Piotr FORMANOWICZ
Instytut Informatyki, Politechnika Poznańska
Instytut Chemii Bioorganicznej, Polska Akademia Nauk

ZŁOŻONOŚĆ OBLICZENIOWA PROBLEMU POSZUKIWANIA PODZBIORÓW TRANZYCJI*

Streszczenie. Do konstrukcji modeli złożonych systemów biologicznych coraz częściej wykorzystuje się sieci Petriego. Mają one wiele zalet w stosunku do innych metod modelowania i analizy tego rodzaju systemów. Jedna z metod analizy modeli wyrażonych jako sieci Petriego oparta jest na t-niezmiennikach. W toku takiej analizy na ogół pojawia się potrzeba identyfikacji pewnych podzbiorów tranzycji, które odpowiadają elementarnym procesom o kluczowym znaczeniu występującym w badanym systemie. Problemy kombinatoryczne związane z poszukiwaniem takich procesów zbadane zostały dotąd w niewielkim stopniu, zwłaszcza od strony złożoności obliczeniowej. W niniejszej pracy przedstawiona została analiza złożoności czasowej jednego z takich problemów.

COMPUTATIONAL COMPLEXITY OF A PROBLEM OF FINDING SUBSETS OF TRANSITIONS

Summary. For constructing models of complex biological systems Petri nets are more and more often used. They have many advantages over other methods of modeling and analysis of such systems. One of the methods of analyzing models expressed as Petri nets is based on t-invariants. In the course of such analysis, there is usually a need to identify certain subsets of transitions that correspond to elementary processes of key importance occurring in the analyzed system. Combinatorial problems associated with the search for such processes have been studied to a small extent, especially from the computational complexity point of view. In this paper an analysis of the time complexity of one of such problems is presented.

1. Wstęp

Szybki postęp dokonujący się w naukach biologicznych powoduje, że coraz istotniejsza staje się analiza matematyczna oraz informatyczna zjawisk i obiektów biologicznych. Obecnie jest oczywiste, że bez niej bardzo trudno byłoby osiągnąć rzeczywiste zrozumienie zasad, na jakich funkcjonują organizmy żywe, tym bardziej, że sta-

*Badania częściowo przeprowadzone w ramach realizacji projektu finansowanego ze środków Narodowego Centrum Nauki przyznanych na podstawie decyzji numer DEC-2012/07/B/ST6/01537.

nowią one bardzo złożone systemy. We wspomnianej analizie kluczową rolę odgrywają matematyczne modele. Mogą być one tworzone za pomocą różnych narzędzi matematycznych, przy czym szczególnie często i z dobrym skutkiem stosowane są równania różniczkowe. Od pewnego czasu jednak coraz szerzej stosowane są metody oparte na różnego rodzaju grafach lub obiektach do nich zbliżonych. Wśród tego rodzaju metod na szczególną uwage zasługują sieci Petriego, które choć grafami nie są, to mają strukturę dwudzielnego ważonego grafu skierowanego. Ponadto, w stosunku do grafów mają istotną zaletę polegającą na tym, że można w stosunkowo prosty sposób modelować za ich pomocą (przynajmniej do pewnego stopnia) dynamikę badanego systemu. Jest ona odzwierciedlona poprzez tokeny przepływające przez sieć. Przepływ taki odpowiada przepływowi substancji, sygnałów itp. przez modelowany system [6, 5]. Istnieje wiele metod i narzędzi analizy sieci Petriego, jednakże w konteksie badania systemów biologicznych szczególnie istotne okazują się te oparte na t-niezmiennikach. Niezmienniki takie odpowiadają podprocesom, które nie zmieniają stanu modelowanego systemu. W toku analizy opartej na nich pojawia się często konieczność znalezienia pewnych podzbiorów tranzycji, które odpowiadają elementarnym procesom o kluczowym znaczeniu występującym w modelowanym systemie. Problemy kombinatoryczne związane z poszukiwaniem takich podzbiorów przebadane zostały dotąd w niewielkim stopniu (por. [2]). W niniejszej pracy przedstawiona jest analiza złożoności obliczeniowej jednego z tego rodzaju problemów.

