Secure Information Sharing in Adversarial Adaptive Diffusion Networks

Konstantinos Ntemos, Jorge Plata-Chaves, *Member, IEEE*, Nicholas Kolokotronis, Nicholas Kalouptsidis, and Marc Moonen, *Fellow, IEEE*

***Abstract*—In this paper, information sharing over adversarial adaptive networks is considered. Defense against adversarial attacks is enforced by an attack detection mechanism that guides the diffusion strategy in the parameter estimation task and the transmission decisions of agents. The latter are taken to maximize long-term rewards, expressed in terms of estimation performance, communication cost, and illegal gains in a nonzero sum repeated game with imperfect action monitoring. In order to track agents' behavior and stimulate cooperation, a reputation protocol is designed, along with the deployment of a set of honest nodes whose behavior is consistent with the reputation protocol. An algorithm that promotes cooperation and can achieve secure parameter estimation is described and illustrated by simulations.**

***Index Terms*—Reputation design, game theory, distributed in–network processing, diffusion algorithm, parameter estimation.**

I. Introduction

In RECENT years, numerous works have focused on modeling the nodes in wireless networks as autonomous "intelligent" agents, able to take decisions independently to maximize their benefits. Although such capabilities of independent and distributed decision–making can enhance network performance in terms of throughput maximization, energy savings and scalability, new security issues arise. The autonomous agents may not belong to the same authority and thus, may have disparate, or even conflicting interests. This allows "misbehavior" issues to appear, as the agents might opt to "misbehaving" in order to achieve their goals. Thus, in networks of autonomous agents, altruistic cooperation among them can not be taken for granted.

Two types of misbehavior can arise in wireless networks, namely selfish and malicious behavior [1], [2]. Selfish agents aim at enjoying the cooperation benefits while minimizing cooperation costs (also called as *free–riders* [5]), whereas malicious agents aim at harming other agents and/or compromise the network functionalities. Malicious agents could be interested in enjoying cooperation benefits as well.

It is shown in [9] that cooperation in a network of autonomous self-interested agents is unlikely to occur without providing *incentives*. For this reason, *trust* and *reputation* models [10]–[18] 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 [22]. The CONFIDANT protocol was proposed in [18] to deal with routing attacks in ad-hoc networks. In [17] a reputation scheme called SORI is employed to deal with selfish behavior and enforce cooperation.

Game Theory (GT) has been utilized along with the design of reputation models to analyze the autonomous agents' behavior in wireless networks. GT provides an appropriate framework to analyze interactions among self-interested and rational decision-makers [20], [21]. Most works on game-theoretic trust and reputation schemes model the agents interactions as variations of the well–known *Prisoner's Dilemma* (PD). In two–player PD, each agent has two choices, to *Cooperate* or not (*to Defect*). The unique *Nash Equilibrium* (NE) in PD is both agents selecting defection if the game is played once (one–shot game), while if both agents cooperate they can have bigger gains. Moreover, if the game is played repeatedly, but for a known number of rounds (*finite time horizon*), then the optimal behavior of the agents is to play the static NE (Defect,Defect) at all rounds. In the *infinite horizon* (discounted) case other NE can arise as well [20], [21]. An in-depth study of the *reputation effects* and the impact of reputation on the emergence of cooperation in non-cooperative games is provided in [22].

In [10], the so-called DARWIN algorithm deals with the imperfect monitoring of agents' actions. More specifically, the scenario whereby an agent's packet forwarding action may be perceived by another agent as packet drop due to not overhearing the transmission is considered. The *indirect reciprocity principle* along with *Evolutionary Game Theory* have been utilized in the design of reputation models for wireless networks as well [11]–[13], [16]. Under the indirect reciprocity principle, a node cooperates with another node because the latter has cooperated with other nodes in the past and not necessarily with the former node itself. An extensive literature survey on trust and reputation models can be found in [2] and [3].

Manuscript received August 9, 2017; accepted October 24, 2017. Date of publication December 28, 2017; date of current version February 19, 2018. This work was supported by the research project HANDiCAMS, which acknowledges the financial support of the Future and Emerging Technologies (FET) programme, within the 7th Framework Programme for Research of the European Commission, under FET-Open Grant 323944. The guest editor coordinating the review of this manuscript and approving it for publication was Dr. Konstantinos N. Plataniotis. *(Corresponding author: Konstantinos Ntemos.)*

K. Ntemos and N. Kalouptsidis are with the Department of Informatics and Telecommunications, University of Athens, Athens 157 72, Greece (e-mail: kdemos@di.uoa.gr; kalou@di.uoa.gr).

J. Plata-Chaves and M. Moonen are with the Department of Electrical Engineering (ESAT-STADIUS), KU Leuven, Leuven 3000, Belgium (e-mail: jorge.plata-chaves@esat.kuleuven.be; marc.moonen@esat.kuleuven.be).

N. Kolokotronis is with the Department of Informatics and Telecommunications, University of Peloponnese, Tripoli 221 00, Greece (e-mail: nkolok@uop.gr).

Digital Object Identifier 10.1109/TSIPN.2017.2787910

2373-776X © 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

The aforementioned research efforts consider the agents' interactions in packet-forwarding tasks. The design of reputation models to foster cooperation for distributed in-network information processing, has only recently received attention [7], [14], [15]. In [14], [15] a distributed reputation model has been proposed to give cooperation incentives to selfish agents in order to estimate an unknown vector of parameters. A repeated game is formulated where autonomous selfish agents are randomly paired. In [7] a centralized trust management system is employed to foster cooperation. The agents make decisions acting as *independent learners*.

As discussed in [7], [14], [15] an important difference compared to the packet-forwarding case stems from the fact that an agent's reward function is time-varying and depends on agent's estimate of the unknown parameter vector at every time instant giving rise to uncertainty regarding the future rewards and game-structures. In this paper we go a step beyond and consider both selfish and malicious agents coexisting in the network under a distributed reputation model.

The presence of malicious agents disseminating intentionally falsified data can deteriorate other agents' estimation performance participating in applications such as parameter estimation tasks [7] and power system state estimation [26]. An interesting discussion on the challenges that stem from malicious behavior in various kinds of networks is provided in [27].

Contributions: We formulate information-sharing for a parameter estimation task among selfish and malicious agents as a repeated game of *imperfect action monitoring* [20]. Unlike most packet-forwarding studies, where the overhearing technique (agents operate in promiscuous mode) provides knowledge on other agents' actions [19], the setup in this paper considers imperfect action information. For this reason, a distributed *detector* is introduced that enables each agent to determine the action of its paired agent. Then, inspired from [15], [16] we propose a reputation model suitable for capturing both selfish and malicious behavior. The reputation model is utilized by the agents in order to estimate the other agents' trustworthiness based on their past interactions and to predict future behavior. In order to stimulate cooperation, along with the deployment of the reputation model, we assume the existence of a set of honest agents in the network which follow a predefined and globally known reputation-based strategy. Finally, we study the decision making process of the agents and we derive an algorithm for secure distributed parameter estimation. We perform extensive simulation experiments and show that agents can successfully estimate the unknown parameter vector. Moreover, misbehaving types of agents can be incentivized to cooperate under suitable conditions.

Organization: The rest of the paper is organized as follows. In Section II we present the network model and system design. In Section III the game-theoretical analysis of the one–shot interactions among the agents and the Pareto inefficiency is discussed. Section IV carries out the analysis of the agents' long-term interactions and the diffusion-based best-response algorithm for secure estimation is developed. The agents' interactions and the implementation of the proposed algorithm are illustrated through simulations in Section V. Finally, concluding remarks are given in Section VI.

Notation: We use boldface letters for random variables and normal fonts for deterministic quantities. The notations $(\cdot)^H$ and $E\{\cdot\}$ stand for the Hermitian transposition and expectation, respectively. For a set, e.g., $\mathcal{X}$, the operator $|\cdot|$ denotes the cardinality. We use the weighted (semi)norm notation $\|x\|^2_{\Sigma} \triangleq x^H \Sigma x$ with a vector x and a Hermitian positive (semi-)definite matrix Σ. The notation $\mathbb{1}_{\{\ldots\}}$ denotes 1 if the condition in the subscript (inside the brackets) is true and 0 otherwise.

II. Adaptive Parameter Estimation Over Adversarial Networks

A. Node Types and Distributed Estimation

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 $\tau \in T = \{\mathsf{M}, \mathsf{S}, \mathsf{H}\}$ and it is unknown to the other nodes in the network. $\mathsf{M}, \mathsf{S}, \mathsf{H}$ 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. Due to lack of knowledge of agents' types the setting is characterized by *incomplete information* [20].

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

$$\mathbf{d}_{k,i} = \mathbf{u}_{k,i} w^o + \mathbf{v}_{k,i} \tag{1}$$

where $\mathbf{v}_{k,i} \in \mathbb{C}$ is a white–noise process with zero mean and power $\sigma^2_{v,k}$ 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., $E\{\mathbf{u}^H_{k,i}\mathbf{u}_{\ell,i'}\} = R_{u,k}\delta_{k,\ell}\delta_{i,i'}$ 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 [25]. 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. Thus, each agent k undertakes the adaptation step

$$\boldsymbol{\psi}_{k,i} = \boldsymbol{\phi}_{k,i-1} + \mu_k \mathbf{u}^H_{k,i}\left[\mathbf{d}_{k,i} - \mathbf{u}_{k,i}\boldsymbol{\phi}_{k,i-1}\right] \tag{2}$$

Agent k, after updating its intermediate estimate $\boldsymbol{\psi}_{k,i}$, utilizes information acquired by its neighbors, in order to form its local estimate of w_0 by undertaking the combination step

$$\boldsymbol{\phi}_{k,i} = \sum_{\ell \in \mathcal{N}_k} z_{k\ell} \boldsymbol{\psi}_{\ell,i}. \tag{3}$$

where, $\boldsymbol{\phi}_{k,i}$ is the local estimate node k has about w^o at time i, $\mu_k > 0$ is the step-size parameter and $\{z_{k\ell} \mid \ell \in \mathcal{N}_k\}$ are non-negative combination coefficients such that

$$\sum_{\ell \in \mathcal{N}_k} z_{k\ell} = 1 \tag{4}$$

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.

B. Admissible Actions and Data Exchange

Nodes exchange data following a distributed random pairing protocol [4], [5], [15]. More specifically, each agent $k \in \mathcal{N}$ is paired with at most one neighboring agent $\ell \in \mathcal{N}_k^-$ when sharing information at time i. Let

$$\mathbb{1}_{k\ell,i} = \mathbb{1}_{\ell k,i} = \begin{cases} 1, & \text{with probability } p_{k\ell} = p_{\ell k}, \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

where $\mathbb{1}_{k\ell,i} = 1$ indicates the pairing of agents k and $\ell \in \mathcal{N}_k^-$, If node k is not paired with any neighbor $\ell \in \mathcal{N}_k^-$ at time i, we use $\mathbb{1}_{kk,i}$. Hence

$$\sum_{\ell \in \mathcal{N}_k} \mathbb{1}_{k\ell,i} = \sum_{\ell \in \mathcal{N}_k} p_{k\ell} = 1, \quad \mathbb{1}_{k\ell,i}\mathbb{1}_{k\ell',i} = 0, \quad \forall \ell \neq \ell' \tag{6}$$

Under the random pairing protocol, the combination step is

$$\boldsymbol{\phi}_{k,i} = z_{kk}\boldsymbol{\psi}_{k,i} + \big(1 - z_{kk}\big) \sum_{\ell \in \mathcal{N}_k} \mathbb{1}_{k\ell,i} \boldsymbol{\psi}_{\ell,i}. \tag{7}$$

where the combination coefficient z_{kk} expresses the weight that agent k gives on his own estimate $\boldsymbol{\psi}_{k,i}$. 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. We elaborate on that later on. 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 information with its paired agent $\ell \in \mathcal{N}_k^-$, we assume that a malicious node k injects some perturbation $\boldsymbol{\eta}_{k\ell} \in \mathbb{C}^{L\times 1}$ to its local estimate $\boldsymbol{\psi}_{k,i}$ as follows

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

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

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

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

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

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.

C. 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

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

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

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

where $s_{k\ell,i}$ is a threshold set by k. We will elaborate on the way k selects this threshold later on. The probabilities of misdetection $p_{k\ell,i}^{[m]}$ and false alarm $p_{k\ell,i}^{[f]}$ of an action taken by node

$\ell \in \mathcal{N}_k^-$ at time i are

$$p_{k\ell,i}^{[m]} = \Pr(\hat{a}_{\ell k,i} = 1 | a_{\ell k,i} = -1) = \Pr(\|\boldsymbol{\psi}_{\ell,i} + \boldsymbol{\eta}_{\ell k} - \widehat{\boldsymbol{\psi}}_{k,i}\|^2 \leq s_{k\ell,i}), \quad (13)$$

$$p_{k\ell,i}^{[f]} = \Pr(\hat{a}_{\ell k,i} = -1 | a_{\ell k,i} = 1) = \Pr(\|\boldsymbol{\psi}_{\ell,i} - \widehat{\boldsymbol{\psi}}_{k,i}\|^2 > s_{k\ell,i}). \quad (14)$$

Thus, after agent k has performed the action detection process, it combines the information received by its paired agent (if any – i.e., $\boldsymbol{a}_{\ell k,i} \neq 0$) only if it has detected this information as non-falsified (i.e., $\hat{\boldsymbol{a}}_{\ell k,i} = 1$), otherwise it ignores the information sent by ℓ. Hence, assuming that $\boldsymbol{a}_{kk,i} = \hat{\boldsymbol{a}}_{kk,i} = 1$, $\boldsymbol{\varphi}_{kk,i} = \boldsymbol{\psi}_{k,i}$ for all $i \geq 1$, the combination step under ADC is

$$\boldsymbol{\phi}_{k,i} = \begin{cases} z_{kk}\boldsymbol{\psi}_{k,i} + \big(1 - z_{kk}\big) \sum_{\ell \in \mathcal{N}_k} \mathbb{1}_{k\ell,i} \boldsymbol{\varphi}_{\ell k,i}, & \text{if } \hat{\boldsymbol{a}}_{\ell k,i} = 1, \\ \boldsymbol{\psi}_{k,i}, & \text{otherwise.} \end{cases} \quad (15)$$

Remark 1: If communication noise is present during the information exchange, changes in the design of the detector are required. More specifically, the detector will have to filter out channel noise before discriminating among the actions of the agents.

