Game-Theoretic Learning for Sensor Reliability Evaluation without Knowledge of the Ground Truth

Anis Yazidi, *Senior Member, IEEE,* Hugo L. Hammer, Konstantin Samouylov, Enrique Herrera-Viedma

***Abstract*—Sensor fusion has attracted a lot of research attention during the last years. Recently, a new research direction has emerged dealing with sensor fusion *without* knowledge of the ground truth. In this paper, we present a novel solution to the latter pertinent problem. In contrast to the first reported solutions to this problem, we present a solution that does not involve any assumption on the group average reliability which makes our results more general than previous works. We devise a strategic game where we show that a perfect partitioning of the sensors into reliable and unreliable groups corresponds to a Nash equilibrium of the game. Furthermore, we give sound theoretical results that prove that those equilibria are indeed the *unique* Nash equilibria of the game. We then propose a solution involving a team of Learning Automata (LA) to unveil the identity of each sensor, whether it is reliable or unreliable, using game-theoretic learning. Experimental results show the accuracy of our solution and its ability to deal with settings that are unsolvable by legacy works.**

***Index Terms*—Unreliable Sensors Identification, Game Theory, Learning Automata, Sensor Fusion.**

I. Introduction

Data fusion from noisy sensors [1], [2], [3], [4] has been an active research topic specially with the emergence of the concept of Internet of Things [5] (IoT).

Data fusion involves combining multiple observations from an environment or phenomenon to produce a more robust, a more accurate or a more complete description about a process being monitored. The underlying idea is to remedy the imperfection of information by exploiting the redundancy or complementarity of the data.

Sensors are known to yield measurement errors due to different physical phenomena that limit their accuracy. The process of fusing measurements from *redundant* unreliable sensors each characterized by some level of fidelity is known to increase the reliability of the aggregated measurement and yields more accurate insight about the process being monitored [2], [1], [6], [7].

A. Yazidi and H. Hammer are with the Department of Computer Science, Oslo Metropolitan University, Oslo, Norway.

Konstantin Samouylov is with the Applied Probability and Informatics Department, Peoples' Friendship University of Russia (RUDN University) and with the Institute of Informatics Problems, Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, Moscow, Russian Federation.

Enrique Herrera-Viedma is with Andalusian Research Institute in Data Science and Computational Intelligence, University of Granada, Granada, Spain. He is also with Department of Electrical and Computer Engineering, Faculty of Engineering, King Abdulaziz University, Jeddah, Saudi-Arabia.

The authors would like to thank FEDER financial support from the Project TIN2016-75850-P. It also was supported by the RUDN University Program 5-100.

The vast majority of research in this direction assumes that the reliability of the sensors can be deduced by comparing their readings against the ground truth. The Weighted Majority Voting (WMV) algorithm [8] is a typical example of an algorithm that operates under the assumption that the ground truth is revealed subsequently to measurement and thus the reliability of the sensors can be deduced. Other algorithms suppose that the reliability of the sensors is known beforehand through an offline training phase where the ground truth is available during that phase or computed based on the physical properties of the sensors. Given the knowledge of the reliability of a sensor, a multitude of conventional fusing approaches can be deployed such as Ordered Weighted Averaging (OWA) method, Bayesian approaches, Dempster Shafer theory and Kalman filters [9], [10], [11], [12]. However, in many real-life applications accessing the ground truth is practically impossible especially in harsh environment [13]. In such settings, assessing the reliability of the sensors is far from being obvious under the absence of the ground truth. Although the problem of assessing the reliability of sensors under the absence of the ground truth is apparently impossible to solve, with few insights, Yazidi et al. [14] have shown that it is possible to solve this seemingly impossible paradox. In [14], Yazidi et al. advocated a solution to the problem motivated by the observation that the agreement between the sensors themselves is a key factor in determining their respective reliability. Similar ideas resorting to the agreement between the source of information as a way to assess their credibility have been reported in the literature [15], [16][1]. However, in contrast to those studies, in our current settings the readings are stochastic and therefore the reliability needs to be learned in an online and gradual manner. In [16], the authors propose to aggregate the decision from different sources of information using a modified average. More precisely and in contrast to the Murphy's approach [18], the weights given to the sources of evidence are not equal. The weights are computed by measuring the similarity between the different bodies of evidence using a so-called Similarity Measure Matrix (SMM). However, this makes the complexity of the algorithm quadratic in terms of number of sensors. Furthermore, the SMM does not take into account the behavior over time of the sensor. In fact, in our settings some sensors systematically deviate from the rest which can not be deduced by only one observation instance at a time as it is the case of [16].

[1]We thank an anonymous reviewer for drawing our attention to seminal references on using agreement between sources of information as a metric to assess their reliability [15], [16] and for pointing an early application of game theory in information fusion [17].

The latter temporal aspect was addressed in a subsequent work [15] where the authors introduce a dynamic reliability measure that is measured by assessing the degree of consensus among the sensors. More precisely, the authors in [15] divide the reliability into a static part which is deduced during a supervised training phase where the ground truth is available, and a dynamic part which is evaluated in an unsupervised manner, i.e., without the knowledge of the ground truth. However, the latter work suffers from an inherent drawback present in [16] since the complexity of computing the dynamic reliability part using SMM is quadratic.

The theoretical results reported in [14] are largely based on the work of Boland [19] who studied a generalized version of the Condorcet Jury Theorem. In fact, while the Condorcet Jury theorem treats the case of homogeneous voters, Boland presents the results for heterogeneous voters belonging to two groups where the two groups have opposite interests expressed in a probabilistic manner. By virtue of analogy with the sensor fusion problem, the approach in [14] foresees two groups of sensors, one group of reliable sensors with the interest in reporting the ground truth and another group of unreliable sensors which has interest in misreporting the truth. In a subsequent work, Yazidi and Herrera-Viedma [20] propose an alternative solution that does not involve the majority voting concept as a way for deducing the reliability of the sensors. Instead of applying a majority-based update such as in [21], Yazidi and Herrera-Viedma [20] propose rather to use a reinforcement learning with continuous feedback as opposed to the binary feedback methodology proposed in [14].

However, the premises of aforementioned two main works [14], [20] for identifying unreliable sensor is a condition according to which the truth prevails over lies expressed using the condition $(N_R-1)p_R+N_Up_U > (N_R+N_U)/2$ where N_R, N_U, P_R and P_U are the number of reliable sensors, number of unreliable sensors, the probability that a reliable sensor reports the truth, and the probability that an unreliable sensor reports the truth respectively. Please note that in the particular case where $(p_R, p_U) = (1, 0)$ meaning that a reliable sensors always reports the truth while an unreliable sensor always misreports the truth, the condition reduces to a simple majority condition where the number of reliable sensors constitutes the majority of the sensors. It is worth mentioning that the main advantage of the work in [20] compared to the original work [14] is that the former is more general and does not require that the total number of sensors $N_R + N_U$ is an even number, and therefore we could say that the advantage is relying on more mild condition.

In this paper, we present a more general solution to the problem of identifying unreliable sensors that does not invoke the condition $(N_R-1)p_R+N_Up_U > (N_R+N_U)/2$ commonly used in the original solutions [20], [14]. To solve the problem under those general settings, we resort to the field of game theory in order to gradually learn the identity of the sensors in a decentralized manner. We apply LA as a learning strategy in order to evolve the game toward a strategic equilibrium state which corresponds to unveiling the identity of the sensors.

