Review of the PhD dissertation of Damian Kurpiewski

Damian Kurpiewski has submitted a dissertation with the title « *E-voting, Card Games, and Drone Teams: Verification of Strategic Properties in Multi-Agent Systems with Imperfect Information* », in accomplishment of his work towards obtaining the doctoral degree in Computer Science from the Institute of Computer Science of the Polish Academy of Sciences.

The subject of this dissertation is the verification of strategic abilities in multi-agent systems, and is related with the problem of automatically synthesizing programs which may be deployed on autonomous agents (robots, UAV, embedded controllers, game-playing agents) in a multi-agent environment, with the aim that the system behaves in a desirable way, as prescribed by some formal specification. This challenging and topical subject has seen some important theoretical advances in the last couple of decades, with the proposal of powerful logical formalisms for the specification of desired agent behaviors and modeling frameworks for the specification of agent capabilities, the exploration of the decidability frontier as well as the analysis of the complexity of their model-checking problem. The landscape is now populated with powerful logics which allow designing fine-grained controllers of autonomous agents satisfying game-theoretical and security inspired properties related with coalition power, absence of information leaks, anonymity or coercion-freeness. However the studies have shown that expressive power comes in pair with larger theoretical complexity, and even limiting hypotheses like « small » memory constraints on agents often induce at least NP-complete or even PSPACE-complete verification or synthesis problems. This limits the applicability of direct implementations of the resulting algorithms and calls for the utilization of efficient data structures, approximation heuristics, partial state-space exploration and abstraction techniques which avoid the state-space explosion problem, techniques which were successfully applied to the verification of distributed systems.

A **summary of the contents** of the thesis is the following :

After an introductive chapter briefly describing the research domain and listing the numerous published papers of the PhD candidate, the dissertation continues with a chapter presenting some background on logics for strategic abilities and their combination with reasoning aboud knowledge, with a particular emphasis on ATL with perfect and imperfect information, which is the topic of the thesis. Also a state of the art of tools for ATL model-checking is provided, and some challenges are presented, which the work of the PhD candidate addressed.

The third chapter focuses on providing the theoretical basis of one of the techniques for improving model-checking ATLir, namely « approximating » the truth value of ATLir formulas. The scope is to avoid the classical « guess and verify » algorithm for checking strategic ability in ATL in the presence of imperfect information, by adapting the fixpoint expansion of the coalition-temporal ATL operators. One type of adaptations is formalized through a series of fixpoint formulas which utilize different coalition-nexttime and/or epistemic operators. In this case, the author shows that these formulas represent either lower- or upper-approximations of the coalition-temporal operators, in the sense that, if the lower approximation of a coalition-temporal formula holds in a state, then the respective coalition-temporal formula also holds, and further if the upper approximation of that formula does not hold, then the respective coalition-temporal formula does not hold either. Another type of adaptation works for asynchronous multi-agent models, where the author proposes projecting the verification process on each agent's transition system, again by suitably translating the ATLir formulas. Again, the author provides guarantees for upper and lower approximation of the coalition-temporal operators. A third type of adaptation reported here is the modification of the model-checking algorithms by optimizing the data structures for representing the epistemic classes. Proposals include the disjoint-set data structure and layered state generation in the presence of stratified transition dependency. This chapter is well written and shows mastering of the semantic variations of temporal epistemic logics.

The fourth and the fifth chapter focus on formalizing a notion of strategic dominance and adapting it for strategy synthesis in the presence of reachability objectives. This notion is then utilized for designing a Depth-First Search-type algorithm, denoted DominoDFS. As a side result, the author shows that checking strategic dominance is co-NP-complete. Then, in the fifth chapter, the author proposes an algorithm for optimizing strategies w.r.t. some quality predicate and some heuristics preorder, and strategic dominance can be used as such a heuristics preorder. Two other types of heuristics are proposed : input and output dominance.

