Proposing an Architecture to Integrate Stochastic Game and Automated Planning Methods into a Comprehensive Framework: CHIP-GT

Jane Jean Kiam,[1] **Régis Sabbadin,** [2] **Caroline P.C. Chanel** [3]

[1] Institute of Flight Systems (IFS), University of the Bundeswehr Munich
[2] Unité de Mathématiques et Informatique Appliquées de Toulouse (MIAT), INRAE
[2] Department of Aerospace Vehicles Design and Control (DCAS), ISAE-SUPAERO
jane.kiam@unibw.de, regis.sabbadin@inrae.fr, caroline.chanel@isae-supaero.fr

Abstract

Game theory and automated planning are complementary in decision making, as the former allows to consider the decision-making behaviors of other non-cooperative agents, while the latter can tackle more diverse problem domains and meanwhile, also ensures feasibility of plans. This paper describes the preliminary concept of an integral framework intended to combine methods arising from game theory and automated planning. Besides putting forth game-theoretic and automated-planning methods from previous works to be integrated into the framework, the interaction between these methods is also explained. Furthermore, extensions of these methods are also anticipated for more applicability of the framework, for example the consideration of multiple agents, and of unreliable communication.

Game-theoretic methods have proven use in determining strategies to be undertaken by considering the decision-making behaviors of other agents in the system (Neyman and Sorin 2003; Jasna, Supriya, and Nambiar 2017; Bondi et al. 2019). Typically, they allow the determination of strategies in cases where there is imperfect information, due either to a *lack of reliable communication* with the other (non-cooperative) players, or due to the nature of the game, where *decision-making agents play against each other*. However, game-theoretic methods can be computationally challenging; therefore, their applicability remains at a much higher abstraction level. In the context of biodiversity conservation, *Green Security Games* (Fang, Stone, and Tambe 2015), in which anti-poaching strategies are determined based on Stackelberg Security Games, the considered time step is monthly and the space is discretized into a small number of targets.

Stochastic Games (Eilon and Vieille 2015) are a framework generalizing both normal form games and Markov Decision Processes (Puterman 1994). They are a model of choice for planning under uncertainty with multiple agents. However, they also suffer from the complexity of computing equilibrium strategies and are limited to abstract models with a rough time/space discretization, or to the computation of stationary long-term equilibrium strategies (Dhamal et al. 2019). By analogy to (Fang et al. 2017), the game-theoretic background will be referred to as *Green Stochastic Games (GSG)*, which are based on Stochastic Games, while exploiting a use case that assimilates biodiversity conservation.

Meanwhile, automated planning has recently moved towards a more exciting landscape. In order to increase applicability of plans, many works have focused on bringing task and motion planning closer (de Silva, Pandey, and Alami 2013; Srivastava et al. 2014; Pecora et al. 2018; Kiam et al. 2020), so that plans are refined enough to be executed by the actuators of the planning agents. Furthermore, planning and acting, a paradigm coined by Ghallab, Nau, and Traverso (2016), has lately also become more in focus, typically to allow for generation of plans in a dynamic world. The Refinement Acting Engine (RAE) by Patra et al. (2020) is a prominent implementation of the paradigm, and uses hierarchical operational models to hierarchically plan for tasks by considering the continuously retrieved information from the dynamic environment. By doing so, it also leverages the advantage of hierarchical planning, i.e. causal-effect reasoning at different abstraction levels.

Combining both game-theoretic methods and automated-planning methods, specifically to enable planning and acting, can therefore be highly beneficial to solve planning problems in an environment with imperfect information, while the consideration of the decision-making behaviors of the non-cooperative agents is necessary. Some previous works, for example (LaValle 2000; Wang, Spica, and Schwager 2018) consider combining game theory and planning; however, their applications are limited to multi-agent motion or path planning.