Układ pracy jest następujący. W rozdziale drugim przedstawione jest bardzo zwięzłe wprowadzenie do sieci Petriego niezbędne do zdefiniowania analizowanego problemu oraz jego sformułowanie. W rozdziale trzecim zamieszczona jest analiza złożoności obliczeniowej badanego problemu. Praca kończy się podsumowaniem zawartym w rozdziale czwartym.

2. Sformułowanie problemu

Sieć Petriego może być zdefiniowana jako piątka $Q = (P, T, F, W, M_0)$, gdzie: $P = \{p_1, p_2, \ldots, p_n\}$ jest skończonym zbiorem miejsc, $T = \{t_1, t_2, \ldots, t_m\}$ jest skończonym zbiorem tranzycji, $F \subseteq (P \times T) \cup (T \times P)$ jest zbiorem łuków, $W : F \to \mathbb{Z}^+$ jest funkcją wagi, $M_0 : P \to \mathbb{N}$ jest oznakowaniem początkowym, $P \cap T = \emptyset \wedge P \cup T \neq \emptyset$ [6].

Sieć taka ma strukturę ważonego dwudzielnego grafu skierowanego. Występują w niej dwa rodzaje wierzchołków, nazywane miejscami i tranzycjami. Te pierwsze na ogół odpowiadają biernym składnikom modelowanego systemu (np. cząsteczkom), a te drugie aktywnym składnikom (np. reakcjom chemicznym). W miejscach mogą znajdować się tokeny reprezentujące liczby odpowiednich składników obecne w danym momencie w systemie. Tokeny przepływają pomiędzy miejscami poprzez tranzycje. Przepływem tym rządzi reguła uruchomienia tranzycji, zgodnie z którą tranzycja jest aktywna, jeżeli w każdym miejscu bezpośrednio ją poprzedzającym znajduje się liczba tokenów równa co najmniej wadze łuku łączącego to miejsce z daną tranzycją. Aktywna tranzycja może zostać uruchomiona, co oznacza, że następuje przepływ tokenów z miejsc bezpośrednio ją poprzedzających do miejsc bezpośrednio po niej następujących, przy czym liczba przepływających tokenów jest równa wadze odpowiedniego łuku [6].

Istotną zaletą sieci Petriego jest ich intuicyjna graficzna reprezentacja, jednak do celów formalnej analizy własności sieci, bardziej odpowiednia jest reprazentacja w postaci macierzy incydencji. Element a_{ij} tego rodzaju macierzy $A = (a_{ij})_{n \times m}$ jest równy różnicy liczb tokenów znajdujących się w miejscu p_i przed i po uruchomieniu tranzycji t_j [6]. Na podstawie macierzy incydencji można wyznaczyć m.in. t-niezmienniki. Niezmiennik taki jest wektorem x liczb całkowitych, który spełnia równanie $A \cdot x = 0$. Z t-niezmiennikiem x jest związany zbiór tranzycji $s(x) = \{t_j : x_j > 0\}$ nazywany jego wsparciem. Celem analizy opartej na t-niezmiennikach jest m.in. poszukiwanie podobnych do siebie niezmienników, które mogą odpowiadać podprocesom oddziałującym na siebie nawzajem w analizowanym systemie. Oddziaływania takie mogą być źródłem istotnych własności badanego systemu biologicznego. W przypadku, gdy liczba t-niezmienników jest duża, na ogół łączy się je w klastry, nazywane t-klastrami, by znaleźć podobieństwa. Podobne do siebie t-niezmienniki posiadają wsparcia, które mają części wspólne. Elementami tych części wspólnych są tranzycje odpowiadające procesom elementarnycm, poprzez które podprocesy oddziałują na siebie. Stąd, tranzycje takie mają istotne znaczenie, a jednym z celów analizy jest ich poszukiwanie [8, 4, 1].