Among recent applications of LA in game theory figure relay selection in cooperative transmission in vehicular ad-hoc networks [22], opportunistic spectrum access in cognitive networks [23], distributed multiuser computation offloading for cloudlet-based mobile cloud computing [24] and user association for heterogeneous networks [25]. The research on the applications of game theory to the field of information fusion is very scarce with few exceptions [17], [26]. A notable work is due to Deng et al. [17] who propose to use evolutionary game theory, and more particularly replicator dynamics, in order to find the most supported evidence in a multievidence system.

In this paper, we will show that a perfect partitioning of the sensors into reliable and unreliable groups corresponds to a Nash equilibrium of an appropriately designed game. We design an LA that is able to converge to this Nash equilibrium through repeated learning.

The contribution of this article can be summarized as follows:

- We present a general solution to the problem of identifying unreliable sensors without the knowledge of the ground truth that requires milder conditions compared to the legacy solutions [20], [14]. In fact, our solution does not impose any condition on the group average reliability.
- The solution is able to converge a perfect partitioning of the sensors even under stochastic deceptive environments [27].
- In order to cope with the stochastic nature of the sensor readings, we use reinforcement learning and model the sensor reliability identification problem as a repeated game. Our work can pave the way towards more research interest in the intersection between game theory and information fusion which is still a fertile area of research.
- We formally prove that, by a careful design of the utility function, the set of Nash equilibria of the game yield an optimal solution to our sensor fusion problem.
- We show that the experimental results are in concordance with the theoretical findings.

The rest of the paper is organized as follows. Section II briefly reviews the theory of LA which is the main tool used in this paper. Section III gives a formal statement of the problem. In Section IV, we present a game-theoretic-based scheme for identifying unreliable sensors in a stochastic environment in the absence of knowledge of the ground truth. Some experimental results that validate our theoretical findings are presented in Section V. Section VI concludes the paper.

II. Stochastic Learning Automata

Learning Automata (LA) is a decision making mechanism for learning under uncertainty and limited information from the environment [21], [28], [29], [30]. The earliest work on LA is due to the Soviet cyberneticist Tsetlin [31] who devised the earliest first learning machines called Tsetlin Automata. The Tsetlin LA was shown to be able to exhibit a self-organizing collective behavior using simple learning rules. Such collective behavior was demonstrated for the case of the Goore game [32] which is a distributed control involving unreliable feedback from the environment. The adoption of the term "Learning Automata" is due to Narendra and Thathachar [28] who built a general family of LA algorithms and established the theoretical fundaments of the so-called variable structure LA schemes.

In simple terms, the LA is a theory according to which a learning agent can gradually learn to interact with a random environment by sequentially choosing actions and receiving feedback about the choices. The LA update loops can be characterized by the learning loop depicted in Figure 1.

In formal terms, a LA is defined by the following quintuple $\langle A, B, Q, F(.,.), G(.)\rangle$, with:

1) $A = \{\alpha_1, \alpha_2, \ldots, \alpha_r\}$ is the set of actions that the LA can select from, and $\alpha(t)$ denotes the actions chosen at time instant t.
2) $B = \{\beta_1, \beta_2, \ldots, \beta_m\}$ is the set of all possible input to the LA subsequent to an action choice. $\beta(t)$ denotes the input at time instant t.
3) $Q = \{q_1, q_2, \ldots, q_s\}$ represents the states of the LA where $Q(t)$ is the state at time instant t.
4) $F(.,.) : Q \times B \mapsto Q$ is a the transition function at time t, such that, $q(t+1) = F[q(t), \beta(t)]$. In other words, $F(.,.)$ gives the next state of the LA at time instant $t+1$ given the current state and the input from the environment both at time t. The next state can be obtained either using a deterministic or stochastic mapping.
5) $G(.)$ defines *output function* and it is a mapping $G : Q \mapsto A$ which determines the action of the LA as a function of the state.

The Environment, E is characterized by :

- $C = \{c_1, c_2, \ldots, c_r\}$ is a set of penalty probabilities, where $c_i \in C$ corresponds to the penalty of action α_i.

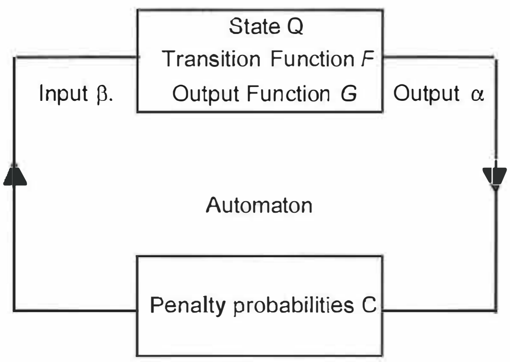

Fig. 1. Feedback Loop of LA.

LA consists of two main streams of approaches: Fixed Structure Stochastic Automata (FSSA) and Variable Structure Stochastic Automata (VSSA). It is worth mentioning that the Tsetlin Automata falls under the class of FSSA. In the VSSA family, the LA maintains a probability vector in the case of finite state action environments, or a probability distribution over the actions space in the case of infinite state action environment that are updated recursively according to the responses from the environment. In the case of FFSA, the decisions of the LA are taken according to a time invariant mapping between the cross product of its internal finite states and the feedback from the environment.

Continuous LA schemes, which by definition operate on a continuous probability space, are known to be slow specially because the larger the probability of one action, the smaller the magnitude of increase of its probability. A breakthrough in the field of LA is the advent of discretized LA [33], [34] which are shown to be significantly faster than their preceding versions of continuous LA. Discretized LA algorithms work with a finite probability space in which the probability of one action takes values from a limited set of values. The discretization of the probability space can be either linear or non-linear depending on whether the finite values of the probability are equi-distant or not.

There are many ways to classify LA algorithms. LA can be classified according to the nature of feedback from the environment into P-Model, Q-Model and S-Model [35]. In the case of P-Model, the feedback of the environment is binary, either reward or penalty. By the way of notation convention 0 corresponds to reward and 1 corresponds to penalty. In the Q-Model, the feedback of the environment can be mapped to a discrete set of values in the interval $[0, 1]$. In the S-Model, the feedback from the environment can be any continuous number in the interval $[0, 1]$. Usually the feedback is normalized in order to be bounded within the interval $[0, 1]$.

LA has found a large set of applications. Those applications include routing problems [36], [37], [38], [39], [40], image processing [41], [42], recommendation systems [43], [44], [45], priority assignment in queueing systems [46], adaptive polling protocols [47], [48], [49], resource allocation under uncertainty [50], to mention a few.

III. Modeling the Problem

We consider a population of N sensors, $\mathcal{S} = \{s_1, s_2, \ldots, s_N\}$. Let $T(t)$ be the unknown ground truth at the time instant t modeled by a binary variable which can take one of the two possible values, 0 and 1. The value of T is unknown and can only be inferred through measurements from sensors. The output from the sensor s_i is referred to as x_i. Let π be the probability of the state of the ground truth, i.e., $T = 0$ with probability π.

We suppose that the probability of the sensor reporting a value erroneously is symmetric. Formally, this reduces to:

$$Prob(x_i = 0|T = 1) = Prob(x_i = 1|T = 0). \qquad (1)$$

Further, let p_i denote the Correctness Probability (CP) of sensor s_i, where: $p_i = Prob(x_i = 0|T = 0) = Prob(x_i = 1|T = 1)$. Let $q_i = 1 - p_i$ denote the Error Probability (EP) of of sensor s_i.

Using the law of total probability it is easy to prove that $Prob(x_i = T)$ is, indeed, p_i.

