AFRL-RI-RS-TR-2022-117

DISTRIBUTED LEARNING AND CONTROLLER DESIGN FOR ASSURED AUTONOMY

UNIVERSITY OF NOTRE DAME

AUGUST 2022

FINAL TECHNICAL REPORT

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

STINFO COPY

AIR FORCE RESEARCH LABORATORY
INFORMATION DIRECTORATE

■ **AIR FORCE MATERIEL COMMAND** ■ **UNITED STATES AIR FORCE** ■ **ROME, NY 13441**

NOTICE AND SIGNATURE PAGE

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

This report is the result of contracted fundamental research deemed exempt from public affairs security and policy review in accordance with SAF/AQR memorandum dated 10 Dec 08 and AFRL/CA policy clarification memorandum dated 16 Jan 09. This report is available to the general public, including foreign nations. Copies may be obtained from the Defense Technical Information Center (DTIC) (http://www.dtic.mil).

AFRL-RI-RS-TR-2022-117 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

FOR THE CHIEF ENGINEER:

/ S /
STEVEN L. DRAGER
Work Unit Manager

/ S /
GREGORY J. HADYNSKI
Assistant Technical Advisor
Computing and Communications
Division, Information Directorate

This report is published in the interest of scientific and technical information exchange, and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

REPORT DOCUMENTATION PAGE

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED	
		START DATE	END DATE
AUGUST 2022	FINAL TECHNICAL REPORT	MARCH 2020	MARCH 2022

4. TITLE AND SUBTITLE

DISTRIBUTED LEARNING AND CONTROLLER DESIGN FOR ASSURED AUTONOMY

5a. CONTRACT NUMBER	5b. GRANT NUMBER	5c. PROGRAM ELEMENT NUMBER
FA8750-20-2-0502	N/A	62303E
5d. PROJECT NUMBER	**5e. TASK NUMBER**	**5f. WORK UNIT NUMBER**
		R302

6. AUTHOR(S)

Vijay Gupta

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University of Notre Dame
Office of Research
940 Grace Hall
Notre Dame IN 46556-5708

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)	11. SPONSOR/MONITOR'S REPORT NUMBER(S)
Air Force Research Laboratory/RITA 525 Brooks Road Rome NY 13441-4505	AFRL/RI	AFRL-RI-RS-TR-2022-117

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited. This report is the result of contracted fundamental research deemed exempt from public affairs security and policy review in accordance with SAF/AQR memorandum dated 10 Dec 08 and AFRL/CA policy clarification memorandum dated 16 Jan 09.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Systematic ways to integrate diverse, heterogeneous, and possibly time-varying components into complex autonomous systems while guaranteeing system level properties define a holy grail in the science of assured autonomy. With much work being done already on topics such as safe machine learning or reinforcement learning to obtain guarantees on performance and safety of learning enabled autonomous systems (including through this program), this research effort focused on the next challenging step: how to provide guarantees on assured autonomy in a multi-agent system where multiple learning components are interacting. The project successfully completed design and analysis of new algorithms for distributed learning in both competitive and cooperative environments.

15. SUBJECT TERMS

Distributed learning, learning in games, safe reinforcement learning, multi-agent systems

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES
a. REPORT	b. ABSTRACT	C. THIS PAGE		
U	U	U	SAR	24

19a. NAME OF RESPONSIBLE PERSON	19b. PHONE NUMBER *(Include area code)*
STEVEN L. DRAGER	N/A

Table of Contents

1. Summary

Systematic ways to integrate diverse, heterogeneous, and possibly time-varying components into complex autonomous systems, while guaranteeing system level properties, define a holy grail in the science of assured autonomy. With much work already performed on topics such as safe machine learning or reinforcement learning to obtain guarantees on performance and safety of learning enabled autonomous systems (including through this program), this seedling research effort focused on the next challenging step: how to provide guarantees on assured autonomy in a multi-agent system where multiple learning components are interacting. The project successfully completed design and analysis of new algorithms for distributed learning in both competitive and cooperative environments.

