Decision-making of Autonomous Agents in Networks with Adversaries

Konstantinos Ntemos *

National and Kapodistrian University of Athens
Department of Informatics and Telecommunications
kdemos@di.uoa.gr

Abstract. In modern networks, agents are characterized by enhanced capabilities of independent decision making. This results in improved network efficiency in terms of throughput maximization, energy consumption and scalability. However, new challenges emerge. Agents may have disparate, or even conflicting interests. As a result, in networks of autonomous "intelligent" agents altruistic cooperation can not be taken for granted. In order to incentivize agents to comply with a behavior that is beneficial for the network, *trust/reputation models* have been proposed in literature.
This thesis aims at investigating the information sharing process of autonomous agents with individual interests able to exhibit *selfish* and *malicious* behavior, as well as the impact of trust/reputation models on the agents' decision-making process. Initially, we study the information sharing process in the most basic network functionality, namely *packet-forwarding*. Then, we move to a higher level and we consider a specific model of information diffusion. More specifically, we study the problem of parameter estimation with selfish and malicious agents. Finally, we make an abstraction and study information sharing in a more general setting. We utilize the notion of *conditional mutual information* to evaluate the shared information and we model agents' interactions as a dynamic game of asymmetric information.
To achieve the research goals of this thesis, methods from Stochastic Control Theory, Game Theory, Signal Processing and Learning algorithms are utilized. The research conducted focuses on the application areas of packet-forwarding and information sharing, which constitute the two basic functionalities of wireless networking.

1 Introduction

Autonomous decision-making can enhance network efficiency in terms of throughput maximization, energy savings and scalability, as there is no need for signaling overhead due to the lack of a centralized authority that coordinates nodes' actions. However, autonomous nodes (or agents) may not belong to the same authority and as a result, they may have disparate, or even conflicting interests.

* Dissertation Advisor: Nicholas Kalouptsidis, Professor

This can lead them to exhibit 'misbehavior', resulting in low network performance and security breaches. The basic 'misbehavior' patterns are *selfish* and *malicious* behavior [1].

A *selfish* agent (also called *free-rider* [2]) aims at exploiting as much as possible network benefits at the minimum possible cost, while malicious agents aim at harming other network agents and/or network infrastructures. Malicious agents could be interested in enjoying cooperation benefits as well.

It is shown in [3] that *cooperation* in a network of autonomous self-interested agents is unlikely to occur without providing *incentives*. For this reason, *trust /reputation* models have been proposed in order to incentivize agents to cooperate, as well as to detect and isolate misbehaving agents. Under such schemes, a node's actions history is used to shape a belief about node's trustworthiness and future expected behavior [5].

This thesis investigates the information sharing process among autonomous agents and consists of three main parts. The first one consists of our study on the interactions of autonomous agents in the most basic form of information sharing, namely in packet-forwarding. We consider agents able to exhibit both selfish and malicious behavior and assume the deployment of a Trust Management System that monitors and evaluates agents' behavior in a setting where agents' actions are *partially observable* [7], [8]. Although the assumption of partial observability of agents' actions is in line with a more realistic modeling of a wireless network environment, it has received significantly less attention in literature. The impact of action monitoring and trust on agents' decision-making process is analyzed. We derive novel conditions, which if satisfied, lead the agents to adhere to the behavior perscribed by the TMS, for their own interest and the need for different treatment between selfish and malicious behavior is highlighted. Such theoretical results can be utilized to design effective trust management systems and enhance network performance and security. These results are discussed in Section 2.

Then, we move on and assume a specific model of information sharing in the second part of the thesis. This part is devoted to our study on *information sharing* over adaptive networks for a distributed parameter estimation task in an *adversarial* setting, where malicious agents exist in the network and have the option to disseminate intentionally *falsified information*. Recently, parameter estimation with selfish agents was studied [4]. We extended this work by considering malicious agents in the network, as well. The malicious agents have the option to disseminate intentionally falsified information to deteriorate the estimation performance of other agents. They try to achieve this goal by adding noise to the parameter estimates they share with the other network agents. The consideration of malicious behavior gives rise to new challenges. To overcome these challenges, we devised appropriate *detection schemes* to identify malicious information and extended the LMS algorithm to enable the agents efficiently estimate the unknown parameters, despite the presence of malicious entities in the network. In addition to the detection schemes, we developed appropriate trust/reputation models to help agents in the estimation task and guide their decision-making process in the aforementioned uncertain and non-stationary set-

ting. The experimental results show that the good performance of the detection schemes, along with the deployment of the trust/reputation models can give rise to high cooperation rates in the network as well, apart from the efficient estimation performance. These results are presented in Section 3.

