturalne pytanie: *W jakich sytuacjach zachodzi równość tych zbiorów?* Dość oczywisty wydaje się:

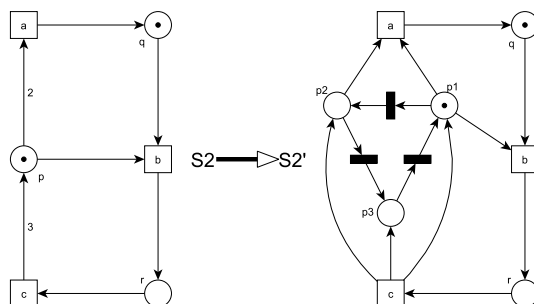
Fakt 3.2.2 Jeśli p/t -sieć $S = (P, T, W, M_0)$ jest siecią trwałą, to zbiór markingów osiągalnych w tej sieci jest równy zbiorowi markingów osiągalnych w odpowiadającej jej sieci pokojowej.

Implikacja w drugą stronę nie jest jednak prawdziwa, co pokazuje Przykład 3.1.6. Sieć przedstawiona w tym przykładzie w sposób oczywisty nie jest siecią trwałą, jednakże wszystkie osiągalne w niej markingi mogą zostać osiągnięte w sposób bezkonfliktowy.

Sieci bez wag i pęteli W kontekście sieci bezkonfliktowych i pokojowych nowego znaczenia nabierają problemy zwyczajności oraz czystości. Zauważmy, że przytoczone w Konstrukcjach 1.2.12 i 1.2.8 klasyczne algorytmy usuwające pętli oraz wagi wprowadzają nowe konflikty. W wielu przypadkach (patrz Przykład 3.2.3) powoduje to nawet, że żadne obliczenie wynikowej sieci nie jest trwałe, mimo że za punkt wyjścia wzięliśmy sieć bezkonfliktową.



Rysunek 3.4: Trwała p/t -sieć z pętelkami, klasyczne usunięcie pęteli powoduje w stanie początkowym powstanie konfliktu dynamicznego między akcjami a oraz b



Rysunek 3.5: Trwała p/t -sieć z wagami, klasyczne usunięcie wag powoduje w stanie początkowym powstanie konfliktu dynamicznego między akcją b a jedną z akcji w pierścieniu powstałym z miejsca p

Przykład 3.2.3 Przedstawione modyfikacje sieci trwałych S_1 oraz S_2 (sieć S_1 jest nawet bezkonfliktowa w rozumieniu Landwebera i Robertsona) pokazują, że zastosowanie Konstrukcji 1.2.12 oraz 1.2.8 do usunięcia odpowiednio pęteli oraz wag dają w wyniku sieci,

w których trzeba rozwiązać konflikt już w stanie początkowym. Wynika to z faktu, że koncepcje miejsca *run* oraz zamiany miejsca na zbiór miejsc połączonych dodatkowymi akcjami w cykl nie zachowują trwałości. Są więc one bezużyteczne w przypadku sieci pokojowych!

3.3 Problemy decyzyjne w sieciach pokojowych

W sekcji tej rozważymy zdefiniowane wcześniej (Podsekcja 1.2.4) problemy decyzyjne związane z markowaniem, tym razem w kontekście sieci pokojowych. Na początek uzasadnimy równoważność decyzyjną wszystkich 7 przedstawionych problemów, starając się tam gdzie to możliwe, zadbać o czystość i zwykłość podanych konstrukcji. Następnie uzasadnimy nierozstrzygalność problemu osiągalności oraz problemu osiągalności pustego markingu w zwykłych i czystych sieciach pokojowych przez redukcję do tych problemów odpowiadających im problemów w sieciach inhibitorowych. Wraz z wcześniej poczynionymi uwagami, da to nierozstrzygalność wszystkich 7 problemów decyzyjnych związanych z markowaniem, zaś po dodaniu przedstawionej dalej konstrukcji pozwalającej symulować obliczenia w sieciach pokojowych przez sieci inhibitorowe, uzasadni to równoważność obliczeniową sieci pokojowych i maszyn Turinga.

Struktura problemów związanych z markowaniem

Przypomnijmy zdefiniowane formalnie w Sekcji 1.2.4 problemy decyzyjne związane z markowaniem sieci Petriego. Podzieliliśmy je na dwie klasy. Do pierwszej klasy powiązanej z pokrywalnością markingu zaliczyliśmy Problem Pokrywalności, Problem Niepustości Miejsca oraz Problem Wykonalności Akcji. Klasę tę określiliśmy jako prostszą - problemy te są rozstrzygalne w p/t-sieciach oraz wielu jej rozszerzeniach, natomiast nierozstrzygalne w sieciach inhibitorowych. Tylko w ostatniej z przytoczonych klas sieci potrafimy uzasadnić konstruktywnie równoważność decyzyjną tych problemów z problemami z drugiej klasy - problemów związanych z osiągalnością. Do tej klasy problemów zaliczamy Problem Osiągalności, Problem Pustości Stanu, Problem Pustości Miejsca oraz Problem Osiągalności Pokrytego Podmarkingu. Uzasadnienie rozstrzygalności problemów z tej klasy dla p/t-sieci zajęło więcej czasu niż uzasadnienie rozstrzygalności pokrywalności. Przy tym dla sieci resetujących oraz podwajających są to problemy nierozstrzygalne (przy rozstrzygalności problemów związanych z pokrywalnością).

W Rozdziale 1 dużo miejsca poświęciliśmy na uzasadnienie równoważności rozważanych problemów decyzyjnych w p/t-sieciach oraz sieciach inhibitorowych. Część z przytoczonych tam argumentów i konstrukcji daje się bez większych przeszkód przenieść na przypadek sieci pokojowych. Są to, w kolejności chronologicznej, zależności oznaczone w Rozdziale 1 jako:

- $P1$ rozstrzygalny $\Rightarrow P3$ rozstrzygalny;
- $P3$ rozstrzygalny $\Rightarrow P1$ rozstrzygalny;
- $P3$ rozstrzygalny $\Rightarrow P2$ rozstrzygalny;
- $P2$ rozstrzygalny $\Rightarrow P3$ rozstrzygalny;

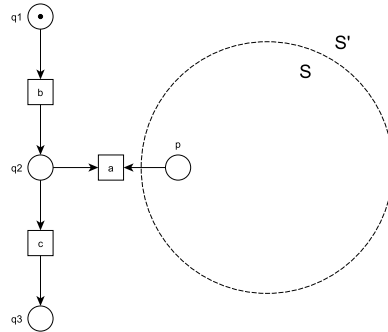
- $O4$ rozstrzygalny $\Rightarrow O1$ rozstrzygalny;
- $O1$ rozstrzygalny $\Rightarrow O2$ rozstrzygalny;
- $O3$ rozstrzygalny $\Rightarrow O2$ rozstrzygalny;

Tylko w przypadku drugiej i ostatniej konstrukcji, użyliśmy wag innych niż 0 lub 1 oraz pętelek, wobec czego pozostałych pięć zależności możemy uzasadnić również w czystych i zwykłych sieciach pokojowych.

Dostosowanie konstrukcji uzasadniającej rozstrzygalność problemu Pustości Miejsca ($P3$ rozstrzygalny $\Rightarrow O3$ rozstrzygalny) w sieciach inhibitorowych, pod warunkiem, że rozstrzygalny jest Problem Niepustości Miejsca wymaga użycia specjalnych własności sieci pokojowych:

Konstrukcja 3.3.1 Niech $S = (P, T, W, M_0)$ będzie dowolną pokojową siecią Petriego z wyróżnionym miejscem p . Skonstruujemy sieć pokojową $S' = (P', T', W', M'_0)$ tak, aby marking w którym miejsce p byłoby puste był osiągalny w sieci S dokładnie wtedy, gdy w sieci S' osiągalny będzie marking, w którym odpowiednie miejsce jest niepuste. Konstrukcję tę ilustruje Rysunek 3.6.

Do sieci S dodajmy trzy nowe miejsca - q_1, q_2 i q_3 oraz trzy nowe akcje - a, b i c . Wejściami do akcji a są miejsca p oraz q_2 . Miejsce q_2 jest ponadto wyjściem z akcji b oraz wejściem do akcji c . Miejsce q_1 jest jedynym wejściem do akcji b , zaś miejsce q_3 jest jedynym wyjściem z akcji c . Wartości markingu początkowego M'_0 na miejscach ze zbioru P są równe wartościom markingu M_0 w tych miejscach. Natomiast w trzech nowych miejscach marking ten przyjmuje odpowiednio wartości $M'_0(q_1) = 1$ oraz $M'_0(q_2) = M'_0(q_3) = 0$.



Rysunek 3.6: Konstrukcja pokojowej sieci S' na podstawie pokojowej sieci S . Osiągalność markingu w którym miejsce p jest puste w sieci S jest równoważna osiągalności markingu w którym miejsce q_3 jest niepuste

W tak skonstruowanej sieci S' akcje b oraz c mogą się wykonać co najwyżej raz. Ich wykonanie spowoduje przerzucenie żetonu z miejsca q_1 do miejsca q_3 . Przeszkodą w tej operacji może okazać się akcja a . Akcja ta, choć może być umożliwiona w sytuacji gdy miejsca q_2 oraz p będą niepuste, to nigdy nie będzie wykonalna - w miejscu q_2 znajdować się może najwyżej jeden żeton, w związku z czym wykonanie akcji a uniemożliwi akcję c , co jest niedopuszczalne w sieciach pokojowych. Z tego powodu nowe akcje mają znikomy wpływ na zachowanie się sieci S' w obrębie jej pierwotnej części S . Umożliwienie akcji a

może co najwyżej uniemożliwić zabranie z miejsca p żetonu. Na podobnej zasadzie akcja c może się wykonać tylko wtedy, gdy miejsce p jest puste (akcja a nie będzie wówczas umożliwiona).

Reasumując, jeśli w sieci S osiągalny jest marking M , w którym miejsce p jest puste, to w sieci S' można wykonać identyczne akcje, powstrzymując się do uzyskania markingu M w miejscach ze zbioru P z wykonaniem akcji b , a następnie wykonać obliczenie bc przerzucające żeton z miejsca q_1 do miejsca q_2 . Z drugiej strony, jeśli w sieci S' osiągalny jest marking w którym miejsce q_3 jest niepuste, to musiała zostać wykonana akcja c . Oznacza to, że w pewnym markingu w sieci S' miejsce p było puste. Wykonując identyczne obliczenie (z dokładnością do usunięcia akcji b) uzyskamy w sieci S marking, w którym miejsce p jest puste.

Ostatecznie, marking w którym miejsce p jest puste jest osiągalny w sieci S dokładnie wtedy, gdy w sieci S' osiągalny jest marking w którym miejsce q_3 jest niepuste.

Konstrukcja 3.3.1, choć dość prosta, jest ciekawa ze względu na inhibitorowy charakter akcji a . Bez użycia łuku inhibitorowego aby wykonać akcję c , musieliśmy pozytywnie przejść test zera w miejscu p . Takiego zachowania nie da się wymusić w klasycznych p/t-sieciach. Ponadto, akcji a nie daje się zakwalifikować jako bezużytecznej, mimo że nigdy nie zostanie ona wykonana. Uzasadnia to konsekwencję z jaką w Rozdziale 1 wiązaliśmy bezużyteczność akcji z brakiem umożliwienia (a nie z brakiem wykonalności).

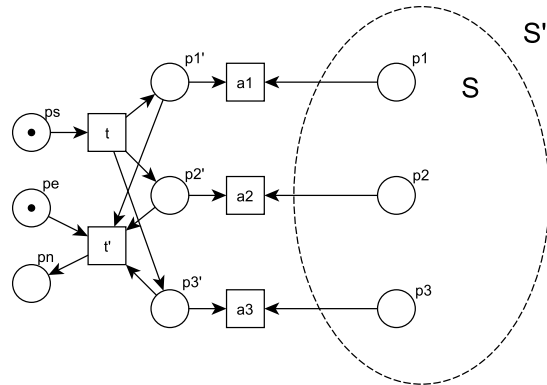
Problemy w pozostałych przypadkach wynikają głównie z zastosowania globalnego miejsca run , czyli koncepcji wyłączania sieci. Pozwala to przeprowadzić różnego rodzaju testy na zamrożonym markingu. Jak pokazuje Przykład 3.2.3, konstrukcje tego typu nie zachowują trwałości obliczeń, w związku z czym nie można ich użyć dla sieci pokojowych.

Trzy kolejne zależności, sprowadzające rozstrzygalność Problemu Pustości Stanu do rozstrzygalności Problemu Niepustości Miejsca, Problemu Pustości Miejsca lub Problemu Wykonalności Akcji, można uzasadnić korzystając z inhibitorowych własności sieci pokojowych [25].

Konstrukcja 3.3.2 Niech $S = (P, T, W, M_0)$ będzie dowolną pokojową siecią Petriego. Skonstruujemy sieć pokojową $S' = (P', T', W', M'_0)$ tak, aby stan pusty w sieci S był osiągalny dokładnie wtedy, gdy odpowiednia akcja będzie wykonalna w sieci S' . Równoważnie będzie wówczas osiągalny marking w którym odpowiednie miejsce w sieci S' będzie niepuste, a inne puste. Tę konstrukcję ilustruje Rysunek 3.7.

Rozszerzmy sieć S dodając dla każdego miejsca $p_i \in P$ parę (a_i, p'_i) , gdzie a_i jest akcją dla której jedynymi wejściami są miejsca p_i oraz p'_i . Ponadto dodajmy dwie akcje t oraz t' i trzy miejsca p_s, p_e oraz p_n . Niech nowe miejsca p'_i będą wyjściami z akcji t oraz wejściami do akcji t' . Miejsce p_s będzie jedynym wejściem do akcji t , miejsce p_e - wejściem do akcji t' zaś miejsce p_n jedynym wyjściem z tej akcji. Marking początkowy M'_0 w miejscach ze zbioru S jest identyczny z markingiem M_0 , w pozostałych miejscach przyjmuje on następujące wartości: $M'_0(p_s) = M'_0(p_e) = 1$ oraz $M'_0(a_i) = M'_0(p_n) = 0$ dla dowolnego i .