2. Introduction

Many systems can be characterized by a collection of interacting subsystems that interact with each other, either explicitly cooperating as a team or competing but in a non-zero sum game manner. Of particular interest to this research are unmanned systems with such a structure. These systems have been envisaged in many different domains ranging from reconnaissance, to search and rescue, to mine detection and sweeping, to tactical missions. Imparting autonomy to these systems will not only decrease human casualties, but also enable an agile and multi-capable force leading to dominance.

In spite of significant advances, achieving assured, long-term autonomy, that will permit systems to respond in uncertain operating conditions for significant intervals of time, where they have to respond autonomously to unexpected changes in the environment, subsystem dynamics or configuration, and even goals and operating constraints, remains challenging. Traditional model-based techniques will likely fail in this quest, since obtaining good models in complex, uncertain, and time-varying environments is too much to ask for. Yet, new learning-based strategies are woefully under-developed at this time in the context of distributed control of teams of agents in a manner that can guarantee overall safety and performance required for high-confidence operation.

This seedling research effort focused on two problems in this context. The first problem focused on the design of a low-level controller that guarantees safety and stability in a compositional manner. Consider a large-scale system in which the dynamics of the subsystems are unknown and thus data driven methods are required to identify the subsystem and design (and update) controllers locally that can ensure that stability constraints are satisfied. While much theory is available for model identification, hardly any guarantees exist if controllers designed using a model are applied to the original system. Viewing this problem as purely a reinforcement learning based control design problem will not be scalable with current centralized or distributed reinforcement learning approaches. The approach taken here was to identify methods for control-oriented learning and compositional controller design such that global properties, such as stability and safety, can be tested for and guaranteed. In the second problem, the focus was on the design of a higher-level controller that identifies the optimal trajectories for exploring the area. Since no model of the environment is available, reinforcement learning is one of the few methods available. Most multi-

agent and distributed reinforcement learning algorithms assume that all agents share their current state, action, and possibly even reward with all other agents at every time step. However, given the harsh communication environment and possible presence of adversaries, algorithms for distributed reinforcement learning where agents share limited information only with neighbors, and where any communication may be maliciously changed, are required. This research effort developed such new distributed reinforcement learning algorithms.

3. Methods, Assumptions, and Procedures

Within the overall scope described above, this research effort considered various problem settings as discussed below:

3.1 Guaranteeing stability and safety in distributed systems with learning based control

For low-level control of distributed systems using learning algorithms, the research effort considered several directions. The main idea was to integrate system theoretic notions such as dissipativity, which has been used for compositional model-based control of large scale systems, with reinforcement learning algorithms. The research developed this idea in various directions.

One direction was to investigate the problem of verifying desired properties such as dissipativity of large-scale cyber-physical systems that operate in uncertain, adversarial environments. This effort proposed learning-based approaches to achieve verification with minimum knowledge of the system dynamics. In order to achieve compositionality in large-scale models, the research distributed the verification process among individual subsystems, which utilize limited information received locally from their immediate neighbors. The need for knowledge of the subsystem parameters was avoided via a novel reinforcement learning-inspired approach that enables decentralized evaluation of appropriate storage functions that can be used to verify dissipativity. The proposed method allows the addition of learning-enabled subsystems to the cyber-physical network while dissipativity is ensured. The research showed how different learning rules could be used to guarantee different properties, such as L2-gain stability or losslessness. Finally, the need for verification of complex properties was addressed by this effort.

Development of reinforcement learning based controllers at the subsystems that ensure stability and robustness for the entire networked system, especially when different agents may not use the same reinforcement learning algorithm, is a central problem in assured autonomy with multiple learning based components, but largely remains open. The effort considered the problem of guaranteeing stability when reinforcement learning is used for distributed control of networked dynamical systems. Specifically, consider a large scale system consisting of many subsystems that are coupled through their inputs and outputs, such as a network of microgrids. Each subsystem designs a local controller based on information about the subsystem state, inputs, and outputs. In

particular, the research assumed that the controller is implemented using a reinforcement learning algorithm, since the dynamics of the subsystems may be unknown. Of note, however, different controllers may potentially use different reinforcement learning algorithms. Which leaves the open research question, how to design controllers that guarantee that the entire system is still stable?