Finally, we studied the dynamic information-sharing in a more general setting without assuming a specific information diffusion model. We utilized the notion of *conditional mutual information* to quantify the information shared among agents. The agents' interactions are modeled as a *dynamic game of asymmetric information*. Self-interested agents observe a hidden Markov Chain (MC) and are called to decide whether they will exchange their observations or not. In the formulated *Dynamic Information Sharing Game* agents' actions affect their peers' *information structure*, which in turn affects their future decisions. The challenges that arise due to the inter-dependence of agents' information structure and decision-making are addressed. For the finite horizon game we prove that agents do not have incentive to share information. In contrast, we show that cooperation can be sustained in the infinite horizon case by devising appropriate *punishment strategies* which are defined over the agents' *beliefs* on the system state. We show that these strategies are closed under the best-response mapping and that cooperation can be the optimal choice in some subsets of the state belief simplex. We characterize these *equilibrium regions*, prove uniqueness of a *maximal* equilibrium region and devise an algorithm for its approximate computation. This study constitutes the third part of this thesis and the results are discussed in Section 4.

2 Trust-based Strategies for Wireless Networks

In this Section, we present our results on information sharing in the most basic form, namely in packet-forwarding. More specifically, we study *trust-based strategies* for autonomous agents in wireless networks in a *partial monitoring* setting [7, 8]. The term *partial monitoring* means that agents' actions are not always perfectly observable. We study the impact of action monitoring and trust on agents' optimal decision-making. Both selfish and malicious behavior is considered. We derive conditions that if satisfied, agents opt to the behavior perscribed by the Trust Management System (TMS).

Here, we briefly present our results on the pairwise interactions of autonomous agents who are called to forward the packets of their peers [8]. The interactions are modeled as a stochastic game with partial action monitoring. We derive conditions that thwart misbehavior and lead the agents to cooperation at equilibrium. Both selfish and malicious behavior is considered.

Regarding, our work on packet-forwarding for Cognitive Radio Networks (CRNs) the interested reader can refer to [7]. In [7] we assume that the Secondary Users (SUs) can exhibit both selfish and malicious behavior and their decision-making process is modeled as a Markov Decision Process (MDP). A trust model is deployed to enhance cooperative behavior and we prove conditions under which, cooperative behavior is optimal.

2.1 System Model

We study the interactions of two agents i, j which want to forward their packets to their intended destinations d_i, d_j, respectively. Node d_i (resp. d_j) is outside the transmission range of i (resp. j). Thus, there is need for i (resp. j) to forward the packets of j (resp. i) to d_j (resp. d_i). Both agents may exhibit selfish or malicious behavior. The set of admissible transmission decisions (i.e. actions) for an agent is

$$A = \{-1, 0, 1\}. \tag{1}$$

The values $-1, 0, 1$ correspond to malicious, selfish and honest action, meaning the agent chooses to launch an attack (i.e. modify the packets and then forward them), not to forward the packets and forward them to the destination, respectively.

Each agent wants its packets to reach the desired destination. This interest is captured by a forwarding benefit $f > 0$ if the other agent forwards its packets (i.e. $a = 1$). Forwarding the other agent's packets incurs a transmission cost $c > 0$. Finally, if an agent launches an undetected attack (i.e. the IDS does not perform sampling), then an illegal gain $e > 0$ would be acquired expressing the gain of a successful attack. In this case, the other agent would suffer a loss $\ell > 0$ from the attack. Agent i does not know whether the IDS will be active in the current time slot, so it uses the sampling probability P_a to form the *instantaneous expected reward*

$$R_i(a_i(t), a_j(t)) = R_i^x(a_i(t)) + R_i^r(a_j(t)), \tag{2}$$

where

$$R_i^x(a_i(t)) = \begin{cases} -c_i, & \text{if } a_i(t) = 1, \\ 0, & \text{if } a_i(t) = 0, \\ (1 - P_a)e_i - c_i & \text{if } a_i(t) = -1 \end{cases} \tag{3}$$

$$R_i^r(a_j(t)) = \begin{cases} f_i, & \text{if } a_j(t) = 1, \\ 0, & \text{if } a_j = 0 \\ -(1 - P_a)\ell_i & \text{if } a_j(t) = -1 \end{cases}. \tag{4}$$