We can define a reliable sensor to be one that has a CP $p_i > 0.5$ and an unreliable sensor as one that has a CP $p_i < 0.5$.

In addition, we assume that every p_i can have one of two possible values from the set $\{p_R, p_U\}$, where $p_R > 0.5$ and $p_U < 0.5$. Then, a sensor s_i is said to be reliable if $p_i = p_R$, and is said be unreliable if $p_i = p_U$. We assume that p_R and p_U are unknown to the algorithm.

Based on the above, the set of reliable sensors is $\mathcal{S}_R = \{s_i | p_i = p_R\}$, and the set of unreliable sensors is $\mathcal{S}_U = \{s_i | p_i = p_U\}$. Furthermore, let $N_R = |\mathcal{S}_R|$ and $N_U = |\mathcal{S}_U|$. Let $q_R = 1 - p_R$ and $q_U = 1 - p_U$ denote the EP of $\mathcal{S}_R$ and $\mathcal{S}_U$ respectively. In order to have a meaningful problem, we suppose that $N_U \geq 1$, $N_R \geq 1$ meaning that there is at least one reliable sensor and one unreliable sensor in the set of sensors. We will use the term identity of a sensor to refer to its type which can be reliable or unreliable. Furthermore, we will use the terms fair sensor and reliable sensor interchangeably.

IV. Solution: Game-Theoretic Learning

We will formulate the problem of sensor reliability evaluation as a repeated strategic game where the aim is to ensure convergence of the unreliable sensors to one action of the game and convergence of the unreliable sensors to the opposite action. We suppose that each sensor can be assimilated to a player in our strategic game. Let a_i denote the action of sensor s_i, referring to the group choice of sensor s_i. We suppose that $a_i(t) \in \{0, 1\}$ where $\{0, 1\}$ corresponds to the set of two groups we are considering, and let $a(t) = \{a_1(t), a_2(t), \ldots, a_N(t)\}$ denote the action profile of the game at time instant t.

We shall now define the reward of player i, $r_i(t)$. For this purpose, let $G_{a_i}(t)$ denote the set of sensors choosing the same action as sensor s_i at time instant t. This is defined formally as $G_{a_i}(t) = \{k \in [1, N]$ such that $a_i(t) = a_k(t)\}$.

Formally $r_i(t)$ is given by

$$r_i(t) = \begin{cases} 1, & \text{if } |G_{a_i}| = 1 \\ \frac{\sum_{k \in G_{a_i}, k \neq i} I\{x_k(t) = x_i(t)\}}{|G_{a_i} \setminus \{i\}|}, & \text{otherwise.} \end{cases} \tag{2}$$

Please note, that according to the above definition, whenever s_i is the only sensor in G_{a_i}, meaning that $G_{a_i}(t) = \{i\}$, we assign 1 to r_i. In the alternative case where $|G_{a_i}(t)| > 1$, $r_i(t)$ reduces to the normalized ratio of the number of players agreeing with sensor i among those sensors that have chosen the same group as s_i at time t. As we will see later in the proof of the theoretical results distinguishing between the latter two cases depending on the cardinally of $G_{a_i}(t)$ is crucial for the convergence of our scheme to a desired equilibrium.

The utility is defined by

$$u_i(a_i, a_{-i}) = \mathbf{E}[r_i | (a_i, a_{-i})] \tag{3}$$

u_i is the utility function of player i which is his expected payoff when he selects his pure strategy a_i while the other players select the profile a_{-i}. Each player in the game aims to maximize his expected payoff. Due to the fact that the sensors provide noisy readings according to an underlying unknown stochastic process, the payoff r_i is a random variable, and therefore the expected payoff is considered. It is worth mentioning that this game is a stochastic game [51] since for a fixed action profile (a_i, a_{-i}) the payoff is not deterministic but rather stochastic. For an example of a stochastic game involving LA we refer the reader to [23] that also considers the expected payoff as in our work.

For $k \in \{0, 1\}$, let $N_{U,k}(t)$ the number of unfair sensors choosing action k at time t. By abuse of notation we denote action k as G_k. Let $N_{R,k}(t)$ be the number of reliable sensors choosing action G_k. Whenever there is no confusion, we will omit the time index t.

A. Construction of the learning automata

Let $p_{i,k}(t)$ denote the probability that the i^{th} sensor takes action k at time t. Note that $k \in \{0, 1\}$.

We give the following update mechanism:

$$p_{i,k}(t+1) \leftarrow p_{i,k}(t) + \lambda r_i(t)\left(1 - p_{i,k}(t)\right), a_i(t) = k$$
$$p_{i,k}(t+1) \leftarrow p_{i,k}(t) - \lambda r_i(t) p_{i,k}(t), a_i(t) \neq k$$

where λ denotes the learning rate that satisfies the condition: $0 < \lambda < 1$. The informed reader observes that each agent or sensor i will choose a group at time instant t. The reinforcement signal is dependent on the readings of the other sensors that have chosen the same group. The reinforcement signal is proportional to the number of sensors agreeing with the sensor s_i and that have chosen the same group as s_i whenever $|G_{a_i}(t)| > 1$. Furthermore, the reinforcement signal is normalized by the size of the group.

The algorithm is given in the form of pseudo-code in Algorithm 1.

Algorithm 1 Distributed Sensor Identification

Require: Initially, for all i, $p_{i,0}(0) = p_{i,1}(0) = 1/2$, $t = 0$.
Require: ϵ convergence parameter, λ learning rate.
while Not all LA converged **do**
  Each sensor s_i senses ground truth and reports $x_i(t)$
  Each sensor s_i chooses an action $a_i(t)$ according to probability vector $[p_{i,0}(t), p_{i,1}(t)]$
  Each sensor s_i receives feedback $r_i(t)$
  for all s_i in the set of sensors **do**

$$p_{i,k}(t+1) \leftarrow p_{i,k}(t) + \lambda r_i(t)\left(1 - p_{i,k}(t)\right), a_i(t) = k$$
$$p_{i,k}(t+1) \leftarrow p_{i,k}(t) - \lambda r_i(t) p_{i,k}(t), a_i(t) \neq k$$

    if $p_{i,k}(t+1) > 1 - \epsilon$ OR $p_{i,k}(t+1) < \epsilon$ **then**
      LA_i has converged
    end if
  end for
  increment time t, $t = t + 1$.
end while

B. Theoretical results

Before we proceed with the main findings of this article, we shall present two theorems that are essentials in order to prove our main theoretical results.

Theorem 1. *Let $s_i \in \mathcal{S}_R$ and suppose that $a_i = k$. Furthermore, we assume that $|G_{a_i}| > 1$ which is equivalent in this case to $N_{R,k} + N_{U,k} > 1$. Then,*

$$u_i(a_i, a_{-i}) = \frac{(N_{R,k} - 1)(p_R^2 + q_R^2) + N_{U,k}(p_U p_R + q_U q_R)}{N_{R,k} + N_{U,k} - 1} \tag{4}$$

Theorem 2. *Let $s_i \in S_U$ and suppose that $a_i = k$. Furthermore, we assume that $|G_{a_i}| > 1$ which is equivalent in this case to $N_{R,k} + N_{U,k} > 1$. Then,*

$$u_i(a_i, a_{-i}) = \frac{(N_{U,k} - 1)(p_U^2 + q_U^2) + N_{R,k}(p_U p_R + q_U q_R)}{N_{R,k} + N_{U,k} - 1} \tag{5}$$

Proof. The proofs of Theorem 1 and Theorem 2 follow the same lines as those of Theorem 1 and Theorem 2 found in [20]. The proofs can be obtained by recurrence and they are omitted for the sake of brevity. □

Theorem 3. *The game admits two pure Nash equilibria where all unreliable sensors converge to the same action, while all reliable sensors converge to the alternative action. Those two pure Nash equilibria satisfy:*

- $a_i = a_j$ *if* s_i *and* s_j *in* S_R *or* s_i *and* s_j *in* S_U.
- $a_i = 1 - a_j$ *if* s_j *in* S_R *and* s_i *in* S_U *or* s_j *in* S_U *and* s_i *in* S_R.