1 Scope and contributions

We propose a framework, namely CHIP-GT (Coordination of Heterogeneous Interacting Planning Agents Using Game Theory), that aims to bring together stochastic game and automated planning methods (for general AI-planning and acting) in order to support a group of *protagonist agents* by devising executable plans against a group of *antagonist agents*. The problem class in question considers the following hypotheses (also see Figure 3 for a simplified illustration in the case of a Green Stochastic Game):

- There are L protagonist agents ($L \geq 1$) and K antagonist agents.
- The agents are capable of decision making and an action of a protagonist agent $l \in \{1, .., L\}$ (resp. antag-

onist agent $k \in \{1,..,K\}$) is denoted a_l^p (resp. a_k^a). The protagonist entities are intelligent and controllable agents[1] for which the planning capabilities are intended. Meanwhile, the antagonist agents affect the system but are non-controllable (by the CHIP-GT framework), and can be perfectly cooperative or perfectly non-cooperative among themselves.

- The protagonist and antagonist players are mobile in a world divided into N cells, denoted by $c \in \{1,..,N\}$. The resources level of each cell is denoted by $\rho_c \in \{1,..,\rho_{max}\}$.
- The protagonist agents communicate among themselves via a central server. Typically, the protagonist agents are cooperative among themselves, but can be non-cooperative, if the communication bridge becomes limited or unavailable. The relation among the antagonist agents is assumed to be cooperative.
- The relation between the protagonist and antagonist agents is non-cooperative, due to their opposing objectives, i.e. an antagonist agent k in cell $c = a_k^a$ "depletes" the resources ρ_c at the cell it is in, while the goal of the protagonist agents is to capture them before all resources are depleted or to capture as many as possible within a given time horizon.
- At the strategic level[2], the dynamics of the system is stochastic. We denote $p_{a^p,a^a}(s'|s)$ as the state transition matrix for the joint strategy (a^p, a^a) undertaken by the protagonist and antagonist agents, where $(a^p, a^a) \in \{1,...,N\}^{L+K}$, state $s \in S$ is a triplet (ρ, s^p, s^a), with ρ, s^a and s^p being the vector of remaining resources at each cell, the vector of positions of protagonist agents and antagonist agents respectively[3]. The transition is inherently stochastic, due to external processes (e.g. weather, dynamics of the resources, escaping capabilities of antagonists, etc.), or can result from non-deterministic decision-making behaviors of the agents (e.g. human behavior). Finally, the stochasticity also results from imperfect/absence of communication among agents in partially observable domains.

2 The CHIP-GT Framework

We conceptualize a preliminary version of the CHIP-GT framework to solve the class of problems that consider the above-mentioned hypotheses. Figure 1 is a representation of the framework in high-level functional blocks. The framework allows planning for multiple protagonist agents in a cooperative relation, but also considers cases in which one or more protagonist agents become non-cooperative with the rest of the group due to communication fall-outs.

In the case where the protagonist agents are in a cooperative relation, i.e. the multiple protagonist agents are coordinated centrally by the module *MA P&E*, that is enabled with a plan-replan logic (see Figure 2 for a more detailed view of the *MA P&E* module). The multi-agent coordination relies on the strategies for the protagonist agents (i.e. a^p in GSG) determined by the *Game-theoretic coordination* module based on the utility functions of the protagonist and antagonist agents. Subsequently, the *MA P&E* module requests to each *SA*$_l$ *P&E* modules of each controllable protagonist agent for calculating (or refining) the expected plan cost for a given macro-action a_l^p (i.e. strategy). The more "refined" cost calculation is performed by refining the action plan at a single-agent level, the *SA*$_l$ *Planning* module. Besides coordinating plans, the *MA P&E* module also monitors 1) the plan execution status by concatenating the plan execution status of each single agent (as monitored by the *SA*$_l$ *P&E* modules), as well as 2) the (detected) exogenous events, and 3) the (sensed) states of the non-controllable agents and the antagonist agents. In case non feasibility is detected during plan execution, either a plan repair (at the multi-agent or single-agent level) can be triggered by the *MA P&E*, or new strategy will be requested from the *Game-Theoretic Coordination* module.