The sixth chapter focuses on abstraction techniques for reducing the state space. The author adapts the may/must techniques, by proposing an upper and a lower abstraction for each multi-agent model. Upper and lower translations of ATLir formulas are also provided, and the respective translations are proved to be sound, in the sense that, when the lower translation holds in the lower abstraction, then the original formula holds in the model, and the dual property for the upper abstraction and translation holds too. Another technique which is adapted from classical temporal verification and formalized here is stuttering equivalence. The author shows that this adaptation is sound and complete, in the sense that satisfiability of formulas in the non-nesting, nexttime-free fragment of ATLir is preserved by stuttering equivalence, and vice-versa.

The seventh chapter provides an important part of the candidate work : a series of different multi-agent system case studies which may be considered as the core of a future benchmark for STV verification. These case studies comprise : a Bridge endplay analysis, a drone patrolling system, a smart production factory, a social explainable AI framework, a simple voting model and a more sofisticated model abstracting some key components of the SELENE voting protocol, two mode « academic-style » case studies (the « castles and workers » and the Tian Ji model), and some randomly generated models. This chapter focuses on the encoding of the agent states, the modeling of agent interactions, capabilities and system dynamics, and the specification of relevant strategic and/or epistemic properties. Models are presented in a pedagogical manner, accompanied with figures which help the reader understand the agent capabilities and system dynamics. In the case of the SELENE voting system, ISPL code is also provided.

The eighth chapter focuses on the presentation of STV. The syntax of the modeling language is shown via a classical train-gate-controller example. This syntax is inspired by guarded commands and its shape seems to be inspired from tools like UppAal and Prism. The chapter also provides some implementation details like class structure and hierarchy.

The ninth chapter presents the second part of the experimental work of the candidate : the experiments on the verification of the relevant properties using the plain model-checking algorithms and the various improved algorithms and heuristics presented in Chapters 3-7 and implemented in the STV tool, accompanied in some cases by modeling and verification of the same case studies in different tools. The results of the verification are evaluated in each case by comparing the performance of the algorithms and heuristics, mainly regarding execution time and memory requirements. The evaluation of the performance of STV on these case studies exposes a tradeoff between different degrees of abstraction and completeness of the verification process, in that, in some cases, over-abstraction may compromise accuracy, leading to inconclusive results. Various methodological insights are provided in some cases, notably validating the efficience of state-space reduction techniques. Also these experiments validate the multi-agent ATL-based approach to the analysis of security protocols.

The thesis ends with a short conclusions chapter, in which the PhD candidate proposes a sketch of a methodology for modeling and verifying multi-agent systems, based on his experience with the implementation of model-checking algorithms and their applicaiton on the case studies.

Originality of the research : The work reported in the dissertation contributes novel and original results in the domain of the verification of strategic abilities in multi-agent systems. The novel contributions consist of the wide range of approximation methods, heuristics, state and data abstractions, and strategic improvement notions which are proposed in chapter 3-7, the production of a verification tool for multi-agent systems, STV, and the benchmark of case studies on which the abstractions and heuristics were experimented and validated. The part of the work which is not only novel but also original consists of the approximation semantics for ATLir and the resulting algorithms for model-checking, the adaptation of strategic dominance and the resulting heuristics for strategic improvement, and the sequence of results and case studies focusing on the specification of security properties and modeling and verification of voting protocols.

The work reported in the thesis has been published in top-ranked international journals (*Artificial Intelligence* 2019, *Journal of Computer Security* 2022, *Fundamenta Informaticae* 2020,), and some of the main conferences in AI and Multi-Agent Systems : IJCAI (2025, 2019), AAMAS (2024, 2022, 2021, 2019, 2017), ECAI (2024), AiML (2022), PRIMA (2022), ICAART (2024, 2023), EUMAS (2023), e-VOTE ID(2020, 2018), fact which shows that the work of the PhD candidate is considered as original and innovative by the research community in a few domains of Computer Science : multi-agent systems (IJCAI, AAMAS, ECAI, PRIMA, ICAART, EUMAS), but also modal logic (AiML), and formal security (e-VOTE ID).

