53rd IEEE Conference on Decision and Control
December 15-17, 2014. Los Angeles, California, USA

Decision Fusion with Corrupted Reports in Multi-Sensor Networks: a Game-Theoretic Approach

A. Abrardo, M. Barni, K. Kallas, B. Tondi
Department of Information Engineering and Mathematics, University of Siena
Via Roma 56, 53100 - Siena, ITALY

Abstract— **Decision fusion in adversarial setting is receiving increasing attention due to its relevance in several applications including sensor networks, cognitive radio, social networks, distributed network monitoring. In most cases, a fusion center has to make a decision based on the reports provided by local agents, e.g. the nodes of a multi-sensor network. In this paper, we consider a setup in which the fusion center makes its decision on the status of an observed system by relying on the decisions made by a pool of local nodes and by taking into account the possibility that some nodes maliciously corrupt their reports to induce a decision error. We do so by casting the problem into a game-theoretic framework and looking for the existence of an equilibrium point defining the optimum strategies for the fusion center and the malicious nodes. We analyze two different strategies for the fusion center: a strategy recently introduced by Varshney et al. in a cognitive radio setup and a new approach based on soft identification of malicious nodes. The superior performance of the new decision scheme are demonstrated by resorting to the game-theoretic framework introduced previously.**

I. INTRODUCTION

We address a distributed decision problem in which a fusion center is required to make a decision about the status of an observed system by relying on the information provided by the nodes of a multi-sensor network. Decision fusion must be carried out in an adversarial setting, that is by taking into account the possibility that some of the nodes malevolently alter their reports to induce a decision error. This is a recurrent problem in many situations wherein the nodes may make a profit from a decision error. As an example, let us consider a cognitive radio system in which users cooperate to sense the frequency spectrum to decide whether the spectrum is free thus allowing them to transmit their data. While cooperation among users allows to make a better decision on the status of the frequency spectrum, it is possible that one or more users deliberately alter their measurements to let the system think that the spectrum is busy, when in fact it is not, and use the available spectrum themselves without sharing such a possibility with the other users [1], [2]. Online reputation systems offer another example. Here a fusion center needs to come out with a final decision (or score) about the reputation of an item like a good or a service by relying on the feedback provided by users. Even in this case, it is possible that malevolent users provide a fake feedback to improve or decrease the reputation of the item under inspection [3], [4]. Other examples come from the emerging field of adversarial signal processing as exemplified in [5].

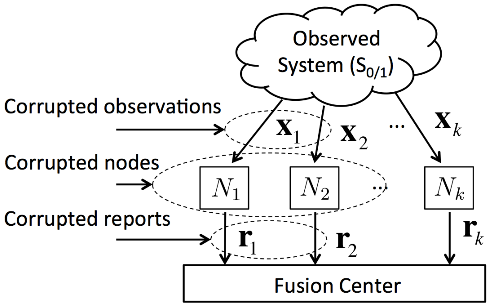


Fig. 1. Decision fusion under adversarial conditions.

A graphical representation of the problem studied in this paper is given in Figure 1. The k nodes of a multi-sensor network observe a system through the vectors $\mathbf{x}_1, \mathbf{x}_2 \dots \mathbf{x}_k$. Based on such vectors, the nodes compute k reports, say $\mathbf{r}_1, \mathbf{r}_2 \dots \mathbf{r}_k$ and send them to a fusion center. The fusion center gathers all the reports and makes a final decision about the status of the observed system. Hereafter, we assume that the system can be only in two states S_0 and S_1. Additionally, we make the simplifying assumption that the reports correspond to local decisions on the system status made by the nodes, i.e. the reports are binary values and $r_i \in \{0, 1\}$ for all i.

The figure depicts three adversarial versions of the above problem. According to the first one, referred to as *decision fusion with corrupted observations*, the adversary (or the adversaries) corrupts the observations seen by the nodes. An asymptotic version of this problem has been studied in [6], however decision fusion with corrupted observations does not fit the scenario addressed in this paper, and hence will not be considered any further. In a second situation (*decision fusion with corrupted nodes*), the fusion center has to tackle with the presence of a number of malevolent nodes, which deliberately alter their reports to induce a decision error. According to a consolidated literature, such nodes are referred to as byzantine nodes or simply Byzantines [7], [8]. Note that a byzantine node can decide to alter its report by relying on its observations of the system, but usually it does not have access to the observations made by the other nodes and their reports[1]. The last case (*decision fusion with corrupted reports*) corresponds to a situation in which the adversary corrupts the reports without having access to the observed sequences. This may correspond to a situation in

[1] When this is the case, we say that the Byzantines are omniscient or that they cooperate among them. In the rest of this paper we will not consider these situations.