Remark 2: The combination step under ADC (see (15)) fuses local estimates of neighboring nodes and seeks to achieve smaller Mean Square Deviation (MSD) compared to the case where the nodes do not cooperate. However, unlike the adaptive LMS combination rules described in [25], the combination step of ADC takes into account the result of the detection process performed by each node k (see (12), (15)). Thus, corrupted estimates sent by malicious neighbors, if correctly detected, are disregarded in the combination step, enabling secure distributed parameter estimation.

D. Adaptive Reputation Update

Nodes maintain records about the trustworthiness or reputation of their neighbors. Let $\boldsymbol{\theta}_{\ell k,i}^{[\mathsf{H}]}, \boldsymbol{\theta}_{\ell k,i}^{[\mathsf{S}]}, \boldsymbol{\theta}_{\ell k,i}^{[\mathsf{M}]}$ denote the reputation values that agent k keeps about agent ℓ at time i (based on the behavior the latter has exhibited in the past) and $\boldsymbol{\theta}_{\ell k,i} = [\boldsymbol{\theta}_{\ell k,i}^{[\mathsf{H}]}, \boldsymbol{\theta}_{\ell k,i}^{[\mathsf{S}]}, \boldsymbol{\theta}_{\ell k,i}^{[\mathsf{M}]}]^T$. $\boldsymbol{\theta}_{\ell k,i}^{[\mathsf{H}]}, \boldsymbol{\theta}_{\ell k,i}^{[\mathsf{S}]}, \boldsymbol{\theta}_{\ell k,i}^{[\mathsf{M}]}$ denote the reputation values 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

$$\boldsymbol{\theta}_{\ell k,i+1} = \boldsymbol{\theta}_{\ell k,i} r_k^{[\hat{\boldsymbol{a}}_{\ell k,i}]} + e^{[\hat{\boldsymbol{a}}_{\ell k,i}]} \left(1 - r_k^{[\hat{\boldsymbol{a}}_{\ell k,i}]}\right) \quad (16)$$

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 reputation that is compatible with the detected action $\hat{\boldsymbol{a}}_{\ell k,i}$. We will adopt the ordering $r_k^{[1]} > r_k^{[0]} > r_k^{[-1]}$, as malicious action is more harmful than selfish and thus more punishment should be enforced.

As stated earlier, honest agents select actions according to a predefined and globally known strategy. The purpose is to reward well–behaved neighbors through information sharing and punish neighbors that have shown selfish or malicious behavior in the past by sending them no information. Thus, the probability that an honest agent k selects action 1 should be in accordance with the reputation value k keeps about the honesty of its paired agent ℓ. Thus, the honest policy, denoted as π_{H}, constitutes a *mixed strategy* and it is defined as follows

$$\boldsymbol{a}_{k\ell,i} = \begin{cases} 1, & \text{with probability } \boldsymbol{\theta}_{\ell k,i}^{[\mathsf{H}]}, \\ 0, & \text{with probability } 1 - \boldsymbol{\theta}_{\ell k,i}^{[\mathsf{H}]} \end{cases} \quad (17)$$

Remark 3: The idea of deploying a type of agents committed to follow a fixed policy through time to stimulate cooperation stems from [29]. The authors in [29] assume a *commitment type* of agents, which follow always the *tit-for-tat* strategy, interacting with normal agents in a repeated PD.

E. Instantaneous Payoffs

The actions of selfish and malicious agents are influenced by instantaneous payoffs which derive from reception benefits and transmission payoffs. Reception benefits result from enhanced estimation performance offered by the data exchange with neighboring nodes. Transmission payoffs entail communication costs incurred in transmission of information for selfish agents and illegal gains offered to malicious agents when their malicious actions pass undetected. Honest nodes, as explained in previous subsection, follow always π_{H} with no reward attached.

More specifically, let $k \in \mathcal{N}$ be paired with $\ell \in \mathcal{N}_k^-$. Then, the profit gained by node k due to the information shared by node ℓ depends on both the true and the detected transmission action taken by ℓ towards k (i.e., $\boldsymbol{a}_{\ell k,i}$ and $\hat{\boldsymbol{a}}_{\ell k,i}$, respectively (see (15), (10)). Moreover, the individual action $\boldsymbol{a}_{k\ell,i}$ and how this action is detected by the paired agent ℓ, i.e., $\hat{\boldsymbol{a}}_{k\ell,i}$, affects the reward as transmitter of agent k. This reward depends on the type of the agent k, i.e., τ_k, because each type has different interests. The reception and transmission rewards of k are denoted by $f_k(\boldsymbol{a}_{\ell k,i}, \hat{\boldsymbol{a}}_{\ell k,i})$ and $g_k(\boldsymbol{a}_{k\ell}, \hat{\boldsymbol{a}}_{k\ell,i}, \tau_k)$, respectively. Thus, agent's k *instantaneous reward* function is

$$R_k(\boldsymbol{a}_{k\ell,i}, \boldsymbol{a}_{\ell k,i}, \hat{\boldsymbol{a}}_{k\ell,i}, \hat{\boldsymbol{a}}_{\ell k,i}, \tau_k) = f_k(\boldsymbol{a}_{\ell k,i}, \hat{\boldsymbol{a}}_{\ell k,i}) + g_k(\boldsymbol{a}_{k\ell,i}, \hat{\boldsymbol{a}}_{k\ell,i}, \tau_k) \quad (18)$$

where

$$f_k(\boldsymbol{a}_{\ell k,i}, \hat{\boldsymbol{a}}_{\ell k,i}) = \begin{cases} -\zeta_k\big(z_{kk}\boldsymbol{\psi}_{k,i} + (1 - z_{kk})\boldsymbol{\psi}_{\ell,i}\big), & \text{if } (1,1), \\ -\zeta_k\big(z_{kk}\boldsymbol{\psi}_{k,i} + (1 - z_{kk})(\boldsymbol{\psi}_{\ell,i} + \boldsymbol{\eta}_{\ell k})\big), & \text{if } (-1,1), \\ -\zeta_k\big(\boldsymbol{\psi}_{k,i}\big), & \text{otherwise} \end{cases} \quad (19)$$

where ζ_k denotes the instantaneous Mean Square Error (MSE) cost associated with node k based on the estimate $\boldsymbol{\phi}_{k,i}$ of w^0. The corresponding *a priori* estimation error is $\mathbf{d}_{k,i+1} - \mathbf{u}_{k,i+1}\boldsymbol{\phi}_{k,i}$ and it measures how close the weight estimate matches the measurements $\{\mathbf{d}_{k,i+1}, \mathbf{u}_{k,i+1}\}$ to each other [15], [25]. Utilizing (1) the *a priori* estimation error is $\mathbf{u}_{k,i+1}(w^0 - \boldsymbol{\phi}_{k,i}) + \mathbf{v}_{k,i+1}$.

Then the MSE cost is given by

$$\begin{aligned}\zeta_k(\boldsymbol{\phi}_{k,i}) &= \mathbb{E}\{|\mathbf{d}_{k,i+1} - \mathbf{u}_{k,i+1}\boldsymbol{\phi}_{k,i}|^2_{R_{u,k}}\} \\ &= \mathbb{E}\{\|w^o - \boldsymbol{\phi}_{k,i}\|^2_{R_{u,k}}\} + \sigma^2_{v,k} \end{aligned} \tag{20}$$

The lower $\zeta_k(\boldsymbol{\phi}_{k,i})$ is, the better the estimation accuracy for agent k at time i becomes and thus, the bigger the reward as receiver is. $f_k(\boldsymbol{a}_{\ell k,i}, \hat{\boldsymbol{a}}_{\ell k,i})$ expresses the gain in terms of estimation accuracy for agent k at time i after $\boldsymbol{\phi}_{k,i-1}$ is updated to $\boldsymbol{\phi}_{k,i}$. The first line of the right–hand side of (19) corresponds to the case in which agent ℓ has sent its estimate to agent k ($\boldsymbol{a}_{\ell k,i} = 1$) and the latter has correctly detected it as a non–falsified estimate ($\hat{\boldsymbol{a}}_{\ell k,i} = 1$) and thus uses it to perform the combination step. In the second line, ℓ sent a falsified estimate ($\boldsymbol{a}_{\ell k,i} = -1$), agent k miss-detected ($\hat{\boldsymbol{a}}_{\ell k,i} = 1$) it and uses it in the combination step. If ℓ sent no information to k ($\boldsymbol{a}_{\ell k,i} = 0$), or k detected the information as falsified ($\hat{\boldsymbol{a}}_{\ell k,i} = -1$), then agent k uses only its own intermediate estimate in the combination step (see (15)). Note that $f_k(\cdot)$ does not depend on node's k type, since all nodes are interested in receiving good estimates to incorporate them in their estimation task.

The transmission reward reflects the cost of transmission for a selfish node and the illegal gains acquired by a malicious node. It is given by

$$g_k(\boldsymbol{a}_{k\ell,i}, \hat{\boldsymbol{a}}_{k\ell,i}, \tau_k) = \begin{cases} m_k, & \text{if } \boldsymbol{a}_{k\ell,i} = -1, \hat{\boldsymbol{a}}_{k\ell,i} = 1, \tau_k = \mathsf{M} \\ -c_k, & \text{if } \boldsymbol{a}_{k\ell,i} = 1, \tau_k = \mathsf{S}, \\ 0, & \text{otherwise} \end{cases} \tag{21}$$

where $c_k > 0$ is the transmission cost and $m_k > 0$ is the illegal gain of a malicious action. Selfish agents seek to minimize transmission cost and maximize reception gains. Thus, if a selfish node k chooses to send its intermediate estimate (i.e., $\boldsymbol{a}_{k\ell,i} = 1$) to a neighboring node ℓ, then a cost c_k is incurred.

A malicious node k aims at harming other nodes and in doing so acquires illegal gain m_k [16]. Since all nodes perform action detection, the illegal gain m_k depends on the detection outcome of its action performed by the paired node ℓ. More specifically, the malicious agent k would acquire this illegal gain only in case of miss-detection of the attack ($\hat{\boldsymbol{a}}_{k\ell,i} = 1$).

A node k can utilize the probabilities of miss-detection and false alarm $p^{[m]}_{k\ell,i}, p^{[f]}_{k\ell,i}, p^{[m]}_{\ell k,i}, p^{[f]}_{\ell k,i}$ to evaluate the following conditional *expected instantaneous reward*

$$\begin{aligned}&\widetilde{R}_k(a_{k\ell,i}, a_{\ell k,i}, \tau_k) \\ &= \mathbb{E}\Big\{R_k\big(a_{k\ell,i}, a_{\ell k,i}, \hat{\boldsymbol{a}}_{k\ell,i}, \hat{\boldsymbol{a}}_{\ell k,i}, \tau_k\big)|\boldsymbol{\phi}_{k,i-1}\Big\} \\ &= \beta_k(a_{\ell k,i}) + \delta_k(a_{k\ell,i}, \tau_k) \end{aligned} \tag{22}$$

where $\beta_k(a_{\ell k,i})$ is

$$\begin{aligned}\beta_k(a_{\ell k,i}) &= \mathbb{E}\big\{f_k(a_{\ell k,i}, \hat{a}_{\ell k,i})|\boldsymbol{\phi}_{k,i-1}\big\} \\ &= \sum_{\hat{a}_{\ell k,i}} \Pr(\hat{a}_{\ell k,i}|a_{\ell k,i}) \mathbb{E}\big\{f_k(a_{\ell k,i}, \hat{a}_{\ell k,i})|\boldsymbol{\phi}_{k,i-1}\big\} \end{aligned} \tag{23}$$