The research effort approached solving this problem by designing distributed controllers to stabilize a class of networked systems, where each subsystem is dissipative and designs a reinforcement learning based local controller to maximize an individual cumulative reward function. The developed solution enforces dissipativity conditions on the local controllers at each subsystem to guarantee stability of the entire networked system. The proposed approach was illustrated on a DC microgrid example; where the objective is to maintain voltage stability of the network using local distributed controllers at each generation unit.

The main contribution of this work is a distributed approach to ensure stability of a networked system with dissipative subsystems when the individual subsystems utilize reinforcement learning to design their own controllers. Beyond the specific stabilization problem that the research focused on, integrating dissipativity (and other input-output) specifications into reinforcement learning-based control is useful since it allows a wide landscape of tools from classical dissipativity theory to be integrated into reinforcement learning-based control design. The proposed algorithm guarantees stability irrespective of the choice of the reinforcement learning algorithm used at each subsystem. In particular, the results also hold for heterogeneous reinforcement learning algorithms being used at each subsystem. It should be noted that, as opposed to most existing literature on multi-agent reinforcement learning, the proposed approach requires only the output from neighboring subsystems to learn the control policy at each subsystem. In other words, to guarantee stability, no information about the states, rewards, or policies of other subsystems is required.

While the above setup was model free learning, the research team considered model-based learning setup as well, in the direction of guaranteeing system stability with learning based controller design. In model-based learning, it is desirable for the learned model to preserve structural properties of the system that may facilitate easier control design or provide performance, stability or safety guarantees. The approach considered an unknown nonlinear system possessing such a structural property - passivity, which can be used to ensure robust stability with a learned

controller. The research developed an algorithm to learn a passive linear model of this nonlinear system from time domain input-output data. The algorithm first learned an approximate linear model of this system using any standard system identification technique. The algorithm then enforced passivity by perturbing the system matrices of the linear model, while ensuring that the perturbed model closely approximates the input-output behavior of the nonlinear system. Finally, the algorithm derived a trade-off between the perturbation size and the radius of the region in which the passivity of the linear model guarantees local passivity of the unknown nonlinear system. This result can be used to ensure stability of the closed loop system when a controller is designed using a model learned via a learning algorithm.

Once stability is guaranteed, performance can be optimized. The problem of control policy design for decentralized state-feedback linear quadratic control with a partially nested information structure was studied for the case when the system model is unknown. A model-based learning solution was proposed, which consisted of two steps. First, the unknown system model was estimated from a single system trajectory of finite length, using least squares estimation. Next, based on the estimated system model, a control policy was designed that satisfied the desired information structure. It was shown that the suboptimality gap between the control policy and the optimal decentralized control policy (designed using accurate knowledge of the system model) scaled linearly with the estimation error of the system model. Using this result, an end-to-end sample complexity result was provided for learning decentralized controllers for a linear quadratic control problem with a partially nested information structure.

3.2 Reinforcement learning in the presence of adversaries

An important component of assured autonomy with learning based components is to develop an inexpensive, automated approach that can answer the question “how to guarantee resilient operation of safety-critical systems under faults or adversarial attacks?” Ensuring that safety-critical cyber-physical systems continue to satisfy correctness and safety specifications even under faults or adversarial attacks is very challenging, especially in the presence of legacy components for which accurate models are unknown to the designer. A major direction for the research was to consider the presence of adversarial agents in this setup.

Multi-agent reinforcement learning is based on cooperation among the agents. Agents seek policies that maximize the sum of utilities and all agents are expected to follow the prescribed algorithms. The first direction of the research was to show that classical multi-agent reinforcement learning algorithms are fragile to misbehaving agents. Recently, many cooperative distributed multi-agent reinforcement learning algorithms have been proposed in the literature. The effect of adversarial attacks on a network that employs a consensus-based multi-agent reinforcement learning algorithm was studied. It was shown that an adversarial agent can persuade all the other agents in the network to implement policies that optimize an objective that it desires. In this sense, the standard consensus-based multi-agent reinforcement learning algorithms are fragile to attacks. This reveals the crucial need to design new resilient multi-agent reinforcement learning algorithms for assured autonomy.

