Game Theory Approach in Multi-agent Resources Sharing

Tangui Le Gléau, Xavier Marjou, Tayeb Lemlouma, Benoît Radier

Game Theory Approach in Multi-agent Resources Sharing

Tangui Le Gléau
Orange Labs
Lannion, France
tangui.legleau@orange.com

Xavier Marjou
Orange Labs
Lannion, France
xavier.marjou@orange.com

Tayeb Lemlouma
IRISA
Lannion, France
tayeb.lemlouma@irisa.fr

Benoit Radier
Orange Labs
Lannion, France
benoit.radier@orange.com

***Abstract*—In multiple real life situations involving several agents, cooperation can be beneficial for all. For example, some telecommunication or electricity providers may cooperate in order to address occasional resources needs by giving to coopetitors some quantities of their own surplus while expecting in return a similar service. However, since agents are a priori egoist, the risk of being exploited is high.**

In this work, we propose to model this kind of situations as a social dilemma (a situation where Nash Equilibrium is non optimal) in which each agent knows only its own state. We design an algorithm modelling the agents whose goal is to make transactions in order to augment their own utility. The algorithm needs to be robust to defection and encourage cooperation.

Our framework modelling each agent consists in iterations divided in four major steps: the communication of demands/needs, the detection of opponent cooperation, the cooperation response policy and finally the allocation of resources.

In this paper, we focus on the cooperation response policy. We propose a new version of tit-for-tat and we evaluate it with metrics such as safety and incentive-compatibility. Several experiments are performed and confirm the relevance of our improvement.

***Index Terms*—Multi-Agent System, Game Theory, Social Dilemma**

I. Introduction

Sharing resources or services between multiple coopetitors with autonomous agents is very common in industrial use cases.

For example, in telecommunications, sharing resources between operators has been suggested, in particular with the arrival of the next generation of telecom (5G) in order to extend and improve capacity and coverage of operators connectivity. A well-suited model for exchanging connectivity resources is the framework called Licensed Shared Access (LSA) [1] which aims at optimising spectrum utilisation. Several previous works address the issue of LSA spectrum sharing, generally with auction mechanisms [2]–[5]. In particular, some of these works study truthfull mechanisms (such as Vickrey-Clarke-Groves mechanism) which has the particularity to have good properties (fairness, incentive-compatibility). This kind of mechanisms involves financial transactions. We instead focus on addressing the issue without money and we follow an utilitarian way, in particular to be able to deal with services which can't be shared financially. In this paper, we assume that network operators can exchange resources to reach an optimal situation and we consider that their personal interest are driven only by their personal utility (which can for example be considered as the quality of experience in telecommunications area). We introduce three major assumptions in the context. First, there is no main regulator, it is impossible for a controller to compute optimal transactions. Secondly, due to strategic issues, agents don't share all their personal state (i.e. the quantities of their under/overused resources), then they have to communicate only a partial state to other agents. Finally, each agent is assumed to be selfish, unlike, for example, consensus optimization problems where each node is encouraged to cooperate. This last assumption is very important as being exploited by a defector has an utility cost and reaching a consensus can be longer than other consensus algorithms due the risk of exploitation.

Games with partial information are generally solved with frameworks involving consensus optimisation issues which has been well studied in literature in optimisation [6]–[10] or synchronisation [11], [12]. In our work, we tackle the issue of self-interested agents, which means that agents have to reach a consensus while taking into account the fact that others agents are a priori selfish.

We propose to formulate the problem as a non-cooperative Markov game and show that it can be viewed as a social dilemma, i.e. a situation where Nash equilibria are non-optimal : agents have no incentive to cooperate despite the fact that mutual cooperation is the optimal global strategy. This typical game theory case has been well studied recently [13]–[17], involving in particular Reinforcement Learning Multi-Agent systems.

In section II, we define the problem as a multi-agent system with different items to be shared whose marginal utility is assumed to be decreasing function (i.e. utility function is concave). The game theory issues naturally raised by the problem are described in section III. To address this sharing problem, we propose an algorithm based on Tit-for-Tat (TFT), known to be robust to iterated dilemmas. We present the architecture of our model in section IV. Some numerical simulations are performed in section V to evaluate some key properties of our algorithm such as efficiency, speed, incentive-compatibility and safety.