Straightforward calculations give

$$\begin{aligned}\beta_k(1) &= -\zeta_k\Big(z_{kk}\boldsymbol{\psi}_{k,i} + \big(1 - z_{kk}\big)\boldsymbol{\psi}_{\ell,i}|\boldsymbol{\phi}_{k,i-1}\Big) \\ &\quad \times \Big[1 - p^{[f]}_{k\ell,i}\Big] - \zeta_k\Big(\boldsymbol{\psi}_{k,i}|\boldsymbol{\phi}_{k,i-1}\Big)p^{[f]}_{k\ell,i} \end{aligned} \tag{24}$$

$$\beta_k(0) = -\zeta_k\Big(\boldsymbol{\psi}_{k,i}|\boldsymbol{\phi}_{k,i-1}\Big) \tag{25}$$

$$\begin{aligned}\beta_k(-1) &= -\zeta_k\Big(\boldsymbol{\psi}_{k,i}|\boldsymbol{\phi}_{k,i-1}\Big)\Big[1 - p^{[m]}_{k\ell,i}\Big] \\ &\quad - \zeta_k\Big(z_{kk}\boldsymbol{\psi}_{k,i} + (1 - z_{kk})\big(\boldsymbol{\psi}_{\ell,i} + \boldsymbol{\eta}_{\ell k}\big)|\boldsymbol{\phi}_{k,i-1}\Big)p^{[m]}_{k\ell,i} \end{aligned} \tag{26}$$

with

$$\zeta_k\Big(\mathbf{x}_{k,i}|\boldsymbol{\phi}_{k,i-1}\Big) = \mathbb{E}\big\{\|w^o - \mathbf{x}_{k,i}\|^2_{R_{u,k}}\big|\boldsymbol{\phi}_{k,i-1}\big\} + \sigma^2_{v,k}. \tag{27}$$

for an arbitrary estimate $\mathbf{x}_{k,i}$ and

$$\begin{aligned}\delta_k(a_{k\ell,i}, \tau_k) &= \mathbb{E}\Big\{g_k(a_{k\ell,i}, \hat{\boldsymbol{a}}_{k\ell,i}, \tau_k)\Big\} \\ &= \sum_{\hat{a}_{k\ell,i}} \Pr(\hat{a}_{k\ell,i}|a_{k\ell,i}) g_k(a_{k\ell,i}, \hat{a}_{k\ell,i}, \tau_k) \\ &= \begin{cases} m_k p^{[m]}_{\ell k,i} & \text{if } a_{k\ell,i} = -1, \tau_k = \mathsf{M}, \\ -c_k & \text{if } a_{k\ell,i} = 1, \tau_k = \mathsf{S}, \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{28}$$

Eq. (24) expresses the expected gain when the paired agent ℓ selects action $\boldsymbol{a}_{\ell k,i} = 1$ at time i. In case of correct detection (with probability $1 - p^{[f]}_{k\ell,i}$), k utilizes estimate $\boldsymbol{\psi}_{\ell,i}$ to perform the combination step and update its new estimate $\boldsymbol{\phi}_{k,i}$. On the other hand, in case of false alarm (with probability $p^{[f]}_{k\ell,i}$), k uses only its own local estimate $\boldsymbol{\psi}_{k,i}$. The same happens if ℓ does not share its estimate with k (i.e., $\boldsymbol{a}_{\ell k,i} = 0$), as it is expressed by (25). Finally, (26) refers to the case when ℓ sends a corrupted estimate to k (i.e., $\boldsymbol{a}_{\ell k,i} = -1$). If k miss-detects it as a good one (with probability $p^{[m]}_{k\ell,i}$), then k utilizes the corrupted estimate $\boldsymbol{\psi}_{\ell,i} + \boldsymbol{\eta}_{\ell k}$ in the combination step, while if k correctly detects it as a corrupted one (with probability $1 - p^{[m]}_{k\ell,i}$), then k uses only its local estimate $\boldsymbol{\psi}_{k,i}$.

The relations among the various parameters affecting the agents' decision-making process are illustrated in Fig. 1.

III. Static Game Analysis

In this section, we consider the static case, we show that all possible interactions of strategic agents are instances of the well-known PD and discuss the Pareto inefficiency of NE.

A. Dominant Strategies and Nash Equilibrium

Given the definition of *expected instantaneous reward*, in this section we analyze the pairwise one–time instant interactions of the agents. The one–shot game (*static game* or *matrix game*) can be represented in matrix form as depicted in Table I(a)–(c) for given agent types.

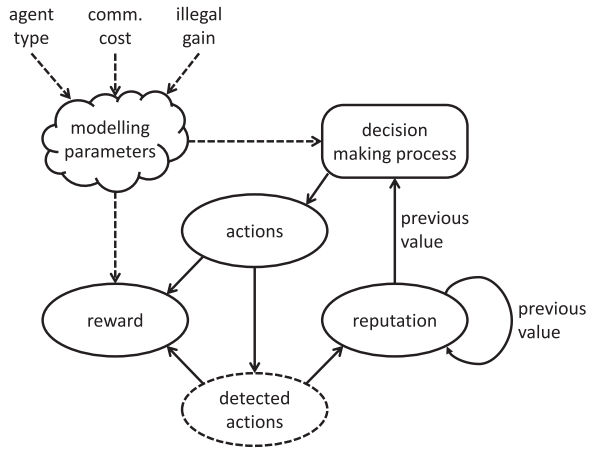


Fig. 1. Relations among the various notions.

TABLE I
MATRIX GAMES FOR STRATEGIC NODES, WHERE WE DEFINE $\widetilde{m}_k = m_k p^{[m]}_{\ell k,i}$ AND $\widetilde{m}_\ell = m_\ell p^{[m]}_{k\ell,i}$

$a_{k\ell,i} \backslash a_{\ell k,i}$	0	1
0	$\beta_k(0), \beta_\ell(0)$	$\beta_k(1), \beta_\ell(0) - c_\ell$
1	$\beta_k(0) - c_k, \beta_\ell(1)$	$\beta_k(1) - c_k, \beta_\ell(1) - c_\ell$

(a) Selfish–selfish matrix game: $(\tau_k, \tau_\ell) = (\mathsf{S}, \mathsf{S})$

$a_{k\ell,i} \backslash a_{\ell k,i}$	-1	1
-1	$\beta_k(-1) + \widetilde{m}_k, \beta_\ell(-1) + \widetilde{m}_\ell$	$\beta_k(1) + \widetilde{m}_k, \beta_\ell(-1)$
1	$\beta_k(-1), \beta_\ell(1) + \widetilde{m}_\ell$	$\beta_k(1), \beta_\ell(1)$

(b) Malicious–malicious matrix game: $(\tau_k, \tau_\ell) = (\mathsf{M}, \mathsf{M})$

$a_{k\ell,i} \backslash a_{\ell k,i}$	-1	1
0	$\beta_k(-1), \beta_\ell(0) + \widetilde{m}_\ell$	$\beta_k(1), \beta_\ell(0)$
1	$\beta_k(-1) - c_k, \beta_\ell(1) + \widetilde{m}_\ell$	$\beta_k(1) - c_k, \beta_\ell(1)$

(c) Selfish–malicious matrix game: $(\tau_k, \tau_\ell) = (\mathsf{S}, \mathsf{M})$

Let us analyze each one of the possible matrix games among strategic agents. These matrix games are the following, selfish–selfish, malicious–malicious and selfish–malicious which are provided in Table I(a)–(c), respectively. In these tables, the expected instantaneous reward for row player k is the first component of each entry, while the second component is the expected reward for the column player $\ell \in \mathcal{N}_k^-$. By observing the entries of the matrix games in which at least one selfish agent is present [i.e.,, Table I(a) and (c)], we can conclude that a selfish agent k acquires always higher reward if it selects the selfish action ($\boldsymbol{a}_{k\ell,i} = 0$) instead of the cooperative action ($\boldsymbol{a}_{k\ell,i} = 1$) regardless of the action and the type of paired node ℓ. Thus, $\boldsymbol{a}_{k\ell,i} = 0$ is the *dominant strategy* for a selfish agent k no matter what the opponent's type is.

Following the same reasoning, from the matrix games in which a malicious agent participates [i.e., Table I(b) and (c)], we can verify that the dominant strategy for a malicious node ℓ is the malicious action ($\boldsymbol{a}_{\ell k,i} = -1$), regardless of the opponent's type and action selected, for $p^{[m]}_{k\ell,i} > 0$, which is verified if the detector is not perfect.[1] Hence, considering that the detector is imperfect, i.e.,, $p^{[m]}_{k\ell,i} > 0$, then for every stage game the unique NE of two paired agents k and ℓ is the action profile (i.e $(\boldsymbol{a}_{k\ell,i}, \boldsymbol{a}_{\ell k,i})$) $(0,0), (-1,-1), (0,-1)$ when their corresponding type profile (τ_k, τ_ℓ) is $(\mathsf{S},\mathsf{S}), (\mathsf{M},\mathsf{M}), (\mathsf{S},\mathsf{M})$, respectively.

Finally, the honest–selfish and honest–malicious interactions are not depicted in table form but they could be as well, with the honest agent having the action choices 0 and 1 and no reward entry. Likewise, with the same reasoning as previously, for a selfish (resp. for a malicious) agent k the optimal action is $\boldsymbol{a}_{k\ell,i} = 0$ (resp. $\boldsymbol{a}_{k\ell,i} = -1$), because they are again dominant strategies.

Remark 4: We note that if agent interactions are not limited to one-on-one scenario and there is the possibility of taking multiple actions simultaneously at a time instant, then in the static case a strategic agent would select again the dominant strategy of not cooperating (i.e., not to send information if it is selfish and attack if it is malicious) towards every agent it interacts. The same stands (in the static case) if the pairing is not symmetrical. In that case of non-symmetrical pairing, the case of only one of the two nodes being able to take action might arise.

B. Pareto Optimality

The matrix games presented in the previous section for given agent types are instances of the well–known PD. In this section, we characterize when the action profile associated with mutual truth–telling cooperation $\big($i.e., $(\boldsymbol{a}_{k\ell,i}, \boldsymbol{a}_{\ell k,i}) = (1,1)\big)$ is Pareto optimal in each matrix game. Let us define the benefit for a node k when the paired agent ℓ is selfish and takes an honest action instead of a selfish action

$$\begin{aligned}\mathbf{b}^{[0]}_{k,i} &= \widetilde{R}_k(a_{k\ell,i}, a_{\ell k,i} = 1, \tau_k) - \widetilde{R}_k(a_{k\ell,i}, a_{\ell k,i} = 0, \tau_k)\\ &= \beta_k(1) - \beta_k(0)\\ &= \big[1 - p^{[f]}_{k\ell,i}\big] \cdot \Big[\zeta_k\Big(\boldsymbol{\psi}_{k,i}\big|\boldsymbol{\phi}_{k,i-1}\Big)\\ &\quad - \zeta_k\Big(z_{kk}\boldsymbol{\psi}_{k,i} + \big(1 - z_{kk}\big)\boldsymbol{\psi}_{\ell,i}\big|\boldsymbol{\phi}_{k,i-1}\Big)\Big] \qquad (29)\end{aligned}$$

We similarly define the benefit for a node k when the paired agent ℓ is malicious and takes an honest action instead of a malicious action

$$\begin{aligned}\mathbf{b}^{[-1]}_{k,i} &= \widetilde{R}_k(a_{k\ell,i}, a_{\ell k,i} = 1, \tau_k) - \widetilde{R}_k(a_{k\ell,i}, a_{\ell k,i} = -1, \tau_k)\\ &= \beta_k(1) - \beta_k(-1)\\ &= \big[1 - p^{[m]}_{k\ell,i} - p^{[f]}_{k\ell,i}\big] \cdot \zeta_k\Big(\boldsymbol{\psi}_{k,i}\big|\boldsymbol{\phi}_{k,i-1}\Big)\\ &\quad - \big[1 - p^{[f]}_{k\ell,i}\big] \cdot \zeta_k\Big(z_{kk}\boldsymbol{\psi}_{k,i} + \big(1 - z_{kk}\big)\boldsymbol{\psi}_{\ell,i}\big|\boldsymbol{\phi}_{k,i-1}\Big)\\ &\quad + \zeta_k\Big(z_{kk}\boldsymbol{\psi}_{k,i} + \big(1 - z_{kk}\big)\big(\boldsymbol{\psi}_{\ell,i} + \boldsymbol{\eta}_{\ell k}\big)\big|\boldsymbol{\phi}_{k,i-1}\Big)p^{[m]}_{k\ell,i}. \qquad (30)\end{aligned}$$

[1] If it is $p^{[m]}_{k\ell,i} = 0$, then a malicious agent ℓ would be indifferent between its two actions in the respective matrix games.