Proof. We will show that the game admits two pure Nash equilibria. According to the above theorem, a Nash equilibrium corresponds to the case where all sensors of the same identity choose the same group while all sensors of the opposite identity choose the opposite groups.

Without loss of generality, we suppose all the reliable sensors in S_R select G_k and all the unreliable sensors in S_U select G_{1-k} where $k \in \{0, 1\}$. We will show that no sensor in G_k or in G_{1-k} can change unilaterally its action without decrease in the utility. By definition, this case corresponds to a Nash equilibrium.

a) Case 1: Let us consider a sensor s_i in G_k, i.e., the group containing exclusively fair sensors.

The utility of s_i is given by

$$u_i(a_i = k, a_{-i}) = \begin{cases} 1, & \text{if } N_{R,k} = 1 \\ \frac{(N_{R,k}-1)(p_R^2+q_R^2)}{(N_{R,k}-1)} = p_R^2 + q_R^2, & \text{if } N_{R,k} > 1 \end{cases} \tag{6}$$

Please note that the above result is obtained by applying Theorem 1 using $N_{U,k} = 0$ and under the condition that $|G_{a_i}| = N_{R,k} > 1$. In the counter-part case where $N_{R,k} = 1$ we have $u_i(a_i = k, a_{-i}) = 1$ which is a consequence of Eq. (2).

We suppose that the sensor changes its action to G_{1-k}. After this change of action, the number of fair sensors in G_{1-k} becomes $N_{R,1-k} = 1$ while the number of unfair sensors in G_{1-k} remains unchanged, i.e., $N_{U,1-k} = N_U$.

Applying Theorem 1, the utility becomes

$$\begin{aligned} u_i(a_i = 1 - k, , a_{-i}) &= \\ \frac{(1-1)(p_R^2 + q_R^2) + N_U(p_U p_R + q_U q_R)}{1 + N_U - 1} & \\ &= p_R p_U + q_R q_U \end{aligned} \tag{7}$$

Let us now consider $u_i(a_i = k, a_{-i}) - u_i(a_i = 1-k, , a_{-i})$ which quantifies the amount of change of the utility of s_i as a consequence of unilaterally switching action to G_{1-k}.

There are two sub-cases to be considered. The first sub-case arises when originally $N_{R,k} = 1$ which implies that we only have one fair sensor among the whole pool of sensors.

$$\begin{aligned} u_i(a_i = k, a_{-i}) - u_i(a_i = 1-k, , a_{-i}) &= \\ 1 - (p_R p_U + q_R q_U) &< 0 \end{aligned} \tag{8}$$

The second sub-case arises when originally $N_{R,k} > 1$ which implies the number of total fair sensors among the whole pool of sensors is strictly larger than 1. After some algebraic simplifications, we obtain

$$\begin{aligned} u_i(a_i = k, a_{-i}) - u_i(a_i = 1-k, , a_{-i}) &= \\ (p_R^2 + q_R^2) - (p_R p_U + q_R q_U) & \\ = p_R(p_R - p_U) + q_R(q_R - q_U) & \\ = p_R(p_R - p_U) + (1 - p_R)(-p_R + p_U) & \\ = (p_R - p_U)(2p_R - 1) & \end{aligned} \tag{9}$$

We know that $p_R > p_U$, and that since $p_R > 1/2$ we also have $2p_R - 1 > 0$. Therefore

$$\begin{aligned} u_i(a_i = k, a_{-i}) - u_i(a_i = 1-k, , a_{-i}) &= \\ (p_R - p_U)(2p_R - 1) &> 0 \end{aligned} \tag{10}$$

Hence the utility decreases in both sub-cases as a consequence of a unilateral change of action.

b) Case 2: Let us consider a sensor s_i in G_{1-k}, i.e., the group containing exclusively unreliable sensors. It is easy to note that s_i is an unreliable sensor. The utility of s_i is given by

$$\begin{cases} 1, & \text{if } N_{U,1-k} = 1 \\ \frac{(N_{U,1-k}-1)(p_U^2+q_U^2)}{(N_{U,1-k}-1)} = p_U^2 + q_U^2, & \text{if } N_{U,1-k} > 1 \end{cases} \tag{11}$$

We suppose that the sensor s_i switches actions by choosing G_k. After this change of action, the number of unfair sensors in G_k becomes $N_{U,k} = 1$ while the number of fair sensors in G_k remains unchanged, i.e., $N_{R,k} = N_R$.

Applying Theorem 2, the new utility of s_i subsequent to action switch becomes

$$\begin{aligned} u_i(a_i = k, a_{-i}) &= \frac{(1-1)(p_U^2 + q_U^2) + N_R(p_U p_R + q_U q_R)}{N_R + 1 - 1} \\ &= p_R p_U + q_R q_U \end{aligned} \tag{12}$$

At this juncture, we consider $u_i(a_i = 1-k, a_{-i}) - u_i(a_i = k, a_{-i})$ which quantifies the amount of change of the utility value of s_i as a consequence of unilaterally switching action.

There are two sub-cases to be considered. The first sub-case is when originally $N_{U,1-k} = 1$ which implies that there is only one unfair sensor among the whole pool of sensors. In this sub-case, we obtain

$$\begin{aligned} u_i(a_i = 1-k, a_{-i}) &- u_i(a_i = k, a_{-i}) \\ &= 1 - (p_R p_U + q_R q_U) < 0 \end{aligned} \tag{13}$$

The second sub-case is when originally $N_{U,1-k} > 1$ which implies the number of total unfair sensors among the whole pool of sensors is strictly larger than 1.

After some algebraic simplifications, we obtain

$$\begin{aligned} u_i(a_i = 1-k, a_{-i}) - u_i(a_i = k, a_{-i}) \\ = (p_U^2 + q_U^2) - (p_R p_U + q_R q_U) \\ = (p_U - p_R)(2p_U - 1) \end{aligned} \tag{14}$$

The above quantity is strictly positive since it is the product of two strictly negative quantities. Therefore the utility of s_i decreases as a consequence of unilaterally changing its action.

Based on the above results, under a Nash equilibrium, all unreliable sensors converge to the same action, while all unreliable sensors converge to the alternative action. A Nash equilibrium satisfies:

- $a_i = a_j$ if both s_i and s_j belong to S_R, or in the case where both s_i and s_j belong to S_U.
- $a_i = 1 - a_j$ if s_j in S_R while s_i in S_U, or in the case where s_j in S_U while s_i in S_R.

It is straightforward to note that there are two Nash equilibria resulting from the latter definition which correspond to 1) when all fair sensors converge to G_0 and all unfair sensors converge to G_1 or 2) vice-versa, i.e., all fair sensors converge to G_1 and all unfair sensors converge to G_0 .

□

Theorem 4. *The Nash equilibria given in Theorem 3 are the unique pure Nash equilibria of the game.*

Proof. We will show that those two Nash equilibria are in fact the *unique* pure Nash of the game by reasoning by contradiction. The informed reader observes that this is a stronger result than the result of Theorem 3 that states that the desirable solutions resulting into perfect partitioning of the sensors are indeed Nash equilibria.