978-1-4673-6090-6/14/$31.00 ©2014 IEEE

which the adversary does not control the nodes but only the communication link between the nodes and the fusion center, or to the case of byzantine nodes which, for some reasons, can not observe the data at the input of the node, or decide not exploit such a knowledge (as strange as it may seem, this is a rather common assumption in the analysis of decision fusion in the presence of byzantine nodes [8]).

In this paper we focus on decision fusion with corrupted reports. We adhere to standard terminology in the literature which refers to nodes with corrupted reports as Byzantines. Despite being the simplest kind of attack, the case of corrupted reports contains all the ingredients of more complex situations, hence its analysis is very instructive and already provides interesting insights into the achievable performance of distributed decision fusion under adversarial conditions.

II. PRIOR WORK AND CONTRIBUTION

Data fusion in a distributed decision framework is a classical subject that has received a steadily increasing attention due to its potential use in a wide variety of wireless sensor network applications [9]. Despite such an interest, the adversarial version of the distributed decision problem has been given a relatively limited attention. In [10], the problem of distributed detection in the presence of Byzantines is considered. The analyzed scheme roughly corresponds to the case of decision fusion with corrupted reports depicted in Figure 1. Both cases of scalar and vector reports are considered. Decision fusion is framed into a Neyman-Pearson setup and the asymptotic performance of the system are analyzed as a function of the percentage of corrupted reports. As a result, the percentage of Byzantines making the decision completely unreliable (blind) is determined. In addition to its asymptotic nature, a noticeable difference between the analysis carried out in [10] and the one presented here is that in [10] the Byzantines are assumed to cooperate among them to infer the exact status of the system under analysis. A more general framework is considered in [6], where both the cases of corrupted observations and corrupted nodes are analyzed. Even in this case, the achievable performance are derived under asymptotic conditions and the adversaries are supposed to perfectly know the status of the observed system. A peculiarity of the analysis carried out in [6] is that the system is attacked only when one of the two hypotheses holds, according to a typical Neyman-Pearson setting. Also, the decision problem is framed into a game-theoretic framework and the performance evaluated at the equilibrium point. As opposed to [10], due to the particular characteristics of the setup studied in [6], in some cases it is possible for the decision fusion center to make reliable decisions even when the number of Byzantines exceeds the number of honest nodes. The works that most closely resemble the present paper are [1], [11], which generalize the analysis carried out in [10]. In these works the authors consider the case of corrupted nodes, even if at the end the nodes act as in the case of corrupted reports, since the Byzantines do not take advantage of the knowledge of the observed sequences $\mathbf{x}_i$. As in the present paper, the adversary does not know the true state of the system and the analysis is not limited to the asymptotic case. The Byzantines act by flipping the local decisions made by the corrupted nodes with a certain probability, while the fusion center first tries to understand which are the byzantine nodes and then makes a decision by discarding the suspect nodes. A game-theoretic formulation is also introduced to devise the optimum strategies for the Byzantines and the fusion center. Another paper in which the decision fusion is casted into a game-theoretical framework is [12]. In such paper, the attacker is supposed to know the system status and choose which subset of reports to attack and deliberately corrupt them.

Contribution. With the above ideas in mind, the contribution of this paper is twofold. First of all, we present a novel soft identification strategy whereby the fusion center can isolate the byzantine nodes from the honest ones. Then, we introduce a game-theoretic formulation of the decision fusion problem with corrupted nodes thus providing a rigorous framework to evaluate the performance achievable by the fusion center and the Byzantines, when both of them play at the equilibrium. The game-theoretic approach is used to compare the new fusion strategy with the one described in [1]. Finally, we demonstrate the superior performance of the soft identification scheme by means of numerical simulations.

III. DECISION FUSION WITH ISOLATION OF BYZANTINE NODES

A. Problem formulation

In the scenario outlined in the previous section a reasonable strategy for the Fusion Center (FC) is to first try to identify which are the corrupted nodes, discard them and then decide about the state of the system by relying only on the remaining reports [8]. In the following we give an exact formulation of such an approach.