W tak skonstruowanej sieci pokojowej S' akcje t oraz t' mogą wykonać się co najwyżej po razie. Ponadto akcja t' może się wykonać dopiero po wykonaniu akcji t , zaś do jej wykonania, sieć S' zachowuje się dokładnie tak, jak sieć S . Żadna z akcji a_i nigdy nie ma szansy się wykonać. Każda z nich posiada dokładnie dwa wejścia, nowododane wejście p'_i jest na początku puste, wypełnione może zostać dopiero po wykonaniu akcji t . Jednocześnie umożliwiona zostaje wtedy akcja t' . Po wykonaniu którejkolwiek z akcji



Rysunek 3.7: Konstrukcja pokojowej sieci S' na podstawie pokojowej sieci S . Marking pusty jest osiągalny w sieci S dokładnie wtedy, gdy w sieci S' wykonalna jest akcja t'

a_i umożliwienie akcji t' zostałyby przerwane - co jest zakazane w sieciach pokojowych. Z kolei, z podobnych powodów, akcja t' może się wykonać tylko wtedy, gdy żadna z akcji a_i nie jest umożliwiona, czyli gdy wszystkie miejsca ze zbioru P są puste.

Ostatecznie więc stan pusty jest osiągalny w sieci pokojowej S dokładnie wtedy akcja t' jest wykonalna w sieci pokojowej S' . Wykonalność tej akcji jest z kolei równoważna możliwości usunięcia żetonu z miejsca p_e oraz zapełnienia miejsca p_n - miejsce p_e jest wejściem tylko do akcji t' , podobnie akcja p_n jest wyjściem tylko z tej akcji.

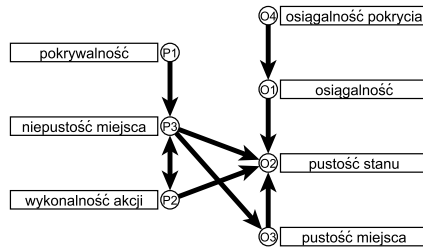
W związku z tym, następujące warunki są równoważne:

- w sieci S marking pusty jest osiągalny,
- w sieci S' wykonalna jest akcja t' ,
- w sieci S' osiągalny jest marking w którym miejsce p_e jest puste,
- w sieci S' osiągalny jest marking w którym miejsce p_n jest niepuste.

Podsumujmy teraz zależności między siedmioma problemami związanymi z markingami w czystych i zwykłych sieciach pokojowych. Przedstawione są one na Rysunku 3.8. Łatwo zauważyć, że rozstrzygalność któregośkolwiek z tych problemów pociągałaby za sobą rozstrzygalność Problemu Pustości Stanu. W kolejnej podsekcji pokażemy, że Problemu Pustości Stanu jest nierozstrzygalny w pokojowych sieciach Petriego.

3.3.1 Nierozstrzygalność problemu osiągalności pustego markingu

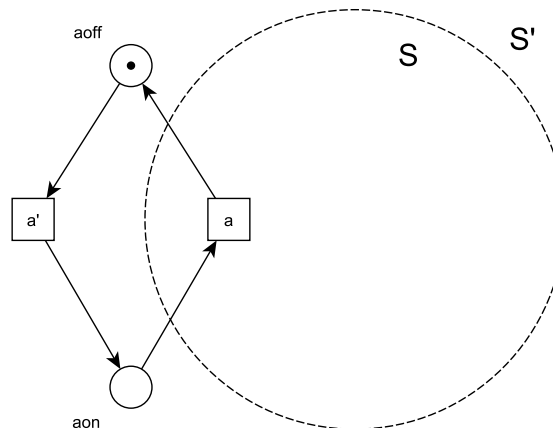
Przypomnijmy, że w sieciach inhibitorowych wszystkie problemy związane z markingiem są konstruktywnie równoważne, a więc nierozstrzygalne (Rysunek 1.17). W szczególności należy do nich Problem Pustości Stanu. Pokażemy teraz w jaki sposób można symulować działanie sieci inhibitorowych przy pomocy sieci pokojowych. Konstrukcję tę przeprowadzimy w trzech krokach. Najpierw pokażemy jak symulować przy pomocy sieci pokojowych klasyczne p/t-sieci (krok 1). Następnie rozszerzymy tę konstrukcję na sieci inhibitorowe



Rysunek 3.8: Struktura problemów związanych z markowaniem w zwykłych i czystych sieciach pokojowych. Krawędź skierowana z wierzchołka odpowiadającego problemowi A do wierzchołka odpowiadającego problemowi B oznacza, że rozstrzygalność problemu A implikuje konstruktywnie rozstrzygalność problemu B

(krok 2). W końcu dodamy odpowiednie mechanizmy, które pozwalałyby rozstrzygać Problem Pustości Stanu w sieciach inhibitorowych, gdybyśmy potrafili rozstrzygać ten problem w sieciach pokojowych (krok 3).

Krok 1 - symulacja p/t-sieci (Rysunek 3.9). Niech $S = (P, T, W, M_0)$ będzie dowolna p/t-siecią. Na jej podstawie utworzymy sieć pokojowa $S' = (P', T', W', M'_0)$, która symuluje w sensie osiągalności markingu działanie sieci S . Jeśli potraktujemy sieć S jako sieć pokojową (zmieniając regułę wykonalności akcji) nie wszystkie akcje wykonalne w danym markingu w sieci S muszą być wykonalne w sposób trwały, a więc pokojowo. Oznacza to, że nie są one wykonalne w tym markingu w sieci S' . Aby poradzić sobie z tym zjawiskiem, opóźnimy wykonanie każdej akcji dodając specjalną pętlę przygotowującą daną akcję do wykonania. Do każdej akcji $a \in T$ dodajemy nową akcję a' oraz dwa miejsca a_{on} i a_{off} . Pierwsze z nich jest wejściem do akcji a i wyjściem z akcji a' , zaś drugie na odwrót, wejściem do akcji a' i wyjściem z akcji a . W markingu początkowym w miejscu a_{off} znajduje się jeden żeton, miejsce a_{on} jest puste. Markingi w sieci S' , w których wszystkie miejsca ze zbioru P zamarkowane są zgodnie z markowaniem M , miejsca a_{on} są puste zaś a_{off} zamarkowane pojedynczym żetonem oznaczają $10M$. W szczególności, $M'_0 = 10M_0$.

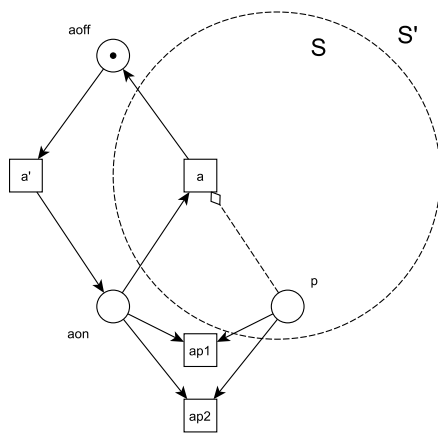


Rysunek 3.9: Dodanie pętli opóźniającej do akcji a

Przy takiej konstrukcji sieci S' , dla każdego obliczenia $u = a_1 \dots a_n$ wykonalne w sieci S istnieje obliczenie dualne $u' = a'_1 a_1 \dots a'_n a_n$ wykonalne w sieci S' . Wykonanie zbyt szybko akcji primowanej a' (bez natychmiastowego wykonania po niej przygotowanej akcji a) może co najwyżej uniemożliwić wykonanie pewnych akcji, nie wpływa natomiast bezpośrednio na miejsca oryginalnej sieci S . W związku z tym, zachodzi dość oczywisty

Fakt 3.3.3 Niech $S = (P, T, W, M_0)$ będzie dowolną p/t-siecią, zaś $S' = (P', T', W', M'_0)$ będzie siecią pokojową skonstruowaną zgodnie z opisaną w Kroku 1 procedurą. Wówczas marking M jest osiągalny w sieci S dokładnie wtedy, gdy marking $10M$ jest osiągalny w sieci S' .

Krok 2 - symulacja sieci inhibitorowych (Rysunek 3.10). Niech S będzie dowolną siecią inhibitorową. Pokażemy jak na podstawie sieci inhibitorowej $S = (P, T, W, I, M_0)$ skonstruować sieć pokojową $S' = (P', T', W', M'_0)$ tak, aby dowolny marking M był osiągalny w sieci S dokładnie wtedy, gdy odpowiadający mu marking $10M$ jest osiągalny w sieci S' (marking $10M$ określamy tak jak w Kroku 1). Konstrukcja ta oparta jest o przedstawioną w kroku pierwszym ideę opóźniania akcji. W tym przypadku musimy sobie poradzić z łukami inhibitorowymi, które nie mogą występować w sieci pokojowej. Łuki te zastąpimy specjalnymi akcjami inhibitorowymi, które choć same nigdy nie będą wykonalne, to ich umożliwienie uniemożliwia będzie wykonanie akcji z oryginalnej sieci S . Podobnie jak w przypadku p/t-sieci dodajemy pętle opóźniające. Ponadto każde wejście inhibitorowe p do akcji a łączymy z miejscem a_{on} dwoma akcjami ap_1 i ap_2 tak, aby oba te miejsca były wejściami do tych akcji.



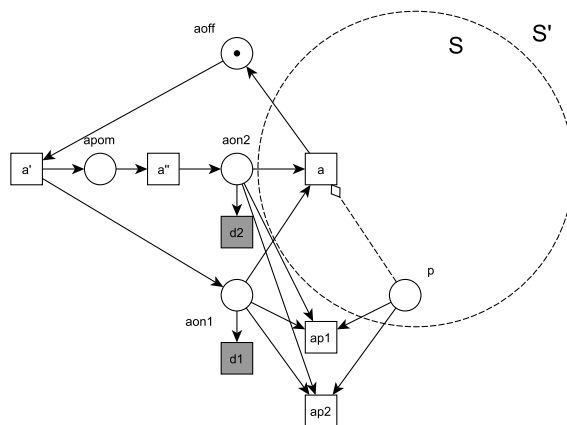
Rysunek 3.10: Dodanie pętli opóźniającej oraz wymiana łuku inhibitorowego (p, a) na akcje inhibitorowe ap_1 i ap_2

W tak skonstruowanej sieci pokojowej przygotowanie do wykonania akcji uniemożliwionej z powodu zamarkowania któregoś z wejść inhibitorowych owocuje jednoczesnym umożliwieniem odpowiednich akcji inhibitorowych, to zaś powoduje uniemożliwienie wykonania właściwej akcji. W przypadku gdy wszystkie wejścia inhibitorowe są puste, a wejścia zwykle zawierają odpowiednią liczbę żetonów, akcja przygotowana może zostać natychmiast wykonana. Dzięki temu, podobnie jak w Kroku 1, dla każdego obliczenia $u = a_1 \dots a_n$ wykonalne w sieci S istnieje obliczenie dualne $u' = a'_1 a_1 \dots a'_n a_n$ wykonalne w sieci S' .

Ponownie odejście od tego schematu spowodować może co najwyżej uniemożliwienie wykonania pewnych akcji (niektórych w sposób stały, przez uwięzienie żetonów w ich wejściach inhibitorowych), nie wpływa zaś w sposób bezpośredni na zawartość miejsc ze zbioru P . W związku z tym prawdziwy jest

Fakt 3.3.4 *Niech $S = (P, T, W, I, M_0)$ będzie dowolną siecią inhibitorową, zaś $S' = (P', T', W', M'_0)$ będzie siecią pokojową skonstruowaną zgodnie z opisaną w Kroku 2 procedurą. Wówczas marking M jest osiągalny w sieci S dokładnie wtedy, gdy marking $10M$ jest osiągalny w sieci S' .*

Krok 3 - Problem Pustości Stanu (Rysunek 3.11). Pokazana w Fakcie 3.3.4 zależność pozwala przypuszczać, że omawiane w tym rozdziale problemy decyzyjne związane z markingiem są nierozstrzygalne w sieciach pokojowych. Pokażemy ten fakt precyzyjnie. Przypomnijmy, że naszkicowane na Rysunku 3.8 zależności między różnymi problemami decyzyjnym w sieciach pokojowych pozwalają skupić się na Problemie Pustości Stanu. Jego nierozstrzygalność oznaczać będzie nierozstrzygalność wszystkich siedmiu rozważanych problemów.



Rysunek 3.11: Rozmnożenie miejsca aktywującego a_{on} i dodanie globalnych akcji czyszczących

Rozważmy dowolną sieć inhibitorową $S = (P, T, W, I, M_0)$, w której występują przynajmniej dwie różne akcji (przypadek z jednoelementowym zbiorem T jest trywialny, choć przedstawione rozumowanie nie jest dla niego poprawne). Na jej podstawie skonstruujemy sieć pokojową $S' = (P', T', W', M'_0)$ w taki sposób, aby stan pusty (wszystkie miejsca puste) był osiągalny w sieci S' dokładnie wtedy, gdy stan pusty jest osiągalny w sieci S . Osiągniemy to przez rozbudowanie sieci pokojowej przedstawionej w Kroku 2 o mechanizm pozwalający wyczyścić wszystkie dodatkowe miejsca w dowolnym markingu $10M$, bez utraty właściwości tej sieci. W tym celu rozmnożymy miejsca aktywujące a_{on} tworząc dwie gałęzie, dodatkowa gałąź zaopatrzona jest w mechanizm opóźniający w postaci akcji a'' . Ponadto, dodamy dwie globalne akcji d_1 oraz d_2 , dla których odpowiednio wszystkie miejsca a_{on1} i a_{on2} są wejściami. Przez marking $10M$ rozumiemy marking, w którym wszystkie miejsca ze zbioru P zamarkowane są zgodnie z markingiem M , wszystkie miejsca aktywujące (a_{on1} , a_{on2} , a_{pom}) są puste, zaś w miejscach a_{off} znajduje się po jednym żetonie. Ponownie $M'_0 = 10M_0$.