We shall consider all possible configurations excluding the Nash cases in Theorem 3 and show by contradiction that they violate the definition of Nash equilibrium. In formal terms if $\mathcal{A}$ represents all possible actions action profiles, which has 2^N possible states as each sensor has two actions, then we need to show that any action profile in the set $\mathcal{A} \setminus \mathcal{A}^*$ is not a Nash equilibrium where $\mathcal{A}^*$ denotes the set of Nash equilibria defined by Theorem 3.

It is easy to note that when $N = 2$, i.e., $N_R = N_U = 1$, then $\mathcal{A} \setminus \mathcal{A}^*$ corresponds to the states where both sensors choose the same action. Clearly, this is not a Nash equilibrium as any sensor which unilaterally deviates by changing action will experience an increase of its utility to 1.

Now let us consider the case where $N > 2$. We can generalize the result from the previous case where $N = 2$ and note when all the sensors converge to one action exclusively leaving one of the groups empty, any sensor deviation, whether this sensor is fair or unfair, by choosing the opposite group increases its utility to 1. Therefore this case is not a Nash equilibrium.

As a consequence, and reasoning by elimination, we are left with the alternative cases where none of the two groups is empty. Furthermore, we must have at least one group containing at least one fair sensor and at least one unfair sensor. In fact, this is true, as we are excluding the actions profiles where the groups are homogeneous which is a Nash equilibrium.

Without loss of generalities, we suppose that the group containing at least one fair and at least one unfair sensor is G_k. As a consequence $N_{R,k} + N_{U,k} \geq 2$. We also have $G_{1-k} \neq \varnothing$, and therefore $N_{R,1-k} + N_{U,1-k} > 0$.

We shall now consider two sub-cases according to whether $\frac{(N_{U,k}-1)p_U + N_{R,k}p_R}{N_{R,k}+N_{U,k}-1}$ is larger or strictly smaller than $\frac{N_{U,1-k}p_U + N_{R,1-k}p_R}{N_{R,1-k}+N_{U,1-k}}$.

c) Sub-case 1: In this first sub-case, we operate with the condition that

$$\frac{(N_{U,k}-1)p_U + N_{R,k}p_R}{N_{R,k}+N_{U,k}-1} \leq \frac{N_{U,1-k}p_U + N_{R,1-k}p_R}{N_{R,1-k}+N_{U,1-k}} \tag{15}$$

We consider a fair sensor s_i that changes action from group G_k to G_{1-k}. Since $p_R > p_U$, we can write

$$(N_{U,k}-1)p_U + N_{R,k}p_R > N_{U,k}p_U + (N_{R,k}-1)p_R \tag{16}$$

In fact, this is true as we have

$$\begin{aligned} (N_{U,k}-1)p_U + N_{R,k}p_R - (N_{U,k}p_U + (N_{R,k}-1)p_R) = \\ -p_U + p_R > 0 \end{aligned} \tag{17}$$

Therefore, using the above result together with the assumption (Eq. (15)) gives

$$\frac{N_{U,k}p_U + (N_{R,k}-1)p_R}{N_{R,k}+N_{U,k}-1} < \frac{N_{U,1-k}p_U + N_{R,1-k}p_R}{N_{R,1-k}+N_{U,1-k}} \tag{18}$$

We consider a fair sensor s_i that unilaterally changes its action from group G_k to G_{1-k}. The original utility of the sensor s_i before switching to the alternative group is given by

$$\begin{aligned} & u_i(a_i = k, a_{-i}) \\ & = \frac{N_{U,k}(p_U p_R + q_R q_R) + (N_{R,k}-1)(p_R^2 + q_R^2)}{N_{R,k}+N_{U,k}-1} \end{aligned} \tag{19}$$

This is can be written as

$$\begin{aligned} & u_i(a_i = k, a_{-i}) \\ & = p_R \frac{N_{U,k}p_U + (N_{R,k}-1)p_R}{N_{R,k}+N_{U,k}-1} + \\ & \quad q_R(1 - \frac{N_{U,k}p_U + (N_{R,k}-1)p_R}{N_{R,k}+N_{U,k}-1}) \end{aligned} \tag{20}$$

After switching action to G_{1-k}, the new value of the utility becomes

$$\begin{aligned} & u_i(a_i = 1-k, a_{-i}) \\ & = \frac{N_{U,1-k}(p_R p_U + q_R q_U) + N_{R,1-k}(p_R^2 + q_R^2)}{N_{R,1-k}+N_{U,1-k}} \\ & = p_R(\frac{N_{U,1-k}p_U + N_{R,1-k}p_R}{N_{R,1-k}+N_{U,1-k}}) \\ & \quad + q_R(1 - \frac{N_{U,1-k}p_U + N_{R,1-k}p_R}{N_{R,1-k}+N_{U,1-k}}) \end{aligned} \tag{21}$$

In order to compare the utility before and after swapping action, let us consider the function $g(.)$ defined as the convex combination

$$g(\rho) = p_R \cdot \rho + q_R \cdot (1-\rho) \tag{22}$$

Let us investigate the dynamics of $g(\rho)$ by studying its derivative function, $g'(\rho)$, which specifically, has the form $g'(\rho) = 2p_R - 1$. Since, by definition, $p_R > 1/2$, we can confirm that $2p_R - 1 > 0$ which is equivalent to stating that $g'(\rho) > 0$. $g(.)$ is thus a *strictly increasing* function. As per inequality (18) we have

$$\frac{N_{U,k}p_U + (N_{R,k}-1)p_R}{N_{R,k}+N_{U,k}-1} < \frac{N_{U,1-k}p_U + N_{R,1-k}p_R}{N_{R,1-k}+N_{U,1-k}} \tag{23}$$

Resorting to the strictly increasing property of the function $g(.)$, we obtain

$$g(\frac{N_{U,k}p_U + (N_{R,k}-1)p_R}{N_{R,k}+N_{U,k}-1}) < g(\frac{N_{U,1-k}p_U + N_{R,1-k}p_R}{N_{R,1-k}+N_{U,1-k}}) \tag{24}$$

Then we can deduce

$$u_i(a_i = k, a_{-i}) < u_i(a_i = 1-k, a_{-i}) \tag{25}$$

This shows that the utility increases by swapping action and therefore this is not a Nash equilibrium.

d) Sub-case 2: In this second sub-case we operate with the condition that

$$\frac{(N_{U,k}-1)p_U + N_{R,k}p_R}{N_{R,k}+N_{U,k}-1} > \frac{N_{U,1-k}p_U + N_{R,1-k}p_R}{N_{R,1-k}+N_{U,1-k}} \tag{26}$$

Let us consider an unreliable sensor in G_k. We will show that its utility increases by unilaterally changing action to G_{1-k}.