Even protagonist agents who work *with* each other can be subject to unreliable communication, resulting thereby in a "non-cooperative" relation among themselves. A protagonist agent that is "detached" from the rest will therefore have to act on its own, which is analogous to the single agent in question being positioned in a non-cooperative game, not only "against" the antagonist agents, but also non-cooperative "with respect to" other protagonist agents, with which the communication (or information sharing) is imperfect. In this case, the strategy to be devised by the *Game-Theoretic Coordination* module will no longer be the one that is stochastically optimal for the ensemble of all protagonist agents, but optimal for a single protagonist agent under the assumption that it plays with (and against) all other non-cooperative protagonist agents (and antagonist agents). Note that the *Game-Theoretic Coordination* module is a functional module that, despite its position with a central interaction to other agents, can be physically hosted either as part of the central multi-agent coordination *MA P&E* or duplicated in every *SA*$_l$ *P&E.*

With CHIP-GT, we intend to put forth an integral framework capable of adopting tractable game theoretic methods, by limiting the use of them at a high-abstraction level, while enabling the refinement of the devised macro-actions into executable plans, by exploiting existing automated planning tools. Furthermore, we also allow the coordination of multiple agents and the consideration of a dynamic world by integrating a planning framework capable of anticipatory planning and anytime planning. Besides the integration of existing tools or framework, CHIP-GT is also intended to allow for extension of these.

[1]These agents can be humans, or mobile (manned or unmanned) vehicles.

[2]At the strategic level, the abstraction level considered is set at the rounds of the game, where the strategy (also referred to as "macro-action" in this paper) undertaken at each round is to move to (or stay at) a cell c_n. "How" to move to a cell is not considered at the strategic level.

[3]Antagonist agents which have been arrested are located at an arbitrary position, say "0", from which they cannot move anymore and influence the game.

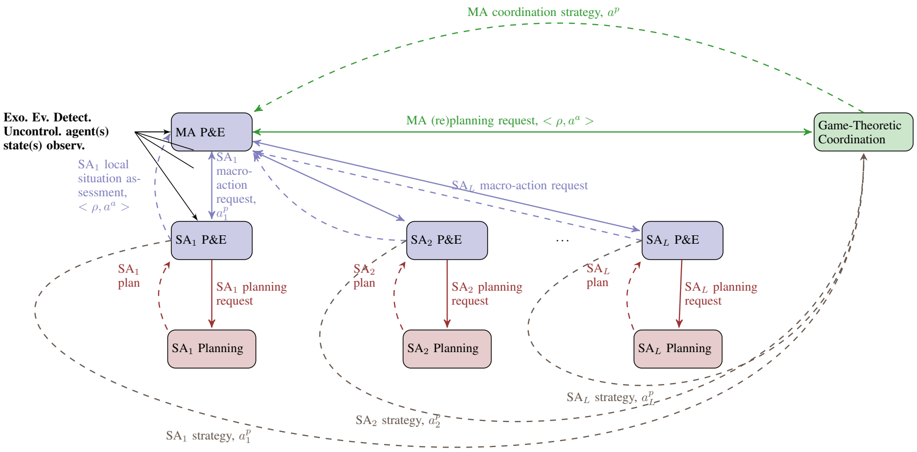

Figure 1: Planning and execution monitoring framework schema.
MA: Multi-Agent; SA_l: l-th Single Agent; P&E: Planning and Execution.

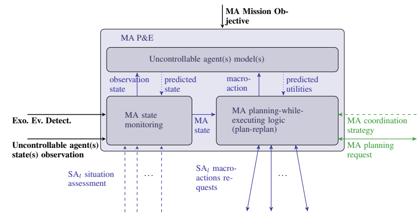

Figure 2: In-depth view of the MA P&E module.

2.1 Game-Theoretic Coordination Module

The *Game-Theoretic Coordination* Module focuses on a high-level multi-agent coordinating system to determine macro-actions for the controlled agents. It embeds the (protagonist and antagonist) agents' preferences in form of reward models, as well as available macro-actions into a game-theoretic framework.

This module leverages game theoretic tools for modelling interactions between agents, expressed in terms of simple actions and rewards, by exploring the concept of mixed Nash equilibrium, in particular, to model rational agents' behavior in a partially cooperative and non-cooperative context. Specific game-theoretic frameworks for sequential decision making, such as (partially-observed) stochastic games (Kumar and Zilberstein 2009), will be adopted in the module.

In addition to exploiting existing game-theoretic methods, multi-agent reinforcement learning tools for the "efficient" computation of approximate Nash equilibria in repeated games with incomplete information such as in (Zhang, Yang, and Başar 2021) can be adapted and exploited in view of enabling the players to learn and update their knowledge about the game model (other players' preferences) while planning for their own goals.