Przy takiej konstrukcji w mocy pozostaje podstawowa własność - dla każdego obliczenia $u = a_1 \dots a_n$ wykonalne w sieci S istnieje obliczenie dualne $u' = a'_1 a''_1 a_1 \dots a'_n a''_n a_n$ wykonalne w sieci S' . Ponadto, w sposób identyczny do opisanego w Kroku 2 działają akcje inhibitorowe ap_1 i ap_2 . Korzystając z podobnej argumentacji uzasadniamy

Fakt 3.3.5 *Niech $S = (P, T, W, I, M_0)$ będzie dowolną siecią inhibitorową, zaś $S' = (P', T', W', M'_0)$ będzie siecią pokojową skonstruowaną zgodnie z opisaną w Kroku 3 procedurą. Wówczas marking M jest osiągalny w sieci S dokładnie wtedy, gdy marking $10M$ jest osiągalny w sieci S' .*

Założmy teraz, że sieć S' znajduje się w dowolnym markingu $10M$. Wykonując po kolei wszystkie akcje a' usuwamy żetony z miejsc a_{off} zapelniając miejsca a_{pom} oraz a_{on1} . Zauważmy, że zanim wykonamy którąkolwiek z akcji bisowanych, umożliwiona i wykonalna jest akcja d_1 . Jej umożliwienie powoduje zatrzymanie obliczeń w sieci S (stąd założenie o odpowiedniej wielkości zbioru T). Możemy ją jednak wykonać, gdyż żadna z akcji potrzebujących żetonu w którymkolwiek z miejsc a_{on1} nie jest umożliwiona. Po wykonaniu akcji d_1 przechodzimy do wykonania wszystkich akcji a'' . Ta operacja przerzuca żetony z miejsc a_{pom} do a_{on2} . Umożliwiona staje się teraz akcja d_2 . Podobnie jak w przypadku akcji d_1 , żadna z akcji dla których którekolwiek z miejsc a_{on2} byłoby wejściem nie jest umożliwiona, gdyż miejsca a_{on1} są puste. W związku z tym możemy wykonać akcję d_2 i tym samym wyczyścić wszystkie dodatkowe miejsca sieci S' . W szczególności procedurę tę możemy wykonać w sytuacji, gdy w obrębie sieci S osiągnęliśmy stan pusty. Łatwo zauważyć, że jedynym sposobem wyczyszczenia sieci S' jest wykonanie akcji d_1 oraz d_2 , po których sieć S' staje się martwa. Zachodzi zatem

Fakt 3.3.6 *Niech $S = (P, T, W, I, M_0)$ będzie dowolną siecią inhibitorową, zaś $S' = (P', T', W', M'_0)$ będzie siecią pokojową skonstruowaną zgodnie z opisaną w Kroku 3 procedurą. Wówczas stan pusty jest osiągalny w sieci S dokładnie wtedy, gdy stan pusty jest osiągalny w sieci S' .*

Pozwala to natychmiast sformułować

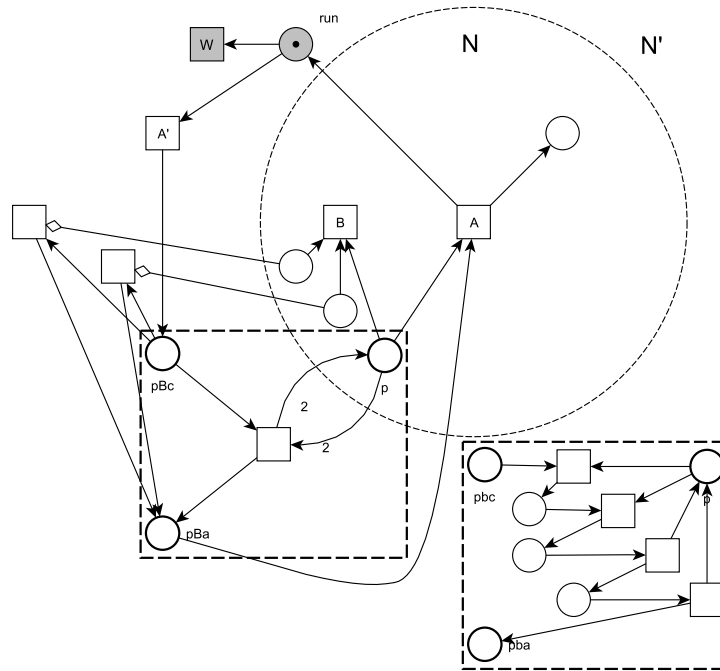
Wniosek 3.3.7 *Problem Pustości Stanu jest nierozstrzygalny w sieciach pokojowych. Wszystkie rozważane problemy decyzyjne związane z markingiem są nierozstrzygalne w sieciach pokojowych.*

3.3.2 Równoważność markingowa sieci pokojowych i inhibitorowych

Drugi krok opisanej w poprzednim paragrafie konstrukcji pokazuje jak symulować działanie sieci inhibitorowych przy pomocy sieci pokojowych. Pozwala to wywnioskować, że sieci pokojowe jako model systemów współbieżnych są przynajmniej tak samo uniwersalne jak sieci inhibitorowe. Powołując się na cytowaną wcześniej pracę Agerwala [2] przypomnijmy, że sieci inhibitorowe są obliczeniowo równoważne maszynom Turinga. Pokażemy teraz konstrukcję pozwalającą symulować działanie sieci pokojowych za pomocą sieci inhibitorowych, co ustali równoważność obliczeniową tych dwóch typów sieci. Jest to zresztą wynik mało zaskakujący, nieformalnie można go wywnioskować z Tezy Churcha-Turinga.

Na potrzeby tej konstrukcji, głównie ze względu na czytelność Rysunku 3.12, akcje oznaczać będziemy wielkimi literami.

Konstrukcja 3.3.8 Symulacja działania zwykłej i czystej sieci pokojowej (Rysunek 3.12) Niech $S = (P, T, W, M_0)$ będzie zwykłą i czystą siecią pokojową. Skonstruujemy sieć inhibitorową $S' = (P', T', W', I, M'_0)$, która symuluje, w sensie osiągalności markingów, działanie sieci S . Na początku dodajmy do sieci S , znane nam już z innych konstrukcji, nowe miejsce run oraz kopię A' każdej akcji $A \in T$. Miejsce run będzie wyjściem z każdej akcji $A \in T$ oraz wejściem do każdej akcji A' . Wykonanie akcji A' wyrażać będzie naszą nadzieję, że akcja A mogła się wykonać w pierwotnej, pokojowej sieci S . Następnie dodajmy fragment sieci odpowiedzialny za testowanie naszej hipotezy. Naj-



Rysunek 3.12: Symulacja działania zwykłych i czystych sieci pokojowych przez sieci inhibitorowe. Na rysunku pokazano alternatywną możliwość wykonania testu na nadmiarową liczbę żetonów w miejscu p

ważniejszym krokiem w tej weryfikacji będzie sprawdzenie, czy inna akcja umożliwiona w tym markingu nie straci tego statusu po wykonaniu akcji A . W tym celu, dla każdej pary $(B, p) \in T \times P$ składającej się z posiadającej wspólne z A wejście akcji B oraz tego wspólnego wejścia, dodajmy dwa miejsca pB_c oraz pB_a . Miejsce pB_c będzie wyjściem z akcji A' zaś miejsce pB_a będzie wejściem do akcji A . Przepływ żetonu między tymi miejscami będzie możliwy, jeśli akcja B nie przeszkadza w wykonaniu akcji A . Mamy dwie możliwości, aby pozytywnie przeprowadzić taki test. Po pierwsze, akcja B może nie być umożliwiona. Ten wariant sprawdzany jest przez nowe akcje połączone z pozostałymi (wszystkimi prócz p) wejściami do akcji B łukami inhibitorowymi. Wyjściem z wszystkich tych akcji jest miejsce pB_a , zaś wejściem miejsce pB_c . Po wtóre, akcja B może być umożliwiona, ale w miejscu p jest wystarczająco dużo żetonów (w rozważanym przypadku sieci czystych i zwykłych przynajmniej dwa), aby po wykonaniu akcji A wystarczyło ich jeszcze dla akcji B . Ten test realizujemy przez dodanie nowej akcji połączonej z miejscem p pętelką o wadze 2 oraz przerzucającej żeton z miejsca pB_c do miejsca pB_a . Przez $10M$

oznaczymy marking, w którym dla miejsc p ze zbioru P mamy $10M(p) = M(p)$, miejsce run zawiera jeden żeton, zaś pozostałe miejsca są puste. Ustalmy wówczas $M'_0 = 10M_0$.

W ten sposób otrzymaliśmy sieć inhibitorową S' , w której każda akcja $A \in T$, oprócz zwykłych warunków umożliwienia związanych z zawartością odpowiednich miejsc wejściowych, musi przejść dodatkowe testy związane z trwałością wykonywanego kroku. Przeanalizujemy zachowanie sieci S' .

W dowolnym markingu $10M$ możemy wykonać dowolną z akcji A' . W dalszej części wykonujemy testy związane z pokojową wykonalnością akcji A w markingu M (żadna z akcji nie biorących udziału w testach nie jest umożliwiona). Jeśli akcja A była pokojowo wykonalna w sieci P (czyli (M, a, M') jest krokiem sieci pokojowej), to testy zostają pomyślnie zakończone i akcja A zostaje umożliwiona w sieci S' , co owocuje uzyskaniem markingu $10M'$. W przeciwnym przypadku, po wykonaniu pozytywnych testów, sieć osiąga marking martwy. W związku z tym każde obliczenie $u = A_1 \dots A_n$ wykonalne w sieci S posiada obliczenie dualne $u = A'_1 \dots A'_n \dots A_n$ wykonalne w sieci S' . Jakakolwiek próba odejścia od tego schematu powoduje natychmiastowe osiągnięcie martwego markingu. W związku z tym zachodzi

Fakt 3.3.9 *Niech $S = (P, T, W, M_0)$ będzie dowolną zwykłą i czystą siecią pokojową, zaś $S' = (P', T', W', I, M'_0)$ będzie siecią inhibitorową skonstruowaną zgodnie z opisaną wyżej procedurą. Wówczas marking M jest osiągalny w sieci S dokładnie wtedy, gdy marking $10M$ jest osiągalny w sieci S' .*

Zauważmy, że zwykłość i czystość sieci S potrzebna była tylko przy uzasadnieniu poprawności testu sprawdzającego, czy po wykonaniu akcji A w miejscu p pozostanie wystarczająco dużo żetonów na podtrzymanie umożliwienia akcji B . Można stąd wywnioskować, że również dla sieci pokojowych które nie są czyste lub zwykłe można przeprowadzić analogiczną konstrukcję. Ostatecznie, rozumowanie to pozwala wysnuć

Wniosek 3.3.10 *Sieci pokojowe są równoważne obliczeniowo sieciom inhibitorowym, a zatem również maszynom Turinga.*

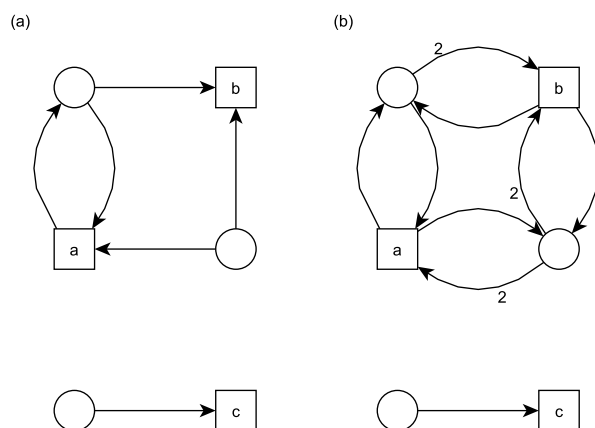
3.3.3 Pokojowe sieci wolnego wyboru

Przedstawione do tej pory wyniki badań nad sieciami pokojowymi zakończył dość zaskakujący wniosek. Na poziomie pojedynczej sieci statycznej N każdy marking osiągalny w sieci pokojowej $P = (N, M_0)$ jest również osiągalny w odpowiadającej jej p/t-sieci $P' = (N, M_0)$. Zbiór markingów osiągalnych w sieci P może być jednak uboższy niż zbiór markingów osiągalnych w sieci P' . Mimo to, patrząc globalnie, rodzina wszystkich możliwych zbiorów markingów osiągalnych w sieciach pokojowych są dużo bogatsza od rodziny zbiorów markingów osiągalnych w p/t-sieciach. Pojawia się natychmiastowe pytanie - czy można tak ograniczyć sieci pokojowe, aby ich siła ekspresji (możliwe zbiory markingów osiągalnych) była równa sile ekspresji zwykłych p/t-sieci? Dotychczasowe rozważania pokazują, że sieci pokojowe nie tracą niczego ze swej ekspresji, jeśli zabronimy stosowania pętelek i wag innych niż zerojedynkowe. Przyjrzyjmy się innemu, naturalnemu kandydatowi na takie ograniczenie - pokojowym sieciom wolnego wyboru, w których problemy osiągalności i pokrywalności markingu są rozstrzygalne. Wynik ten uzyskamy przez wskazanie konstrukcji pozwalającej symulować działanie pokojowych sieci wolnego wyboru za

pomocą klasycznych p/t-sieci. Nie bez znaczenia jest tutaj zarówno warunek określający współdzielenie zbiorów miejsc wejściowych jak i zwykłość sieci wolnego wyboru.