II. Problem Formulation

We consider a game with N agents $A_1, ..., A_N \in \mathcal{I}$ and M items $B_1, ..., B_M \in \mathcal{K}$. To simplify the notation, we then consider that $\mathcal{I}$ and $\mathcal{K}$ are confounded with sets $[0, .., N-1]$ and $[0, .., M-1]$. $\mathcal{S}$ is the state set which corresponds to agents resources: at each time t, the state is defined by $s(t) = \{s_{i,k}(t),\ \forall i \in \mathcal{I}\ \forall k \in \mathcal{K}\}$, where $s_{i,k}(t)$ corresponds to the quantity of item B_k owned by agent A_i. We assume that the utility function of each agent A_i is $f^{(i)}(s(t))$ defined by:

$$f^{(i)}(s(t)) = \sum_{k=0}^{M-1} f_k^{(i)}(s_{i,k}(t)) \tag{1}$$

where each $f_k^{(i)}$ is considered monotonically increasing and concave due to the assumption of decreasing marginal utility of resource.

At each time t, each agent A_i executes an action $a^{(i)}(t)$ which is a set of transactions:

$$a^{(i)}(t) = (u^{(1)}, ..., u^{(m)})$$

where $u^{(k)} \in \mathcal{K} \times \mathcal{I} \times \mathbb{R}$ is a transaction (give one quantity of one item to one agent).
To simplify, we consider that actions are gathered in a joint action $X = (X_1, ..., X_M)$ where each $X_k \in \mathcal{M}_N(\mathbb{R})$ sums up the transactions of item B_k and is then anti-symmetric (giving Δ to another agent means receiving $-\Delta$ from him).
The transition function $\mathcal{T}$ maps states and actions with next states by executing transactions:

$$s(t+1) = \mathcal{T}(s(t), X = (X_1, ..., X_M)) \tag{2a}$$

$$s_{i,k}(t+1) = s_{i,k}(t) + \sum_{j=0}^{N-1} (X_k)_{i,j} \tag{2b}$$

Finally, we assume a game with partial observation: at each time t, agent A_i observes a part of the total state $\mathcal{O}(s(t), i) = \{s_{i,k}(t),\ \forall k \in \mathcal{K}\} \cup \{(X_k)_{i,j}(t'),\ \forall j,\ \forall t' < t\}$.

The objective is to maximise the social welfare which is the sum of utilities:

$$\max \sum_{i=0}^{N-1} f^{(i)}(s(t)) \tag{3}$$

where each agent A_i wants to maximise independently its own utility function $f^{(i)}(s(t))$ which leads to a non-optimal Nash equilibrium issues that will be described in the next section.

III. Game Theory

Our problem deals with game theory due to personal interests of agents. In this section, we introduce some game theory concepts and formulate our problem as a multi-agents social dilemma i.e. a situation where it is more profitable to cooperate (accept to exchange resources) but in a personal point of view, cooperating leads to the risk of being exploited.

A. Nash Equilibrium

Let us begin by defining a strategy, it is a function that maps personal states ($s^{(i)}$) to actions (quantities allocated to others agents):

$$\pi_i : s^{(i)} \mapsto (X_k)_{i,j} \tag{4}$$

If $G^{(i)}(\pi_i)$ is the payoff of agent A_i, a joint strategy $(\pi_i^*)_{i \in \mathcal{I}}$ is said to be a Nash Equilibrium if [18]:

$$\begin{aligned} &\forall i \in \mathcal{I},\ \forall \pi_i,\ G^{(i)}(\pi_i^*, \pi_{-i}^*) \geq G^{(i)}(\pi_i, \pi_{-i}^*) \\ &\text{with } \pi_{-i} = [\pi_0, ..., \pi_{i-1}, \pi_{i+1}, ..., \pi_{N-1}] \end{aligned} \tag{5}$$

B. Social Dilemmas

In the following section, we model our exchange problem as a sequential social dilemma [13]. Before defining the multi-agents continuous version, let us consider the case with $N = 2$ agents and with simply two possible actions: cooperation and defection. We can establish a payoff matrix of the game as follows:

	Cooperate	Defect
Cooperate	R , R	S , T
Defect	T , S	P , P

Table I: Payoffs of agent A (left) and B (up) in Social Dilemma

Such a game is called social dilemma if the following inequalities are verified:

- $R > P$ (1)
- $R > S$ (2)
- at least one of these two inequalities:
  - $T > R$: greed (3a)
  - $P > S$: fear (3b)

(1) means that mutual cooperation is better than mutual defection and (2) that mutual cooperation is better than being exploited. If such a game is iterated, it is called Sequential Social Dilemma (SSD), then it is relevant to add a condition:

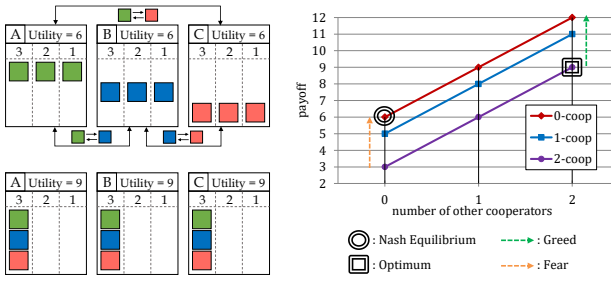
- $R > \frac{1}{2}(S + T)$: mutual cooperation is better than equiprobable different choice (4)

Social dilemma admit non-optimal Nash equilibria in particular (*Defect, Defect*) in Prisoner's Dilemma (where greed (3a) and fear (3b) are verified).

C. Continuous Multi-Agents Social Dilemmas

Let us introduce the case with $N \geq 2$ agents. Each agent is free to cooperate or not with all other agents. We formalise this type of action by a cooperation rate $c_{i,j}$ which defines the degree of cooperation of agent A_i with respect to agent A_j (0 for total defection and 1 for total cooperation).

Similarly to [16], we propose a definition for a multi-agent continuous social dilemma. We assume that cooperation is continuous and asymmetric. Let us suppose that each agent A_i earns a payoff defined by an gain function $G^{(i)}(c_{i,0}, ..., c_{i,N-1}, c_{0,i}, ..., c_{N-1,i})$ which depends of all cooperation degrees from him toward other agents ($c_{i,k} \forall k \neq i$)



(a) Optimal transactions (b) Schelling Diagram

Figure 1: A simple discrete dilemma where 3 agents have 3 different kinds of items whose marginal utility is $4-k$. Figure 1a shows the optimal sharing where each personal utility goes from 6 to 9 and Figure 1b shows the corresponding Schelling diagram (for a better understanding, see [13] or [19]).

and from other agents towards him ($c_{k,i} \forall k \neq i$).
Let us extend inequalities of III-B to define social dilemmas in a multi-agent continuous context. The situation is said to be a dilemma if:

- $\forall j$, $G^{(i)}$ is decreasing w.r.t $c_{i,j}$
- It exists $c_{i,j} \neq 0$ such as $G^{(i)}(c_{i,j}, c_{j,i}) > G^{(i)}(0,0)$

The first inequality states that independently, each agent is not interested in cooperating. This can be considered either as fear (for sufficiently small value of $c_{j,i}$) or greed (for sufficiently large value of $c_{j,i}$). The second inequality corresponds to the fact that the Nash Equilibrium ($c_{j,i} = 0 \ \forall i,j$ due to the first inequality) is not optimal hence the presence of dilemma.

For clarity, let us briefly detail the example with $n = 2$ agents. A_1 and A_2 earn respectively $G^{(1)}(c_{1,2}, c_{2,1})$ and $G^{(2)}(c_{2,1}, c_{1,2})$. We can define particular values such as, the problem becomes the classic social dilemma seen in III-B:

$$G^{(1)}(0,0) = G^{(2)}(0,0) = P \tag{6a}$$
$$G^{(1)}(1,1) = G^{(2)}(1,1) = R \tag{6b}$$
$$G^{(1)}(0,1) = G^{(2)}(0,1) = S \tag{6c}$$
$$G^{(1)}(1,0) = G^{(2)}(1,0) = T \tag{6d}$$

Note that for $n > 2$, it is difficult to show payoffs with a table. However, it is possible to represent the payoffs thanks to Schelling diagram representing payoffs in function of choice of cooperation and number of other cooperators. An example is shown in Fig 1b.