W niniejszej pracy rozważany jest jeden z problemów poszukiwania zbiorów tego rodzaju tranzycji. Rozwiązaniem tego problemu jest odpowiednio duży podzbiór tranzycji, które występują we wsparciach t-niezmienników, które są elementami odpowiednio dużej liczby t-klastrów. Problem ten formalnie zdefiniowany jest w następujący sposób [2]:

PROBLEM WTWT-K (WYSTĘPOWANIE TRANZYCJI W T-KLASTRACH) – WERSJA PRZESZUKIWANIA 1
INSTANCJA: zbiór $T = \{t_1, t_2, \ldots, t_m\}$ tranzycji, zbiór $S = \{s_1, s_2, \ldots, s_p\}$ wsparć t-niezmienników, gdzie $\forall_{i=1,2,\ldots,p} s_i \subseteq T$, zbiór $C = \{c_1, c_2, \ldots, c_q\}$ t-klastrów, gdzie $\forall_{c_i \in C} c_i = \{s_{i_1}, s_{i_2}, \ldots, s_{i_{|C_i|}}\}$, $\forall_{i=1,2,\ldots,q; j=1,2,\ldots,q; i \neq j} c_i \cap c_j = \emptyset$, liczby $K \in \mathbb{Z}^+$ i $R \in \mathbb{Z}^+$.
ODPOWIEDŹ: zbiór $W = \{w_1, w_2, \ldots, w_r\} \subseteq T$ taki, że $\exists_{c_{i_1}, c_{i_2}, \ldots, c_{i_k} \in C} W \subseteq d_{i_1}, W \subseteq d_{i_2}, \ldots, W \subseteq d_{i_k}$, gdzie $\forall_{c_i \in C} d_i = \bigcup_{j=1}^{|c_i|} s_{i_j}$ oraz $k \geqslant K$ i $r \geqslant R$.

Zbiory d_i występujące w powyższej definicji nazywane są wsparciami t-klastrów i zawierają one tranzycje należące do wsparć t-niezmienników będących elementami danego t-klastra. W dalszej części niniejszej pracy wykazana zostanie silna **NP**-trudność powyższego problemu.

3. Złożoność obliczeniowa problemu WTwt-K

Do przeprowadzenia dowodu silnej **NP**-trudności PROBLEMU WTWT-K wykorzystany zostanie PROBLEM ZRÓWNOWAŻONEGO PEŁNEGO PODGRAFU DWUDZIELNEGO, który w wersji decyzyjnej może zostać zdefiniowany w następujący sposób [3]:

PROBLEM ZRÓWNOWAŻONEGO PEŁNEGO PODGRAFU DWUDZIELNEGO
INSTANCJA: Graf dwudzielny $G = (V, E)$, całkowita liczba dodatnia $\mathcal{K} \leqslant |V|$.
ODPOWIEDŹ: TAK, jeżeli istnieją rozłączne podzbiory $V_1 \subseteq V$ i $V_2 \subseteq V$, takie że $|V_1| = |V_2| = \mathcal{K}$ oraz $u \in V_1 \wedge v \in V_2 \Rightarrow \{u, v\} \in E$; NIE w przeciwnym przypadku.

Problem ten zostanie przekształcony za pomocą transformacji wielomianowej [3, 7] do wersji decyzyjnej pewnego problemu dotyczącego poszukiwania odpowiednio dużego podzbioru elementów w kolekcji zbiorów, tutaj nazwanego PROBLEMEM PODZBIORÓW, który zdefiniowany jest w następujący sposób [2]:

PROBLEM PODZBIORÓW – WERSJA DECYZYJNA
INSTANCJA: $T = \{t_1, t_2, \ldots, t_m\}$, kolekcja $C = \{c_1, c_2, \ldots, c_q\}$ podzbiorów zbioru T, liczby $K \in \mathbb{Z}^+$ i $U \in \mathbb{Z}^+$.
ODPOWIEDŹ: TAK, jeżeli istnieją zbiór $A \subseteq T$ i podkolekcja $C' = \{c_{i_1}, c_{i_2}, \ldots, c_{i_k}\}$ kolekcji C takie, że $A \subseteq c_{i_1}, A \subseteq c_{i_2}, \ldots, A \subseteq c_{i_k}$, przy czym $|A| \geqslant U$ i $k \geqslant K$; NIE w przeciwnym przypadku.

Lemat
PROBLEM PODZBIORÓW jest **NP**-zupełny.

Dowód
Można przyjąć, że instancja PROBLEMU ZRÓWNOWAŻONEGO PEŁNEGO PODGRAFU DWUDZIELNEGO dana jest w postaci: $G = (V, E)$, V_A, V_B, gdzie $V = V_A \cup V_B \wedge V_A \cap V_B = \emptyset$ oraz $\mathcal{K} \leqslant |V|$.