Konstrukcja 3.3.11 Niech $S = (P, T, W, M_0)$ będzie pokojową siecią wolnego wyboru. Skonstruujemy p/t-sieć $S' = (P, T, W', M_0)$, która symuluje w sensie osiągalności markingów działanie sieci S . Przypomnijmy, że sieci wolnego wyboru są sieciami zwykłymi (funkcja wagowa jest zerojedynkowa). W związku z tym wejście do akcji a może być zwykłym wejściem lub być połączone z nim pętelką (być jednocześnie wejściem i wyjściem). Poza tym zwykle wejścia mogą być wejściami do jednej lub wielu różnych akcji. Potraktujemy te trzy przypadki w różny sposób:

1. jeśli miejsce p jest wejściem do jednej akcji a , to niczego nie zmieniamy ($W'(a, p) = W(a, p) = 0 \wedge W'(p, a) = W(p, a) = 1$);
2. jeśli miejsce p jest połączone z akcją a pętelką, to również niczego nie zmieniamy ($W'(a, p) = W(a, p) = 1 \wedge W'(p, a) = W(p, a) = 1$);
3. jeśli miejsce p jest wejściem do kilku różnych akcji, to zwiększamy dla tej pary wagi o jeden ($W'(a, p) = W(a, p) + 1 = 1 \wedge W'(p, a) = W(p, a) + 1 = 2$), przez co powstają nowe pętelki.



Rysunek 3.13: Symulacja działania pokojowych sieci wolnego wyboru przez p/t-sieci. Na Rysunku a) pokazano trzy przypadki zależności między akcją a wejściem do niej w sieci pokojowej, na rysunku b) - odpowiadające im przekształcenia sieci

Łatwo widać, że jeśli jakaś akcja nie była wykonalna w sieci S , to nie jest wykonalna w sieci S' . Zauważmy, że jeśli akcje a i b mają jakiegokolwiek wspólne wejście, to ponieważ zbiory wejść są identyczne - to w zadanym markingu albo obie są umożliwiające w sieci S albo obie nie są umożliwiające w tej sieci. Możliwe jest jednak, że tylko jedna z nich jest wykonalna (ze względu na istniejące pętelki). Jeśli w danym markingu M akcja a jest wykonalna w sieci S , to jej wykonanie nie odbierze żetonów żadnej innej akcji, co oznacza, że wejście do niej może zostać wyczyszczone tylko w przypadku, gdy nie dzieli go z żadną inną akcją. W związku z tym każda wykonalna w markingu M w sieci S akcja jest też umożliwiające w sieci S' . W związku z tym sieci S oraz S' działają identycznie. Nasuwa to natychmiastowy

Wniosek 3.3.12 *Zbiory markingów osiągalnych w sieciach S oraz S' są identyczne. Pokojowe sieci wolnego wyboru mają moc obliczeniową nie większą niż klasyczne p/t-sieci.*

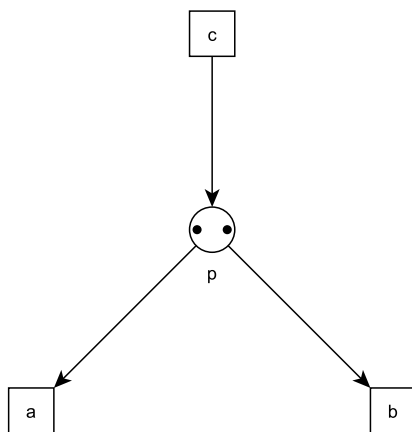
Zauważmy, że konstrukcja przedstawiona na Rysunku 3.9 nie daje w wyniku sieci wolnego wyboru (chyba, że zbiór akcji T jest jednoelementowy). Problemem otwartym pozostaje pytanie, czy każdą p/t-sieć można zasymulować przy pomocy pokojowej sieci wolnego wyboru.

3.4 Uogólnienia sieci pokojowych

Konflikty w sieciach pokojowych

Przypomnijmy motywację dla badań nad sieciami pokojowymi. Języki obliczeń w tych sieciach są tożsame z językami obliczeń trwałych w odpowiadających im p/t-sieciach. Nie oznacza to, że w sieciach pokojowych nie ma konfliktów. Przypomnijmy, że pojęcie konfliktu dynamicznego zdefiniowaliśmy (w Definicji 3.1.2) bardzo ogólnie. Zastanówmy się, czy w myśl takiej definicji w sieciach pokojowych występują jakiegokolwiek konflikty.

Przykład 3.4.1 Przywołajmy sieć statyczną z Przykładu 3.1.6. Składa się ona z jednego miejsca p , dwóch akcji a, b dla których miejsce p jest wspólnym wejściem oraz akcji c , dla której miejsce p jest wyjściem. Przyjmijmy ponadto, że w markingu początkowym M_0 w miejscu p znajdują się dwa żetony. Zauważmy, że w tym markingu wszystkie akcje



Rysunek 3.14: Konflikt w sieciach pokojowych

są możliwe i wykonalne. Problem pojawia się w sytuacji, kiedy wykonamy akcję a (lub symetrycznie b , ale skoncentrujmy się na akcji a). Po jej wykonaniu w miejscu p pozostaje jeden żeton. Cały czas wszystkie akcje są możliwe, czyli akcja a wykonała się pokojowo, jednak w tej chwili wykonanie akcji b spowoduje uniemożliwienie całej akcji a . Reasumując, w przytoczonej sieci akcje a oraz b są pokojowo wykonalne, natomiast obliczenia ab i ba nie. Oznacza to (wprost z Definicji 3.1.2), że jeśli sieć S potraktujemy jako sieć pokojową, to akcje a i b są w markingu M_0 w konflikcie dynamicznym.

Sieci 2-pokojowe Przytoczony przykład pokazuje, że sieci pokojowe nie są wolne od konfliktów. Można się zastanowić nad bezkonfliktowymi przebiegami w sieciach pokojowych, co prowadzi do

Definicja 3.4.2 *2-pokojową siecią Petriego* nazywamy sieć o strukturze p/t-sieci (czyli czwórkę $S = (P, T, W, M_0)$) ze zmienionymi definicjami umożliwienia i wykonalności. Ponownie rozdzielamy te dwa pojęcia. Akcja $a \in T$ jest *umożliwiona* w markingu M w sieci 2-pokojowej S (czyli $M\mathcal{U}a$), o ile jest ona wykonalna w sieci pokojowej o identycznej z siecią S strukturze. Akcja umożliwiona w markingu M jest *wykonalna* w tym markingu w sieci 2-pokojowej (czyli $M\mathcal{W}a$) o ile jej wykonanie nie uniemożliwi wykonania innej pokojowo wykonalnej (2-pokojowo umożliwionej) akcji, innymi słowy nie wprowadzi p/t-sieci o strukturze identycznej jak sieć S w stan konfliktowy. Formalnie zapisać to można

$$\forall_{b \neq a} M\mathcal{W}b \Rightarrow (Ma)\mathcal{W}b,$$

gdzie relacja \mathcal{W} oraz marking wynikowy Ma rozumiane są w sensie sieci pokojowej o identycznej jak sieć S strukturze. Samo wykonanie akcji powoduje identyczną jak w przypadku p/t-sieci i sieci pokojowych zmianę markingu.

Uwaga 3.4.3 *Zauważmy, że analizowane konfuzje w których pojawiają się nowe konflikty związane są z akcjami prowadzącymi do konfliktów, a więc nie są 2-wykonalne. Część konfuzji jest w związku z tym rodzajem konfliktów wyższych rzędów.*

Zauważmy ponadto, że za pomocą sieci 2-pokojowych możemy symulować zachowanie p/t-sieci. Konstrukcja jest analogiczna jak w przypadku klasycznych sieci pokojowych (Rysunek 3.9). Odroczenie wykonania akcji powoduje umożliwienie jej trwałego wykonania, a zabronienie wprowadzania sieci w stany konfliktowe może co najwyżej uniemożliwić wykonanie dwóch akcji primowanych z rzędu, nie wpływa jednak na markingi osiągalne w tak skonstruowanej sieci. Nie może też - wprost z definicji - umożliwić wykonania żadnej akcji, która nie byłaby wykonalna w odpowiadającej jej sieci pokojowej!

Dla sieci 2-pokojowych działa również konstrukcja z drugiego kroku dowodu nierozstrzygalności Problemu Pustości Stanu w sieciach pokojowych, czyli symulacja sieci inhibitorowych przedstawiona na Rysunku 3.10. Podobnie jak w przypadku tej konstrukcji, obliczenia prowadzone zgodnie z wykonalnymi w wyjściowej sieci inhibitorowej (zamiast akcji a , akcje aa') są umożliwiające w powstałej sieci 2-pokojowej. Pożądanym efektem ubocznym jest niemożliwość przygotowania do wykonania akcji a posiadającej żetony w swoich wejściach inhibitorowych. Wykonanie akcji a' w takiej sytuacji wprowadziłoby sieć w stan konfliktowy, co jest zabronione w sieciach 2-pokojowych.

Konstrukcje te okażą się jednak niewystarczające w przypadku sieci pokojowych wyższych rzędów (patrz Definicja 3.4.7 i Przykład 3.4.19). W związku z tym podamy alternatywną konstrukcję, będącą iteracją rozumowania pozwalającego symulować p/t-sieci i sieci inhibitorowe przy pomocy sieci pokojowych.

Konstrukcja 3.4.4 Niech $S = (P, T, W, M_0)$ będzie siecią Petriego. Skonstruujemy na jej podstawie sieć pokojową, a następnie na jej podstawie sieć 2-pokojowa $S' = (P', T', W', M'_0)$, które symulują w sensie osiągalności markingów działanie sieci S . W obu etapach postąpimy zgodnie z instrukcjami zawartymi w pierwszym kroku konstrukcji dla sieci pokojowych (Rysunek 3.9). Ostatecznie więc, do zbioru miejsc oprócz miejsc należących do

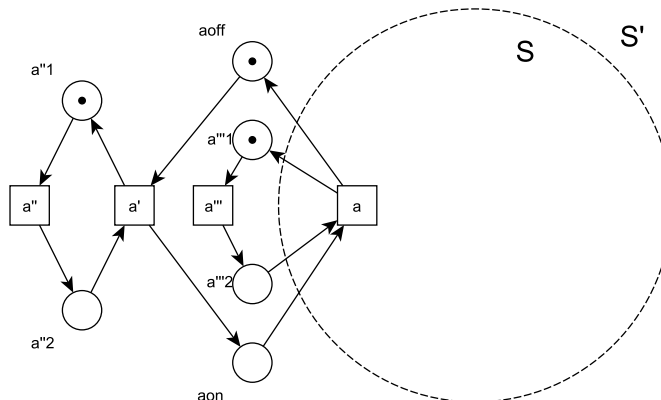
oryginalnej sieci S należeć będą po dwa miejsca powstałe dla każdej akcji $a \in T$, oznaczane a_{off} oraz a_{on} a także po cztery dodatkowe miejsca powstałe dla każdej akcji $a \in T$, oznaczone przez a''_1, a''_2, a'''_1 i a'''_2 . Poza tym, zbiór akcji T' zawierał będzie cały zbiór T oraz po trzy nowe akcje dla każdej akcji $a \in T$. Będą to akcja a' , która powstała w pierwszym etapie konstrukcji oraz akcje a''' i a'' powstałe w drugim etapie. Funkcja wagowa w obrębie oryginalnej sieci S pozostanie niezmienną. Poza nią będzie wynosić zero, za wyjątkiem zdefiniowanych dla każdego $a \in T$ następujących par:

- $W(a, a_{off}) = W(a_{on}, a) = 1$
- $W(a', a_{on}) = W(a_{off}, a') = 1$
- $W(a, a'''_1) = W(a'''_2, a) = 1$
- $W(a'''_1, a'''_2) = W(a'''_1, a''_1) = 1$
- $W(a', a''_1) = W(a''_2, a') = 1$
- $W(a'', a''_2) = W(a''_1, a'') = 1$.

Marking $10M : P' \rightarrow \mathbb{N}$ odpowiadający markingowi $M : P \rightarrow \mathbb{N}$, określony tak aby $M'_0 = 10M_0$ był markingiem początkowym sieci S' odpowiadającym markingowi M_0 w sieci S , dany jest formułą:

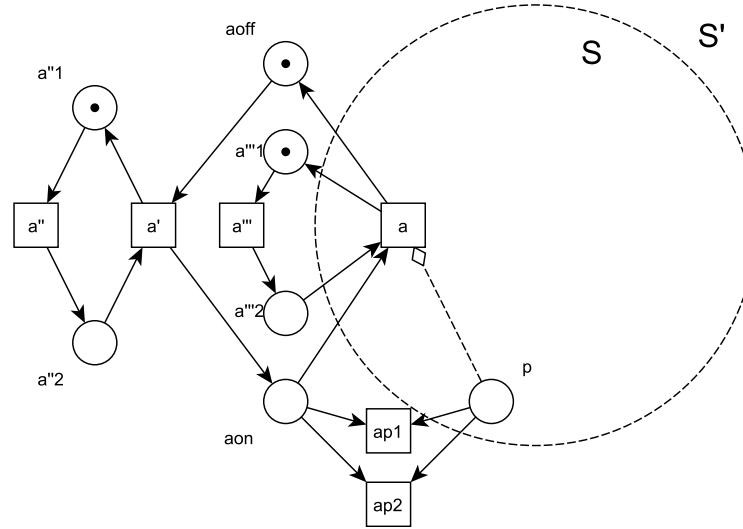
$$10M(p) = \begin{cases} M(p) & \text{dla } p \in P \\ 1 & \text{dla } p = a_{off} \text{ dla } a \in T \\ 0 & \text{dla } p = a_{on} \text{ dla } a \in T \\ 1 & \text{dla } p = a''_1 \text{ dla } a \in T \\ 0 & \text{dla } p = a''_2 \text{ dla } a \in T \\ 1 & \text{dla } p = a'''_1 \text{ dla } a \in T \\ 0 & \text{dla } p = a'''_2 \text{ dla } a \in T \end{cases} .$$

Ideę całej tej konstrukcji przedstawiono na Rysunku 3.15.