2.2 Results

Theorem 1. *Suppose $P_a, p, \phi > 0$ and $q >= \max\{p, \phi\}$. Suppose further that $p > \phi$. Then the honest strategy profile $(\pi_{\text{H}}, \pi_{\text{H}})$ is PPE if and only if the following hold for both agents*

$$f \geq \max\{k_1(1 - P_a)e - k_2 c, k_3(1 - P_a)e, k_3 c\}, \tag{5}$$

where

$$k_1 = \frac{(1 - \delta + \delta P_a p)(1 - \delta + 2\delta P_a \phi - \delta P_a \phi^2)}{\delta P_a \phi (1 - \delta + \delta P_a p + \delta P_a \phi - \delta P_a p \phi)},$$

$$k_2 = \frac{(1 - \delta + \delta P_a \phi)(1 - \delta + \delta P_a p + \delta P_a \phi - \delta P_a \phi^2)}{\delta P_a \phi (1 - \delta + \delta P_a p + \delta P_a \phi - \delta P_a p \phi)},$$

$$k_3 = \frac{1 - \delta + \delta P_a p}{\delta P_a p}.$$

Furthermore, if $p \leq \phi$, $(\pi_{\text{H}}, \pi_{\text{H}})$ is PPE if and only if the following hold for both agents

$$f \geq \max\{k_3(1-P_a)e, k_3 c\}. \tag{6}$$

Proposition 1. *The set*

$$K(e) = \left\{(c,f) \in \mathbb{R}^2 : c, f \geq 0 \text{ and } (5) \text{ holds}\right\}, \text{ if } p > \phi$$
$$K(e) = \left\{(c,f) \in \mathbb{R}^2 : c, f \geq 0 \text{ and } (6) \text{ holds}\right\}, \text{ if } p \leq \phi$$

is decreasing in e, i.e. $K(e') \subset K(e)$ for $e' > e$.

3 Information Sharing in Diffusion Adaptive Networks with Malicious Agents

In this Section, we study information sharing at a higher level. We are not only interested on whether agents shared information or not (as was the case in packet-forwarding), but on what information they shared, as well. More specifically, we study information sharing in parameter estimation in an *adversarial* setting where malicious agents may disseminate falsified information.

In [6] an action detection scheme based on k-means algorithm is developed, along with a centralized trust management system to assist agents in their estimation process. In [9] we extend the aforementioned work by devising distributed detection and reputation systems leading to the development of a scalable algorithm for the agents' estimation and decision-making processes which is robust to malicious information dissemination. In this Section we briefly present only our work [9].

3.1 System Model

We consider a set of N agents with sensing, computing, and communication capabilities coexisting in a network. Let $\mathcal{G} = \langle \mathcal{N}, \mathcal{E} \rangle$ be the associated graph, where the agents are labeled by the elements of $\mathcal{N}$ and $\mathcal{E}$ indicates bi-directional links between two agents. Moreover, the nodes linked to $k \in \mathcal{N}$ (incl. node k) form its neighborhood, which is denoted by $\mathcal{N}_k$. Let $N = |\mathcal{N}|$ and $\mathcal{N}_k^- = \mathcal{N}_k \backslash \{k\}$. The terms node and agent will be used interchangeably.

The nodes can exchange information with their neighbors. Each node's type is denoted as $\tau \in T = \{\text{M}, \text{S}, \text{H}\}$ and it is unknown to the other nodes in the network. $\text{M}, \text{S}, \text{H}$ stand for the malicious, selfish and honest type, respectively. The agent's type determines the node's reward function and the set of admissible actions towards its neighbors.

All types of nodes want to estimate an unknown vector of parameters, denoted $w_0 \in \mathbb{C}^{L \times 1}$. To do so, they utilize information acquired by their own sensors and information received by their neighbors. At each discrete time instant i, each node k, has access to data $\{d_{k,i}, u_{k,i}\}$, corresponding to realizations of zero-mean random processes $\{\mathbf{d}_{k,i}, \mathbf{u}_{k,i}\}$. These data are related to the true vector of parameters w^o via

$$\mathbf{d}_{k,i} = \mathbf{u}_{k,i} w^o + \mathbf{v}_{k,i} \tag{7}$$

where $\mathbf{v}_{k,i} \in \mathbb{C}$ is a white–noise process with zero mean and power $\sigma_{v,k}^2$ and $\mathbf{u}_{k,i} \in \mathbb{C}^{1\times L}$ is the regression vector. The regressors $\{\mathbf{u}_{k,i}\}_{k=1}^{N}$ are temporally and spatially independent with covariance matrix $R_{u,k}$, i.e., $E\{\mathbf{u}_{k,i}^H \mathbf{u}_{\ell,i'}\} = R_{u,k}\delta_{k,\ell}\delta_{i,i'}$ where $\delta_{k,\ell}$ is the Kronecker symbol. $\mathbf{u}_{k,i}$ and $\mathbf{v}_{\ell,i'}$ are independent for all ℓ, k and i, i'.