Na podstawie instancji PROBLEMU ZRÓWNOWAŻONEGO PEŁNEGO PODGRAFU DWUDZIELNEGO tworzona jest następująca instancja PROBLEMU PODZBIORÓW: $T = V_A$, $C = V_B$, $\forall_{v_i \in V_B} c_i = \{v_j : \{v_i, v_j\} \in E\}$, $K = \mathcal{K}$, $U = \mathcal{K}$.

$\Rightarrow$ Jeżeli odpowiedzią dla PROBLEMU ZRÓWNOWAŻONEGO PEŁNEGO PODGRAFU DWUDZIELNEGO jest TAK, to istnieją odpowiednie podzbiory $V_1 \subseteq V_A$ i $V_2 \subseteq V_B$ definiujące odpowiedni podgraf pełny. Oznacza to, że w zbiorze wierzchołków incydentnych każdego z wierzchołków ze zbioru V_2 znajduje się ten sam podzbiór wierzchołków V_1. Oba zbiory mają tę samą liczność K. Każdy element v_i zbioru V_2 odpowiada zbiorowi c_i z kolekcji C', a zbiór ten zawiera wszystkie wierzchołki, z którymi jest incydentny wierzchołek v_i, czyli co najmniej wszystkie wierzchołki ze zbioru V_1. A zatem w każdym ze zbiorów c_i zawarty jest zbiór V_1, który jest równy zbiorowi A z instancji PROBLEMU PODZBIORÓW. Zarówno kolekcja C', jak i zbiór A mają moc K, czyli odpowiedź dla PROBLEMU PODZBIORÓW brzmi TAK.

$\Leftarrow$ Jeżeli odpowiedzią dla PROBLEMU PODZBIORÓW jest TAK, oznacza to, że istnieje podkolekcja $C' = \{c_{i_1}, c_{i_2}, \ldots, c_{i_K}\}$ taka, że każdy z jej elementów zawiera zbiór A, przy czym zarówno C', jak i A mają moc równą K. Oznacza to, że w grafie G istnieje podzbiór wierzchołków V_2 odpowiadających elementom podkolekcji C'. Każdy z wierzchołków $v_i \in V_2$ jest incydentny z wierzchołkami odpowiadającymi elementom zbioru c_i, a ponieważ każdy z tych zbiorów zawiera zbiór A, to każdy wierzchołek $v_i \in V_2$ jest incydentny z wierzchołkami odpowiadającymi elementom zbioru A. Oznacza to jednocześnie, że każdy wierzchołek odpowiadający elementom zbioru A jest incydentny ze wszystkimi wierzchołkami ze zbioru V_2. A zatem zbiór A jest równy zbiorowi V_1, który, podobnie jak V_2, ma liczność K i razem z nim definiuje żądany podgraf dwudzielny pełny. Stąd, odpowiedź dla PROBLEMU ZRÓWNOWAŻONEGO PEŁNEGO PODGRAFU DWUDZIELNEGO brzmi TAK. □

W celu wykazania **NP**-trudności PROBLEMU WTWT-K w wersji przeszukiwania 1 można przetransformować wielomianowo PROBLEM PODZBIORÓW do decyzyjnej wersji PROBLEMU WTWT-K, która zdefiniowana jest w następujący sposób:

PROBLEM WTWT-K – WERSJA DECYZYJNA 1
INSTANCJA: zbiór $T = \{t_1, t_2, \ldots, t_m\}$ tranzycji, zbiór $S = \{s_1, s_2, \ldots, s_p\}$ wsparć t-niezmienników, gdzie $\forall_{i=1,2,\ldots,p} s_i \subseteq T$, zbiór $C = \{c_1, c_2, \ldots, c_q\}$ t-klastrów, gdzie $\forall_{c_i \in C} c_i = \{s_{i_1}, s_{i_2}, \ldots, s_{i_{|C_i|}}\}$, $\forall_{i=1,2,\ldots,q; j=1,2,\ldots,q; i \neq j} c_i \cap c_j = \emptyset$ oraz liczby $K \in \mathbb{Z}^+$ i $R \in \mathbb{Z}^+$.
ODPOWIEDŹ: TAK, jeżeli $\exists\ W = \{w_1, w_2, \ldots, w_r\} \subseteq T$ taki, że $\exists_{c_{i_1}, c_{i_2}, \ldots, c_{i_k} \in C} W \subseteq d_{i_1}, W \subseteq d_{i_2}, \ldots, W \subseteq d_{i_k}$, gdzie $\forall_{c_i \in C} d_i = \bigcup_{j=1}^{|c_i|} s_{i_j}$ oraz $k \geqslant K$ i $r \geqslant R$; NIE w przeciwnym przypadku.