Rysunek 3.15: Idea konstrukcji sieci 2-pokojowej S' symulującej działanie p/t-sieci S

Zauważmy na początek, że przy takiej konstrukcji, dla każdego obliczenia $u = a_1 \dots a_n$ wykonalnego w p/t-sieci S istnieje obliczenie dualne $u' = a_1''' a_1'' a_1' a_1 \dots a_n''' a_n'' a_n' a_n$ wykonalne w sieci 2-pokojowej S' . Wykonanie akcji w innej niż zaproponowana kolejności może co najwyżej uniemożliwić wykonanie pewnych akcji pochodzących ze zbioru T , nie wpływa natomiast na miejsca oryginalnej sieci S (w szczególności, w sieci 2-pokojowej S' nie może zostać umożliwione nic, co nie byłoby możliwe w obciętych do P markingów w p/t-sieci S). W związku z tym prawdziwy jest następujący fakt: Marking M jest osiągalny w p/t-sieci S dokładnie wtedy, gdy odpowiadający mu marking $10M$ jest osiągalny w sieci 2-pokojowej S' .



Rysunek 3.16: Idea konstrukcji sieci 2-pokojowej S' symulującej działanie sieci inhibitorowej S

Na podobnej zasadzie możemy przeprowadzić drugi krok (patrz Rysunek 3.16), czyli symulację działania sieci inhibitorowych. Weźmy teraz sieć inhibitorową $S = (P, T, W, I, M_0)$. Za punkt wyjścia przyjmijmy sieć 2-pokojową S'' skonstruowaną z sieci S w identyczny jak w kroku pierwszym sposób, ze zdefiniowanymi identycznie markingami $10M$ odpowiadającymi markingom M , których są okrojonymi do dziedziny będącej zbiorem P wersjami. Dodajmy jeszcze nowe akcje, odpowiadające za inhibitorowe działanie tej sieci. Ponownie, dla każdego łuku inhibitorowego (p, a) dodamy po dwie akcje $ap1$ oraz $ap2$. Wejściami dla tych akcji będą dodane w pierwszym kroku wejście do akcji a (czyli a_{on}) a także miejsce inhibitorowe p z oryginalnej p/t-sieci S . Podobnie jak w przypadku sieci pokojowych, w sieci 2-pokojowej S' możliwe jest w markingu $10M$ wykonanie akcji, która była umożliwiona w sieci S w markingu M . Inhibitory działają zgodnie z intuicją – jeśli w markingu M w sieci inhibitorowej S akcja a ma wystarczającą liczbę zasobów (żetonów w odpowiednich miejscach wejściowych) aby się wykonać, ale nie jest umożliwiona ze względu na niepustość miejsca inhibitorowego p , to przy próbie jej wykonania, czyli wykonania obliczenia $a''' a'' a'$ a później akcji a , doszlibyśmy do markingów w którym jednocześnie umożliwione są (w sensie umożliwienia w p/t-sieci o identycznej jak S' strukturze) akcje a , $ap1$ oraz $ap2$. Wykonanie którejkolwiek z nich spowodowałoby uniemożliwienie (w sensie umożliwienia w p/t-sieci o identycznej jak S' strukturze) obu pozostałych akcji, czyli akcja a w takim markingu jest umożliwiona, ale nie jest wykonalna w sieci pokojowej o identycznej jak

sieć S' strukturze, a zatem nie jest nawet umożliwiona w sieci 2-pokojowej S' . Istotą tego uniemożliwienia jest jednoczesne pojawienie się żetonów w miejscach a_{on} oraz p i nie ma znaczenia, czy taki marking jest w ogóle osiągalny w sieci 2-pokojowej S' . Poza tym, mimo że wejściem do akcji ap_1 oraz ap_2 jest miejsce p należące do oryginalnej p/t-sieci S , to i tak żadna z nich nie będzie nigdy wykonalna i nie zmieni wartości markingu okrojonego do zbioru P . Można stąd wywnioskować, że w nowopowstałej sieci 2-pokojowej S' nie da się osiągnąć markingu $10M$ jeśli w sieci inhibitorowej S nie można było osiągnąć markingu M .

Ponadto, działanie sieci 2-pokojowych możemy symulować za pomocą sieci inhibitorowych. 2-pokojowa wykonalność ma, podobnie jak trwała wykonalność stosowana w sieciach pokojowych, charakter lokalny i można ją przetestować badając topologiczne sąsiedztwo akcji (z wykorzystaniem łuków inhibitorowych). Jest to jednak operacja dużo bardziej złożona niż przedstawiona na Rysunku 3.12 symulacja sieci pokojowych. Nie będziemy w związku z tym przytaczać szczegółów. Za wystarczający argument na poparcie tej tezy uznamy Tezę Churcha-Turinga (gdyby sieci 2-pokojowych nie dało się symulować za pomocą sieci inhibitorowych, to uzyskalibyśmy w ten sposób model obliczeń ogólniejszy niż maszyny Turinga). Pozwala to sformułować natychmiastowy

Wniosek 3.4.5 *Sieci 2-pokojowe są równoważne obliczeniowo sieciom pokojowym, a zatem również maszynom Turinga.*

3.4.1 Hierarchia sieci pokojowych

Przykład 3.4.1 jest na tyle uniwersalny, że dodając w markingu początkowym dodatkowy żeton do miejsca p otrzymujemy sieć 2-pokojową w której w markingu początkowym występuje konflikt. Zaproponujemy teraz całą hierarchię pokojowej wykonalności (a co za tym idzie - sieci pokojowych).

Definicja 3.4.6 Niech $S = (P, T, W, M_0)$ będzie p/t-siecią. W sieci tej zwykłą wykonalność nazwiemy *0-wykonalnością*, zaś wykonalność trwałą (wykorzystaną w sieciach pokojowych) - *1-wykonalnością*. Bardziej ogólnie, weźmy marking M , akcję a oraz $k > 0$. Zbiór wszystkich obliczeń k -wykonalnych w markingu M oznaczają będziemy $EX_k(M)$, zaś zbiór wszystkich markingów osiągalnych w sieci k -pokojowej przez $[M_0 >_k$. Mówimy, że akcja a jest *$k+1$ -umożliwiona* w markingu M o ile jest k -wykonalna w tym markingu. Akcja a jest *$k+1$ -wykonalna* w markingu M , o ile jest $k+1$ -umożliwiona w tym markingu, a ponadto jej wykonanie nie uniemożliwi (w sensie k -wykonalności) wykonania innej, $k+1$ -umożliwionej akcji. Bardziej formalnie:

$$a \in EX_{k+1}(M) \Leftrightarrow a \in EX_k(M) \wedge EX_k(M) \setminus \{a\} \subseteq EX_k(Ma).$$

Mając tak zdefiniowaną k -osiągalność, podamy rekurencyjną definicję sieci k -pokojowych.

Definicja 3.4.7 *0-pokojowa sieć Petriego* nazywamy p/t-sieć (w myśl Definicji 1.2.5). *k -pokojową sieć Petriego* nazywamy sieć o strukturze p/t-sieci (czyli czwórkę $S = (P, T, W, M_0)$), zachowującą definicje kroku oraz obliczenia, w której relacja umożliwienia

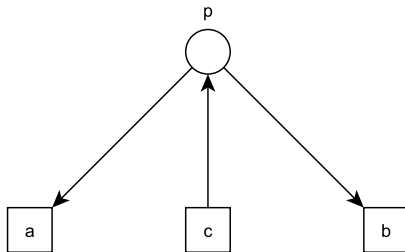
$\mathcal{U} \subseteq \mathbb{N}^P \times T$ zdefiniowana jest jako relacja wykonalności w sieci k-1-pokojowej o identycznej strukturze. Natomiast relacja wykonalności w markingu M akcji a określona jest następująco:

$$M\mathcal{W}a \Leftrightarrow M\mathcal{U}a \wedge \forall_{b \neq a} M\mathcal{U}b \Rightarrow (Ma)\mathcal{U}b,$$

gdzie przez Ma rozumiemy marking wynikowy kroku (M, a, M') w sieci k-1-pokojowej o identycznej strukturze.

Uwaga 3.4.8 *Definicja 3.4.6 odnosi się do własności p/t-sieci, natomiast Definicja 3.4.7 wprowadza nowe typy sieci Petriego. Są to bliźniacze konstrukcje rekurencyjne i k-wykonalność w p/t-sieci S jest tożsama umożliwieniu w sieci k-pokojowej o strukturze identycznej jak sieć S .*

Przykład 3.4.9 Rozważmy jeszcze raz sieć S z Przykładów 3.1.6 oraz 3.4.1 z markingiem początkowym $M_0 = (0)$:



Rysunek 3.17: Konflikt w sieciach pokojowych

Potraktujmy tę sieć jako p/t-sieć (czyli sieć 0-pokojową). Wówczas po każdym wykonaniu akcji c można wykonać akcję a lub b . W szczególności, wykonalne jest nieskończone obliczenie $u_0 = (ca)^\omega$. Jeśli sieć S potraktowana zostanie jako sieć pokojowa (czyli 1-pokojowa), to obliczenie u_0 nie jest już w niej wykonalne (gdyż nie jest obliczeniem trwałym). Wykonalne natomiast jest obliczenie $u_1 = c(ca)^\omega$, rozwiązujące nieskończenie wiele razy konflikt dynamiczny przy markingu M , w którym $M(p) = 2$. Wynika stąd natomiast, że nie jest to obliczenie wykonalne w sieci S traktowanej jako sieć 2-pokojowa. Kontynuując powyższe rozumowanie, uzyskujemy ciąg obliczeń $u_n = c^n(ca)^\omega$, które są wykonalne w każdej sieci S traktowanej jako sieć k-pokojowa dla k nie przekraczających n , zaś nie są wykonalne dla większych k . Nie zmienia to faktu, że niezależnie od przyjętego typu pokojowej wykonalności, zbiór markingów osiągalnych to zbiór funkcji przyporządkowujących jednemu argumentowi p dowolną, ustaloną liczbę naturalną.

Przeprowadzona w przykładzie dyskusja prowadzi do sformułowania faktów określającego zmniejszanie się, wraz ze wzrostem k , rozmiaru języków obliczeń wykonalnych oraz zbiorów markingów osiągalnych w sieciach k-pokojowych:

Fakt 3.4.10 Niech $S = (P, T, W, M_0)$ będzie p/t -siecią. Wówczas, rozważając sieci k -pokojowe S_k o identycznej jak S strukturze, mamy

$$L(S_0) \supseteq L(S_1) \supseteq \dots \supseteq L(S_k) \supseteq \dots$$

Fakt 3.4.11 Niech $S = (P, T, W, M_0)$ będzie p/t -siecią. Wówczas, rozważając sieci k -pokojowe S_k o identycznej jak S strukturze, mamy

$$[M_0 > \supseteq [M_0 >_1 \supseteq \dots \supseteq [M_0 >_k \supseteq \dots$$

Możemy również rozszerzyć pojęcie klasycznej trwałości:

Definicja 3.4.12 Niech $S = (P, T, W, M_0)$ będzie p/t -siecią. Powiemy, że osiągalny w sieci S marking $M \in [M_0 >$ jest k -trwały dokładnie wtedy, gdy każda akcja umożliwiona w tym markingu jest k -wykonalna. Bardziej formalnie

$$\forall a \in EX_0(M) \Rightarrow a \in EX_k(M).$$

Powiemy, że sieć S jest k -trwała jeśli każdy marking w niej osiągalny jest k -trwały.

Zgodnie z powyższą definicją, klasyczna trwałość jest tożsama z 1-trwałością. Zdefiniowana w ten sposób hierarchia sieci trwałych jest płaska, to znaczy zachodzi

Lemat 3.4.13 P/t -sieć $S = (P, T, W, M_0)$ jest siecią $k+1$ -trwałą dokładnie wtedy, gdy jest siecią k -trwałą (dla $k > 0$).

Dowód. Załóżmy, że sieć S jest k -trwała, ale nie jest $k+1$ -trwała. Oznacza to, że istnieje taki marking M osiągalny w tej sieci i akcja a k -wykonalna w tym markingu, że $a \notin EX_{k+1}(M)$. Oznacza to, że $a \notin EX_k(M)$ lub $\exists_{b \neq a} b \in EX_k(M) \wedge b \notin EX_k(Ma)$. Pierwszy składnik tej koniunkcji jest sprzeczny z k -wykonalnością akcji a . Zauważmy ponadto, że $b \in EX_{k-1}(M)$ oraz $a \in EX_{k-1}(M)$. Ponieważ mamy też, że $a \in EX_k(M)$, to $b \in EX_k(Ma)$, co prowadzi do sprzeczności i kończy dowód. \square

3.4.2 Lematy o propagacji i genezie konfliktów

W poprzednim paragrafie wprowadziliśmy różne poziomy wykonalności akcji w danym markingu (Definicja 3.4.6). Przyjrzyjmy się teraz dwóm lematom dającym praktyczne narzędzia do badania k -wykonalności. Pierwszy z nich, nazwany lematem o propagacji konfliktu, mówi o wstecznym (w rozumieniu porządku topologicznego na grafie osiągalności) propagowaniu klasycznych konfliktów, które powodują przerwanie w markingu M wykonalności akcji na poziomie 1. Przy spełnieniu odpowiednich warunków dodatkowych, takie przerwanie, na poziomie o jeden wyższym, może się propagować do markingów, z których da się bezpośrednio osiągnąć marking M .