D. Dilemmas due to concavity of utility functions

In our problem formulation, we can assume that $f^{(i)}(s_{i,k}(t) + \sum_{j=0}^{N-1}(X_k)_{i,j}) \sim G^{(i)}(c_{i,j}, c_{j,i})$ where $c_{i,j}$ can be estimated as an increasing function with respect to $-(X_k)_{i,j}$. Moreover, we can show that for well-balanced initial resources, due the concavity of utility functions, it exists

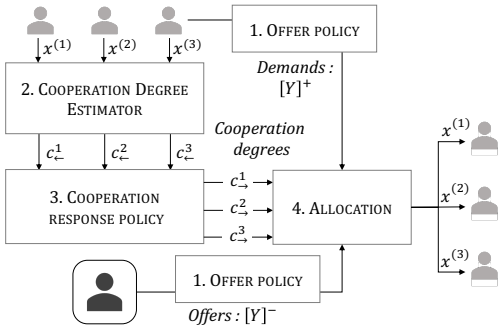


Figure 2: Architecture of our agent model

optimal sum of utility functions which doesn't correspond to null transactions. The two latter statements correspond to the two inequalities in III-C which define continuous multi-agent dilemmas.

IV. Model

The goal is to model an agent able to choose transactions to reach the optimum situation, while taking into account of the selfishness of others agents.

A. Architecture

As mentioned in the introduction, our agent model consists in four-step iterations:

1) Offer Policy: compute optimal demands and offers. Only the demands are communicated to other agents (section IV-B)
2) Cooperation Degree Estimator: according to the previous allocations from others agents, each agent estimates a cooperation degree from every agent (section IV-C)
3) Cooperation Response Policy: According to the estimated cooperation degrees, the policy computes response cooperation degrees (section IV-D).
4) Allocation: According to the cooperation degrees, demands from other agents and personal offers, an allocation of resources for each agent and item is computed.

B. Offers/Demands

At each iteration of the algorithm, agents have to communicate their needs to the other agents and know their offers (what they are willing to give). Though it is an important part, we don't focus on that in this paper. Nevertheless, one legitimate approach for an agent could be to demand a cumulative quantity equal to what he could give (with the idea to cooperation). Therefore, we simply propose that each agent optimise an expected "virtual" gain of utility:

$$\max_{Y^{(i)}} \sum_{k=1}^{M} f_k^{(i)}(s_{i,k} + Y_k^{(i)}) \tag{7}$$

$Y^{(i)}$ is the vector of demands/offers for agent A_i, where a positive (resp. negative) component corresponds to a demand (resp. offer).

C. Cooperation degree detection

When collaborating, a major problem in sharing resources is to detect whether the opponents are cooperative.
There is no distinction between capacity and willingness to cooperate since collaboration between agents means exchanging equivalent quantities of resources. Then, if an agent can't cooperate with another agent anymore, there is no point to continue to cooperate with him anymore.

At each round, each agent A_i computes a cooperation degree $c_{i,j}(t)$ toward every other agent A_j according to his interest, just by comparing what A_j gave with what he (A_i) was willing to give at previous round $t-1$. We introduce the notation of maximal offer $\Gamma_k^{(i)} = -[Y_k^{(i)}]^-$ that A_i can propose for item B_k (computed by A_i in step 2 IV-B). We can then define $c_{i,j}(t)$ with :

$$c_{i,j}(t) = \frac{(N-1)\sum_{k=0}^{M-1}(X_k)_{j,i}}{\sum_{k=0}^{M-1}\Gamma_k^{(i)}} \tag{8}$$