=

$$\begin{aligned} &u_i(a_i = k, a_{-i}) \\ &= \frac{(N_{U,k}-1)(p_U^2+q_U^2) + N_{R,k}(p_Up_R + q_Rq_R)}{N_{R,k}+N_{U,k}-1} \end{aligned} \tag{27}$$

This gives

$$\begin{aligned} &u_i(a_i = k, a_{-i}) \\ &= p_U\frac{(N_{U,k}-1)p_U + N_{R,k}p_R}{N_{R,k}+N_{U,k}-1} \\ &+q_R(1 - \frac{(N_{U,k}-1)p_U + N_{R,k}p_R}{N_{R,k}+N_{U,k}-1}) \end{aligned} \tag{28}$$

Now we consider utility $u_i(a_i = 1-k, a_{-i})$ when the unreliable sensor switches to G_{1-k}.

$$\begin{aligned} &u_i(a_i = 1-k, a_{-i}) \\ &= \frac{N_{U,1-k}(p_U^2+q_U^2) + N_{R,1-k}(p_Up_R + q_Rq_R)}{N_{R,1-k}+N_{U,1-k}} \end{aligned} \tag{29}$$

$$\begin{aligned} &= p_U\frac{N_{U,1-k}p_U + N_{R,1-k}p_R}{N_{R,1-k}+N_{U,1-k}} \\ &+q_U(1 - \frac{N_{U,1-k}p_U + N_{R,1-k}p_R}{N_{R,1-k}+N_{U,1-k}}) \end{aligned} \tag{30}$$

Let us consider the function $h(.)$ defined by

$$h(\rho) = p_U \cdot \rho + q_U \cdot (1-\rho) \tag{31}$$

and investigate the dynamics of $h(\rho)$ by studying its derivative, $h'(\rho)$. Since $h'(\rho) = 2p_U - 1$, and $p_U < 1/2$, we see that $2p_U - 1 < 0$ which is equivalent to the conclusion that $h'(\rho) < 0$. Therefore $h(.)$ is a strictly *decreasing* function.

Because of inequality (26) and because of the strictly decreasing nature of $h(.)$, we have

$$h(\frac{(N_{U,k}-1)p_U + N_{R,k}p_R}{N_{R,k}+N_{U,k}-1}) > \cancel{h}(\frac{N_{U,1-k}p_U + N_{R,1-k}p_R}{N_{R,1-k}+N_{U,1-k}}) \tag{32}$$

This gives

$$u_i(a_i = 1-k, a_{-i}) > u_i(a_i = k, a_{-i})$$

□

Theorem 5. *With a sufficiently small step size λ, the proposed LA algorithm converges to one of the Nash equilibria of the game.*

The result is a consequence of the work of [51] on multi-person discrete game where the payoff after each play is stochastic. The LA game is known to converge in this case to one pure Nash equilibrium. As we have proven in Theorem 4 the game admits only two Nash equilibria which correspond to the reliable sensors converge to one group and the unreliable sensor converge to the opposite group. We have also shown that those Nash equilibria are the optimal and desirable solutions in our sensor identification problem.

V. Experiments

In this section, we report some experimental results that demonstrate the efficiency of our approach. Furthermore, the aim of this section is to verify the theoretical findings that we have derived in the previous Section.

a) Convergence speed and accuracy under varying number of sensors: In this experiment, we investigate the convergence speed of the algorithm and accuracy by varying the total number of sensors from 6 til 1200, i.e. by a factor of 200. The LA is deemed to have converged if one of its action probabilities attained the value $1-\epsilon$, where the value of ϵ was set to 0.01. In Table I, we report the average convergence time for an ensemble of 1000 experiments together with the 95% confidence interval. Please note that the convergence time for an experiment is recorded as the time required to reach convergence for the whole pool of sensors, which is the time required by the last un-converged LA in the pool to converge. We fix $(p_R, p_U) = (0,8,0.1)$ and vary the number of sensors by a multiplicative factor (2, 20 and 200). The learning rate is fixed to 0.01 in all experiments unless specified differently. Two observations are worth mentioning from Table I. First, as we increase the number of sensors, the average convergence time does not increase at the same pace. When we increase (N_R, N_U) from $(5,1)$ to $(1000, 200)$, .i.e., by an order of magnitude of 200 times, the time required for achieving convergence only increased by less than 4 times

(N_R, N_U)	Convergence time
(5,1)	2806 (2757, 2856)
(10,2)	3633 (3573, 3694)
(100,20)	4982 (4917, 5048)
(1000,200)	13744 (13569, 13920)

TABLE I

CONVERGENCE TIME FOR $(p_R, p_U) = (0.8, 0.1)$ UNDER VARYING NUMBER OF SENSORS. THE VALUES IN PARENTHESES REFER TO 95% CONFIDENCE INTERVALS.

from 2806 to 13744. This is a desirable property that shows that the convergence time scales with the number of sensors.

The second remark concerns the accuracy of the scheme. In Table II, we report the accuracy of our scheme under the same number of sensors as in Table I. Interestingly, the error is negligible which demonstrates the high performance of our scheme in differentiating between reliable and unreliable sensors. From Table II, we observe that the error increases as we increase the number of sensors. This is understandable as for a larger pool of sensors there is a higher likelihood of wrong convergence compared to a smaller one.

(N_R, N_U)	error rate
(1,5)	0
(2,10)	2.5 10^{-4}
(20,100)	8.41 10^{-4}
(200,1000)	0.0122

TABLE II

CONVERGENCE ERROR FOR $(p_R, p_U) = (0, 8, 0.1)$ UNDER VARYING NUMBER OF SENSORS

In Figure 2, we report the evolution of the actions probabilities of a team of LA. We suppose that the number of sensors is 7 where $s_1, s_2 \ldots, s_5$ are fair while s_6, s_7 are unfair. We fix $(p_R, p_U) = (0, 8, 0.2)$ and we also fix the learning rate $\lambda = 0.01$. We expect that sensors s_6, s_7 will converge to the same action, either action 0 or 1. In this experiment, we see that they converge to action 1. On the other hand, the rest of the sensors $s_1, s_2 \ldots, s_5$ converge to the opposite action.

b) Increasing the difficulty of the environment: In this experiment, we increase the difficulty of the environment compared to the previous experiment by making it more difficult to differentiate between a reliable sensor and an unreliable one via decreasing p_R from 0.8 to 0.7 and increasing p_U from 0.1 to 02. Please note that by making p_R decrease to a slightly larger value than 0.5 and by increasing p_U to a lower value than 0.5, the environment turns to be more difficult as it becomes harder to differentiate between reliable and unreliable sensors. This is a consequence of the fact that the probability that a reliable and an unreliable sensor disagree decreases in the latter case. By comparing Table I to Table III, we see that the convergence time increased. For example, when $(N_R, N_U) = (5, 1)$, the convergence increased from 2806 iterations in average to 5311. We observe too that the convergence error increased too by comparing Table II to Table IV.

c) Comparing against legacy works: In this experiment, we report comparisons results of the convergence time of our scheme for $\lambda = 0.01$ against the other two schemes in the literature L_{RI} presented in [14] and S-LA [20]. We use a large number of sensors namely 400, where the number of

(N_R, N_U)	Convergence time
(5,1)	5311 (5030, 5593)
(10,2)	6607 (6495, 6720)
(100,20)	10230 (10076, 10384)
(1000,200)	13896 (12903, 14888)

TABLE III

CONVERGENCE TIME FOR $(p_R, p_U) = (0, 7, 0.2)$ UNDER VARYING NUMBER OF SENSORS. THE VALUES IN PARENTHESES REFER TO 95% CONFIDENCE INTERVALS.

(N_R, N_U)	error rate
(5,1)	0.00116
(10,2)	5.833 10^{-4}
(100,20)	0.00497
(1000,200)	0.01187

TABLE IV

CONVERGENCE TIME AND ERROR FOR $(p_R, p_U) = (0, 7, 0.2)$ UNDER VARYING NUMBER OF SENSORS

fair sensors is equal to the number of unfair sensors. We vary the environment (p_R, p_U), we observe that our scheme has more convergence time than the L_{RI} and S-LA. However, our scheme is still comparable to the S-LA in terms of convergence speed. For example, according to Table V, for $(p_R, p_U) = (0.75, 0.35)$, our approach converges in 14385 iterations in average while the S-LA takes 6481 iterations.