Objectives and methods : Mr. Kurpiewski's thesis contributes to the effort of improving the model-checking algorithms for Alternating-time Temporal Logic with imperfect information (ATLir). The author's contributions are located at the theoretical level, at the implementation level and at the experiments level. At the theoretical level, Mr. Kurpiewski proposes semantic formulations of the approximation of the coalition modalities in ATL, based on fixpoint operators, strategy dominance and optimization, may/must abstraction and partial-order reductions. At the implementation level, the dissertation presents the development of a new model-checking tool for ATLir called *StraTegic Verifier* (STV). At the experiments level, the author reports on his experience on both modeling and verification for numerous case studies, stemming from voting protocols, UAV and robot control, social explainable AI, and game playing.

To this end, on one hand, Mr. Kurpiewski's work adapts various methods which are known to lead to improve the time and memory consumption of model-checking algorithms in the case of distributed systems and have never been applied to the verification of epistemic strategy properties of multi-agent systems : various types of state and data abstractions, semantic approximations or approximation heuristics, and adaptations of efficient data structures. On the other hand, the PhD candidate also explores proposes an adaptation of the notion of strategic dominance from Game Theory and studies the improvements it brings to the verification process. The author also adapts the methodology for validating implementations of model-checking algorithms from distributed systems by proposing a wide range of parameterized case studies and studying their performance by varying parameter values and observing the time and memory consumptions.

The results shows mastering of the intricate dependencies between algorithmic efficiency and semantics of multi-agent logics, which lies at the basis of the techniques proposed in the thesis for taming the state-space explosion problem. The dissertation also witnesses an important engineering effort during the PhD for the development of a new verification tool for ATLir.

Comments and remarks : As a general remark, the theoretical part from chapters 3-7 can be regarded as a reference of the heuristics for taming the state-space explosion problem in verification of strategic properties of multi-agent systems. Also the reviewer appreciates the important pedagogic effort in deployed while presenting the details of both modeling in most of the case studies, and on the specification of properties of interest, especially in the case of subtle variations in the ATL formulas for the voting protocols.

On a more per-chapter level, I appreciated the work in the seventh chapter, which shows the mastering of what some may consider as the « art of modeling » multi-agent systems which allows for accurate modeling and semantic fidelity. On the other hand, it would have been nice if the author was presenting also some methodology for practitioners which would help them towards ensuring accuracy of the models and properties to be verified, even by presenting the tricky parts of the modeling in each case, where they could come from and how the candidate managed to avoid or correct various modeling or specification traps.

Concerning the ninth chapter, the results of the extensive experiments reporting there are promising and suggest that, with appropriate support, the STV tool may become a standard in verification of multi-agent systems and a good basis for technology transfer in industry. On the other hand, the choice of focusing exclusively on performance is somehow limiting the usability of the reported results by practitioners. The reviewer would have expected that the PhD candidate synthesizes some guidelines for practitioners, in the sense of promoting the utilization of some of the heuristics or abstractions in the presence of specific system characteristics extracted from the various case studies. Examples are provided in the conclusion section – like the suggestion of a « best practice » to use hierarchical state representation as an abstraction technique, which works for Bridge endplay – or can be identified from the chapters 3-7 where the heuristics are presented. However a more systematic presentation of these « best practices » is missing from this chapter. Some critiques that can be made on the eigth chapter which does not provide some class diagram accompanying the implementation details, along with a guidelines of the correspondence between the algorithms presented in the previous chapters and the implemented methods. Also the syntax of the modeling langage and the accepted ATL formulas is only provided through examples. The reviewer encourages the candidate to prepare and publish a manual and a tutorial for STV users.

Conclusion : Based on the above arguments, I am happy to conclude that the doctoral thesis of Mr. Damian Kurpiewski:

1. Presents the candidate’s general theoretical knowledge in the discipline of Computer Science;
2. Presents the ability of the candidate to conduct independent scientific work;
3. Provides an original solution to a scientific problem.

Therefore, I recommend with no doubt the full acceptance of this PhD dissertation in accomplishment of Damian Kurpiewski’s work towards obtaining the degree of Doctor in Computer Science.

Créteil, April 23rd 2026

Catalin Dima,

Professeur des Universités,

Laboratoire d’Algorithmique, Complexité et Logique

Université Paris-Est Créteil