As we said, we are considering the case of binary reports. Specifically, each node makes a local decision about the state of the observed system and forwards its one-bit decision to FC, which must decide between hypothesis H_0 (system is in state S_0) and hypothesis H_1 (system is in state S_1). We assume that a fraction α of the k nodes (or links, according to the definitions given in Figure 1) is under the control of byzantine attackers which, in order to make the fusion process fail, corrupt the reports by flipping the one-bit local decisions with probability P_{mal} (as in [1], [11], we assume a symmetric attacking strategy). By referring to Figure 1, the above attack corresponds to the insertion of a binary symmetry channel with crossover probability P_{mal} in the attacked links.

The strategy adopted by the fusion center consists in trying to identify the corrupted nodes and remove the corresponding reports from the fusion process. To do so, the FC observes the decisions taken by the nodes over a time period T, and makes the final decision on the state of the system at each instant t only at the end of T. To elaborate, for each instant t, we indicate the reports received from the nodes as $r^k(t) = (r_1(t), r_2(t), ..., r_k(t))$ where $r_i(t) \in \{0, 1\}$. The fusion center applies an l-out-of-k fusion rule[2] to $r^k(t)$ to make an intermediate decision on the status of the system at time t. Let us indicate such a decision as $d_{int}(t)$. In order to distinguish the behavior of the honest nodes from that of the corrupted ones, it is convenient to introduce the local decisions $u_i(t)$ made at the node level. More specifically, the local decisions made by the i-th node over the time window T, are denoted as $\mathbf{u}_i = (u_i(1), u_i(2), \ldots, u_i(T))$. The relationship between $u_i(t)$ and the status of the system as time t is ruled by the following equations, which take into account the probability of a decision error by the local node:

$$P\left(u_i(t) = 1 | H_1\right) = P_{d_i} \tag{1}$$

$$P\left(u_i(t) = 1 | H_0\right) = P_{fa_i}, \tag{2}$$

where P_{d_i} and P_{fa_i} are, respectively, the probability of correct detection and false alarm for node i. Note that in accordance to previous works, we adopted a terminology typical of detection theory, even if the analysis presented here focuses mainly on a scenario in which H_0 and H_1 plays a symmetric role. In the following, we assume that the states assumed by the system over subsequent instants are independent of each other. Errors at different nodes and different times are also assumed to be independent.

By assuming that transmission takes place over error-free channels, for honest nodes we have $r_i(t) = u_i(t)$, while for the corrupted links we have $r_i(t) \neq u_i(t)$ with probability P_{mal}. Then, for the corrupted reports we have:

$$P(r_i(t) = 1 | H_1) = P_{mal}(1 - P_{d_i}) + (1 - P_{mal})P_{d_i}, \tag{3}$$

$$P(r_i(t) = 1 | H_0) = P_{mal}(1 - P_{fa_i}) + (1 - P_{mal})P_{fa_i}. \tag{4}$$

Given the observation vector $r^k(t)$ for each t ($t = 1, .., T$), in order to remove the fake reports from the data fusion process, FC proceeds as follows: it associates to each link i a *reputation score* Γ_i, based on the consistency of the reports received from that node with the intermediate decisions $d_{int}(t)$ over the entire time window T. Then, FC isolates the nodes whose reputation is lower than a threshold η and decides about the system state by fusing only the remaining reports.

B. Byzantine Identification: hard reputation measure

A simple Byzantine isolation scheme has been proposed by Rawat et al. in [1], in a scenario related to Collaborative Spectrum Sensing for Cognitive Radio Networks. In such a scenario, each node of the network decides about the presence (S_1) or absence (S_0) of the primary transmitter; on their side, the malicious nodes may send to the FC false data by flipping their local decisions. In the identification scheme proposed in [1], the FC computes for each node i a reputation score by simply counting the number of times that the reports received from that node are different from the intermediate decisions $d_{int}(t)$ during the sensing period T. The *reputation score* $\Gamma_{H,i}$ is hence defined as $\Gamma_{H,i} = \sum_{i=1}^{T} \mathcal{I}(r_i(t) = d_{int}(t))$ where $\mathcal{I}(x)$ (indicator function) is equal to 1 when its argument its true and 0 otherwise. Accordingly, the nodes whose reputation is lower than a threshold η are removed from the fusion process. For each t, the final decision is taken by relying on an l'-out-of-k' rule, where l' is the final decision threshold and k' is the number of nodes remaining after that the thought-to-be byzantine nodes have been discarded.

In [1], the above scheme is shown to be able to mitigate the effect of byzantine attacks when $\alpha < 0.5$, a situation in which the Byzantines are not able to blind the FC by attacking the network independently (referred to as Independent Malicious Byzantine Attacks (IMBA) in [1]), which is the only case considered in this paper.