D. Cooperation response policy

To reach a safe negotiation among agents with personal interests, we introduce a cooperation policy. It maps incoming cooperation degrees to outgoing degrees.
In [20], a robust algorithm called tit-for-tat (TFT) is introduced to solve the iterated prisoner dilemma. The principle is simple: begin by cooperating and then copy the previous opponent action. In continuous version of iterated Prisoner Dilemma, [21] propose several adaptations of continuous TFT.
We propose a new version of TFT with an inertia component and an adaptive incentive rate. At step t, agent A (facing agent B of degree b_{t-1}) chooses a cooperation rate a_t as follows:

$$a_0 = 0 \tag{9a}$$
$$a_t = \beta a_{t-1} + (1-\beta)(r_{t-1} + (1-r_{t-1})b_{t-1}) \tag{9b}$$
$$r_0 = r \tag{9c}$$
$$r_t = [r_{t-1} + \alpha\Delta_{t-1}]^+ \tag{9d}$$
$$\text{with } \Delta_{t-1} = b_{t-1} - a_{t-1}$$

where β is an inertia coefficient and r_t is a coefficient to incentivise cooperation which is considered constant in [21]. However, to reinforce the robustness to defectors, we improve it as an adaptive incentive coefficient with a rate α. The benefit of this last modification is to adapt the incentive to cooperate according to the last response of opponent: decrease it if the opponent doesn't answer positively and keep it constant (even increase it) if the answer is positive. Then, the pure defectors won't exploit much long this incentive behaviour.
To conclude, we can notate the generic algorithm as TFT(β,r,α). Let us note that a pure defector and pure cooperator correspond respectively to TFT($\beta = 1$, r, α) and TFT($\beta = 0$, $r = 1$, $\alpha = 0$).

E. Allocation

The main idea of allocation is that agent A_i computes for each item k the maximal part of spectrum he can offer $\Gamma_k^{(i)} = -[Y_k^{(i)}]^-$ and allocates to each agent A_j a part of this offer proportional to the clipped demand of agent A_j: $\tilde{D}_k^{(j)} = \min(Y_k^{(j)}, \Gamma_k^{(i)})$ and $c_{i,j}$ the cooperation degree between A_i and A_j:

$$(X_k)_{(i,j)} = \frac{c_{i,j}\tilde{D}_k^{(j)}}{\sum_{j=0,j\neq i}^{N-1}\tilde{D}_k^{(j)}}\Gamma_k^{(i)} \tag{10}$$

With this allocation, we ensure that $(X_k)_{(i,j)} \leq \Gamma_k^{(i)}$ and that it is increasing w.r.t $c_{i,j}$ and $\tilde{D}_k^{(j)}$.

V. Experiments

Different algorithms for sharing resources are evaluated[1]. We introduce several metrics to study the safety, the incentive-compatibility and the speed of convergence. We test the parameters of our algorithm on a fixed test presented in Fig 3. For the simulation, we use a simple and unique concave function for every utility: $\forall k,\ \forall i,\ f_k^{(i)} : x \mapsto \ln(x+2)$.

In Figures 4, 5 and 6, we show a simulation of our model. In Fig 4, all agents use TFT and we observe that social welfare and cooperation converge. After convergence, the cooperation detection, linked to previous transactions which are then very low causes an unstable cooperation response. In Fig 5, we introduce an selfish agent (defector) and we observe that other agents TFT($\beta = 0.1$,$r = 0.1$,$\alpha = 0$) are exploited since the defector earns more. At last, in Fig 6, we evaluate the relevance of the adaptive incentive rate α on the previous situation and we observe that this rate is relevant since the agents TFT are not exploited anymore.

In what follows, we evaluate the risk of being exploited and the incentive one agent has to cooperate. in this purpose, we define some metrics (in V-A), and we use them to evaluate the parameters (β, r, α) of our TFT model.

A. Metrics

1) Efficiency: First, let us define the global social welfare which corresponds to the sum of all agents utilities:

$$SW(t) = \sum_{i=0}^{N-1} f^{(i)}(s(t)) = \sum_{i=0}^{N-1}\sum_{k=0}^{M-1} f_k^{(i)}(s_{i,k}(t)) \tag{11}$$

Secondly, we compute optimal social welfare SW_{opt} with a PSO algorithm [22]. Finally, we define efficiency E as the ratio:

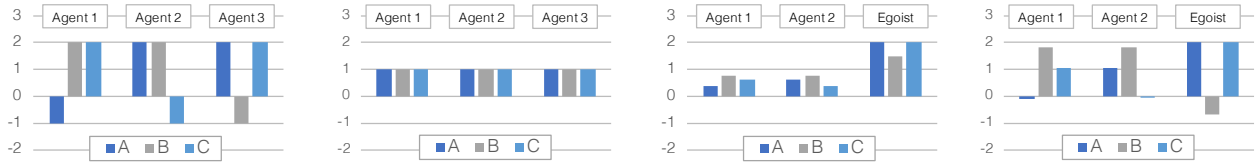
$$E(t) = \frac{SW(t) - SW_0}{SW_{opt} - SW_0} \tag{12}$$

2) Speed: We define a metric measuring the speed of convergence to maximal efficiency:

$$Sp = \frac{1}{TE_{max}}\int_0^T E(t)dt \tag{13}$$

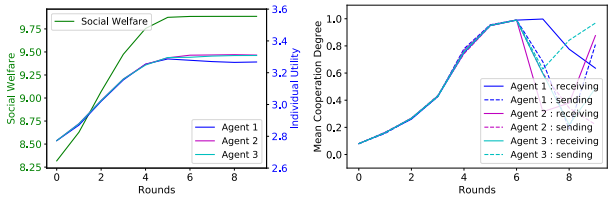
Then, we have $Sp \in [0,1]$ ($Sp = 1$ corresponding to the case where all optimal transactions are made at the first step).

[1]The source code of our framework is available on GitHub: https://github.com/tlgleo/sharing_resources



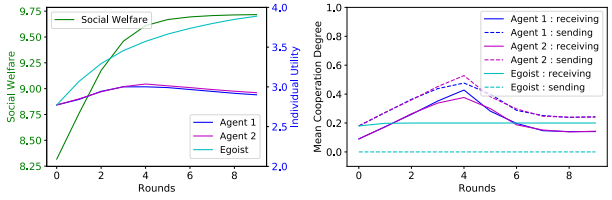
(a) Initial state (b) Optimal transactions (c) Exploited by one egoist (d) With our model

Figure 3: Simple example of sharing resources between three agents with three items of concave utilities. Fig 3a shows the initial state and figures 3b, 3c and 3d show three possible outcomes. Three simple TFT agents are able to reach the optimal situation (3b and Fig 4), but two TFT agents can be exploited by a pure defector (3c and Fig 5) due to the constant incentive coefficient. At last, with our adaptive incentive rate, our TFT agents are not so exploited anymore (3d and Fig 6).



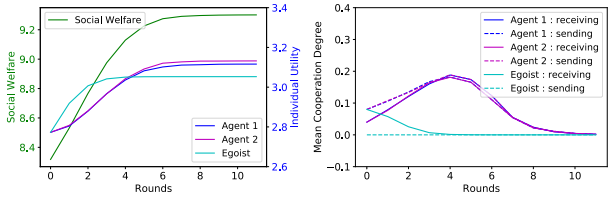
(a) Evolution of social welfare and individual utilities (b) Evolution of incoming and outgoing cooperation degrees

Figure 4: Simulation with the example in Fig 3 with three agents TFT($\beta = 0.5$, $r = 0.02$, $\alpha = 0$)



(a) Evolution of social welfare and individual utilities (b) Evolution of incoming and outgoing cooperation degrees

Figure 5: Simulation with example in Figure 3 with two TFT($\beta = 0.5$, $r = 0.2$, $\alpha = 0$) and one pure defector



(a) Evolution of social welfare and individual utilities (b) Evolution of incoming and outgoing cooperation degrees

Figure 6: Same simulation as in Fig 5 but with incentive adaptive rate $\alpha = 0.3$ to prevent exploitation from selfish agent: TFT($\alpha = 0.5$, $r = 0.2$, $\alpha = 0.3$)

3) Incentive-Compatibility: We adopt a metric proposed in [14] to measure incentive-compatibility, which measures how agents are encouraged to cooperate. Then, we define $IC(\pi_i)$ as the difference:

$$IC(\pi_i) = G^{(i)}(\pi_i, \pi_j) - G^{(i)}(D, \pi_j) \tag{14}$$

where D is the pure defection policy. In other words, $IC(\pi_i)$ computes the preference of an agent between defecting and following the policy π_i when he faces agents following the same policy π_j.

