Streszczenie

Niniejsza rozprawa doktorska, zatytułowana "State Space Reductions for Multi-agent Systems" i napisana pod kierunkiem prof. dr hab. inż. Wojciecha Penczka, przedstawia kilka technik ograniczających poważny problem w formalnej weryfikacji systemów wieloagentowych, jakim jest eksplozja przestrzeni stanów i tranzycji. Problem ten dotyczy w szczególności systemów asynchronicznych, w których konieczne jest uwzględnienie wszystkich możliwych ciągów akcji agentów w dowolnej kolejności. Jednak właśnie taka semantyka wykonania często jest formalizmem lepiej odpowiadającym potrzebom modelowania i weryfikacji rzeczywistych systemów, procesów i protokołów, które są zwykle, przynajmniej na pewnym poziomie abstrakcji, asynchroniczne. Tym bardziej świadczy to, że efektywne metody redukcji modeli są w praktyce niezbędne, zwłaszcza gdy mamy do czynienia z weryfikacją *zdolności strategicznej* autonomicznych agentów, charakteryzującą się znacząco wyższą złożonością obliczeniową niż w przypadku własności czysto temporalnych.

Rozdział 1 pokrótce przedstawia rys historyczny technik formalnej weryfikacji, w tym weryfikacji modelowej (ang. *model checking*) i uzasadnia potrzebę ich praktycznego stosowania. Ponadto omawia problem eksplozji przestrzeni stanów i tranzycji i wprowadza zagadnienie redukcji modeli.

Rozdział 2 zawiera podstawy teoretyczne niezbędne do późniejszego zdefiniowania metod redukcji, w szczególności definicje asynchronicznych systemów wieloagentowych (AMAS) i ich semantyki wykonania, składnię i semantykę logiki temporalnej czasu alternującego ($\mathbf{ATL}^*$), oraz kluczowe zagadnienia dotyczące zdolności strategicznej agentów, takie jak pojęcia strategii i jej wyniku.

Rozdział 3 zaczyna się od przywołania istniejącej formalizacji redukcji częściowoporządkowych (POR) dla logiki temporalnej czasu liniowego ($\mathbf{LTL}$), a następnie przedstawia adaptację tej techniki do $\mathbf{sATL}^*$, podzbioru $\mathbf{ATL}^*$ o nadal znacząco większej wyrażalności niż $\mathbf{LTL}$. Poprawność uzyskanych redukcji jest udowodniona w różnych wariantach semantycznych, uwzględniających zarówno strategie z pełną historią stanów jak i bezpamięciowe, subiektywną definicję zdolności strategicznej, a także rozszerzenie $\mathbf{sATL}^*$ o operator epistemiczny. Uzyskane wyniki są istotne zarówno z teoretycznego, jak i praktycznego punktu widzenia. Pozwalają bowiem na uzyskanie znaczącej redukcji modeli w nowym zastosowaniu, tj. weryfikacji własności dotyczących zdolności strategicznej agentów, nie zwiększając przy tym złożoności obliczeniowej algorytmu redukcji w stosunku do $\mathbf{LTL}$. Co więcej, możliwe jest wykorzystanie już

istniejących narzędzi implementujących algorytmy dla tej ostatniej logiki.

Rozdział 4 definiuje dwie wyspecjalizowane techniki redukcji w oparciu o wzorce (ang. *pattern-based reduction*) oraz o warstwy (ang. *layer-based reduction*). W odróżnieniu od redukcji częściowoporządkowych, można je zastosować dla dużo mniejszej klasy modeli. Pozwala to jednak na wykorzystanie ich charakterystycznych własności do uzyskania bardziej efektywnej redukcji niż umożliwiana przez POR. Jako przykład modeli o pożądanych własnościach, przede wszystkim z topologią synchronizacji pomiędzy komponentami będącą drzewem, zostają wykorzystane scenariusze bezpieczeństwa reprezentowane jako w postaci drzew ataku-obrony (ang. *attack-defence trees*, ADTree) po translacji do formalizmu automatowego (ang. *guarded update systems*, $\mathcal{GUS}$).