IV. DECISION FUSION WITH SOFT IDENTIFICATION OF MALICIOUS NODES

In this section, we propose a different isolation strategy which removes the Byzantines from the network according to a soft [3] reliability measure. For any instant t and given the vector $r^k(t)$ with the reports, the new isolation strategy relies on the estimation of the following probabilities:

$$\begin{aligned} &P\left(u_i(t) = 1, r^k(t)\right), \\ &P\left(u_i(t) = 0, r^k(t)\right). \end{aligned} \tag{5}$$

For a honest node, in fact, such probabilities are very different from each other, since the expression for which $r_i(t) = u_i(t)$ is close to 1, while the other is very close to 0. On the contrary, for a byzantine node, the above probabilities tend to be closer. For this reason, we propose to measure the reputation score of a node as follows. For each t we first compute:

$$R_i(t) = \left| \log \left[\frac{P\left(u_i(t) = 0, r^k(t)\right)}{P\left(u_i(t) = 1, r^k(t)\right)} \right] \right|, \tag{6}$$

that is the absolute value of the log-ratios of the two probabilities. Then we set:

$$\Gamma_{S,i} = \sum_{i=1}^{T} R_i(t). \tag{7}$$

To evaluate (6), we start rewriting the joint probabilities within the log as follows (for notation simplicity, we omit the index t):

$$\begin{aligned} P\left(u_i, r^k\right) = &P\left(r^k | u_i, H_0\right) P\left(u_i, H_0\right) \\ &+ P\left(r^k | u_i, H_1\right) P\left(u_i, H_1\right). \end{aligned} \tag{8}$$

To proceed, we make the simplifying assumptions that the reports received by the FC from different nodes are conditionally independent[4]. This is only approximately true since in our scenario we operate under a fixed number of Byzantines, and then the probability that a node is Byzantine depends (weakly) on the state of the other nodes when their

[2] In other words, the fusion center decides in favor of H_1 if l out k nodes decided for such an hypothesis.

[3] We point out that our method is soft for identification of the Byzantines, but is not used in the final decision step.

[4] That is they are independent when conditioning to H_0 or H_1.

number is large enough. Such dependence decreases when the number of nodes increases and disappears asymptotically by the law of large numbers.

Let us now consider the quantity $P(r_j|u_i, H_0)$. When $i = j$, we can omit the conditioning to H_0 since r_i depends on the system status only through u_i. On the other side, when $i \neq j$, we can omit the conditioning to u_i, due to the conditional independence of node reports. A similar observation holds under H_1. Then we can write:

$$P\left(u_i, r^k\right) = P\left(r_i|u_i\right)\left\{P\left(u_i|H_0\right)P\left(H_0\right)\prod_{j\neq i}P\left(r_j|H_0\right) + P\left(u_i|H_1\right)P\left(H_1\right)\prod_{j\neq i}P\left(r_j|H_1\right)\right\}, \tag{9}$$

where $P\left(r_i|u_i\right) = (1 - \alpha P_{mal})$ if $r_i = u_i$, and αP_{mal}, otherwise. Moreover, we have $P(u_j = 1|H_1) = P_{d_j}$ and $P(u_j = 1|H_0) = P_{fa_j}$. In addition:

$$P\left(r_j|H_0\right) = (1 - \alpha P_{mal})P\left(u_j = r_j|H_0\right) + \alpha P_{mal}P\left(u_j \neq r_j|H_0\right) \tag{10}$$

$$P\left(r_j|H_1\right) = (1 - \alpha P_{mal})P\left(u_j = r_j|H_1\right) + \alpha P_{mal}P\left(u_j \neq r_j|H_1\right). \tag{11}$$

By inserting the above expressions in (8) and (6), we can compute the soft reputation score $\Gamma_{S,i}$. Then, the FC relies on $\Gamma_{S,i}$ to distinguish honest nodes from Byzantine ones. Specifically, the distinction is made by isolating those nodes whose reputation score $\Gamma_{S,i}$ is lower than a threshold η (hereafter, we will set $P_{fa_i} = P_{fa}$ and $P_{d_i} = P_d$ $\forall i$).

We conclude this section by observing that, strictly speaking, FC is required to know α and the flipping probability P_{mal}. With regard to α, we assume that FC knows it. As to P_{mal}, in the next sections, we will see that choosing $P_{mal} = 1$ is always the optimum strategy for the attackers, and hence FC can assume that $P_{mal} = 1$.