Furthermore, as the distance between p_R and p_U decreases, we observe a decrease in the convergence speed. When p_R gets closer and closer to its minimal value 0.5 while p_U increases gradually towards its maximal value 0.5, distinguishing between the readings of a reliable sensor and an unreliable one becomes harder. For example, consider the case where $p_R = 0.7$: as p_U increases from 0.3 to 0.45 covering the set $\{0.3, 0.35, 0.4, 0.45\}$ we observe that with each increase of p_U the convergence time increases too. The same applies when we consider fixed $p_R = 0.8$, as p_U increases from 0.3 to 0.45 the required convergence time increases too. Furthermore, similar conclusions emerge if we compare the convergence time for $p_R = 0.8$ in one hand and $p_R = 0.7$ in the other hand under the same p_U. For example, when $p_U = 0.35$, the convergence time increases from 11504 to 14385 as p_R decreases from $p_R = 0.8$ to $p_R = 0.7$.

However, as we have emphasized previously, our scheme is more general than the compared approaches. In fact, our scheme not only operates under milder conditions compared to the state of the art, but also is able to solve the sensor type identification problem even under deceptive environment. As a future work, it will be interesting to investigate boosting the convergence speed of our scheme by adopting for example discretized LA design [33], [34].

d) Varying the learning rate: We vary the learning rate in this experiment from $\lambda = 0.01$ to $\lambda = 0.001$ and report the convergence time along the error rate in Table VI. We observe for $\lambda = 0.001$, it takes almost 10 times more iterations to achieve convergence compared to $\lambda = 0.01$. As seen in Table VI, the convergence time increases from 10230 to 103333. We observe that when the learning rate is as low as $\lambda = 0.001$ the error is 0 compared to 0.00497 when $\lambda = 0.01$. This is interesting remark as we know according to the theory of LA that there is a trade-off between the convergence speed and

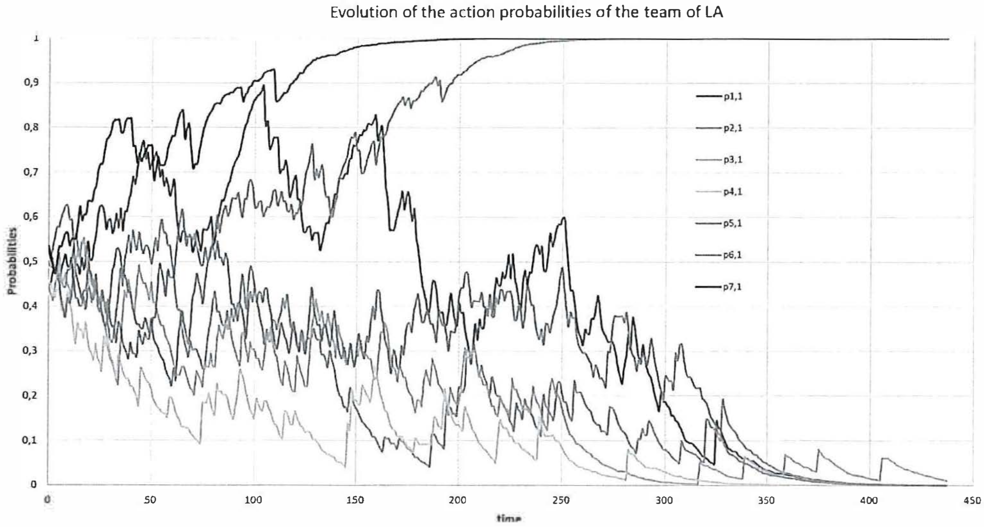

Fig. 2. Evolution of the actions probabilities of a team of LA with 5 fair sensors and 2 unfair sensors with learning rate $\lambda = 0.01$.

(p_R, p_U)	Game LA	S-LA	L_{RI}
(0.75 , 0.45)	52224	7806	2148
(0.75 , 0.4)	23851	6437	966
(0.75 , 0.35)	14385	6481	638
(0.75 , 0.3)	10069	8443	622
(0.8 , 0.45)	42603	6976	2146
(0.8 , 0.4)	17749	5177	975
(0.8 , 0.35)	11504	4650	613
(0.8 , 0.3)	8680	5198	460

TABLE V
AVERAGE CONVERGENCE TIME FOR THE CASE WHEN $(N_R, N_U) = (200, 200)$

the error rate. In fact, by choosing a low learning rate, the convergence speed increases usually in the detriment of the convergence accuracy and vice-versa.

learning rate	convergence time (95% conf. int.)	error rate
0.01	10230 (10075, 10383)	0.004975
0.001	103333 (102294, 104372)	0

TABLE VI
CONVERGENCE TIME WITH 95% CONFIDENCE INTERVAL AND ERROR FOR THE CASE WHEN $(N_R, N_U) = (100, 20)$ AND $(p_R, p_U) = (0, 7, 0.2)$ UNDER VARYING LEARNING RATE

e) Working under the case of deceptive environment: The premises of the legacy work for identifying unreliable sensor is that the truth prevails over lies expressed using the condition $(N_R - 1)p_R + N_U p_U > (N_R + N_U)/2$. In this experiment, we perform tests where this condition is violated and therefore the environment is deemed deceptive [27] as opposed to informative.

In Table VII, we report the average convergence time, together with the confidence intervals for 20 experiments as well as the convergence error under varying number of sensors under a fixed $(p_R, p_U) = (0, 7, 0.2)$. Interestingly, and as expected, the scheme converges with high accuracy even if the environment is not informative. Please note that as we increase the number of sensors from $(N_R, N_U) = (1, 5)$ to $(N_R, N_U) = (200, 1000)$, i.e., by an order of magnitude of 200 times, the average convergence time only doubles from 7279 to 14698. It seems that independently of whether the environment is informative or deceptive the scheme exhibits similar behavior in terms of convergence speed and error rate. In fact, if we replace in Theorem 1, p_R by $1 - p_U$ which is larger than 0.5 and p_U by $1 - p_R$ which is less than 0.5, and exchange the number of reliable and unreliable sensors, the utility function turns out to be identical. In other terms, the settings of the informative environment and those of its constructed counter-part deceptive environment produce the same utility function. Following a similar reasoning, it is easy to note that Theorem 1 and Theorem 2 are symmetric. We should emphasize that theoretical results obtained for all other legacy works [14], [20] operate under the assumption of informative environment. Therefore, our algorithm presented in this paper can be considered as the most general solution to the sensor identification problem found in the literature.

(N_R, N_U)	Convergence time (95% conf. int.)	error rate
(1,5)	7279 (5725, 8833)	0
(2,10)	8711 (7626, 9795)	0
(200, 100)	11011 (10253, 11768)	0.00125
(200, 1000)	14698 (14111, 15285)	0.002374

TABLE VII
CONVERGENCE TIME WITH 95% CONFIDENCE INTERVAL AND ERROR FOR $(p_R, p_U) = (0, 7, 0.2)$ UNDER VARYING NUMBER OF SENSORS FOR THE CASE OF DECEPTIVE ENVIRONMENT.