Rozdział 5 powraca do formalizmu ADTree, ale rozpatruje zagadnienie redukcji modeli z innej perspektywy, a mianowicie dążąc do zminimalizowania liczby agentów w systemie wieloagentowym. W tym celu formalizm AMAS zostaje rozszerzony, aby umożliwić reprezentowanie scenariuszy bezpieczeństwa z ADTree. W ten sposób zostaje uogólniona automatowa reprezentacja w postaci $\mathcal{GUS}$, rozważana w poprzednim rozdziale, która dotychczas nie brała pod uwagę agentów. Translacja z ADTree do AMAS pozwala na rozważanie scenariuszy ataku/obrony w nowym kontekście, uwzględniającym liczbę agentów w przeciwnych koalicjach oraz ich konkretny przydział do poszczególnych zadań. Zostaje zaproponowany algorytm optymalnego planowania zadań z wykorzystaniem minimalnej liczby agentów, a jego poprawność udowodniona i oceniona eksperymentalnie.

Rozdział 6 stanowi podsumowanie rozprawy doktorskiej oraz wytycza potencjalne kierunki dalszych badań naukowych.

Teofil Sidoruk

Abstract

The Ph.D. thesis "State Space Reductions for Multi-agent Systems", written under the supervision of Prof. Wojciech Penczek, discusses several approaches to mitigating the problem of state- and transition-space explosion, which is a major obstacle in the formal verification of multi-agent systems. Asynchronous multi-agent systems exacerbate this issue, requiring to account for all possible sequences of agents' actions, in any order. At the same time, they are often preferable to synchronous formalisms for modeling real-world systems, processes, and protocols. This emphasises the need for efficient reduction techniques, especially when dealing with the verification of *strategic abilities* of autonomous agents, whose computational complexity is significantly higher compared to the case of properties concerning just the temporal evolution of a system.

Chapter 1 recalls the historical background of formal verification techniques, including model checking, and discusses their relevance and necessity in practical applications. The associated issue of state explosion is introduced, as well as the idea of model reductions.

Chapter 2 introduces the necessary theoretical background, in particular the definitions of Asynchronous Multi-agent Systems (AMAS) and their execution semantics, the syntax and semantics of Alternating-time Temporal Logic (**ATL***), and key notions related to the strategic ability of agents.

Chapter 3 recalls the technique of partial order reduction (POR) as previously defined for Linear Temporal Logic (**LTL**) and subsequently demonstrates how it can be adapted to **sATL***, a subset of **ATL*** that remains significantly more expressive than **LTL**. The correctness of reductions is proven in various semantical settings, including both memoryless and perfect recall strategies, subjective strategic ability, and an epistemic extension of **sATL***. The obtained results are notable both in theory and practice, as they provide significant model reductions for the new purpose of verifying strategic ability properties at no additional computational cost, and potentially using existing tools implementing algorithms previously designed for **LTL**.

Chapter 4 defines two specialised techniques, called pattern-based reduction and layer-based reduction. While applicable to a much smaller class of models than POR, they are able to exploit specific characteristics of these models, thus yielding additional gains in model reduction. Attack-defence trees (ADTrees) are introduced and translated to a multi-agent setting (Guarded Update Systems, $\mathcal{GUS}$) in order to provide an example of models exhibiting the desired

characteristics, namely a tree synchronisation topology between components.

Chapter 5 revisits the setting of ADTrees, but looks at model reductions from a different perspective, aiming to minimise the number of agents in a multi-agent system. To that end, the formalism of AMAS is extended to represent ADTrees, superseding the previous formulation of $\mathcal{GUS}$ which did not include agents. Representing security scenarios from ADTrees as extended AMAS allows for studying a new aspect of these models, now considered in the context of two opposing agent coalitions of specific size and assignment to particular tasks. Therefore, an algorithm that synthesises an optimal assignment using a minimal number of agents is proposed, proved to be correct, and evaluated experimentally.

Chapter 6 concludes the thesis, summarising it and identifying several potential avenues of further research.

Teofil Sidoruk