V. A GAME-THEORETICAL APPROACH TO THE DECISION FUSION PROBLEM

In this section, we evaluate the performance achieved by using the soft Byzantine isolation strategy defined in the previous section and compare it with the hard identification strategy described in [1]. To do so, we use a game-theoretic approach in such a way to analyze the interplay between the choices made by the attackers and the fusion center.

A. Game theory in a nutshell

A 2-player game is defined as a 4-uple $G(\mathcal{S}_1, \mathcal{S}_2, v_1, v_2)$, where $\mathcal{S}_1 = \{s_{1,1} \ldots s_{1,n_1}\}$ and $\mathcal{S}_2 = \{s_{2,1} \ldots s_{2,n_2}\}$ are the set of strategies the first and the second player can choose from, and $v_l(s_{1,i}, s_{2,j}), l = 1, 2$, is the payoff of the game for player l, when the first player chooses the strategy $s_{1,i}$ and the second chooses $s_{2,j}$. When $v_1(s_{s1,i}, s_{2,j}) = -v_2(s_{1,i}, s_{2,j})$, the two players have opposite payoffs and the game is said to be a zero-sum game. In this paper, we consider a strategic game, meaning that the players choose their strategies before starting the game without knowing the strategy chosen by the other player.

Game theory aims at determining the existence of equilibrium points, i.e. pair of strategies that in *some sense* represent a *satisfactory* choice for both players [13]. The most famous equilibrium notion is due to Nash. Intuitively, a profile is a Nash equilibrium if each player does not have any interest in changing its choice assuming the other does not change its strategy. A stronger equilibrium notion passes through the definition of dominant strategy. A strategy is said to be strictly dominant for one player if it is the best strategy for the player, no matter how the other player decides to play. In many cases dominant strategies do not exist, however when one such strategy exists for one of the players, he will surely adopt it. The other player, in turn, can choose his strategy anticipating that the first player will play the dominant strategy. In this way, when a dominant strategy exists, the game is dominance solvable and the players have only one rational choice called the only rationalizable equilibrium of the game [14].

B. The Decision Fusion Game: definition

A first attempt to cast the decision fusion process under byzantine attacks into a game-theoretic framework can be found in [11]. In that paper, the FC is given the possibility of setting the local sensor threshold for the hypothesis testing problem at the nodes and the fusion rule, while the Byzantines can choose the flipping probability P_{mal}.

With respect to [11], we study a more general version of the decision fusion game which includes the isolation scheme described in Section III. To this purpose, the FC is endowed with the possibility of setting the isolation threshold η, as well as the final fusion rule after removal of byzantine nodes. Finally, the performance are evaluated in terms of overall error probability after the removal step. We suppose that FC does not act strategically on the local sensor threshold; then P_d and P_{fa} are fixed and known to FC. With regard to the Byzantines (B), they are free to decide the flipping probability P_{mal}.

With the above ideas in mind, we define the general decision fusion game as follows:

Definition 1: The $DF(\mathcal{S}_{FC}, \mathcal{S}_B, u)$ game is a zero-sum strategic game, played by FC and B, defined by the following strategies and payoff.

- The set of strategies available to FC is given by all the possible isolation thresholds η, and the values of l and l' in the l-out-of-k intermediate and final decision rules:
$$\mathcal{S}_{FC} = \ \{(l, \eta, l');\, l, l' = 1, .., k,\, \eta_{\min} \leq \eta \leq \eta_{\max}\}, \tag{12}$$
where $\eta_{\min}$ and $\eta_{\max}$ depend on the adopted isolation scheme.
- The set of strategies for B are all the possible flipping probabilities for the corrupted nodes:
$$\mathcal{S}_B = \{P_{mal},\, 0 \leq P_{mal} \leq 1\}. \tag{13}$$
- The payoff is the final error probability after malicious node removal, namely $P_{e,ar}$. Of course, FC wants to minimize $P_{e,ar}$, while B tries to maximize it.

Applying the above definition to the identification schemes introduced so far, we see that for the case of hard reputation measure (DF_H game), the values of the isolation threshold η range in the set of integers from 0 to T, while for the scheme based on the soft removal of the malicious nodes (DF_S game) η may take all the continuous values between $\eta_{\min} = \min_{i=1,..,k} R_i(t)$ and $\eta_{\max} = \max_{i=1,..,k} R_i(t)$.