Nodes compute parameter estimates adaptively in a distributed manner based on data exchange with neighboring nodes. Agents can mutually benefit in terms of estimation performance by exchanging their local estimates and cooperatively estimate w_0. In doing so, agents utilize the diffusion-based Least Mean Squares (LMS) algorithm [4]. Under the *adapt-then-combine* (ATC) implementation of the diffusion-based LMS an agent k utilizes the new information $d_{k,i}$ acquired by its own sensors in order to update its local intermediate estimate of w_0 at time i. To avoid harmful adversarial attacks (intentionally falsified data) by potential malicious agents, estimation tasks are augmented by a reputation model and an attack detection model. As a consequence, the classic adapt-then-combine (ATC) adaptation scheme is replaced by an *adapt-detect-combine* (ADC) strategy (explained in the sequel) while transmission actions are shaped by both short and long term rewards as well as the reputations of the recipients.

Admissible Actions and Data Exchange. Nodes exchange data following a distributed random pairing protocol [2], [4]. Nodes exchange estimation-related data based on their type. Honest nodes always apply a predetermined policy regarding the sharing of their local estimates based on recipient's reputation. On the other hand, selfish and malicious nodes are *strategic agents* that exchange data with neighboring nodes based on the expected rewards resulting from their actions. Payoffs for a selfish node k trade off communication cost and improvement in estimation performance.

A malicious node k can either share its local estimates $\boldsymbol{\psi}_{k,i}$ with its paired agent or send intentionally falsified information in order to degrade the other node's estimation performance. To model this malicious behavior, before sharing the information with its paired agent $\ell \in \mathcal{N}_k^-$, we assume that a malicious node k injects some perturbation $\eta_{k\ell} \in \mathbb{C}^{L\times 1}$ to its local estimate $\boldsymbol{\psi}_{k,i}$ as follows

$$\boldsymbol{\varphi}_{k\ell,i} = \boldsymbol{\psi}_{k,i} + \eta_{k\ell} \tag{8}$$

Let $\boldsymbol{a}_{k\ell,i} \in \mathcal{A}(\tau_k)$ be the transmission action selected by node k towards agent $\ell \in \mathcal{N}_k^-$. The set of admissible actions of a node k depends on its type τ_k and it is defined as

$$\mathcal{A}(\tau_k) = \begin{cases} \{-1, 1\}, & \text{if } \tau_k = \text{M} \\ \{0, 1\}, & \text{otherwise.} \end{cases} \tag{9}$$

where $\{1, 0, -1\}$ represent honest (sends its intermediate estimate to the paired agent), selfish (sends nothing) and malicious action (sends an erroneous estimate), respectively. Note that the malicious action can be applied only by nodes of adversarial type.

The various possible messages shared by agent k with its paired agent $\ell \in \mathcal{N}_k^-$ at time i mentioned above can be captured by

$$\boldsymbol{\varphi}_{k\ell,i} = \boldsymbol{a}_{k\ell,i}^2 \big[\boldsymbol{\psi}_{k,i} + \tfrac{1}{2}\big(1 - \boldsymbol{a}_{k\ell,i}\big) \cdot \eta_{k\ell}\big] \tag{10}$$

where we assume that $\boldsymbol{\varphi}_{k\ell,i}$ is an all-zero vector when agent k sends no information to agent ℓ (i.e. $\boldsymbol{a}_{k\ell,i} = 0$). For simplicity, we consider noise free communications during the exchange of information among the agents.

Adapt-Detect-Combine Strategy. Nodes can not tell with certainty if a neighbor sends them a reliable estimate or an intentionally falsified one. Therefore, the actions of agents are not fully observable. Because of that, each node k is equipped with a detection mechanism that takes a decision $\hat{\boldsymbol{a}}_{\ell k,i}$ about the action of node ℓ $\boldsymbol{a}_{\ell k,i}$ given the received vector $\boldsymbol{\varphi}_{\ell k,i}$. Under the assumption of noise free information exchange, the decision rule is straightforward for selfish action (i.e. $\hat{\boldsymbol{a}}_{\ell k,i} = 0$ if $\boldsymbol{a}_{\ell k,i} = 0$ – meaning that the selfish action is observable). In the remaining cases (i.e. $\boldsymbol{a}_{\ell k,i} = 1, -1$) the detection decision relies on private information of agent k. More specifically, to estimate the action taken by its paired node $\ell \in \mathcal{N}_k^-$ at time instant i, each node $k \in \mathcal{N}$ implements an additional *stand–alone* LMS, which is given by

$$\widehat{\boldsymbol{\psi}}_{k,i} = \widehat{\boldsymbol{\psi}}_{k,i-1} + \mu_k \, \mathbf{u}_{k,i}^H \left[\mathbf{d}_{k,i} - \mathbf{u}_{k,i} \widehat{\boldsymbol{\psi}}_{k,i-1} \right] \tag{11}$$