Given the previous definitions of $\mathbf{b}_{k,i}^{[0]}$ and likewise for $\mathbf{b}_{\ell,i}^{[0]}$, when both paired agents k and ℓ are selfish [see Table I(a)], the action profile $\left(\boldsymbol{a}_{k\ell,i}, \boldsymbol{a}_{\ell k,i}\right) = (1,1)$ is Pareto superior to the action profile $(0,0)$ and this is denoted as $(1,1) \overset{P}{\succ} (0,0)$ if the following two conditions hold

$$\mathbf{b}_{k,i}^{[0]} > c_k \text{ and } \mathbf{b}_{\ell,i}^{[0]} > c_\ell. \tag{31}$$

Clearly, when both paired agents are selfish, the action profile $\left(\boldsymbol{a}_{k\ell,i}, \boldsymbol{a}_{\ell k,i}\right) = (0,0)$ is Pareto superior to the action profile $(1,1)$, i.e., $(0,0) \overset{P}{\succ} (1,1)$, if

$$\mathbf{b}_{k,i}^{[0]} < c_k \text{ and } \mathbf{b}_{\ell,i}^{[0]} < c_\ell. \tag{32}$$

Following a similar reasoning, when both paired nodes k and ℓ are malicious [see Table I(b)] we have

$$\begin{aligned}
&(1,1) \overset{P}{\succ} (-1,-1) \text{ if } \left[\mathbf{b}_{k,i}^{[-1]}, \mathbf{b}_{\ell,i}^{[-1]}\right] \succ \left[m_k p_{\ell k,i}^{[m]}, m_\ell p_{k\ell,i}^{[m]}\right] \\
&(-1,-1) \overset{P}{\succ} (1,1) \text{ if } \left[\mathbf{b}_{k,i}^{[-1]}, \mathbf{b}_{\ell,i}^{[-1]}\right] \prec \left[m_k p_{\ell k,i}^{[m]}, m_\ell p_{k\ell,i}^{[m]}\right]
\end{aligned} \tag{33}$$

where, for arbitrary vectors $\mathbf{x}, \mathbf{y} \in \mathbb{C}^{L\times 1}$, the operator $\mathbf{x} \succ \mathbf{y}$ (resp. $\mathbf{x} \prec \mathbf{y}$) denote that the element-wise inequality $x_j > y_j$ (resp. $x_j < y_j$) hold for all $j \in \{1,2,\ldots\}$. Moreover, when a selfish node k and a malicious node ℓ are paired, [see Table I(c)] we can note that

$$\begin{aligned}
&(1,1) \overset{P}{\succ} (0,-1) \text{ if } \left[\mathbf{b}_{k,i}^{[-1]}, \mathbf{b}_{\ell,i}^{[0]}\right] \succ \left[c_k, m_\ell p_{k\ell,i}^{[m]}\right], \\
&(0,-1) \overset{P}{\succ} (1,1) \text{ if } \left[\mathbf{b}_{k,i}^{[-1]}, \mathbf{b}_{\ell,i}^{[0]}\right] \prec \left[c_k, m_\ell p_{k\ell,i}^{[m]}\right].
\end{aligned} \tag{34}$$

Although the action profiles $(0,1)$, or $(1,0)$ for two selfish paired agents, $\{(-1,1),(1,-1)\}$ for two malicious and $\{(0,1),(1,-1)\}$ for a selfish paired with a malicious can also be Pareto optimal, they are beneficial only for one agent.

IV. Long–Term Rewards, Best–Response Policies and Learning

Although cooperation is not the optimal choice for strategic agents in the static case, repeated interactions can give rise to honest behavior. In this section, we explore conditions that enforce strategic agents to cooperate. In doing so, we utilize the distributed reputation model described in Section II-D and design an algorithm that combines optimal long-term decision making with secure parameter estimation in an adversarial setting. We show that truth–telling cooperation is achievable under suitable conditions.

A. Long–Term Rewards

In the case of repeated interactions (beginning from time i), each agent k maximizes the expected sum of discounted instantaneous rewards

$$\sum_{n=i}^{\infty} \lambda_k^{n-i} \mathbb{E}\left\{ R_k\left(\boldsymbol{a}_{k\ell,n}, \boldsymbol{a}_{\ell k,n}, \hat{\boldsymbol{a}}_{k\ell,n}, \hat{\boldsymbol{a}}_{\ell k,n}, \tau_k\right) \middle| \boldsymbol{\phi}_{k,i-1} \right\} \tag{35}$$

where $\boldsymbol{a}_{k\ell,n} \in \mathcal{A}(\tau_k)$, the expectation is taken over all random variables for $n \geq i$, and $\lambda_k \in (0,1)$ is a *discount factor* that defines the relative importance given by agent k to short–term rewards. The complexities of the above formulation are aggravated by the randomness in future pairings, future action selections of paired agents and future action detection outcomes, in conjunction with the uncertainties associated with the future values of parameter estimates and detection probabilities. To create an amenable computational setting, we employ the *bounded rationality* assumption [15], [28] which implies that agents have computational limitations. Thus, under the assumption that agents are boundedly rational, we impose three simplifying assumptions: (i) stationarity of reward maximizers, (ii) stationarity in future pairings, local LMS estimates, detection probabilities and reputations (iii) reputation-based beliefs about the identity of paired agent. Following the standard practice in multi-agent learning algorithms [23], [28], stationary actions are confined to time invariant forms (assumption (i))

$$\boldsymbol{a}_{k\ell,n} = a_{k\ell,i}, \quad \forall\, n \geq i \tag{36}$$

Under assumption (ii) each agent solves (35) assuming that

$$\boldsymbol{\phi}_{k,n} = \boldsymbol{\phi}_{k,i-1} \tag{37}$$

$$\mathbb{1}_{k\ell,n} = \mathbb{1}_{k\ell,i} \tag{38}$$

$$p_{k\ell,n}^{[m]} = p_{k\ell,i}^{[m]},\ p_{\ell k,n}^{[m]} = p_{\ell k,i}^{[m]} \tag{39}$$

$$p_{k\ell,n}^{[f]} = p_{k\ell,i}^{[f]},\ p_{\ell k,n}^{[f]} = p_{\ell k,i}^{[f]} \tag{40}$$

$$\boldsymbol{\theta}_{\ell k,n} = \boldsymbol{\theta}_{\ell k,i}. \tag{41}$$

for all $n \geq i$ and $\ell \in \mathcal{N}_k^-$. Bounded rationality is a further step towards building a stationary landscape for the decision-making process of the agents and it should be noted that although these assumptions are made by a strategic agent at each time instant to decide on its best available action, these quantities are still time-varying. The third assumption (i.e., assumption (iii)) copes with the uncertainty in future action $\boldsymbol{a}_{\ell k,n}$ of the paired agent ℓ. Reputation values are employed as beliefs about the unknown type of agent ℓ. We point out that these proxies should not be conceived as a Bayesian belief mechanism [20], but as enablers for the scalable algorithm described below. In more specific terms, the beliefs of agent k about the future actions of agent ℓ are formed as follows

$$B(\boldsymbol{a}_{\ell k,n}) = \begin{cases} \boldsymbol{\theta}_{\ell k,n}^{[\mathsf{H}]} \boldsymbol{\theta}_{k\ell,n}^{[\mathsf{H}]}, & \boldsymbol{a}_{\ell k,n} = 1 \\ \boldsymbol{\theta}_{\ell k,n}^{[\mathsf{H}]} (1 - \boldsymbol{\theta}_{k\ell,n}^{[\mathsf{H}]}) + \boldsymbol{\theta}_{\ell k,n}^{[\mathsf{S}]}, & \boldsymbol{a}_{\ell k,n} = 0 \\ \boldsymbol{\theta}_{\ell k,n}^{[\mathsf{M}]}, & \boldsymbol{a}_{\ell k,n} = -1 \end{cases} \tag{42}$$

The above formula takes into account that the honest policy π_{H} is common knowledge. Thus, if k has belief $\boldsymbol{\theta}_{\ell k,n}^{[\mathsf{H}]}$ that ℓ is honest, then it expects the latter to follow π_{H} (i.e., $\boldsymbol{a}_{\ell k,n} = 1$ with probability $\boldsymbol{\theta}_{k\ell,n}^{[\mathsf{H}]}$). Moreover, guided by the static NE analyzed in Section III, agent k would expect a selfish agent ℓ to select $\boldsymbol{a}_{\ell k,n} = 0$ and a malicious node ℓ to select $\boldsymbol{a}_{\ell k,n} = -1$ because these are the dominant strategies of strategic agents in the respective matrix games. This rationale is in accordance with the minimax principle which has been widely utilized in

multi-agent learning algorithms [30] and serves as a worst–case scenario (or safety level) on the expectation of other agents' behavior. This means that each agent k expects its (strategic) opponents to select the action minimizing agent's k payoff. This principle helps building scalable opponent-independent multi-agent learning algorithms as in [31].

Based on the above simplifications, k chooses the action $a_{k\ell,i} \in A(\tau_k)$ to maximize

$$J_k(a_{k\ell,i}, \tau_k) = \sum_{n=i}^{\infty} \lambda_k^{n-i} \mathbb{E}\{\beta_k(\boldsymbol{a}_{\ell k,n}) + \delta_k(a_{k\ell,i}, \tau_k) | \mathcal{M}_{k,i}\}. \quad (43)$$

where $\mathcal{M}_{k,i} = \{\boldsymbol{\phi}_{k,i-1}, \boldsymbol{\theta}_{\ell k,i}, \boldsymbol{\theta}_{k\ell,i}, B(\boldsymbol{a}_{\ell k,n}), \pi_{\mathsf{H}}\}$ denotes the information k possesses at time i. Then, the best response rule for a selfish agent k paired with an agent $\ell \in \mathcal{N}_k^-$ at time i is

$$a_{k\ell,i} = \begin{cases} 1, & \text{if } J_k(a_{k\ell,i} = 1, \tau_k = \mathsf{S}) > \\ & J_k(a_{k\ell,i} = 0, \tau_k = \mathsf{S}), \\ 0, & \text{otherwise.} \end{cases} \quad (44)$$

Similarly, the best response rule for a malicious agent k paired with an agent $\ell \in \mathcal{N}_k^-$ at time i is

$$a_{k\ell,i} = \begin{cases} 1, & \text{if } J_k(a_{k\ell,i} = 1, \tau_k = \mathsf{M}) > \\ & J_k(a_{k\ell,i} = -1, \tau_k = \mathsf{M}), \\ -1 & \text{otherwise.} \end{cases} \quad (45)$$

The best response of each strategic agent k reduces to a threshold-based policy as demonstrated in the next theorem.

Theorem 1: Under assumptions (36)–(42), the best-response rule of a selfish agent k at time i is

$$a_{k\ell,i} = \begin{cases} 1, & \text{if } \mathbf{b}_{k,i}^{[0]} \boldsymbol{x}_{k,i}^{[\mathsf{S}]} \boldsymbol{\theta}_{\ell k,i}^{[\mathsf{H}]} > c_k, \\ 0, & \text{otherwise.} \end{cases} \quad (46)$$

Similarly, the best-response rule of a malicious agent is given by

$$a_{k\ell,i} = \begin{cases} 1, & \text{if } \mathbf{b}_{k,i}^{[0]} \boldsymbol{x}_{k,i}^{[\mathsf{M}]} \boldsymbol{\theta}_{\ell k,i}^{[\mathsf{H}]} > p_{\ell k,i}^{[m]} m_k \\ -1, & \text{otherwise.} \end{cases} \quad (47)$$

where

$$\boldsymbol{x}_{k,i}^{[\mathsf{S}]} = (1-\lambda_k)\boldsymbol{\theta}_{k\ell,i}^{[\mathsf{H}]} \left(\frac{\lambda_k A_{k,i}}{1-\lambda_k A_{k,i}} - \frac{\lambda_k r_\ell^{[0]}}{1-\lambda_k r_\ell^{[0]}} \right) + \frac{\lambda_k \Gamma_{k,i}}{(1-\lambda_k A_{k,i})} \quad (48)$$

$$\boldsymbol{x}_{k,i}^{[\mathsf{M}]} = (1-\lambda_k)\boldsymbol{\theta}_{k\ell,i}^{[\mathsf{H}]} \left(\frac{\lambda_k A_{k,i}}{1-\lambda_k A_{k,i}} - \frac{\lambda_k B_{k,i}}{1-\lambda_k B_{k,i}} \right) + \frac{\lambda_k \Gamma_{k,i}}{(1-\lambda_k A_{k,i})} - \frac{\lambda_k Z_{k,i}}{(1-\lambda_k B_{k,i})} \quad (49)$$

and

$$A_{k,i} = \left(1 - p_{\ell k,i}^{[f]}\right) r_\ell^{[1]} + p_{\ell k,i}^{[f]} r_\ell^{[-1]}, \quad (50)$$

$$\Gamma_{k,i} = \left(1 - p_{\ell k,i}^{[f]}\right) \left(1 - r_\ell^{[1]}\right), \quad (51)$$

$$B_{k,i} = r_\ell^{[-1]} + p_{\ell k,i}^{[m]} \left(r_\ell^{[1]} - r_\ell^{[-1]}\right), \quad (52)$$

$$Z_{k,i} = p_{\ell k,i}^{[m]} \left(1 - r_\ell^{[1]}\right). \quad (53)$$

Proof: See the Appendix. ■

B. Estimation of Detection Performance

Obtaining the exact values, or even bounds on the true probabilities of miss–detection and false alarm in (13), (14) is difficult for both the transient phase as well as for the steady–state of LMS. An in-depth analysis of a similar detection scheme takes place in [24] deriving bounds on the detection performance at the steady-state. Moreover, even if such a characterization was available, it would require knowledge about the other agents' private information such as their regressors and noise variances. In a realistic scenario, where the agents perform distributed decision–making based on their imperfect local observations, an approximation of the false alarm and miss-detection probabilities is needed in order to predict the future detection outcomes of the actions. To do that, the agents need more information. Thus, we let the agents exchange at the end of every time instant i the detected outcomes of each other's action (i.e., for paired agents k, ℓ, k informs ℓ about $\hat{\boldsymbol{a}}_{\ell k,i}$ and ℓ informs k about $\hat{\boldsymbol{a}}_{k\ell,i}$). Hence an agent k at the beginning of time i has access to the history of its own past actions and the detected outcomes of its actions for each agent paired with it in the past. These histories are defined as

$$\mathcal{H}_{k,i} = \{\boldsymbol{a}_{k\ell,j} : \ell \in \mathcal{N}_k^-, j < i\}, \quad (54)$$

$$\widehat{\mathcal{H}}_{k,i} = \{\hat{\boldsymbol{a}}_{k\ell,j} : \ell \in \mathcal{N}_k^-, j < i\}. \quad (55)$$