C. The Decision Fusion Game: equilibrium point

With regard to the optimum choice for the Byzantines, previous works have either conjectured or demonstrated (in particular cases) that $P_{mal} = 1$ is a dominant strategy [1], [11]. Even in our case, the simulations we carried out, some of which are described in the next section, confirms that $P_{mal} = 1$ is indeed a dominant strategy for both the hard and the soft identification schemes. This means that, notwithstanding the introduction of an identification scheme for discarding the reports of malicious nodes from the fusion process, the optimum for the Byzantines is (still) always flipping the local decisions before transmitting them to FC. This means that for the Byzantines it is better to use all their power ($P_{mal} = 1$) in order to make the intermediate decision fail than to use a lower P_{mal} to avoid being identified.

As a consequence of the existence of a dominant strategy for B, the optimum strategy for FC is the triple (l^*, η^*, l'^*) which minimizes $P_{e,ar}$ when $P_{mal} = 1$. By exploiting a result derived in [9] for the classical decision fusion problem and later adopted in [11] in presence of Byzantines, the optimal value l^* determining the intermediate fusion rule is given by

$$l^* = \frac{\ln\left[(P(H_0)/P(H_1))\{(1-p_{10})/(1-p_{11})\}^k\right]}{\ln\left[\{p_{11}(1-p_{10})\}/\{p_{10}(1-p_{11})\}\right]}, \quad (14)$$

where $P(H_0)$ and $P(H_1)$ are the a-priori probabilities of H_0 and H_1, while $p_{10} = p(r = 1|H_0)$ and $p_{11} = p(r = 1|H_1)$, evaluated for $P_{mal} = 1$. With regard to η and l', we have:

$$(\eta^*, l'^*) = \arg\min_{(\eta, l')} P_{e,ar}((l^*, \eta, l'), P_{mal} = 1). \quad (15)$$

Depending on the adopted isolation scheme, we have a different expression for $P_{e,ar}$ and then different η^*'s and l'^*'s as well. The minimization problem in (15) is solved numerically for both hard and soft isolation in the next section. According to the previous analysis, $((l^*, \eta^*, l'^*), P^*_{mal})$ is the only *rationalizable equilibrium* for the DF game, thus ensuring that any rational player will surely choose these strategies. The value of $P_{e,ar}$ at the equilibrium represents the achievable performance for FC and is used to compare the effectiveness of data fusion based on soft and hard Byzantine isolation.

VI. PERFORMANCE ANALYSIS

In this section, we evaluate the performances at the equilibrium for the two games DF_H and DF_S, showing that the soft strategy outperforms the one proposed in [1], in terms of $P_{e,ar}$. We also give a comparison of the two schemes in terms of isolation error probability.

In all our simulations, we consider a multi-sensor network with $k = 100$ nodes. We assume that the probability of the two states S_0 and S_1 are the same. We run the experiments with the following settings: $P_d = 1 - P_{fa}$ takes values in the set $\{0.8, 0.9\}$ and $\alpha \in [0.4, 0.49]$, corresponding to a number of honest nodes ranging from 51 to 60. The observation window T is set to 4 (such a value determines the delay of the decision at the FC and then reasonably it must be kept low in practical applications). For each setting, the probability of error $P_{e,ar}$ of the two schemes is estimated over 50000 simulations.

Due to the symmetry of the experimental setup with respect the two states, we have that $p_{10} = p_{01} = 1 - p_{11}$. Accordingly, from (14) we get that $l^* = k/2$ and then the majority rule is optimal for any P_{mal} (not only at the equilibrium). Besides, still as a consequence of the symmetric setup, the optimality of the majority rule is experimentally proved also for the final fusion rule, regardless of the values of η and P_{mal}. Then, in order to ease the graphical representation of the game in normal form, we fix $l^* = 50$ and $l'^* = k'/2$ and remove these parameters from the strategies available to the FC.