Given this fragility of current multi-agent reinforcement learning algorithms, a robust multi-agent reinforcement learning algorithm was designed. A fully decentralized network was considered, where each agent receives a local reward and observes the global state and action. A resilient consensus-based actor-critic algorithm was proposed, whereby each agent estimates the team-average reward and value function, and communicates the associated parameter vectors to its immediate neighbors. It was shown that in the presence of Byzantine agents, whose estimation and communication strategies are completely arbitrary, the estimates of the cooperative agents converge to a bounded consensus value with probability one, provided that there are at most H Byzantine agents in the neighborhood of each cooperative agent and the network is (2H+1)-robust. Furthermore, it has been proven that the policy of the cooperative agents converges with probability one to a bounded neighborhood around a local maximizer of their team-average objective function under the assumption that the policies of the adversarial agents asymptotically become stationary.

Current techniques for secure-by-design systems engineering do not provide an end-to-end methodology for a designer to provide real-time assurance for safety-critical systems by identifying system dynamics and updating control strategies in response to newly discovered faults, attacks or other changes such as system upgrades. A new methodology was proposed, along with an integrated software framework implemented to guarantee the resilient operation of safety-critical systems with unknown dynamics. The proposed framework consists of three main

components. The runtime monitor evaluates the system behavior on-the-fly against its correctness specifications expressed as signal temporal logic formulas. The model synthesizer incorporates a sparse identification approach that is used to continually update the plant model and control policies to adapt to any changes in the system or the environment. The decision and control module designs a controller to ensure that the correctness specifications are satisfied at runtime. For evaluation, the proposed framework was applied to ensure the resilient operations of two case studies.

3.3 Learning for agents interacting in a game set up

If agents are not cooperative, interactions among them can be considered in the form of a game. The strategies that the agents should follow (e.g. in a Nash equilibrium setup) can now be learned through suitable learning algorithms. This effort considered such a set up extensively.

One research direction was in the form of a meta-learning framework for games between adapting players. An agent with increased cognitive abilities was augmented with a structure that allows them to identify the way that their opponents learn during the game. This was achieved via approximators that are tuned online leveraging only observed actions from the environment. It was shown that knowledge of the utilities of the opponents enable asymptotic convergence of the approximation weights. The framework was then extended via backpropagation through time such that knowledge of the utilities was not necessary and convergence of the errors to a residual set was shown. Finally, simulations of players learning in a penny matching game demonstrated the efficacy of this approach.

Once learning algorithms of the opponents are identified, this information can be utilized to further gain utility in the game. Fictitious play is a popular learning algorithm in which players that utilize the history of actions played by the players and the knowledge of their own payoff matrix can converge to the Nash equilibrium under certain conditions on the game. The presence of an intelligent player that has access to the entire payoff matrix for the game was considered. It was shown that by not conforming to fictitious play, such a player can achieve a better payoff than the one at the Nash Equilibrium. This result can be viewed both as a fragility of the fictitious play algorithm to a strategic intelligent player and an indication that players should not throw away additional information they may have, as suggested by classical fictitious play. The main outcome

of this research path is that learning algorithms used by strategic agents in a competitive setting themselves can be utilized by adversaries to degrade performance.

Another research direction was to consider the possibility of collusion and incentives. Many scenarios in distributed systems require a system supervisor or operator to incentivize self-interested agents to exert costly effort to make decisions that align with the goal of the operator. For instance, in participatory sensing, a system operator requires many autonomous sensors to take measurements to allow estimation of a global quantity. The operator cannot observe directly the effort of each sensing agent (possibly for privacy reasons) and the agent might not benefit directly from the goal of the operator and thus needs to be compensated based on noisy outputs. This research considered the situation where the compensation or incentive needs to be designed by a learning algorithm. Specifically, the effort studied such a setup where a principal incentivizes multiple agents of different types who can collude with each other to derive rent. The principal cannot observe the efforts exerted directly, but only the outcome of the task, which is a noisy function of the effort. The type of each agent influences the effort cost and task output. For a duopoly in which agents are coupled in their payments, it was shown that if the principal and the agents interact finitely many times, the agents can derive rent by colluding, even if the principal knows the types of the agents. However, if the principal and the agents interact infinitely often, the principal can disincentivize agent collusion through a suitable learning-based contract.