From (54), (55) the histories of the actions and the corresponding detected outcomes towards a particular agent $\ell \in \mathcal{N}_k^-$ are defined

$$\mathcal{H}_{k\ell,i} = \{\boldsymbol{a}_{ky,j} \in \mathcal{H}_{k,i} : y = \ell, j < i\}, \quad (56)$$

$$\widehat{\mathcal{H}}_{k\ell,i} = \{\hat{\boldsymbol{a}}_{ky,j} \in \widehat{\mathcal{H}}_{k,i} : y = \ell, j < i\}. \quad (57)$$

Furthermore, from (56), (57) the partial histories of a particular action a exercised by agent k towards an agent ℓ and the respective detected outcomes are defined as

$$\mathcal{H}_{k\ell,i}^{[a]} = \{\boldsymbol{a}_{kl,j} \in \mathcal{H}_{k\ell,i} : \boldsymbol{a}_{kl,j} = a, j < i\}, \quad (58)$$

$$\widehat{\mathcal{H}}_{k\ell,i}^{[a]} = \{\hat{\boldsymbol{a}}_{kl,j} \in \widehat{\mathcal{H}}_{k\ell,i} : \boldsymbol{a}_{kl,j} \in \mathcal{H}_{kl,i}^{[a]}, j < i\}. \quad (59)$$

Let $\rho_{k\ell,i} = |\mathcal{H}_{k\ell,i}^{[-1]}|$, $\xi_{k\ell,i} = |\mathcal{H}_{k\ell,i}^{[1]}|$ and $\hat{\boldsymbol{a}}^{[n]}$ denotes the nth element of $\widehat{\mathcal{H}}_{k\ell,i}^{[a]}$. Based on the observations contained in

$\mathcal{H}^{[a]}_{k\ell,i}, \widehat{\mathcal{H}}^{[a]}_{k\ell,i}$ agent k can estimate $p^{[m]}_{\ell k,i}$ and $p^{[f]}_{\ell k,i}$ as follows

$$\widetilde{p}^{[m]}_{\ell k,i} = \mathbb{1}_{k\ell,i} \frac{\sum_{n=1}^{\rho_{k\ell,i}} \gamma_k^{\rho_{k\ell,i}-n} \mathbb{1}_{\{\hat{\boldsymbol{a}}^{[n]}=1, \hat{\boldsymbol{a}}^{[n]} \in \widehat{\mathcal{H}}^{[-1]}_{k\ell,i}\}} + \gamma_k^{\rho_{k\ell,i}} h_{1,k}}{\sum_{n=1}^{\rho_{k\ell,i}} \gamma_k^{\rho_{k\ell,i}-n} + \gamma_k^{\rho_{k\ell,i}} h_{2,k}} + (1-\mathbb{1}_{k\ell,i}) \widetilde{p}^{[m]}_{\ell k,i-1} \tag{60}$$

$$\widetilde{p}^{[f]}_{\ell k,i} = \mathbb{1}_{k\ell,i} \frac{\sum_{n=1}^{\xi_{k\ell,i}} \gamma_k^{\xi_{k\ell,i}-n} \mathbb{1}_{\{\hat{\boldsymbol{a}}^{[n]}=-1, \hat{\boldsymbol{a}}^{[n]} \in \widehat{\mathcal{H}}^{[1]}_{k\ell,i}\}} + \gamma_k^{\xi_{k\ell,i}} h_{1,k}}{\sum_{n=1}^{\xi_{k\ell,i}} \gamma_k^{\xi_{k\ell,i}-n} + \gamma_k^{\xi_{k\ell,i}} h_{2,k}} + (1-\mathbb{1}_{k\ell,i}) \widetilde{p}^{[f]}_{\ell k,i-1} \tag{61}$$

Note that if k and ℓ are not paired at time i, then the empirical probabilities $\widetilde{p}^{[m]}_{\ell k,i}, \widetilde{p}^{[f]}_{\ell k,i}$ are not updated. The reason for the use of the discounting rule is that the agents weigh more their recent observations. Moreover, in order to prevent $\widetilde{p}^{[m]}_{\ell k,n}, \widetilde{p}^{[f]}_{\ell k,n}$ from taking the extreme values (i.e., 1 or 0) during the first iterations when the observations are few we use the constants $h_{1,k}, h_{2,k}$. Since, initially, no information is available to the agents and they are completely uncertain about the detection performance of the other agents, $\widetilde{p}^{[m]}_{\ell k,1}, \widetilde{p}^{[f]}_{\ell k,1}$ should be equal to 0.5 and thus we set $h_{1,k}=0.5, h_{2,k}=1$. Note that due to the discounting, this initial bias vanishes as time proceeds.

C. Reward Prediction and Adaptive Threshold Selection

A strategic agent has to predict the benefit $\mathbf{b}^{[0]}_{k,i}$ in order to decide which action to perform using the best–response rule (i.e., (46) or (47)). This is performed by following the same approach as in [15]. More specifically, we make the assumption that the agents are *risk–taking*, in the sense that they assume that the benefit acquired by the estimate sent by the paired agent will yield the maximum possible estimation improvement. In this way, for an agent k paired with agent ℓ, the predicted benefit, denoted as $\bar{\mathbf{b}}^{[0]}_{k,i}$, will be the result of the following maximization problem

$$\begin{aligned}\bar{\mathbf{b}}^{[0]}_{k,i} &= \max_{\boldsymbol{\psi}_{\ell,i}} \mathbf{b}^{[0]}_{k,i} = \max_{\boldsymbol{\psi}_{\ell,i}} \big\{ \big[1 - p^{[f]}_{k\ell,i}\big] \\ &\times \big[\mathbb{E}\big\{||w^o - \boldsymbol{\psi}_{k,i}||^2_{R_{u,k}} | \boldsymbol{\phi}_{k,i-1}\big\} \\ &- \mathbb{E}\big\{||w^o - z_{kk}\boldsymbol{\psi}_{k,i} - (1-z_{kk})\boldsymbol{\psi}_{\ell,i}||^2_{R_{u,k}} | \boldsymbol{\phi}_{k,i-1}\big\}\big]\big\}\end{aligned} \tag{62}$$

We observe from (14) that given $\boldsymbol{\psi_{\ell,i}}$, an agent k can determine the value of $p^{[f]}_{k\ell,i}$ because agent k has access to $\widehat{\boldsymbol{\psi}}_{k,i}$ and $s_{k\ell,i}$. More specifically for a given $\boldsymbol{\psi}_{\ell,i}$, it is

$$p^{[f]}_{k\ell,i} = \mathbb{1}_{\{\|\widehat{\boldsymbol{\psi}}_{k,i} - \boldsymbol{\psi}_{\ell,i}\|^2 \le s_{k\ell,i}\}}$$

Thus, the assumption of risk-taking agents implies that an agent k assumes that it will detect the acquired $\boldsymbol{\psi}_{\ell,i}$, which maximizes the estimation benefit, correctly as non-erroneous estimate. This can be done by letting each agent k to adjust the detection threshold $s_{k\ell,i}$ adaptively in order to acquire the maximum possible benefit. Then, the maximization problem (62) is re-formulated as

$$\begin{aligned}\bar{\mathbf{b}}^{[0]}_{k,i} &= \max_{\boldsymbol{\psi}_{\ell,i}, s_{k\ell,i}} \mathbf{b}^{[0]}_{k,i} \\ &= \max_{\boldsymbol{\psi}_{\ell,i}, s_{k\ell,i}} \big\{\big(1 - \mathbb{1}_{\{\|\widehat{\boldsymbol{\psi}}_{k,i} - \boldsymbol{\psi}_{\ell,i}\|^2 \le s_{k\ell,i}\}}\big) \\ &\times \big(\mathbb{E}\big\{||w^o - \boldsymbol{\psi}_{k,i}||^2_{R_{u,k}} | \boldsymbol{\phi}_{k,i-1}\big\} \\ &- \mathbb{E}\big\{||w^o - z_{kk}\boldsymbol{\psi}_{k,i} - (1-z_{kk})\boldsymbol{\psi}_{\ell,i}||^2_{R_{u,k}} | \boldsymbol{\phi}_{k,i-1}\big\}\big)\big\}\end{aligned} \tag{63}$$

The solution to the above problem is $\boldsymbol{\psi}^*_{\ell,i} = \frac{w^o - z_{kk}\boldsymbol{\psi}_{k,i}}{1-z_{kk}}$ and any $s^*_{kl,i} \ge \|\widehat{\boldsymbol{\psi}}_{k,i} - \boldsymbol{\psi}^*_{\ell,i}\|^2$. We select $s^*_{kl,i} = \|\widehat{\boldsymbol{\psi}}_{k,i} - \boldsymbol{\psi}^*_{\ell,i}\|^2$. Thus it is

$$\bar{\mathbf{b}}^{[0]}_{k,i} = \mathbb{E}\big\{||w^o - \boldsymbol{\psi}_{k,i}||^2_{R_{u,k}} | \boldsymbol{\phi}_{k,i-1}\big\} \tag{64}$$

Still $\bar{\mathbf{b}}^{[0]}_{k,i}$ cannot be calculated, as w^o, $R_{u,k}$ are unknown. To proceed we use the approximation presented in [15]

$$\bar{\mathbf{b}}^{[0]}_{k,i} \approx (1-\mu_k||\boldsymbol{u}_{k,i}||^2)^2 \big(\widehat{\boldsymbol{\phi}}_{k,i-1}\big)^H \boldsymbol{u}^H_{k,i}\boldsymbol{u}_{k,i}\widehat{\boldsymbol{\phi}}_{k,i-1} \tag{65}$$

where $R_{u,k}$ was replaced with the instantaneous approximation $R_{u,k} \approx \boldsymbol{u}^H_{k,i}\boldsymbol{u}_{k,i}$ and $\widehat{\boldsymbol{\phi}}_{k,i-1}$ stands for an approximation of the estimation error $w^0 - \boldsymbol{\phi}_{k,i-1}$, acquired by using a moving average filter as

$$\bar{\boldsymbol{w}}_{k,i} = (1-\alpha)\bar{\boldsymbol{w}}_{k,i-1} + \alpha\boldsymbol{\psi}_{k,i} \tag{66}$$

$$\widehat{\boldsymbol{\phi}}_{k,i-1} \approx \bar{\boldsymbol{w}}_{k,i} - \boldsymbol{\phi}_{k,i-1} \tag{67}$$

where $\alpha \in (0,1)$ is a forgetting factor. For ease of reference, the commonly used notations are summarized in Table II.

D. Agents' Interactions and Learning Algorithm

Based on the steps outlined above, we arrive at the diffusion-based secure LMS algorithm (DS-LMS).

Remark 5: We note that the algorithm is scalable, as the time complexity of the DS–LMS executed by an agent k (after the random–matching process has been completed) is $O(T_h \cdot \max\{|\mathcal{N}^-_k|, L\})$, where T_h is the time horizon of the execution of the algorithm, L is the size of the unknown vector of parameters w_0 and $|\mathcal{N}^-_k|$ is the number of neighbors of agent k.

V. Simulations

In this section, we present simulations showing the interactions of the nodes when they use the DS-LMS algorithm to perform their decision–making process. The network consists of 20 nodes with 8 malicious, 8 selfish and 4 honest nodes. The unknown parameter vector has size $L=10$ and it is $w^0 = [0.1, 0.2, 0.25, 0.05, 0.15, 0.16, 0.22, 0.06, 0.12, 0.19]^T$. The noise variance for each node is shown in first sub-figure in Fig. 2 and $R_{u,k} = R_u$ is uniform and diagonal with entries shown in the second sub-figure in Fig. 2. The step–sizes are $\mu_k = 0.01$ and combination weights are $z_{kk} = 0.5$, for every node $k \in \mathcal{N}$. We consider same transmission cost $c_k = c$ and same illegal gains $m_k = m$ for all selfish and malicious agents, respectively. The discount factor is $\lambda_k = 0.99$ and

TABLE II
LIST OF VARIABLES

Variable	Meaning
$\mathcal{N}$	Set of agents in the network.
N	Number of agents in the network.
$\mathcal{N}_k$	Set of neighboring agents of k.
$\mathcal{N}_k^-$	Set of neighboring agents of k without k.
τ_k	Type of k.
w_0	Unknown vector of parameters.
$\boldsymbol{u}_{k,i}$	Regression vector of k at time i.
$\boldsymbol{v}_{k,i}$	Noise vector of k at time i.
$\sigma_{v,k}^2$	Noise variance of k.
$\boldsymbol{d}_{k,i}$	Observation data of k at time i.
$R_{u,k}$	Covariance matrix of k.
$\boldsymbol{\psi}_{k,i}$	Intermediate estimate of LMS of k at time i.
$\widehat{\boldsymbol{\psi}}_{k,i}$	Intermediate estimate of *standalone* LMS of k at time i.
$\boldsymbol{\phi}_{k,i}$	Local estimate of k at time i.
μ_k	Step-size parameter of k.
$z_{k\ell}$	Combination coefficient of k towards $\ell \in \mathcal{N}_k$.
$\boldsymbol{\eta}_{k\ell}$	Noise added of malicious agent k towards ℓ.
π_{H}	Honest policy.
$\boldsymbol{a}_{k\ell,i}$	Action selected by agent k towards ℓ at time i.
$\hat{\boldsymbol{a}}_{k\ell,i}$	Detection outcome of action $\boldsymbol{a}_{k\ell,i}$ by ℓ at time i.
$p_{k\ell,i}^{[m]}$	Probability of miss-detection of agent k interacting with ℓ at time i.
$p_{k\ell,i}^{[f]}$	Probability of false alarm of agent k interacting with ℓ at time i.
$\widetilde{p}_{\ell k,i}^{[m]}$	Probability of miss-detection of ℓ interacting with k at time i as it is estimated by k.
$\widetilde{p}_{\ell k,i}^{[f]}$	Probability of false alarm of ℓ interacting with k at time i as it is estimated by k.
c_k	Transmission cost of k
m_k	Illegal gain of k
$s_{k\ell,i}$	Detection threshold set by k interacting with ℓ at time i.
$\boldsymbol{b}_{k,i}^{[0]}$	Estimation benefit of k when paired agent selects $\boldsymbol{a}_{\ell k,i}=1$ instead of $\boldsymbol{a}_{\ell k,i}=0$.
$\boldsymbol{b}_{k,i}^{[-1]}$	Estimation benefit of k when paired agent selects $\boldsymbol{a}_{\ell k,i}=1$ instead of $\boldsymbol{a}_{\ell k,i}=-1$.
$\bar{\boldsymbol{b}}_{k,i}^{[0]}$	Maximum estimation benefit of k when paired agent selects $\boldsymbol{a}_{\ell k,i}=1$ instead of $\boldsymbol{a}_{\ell k,i}=0$.
$\boldsymbol{\theta}_{k\ell,i}$	Reputation vector that ℓ keeps about k at time i.