4) Safety: In the same vein, we measure the risk an agent takes by cooperating. We define the safety $Sf(\pi_i)$ as the difference:

$$Sf(\pi_i) = G^{(i)}(\pi_i, D) - G^{(i)}(D, D) \tag{15}$$

Simply put, $Sf(\pi_i)$ computes the lost difference if an agent is exploited.

B. Evaluation

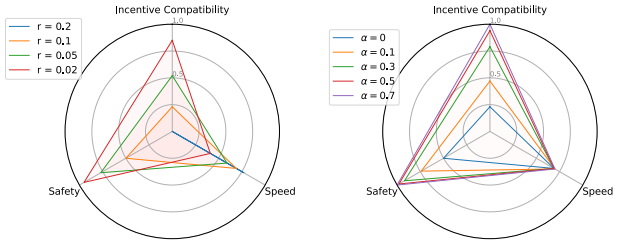
We evaluate several parameters of our proposed algorithm TFT(β, r, α) thanks to the metrics described in V-A.

β	r	α	Speed	IC	Safety
0.1	0.2	0	0.77	-0.73	-1.21
0.1	0.1	0	0.69	-0.5	-0.64
0.1	0.05	0	0.597	-0.21	-0.33
0.1	0.02	0	0.41	0.12	-0.13
0.3	0.1	0	0.66	-0.49	-0.63
0.5	0.1	0	0.6	-0.47	-0.6
0.1	0.1	0.1	0.7	-0.26	-0.36
0.1	0.1	0.3	0.69	0.06	-0.16
0.1	0.1	0.5	0.69	0.21	-0.09
0.1	0.1	0.7	0.7	0.28	-0.07

Table II: Results of experience showing the influence of parameters β, r, α

C. Discussion

The principle of the classic discrete TFT is to copy the previous action of an opponent after beginning with pure cooperation. However, when cooperation is continuous, discrete actions (cooperate/defect) are replaced by a cooperation degree. Then, the algorithm must numerically compute a cooperation degree in response. The stakes are to incentivise cooperation which corresponds to coefficient r and to be robust



(a) Influence of r (b) Influence of α

Figure 7: Influence of parameters r and α evaluated by the speed of convergence, safety and incentive-compatibility. The two latter metrics are normalised.

to defection i.e. the algorithm does not need to be too fast. In figure 7a, we notice that increasing r allows to gain speed of convergence but safety and incentive-compatibility fall, meaning the risk of being exploited is very high. In our test (Table II), the inertia coefficient β is not enhanced. It is probably due to stability of agents behaviours. But we are convinced that it would be suited to dynamic behaviours (which could be studied in further works). In contrast, our modification over r incentive coefficient is rather relevant (Figure 7b) since increasing α allows to reach very satisfying safety and incentive-compatibility without altering convergence speed.

VI. Conclusion and perspectives

We presented a formalisation of a non-cooperative game to exchange resources of decreasing marginal utilities. We view the problem as a social dilemma: nobody is incentivised to cooperate (exchanging resources) since utilities functions are monotonically increasing but doing nothing is non optimal due to the concavity of utilities functions of agents. We solved this problem with tit-for-tat algorithms mixed with designed algorithms for allocation of resources. We first formulated the problem of exchanging resources with concave utility functions in sequential social dilemmas. We then proposed extensions to definitions for social dilemmas in a Markov Game with more than two agents and with continuous cooperation. We proposed an adaptation of a continuous version of tit-for-tat with adaptive incentive coefficient. We designed a simple algorithm for allocating resources to other agents without pricing, only with negotiation through tit-for-tat algorithms. We finally adopted some metrics to study key properties such as efficiency, speed, safety and incentive-compatibility. The results show that our algorithm is efficient: the transactions reach the optimal consensus and the TFT agent is robust to defection and incentivised to cooperate. Our model solves iterated tit-for-tat between multiple agents in homogeneous well-balanced initial states. Future work should strengthen it with a finer control of the offers and demands and a better cooperation response policy, possibly with techniques like self-play reinforcement learning.