Tables I and II show the payoff matrix for the DF_H and DF_S games when $\alpha = 0.46$ and $P_d = 0.8$ (very similar results are obtained for different values of these parameters). For the DF_S game, the threshold values are obtained from the reliability interval $[\eta_{S,\min}, \eta_{S,\max}]$. Since the reliability measures take different values for different P_{mal} a large number of thresholds have been considered, however for sake of brevity, we show the results obtained with a rather coarse quantization interval, especially far from the equilibrium point. As to the strategy of the Byzantines, the simulation results confirm the dominance of $P_{mal} = 1$ for both games. Looking at the performance at the equilibrium, we see that the DF_S game is more favorable to the FC, with a $P_{e,ar}$ at the equilibrium equal to 0.1375 against 0.1982 for the DF_H game. In Figure 2, the two games are compared by plotting the corresponding payoffs at the equilibrium for various values of α in the interval $[0.4, 0.49]$. Upon inspection of the figure, the superiority of the soft isolation scheme is confirmed. Finally, we compared the two schemes in terms of capability of isolation of the byzantine nodes. The ROC curve with the probability of correct isolation (P^B_{ISO}) versus the erroneous isolation of honest nodes (P^H_{ISO}), obtained by varying η, is depicted in Figure 3 for both schemes. The curves correspond to the case in which $\alpha = 0.46$ and $P_d = 0.8$. As we can see, soft isolation allows to obtain a slight improvement of the isolation performance with respect to isolation based on a hard reputation score.

VII. CONCLUSIONS

We presented a new scheme for decision fusion in the presence of Byzantine nodes, relying on a soft reputation measure for the identification of nodes. In order to evaluate the performance of the new scheme and compare it against prior art based on a hard reputation measure, we have introduced a game theoretic framework which is particularly suited to analyze the interplay between the fusion center and

η_H / P_{mal}	0.6	0.7	0.8	0.9	1
4	0.0016	0.0087	0.0354	0.1109	0.2746
3	0.0015	0.0078	0.0262	0.06628	**0.1982**
2	0.0016	0.0080	0.0281	0.0726	0.1998
1	0.0016	0.0087	0.0354	0.1109	0.2746
0	0.0016	0.0087	0.0354	0.1109	0.2746

TABLE I

PAYOFF OF THE DF_H GAME FOR $\alpha = 0.46$ AND $P_d = 80$, $P_{fa} = 0.2$.

η_S / P_{mal}	0.6	0.7	0.8	0.9	1
$\eta_{S,\min}$	0.0009	0.0035	0.0131	0.0596	0.2253
·	0.0009	0.0035	0.0131	0.0596	0.1889
·	0.0009	0.0035	0.0131	0.0596	0.1589
·	0.0009	0.0035	0.0131	0.0596	0.1401
·	0.0009	0.0035	0.0131	0.0596	0.1405
·	0.0009	0.0035	0.0131	0.0596	**0.1375**
·	0.0009	0.0035	0.0131	0.0596	0.1528
·	0.0009	0.0035	0.0131	0.0596	0.1801
·	0.0009	0.0035	0.0131	0.0596	0.2192
·	0.0009	0.0035	0.0131	0.0596	0.2742
·	0.0009	0.0035	0.0131	0.0361	0.2742
·	0.0009	0.0035	0.0131	0.0209	0.2742
·	0.0009	0.0035	0.0131	0.0586	0.2742
·	0.0009	0.0035	0.0131	0.1108	0.2742
·	0.0009	0.0035	0.0088	0.1108	0.2742
·	0.0009	0.0035	0.0054	0.1108	0.2742
·	0.0008	0.0021	0.0355	0.1108	0.2742
$\eta_{S,\max}$	0.0006	0.0011	0.0355	0.1108	0.2742

TABLE II

PAYOFF OF THE DF_S GAME FOR $\alpha = 0.46$ AND $P_d = 80$, $P_{fa} = 0.2$.

the Byzantines. We evaluated the equilibrium point of the game by means of simulations and used the payoff at the equilibrium to assess the validity of the new soft reputation metric.

Future work will focus on two research directions. On one side we will try to derive a MAP fusion rule to further improve the performance of the fusion center. On the other side we will improve the performance of the Byzantines by letting them exploit the knowledge of the observation vectors (decision fusion with corrupted nodes).

REFERENCES

[1] A. S. Rawat, P. Anand, H. Chen, and P. K. Varshney, "Collaborative spectrum sensing in the presence of byzantine attacks in cognitive radio networks," *IEEE Transactions on Signal Processing*, vol. 59, no. 2, pp. 774–786, February 2011.