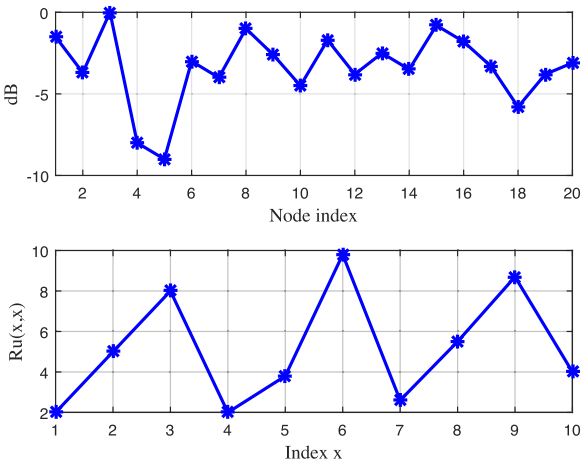


Fig. 2. Noise variance of nodes and diagonal entries of R_u.

Diffusion-based Secure LMS (DS-LMS).

- For each node k, start with any initial guesses $\boldsymbol{\phi}_{k,0}, \widehat{\boldsymbol{\psi}}_{k,0}, \bar{\boldsymbol{w}}_{k,0}$ and set $\boldsymbol{\theta}_{\ell k,0}=[1,0,0]^T, \widetilde{p}_{\ell k,0}^{[m]}=\widetilde{p}_{\ell k,0}^{[f]}=0.5$ for every $\ell \in \mathcal{N}_k^-$.
- At each time $i \geq 1$, generate $\{\mathbb{1}_{k\ell,i}\}$ for all k, ℓ and for each node $k \in \mathcal{N}$ execute
  - Calculate $\boldsymbol{\psi}_{k,i}$ using (2) (*Adaptation step*).
  - Calculate $\widehat{\boldsymbol{\psi}}_{k,i}$ using (11) (*Adaptation step of stand–alone LMS*).
  - Calculate $\widetilde{p}_{\ell k,i}^{[m]}, \widetilde{p}_{\ell k,i}^{[f]}$ and set $p_{\ell k,i}^{[m]}=\widetilde{p}_{\ell k,i}^{[m]}, p_{\ell k,i}^{[f]}=\widetilde{p}_{\ell k,i}^{[f]}$ for every $\ell \in \mathcal{N}_k^-$.
  - Calculate $\bar{\boldsymbol{w}}_{k,i}, \widehat{\boldsymbol{\phi}}_{k,i-1}$ via (66), (67), respectively.
  - Calculate $s_{k\ell,i}=\|\widehat{\boldsymbol{\psi}}_{k,i}-\frac{\bar{\boldsymbol{w}}_{k,i}-z_{kk}\boldsymbol{\psi}_{k,i}}{1-z_{kk}}\|^2$, calculate $\bar{\mathbf{b}}_{k,i}^{[0]}$ and set $\mathbf{b}_{k,i}^{[0]}=\bar{\mathbf{b}}_{k,i}^{[0]}, \ell \in \mathcal{N}_k^- : \mathbb{1}_{k\ell,i}=1$.
  - *Selection of action* $\boldsymbol{a}_{k\ell,i}, \ell \in \mathcal{N}_k^- : \mathbb{1}_{k\ell,i}=1$. Apply the honest policy π_{H} (17) if $\tau_k=\mathsf{H}$ and the best–response rule provided in (46), (47) if $\tau_k=\mathsf{S}$, M, respectively.
  - *Information sharing*. Transmit $\boldsymbol{\varphi}_{k\ell,i}$ to ℓ according to (10), $\ell \in \mathcal{N}_k^- : \mathbb{1}_{k\ell,i}=1$.
  - *Detection of action* $\hat{\boldsymbol{a}}_{\ell k,i}$ by applying (12), $\ell \in \mathcal{N}_k^- : \mathbb{1}_{k\ell,i}=1$.
  - *Combination step*. Execute (15) to generate the updated estimate $\boldsymbol{\phi}_{k,i}$.
  - *Update of reputation vector*. Update $\boldsymbol{\theta}_{\ell k,i+1}$ via (16), if $\mathbb{1}_{k\ell,i}=1, \ell \in \mathcal{N}_k^-$ and $\boldsymbol{\theta}_{\ell k,i+1}=\boldsymbol{\theta}_{\ell k,i}$ for all $\ell \in \mathcal{N}_k^- : \mathbb{1}_{k\ell,i}=0$.
  - *Exchange detected action* $\hat{\boldsymbol{a}}_{\ell k,i}$ with agent $\ell \in \mathcal{N}_k^- : \mathbb{1}_{k\ell,i}=1$.

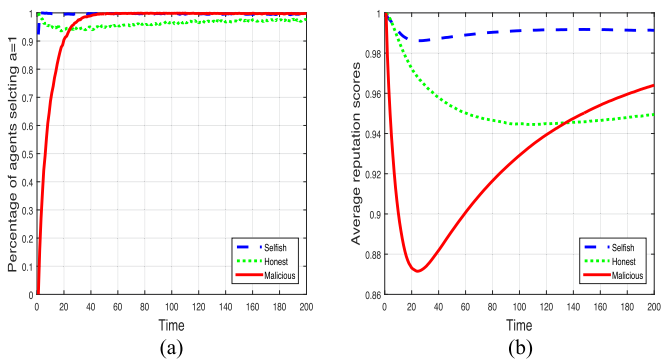


Fig. 3. (a) Action selection, (b) evolution of reputations, for small c and m.

$[r_k^{[-1]}, r_k^{[0]}, r_k^{[1]}]=[0.85, 0.9, 0.95]$ for every $k \in \mathcal{N}$. The forgetting factor in (66) is $\alpha=0.01$ and in (60), (61) $\gamma_k=0.5$ for all k. The noise added by the malicious agents $\boldsymbol{\eta}_{k\ell}$ when they choose to do so, is generated initially by a Gaussian distribution with mean value and standard deviation 0.5 and 0.05, respectively, for every node pair $k \in \mathcal{N}$, $\ell \in \mathcal{N}_k^-$. All experiments are performed for 200 time instances and 1000 Monte Carlo simulations were performed, except for Fig. 9 where 100 Monte Carlo simulations were performed. We

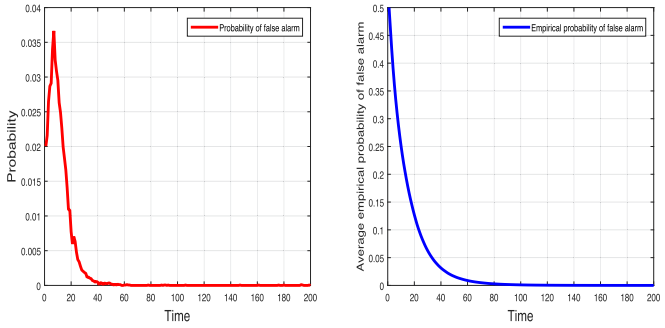


Fig. 4. (a) (Actual) Probability of false alarm, (b) average empirical probability of false alarm, for small c and m.

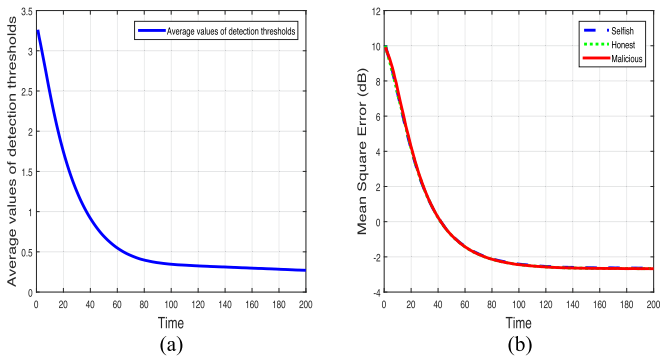


Fig. 5. (a) Average values of detection thresholds, (b) MSE, for small c and m.

generate a random connected graph at every iteration of the Monte Carlo simulations in order not to give rise to issues like the topology and connectivity of the different types of nodes in the network. Every node has 2 to 8 neighbors and this number is randomly generated.

Figs. 3–5 refer to the case with small transmission cost and illegal gains, $c = m = 10^{-7}$. In Fig. 3(a) the percentage of each type of agent choosing to cooperate (i.e., $\boldsymbol{a} = 1$) is shown. We observe that both strategic types of agents choose cooperation as the optimal choice at a high rate. The percentage of the honest nodes that choose to cooperate is high as well, because as selfish and malicious behave cooperatively, their average reputation values (reputation values of honest behavior, i.e., $\boldsymbol{\theta}^{[\mathsf{H}]}$) are large as can be seen in Fig. 3(b). The drop in reputations during the first time instances is due to the false alarm incidents occurring during the initial phase of interactions as can be seen in Fig. 4(a). After that phase, as the probability of false alarm drops, the reputation values recover. In Fig. 4(b) the average empirical probability of false alarm calculated by the agents via (61) is depicted and in Fig. 5(a) the average values of detection thresholds $s_{k\ell,n}$ set by the agents during the process are shown. The probability of miss-detection is not shown in this experiment as the number of malicious actions is negligible. Finally, the MSE achieved is depicted in Fig. 5(b).

Figs. 6–8 refer to the case where the transmission cost and illegal gain are large with values $c = 0.5$ and $m = 50$. In contrast with the previous case, the percentage of selfish agents that

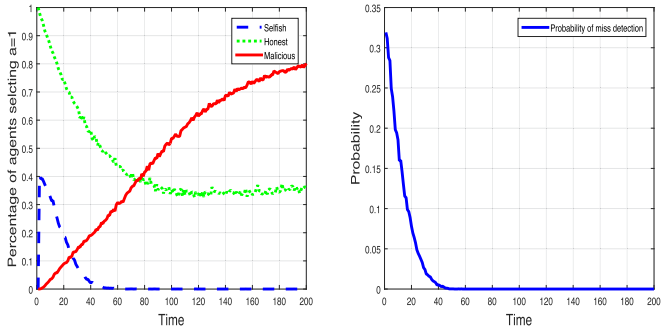


Fig. 6. (a) Action selection, (b) (actual) probability of miss–detection, for large c and m.

Fig. 7. (a) Average empirical probability of miss-detection, (b) evolution of reputations, for large c and m.

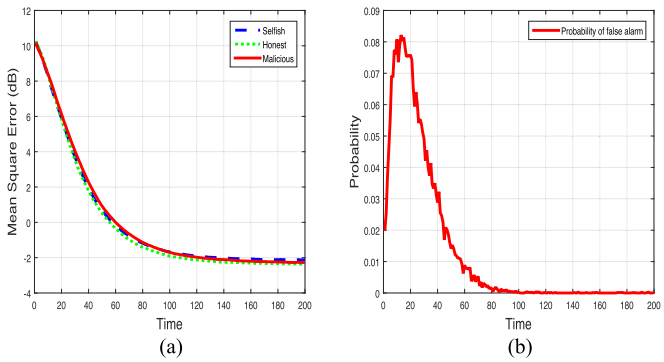


Fig. 8. (a) MSE, (b) (actual) probability of false alarm, for large c and m.