References

[1] Karsten Buckwitz, Jan Engelberg, and Gernot Rausch. Licensed shared access (LSA)—Regulatory background and view of administrations. In *9th International Conference on Cognitive Radio Oriented Wireless Networks and Communications*, pages 413–416. IEEE, 2014.

[2] Ayman Chouayakh, Aurelien Bechler, Isabel Amigo, et al. Auction mechanisms for Licensed Shared Access: reserve prices and revenue-fairness trade offs. *ACM SIGMETRICS Performance Evaluation Review*, 46(3):43–48, 2019.

[3] Xia Zhou and Haitao Zheng. TRUST: A general framework for truthful double spectrum auctions. In *IEEE INFOCOM*, pages 999–1007, 2009.

[4] Yanjiao Chen, Jin Zhang, Kaishun Wu, and Qian Zhang. Tames: A truthful auction mechanism for heterogeneous spectrum allocation. In *Proceedings IEEE INFOCOM*, pages 180–184, 2013.

[5] Huiyang Wang, Eryk Dutkiewicz, Gengfa Fang, et al. Spectrum sharing based on truthful auction in licensed shared access systems. In *IEEE 82nd Vehicular Technology Conference*, pages 1–5, 2015.

[6] Wei Ren, Randal W Beard, and Ella M Atkins. A survey of consensus problems in multi-agent coordination. In *Proceedings of the American Control Conference*, pages 1859–1864. IEEE, 2005.

[7] Ren, Wei and Beard, Randal W and Atkins, Ella M. Information consensus in multivehicle cooperative control. *IEEE Control systems magazine*, 27(2):71–82, 2007.

[8] Ruiliang Zhang and James Kwok. Asynchronous distributed ADMM for consensus optimization. In *International conference on machine learning*, pages 1701–1709, 2014.

[9] Reza Olfati-Saber and Richard M Murray. Consensus problems in networks of agents with switching topology and time-delays. *IEEE Transactions on automatic control*, 49(9):1520–1533, 2004.

[10] Anup Menon and John S Baras. Collaborative extremum seeking for welfare optimization. In *53rd IEEE Conference on Decision and Control*, pages 346–351, 2014.

[11] Yi Dong and Jie Huang. A leader-following rendezvous problem of double integrator multi-agent systems. *Automatica*, 49(5):1386–1391, 2013.

[12] Gary Cheng, Kabir Chandrasekher, and Jean Walrand. Static & Dynamic Appointment Scheduling with Stochastic Gradient Descent. In *American Control Conference (ACC)*, pages 2092–2099. IEEE, 2019.

[13] Joel Z Leibo, Vinicius Zambaldi, Marc Lanctot, Janusz Marecki, et al. Multi-agent reinforcement learning in sequential social dilemmas. In *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*, pages 464–473. International Foundation for Autonomous Agents and Multiagent Systems, 2017.

[14] Adam Lerer and Alexander Peysakhovich. Maintaining cooperation in complex social dilemmas using deep reinforcement learning. 2017.

[15] Natasha Jaques, Angeliki Lazaridou, Edward Hughes, et al. Intrinsic social motivation via causal influence in multi-agent RL. 2018.

[16] Edward Hughes, Joel Z Leibo, Matthew Phillips, et al. Inequity aversion improves cooperation in intertemporal social dilemmas. In *Advances in Neural Information Processing Systems*, pages 3326–3336, 2018.

[17] Shayegan Omidshafiei, Dong-Ki Kim, Miao Liu, Gerald Tesauro, et al. Learning to teach in cooperative multiagent reinforcement learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 6128–6136, 2019.

[18] John Nash. Non-cooperative games. *Annals of mathematics*, pages 286–295, 1951.

[19] Thomas C Schelling. *Micromotives and macrobehavior*. WW Norton & Company, 2006.

[20] Robert Axelrod and William Donald Hamilton. The evolution of cooperation. *science*, 211(4489):1390–1396, 1981.

[21] Tom Verhoeff. The trader's dilemma: A continuous version of the prisoner's dilemma. *Computing Science Notes*, 93(02), 1998.

[22] Russell Eberhart and James Kennedy. Particle swarm optimization. In *Proceedings of the IEEE international conference on neural networks*, volume 4, pages 1942–1948. Citeseer, 1995.