4. Results and Discussion

The technical output of this research was summarized in quarterly reports and in the publications [SAG20, KFVG20, FKG21, KVGA21, NG21, FLLG21, KSS+21, YZG21, VKGV21, AVG22]. The results obtained and their significance for the problem settings mentioned in Section 3 are discussed below.

4.1 Design of Learning-based Distributed Controller Algorithms for Guaranteed Stability

For systems that consist of multiple interacting subsystems, control design to guarantee stability, performance or safety is a hard problem even when models are known. Some system theoretic properties such as dissipativity have proven useful for this problem; however, traditionally verification and controller design for guarantee of such properties have assumed accurate knowledge of models. The overarching contribution of this research was design of learning-based distributed controller algorithms that can be combined with dissipativity and similar properties for guaranteed stability.

In [KVGA21], the large scale system was modeled as cascade interconnections of linear time invariant subsystems. First, some system theoretic results were derived that may be of independent interest. Conditions to guarantee system stability based on L2-gain stability theorems were derived and those conditions were then manipulated to express the centralized conditions through a decentralized counterpart in which the properties of the subsystem can be considered individually. The reinforcement learning algorithm considered was Q-learning. Connections between Q-learning and dissipativity conditions were leveraged to restate the properties required on the L2 gain in a model free fashion. This is a very interesting result since learning-based function approximation can now be used to verify the properties on the individual subsystems and their coupling with their neighbors to guarantee the stability of the original system. In performing a centralized analysis, the L2 gain conditions can be relaxed to more general passivity conditions to expand the scope of the systems that can be considered. This work highlighted how verification of stability can be done using dissipativity properties of dynamic systems even with model-free reinforcement learning algorithms.

In [KSS+21], the research took the next step from verification of dissipativity (and hence stability) to design of controllers that guarantee dissipativity for subsystems, which in turn leads to stability of the overall system. Specifically, the problem of guaranteeing stability when reinforcement learning is used for distributed control of networked dynamical systems was considered. Consider a large scale system consisting of many subsystems that are coupled through their inputs and outputs, such as a network of microgrids. Each subsystem designs a local controller based on information about the subsystem state, inputs, and outputs using a reinforcement algorithm, since the dynamics of the subsystems may be unknown. Of note, however, different controllers may potentially use different reinforcement algorithms. How does one design the controllers that guarantee that the entire system is still stable? There are at least two challenges here. First, the control strategy should be distributed. While there exists a wide literature on reinforcement learning techniques for multi-agent systems, distributed control strategies using such techniques that provide guarantees like stability, safety, and robustness are still scant. Works that consider the problem of guaranteeing stability and robustness with reinforcement controllers have largely been limited to contexts such as model-based reinforcement learning and linear quadratic regulator designs for single-agent systems. Second, most available literature on multi-agent reinforcement learning considers the case when all subsystems implement the same algorithm and further share information such as a global state or rewards with other subsystems. Development of learning-based controllers at the subsystems that ensure stability and robustness for the entire networked system, especially when different agents may not use the same reinforcement learning algorithm, largely remains an open problem.

This research developed a reinforcement learning based distributed control design approach that exploits a dissipativity property of individual subsystems to guarantee stability of the entire networked system. The proposed approach is to use a control barrier function to characterize the set of controllers that enforce a dissipativity condition at each subsystem. This approach imposes a minimal energy perturbation on the control input learned by the reinforcement learning algorithm to project it to an input in this set. Together, these results guarantee the stability of the entire networked system, even when the subsystems utilize potentially heterogeneous reinforcement learning algorithms to design their local controllers.