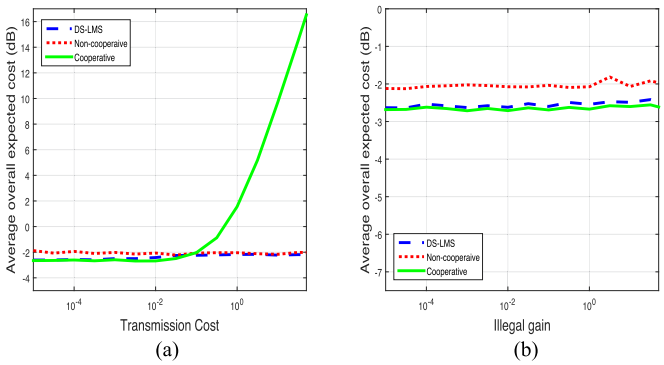


Fig. 9. Average overall costs for selfish nodes (a) and malicious (b).

choose to cooperate is small as it is shown in Fig. 6(a) where initially as the estimation benefits are high they cooperate up to a some point. Malicious behavior though, is different. The percentage of malicious agents that choose to cooperate is initially small, as expected, due to the high values of expected illegal gain, but gradually as the probability of miss–detection drops [see Fig. 6(b)], the estimated (via (60)) empirical probability of miss-detection drops as well [see Fig. 7(a)]. Because of that, their expected illegal gain becomes very low as time progresses and the malicious agents have no more incentives to misbehave (see (47)) and choose to cooperate instead. The average reputation values of honesty (i.e., $\boldsymbol{\theta}^{[\mathsf{H}]}$) are shown in Fig. 7(b) where we observe that as malicious do not adhere to misbehaving eventually, their reputation recovers. In Fig. 8(a) the MSE performance is depicted and we observe that it is worse than the MSE achieved in the case with small transmission costs and illegal gains for all types of agents because the agents are less cooperative. However, MSE converges because the detection scheme exhibits small number of miss-detections as time progresses [see Fig. 6(b)]. The probability of false alarm similarly drops as time progresses as shown in Fig. 8(b).

In Fig. 9(a) [resp. Fig. 9(b)] the average overall expected cost of selfish (resp. malicious) agents versus the transmission cost (resp. illegal gains) at the steady-state is plotted. The overall expected cost is associated with the estimation accuracy in terms of instantaneous evaluation of the estimation performance of agent k (based on the estimate k possesses at the beginning of time i - i.e., $\boldsymbol{\phi}_{k,i-1}$) and the expected cost from transmission action at time i (i.e., $\boldsymbol{a}_{k\ell,i}$). This average overall cost for each type of agents is given by

$$C_{\tau}^{\mathrm{tot}} = \frac{1}{N_{\tau}} \sum_{k=1}^{N} \mathbb{1}_{\{\tau_k=\tau\}} \cdot \Big(\zeta_k(\boldsymbol{\phi}_{k,i-1}) - \delta_k(a_{k\ell,i}, \tau_k) \Big) \quad (68)$$

where N_{τ} is the number of agents of type τ. We employ cost instead of reward to express values in dB. The three curves in Fig. 9 refer to the average overall expected cost when agents always cooperate (all agents select $\boldsymbol{a} = 1$), never cooperate (honest and selfish agents select $\boldsymbol{a} = 0$ and malicious $\boldsymbol{a} = -1$) and when they follow the DS–LMS algorithm to select actions at every time instant. Using the proposed algorithm, we observe that both selfish and malicious agents track the scheme (non–cooperative or cooperative) that minimizes their corresponding type-specific average overall cost for the various values of transmission cost and illegal gains, respectively, at the steady state. We note that malicious agents choose to cooperate at the steady–state because the probability of miss-detection becomes negligible at steady-state (as explained earlier) and thus the expected illegal gains are very small.

Finally, the impact of different distributions of agent types on the behavior of honest, selfish and malicious agents is highlighted in Figs. 10(a), (b) and 11, respectively. The transmission cost and illegal gains are set to $c = 0.1$ and $m = 50$. The reason why big transmission costs and illegal gains are selected is because for small values of these parameters all agents have almost the same behavior (i.e., they all cooperate) through time, as explained earlier [see Fig. 3(a)]. Three different distributions

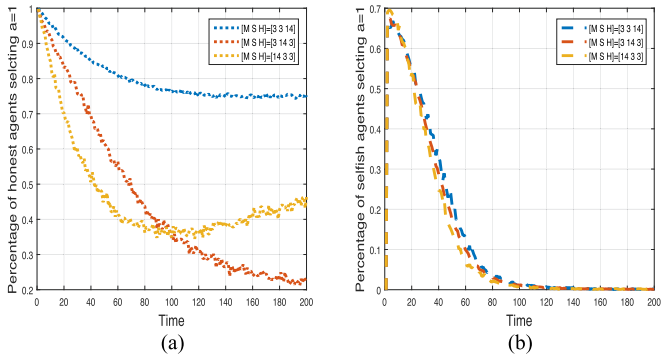


Fig. 10. (a) Honest agents' behavior, (b) selfish agents' behavior for different types distribution.

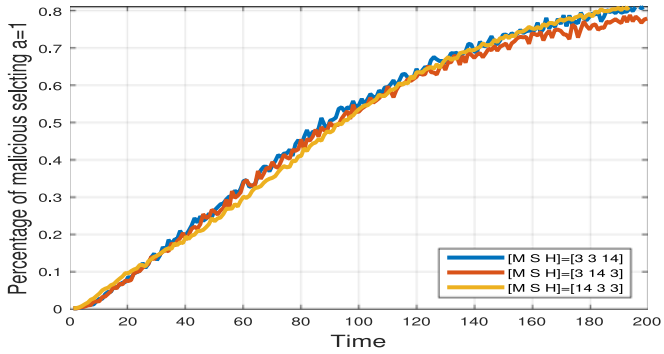


Fig. 11. Malicious agents' behavior for different types distribution.

of agent types are considered, one with the majority of agents being honest (i.e., $[N_{\mathsf{m}}\ N_{\mathsf{S}}\ N_{\mathsf{H}}] = [3\ 3\ 14]$), one with the majority being selfish (i.e., $[N_{\mathsf{M}}\ N_{\mathsf{S}}\ N_{\mathsf{H}}] = [3\ 14\ 3]$) and one with the majority being malicious (i.e., $[N_{\mathsf{M}}\ N_{\mathsf{S}}\ N_{\mathsf{H}}] = [14\ 3\ 3]$). In Fig. 10(a) we observe that honest nodes cooperate more when most agents are honest, as expected. Moreover, the cooperation rate is higher in the distribution with more selfish compared to the one with more malicious agents during the initial stages. This happens because malicious agents' reputation drops more than the reputation of selfish nodes as the punishment of malicious behavior is more severe. The cooperation rate in the distribution with more malicious agents goes up at the later stages because malicious adhere to truth-telling [as explained in Fig. 6(a)] and as a result their reputation raises and honest agents are more likely to cooperate with them. Regarding selfish agents' behavior, we observe in Fig. 10(b) that cooperation rate is higher when honest agents constitute the majority and lower when the malicious agents are the majority during the initial stages. At the later stages, selfish agents do not cooperate for large transmission costs [as explained in Fig. 6(a)] regardless of the distribution. Finally, in Fig. 11 we observe that malicious agents cooperate less when the majority of agents are malicious during the initial stages, but later cooperation rate increases as the probability of miss-detection drops and truth-telling dominates for malicious agents [as explained in Fig. 6(a)].

One important remark that stems from the examination of the agents behavior is the difference between selfish and malicious agents. If the detection scheme achieves very small prob-

ability of miss-detection and malicious agents become aware of that, attacking is not beneficial for malicious agents and dissemination of falsified information can be thwarted for malicious, but rational, decision-makers as the time progresses.

VI. Conclusion

In this paper the decision–making process of the two basic misbehaving types of agents found in wireless networks, involved in a parameter estimation task was studied. We derived an algorithm for secure parameter estimation that combines a suitable detection scheme, along with a reputation protocol in order to combat adversarial falsified data injection attacks and provide incentives to misbehaving agents for adhering to truth-telling cooperation. It is shown that truth-telling cooperation can emerge under suitable conditions. As future work, we aim at investigating other types of malicious behavior patterns, along with the appropriate design of suitable detection and reputation models, as well as the case of information exchange among agents over noisy links.

Appendix

A selfish agent k paired with a neighbor $\ell \in \mathcal{N}_k^-$ at time i selects the action $a_{k\ell,i}=1$ as the optimal one instead of $a_{k\ell,i}=0$ if

$$J_k(a_{k\ell,i}=1,\tau_k=\mathsf{S}) > J_k(a_{k\ell,i}=0,\tau_k=\mathsf{S}) \tag{69}$$

By utilizing (43) and (28) we can re-write (69) as follows

$$\begin{aligned}&-\sum_{n=i}^{\infty}\lambda_k^{n-i}c_k\\&+\sum_{n=i+1}^{\infty}\lambda_k^{n-i}\mathbb{E}\big\{\beta_k(\boldsymbol{a}_{\ell k,n})|\boldsymbol{a}_{k\ell,i}=1,\mathcal{M}_{k,i}\big\}\\&-\sum_{n=i+1}^{\infty}\lambda_k^{n-i}\mathbb{E}\big\{\beta_k(\boldsymbol{a}_{\ell k,n})|\boldsymbol{a}_{k\ell,i}=0,\mathcal{M}_{k,i}\big\}>0\end{aligned} \tag{70}$$

Note that agent k is confined to use stationary strategies (see (36)). We use (42) to get

$$\begin{aligned}&\mathbb{E}\big\{\beta_k(\boldsymbol{a}_{\ell k,n})|\boldsymbol{a}_{k\ell,i}=j,\mathcal{M}_{k,i}\big\}\\&=\sum_{a_{\ell k,n}}\mathbb{E}\big\{\beta_k(a_{\ell k,n})|\boldsymbol{a}_{k\ell,i}=j,\mathcal{M}_{k,i}\big\}B(a_{\ell k,n})\end{aligned} \tag{71}$$

Now, let $\boldsymbol{\theta}_{k\ell,n}^{[\mathsf{H}|j]}$ denote the expected value of $\boldsymbol{\theta}_{k\ell,n}^{[\mathsf{H}]}$ at time n when agent k selects the action $\boldsymbol{a}_{k\ell,i}=j$. Then, (42) can be employed along with assumption $\boldsymbol{\theta}_{\ell k,n}=\boldsymbol{\theta}_{\ell k,i}$ to get

$$B(\boldsymbol{a}_{\ell k,n}=1)=\boldsymbol{\theta}_{k\ell,n}^{[\mathsf{H}|j]}\boldsymbol{\theta}_{\ell k,i}^{[\mathsf{H}]} \tag{72}$$

$$B(\boldsymbol{a}_{\ell k,n}=0)=\left(1-\boldsymbol{\theta}_{k\ell,n}^{[\mathsf{H}|j]}\right)\boldsymbol{\theta}_{\ell k,i}^{[\mathsf{H}]}+\boldsymbol{\theta}_{\ell k,i}^{[\mathsf{S}]} \tag{73}$$

$$B(\boldsymbol{a}_{\ell k,n}=-1)=\boldsymbol{\theta}_{\ell k,i}^{[\mathsf{M}]} \tag{74}$$

The update of $\boldsymbol{\theta}_{k\ell,i}=[\boldsymbol{\theta}_{k\ell,i}^{[\mathsf{H}]},\boldsymbol{\theta}_{k\ell,i}^{[\mathsf{S}]},\boldsymbol{\theta}_{k\ell,i}^{[\mathsf{M}]}]^T$ depends on the detected action $\hat{\boldsymbol{a}}_{k\ell,i}$. Thus, for a given action $\boldsymbol{a}_{k\ell,i}=j$ and given $\boldsymbol{\theta}_{k\ell,i}$, the expected value of the reputation vector $\boldsymbol{\theta}_{k\ell,i+1}$ that ℓ keeps about k at time instant $i+1$ is

$$\begin{aligned}\boldsymbol{\theta}_{k\ell,i+1}^{[j]}=&\sum_{\hat{a}_{k\ell,i}}\Pr(\hat{a}_{k\ell,i}|\boldsymbol{a}_{k\ell,i}=j)\\&\times\left[\boldsymbol{\theta}_{k\ell,i}r_k^{[\hat{a}_{k\ell,i}]}+e^{[\hat{a}_{k\ell,i}]}(1-r_k^{[\hat{a}_{k\ell,i}]})\right]\end{aligned} \tag{75}$$

Thus, by utilizing assumption $\mathbb{1}_{k\ell,n}=\mathbb{1}_{k\ell,i}$ for every $n\geq i$ along with assumptions (36), (40), (41) and the reputation update rule (16), we can compute $\boldsymbol{\theta}_{k\ell,n}^{[\mathsf{H}|0]}$ and $\boldsymbol{\theta}_{k\ell,n}^{[\mathsf{H}|1]}$ for a time instant $n\geq i$ as follows

$$\boldsymbol{\theta}_{k\ell,n}^{[\mathsf{H}|1]}=A_{k,i}^{n-i}\boldsymbol{\theta}_{k\ell,i}^{[\mathsf{H}]}+\Gamma_{k,i}\sum_{p=0}^{n-i-1}A_{k,i}^{p}, \tag{76}$$

$$\boldsymbol{\theta}_{k\ell,n}^{[\mathsf{H}|0]}=\boldsymbol{\theta}_{k\ell,i}^{[\mathsf{H}]}(r_\ell^{[0]})^{n-i}, \tag{77}$$

where

$$A_{k,i}=(1-p_{\ell k,i}^{[f]})r_\ell^{[1]}+p_{\ell k,i}^{[f]}r_\ell^{[-1]} \tag{78}$$

$$\Gamma_{k,i}=(1-p_{\ell k,i}^{[f]})(1-r_\ell^{[1]}) \tag{79}$$

Then (70), using (71)–(74), becomes

$$\begin{aligned}&-\sum_{n=i}^{\infty}\lambda_k^{n-i}c_k\\&\quad+\sum_{n=i+1}^{\infty}\lambda_k^{n-i}\mathbf{b}_{k,i}^{[0]}\boldsymbol{\theta}_{\ell k,i}^{[\mathsf{H}]}\left(\boldsymbol{\theta}_{k\ell,n}^{[\mathsf{H}|1]}-\boldsymbol{\theta}_{k\ell,n}^{[\mathsf{H}|0]}\right)>0\end{aligned} \tag{80}$$