As the stand–alone LMS does not use the estimates of the other agents, no combination step is included. This means that $\widehat{\boldsymbol{\psi}}_{k,i}$ is not corrupted by possibly falsified information injected by malicious nodes. A node k can utilize this trustworthy estimate $\widehat{\boldsymbol{\psi}}_{k,i}$ to reliably determine if a paired agent ℓ has shared a truthful or an intentionally corrupted estimate. More precisely, to determine $\hat{\boldsymbol{a}}_{\ell k,i}$ in case $\boldsymbol{a}_{\ell k,i} \neq 0$, agent k applies the following cutoff rule

$$\hat{a}_{\ell k,i} = \begin{cases} 1, & \text{if } \|\boldsymbol{\varphi}_{\ell k,i} - \widehat{\boldsymbol{\psi}}_{k,i}\|^2 \leq \boldsymbol{s}_{k\ell,i} \\ -1, & \text{otherwise.} \end{cases} \tag{12}$$

where $\boldsymbol{s}_{k\ell,i}$ is a threshold set by k.

Adaptive Reputation Update Nodes maintain records with their beliefs about the trustworthiness or reputation of their neighbors. Beliefs are represented as probabilities that a given neighbor is honest, selfish or malicious based on the behavior the latter has exhibited in the past. Let $\boldsymbol{\theta}_{\ell k,i}^{[\mathrm{H}]}, \boldsymbol{\theta}_{\ell k,i}^{[\mathrm{S}]}, \boldsymbol{\theta}_{\ell k,i}^{[\mathrm{M}]}$ denote these belief measures of agent k about agent ℓ at time i and $\boldsymbol{\theta}_{\ell k,i} = [\boldsymbol{\theta}_{\ell k,i}^{[\mathrm{H}]}, \boldsymbol{\theta}_{\ell k,i}^{[\mathrm{S}]}, \boldsymbol{\theta}_{\ell k,i}^{[\mathrm{M}]}]^T$. $\boldsymbol{\theta}_{\ell k,i}^{[\mathrm{H}]}, \boldsymbol{\theta}_{\ell k,i}^{[\mathrm{S}]}, \boldsymbol{\theta}_{\ell k,i}^{[\mathrm{M}]}$ denote the belief that agent k has about the honesty, selfishness and malice of agent ℓ at instant i, respectively and they add up to one.

If at time i, k is not paired with $\ell \in \mathcal{N}_k^-$ the reputation vector remains the same, i.e. $\boldsymbol{\theta}_{\ell k,i+1} = \boldsymbol{\theta}_{\ell k,i}$. If k is paired with $\ell \in \mathcal{N}_k^-$ the reputation vector is updated in accordance with the detected action $\hat{\boldsymbol{a}}_{\ell k,i}$ as follows

$$\boldsymbol{\theta}_{\ell k,i+1} = \boldsymbol{\theta}_{\ell k,i} r_k^{[\hat{\boldsymbol{a}}_{\ell k,i}]} + e^{[\hat{\boldsymbol{a}}_{\ell k,i}]} (1 - r_k^{[\hat{\boldsymbol{a}}_{\ell k,i}]}) \tag{13}$$

where $e^{[1]} = [1,0,0]^T, e^{[0]} = [0,1,0]^T, e^{[-1]} = [0,0,1]^T$ and $r_k^{[\hat{\boldsymbol{a}}_{\ell k,i}]} \in (0,1)$ denoting a positive coefficient that reinforces the belief that is compatible with the detected action $\hat{\boldsymbol{a}}_{\ell k}$.

$$a_{k\ell,i} = \begin{cases} 1, & \text{with probability } \boldsymbol{\theta}^{[\mathrm{H}]}_{\ell k,i}, \\ 0, & \text{with probability } 1 - \boldsymbol{\theta}^{[\mathrm{H}]}_{\ell k,i} \end{cases} \tag{14}$$

3.2 Results

The honest agents are not *strategic*, meaning that they do not try to maximize a utility function, but they behave in a pre-defined and known way. An honest agent k shares its intermediate estimate with a paired neighbor ℓ with probability equal to $\boldsymbol{\theta}^{[\mathrm{H}]}_{\ell k,i}$ and sends nothing with probability $1 - \boldsymbol{\theta}^{[\mathrm{H}]}_{\ell k,i}$. Agents try to maximize their long-term expected payoffs. To combat the uncertainties arising in the maximization problem we utilize the *bounded rationality assumption*, which implies that agents have computational limitations. Under the bounded rationality assumption we derive a cut-off action selection rule for agents' decision-making process. We develop DS-LMS by utilizing ADC combination strategy and the action selection rule and present experimental results demonstrating that agents using DS-LMS achieve efficient estimation performance and decision-making. Moreover, cooperation stimulation can arise under certain conditions.