To the team's knowledge, this is the first distributed approach to ensure stability of a networked system with dissipative subsystems when the individual subsystems utilize reinforcement learning to design their own controllers. Beyond the specific stabilization problem that the research focused on, integrating dissipativity (and other input-output) specifications into learning-based control is useful since it allows a wide landscape of tools from classical dissipativity theory to be integrated into learning-based control design. The proposed algorithm guarantees stability irrespective of the choice of the learning algorithm used at each subsystem. Further, the proposed approach requires only the output from neighboring subsystems to learn the control policy at each subsystem. In other words, to guarantee stability, no information about the states, rewards, or policies of other subsystems is required.

In [SAG20], this project considered the complementary problem when the reinforcement learning algorithms first learn a model of the system. As was shown above, dissipativity in the system model can be used to guarantee stability. Thus, the problem reduces to the following: Can one identify a system model that is dissipative and further the dissipativity level of the learned model provide some worst case guarantee on the dissipativity level of the true unknown system?

The problem of identifying a dissipative linear model of an unknown dissipative nonlinear dynamical system was solved using given time-domain input-output data. First, the approach learned an approximate linear model of the system, referred to as a baseline model, using standard system identification techniques. Next, the system matrices of this baseline linear model was perturbed to enforce quadratic dissipativity. It was shown that this perturbation could be chosen to ensure that the input-output behavior of the dissipative linear approximation closely approximates that of the original nonlinear system, provided that the baseline linear model closely approximates the nonlinear system dynamics in the input-output sense. Further, an analytical condition was provided relating the size of the perturbation to the radius in which local quadratic dissipativity properties of the nonlinear system can be guaranteed by the dissipative linear model. This relationship formalizes the intuition that larger perturbations lead to poorer approximations; in other words, the radius of local dissipativity of the nonlinear system decreases as the size of the perturbation is increased. Thus, the problem posed above is completely solved. While the proposed approach is offline, it is promising to extend the perturbation approach to quickly identify dissipative models in an online setting, where a baseline model is typically already available.

In [YZG21], the research took the first steps towards moving to performance guarantees. It is well known that optimal design of distributed controllers is a different problem, even when models are perfectly known. For learning-based controller designs, the project thus has to limit the approach to particular information structures and dynamics.

The research project, thus, considered a decentralized infinite-horizon state-feedback Linear Quadratic Regulator control problem with a partially nested information structure and assumed that the controllers do not have access to the system model. A model-based learning approach was utilized, where the system model was first identified, and then used to design a control policy that satisfies the prescribed information constraints. Using this approach, provided an end-to-end sample complexity result, which relates the number of data samples used for estimating the system model to the performance of the control policy. The performance of the control policy is characterized by the gap between the infinite-horizon cost of the control policy and that of the optimal control policy for the partially nested information structure, when the system model is known a priori. Surprisingly, despite the existence of the information constraint and the fact that the optimal controller is a linear dynamic controller, the sample complexity result matches with that of learning centralized control design without any information constraints.

4.2 Presence of Adversaries and Faults in Learning-based Controllers

The second big direction of research was to consider the presence of adversaries and faults when learning-based controllers were being used. This is an important direction both because in multi-agent systems, absence of strategic agents is a very strong assumption, but also because such controllers may be used in safety-critical systems where presence of faults can be catastrophic.

Research began in [FKG21] by showing that standard multi-agent reinforcement learning algorithms proposed in the literature are fragile to the presence of even one strategic agent. Specifically, the research considered a consensus-based multi-agent reinforcement learning algorithm with discounted rewards in the objective function. The attacks considered are different from the commonly studied data poisoning attacks in reinforcement learning, which seek to understand if changing the data or rewards by an external agent can degrade the performance of the learning algorithms. Instead, the project considered a setting where a participating agent itself is malicious. Specifically, the question asked was whether a single adversarial agent can either

prevent convergence of the algorithm, or even worse, lead the other agents to optimize a utility function that it chooses. It has been shown that the answer to this question is in the affirmative by designing a suitable attack and analyzing the convergence of the algorithm under it.