2.2 The Planning Modules

While the *Game-Theoretic Coordination* module deals with the determination of macro-actions at the strategic level, in order to obtain executable actions for the controllable protagonist agents, the refinement of the macro-actions into executable action plans π_l for each controllable protagonist agent l is performed by the *SA_l Planning* module. Note that the plan π_l is a sequence of time-stamped executable actions.

The refinement of macro-actions will be performed similarly to (Kiam et al. 2020), in which task and motion planning for multiple agents are done in an interleaved manner. Here, we will proceed instead by interleaving the request of a macro-action to the *MA P&E* module and the refinement of the macro-action into executable actions in the *SA_l planning* module.

Another manner to refine the macro-action is by exploiting hierarchical operational models, similar to the approach adopted by Patra et al. (2020), with which the online incremental refinement will enable the consideration of the dynamics in the environment the agent acts upon.

2.3 Interleaving Game Theory and Automated Planning

Due to partial observability and lack of information sharing, game-theoretic models constructed in the *Game-Theoretic Coordination* module may not correspond to reality. Moreover, in the case where non-controllable (antagonist) agents are considered, their behaviors can at best be predicted only at the decision phase. In order to ensure robustness of CHIP-GT, discrepancy between the truth and the models caused either by simplified assumptions or by the versatility of the system must be coped with.

These discrepancies will be dealt with by introducing a planning and execution (P&E) monitoring module that will compare the evolution of the world state predicted at the decision/planning time instant to the real evolution. The methods to be included in the *P&E* modules will be devised to: (i) evaluate the discrepancies, (ii) alarm the protagonists in case a re-coordination is necessary, and depending on the situation, (iii) send new planning or re-planning requests to the high-level *Game-Theoretic Coordination* module or to the single-agent planners, *SA_l Planning*. The AMPLE framework by (Chanel et al. 2019) will be integrated as part of the *SA P&E* modules, and extended for the *MA P&E* module, to support the P&E function in the coordination of multiple agents to check for the inter-dependency between protagonist agents. AMPLE is a framework where planning and execution are not interleaved but run in two parallel concurrent processes. The planning thread receives planning requests and delegates action selection to the embedded planning software, which is strictly anytime, reactive and conditional. On the other hand, the execution thread orchestrates these planning requests by applying *planning-while-executing* logics regarding hypothetical future states as well as action execution and state monitoring. The *MA P&E* module will also be extended to cope with local plan repair (when at least one predicate describing the world state becomes false) or pro-active planning (based on hypotheses about future states, resource conflicts among agents, future exogenous event, etc.).

In the ideal scenario, all protagonist agents are in communication with each other through the central *MA P&E* block. In most realistic settings, this is not always possible, resulting in either the fall-out of a single or several protagonist agents from the centralized communication network. The fallen-out protagonist agent is in this case in non-cooperative relation with respect to other protagonist agents as well. Therefore, it is necessary to include coordination and planing (as well as plan execution) capabilities to the individual protagonist agents.

Each fallen out protagonist agent's *SA P&E* module will continue to monitor the last retrieved *plan memory*, i.e. plans to be executed by the other protagonist agents before the fall-out happens, and will continue to monitor the executability of the plan, which may no longer hold true due to a number of causes, among which are 1) the fallen out cooperative relation, e.g. if the reward of the ensemble of cooperative protagonist agents in the game model is reduced due to the fall out, 2) if the inter-dependency of executable actions between the fallen out protagonist agent and the others can no longer be verified, 3) if the fallen out agent has now a different reward function than when it was in a cooperative relation with the other protagonist agents, etc. Once infeasibility is identified, the *SA P&E* module will request for new coordination strategies to be computed by the *Game Theoretic Coordination* module (via the interactions marked in gray in Figure 1). The latter will exploit another game model by assuming a *non-cooperative* relation between this single fallen out protagonist agent and the other protagonist agents. The new strategy will be used to repair the plan of the fallen out protagonist agent, by considering also the irreversible effects and incomplete non-primitive actions.