4 Dynamic Information Sharing and Punishment Strategies

In this Section we study the problem of information sharing among rational self-interested agents as a dynamic game of asymmetric information. We assume that the agents imperfectly observe a Markov chain and they are called to decide whether they will share their noisy observations or not. We utilize the notion of *conditional mutual information* to evaluate the information being shared among the agents. The challenges that arise due to the inter-dependence of agents' information structure and decision-making are addressed. For the finite horizon game we prove that agents do not have incentive to share information. In contrast, we show that cooperation can be sustained in the infinite horizon case by devising appropriate punishment strategies which are defined over the agents' *beliefs* on the system state. We show that these strategies are closed under the best-response mapping and that cooperation can be the optimal choice in some subsets of the state belief simplex. We characterize these *equilibrium regions*, prove uniqueness of a *maximal* equilibrium region and we devise an algorithm for its approximate computation. The interested reader can refer to [10].

4.1 System Model

State dynamics and observation models. The *system state* evolves in time as a Markov chain, i.e.

$$\mathbb{P}(X_{t+1}|X_{0:t}, Y^n_{0:t-1}, Y^{-n}_{0:t-1}, A^n_{0:t}, A^{-n}_{0:t}) = \mathbb{P}(X_{t+1}|X_t). \tag{15}$$

where t is the time index and X_t takes values in a finite set $\mathcal{X}$ at every time instant t. The initial state probability distribution vector is π_0. Each agent $n \in \mathcal{N} = \{1, 2\}$ receives observation $Y_t^n \in \mathcal{Y}^n$ according to observation probabilities that satisfy

$$\mathbb{P}(Y_t^n | X_{0:t}, Y_{0:t-1}^n, Y_{0:t-1}^{-n}, A_{0:t}^n, A_{0:t}^{-n}) = \mathbb{P}(Y_t^n | X_t). \tag{16}$$

From here on, we refer to the other agent as $-n$ (i.e., when $n = 1$, it is $-n = 2$ and vice versa).

Data exchange. The data exchange mechanism is materialized by the two received sequences Z_t^{-n} and Z_t^n by agent n and $-n$, respectively. More precisely, at each time t agent n (resp. $-n$) receives Z_t^{-n} (resp. Z_t^n) which is a deterministic function of the agent's observation Y_t^{-n} (resp. Y_t^n) and the action of agent $-n$ (resp. n). The action set for both agents is $\mathcal{A} = \{1, 0\}$. $A_t^n = 1$ means that agent n sends its observation Y_t^n to $-n$, while $A_t^n = 0$ means that n sends no data. Hence, data exchange is modeled as follows

$$Z_t^n(Y_t^n, A_t^n) = \begin{cases} Y_t^n, & \text{if} \quad A_t^n = 1 \\ \epsilon, & \text{if} \quad A_t^n = 0 \end{cases} \tag{17}$$

where the symbol ϵ denotes that no information is sent to the other agent. $Z_t^n \in \tilde{\mathcal{Y}}_t^n$, where $\tilde{\mathcal{Y}}_t^n = \mathcal{Y}_t^n \cup \{\epsilon\}$.

Information sets. Agent n at time t has access to information

$$I_t^n = (I_t^{n,p}, I_t^c). \tag{18}$$

I_t^n is comprised of agent's n *private history* $I_t^{n,p}$ and the *common history* I_t^c. The common history is known to both agents and consists of the agents' actions (i.e., $A_{1:t-1}^{1:2}$) and the history of the exchanged signals between the agents (i.e., $Z_{1:t-1}^{1:2}$), while the private history $I_t^{n,p}$ is known only to agent n and includes all the observations that agent n has decided not to share until the present time t. These histories at the beginning of time t are defined as follows

$$I_t^c = (Z_{0:t-1}^1, Z_{0:t-1}^2, A_{0:t-1}^1, A_{0:t-1}^2) \tag{19}$$

$$I_t^{n,p} = (Y_k^n | A_k^n = 0,\ 0 \le k < t) \tag{20}$$

Let $\mathcal{I}_t^n, \mathcal{I}_t^{n,p}, \mathcal{I}_t^c$ be the sets of all possible player's n *histories*, player's n *private histories* and *common histories* at time t, respectively. Initially, at time $t = 0$ the common information is $I_0^c = \tilde{\pi}_0$, where $\tilde{\pi}_0$ is the *common prior* belief on state X_0, and based on agents' actions, it evolves as

$$I_{t+1}^c = \begin{cases} (I_t^c, A_t^1, A_t^2), & \text{if} \ \ A_t^1 = A_t^2 = 0 \\ (I_t^c, A_t^1, A_t^2, Y_t^1), & \text{if} \ \ A_t^1 = 1, A_t^2 = 0 \\ (I_t^c, A_t^1, A_t^2, Y_t^2), & \text{if} \ \ A_t^1 = 0, A_t^2 = 1 \\ (I_t^c, A_t^1, A_t^2, Y_t^1, Y_t^2), & \text{if} \ \ A_t^1 = A_t^2 = 1. \end{cases} \tag{21}$$