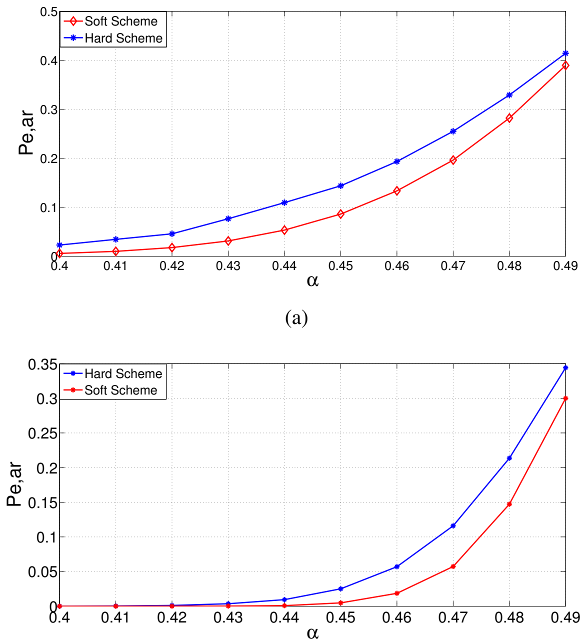
[2] W. Wang, H. Li, Y. Sun, and Z. Han, "Securing collaborative spectrum sensing against untrustworthy secondary users in cognitive radio networks," *EURASIP Journal on Advances in Signal Processing*, vol. 2010, p. 4, 2010.

[3] K. Hoffman, D. Zage, and C. Nita-Rotaru, "A survey of attack and defense techniques for reputation systems," *ACM Computing Surveys (CSUR)*, vol. 42, no. 1, p. 1, 2009.

[4] Y. Sun and Y. Liu, "Security of online reputation systems: The evolution of attacks and defenses," *IEEE Signal Processing Magazine*, vol. 29, no. 2, pp. 87–97, March 2012.

[5] M. Barni and F. Pérez-González, "Coping with the enemy: advances in adversary-aware signal processing," in *ICASSP 2013, IEEE International Conference on Acoustics, Speech and Signal Processing*, Vancouver, Canada, May 2013.

[6] M. Barni and B. Tondi, "Multiple-observation hypothesis testing under adversarial conditions," in *Proc. of WIFS'13, IEEE International Workshop on Information Forensics and Security*, Guangzhou, China, Nov 2013, pp. 91–96.



(a)

(b)

Fig. 2. Error probability $P_{e,ar}$ at the equilibrium for $P_d = 0.8$ (a) and $P_d = 0.9$ (b).

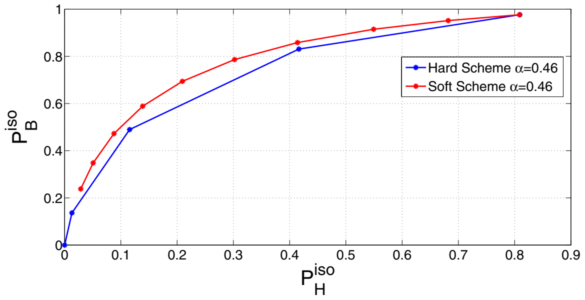


Fig. 3. P_{iso}^H vs. P_{iso}^B at $P_{mal} = 1.0$, for $\alpha = 0.46$ and $P_d = 0.8$. For the soft scheme, 10 thresholds are taken.

[7] L. Lamport, R. Shostak, and M. Pease, "The Byzantine generals problem," *ACM Trans. Program. Lang. Syst.*, vol. 4, no. 3, pp. 382–401, 1982.

[8] A. Vempaty, T. Lang, and P. Varshney, "Distributed inference with byzantine data: State-of-the-art review on data falsification attacks," *IEEE Signal Processing Magazine*, vol. 30, no. 5, pp. 65–75, Sept 2013.

[9] P. K. Varshney, *Distributed Detection and Data Fusion*. Springer-Verlag, 1997.

[10] S. Marano, V. Matta, and L. Tong, "Distributed detection in the presence of byzantine attacks," *IEEE Transactions on Signal Processing,*, vol. 57, no. 1, pp. 16–29, 2009.

[11] B. Kailkhura, S. Brahma, Y. S. Han, and P. K. Varshney, "Optimal distributed detection in the presence of byzantines," in *ICASSP 2013, IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, Vancouver, Canada, 27-31 May 2013.

[12] K. G. Vamvoudakis, I. J. P. Hespanha, F. Ieee, B. Sinopoli, and I. Y. Mo, "Adversarial detection as a zerosum game," in *Proc. IEEE Conference on Decision and Control*, 2012, pp. 7133–7138.

[13] M. J. Osborne and A. Rubinstein, *A Course in Game Theory*. MIT Press, 1994.

[14] Y. C. Chen, N. Van Long, and X. Luo, "Iterated strict dominance in general games," *Games and Economic Behavior*, vol. 61, no. 2, pp. 299–315, November 2007.