3 Example Use Case: Green Stochastic Game

The CHIP-GT framework can be applied to a Green Stochastic Game, in which a group of protagonist agents consisting of rangers and patrolling unmanned aerial vehicles defend animals in conservation areas against poachers, while the state transition follows a stochastic model (see Figure 3). The Green Security Game framework (Fang, Stone, and Tambe 2015) and its variants (Fang et al. 2017; Bondi et al. 2019) have received much attention lately. However, to the best of our knowledge, stochastic games approaches to anti-poaching games have not been considered yet. The latter framework differs both in the kind of equilibrium searched for (Nash vs Stackelberg) and the inherently multi-stage nature of the game. Moreover, embedding planning capabilities within a game-theoretic strategic framework, as we suggest, will allow to go beyond strategic planning at a high abstraction level, ensuring thereby more plan feasibility.

Besides anti-poaching, similar problems in fighting crimes involving the smuggling of drugs and goods (Krebs, Costelloe, and Jenks 2003), or event security patrolling (Jain et al. 2010) are use cases that CHIP-GT can be applied to.

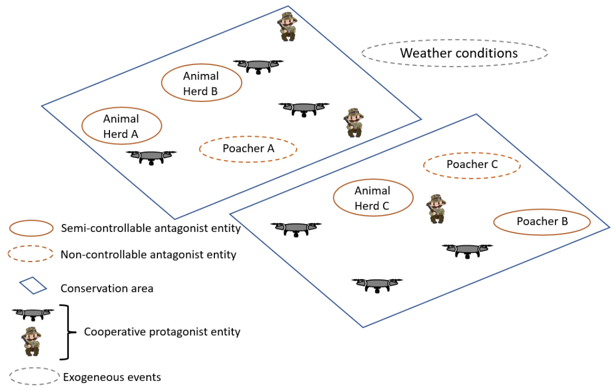

Figure 3: An example GSG-scenario of the patrolling operation in a nature conservation area.

4 Conclusion, Discussion and Future Work

A framework architecture is proposed in this paper to integrate stochastic-game methods and automated planning so that planning problems that include the presence of adversarial non-cooperative agents can be solved by leveraging also game-theoretic methods. Furthermore, the proposed framework CHIP-GT intends to be more comprehensive; therefore, preliminary considerations to include multi-agent planning and execution, as well as interleaving *MA P&E* and *SA P&E* in the case of unreliable communication among the protagonist agents are described.

Compared to other methods for multi-agent planning under uncertainties, for example Dec-POMDP by Spaan, Oliehoek, and Vlassis (2008) which considers the delay in the communication of information among multiple cooperative agents, as well as variants of Dec-POMDP combining planning and learning as described in Amato (2018), CHIP-GT does not only allow to consider an imperfect communication network among multiple controllable protagonist agents, but also a lack of information, or even deception from the non-cooperative and non controllable antagonist agents, who play against the controllable protagonist agents.

In addition to the integration of existing game theoretic and automated planning methods mentioned in Section 2.1 to 2.3, as well as the extension of these in CHIP-GT, as discussed in the previous sections, another important, yet challenging aspect to consider is a proper formalism, which will facilitate the development of a suitable domain definition interface, so that CHIP-GT can be (re-)usable for different problems.

References

Amato, C. 2018. Decision-Making Under Uncertainty in Multi-Agent and Multi-Robot Systems: Planning and Learning. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI)*, 5662–5666.

Bondi, E.; Oh, H.; Xu, H.; Fang, F.; Dilkina, B.; and Tambe, M. 2019. Biodiversity Conservation with Drones: Using Uncertain Real-Time Information in Signaling Games to Prevent Poaching. In *International Joint Conference on Artificial Intelligence (IJCAI) -Workshop on Biodiversity Conservation*.

Chanel, C. P. C.; Albore, A.; T'Hooft, J.; Lesire, C.; and Teichteil-Königsbuch, F. 2019. AMPLE: an anytime planning and execution framework for dynamic and uncertain problems in robotics. *Autonomous Robots*, 43(1): 37–62.

de Silva, L.; Pandey, A. K.; and Alami, R. 2013. An interface for interleaved symbolic-geometric planning and backtracking. In *International Conference on Intelligent Robots and Systems (IROS), IEEE.*

Dhamal, S.; Ben-Ameur, W.; Chahed, T.; Altman, E.; and Sunny, A. 2019. A stochastic game framework for analyzing computational investment strategies in distributed computing. Al-01870871v5.