The private information of agent n at time $t = 0$ is $I_0^{n,p} = \emptyset$ for all n and it is updated as

$$I_{t+1}^{n,p} = \begin{cases} I_t^{n,p}, & \text{if} \ \ A_t^n = 1 \\ (I_t^{n,p}, Y_t^n), & \text{if} \ \ A_t^n = 0 \end{cases} \tag{22}$$

4.2 Results

We prove that in the finite horizon game information sharing is ruled out, as the only strategy that is *sequentially rational* is never to cooperate for avery agent. This result makes us to focus on the infinite horizon game to search for strategies that can sustain cooperation and information sharing. In the following we define *Constrained Grim-Trigger* strategies.

Definition 1. *The* Constrained Grim Trigger *(CGT) strategy is defined as follows. Let $\mathcal{F}_{\mathcal{X}}$ denote the space of mappings $\sigma : \{0,1\} \times \Delta(\mathcal{X}) \to \Delta(\mathcal{A})$. Define the CGT map $\sigma^{n,\cdot} : \mathcal{P}(\Delta(\mathcal{X})) \to \mathcal{F}_{\mathcal{X}}$ for agent n, by*

$$\sigma^{n,\Pi^{n,c}}(s_t, \pi_t^n)(a_t^n = 1) = \begin{cases} 1, & \text{if } \ s_t = 1 \text{ and } \pi_t^n \in \Pi^{n,c} \\ 0, & \text{otherwise} \end{cases}$$

where π_t^n is the belief vector *over system states with elements $\pi_t^n(X_t = x) = \mathbb{P}^{\sigma^{n,\Pi^{n,c}},\sigma^{-n,\Pi^{-n,c}}}(X_t = x | i_t^n), x \in \mathcal{X}$ and $\mathcal{P}(\Delta(\mathcal{X}))$ denotes the powerset of the simplex $\Delta(\mathcal{X})$. Elements of the image of σ^n are called* CGT strategies *for player n.*

We briefly present the main results in the following.

Theorem 2. *Given agent $-n$ follows a CGT strategy $\sigma^{-n,\Pi^{-n,c}}$, agent's n best-response problem is a POMDP. Moreover, (S_t, Π_t^n) is an* information state*.*

Theorem 3. *The CGT strategies are closed under the best-response mapping.*

The importance of this result is that an agent can respond to a CGT strategy with a CGT strategy without *loss of optimality*. Then, we moved on to prove structural results of agents' cooperation regions $(\Pi^{1,c}, \Pi^{2,c})$.

Definition 2. *A pair of regions $(\Pi^{1,c}, \Pi^{2,c}) \in \mathcal{P}(\Delta(\mathcal{X})) \times \mathcal{P}(\Delta(\mathcal{X}))$ is in* cooperation equilibrium*, if the following conditions are satisfied*

$$\Pi^{n,c} = O^n(\Pi^{-n,c}), n \in \{1,2\}$$

where $O^n()$ denotes the operator that takes as input other agent's strategy $\sigma^{-n,\Pi^{-n,c}}$, solves agent's n POMDP (best-response problem) and returns the optimal cooperation region $\Pi^{n,c}$ for agent n. The most important results are the following.

Proposition 2. *If a pair of regions $(\Pi^{1,c}, \Pi^{2,c})$ is in* cooperation equilibrium*, then the two regions coincide, that is $\Pi^{1,c} = \Pi^{2,c} = \Pi^c$. Then, Π^c is called* equilibrium region*.*

Definition 3. *Let $\mathcal{E}$ be the set containing all equilibrium regions.*

Theorem 4. *There exists a unique* maximal *equilibrium region $\Pi^* \in \mathcal{E}$.*

We devised the following iterative algorithmic scheme to approximate Π^*.