Należy zauważyć, że problem ten może zostać sformułowany w następujący, równoważny z punktu widzenia złożoności obliczeniowej, sposób:

PROBLEM WTWT-K – WERSJA DECYZYJNA 1 BIS
INSTANCJA: zbiór $T = \{t_1, t_2, \ldots, t_m\}$ tranzycji, zbiór $D = \{d_1, d_2, \ldots, d_q\}$ wsparć t-klastrów, liczby $K \in \mathbb{Z}^+$ i $R \in \mathbb{Z}^+$.
ODPOWIEDŹ: TAK, jeżeli $\exists\ W = \{w_1, w_2, \ldots, w_r\} \subseteq T$ taki, że $\exists_{D' = \{d_{i_1}, d_{i_2}, \ldots, d_{i_k}\} \subseteq D} W \subseteq d_{i_1}, W \subseteq d_{i_2}, \ldots, W \subseteq d_{i_k}$, takie że $k \geqslant K$ i $r \geqslant R$; NIE w przeciwnym przypadku.

Jest tak dlatego, że wyznaczenie zbioru D występującego w instancji wersji decyzyjnej 1 bis PROBLEMU WTWT-K oraz w odpowiedzi wersji decyzyjnej 1 tego problemu na podstawie zbiorów S i C występujących w instancji tego ostatniego problemu jest trywialne, stąd, z punktu widzenia złożoności obliczeniowej, nie ma znaczenia, czy w instancji znajdą się zbiory S i C, czy zbiór D.

Zauważmy dalej, że PROBLEM WTWT-K w wersji decyzyjnej 1 bis jest równoważny PROBLEMOWI PODZBIORÓW, którego **NP**-zupełność została powyżej udowodniona. Można zatem sformułować następujące twierdzenie.

Twierdzenie 1
PROBLEM WTWT-K w wersji decyzyjnej 1 bis jest **NP**-zupełny.

Dowód
NP-zupełność PROBLEMU WTWT-K w wersji decyzyjnej 1 bis wynika wprost z **NP**-zupełności PROBLEMU PODZBIORÓW. □

Łatwo też zauważyć, na podstawie wcześniejszych rozważań, że zachodzi następujące twierdzenie dotyczące wersji decyzyjnej 1 PROBLEMU WTWT-K.

Twierdzenie 2

PROBLEM WTWT-K w wersji decyzyjnej 1 jest **NP**-zupełny.

Dowód
NP-zupełność PROBLEMU WTWT-K w wersji decyzyjnej 1 wynika wprost z **NP**-zupełności PROBLEMU WTWT-K w wersji decyzyjnej 1 bis. □

Oczywiście wersja przeszukiwania PROBLEMU WTWT-K jest niełatwiejsza niż jego wersja decyzyjna (por. [3]), a zatem zachodzi następujące twierdzenie dotyczące tej wersji.

Twierdzenie 3
PROBLEM WTWT-K w wersji przeszukiwania 1 jest **NP**-trudny.

Dowód
NP-trudność PROBLEMU WTWT-K w wersji przeszukiwania 1 wynika z **NP**-zupełności PROBLEMU WTWT-K w wersji decyzyjnej 1. □

Warto też zauważyć, że zachodzi jeszcze jedno twierdzenie dotyczące PROBLEMU WTWT-K w wersji przeszukiwania 1.

Twierdzenie 4
PROBLEM WTWT-K w wersji przeszukiwania 1 jest silnie **NP**-trudny.