Lemat 3.4.14 (o propagacji) Niech $S = (P, T, W, M_0)$ będzie p/t -siecią. Wówczas jeśli w osiągalnym w sieci S markingu M ma miejsce przerwanie wykonalności akcji a ($a \in EX_k(M) \wedge a \notin EX_{k+1}(M)$) oraz istnieje akcja b różna od a prowadząca z markingu M' do markingu M ($M'bM$) a ponadto akcje a, b są $k+1$ -wykonalne w markingu M' , to w

tym markingu następuje przerwanie wykonalności akcji b , to znaczy nie jest ona $k+2$ -wykonalna. Bardziej formalnie:

$$\left. \begin{array}{l} a \in EX_k(M) \\ a \notin EX_{k+1}(M) \\ M'bM \\ a \in EX_{k+1}(M') \\ b \in EX_{k+1}(M') \end{array} \right\} \Rightarrow b \notin EX_{k+2}(M').$$

Dowód. Załóżmy, że $b \in EX_{k+2}(M')$, czyli równoważnie

$$b \in EX_{k+1}(M' \wedge EX_{k+1}(M') \setminus \{b\} \subseteq EX_{k+1}(M'b = M).$$

Ale $a \in EX_{k+1}(M') \setminus \{b\}$, a jednocześnie $a \notin EX_{k+1}(M)$. W związku z uzyskaną sprzecznością wnioskujemy, że $b \notin EX_{k+2}(M')$, co jest tezą. \square

Sam lemat zawiera pokąźną listę warunków i wygląającą przy niej skromnie tezę. Jego siła ukazuje się dopiero wraz z drugim lematem, nazwanym lematem o genezie konfliktów. Lemat ten mówi, że każde przerwanie akcji w danym markingu jest spowodowane w opisany w lemacie o propagacji sposób.

Lemat 3.4.15 (o genezie) *Niech $S = (P, T, W, M_0)$ będzie p/t -siecią, zaś M markin-
giem osiągalnym w tej sieci. Wówczas każde przerwanie wykonalności akcji a w tym mar-
kingu ma swoją genezę w przerwaniu (na wyższym poziomie) pewnej, różnej od a , akcji w
markingu Ma . Bardziej formalnie:*

$$\left. \begin{array}{l} a \in EX_{k+1}(M) \\ a \notin EX_{k+2}(M) \end{array} \right\} \Rightarrow \exists_b \left\{ \begin{array}{l} b \in EX_k(Ma) \\ b \notin EX_{k+1}(Ma) \\ b \in EX_{k+1}(M) \end{array} \right. .$$

Dowód. Ponieważ jednocześnie $a \in EX_{k+1}(M)$ oraz $a \notin EX_{k+2}(M)$, to z definicji k -wykonalności mamy, że zbiór $(EX_{k+1}(M) \setminus \{a\}) \setminus EX_{k+1}(Ma)$ jest niepusty. Weźmy $b \in (EX_{k+1}(M) \setminus \{a\}) \setminus EX_{k+1}(Ma)$. Otrzymujemy wówczas, że $b \notin EX_{k+1}(Ma)$ oraz $b \in EX_{k+1}(M)$. Ponieważ $a, b \in EX_{k+1}(M)$, to również $a, b \in EX_k(M)$. gdyby przy tym $b \notin EX_k(Ma)$, to, z definicji, a nie byłoby $k+1$ -wykonalne w markingu M . Uzyskana sprzeczność daje trzecią część tezy i kończy dowód. \square

Z powyższego lematu możemy wysnuć natychmiastowy, indukcyjny

Wniosek 3.4.16 *Niech $S = (P, T, W, M_0)$ będzie dowolną p/t -siecią, zaś M dowolnym
markingiem osiągalnym w tej sieci. Wówczas każde przerwanie wykonalności akcji a w
tym markingu ma swoją (niezbyt odległą) genezę w klasycznym konflikcie dynamicznym.
Bardziej formalnie:*

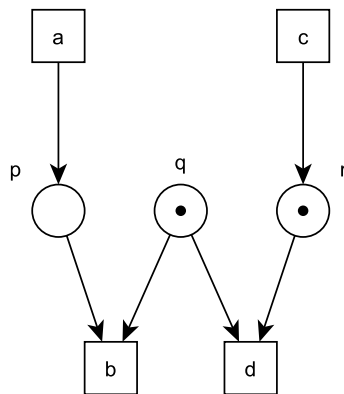
$$\left. \begin{array}{l} a \in EX_{k+1}(M) \\ a \notin EX_{k+2}(M) \end{array} \right\} \Rightarrow \exists_{w \in \Sigma^\infty, b \in \Sigma} \left\{ \begin{array}{l} b \in EX_0(Maw) \\ b \notin EX_1(Maw) \\ |w| < k \end{array} \right. .$$

Ponadto, znając pełną charakteryzację konfliktów wyższych rzędów (w odpowiednich sie-
ciach k -pokojowych), można sformułować

Wniosek 3.4.17 $S = (P, T, W, M_0)$ będzie dowolną p/t-siecią, zaś M dowolnym markingiem osiągalnym w tej sieci. Jeśli w markingu M jest umożliwiona tylko jedna akcja a , to jest ona dla dowolnego k k -pokojuowo wykonalna.

Na koniec pokażemy jeszcze, na przykładzie, praktyczne zastosowanie sformułowanych lematów:

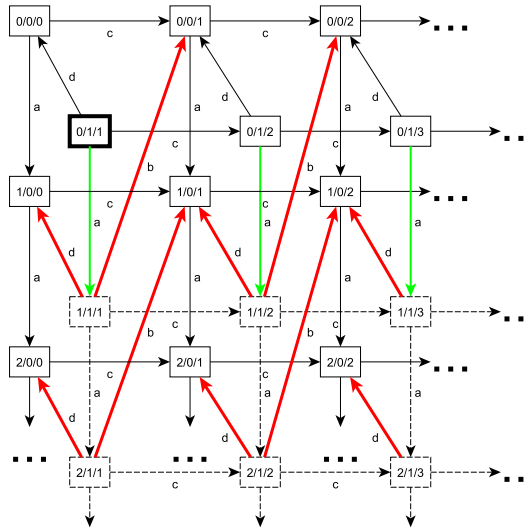
Przykład 3.4.18 Rozważmy przedstawioną na Rysunku 3.18 p/t-sieć $S = (P, T, W, M_0)$ składającą się z czterech akcji i trzech miejsc. Występuje w niej jeden konflikt strukturalny - akcje b oraz d rywalizują o żeton znajdujący się w miejscu q . W markingach pokrywających marking $(1, 1, 1)$ konflikt ten staje się dynamiczny. Sieć może zostać wprowadzona w taki stan konfliktowy przez wykonanie akcji a , pod warunkiem, że wcześniej nie została wykonana akcja d . Zauważmy, że powoduje to dość dziwne zjawisko - akcja a nie jest 2-wykonalna w markingu początkowym. Bez znaczenia jest tu fakt, że zbiór miejsc wejściowych tej akcji jest pusty, czyli nie zużywa ona żadnych zasobów dostępnych w sieci!



Rysunek 3.18: Sieć S

Na kolejnym rysunku (Rysunek 3.19) pokazany jest fragment grafu osiągalności omawianej sieci. Jest to nieskończony graf acykliczny o bardzo regularnej budowie. Składa się z markingów ułożonych w dwóch płaszczyznach, w pierwszej z nich (na Rysunku 3.19 na zewnątrz) miejsce q jest zamarkowane, w drugiej (wewnętrznej) miejsce q jest puste. Wraz z łukami poziomymi (akcja c) rosną liczby żetonów w miejscu r , zaś wraz z łukami pionowymi (akcja a) rosną liczby żetonów w miejscu p . Akcje d oraz b opisują łuki prowadzące od płaszczyzny zewnętrznej do wewnętrznej.

Korzystając z przedstawionych lematów i odnalezionych w osiągalnych markingach postaci $(m, 1, n)$, gdzie $m, n > 0$, konfliktów dynamicznych w p/t-sieci S uzyskujemy pełen opis odpowiadających sieci S sieci k -pokojuowych (dla dowolnego k). W sieci 1-pokojuowej $S_1 = (P, T, W, M_0)$ osiągalne są wszystkie markingi osiągalne w sieci S , natomiast wykonalne wszystkie akcje, poza tymi oznaczonymi kolorem czerwonym (wszystkie akcje b oraz część d). W sieci 2-pokojuowej $S_2 = (P, T, W, M_0)$ osiągalne są markingi obwiedzione linią ciągłą, zaś wykonalne akcje oznaczone ciągłymi, czarnymi łukami (łuki przerywane odpowiadają akcjom umożliwiającym w stanach nieosiągalnych). Dla $k > 2$ sieci k -pokojuowe są identyczne jak sieć 2-pokojuowa.



Rysunek 3.19: Fragment grafu osiągalności sieci S . Łukom pionowym odpowiadają akcje a , poziomym - c zaś ukośnym d (krótkie) i b (długie). Na czerwono zaznaczono akcje konfliktowe (a więc nie należące do EX_1), na zielono - akcje prowadzące do konfliktów (a więc nie należące do EX_2). Markingi i akcje nieosiągalne w sieci S traktowanej jako sieć 2-pokojowa zaznaczono liniami przerywanymi, marking początkowy pogrubiono

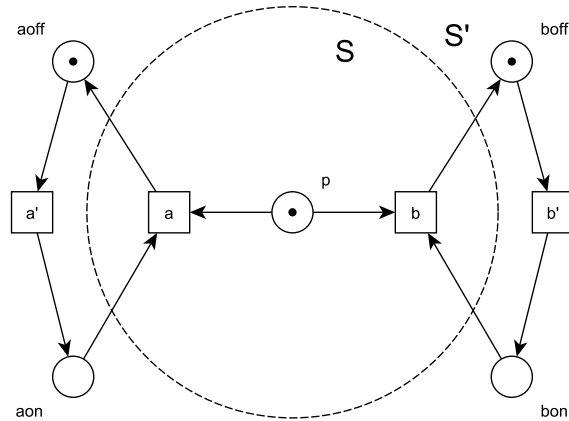
Zauważmy jeszcze, że dla markingu początkowego $M_0 = (0, 1, 0)$ 2-pokojowa sieć S' jest martwa!

3.4.3 Sieci ω -pokojowe

Konstrukcja symulująca działanie p/t-sieci przedstawiona na Rysunku 3.9 nie działa dla sieci 3-pokojowych! Rozważmy w tym celu następującą sieć, będącą zgodnym z przytoczoną konstrukcją rozszerzeniem sieci podobnej do tej z Przykładu 3.1.6:

Przykład 3.4.19 Rozważmy p/t-sieć $S = (P, T, W, M_0)$, przedstawioną na Rysunku 3.20. Zastosujemy konstrukcję, która pozwala symulować tę sieć za pomocą 1-pokojowej (lub 2-pokojowej) sieci S' . Rozważmy sieć S' (jako p/t-sieć) i zbadajmy różne poziomy wykonalności akcji w tej sieci. W markingu początkowym wszystkie akcje primowane są 0-wykonalne, 1-wykonalne oraz 2-wykonalne. Po wykonaniu akcji a' , akcja b' przestaje być 2-wykonalna. W związku z tym, w sieci S' traktowanej jako sieć 2-pokojowa ma miejsce konflikt dynamiczny w stanie M_0 , a co za tym akcja a' nie jest 3-wykonalna. W związku z tym sieć S' jest martwa i z pewnością nie symuluje działania sieci S .

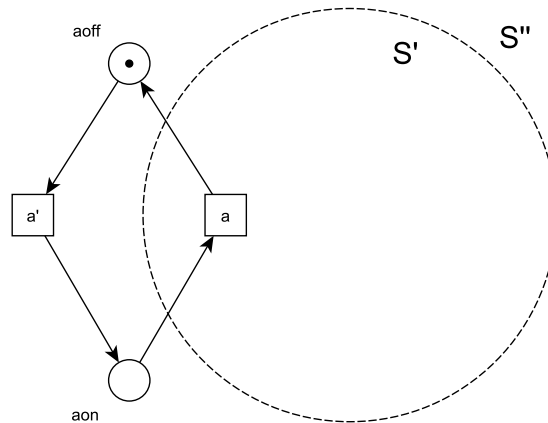
Problem ten można rozwiązać stosując podobne do przedstawionych w Konstrukcji 3.4.4 narzędzia. W przypadku symulowania p/t-sieci $S = (P, T, W, M_0)$ za pomocą sieci 3-pokojowej S' , dla każdej akcji $a \in P$ tworzyć będziemy nie 3, ale 7 nowych akcji oraz nie 6, ale 14 nowych miejsc. Zapewni to wystarczające opóźnienie przy wykonywaniu akcji a i każde obliczenie wykonalne w sieci S będzie również wykonalne (po dodaniu przed każdą akcją $a \in P$ nowych, skonstruowanych dla niej akcji w odpowiedniej kolejności) w sieci S' . W przypadku symulacji sieci inhibitorowej również postępujemy zgodnie z zawartą w Konstrukcji 3.4.4 ideą i nowe akcje inhibitorowe łączymy ze wszystkimi nowymi wejściami do akcji a oraz odpowiednim miejscem inhibitorowym.



Rysunek 3.20: Problem w sieciach 3-pokojowych

Zaproponowane dla sieci 2-pokojowych i przytoczone dla sieci 3-pokojowych rozwiązanie rozszerzymy teraz na sieci k-pokojowe. Przedstawiona poniżej konstrukcja będzie miała charakter rekurencyjny:

Konstrukcja 3.4.20 Niech $S = (P, T, W, M_0)$ będzie p/t-siecią, zaś $S' = (P', T', W', M'_0)$ siecią k-pokojową symulującą w sensie osiągalnych markingów (to jest taką, że w sieci S osiągalny jest marking M dokładnie wtedy gdy w sieci S' osiągalny jest marking $10M$ markujący miejsca nie będące wyjściami z nowych akcji, a obcięty do zbioru P równy markingowi M). Skonstruujemy sieć k+1-pokojową $S'' = (P'', T'', W'', M''_0)$, która na podobnej zasadzie symulować będzie p/t-sieć S .