Iterative Refinement Algorithm (ItRA)

Input: k (number of iterations), $\Pi^{n,c} = \Delta(\mathcal{X})$

– for k iterations do:
  • $\Pi^{n,c} \leftarrow F^n(\Pi^{n,c})$

---

where

$$F^n(C) = O^n(O^{-n}(C))$$

Proposition 3. *$\forall k > 0$, $\Pi^* \subseteq ItRA(k)$ and $ItRA(k+1) \subseteq ItRA(k)$.*

We also presented experimental results that empirically verify the existence of regions where information sharing under CGT strategies is optimal.

5 Conclusions and future work

In this thesis, information sharing among autonomous agents, that can exhibit selfish or malicious behavior, was investigated. We were interested in agents' decision-making process, as well as in cooperation stimulation methods and their impact on agents' decisions.

Initially, we considered the most basic application of information sharing, namely packet forwarding. In our study we assumed Markovian trust update mechanisms and studied their impact on agents' optimal decision-making process. Conditions that lead agents to a cooperative behavior were derived.

Then, we moved on to a higher level and considered a specific information sharing model. More specifically, we investigated information sharing in the problem of distributed parameter estimation in an adversarial setting, where malicious agents exist in the network.

Finally, we studied the information sharing process in a more general setting and utilized the notion of *conditional mutual information* to quantify the value of shared information. We modelled agents' interactions as a dynamic game of asymmetric information and studied jointly the information sharing and decision-making processes. We proved that in the finite horizon information sharing can never occur and devised appropriate *punishment strategies*, which we call CGT strategies, for the infinite horizon problem. We proved that cooperation is sustainable under the proposed strategies and derived structural results for the cooperation regions. Based on these results, we devised an algorithm for the approximate computation of the maximal cooperation region.

Finally, we modeled the information sharing process as a dynamic game of asymmetric information and studied jointly the information sharing and decision-making processes. We modeled the *value of shared information* using the notion of *conditional mutual information*. We showed that in the finite horizon information sharing can never occur and devised appropriate *punishment strategies*, which we call CGT strategies, for the infinite horizon problem. We proved that cooperation is sustainable under the proposed strategies and derived structural results for the cooperation regions. Based on these results, we devised an algorithm for the approximate computation of the maximal cooperation region.

As a closing remark, I would like to make a statement that escapes the technical results of this thesis. *Asymmetries* in information create complexity.

A big part of this thesis was devoted to the study of how to impose *truth-telling strategies* (in a sense *honesty*) among the agents. When it comes to human relationships, honesty could be imposed by the following realization. The virtue of honesty, which implies the establishment of *symmetric information* among the parties, is not valuable because of mere ethical and humanitarian reasons, but also because it facilitates simplicity.

References

1. P. Michiardi, R. Molva, "Simulation-based Analysis of Security Exposures in Mobile Ad Hoc Networks," in Proc. *European Wireless Conference*, Feb. 2002, pp 15–17.
2. J. Xu and M. Van der Schaar, "Social norm design for information exchange systems with limited observations," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 11, pp. 2126–2135, 2012.
3. M. Felegyhazi, J-P. Hubaux, and L. Buttyan, "Nash equilibria of packet forwarding strategies in wireless ad hoc networks," *IEEE Transactions on Mobile Computing*, vol. 5, no. 5, pp. 463-476, 2006.
4. C.-K. Yu, M. van der Schaar, and A. H. Sayed, "Information-sharing over adaptive networks with self-interested agents," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 1, no. 1, pp. 2–19, 2015.
5. G. J. Mailath, and L. Samuelson, "Repeated Games and Reputations: Long-Run Relationships," London, U.K.: Oxford Univ. Press, 2006.
6. K. Ntemos, N. Kalouptsidis, and N. Kolokotronis. "Managing trust in diffusion adaptive networks with malicious agents," in Proc. *European Signal Processing Conference, 2015. EUSIPCO 2015*, pp 91–95, 2015.
7. K. Ntemos, N. Kolokotronis, and N. Kalouptsidis, "Using trust to mitigate malicious and selfish behavior of autonomous agents in CRNs," in Proc. IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), IEEE, 2016.
8. K. Ntemos, N. Kalouptsidis, and N. Kolokotronis, "Trust-based strategies for wireless networks under partial monitoring." in Proc. 25th European Signal Processing Conference (EUSIPCO), IEEE, 2017.
9. K. Ntemos, J. Plata-Chaves, N. Kolokotronis, N. Kalouptsidis, and M. Moonen, "Secure information sharing in adversarial adaptive diffusion networks." IEEE Transactions on Signal and Information Processing over Networks vol. 4, no. 1, pp. 111-124, 2017.
10. K. Ntemos, G. Pikramenos, N. Kalouptsidis, and N. Kolokotronis, "Dynamic Information Sharing and Punishment Strategies", *under review* in IEEE Transactions on Automatic Control.