This work is important since it considers networks with adversaries that can compromise the consensus and critic updates and transmit corrupted signal values to its neighbors. It has been shown that when the malicious agent greedily attempts to maximize its own well-defined objective function, all other agents in the network end up maximizing the adversary's objective function as well. This study motivates the development of resilient multi-agent reinforcement learning algorithms.

In [FLLG21], such resilient multi-agent reinforcement learning algorithms were presented. The question considered was whether it was possible to design a consensus-based actor-critic multi-agent reinforcement learning algorithm with parametric function approximation for decentralized learning that is provably resilient to adversarial attacks, in the sense that the cooperative agents learn optimal policies in an environment influenced by the adversarial agents? It is important to note that the adversarial agents considered impact the other agents both due to the information they communicate to them as well as through implementation of control policies that affect the evolution of the state of the environment. To achieve resilience against adversarial attacks on control policies is difficult in the specified setting, as it does not assume that the agents are aware of the control policies of one another. The objective is to design a resilient algorithm that leads the cooperative agents to learn near-optimal policies in an environment that is affected by the adversarial agents. This is still a unique challenge because the adversarial agents can model attacks on communication channels that seek to degrade the network performance.

A novel resilient projection-based consensus method was introduced for decentralized actor-critic multi-agent reinforcement learning in which the cooperative agents estimate the critic and team average reward function that are crucial in the approximation of the true policy gradient. The algorithm includes two important steps that jointly facilitate a high degree of resilience in the critic and team-average reward function. In the first step, the received parameters are projected into the feature vectors that are the same for all agents, since the agents train linear models using the same basis functions. In the second step, the cooperative agents perform resilient aggregation in the

space of the estimated neighbors' estimation errors and apply the aggregated estimation error in the stochastic gradient descent update, which ensures diffusion of local data across the network. Both linear and nonlinear function approximations were considered. The proposed algorithms significantly reduce the impact of attacks on communication channels in training, and hence allow the cooperative agents to learn policies that maximize their team-average objective function.

4.3 Non-Cooperative Agents

The third major setting considered by the research was when the agents are not cooperative. If every agent has a different utility function, then their interactions can be considered as a game and concepts such as Nash equilibrium are more suitable to identify the optimal policies for the agents. Since these policies are usually difficult to identify, learning algorithms that converge to such policies have been proposed. This is a natural setting for distributed learning problem set up that was a focus on in this proposal, even though it was not a part of the originally proposed work.

The problem of agents using heterogeneous learning algorithms is even more important in a game setting. This makes it all the more surprising that almost all results for learning in a game setting assume homogeneous learning algorithms among the agents. In [KFVG20], the research addressed this problem. Specifically, a learning algorithm was formulated that enables agents to adapt their strategies while playing a repeated game in response to what other agents are playing. Subsequently, the meta-learning framework was formulated – a framework of gaining understanding of the learning algorithm – of an intelligent player via tuning algorithms that identify the decision making mechanism of the opponents. Finally, this algorithm was extended using backpropagation through time such that both the decision mechanism and the utility are learned. This is a significant contribution since this framework also allows the introduction of heterogeneity in cognitive abilities – in the vein of bounded rationality – to learning in games.

If the agents are not cooperating with each other, the assumption that they will faithfully relay information such as their utilities also becomes questionably. In [VKGV21], it was considered how this assumption can be removed. The research focused specifically on fictitious play for interaction between n + 1 players that play a matrix stage game repeatedly. The players are classified based on their information level where the first class consists of a single intelligent player who is aware of the complete game. The second class contains all the remaining players, referred

to as opponents, which are limited to the knowledge of their own payoffs for different strategy vectors. When all players employ fictitious play, under suitable conditions, the players converge to the Nash equilibrium. However, the intelligent player need not adhere to Fictitious Play. The question was asked: Can the intelligent player obtain a higher than Nash equilibrium payoff by deviating from fictitious play? Further, if there exists such a strategy profile, how does the intelligent player enforce it when the opponents are implementing fictitious play?