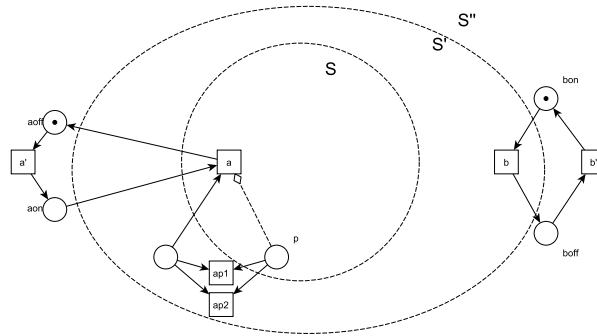
Rysunek 3.21: Symulowanie sieci k-pokojowej S' przez sieć k+1-pokojową S''

Dla każdej akcji $a \in T'$ ze skonstruowanej sieci k-pokojowej S' dodajmy nową akcję a' oraz dwa nowe miejsca a_{on} oraz a_{off} . Podobnie jak w szczególnych przypadkach tej konstrukcji, miejsce a_{on} będzie wyjściem z akcji a oraz wejściem do akcji a' , zaś miejsce a_{off} – wyjściem z akcji a' i wejściem do akcji a . Przez marking $1010M$ oznaczmy marking w sieci k+1-pokojowej S'' , który odpowiada markingowi $10M$ w sieci k-pokojowej S' oraz markingowi

M w p/t-sieci S w następujący sposób:

$$1010M(p) = \begin{cases} 10M(p) & \text{dla } p \in P' \\ 1 & \text{dla } p = a_{off} \text{ dla } a \in T' \\ 0 & \text{dla } p = a_{on} \text{ dla } a \in T' \end{cases} .$$

Za marking początkowy w sieci S'' przyjmujemy marking $M''_0 = 1010M_0$. Rozważmy teraz marking postaci $10M$ w sieci k-pokojowej S' oraz odpowiadający mu marking $1010M$ w sieci k+1-pokojowej S'' . Jeśli akcja a jest wykonalna w markingu $10M$ (w sensie wykonalności w sieci k-pokojowej), to w sieci k-pokojowej o identycznej jak S'' strukturze w markingu $1010M$ wykonalne są akcje a' a następnie a . Są więc one umożliwiające (w odpowiednich markingach) w sieci k+1-pokojowej S'' . Zarówno wykonanie w rozważanych markingach akcji a' jak i akcji a nie może uniemożliwić żadnej innej akcji wykonalnej w tych markingach w sieci k-pokojowej o identycznej jak sieć S' strukturze, gdyż żadna z akcji należących do P' nie jest wykonalna, a nowe akcje primowane nie mają wspólnych wejść z akcjami a' oraz a . Zauważmy, że w związku z tym, jeśli obliczenie $u = a_1a_2 \dots a_n$ jest wykonalne w sieci k-pokojowej S' , to obliczenie $u' = a'_1a_1a'_2a_2 \dots a'_na_n$ jest wykonalne w sieci k+1-pokojowej S'' . W związku z tym, dla każdego obliczenia $w = a_1a_2 \dots a_n$ wykonalnego w p/t-sieci S istnieje obliczenie $w' = v'_1a_1v'_2a_2 \dots v'_na_n$ wykonalne w sieci k+1-pokojowej S'' , gdzie v'_i są odpowiednimi obliczeniami złożonymi z akcji należących do zbioru $T'' \setminus T$. Poza tym, ze względu na brak bezpośredniej ingerencji przez akcje ze zbioru $T'' \setminus T$ w zawartość miejsc ze zbioru P , jeśli w sieci k+1-pokojowej S'' wykonalne jest obliczenie σ , to w p/t-sieci S wykonalne jest obliczenie będące rzutem obliczenia σ na zbiór T . Dowodzi to, że marking M jest osiągalny w p/t-sieci S dokładnie wtedy, gdy marking $1010M$ jest osiągalny w sieci S'' .



Rysunek 3.22: Symulowanie sieci inhibitorowej S przez sieci k+1-pokojową S'' (z pośrednictwem sieci k-pokojowej S'). Zaznaczono akcję a z oryginalnej sieci inhibitorowej S , akcje inhibitorowe ap_1 oraz ap_2 , które powstały w celu symulacji luków inhibitorowych przy pierwszym kroku operacji prowadzącej w k+1 krokach do sieci S'' (i nie są potem dodatkowo opóźniane), oraz akcję $b \in T' \setminus T$, powstałą w jednym z pośrednich kroków konstrukcji. Nowe akcje a' oraz b' zapewniają wymagane przy wykonalności odpowiednich akcji w sieci k+1-pokojowej S'' opóźnienie

Weźmy teraz sieć inhibitorową $S = (P, T, W, I, M_0)$ oraz symulującą ją sieć k-pokojowa $S' = (P', T', W', M'_0)$. Skonstruujemy sieć k+1-pokojową $S'' = (P'', T'', W'', M''_0)$, która symuluje działanie sieci inhibitorowej S . Nowe akcje, należące do zbioru $T' \setminus T$ można podzielić na dwie grupy - akcje zapewniające opóźnienie wykonania akcji ze zbioru T

oraz akcje zapewniające inhibitorowe działanie miejsc ze zbioru P . Dla każdej akcji z pierwszej grupy dodajemy, analogicznie jak w przypadku p/t-sieci, po jednej nowej akcji i po dwa nowe miejsca. Daje nam to takie samo zachowanie sieci $k+1$ -pokojowej S'' jak w przypadku symulacji p/t-sieci, pozostawiając działające akcje inhibitorowe odziedziczone po sieci k -pokojowej S' . Tak więc przy identycznym jak w przypadku symulacji p/t-sieci oznaczeniu markingu $1010M$, możemy wywnioskować, że marking M jest osiągalny w sieci inhibitorowej S dokładnie wtedy, gdy marking $1010M$ jest osiągalny w sieci S'' .

Ostatecznie więc, wydłużając opóźnienia, możemy symulować działanie p/t-sieci i sieci inhibitorowych za pomocą sieci k -pokojowych dla dowolnego k . Daje to nam płaską (powołując się na Tezę Churcha-Turinga), w sensie mocy obliczeniowej, hierarchię sieci pokojowych. Naturalne wydaje się zdefiniowanie, w sposób graniczny, pokojowej super-wykonalności (ω -wykonalności):

Definicja 3.4.21 Niech $S = (P, T, W, M_0)$ będzie p/t-siecią, zaś M markingiem osiągalnym w tej sieci. Mówimy, że akcja a jest ω -wykonalna w markingu M , o ile jest w tym markingu k -wykonalna dla dowolnie dużego k . Bardziej formalnie:

$$a \in EX_\omega(M) \Leftrightarrow \forall k a \in EX_k(M).$$

Definicja 3.4.22 Super-pokojową (ω -pokojową) siecią Petriego nazywamy sieć o strukturze p/t-sieci (czyli czwórkę $S = (P, T, W, M_0)$) zachowującą definicje kroku oraz obliczenia, gdzie marking M jest w relacji umożliwienia z akcją a wtedy i tylko wtedy, gdy dla dowolnego k mamy $M \mathcal{U} a$ w sieci k -pokojowej o identycznej jak S strukturze. Ponadto marking M jest w relacji wykonalności z akcją a wtedy i tylko wtedy, gdy dla dowolnego k mamy $M \mathcal{W} a$ w sieci k -pokojowej o identycznej jak S strukturze.

Definicja 3.4.23 Marking p/t-sieci jest ω -trwały, jeśli każda umożliwiona w nim akcja jest ω -wykonalna, zaś cała p/t-sieć jest ω -trwała, jeśli każdy marking w niej osiągalny jest ω -trwały.

Zauważmy, że w przypadku sieci ω -pokojowych relacje umożliwienia i wykonalności są tożsame. Ponadto, nie da się dostosować przytaczanych konstrukcji symulujących działanie p/t-sieci i sieci inhibitorowych. Próba przeniesienia ich na ten przypadek prowadzi do sytuacji, kiedy zbiory akcji i miejsc stają się nieskończone - co jest sprzeczne z definicją sieci. Pozostaje więc związany z tym otwarty problem:

Problem 3.4.24 Jaka jest moc obliczeniowa sieci ω -pokojowych?

Prawdziwy natomiast jest fakt o sieciach ω -trwałych, będący prostym wnioskiem z Lematu 3.4.13:

Wniosek 3.4.25 P/t-sieć $S = (P, T, W, M_0)$ jest siecią ω -trwałą (k -trwałą dla dowolnego k) dokładnie wtedy, gdy jest siecią trwałą.

Ponadto, dla sieci ω -pokojowych można sformułować odpowiednią wersję twierdzenia Kellera (zwanego też zasadą diamentu) i płynący bezpośrednio z tego twierdzenia wniosek (zwany też zasadą dużego diamentu):

Twierdzenie 3.4.26 (Kellera dla sieci ω -pokojowych) *Niech $S = (P, T, W, M_0)$ będzie siecią ω -pokojową. Wówczas, jeśli dwie różne akcje a, b są wykonalne w osiągalnym w sieci S markingu M , to akcja a jest wykonalna w markingu Mb zaś akcja b jest wykonalna w markingu Ma , a ponadto $(Ma)b = (Mb)a$.*

Dowód. Przez S' oznaczymy p/t-sieć odpowiadającą sieci S . Załóżmy, że akcje a, b są wykonalne w sieci ω -pokojowej S w markingu M , ale akcja b nie jest wykonalna w tej sieci w markingu Ma . Zauważmy na początku, że akcja b musi być umożliwiona w sieci S' w markingu Ma . W innej sytuacji w markingu M występowałby klasyczny konflikt dynamiczny między akcjami a i b , co powodowałoby, że akcja a nie byłaby nawet 1-wykonalna w markingu M , a to jest sprzeczne z założeniem o jej ω -wykonalności. Ponadto, przyjęte założenia oznaczają, że akcje a, b są dla dowolnego k k-wykonalne w sieci S' oraz istnieją takie n , że akcja b nie jest n -wykonalna w sieci S' , choć jest w niej $n-1$ -wykonalna. W związku z tym, wprost z definicji k -wykonalności mamy, że wykonanie w markingu M akcji a powoduje przerwanie n -wykonalności akcji b , czyli akcja a nie jest w markingu M $n+1$ -wykonalna. To z kolei jest sprzeczne z założeniami. Wnioskujemy stąd, że jeśli akcje a, b są ω -umożliwione w markingu M , to akcja b jest ω -umożliwiona w markingu Ma . Analogicznie, akcja a jest ω -umożliwiona w markingu Mb . Ostatnia część tezy wynika z ogólnej własności p/t-sieci - markingowy wynikowy nie zależy od kolejności wykonywania akcji (ważne jest tylko, aby w stanach pośrednich wszystkie one były wykonalne). \square

Wniosek 3.4.27 *Niech $S = (P, T, W, M_0)$ będzie siecią ω -pokojową. Wówczas, jeśli dwa różne obliczenia u, v są wykonalne w osiągalnym w sieci S markingu M , to istnieje takie obliczenie w wykonalne w sieci S w markingu M , że obliczenia u oraz v są jego śladowymi prefiksami. Bardziej formalnie:*

$$Mu \wedge Mv \Rightarrow \exists_w Mw \wedge [u] \sqsubseteq^{\equiv} [w] \wedge [v] \sqsubseteq^{\equiv} [w].$$

Warto zauważyć, że największą podklasą p/t-sieci spełniającą analogiczne twierdzenie, są sieci trwałe, których moc obliczeniowa jest zdecydowanie mniejsza niż całej klasy p/t-sieci (patrz Fakt 3.1.15). Wróćmy na chwilę do Przykładu 3.4.18:

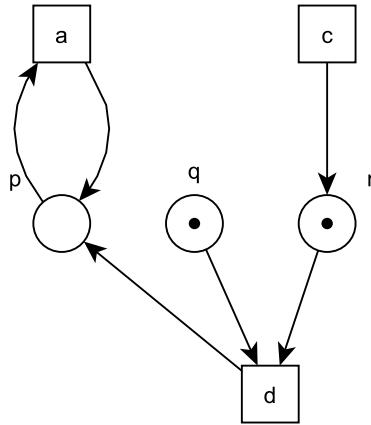
Przykład 3.4.28 Przypomnijmy, że sieć o przedstawionej w Przykładzie 3.4.18 strukturze, traktowana jako sieć k -pokojowa, zachowuje się identycznie dla dowolnych $k > 1$. Oznacza to, że zachowuje się tak również jako sieć ω -pokojowa. Zauważmy, że język obliczeń ω -wykonalnych w tej sieci jest językiem regularnym zadany wyrażeniem $c^*d(a \cup c)^*$, zaś zbiór markingów osiągalnych to zbiór półliniowy

$$[M_0]_{>\omega} = (q + r + k_1r) \cup (k_2p + k_3r),$$

gdzie k_i są dowolnymi liczbami naturalnymi, zaś q (i analogicznie p oraz r) rozumiemy jako funkcję działającą ze zbioru miejsc do liczb naturalnych (tak jak marking) o wartości jeden dla miejsca q oraz wartościach dla pozostałych argumentów równych zero.

Ponadto można skonstruować sieć ω -trwałą (a więc na mocy Wniosku 3.4.25 trwałą w klasycznym sensie), która ma identyczny język obliczeń wykonalnych oraz zbiór osiągalnych markingów:

Przedstawione powyżej fakty nasuwają dość naturalną hipotezę:



Rysunek 3.23: Sieć (super-)trwała działająca identycznie jak sieć ω -pokojowa z Rysunku 3.18

Hipoteza 3.4.29 Sieci ω -pokojowe są w sensie mocy obliczeniowej słabsze od maszyn Turinga, a nawet od p/t-sieci.