Hence, substituting (76) and (77) into (80) we get

$$\begin{aligned}&-\frac{c_k}{1-\lambda_k}+\mathbf{b}_{k,i}^{[0]}\boldsymbol{\theta}_{\ell k,i}^{[\mathsf{H}]}\sum_{n=i+1}^{\infty}\lambda_k^{n-i}\\&\times\left[\boldsymbol{\theta}_{k\ell,i}^{[\mathsf{H}]}A_{k,i}^{n-i}+\Gamma_{k,i}\sum_{p=0}^{n-i-1}A_{k,i}^{p}-\boldsymbol{\theta}_{k\ell,i}^{[\mathsf{H}]}(r_\ell^{[0]})^{n-i}\right]>0\end{aligned} \tag{81}$$

We observe that $0<A_{k,i}<1$ and $0\leq\Gamma_{k,i}<1$ as it is $r_\ell^{[\hat{\boldsymbol{a}}]}\in(0,1)$ for every detected action $\hat{\boldsymbol{a}}$. Furthermore, it is $\sum_{n=i+1}^{\infty}\delta^{n-i}\sum_{p=0}^{n-i-1}d^p=\frac{\delta}{(1-\delta)(1-\delta d)}$ for $0\leq\delta,d<1$. Hence, inequality (81) yields

$$\mathbf{b}_{k,i}^{[0]}\boldsymbol{x}_{k,i}^{[\mathsf{S}]}\boldsymbol{\theta}_{\ell k,i}^{[\mathsf{H}]}>c_k \tag{82}$$

where $\boldsymbol{x}_{k,i}^{[\mathsf{S}]}$ is given by (48).

Similarly, a malicious agent k paired with a neighbor $\ell\in\mathcal{N}_k^-$ at time i will select the action $a_{k\ell,i}=1$ as the optimal one instead of $a_{k\ell,i}=-1$ if

$$J_k(a_{k\ell,i}=1,\tau_k=\mathsf{M})>J_k(a_{k\ell,i}=-1,\tau_k=\mathsf{M}) \tag{83}$$

By utilizing (28) and (43), we can re-write (83) as follows

$$-\sum_{n=i}^{\infty}\lambda_k^{n-i}p_{\ell k,n}^{[m]}m_k + \sum_{n=i+1}^{\infty}\lambda_k^{n-i}\mathbb{E}\{\beta_k(\boldsymbol{a}_{\ell k,n})|\boldsymbol{a}_{k\ell,i}=1,\mathcal{M}_{k,i}\} - \sum_{n=i+1}^{\infty}\lambda_k^{n-i}\mathbb{E}\{\beta_k(\boldsymbol{a}_{\ell k,n})|\boldsymbol{a}_{k\ell,i}=-1,\mathcal{M}_{k,i}\}>0 \quad (84)$$

Eqs. (71)–(76) are the same for the malicious case, as well. Moreover, under assumptions (36), (38), (39), (41) and the reputation update rule (16), the value of $\boldsymbol{\theta}_{k\ell,n}^{[\mathsf{H}|-1]}$ for a time instant $n \geq i$ is

$$\boldsymbol{\theta}_{k\ell,n}^{[\mathsf{H}|-1]} = B_{k,i}^{n-i}\boldsymbol{\theta}_{k\ell,i}^{[\mathsf{H}]} + Z_{k,i}\sum_{q=0}^{n-i-1}B_{k,i}^{q}, \quad (85)$$

where

$$B_{k,i} = r_{\ell}^{[-1]} + \left(r_{\ell}^{[1]} - r_{\ell}^{[-1]}\right)p_{\ell k,i}^{[m]}, \quad (86)$$

$$Z_{k,i} = \left(1 - r_{\ell}^{[1]}\right)p_{\ell k,i}^{[m]}. \quad (87)$$

It is $0 < B_{k,i} < 1$ and $0 \leq Z_{k,i} < 1$ as $r_{\ell}^{[\hat{\boldsymbol{a}}]} \in (0,1)$ for every detected action $\hat{\boldsymbol{a}}$. Then (84), using (71)–(74), becomes

$$-\sum_{n=i}^{\infty}\lambda_k^{n-i}p_{\ell k,i}^{[m]}m_k + \sum_{n=i+1}^{\infty}\lambda_k^{n-i}\mathbf{b}_{k,i}^{[0]}\boldsymbol{\theta}_{\ell k,i}^{[\mathsf{H}]} \times\left(\boldsymbol{\theta}_{k\ell,n}^{[\mathsf{H}|1]} - \boldsymbol{\theta}_{k\ell,n}^{[\mathsf{H}|-1]}\right) > 0 \quad (88)$$

Hence, substituting (76) and (85) into (88) we get

$$-\frac{p_{\ell k,i}^{[m]}m_k}{1-\lambda_k} + \mathbf{b}_{k,i}^{[0]}\boldsymbol{\theta}_{\ell k,i}^{[\mathsf{H}]}\Bigg[\boldsymbol{\theta}_{k\ell,i}^{[\mathsf{H}]}\sum_{n=i+1}^{\infty}\lambda_k^{n-i}\left(A_{k,i}^{n-i} - B_{k,i}^{n-i}\right) + \Gamma_{k,i}\sum_{n=i+1}^{\infty}\lambda_k^{n-i}\sum_{p=0}^{n-i-1}A_{k,i}^{p} - Z_{k,i}\sum_{n=i+1}^{\infty}\lambda_k^{n-i}\sum_{p=0}^{n-i-1}B_{k,i}^{p}\Bigg] > 0 \quad (89)$$

Likewise, after some manipulations, (89) yields

$$\mathbf{b}_{k,i}^{[0]}\boldsymbol{x}_{k,i}^{[\mathsf{M}]}\boldsymbol{\theta}_{\ell k,i}^{[\mathsf{H}]} > p_{\ell k,i}^{[m]}m_k \quad (90)$$

where $\boldsymbol{x}_{k,i}^{[\mathsf{M}]}$ is given by (49).

REFERENCES

[1] P. Michiardi and R. Molva, "Simulation-based analysis of security exposures in mobile ad hoc networks," in *Proc. Eur. Wireless Conf.*, Feb. 2002, pp. 15–17.

[2] A. Ahmed, K. A. Bakar, M. I. Chana, K. Haseeb, and A. W. Khan, "A survey on trust based detection and isolation of malicious nodes in ad-hoc and sensor networks," *Front. Comput. Sci.*, vol. 9, no. 2, pp. 280–296, 2015.

[3] M. Momani and S. Challa, "Survey of trust models in different network domains," *Int. J. Ad Hoc Sens. Ubiquitous Comput.*, vol. 1, no. 3, pp. 1–19, 2010.

[4] C. D. Aliprantis, G. Camera, and D. Puzzello, "A random matching theory," *Games Econ. Behav.*, vol. 59, no. 1, pp. 1–16, 2007.

[5] J. Xu and M. Van der Schaar, "Social norm design for information exchange systems with limited observations," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 11, pp. 2126–2135, Dec. 2012.

[6] G. Gigerenzer and R. Selten, *Bounded Rationality: The Adaptive Toolbox*. Cambridge, MA, USA: MIT Press, 2002.

[7] K. Ntemos, N. Kalouptsidis, and N. Kolokotronis, "Managing trust in diffusion adaptive networks with malicious agents," in *Proc. Eur. Signal Process. Conf.*, 2015, pp. 91–95.

[8] H. Park and M. Van der Schaar, "On the impact of bounded rationality in peer-to-peer networks," *IEEE Signal Process. Lett.*, vol. 16, no. 8, pp. 675–678, Aug. 2009.

[9] M. Felegyhazi, J.-P. Hubaux, and L. Buttyan, "Nash equilibria of packet forwarding strategies in wireless ad hoc networks," *IEEE Trans. Mobile Comput.*, vol. 5, no. 5, pp. 463–476, May 2006.

[10] J. J. Jaramillo and R. Srikant, "DARWIN: Distributed and adaptive reputation mechanism for wireless ad-hoc networks," in *Proc. Int. Conf. Mobile Comput. Netw.*, 2007, pp. 87–98.

[11] Y. Chen and K. J. R. Liu, "Indirect reciprocity game modelling for cooperation stimulation in cognitive networks," *IEEE Trans. Commun.*, vol. 59, no. 1, pp. 159–168, Jan. 2011.

[12] Y. Gao, Y. Chen, and K. J. R. Liu, "Cooperation stimulation in cooperative communications: An indirect reciprocity game," in *Proc. IEEE Int. Conf. Commun.*, 2012, pp. 5163–5167.

[13] C. Tang, A. Li, and X. Li, "When reputation enforces evolutionary cooperation in unreliable MANETs," *IEEE Trans. Cybern.*, vol. 45, no. 10, pp. 2190–2201, Oct. 2015.

[14] C.-K. Yu, M. van der Schaar, and A. H. Sayed, "Reputation design for adaptive networks with selfish agents," in *Proc. IEEE 14th Workshop Signal Process. Adv. Wireless Commun.*, 2013, pp. 160–164.

[15] C.-K. Yu, M. van der Schaar, and A. H. Sayed, "Information-sharing over adaptive networks with self-interested agents," *IEEE Trans. Signal Inf. Process. Over Netw.*, vol. 1, no. 1, pp. 2–19, Mar. 2015.

[16] L. Xiao, Y. Chen, W. S. Lin, and K. J. R. Liu, "Indirect reciprocity security game for large-scale wireless networks," *IEEE Trans. Inf. Forensics Secur.*, vol. 7, no. 4, pp. 1368–1380, Aug. 2012.

[17] Q. He, D. Wu, and P. Khosla, "SORI: A secure and objective reputation-based incentive scheme for ad-hoc networks," in *Proc. Wireless Commun. Netw. Conf.*, 2004, pp. 825–830.

[18] S. Buchegger and J.-Y. Le Boudec, "Performance analysis of the CONFIDANT protocol (cooperation of nodes-fairness in dynamic ad-hoc networks)," in *Proc. 3RD ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, 2002, pp. 226–236.

[19] S. Marti *et al.*, "Mitigating routing misbehavior in mobile ad hoc networks." in *Proc. Int. Conf. Mobile Comput. Netw.*, 2000, pp. 255–265.

[20] D. Fudenberg and J. Tirole, *Game Theory*. Cambridge, MA, USA: MIT Press, 1991, p. 393.

[21] N. Nisan *et al.*, Eds. *Algorithmic Game Theory*, vol. 1. Cambridge U.K.: Cambridge Univ. Press, 2007.

[22] G. J. Mailath and L. Samuelson, *Repeated Games and Reputations: Long-Run Relationships*. London, U.K.: Oxford Univ. Press, 2006.

[23] K. Tuyls and G. Weiss, "Multiagent learning: Basics, challenges, and prospects," *AI Mag.*, vol. 33, no. 3, pp. 41–52, 2012.

[24] X. Zhao and A. H. Sayed, "Distributed clustering and learning over networks," *IEEE Trans. Signal Process.*, vol. 63, no. 13, pp. 3285–3300, Jul. 2015.

[25] A. H. Sayed, "Adaptive networks," *Proc. IEEE*, vol. 102, no. 4, pp. 460–497, Apr. 2014.

[26] V. Kekatos and G. B. Giannakis, "Distributed robust power system state estimation," *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 1617–1626, May 2013.

[27] J. Plata-Chaves, A. Bertrand, M. Moonen, S. Theodoridis, and A. M. Zoubir, "Heterogeneous and multitask wireless sensor networks–algorithms, applications, and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 3, pp. 450–465, Apr. 2017.

[28] Y. Shoham and K. Leyton-Brown, *Multiagent Systems: Algorithmic, Game Theoretic and Logical Foundations*. Cambridge, U.K.: Cambridge Univ. Press, 2008.

[29] D. M. Kreps, P. Milgrom, J. Roberts, and R. Wilson, "Rational cooperation in the finitely repeated prisoners' dilemma," *J. Econ. Theory*, vol. 27, no. 2, pp. 245–252, 1982.

[30] L. Busoniu, R. Babuska, and B. De Schutter, "A comprehensive survey of multiagent reinforcement learning," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 38, no. 2, pp. 156–172, Mar. 2008.

[31] M. L. Littman, "Value-function reinforcement learning in Markov games," *J. Cogn. Syst. Res.*, vol. 2, no. 1, pp. 55–66, 2001.

Authors' photographs and biographies not available at the time of publication.