Under such a setting, strategies were identified that can deliver an expected payoff greater than the Nash and the Stackelberg equilibrium payoff for the intelligent player. For the case when there are 2 players in the game, the strategies that were identified are optimal for the intelligent player. For the general case of n + 1 players, a more tractable class of strategies were provided, termed as convergence based mixed strategies that may be sub-optimal, yet can provide an expected payoff greater than the Nash and the Stackelberg payoff for the intelligent players. A Linear Programming formulation was also provided that determines the strategy identified above without having to explore actions of all opponents at every time instant. Finally, a pure action trajectory was determined for the intelligent player that reaches the desired mixed strategy probabilities, while keeping the opponents in their fictitious play determined strategies.

This is an interesting contribution since it can be viewed both as a fragility of the fictitious play algorithm to a strategic intelligent player and an indication that players should not throw away additional information they may have, as suggested by classical fictitious play.

5. Conclusions

This project considered assured autonomy in a multi-agent system where multiple learning components are interacting. New algorithms were designed and analyzed in three main directions:

- algorithms to verify and guarantee properties such as stability in cooperative distributed control,
- algorithms to guarantee continued operation in the face of adversaries in multi-agent reinforcement control,
- and algorithms for learning in games.

New insights on how to integrate diverse learning components into complex autonomous systems while guaranteeing system level properties were obtained and the project was completed successfully. Various new directions for research have been noted for follow up.

6. References

[AVG22] Nayara Aguiar, Parv Venkitasubramaniam, and Vijay Gupta. Data-driven contract design for multi-agent systems with collusion detection. IEEE Signal Processing Letters, 29:1002–1006, 2022.

[FKG21] Martin Figura, Krishna Chaitanya Kosaraju, and Vijay Gupta. Adversarial attacks in consensus-based multi-agent reinforcement learning. In 2021 American Control Conference (ACC), pages 3050–3055. IEEE, 2021.

[FLLG21] Martin Figura, Yixuan Lin, Ji Liu, and Vijay Gupta. Resilient consensus-based multi-agent reinforcement learning. arXiv preprint arXiv:2111.06776, 2021.

[KFVG20] Aris Kanellopoulos, Filippos Fotiadis, Kyriakos G Vamvoudakis, and Vijay Gupta. A metalearning and bounded rationality framework for repeated games in adversarial environments. In 2020 59th IEEE Conference on Decision and Control (CDC), pages 1640–1645. IEEE, 2020.

[KSS+21] Krishna Chaitanya Kosaraju, Seetharaman Sivaranjani, Wesley Suttle, Vijay Gupta, and Ji Liu. Reinforcement learning based distributed control of dissipative networked systems. IEEE Transactions on Control of Network Systems, 2021.

[KVGA21] Aris Kanellopoulos, Kyriakos G Vamvoudakis, Vijay Gupta, and Panos Antsaklis. Dissipativity-based verification for autonomous systems in adversarial environments. In Handbook of Reinforcement Learning and Control, pages 273–291. Springer, 2021.

[NG21] Luan Nguyen and Vijay Gupta. Towards a framework of enforcing resilient operation of cyberphysical systems with unknown dynamics. IET Cyber-Physical Systems: Theory & Applications, 6(3):125–138, 2021.

[SAG20] S Sivaranjani, Etika Agarwal, and Vijay Gupta. Data-driven identification of approximate passive linear models for nonlinear systems. In Learning for Dynamics and Control, pages 338–339. PMLR, 2020.

[VKGV21] Bhaskar Vundurthy, Aris Kanellopoulos, Vijay Gupta, and Kyriakos Vamvoudakis. Intelligent players in a fictitious play framework. arXiv preprint arXiv:2110.05939, 2021.

[YZG21] Lintao Ye, Hao Zhu, and Vijay Gupta. On the sample complexity of decentralized linear quadratic regulator with partially nested information structure. arXiv preprint arXiv:2110.07112, 2021.

List of Symbols, Abbreviations, and Acronyms

Document contains none