Podsumowanie

Rozprawa podejmuje temat nieskończonych procesów współbieżnych i bada go przez pryzmat jednego z najistotniejszych dla tych procesów zjawisk - konfliktu. W dyskusji wykorzystane są dwa matematyczne modele związane ze współbieżnością - ślady Mazurkiewicza opisujące procesy współbieżne i ich języki oraz sieci Petriego, opisujące systemy współbieżne. Oba te modele są szczegółowo i z dużą dbałością o formalizm matematyczny wprowadzone w rozdziale pierwszym.

Rozdział drugi koncentruje się na jednym z podejść do śladów Mazurkiewicza i bada zbiory rzutów takich śladów na podalfabety zależne. Jest to podejście znane z teorii śladów skończonych, ale mało w niej popularne. W przypadku badania procesów nieskończonych staje się ono bardziej efektywne - pozwala zredukować jeden wymiar nieskończony. Zamiast badać nieskończone zbiory (klasy równoważności) nieskończonych ciągów (linearyzacji procesu) możemy skupić się na odpowiedniej podklasie skończonych zbiorów nieskończonych ciągów (rzutów procesu). Nie tracimy przy tym żadnych informacji o samym procesie, co pokazuje przedstawiona w rozprawie procedura rekonstrukcji śladu z reprezentacji rzutowej. Wszegobecny w rozprawie konflikt, pojawia się tu zarówno w strukturalnej definicji wewnątrznie bezkonfliktowych zbiorów rzutowych, jak i w samym procesie rekonstrukcji, gdzie jego nieporządkanych efektów unikamy wprowadzając pojęcie warstw akcji.

Rozdział trzeci w całości skupia się na pojęciu konfliktu w sieciach Petriego. Myślą przewodnią tego rozdziału są procesy bezkonfliktowe p/t-sieci. Na początku rozdziału zostaje uporządkowana nomenklatura i rozdzielone dwa pojęcia - konfliktu statycznego i dynamicznego. W dalszej części przeprowadzona zostaje analiza procesów trwałych - prowadzi ona do zdefiniowania pojęcia sieci pokojowej, czyli p/t-sieci nierozwiązującej w czasie swojego działania żadnych konfliktów. Okazuje się, że tak zdefiniowany typ sieci daje możliwość synchronizacji i testowania pustości miejsca, a w efekcie symulowania sieci inhibitorowych. Zbadana zostaje też podklasa pokojowych sieci wolnego wyboru. Podobnie jak w przypadku klasycznym (p/t-sieci) okazuje się, że ograniczenie się do takich sieci statycznych, w których akcje mają wspólne lub rozłączne zbiory miejsc wejściowych powoduje zmniejszenie siły modelu. Argumentem świadczącym na korzyść tej tezy jest możliwość symulacji pokojowych sieci wolnego wyboru przez p/t-sieci. Jednocześnie pojawia się w tym miejscu pierwszy problem, który nie został rozwiązany w ramach rozprawy - czy pokojowe sieci wolnego wyboru pozwalają zasymulować działanie dowolnych p/t-sieci?

Dalsza analiza zachowań sieci pokojowych prowadzi do odnalezienia w nich sytuacji konfliktowych i, przez powtórzenie rozumowania z pierwszej części rozdziału, do zdefiniowania sieci 2-pokojowych. Procesy opisywane przez te sieci nie tylko nie rozwiązują żadnych konfliktów, ale nie występuje tam również znane z badań nad systemami

współbieżnymi zjawisko konfuzji. Znajduje ono swoje naturalne uogólnienie w postaci k -wykonalności (a właściwie przerwania wykonalności) akcji, ta zaś prowadzi do definicji sieci k -pokojowych. Przytoczony przykład pokazuje, że dana sieć, na kolejnych poziomach pokojowości, może mieć istotnie mniejsze zbiory wykonalnych obliczeń. Nie jest natomiast jasne, jak wygląda kwestia zbiorów osiągalnych markingów. Te, przynajmniej w rozwiązanych przykładach, stabilizują się. Może to oznaczać, że sieć ma zawsze swój skończony poziom konfliktowości (wyznaczony przez poziom k -wykonalności powyżej której zbiór markingów osiągalnych stabilizuje się).

Najciekawszych wniosków dostarcza ostatnia część rozdziału trzeciego. Jako efekt nieskończonego procesu pozbywania się konfliktów coraz wyższych rzędów, zdefiniowana zostaje klasa sieci super-pokojowych. Przystają tu jednak działać wypracowane wcześniej metody, a wyniki przeprowadzonych badań pozwalają sądzić, że podobnie jak pokojowe sieci wolnego wyboru, sieci super pokojowe nie są uniwersalnym modelem obliczeń. Wiele wskazuje na to, że ich siła wyrazu spada nie tylko do poziomu p/t -sieci, ale nawet do sieci trwałych, których zbiory markingów osiągalnych są półliniowe. Ustalenie faktycznej mocy obliczeniowej sieci super-pokojowych jest kolejnym ciekawym problemem, który nie doczekał się rozwiązania w ramach niniejszej rozprawy.

Postawione problemy pokazują, że przedstawiona w rozprawie dyskusja nie wyczerpuje tematu konfliktów w nieskończonych procesach współbieżnych, a stanowi jedynie wstęp do dalszych badań. Poza kierunkami wynikającymi bezpośrednio z przeprowadzonej analizy, zastanović się można nad tym, czy spełnianie twierdzenia Kellera (w wykorzystanej w rozprawie postaci) zawsze ogranicza możliwości danej klasy sieci? Warto przyjrzeć się temu problemowi przez pryzmat trwałych sieci inhibitorowych.

Bibliografia

- [1] Samson Abramsky and Achim Jung. Domain theory. In S. Abramsky, D.M. Gabbay, and T.S.E. Maibaum, editors, *Handbook of Logic in Computer Science*, pages 1–168. Clarenton Press, Oxford, 1994. (wersja poprawiona i rozszerzona).
- [2] Tilak Agerwala. A complete model for representing the coordination for asynchronous processes. Technical report, Hopkins Computer Research, 1974.
- [3] Toshiro Araki and Tadao Kasami. Some decision problems related to the reachability problem for Petri nets. *Theoretical Computer Science*, 3(1):85–104, 1976.
- [4] Kamila Barylska, Łukasz Mikulski, and Edward Ochmański. Nonviolence Petri nets. In *Proceedings of CS&P'09*, volume 1, pages 50–59. Warsaw University, 2009.
- [5] Kamila Barylska, Łukasz Mikulski, and Edward Ochmański. On persistent reachability in Petri nets. In *Proceedings of APNOC'10*, pages 1–12. Universidade do Minho, 2010. (wydane elektronicznie).
- [6] Kamila Barylska and Edward Ochmański. Levels of persistency in place/transition nets. *Fundamentae Informatica*, 93(1-3):33–43, 2009.
- [7] Hans-Dieter Burkhard. Ordered firing in Petri nets. *Elektronische Informationsverarbeitung und Kybernetik*, 17(2/3):71–86, 1981.
- [8] Hans-Dieter Burkhard. On priorities of parallelism: Petri nets under the maximum firing strategy. In *Logics of Programs and Their Applications*, volume 148 of *Lecture Notes in Computer Science*, pages 86–97. Springer, 1983.
- [9] Pierre Cartier and Dominique Foata. *Problèmes combinatoires de commutation et réarrangements*, volume 85 of *LNM*. Springer, Berlin, 1969.
- [10] Piotr Chrzastowski-Wachtel. Testing undecidability of the reachability in Petri nets with the help of 10th Hilbert problem. In *Application and Theory of Petri Nets 1999*, volume 1639 of *Lecture Notes in Computer Science*, pages 268–281. Springer, 1999.
- [11] Mireille Clerbout, Michel Latteux, and Yves Roos. Semi-commutations. In V. Diekert and G. Rozenberg, editors, *The Book of Traces [16]*, chapter 12, pages 487–552. World Scientific, Singapore, 1995.
- [12] Mireille Clerbout, Yves Roos, and Isabelle Ryl. Semi-commutations and partial commutations. *ITA*, 34(4):307, 2000.

- [13] Joerg Desel, Jetty Kleijn, and Wolfgang Reisig. The Petri net course - basic net classes. Tutorial presented at International Conference on Application and Theory of Petri Nets and other models of concurrency, 2009.
- [14] Jörg Desel and Javier Esparza. *Free choice Petri nets*, volume 40 of *Cambridge tracts in theoretical computer science*. Cambridge University Press, Cambridge, 1995.
- [15] Volker Diekert. On the concatenation of infinite traces. *Theoretical Computer Science*, 113(1):35–54, 1993.
- [16] Volker Diekert and Grzegorz Rozenberg, editors. *The Book of Traces*. World Scientific, Singapore, 1995.
- [17] Catherine Dufourd, Alain Finkel, and Philippe Schnoebelen. Reset nets between decidability and undecidability. In *Automata, Languages and Programming 1998*, volume 1443 of *Lecture Notes in Computer Science*, pages 103–115. Springer, 1998.
- [18] Javier Esparza and Mogens Nielsen. Decidability issues for Petri nets - a survey. *Bulletin of the EATCS*, 52:244–262, 1994.
- [19] Alain Finkel. The minimal coverability graph for Petri nets. *Lecture Notes in Computer Science*, 674:210–243, 1993.
- [20] Paul Gastin. Infinite traces. In *Semantics of Systems of Concurrent Processes*, number 469 in *Lecture Notes in Computer Science*, pages 277–308, Berlin - Heidelberg - New York, 1990. Springer.
- [21] Paul Gastin and Antoine Petit. Infinite traces. In V. Diekert and G. Rozenberg, editors, *The Book of Traces [16]*, chapter 11, pages 393–486. World Scientific, Singapore, 1995.
- [22] Jan Grabowski. The decidability of persistence for vector additions systems. *Information Processing Letters*, 11(1):20–23, 1980.
- [23] Michel H. T. Hack. Decidability questions for Petri nets. Technical Report 161, MIT, June 1976.
- [24] Ryszard Janicki and Maciej Koutny. Semantics of inhibitor nets. *Information and Computation*, 123(1):1–16, 1995.
- [25] Szymon Kaniewski, Łukasz Mikulski, and Edward Ochmański. Nonviolence Petri nets without weights and selfloops. (ukáže się w materiałach pokonferencyjnych warsztatów MASYW’10), 2010.
- [26] Richard M. Karp and Raymond E. Miller. Parallel program schemata. *JCSS*, 3:147–195, 1969.
- [27] Robert M. Keller. A fundamental theorem of asynchronous parallel computation. In *Parallel processing*, volume 24 of *Lecture Notes in Computer Science*, pages 102–112. Springer, 1975.

- [28] S. Rao Kosaraju. Decidability of reachability in vector addition systems. In *ACM Symposium on Theory of Computing (STOC '82)*, pages 267–281. ACM Press, 1982.
- [29] Marta Z. Kwiatkowska. Defining process fairness for non-interleaving concurrency. *FSTTCS: Foundations of Software Technology and Theoretical Computer Science*, 10:286–300, 1990.
- [30] L. H. Landweber and E. L. Robertson. Properties of conflict-free and persistent Petri nets. *Journal of ACM*, 25:352–264, 1978.
- [31] Yuri V. Matiyasevich. Diofantovost' perechislimykh mnozhestv. *Doklady Akademii Nauk SSSR*, 191(2):279–282, 1970. (po rosyjsku).
- [32] Ernst W. Mayr. An algorithm for the general Petri net reachability problem. In *ACM Symposium on Theory of Computing (STOC '81)*, pages 238–246. ACM Press, May 1981.
- [33] Ernst W. Mayr. Persistence of vector replacement systems is decidable. *Acta Informatica*, 15:309–328, 1981.
- [34] Antoni Mazurkiewicz. Concurrent program schemes and their interpretations. Daimi report pb-78, Aarhus University, 1977.
- [35] Antoni Mazurkiewicz. Introduction to trace theory. In V. Diekert and G. Rozenberg, editors, *The Book of Traces [16]*, chapter 1, pages 3–42. World Scientific, Singapore, 1995.
- [36] Łukasz Mikulski. Projection representation of Mazurkiewicz traces. *Fundamenta Informaticae*, 85:399–408, 2008.
- [37] Łukasz Mikulski. Internal conflict-free projection sets. *Artificial Intelligence and SC*, 59:497–504, 2009.
- [38] Tadao Murata. Petri nets: Properties, analysis and applications. *Proceedings of the IEEE*, 77(4):541–580, April 1989.
- [39] James L. Peterson. Petri nets. *ACM Computing Surveys*, 9(3):223–252, September 1977.
- [40] Carl A. Petri. *Kommunikation mit Automaten*. PhD thesis, University of Bonn, Bonn, Germany, 1962. (po niemiecku).
- [41] Carl A. Petri. Communication with automata. Technical report, Princeton, 1966.
- [42] Carl A. Petri and Wolfgang Reisig. Petri nets, 2008. <http://www.scholarpedia.org>.
- [43] Roman R. Rędziejowski. Associative omega-product of traces represented by projections. In *Proceedings of CS&P'09*, volume 2, pages 505–513. Warsaw University, 2009.
- [44] Wolfgang Reisig. *Sieci Petriego*. WNT, Warszawa, 1988.

- [45] Grzegorz Rozenberg and P. S. Thiagarajan. Petri nets: Basic notions, structure and behaviour. Number 224 in *Lecture Notes in Computer Science*, pages 585–668, Heidelberg, 1986. Springer-Verlag.
- [46] Mike W. Shields. Concurrent machines. *The Computer Journal*, 28(5):449–465, 1985.
- [47] Peter H. Starke. *Sieci Petri*. PWN, Warszawa, 1987.
- [48] Rüdiger Valk. On the computational power of extended Petri nets. In *Mathematical Foundations of Computer Science*, volume 64 of *Lecture Notes in Computer Science*, pages 526–535. Springer, 1978.