Dowód
Zauważmy, że mimo iż w instancji PROBLEMU WTWT-K występują liczby całkowite dodatnie K i R, które mogą być dowolnie duże, to w przypadku, gdy K jest większe od q, czyli liczby t-klastrów (które są dane w instancji) lub R jest większe od m, czyli liczby tranzycji (które również są dane w instancji) problem jest trywialny, a odpowiedź brzmi NIE. Zatem rozważany problem można podzielić na dwa podproblemy – problem P_1 składający się z instancji, dla których zachodzi $K > q \lor R > m$ oraz problem P_2, który składa się z instancji, dla których zachodzi $K \leqslant q \land R \leqslant m$. Zauważmy, że problem P_2 jest problemem nieliczbowym, gdyż wszystkie liczby występujące w jego instancji są ograniczone od góry przez wielomian rozmiaru tejże instancji [3]. Maszyna Turinga rozwiązująca problem PROBLEM WTWT-K w wersji decyzyjnej 1 powinna najpierw sprawdzić, czy zachodzi $K > q \lor R > m$. Jeżeli tak, powinna zwrocić odpowiedź NIE. W przeciwnym przypadku powinna przejść do rozwiązania problemu P_2. Widać, że to ten problem powoduje, iż PROBLEM WTWT-K w wersji decyzyjnej 1 jest **NP**-zupełny, czyli oczywiście on sam musi być **NP**-zupełny. Ponieważ jednak problem P_2 jest **NP**-zupełny i nie jest problemem liczbowym, jest silnie **NP**-zupełny [3]. A zatem, również PROBLEM WTWT-K w wersji decyzyjnej 1 jest silnie **NP**-zupełny, z czego oczywiście wynika, że problem ten w wersji przeszukiwania jest silnie **NP**-trudny. □

4. Podsumowanie

W pracy przedstawiono dowód silnej **NP**-trudności jednego z problemów związanych z poszukiwaniem podzbiorów tranzycji w modelach złożonych systemów biologicznych opartych na sieciach Petriego. Tranzycje będące elementami takich podzbiorów mogą odpowiadać elementarnym procesom modelowanego systemu, gdyż występują w wielu grupach podprocesów, które się na niego składają. W toku analizy modeli wyrażonych jako sieci Petriego opartej na t-niezmiennikach często zachodzi potrzeba identyfikowania tego rodzaju kluczowych procesów elementarnych. Do niedawna było to jednak zajęcie żmudne (i jest tak w dużym stopniu nadal), gdyż nie istniały algorytmy wspierające poszukiwanie takich procesów (tranzycji). Co więcej, problemy kombinatoryczne, które są z tym związane tylko w niewielkim stopniu były przebadane od strony złożoności obliczeniowej (por. [2]). Problem rozważany w niniejszej pracy należy do jednego z ważniejszych z punktu widzenia praktycznych zastosowań, stąd wynik zaprezentowany w niej jest istotny, gdyż ukierunkowuje prace związane z konstrukcją odpowiednich algorytmów w stronę algorytmów przybliżonych.

LITERATURA

1. Formanowicz D., Kozak A., Głowacki T., Radom M., Formanowicz P.: Hemojuvelin-hepcidin axis modeled and analyzed using Petri nets. Journal of Biomedical Informatics 46, 2013, p. 1030–1043.
2. Formanowicz P.: Kombinatoryczne aspekty analizy t-niezmienników w modelach systemów biologicznych opartych na sieciach Petriego [w:] Świerniak A., Krystek J., (Red.): Automatyzacja procesów dyskretnych, Teoria i zastosowania", Wydawnictwo Pracowni Komputerowej Jacka Skalmierskiego, Vol II, Gliwice, 2016, p. 65–76.
3. Garey M.R., Johnson D.S.: Computers and Intractability. A guide to the Theory of NP-completeness. W. H. Freeman, New York 1979.
4. Grafahrend-Belau E., Schreiber F., Heiner M., Sackmann A., Junker B.H., Grunwald S. Speer A., Winder K., Koch I.: Modularization of biochemical networks based on classification of Petri net t-invariants, BMC Bioinformatics, 9, 2008, 90.
5. Koch I., Reisig W., Schreiber F. (Red.): Modeling in systems biology: the Petri net approach', Springer, London, 2011.
6. Murata T.: Petri Nets: Properties, Analysis and Applications. Proceedings of the IEEE, 77, 1989, p. 541–580.
7. Papadimitriou C.H.: Computational Complexity. Addison-Wesley Publishing Company, Reading, Massachusetts 1994.
8. Sackmann A., Heiner M., Koch I.: Application of Petri net based analysis techniques to signal transduction pathways BMC Bioinformatics, 7, 2006, 482.