Eilon, S.; and Vieille, N. 2015. Stochastic Games. *Proceedings of the National Academy of Sciences*, 45(112): 13743–6.

Fang, F.; Nguyen, T.; Sinha, A.; Gholami, S.; Plumptre, A.; Joppa, L.; Tambe, M.; Driciru, M.; Wanyama, F.; Rwetsiba, A.; Critchlow, R.; and Beale, C. 2017. Predicting poaching for wildlife Protection. *IBM Journal of Research and Development*, 61: 3:1–3:12.

Fang, F.; Stone, P.; and Tambe, M. 2015. When Security Games Go Green: Designing Defender Strategies to Prevent Poaching and Illegal Fishing. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI)*, 2589–2595.

Ghallab, M.; Nau, D.; and Traverso, P. 2016. *Automated Planning and Acting*. Cambridge University Press.

Jain, M.; Tsai, J.; Pita, J.; Kiekintveld, S., C.and Rathi; Ordóñez, F.; and Tambe, M. 2010. Software Assistants for Randomized Patrol Planning for The LAX Airport Police and The Federal Air Marshals Service. *Interfaces*, 40(4): 267–290.

Jasna, S.; Supriya, P.; and Nambiar, T. 2017. Application of game theory in path planning of multiple robots. In *International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT).*

Kiam, J. J.; Scala, E.; Ramirez Javega, M.; and Schulte, A. 2020. An AI-Based Planning Framework for HAPS in a Time-Varying Environment. In *Proceedings of the International Conference on Automated Planning and Scheduling*.

Krebs, C.; Costelloe, M.; and Jenks, D. 2003. Drug Control Policy and Smuggling Innovation: A Game-Theoretic Analysis. *Journal of Drug Issues*, 33(133): 133–160.

Kumar, A.; and Zilberstein, S. 2009. Dynamic programming approximations for partially observable stochastic games. In *Proceedings of FLAIRS'09*.

LaValle, S. 2000. Robot Motion Planning: A Game-Theoretic Foundation. *Algorithmica*, 26: 430–465.

Neyman, A.; and Sorin, S. 2003. *Stochastic Games and Applications*. Kluwer Academic Press.

Patra, S.; Mason, J.; Kumar, A.; Ghallab, M.; Traverso, P.; and Nau, D. 2020. Integrating Acting, Planning, and Learning in Hierarchical Operational Models. In *Proceedings of the International Conference on Automated Planning and Scheduling*.

Pecora, F.; Andreasson, H.; Mansouri, M.; and Petkov, V. 2018. A Loosely-Coupled Approach for Multi-Robot Coordination, Motion Planning and Control. In *International Conference on Automated Planning and Scheduling (ICAPS).*

Puterman, M. 1994. *Markov Decision Processes. Discrete stochastic dynamic programming*. Wiley-Interscience.

Spaan, M.; Oliehoek, F.; and Vlassis, N. 2008. Multiagent Planning under Uncertainty with Stocahstic Communication Delays. In *Proceedings of the 18th International Conference on Automated Planning and Scheduling (ICAPS)*, 338–345. AAAI Press.

Srivastava, S.; Fang, E.; Riano, L.; Chitnis, R.; Russell, S.; and Abbeel, P. 2014. Combined task and motion planning through an extensible planner-independent interface layer. In *IEEE International Conference in Robotics and Automation (ICRA).*

Wang, Z.; Spica, R.; and Schwager, M. 2018. Game Theoretic Motion Planning for Multi-Robot Racing. In *Proceedings of the International Symposium on Distributed Autonomous Robotics Systems (DARS).*

Zhang, K.; Yang, Z.; and Başar, T. 2021. *Vamvoudakis K.G., Wan Y., Lewis F.L., Cansever D. (eds) Handbook of Reinforcement Learning and Control*, volume 325 of *Studies in Systems, Decision and Control*, chapter Multi-Agent Reinforcement Learning: A Selective Overview of Theories and Algorithms. Springer.