VI. CONCLUSION

In this paper, we study the problem of sensor type inference without knowledge of the ground truth in a stochastic environment. We present a game-theoretic-based solution to the problem based on the theory of Learning Automata. We show that by carefully designing the utility function of the game, the optimal solution to our sensor reliability evaluation problem corresponds to the pure Nash equilibria of the game. The advantage of the current work compared to the literature is the fact it does not require the condition used in the literature that according to which truth prevails over lies. Thus, our solution converges even under deceptive environments. We provide sound theoretical results that prove the convergence of our scheme. Our experimental results are in concordance with our theoretical findings. Our work constitutes some of the limited attempts in literature to apply game theory to the field of information fusion. Therefore, we hope that this study can fuel more research interest in the applications of game theory to sensor fusion. As a future work, we would like to investigate extending our work by taking into account a static reliability as suggested in [15] that can be extracted during a training phase.

REFERENCES

[1] R. Gravina, P. Alinia, H. Ghasemzadeh, G. Fortino, Multi-sensor fusion in body sensor networks: State-of-the-art and research challenges, Information Fusion 35 (2017) 68–80.

[2] S. Sun, H. Lin, J. Ma, X. Li, Multi-sensor distributed fusion estimation with applications in networked systems: A review paper, Information Fusion 38 (2017) 122 – 134. doi:https://doi.org/10.1016/j.inffus.2017.03.006.

[3] X. Wang, J. Zhu, Y. Song, L. Lei, Combination of unreliable evidence sources in intuitionistic fuzzy mcdm framework, Knowledge-Based Systems 97 (2016) 24–39.

[4] H. Jiang, R. Wang, J. Gao, Z. Gao, X. Gao, Evidence fusion-based framework for condition evaluation of complex electromechanical system in process industry, Knowledge-Based Systems 124 (2017) 176–187.

[5] V. K. Solanki, M. Venkatesan, S. Katiyar, Conceptual model for smart cities: Irrigation and highway lamps using iot., IJIMAI 4 (3) (2017) 28–33.

[6] J. Frolik, M. Abdelrahman, P. Kandasamy, A confidence-based approach to the self-validation, fusion and reconstruction of quasi-redundant sensor data, IEEE Transactions on Instrumentation and Measurement 50 (6) (2001) 1761–1769.

[7] T. Tian, S. Sun, N. Li, Multi-sensor information fusion estimators for stochastic uncertain systems with correlated noises, Information Fusion 27 (2016) 126–137.

[8] R. Polikar, Ensemble learning, in: Ensemble machine learning, Springer, 2012, pp. 1–34.

[9] M. J. del Moral, F. Chiclana, J. M. Tapia, E. Herrera-Viedma, A comparative study on consensus measures in group decision making, International Journal of Intelligent Systems 33 (8) (2018) 1624–1638. doi:10.1002/int.21954.

[10] I. J. Pérez, F. J. Cabrerizo, S. Alonso, E. Herrera-Viedma, A new consensus model for group decision making problems with non-homogeneous experts, IEEE Transactions on Systems, Man, and Cybernetics: Systems 44 (4) (2014) 494–498.

[11] T. Denœux, S. Li, S. Sriboonchitta, Evaluating and comparing soft partitions: An approach based on dempster–shafer theory, IEEE Transactions on Fuzzy Systems 26 (3) (2018) 1231–1244.

[12] M. Liggins II, D. Hall, J. Llinas, Handbook of multisensor data fusion: theory and practice, CRC press, 2017.

[13] H. Jin, X. Chen, J. Yang, H. Zhang, L. Wang, L. Wu, Multi-model adaptive soft sensor modeling method using local learning and online support vector regression for nonlinear time-variant batch processes, Chemical Engineering Science 131 (2015) 282–303.

[14] A. Yazidi, B. J. Oommen, M. Goodwin, On solving the problem of identifying unreliable sensors without a knowledge of the ground truth: The case of stochastic environments, IEEE Transactions on Cybernetics 47 (7) (2017) 1604–1617.

[15] H. Guo, W. Shi, Y. Deng, Evaluating sensor reliability in classification problems based on evidence theory, IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 36 (5) (2006) 970–981.

[16] D. Yong, S. WenKang, Z. ZhenFu, L. Qi, Combining belief functions based on distance of evidence, Decision support systems 38 (3) (2004) 489–493.

[17] X. Deng, D. Han, J. Dezert, Y. Deng, Y. Shyr, Evidence combination from an evolutionary game theory perspective, IEEE transactions on cybernetics 46 (9) (2015) 2070–2082.

[18] C. K. Murphy, Combining belief functions when evidence conflicts, Decision support systems 29 (1) (2000) 1–9.

[19] P. J. Boland, Majority systems and the condorcet jury theorem, The Statistician (1989) 181–189.

[20] A. Yazidi, E. Herrera-Viedma, A new methodology for identifying unreliable sensors in data fusion, Knowledge-Based Systems 136 (2017) 85–96.

[21] M. Agache, B. J. Oommen, Generalized pursuit learning schemes: New families of continuous and discretized learning automata, IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics 32 (6) (2002) 738–749.

[22] D. Tian, J. Zhou, Z. Sheng, M. Chen, Q. Ni, V. C. Leung, Self-organized relay selection for cooperative transmission in vehicular ad-hoc networks, IEEE Transactions on Vehicular Technology 66 (10) (2017) 9534–9549.

[23] H. Cao, J. Cai, Distributed opportunistic spectrum access in an unknown and dynamic environment: A stochastic learning approach, IEEE Transactions on Vehicular Technology 67 (5) (2018) 4454–4465.

[24] H. Cao, J. Cai, Distributed multiuser computation offloading for cloudlet-based mobile cloud computing: A game-theoretic machine learning approach, IEEE Transactions on Vehicular Technology 67 (1) (2018) 752–764.

[25] A. Hajijamali Arani, M. J. Omidi, A. Mehbodniya, F. Adachi, A distributed learning–based user association for heterogeneous networks, Transactions on Emerging Telecommunications Technologies 28 (11) (2017) e3192.

[26] X. Deng, W. Jiang, J. Zhang, Zero-sum matrix game with payoffs of dempster-shafer belief structures and its applications on sensors, Sensors 17 (4) (2017) 922.

[27] B. J. Oommen, G. Raghunath, B. Kuipers, Parameter learning from stochastic teachers and stochastic compulsive liars, IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 36 (4) (2006) 820–834.

[28] K. S. Narendra, M. A. L. Thathachar, Learning Automata: An Introduction, Prentice-Hall, New Jersey, 1989.

[29] A. S. Poznyak, K. Najim, Learning Automata and Stochastic Optimization, Springer-Verlag, Berlin, 1997.

[30] M. A. L. Thathachar, P. S. Sastry, Networks of Learning Automata: Techniques for Online Stochastic Optimization, Kluwer Academic, Boston, 2003.

[31] M. L. Tsetlin, Automaton Theory and the Modeling of Biological Systems, Academic Press, New York, 1973.

[32] B. Tung, L. Kleinrock, Distributed control methods, in: High Performance Distributed Computing, 1993., Proceedings the 2nd International Symposium on, IEEE, 1993, pp. 206–215.

[33] B. J. Oommen, M. Agache, Continuous and discretized pursuit learning schemes: Various algorithms and their comparison, IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics 31 (2001) 277–287.

[34] M. A. L. Thathachar, B. J. Oommen, Discretized reward-inaction learning automata, Journal of Cybernetics and Information Science (1979) 24–29.

[35] L. Mason, An optimal learning algorithm for s-model environments, IEEE Transactions on Automatic Control 18 (5) (1973) 493–496.

[36] S. Misra, B. J. Oommen, GPSPA: A new adaptive algorithm for maintaining shortest path routing trees in stochastic networks, International Journal of Communication Systems 17 (2004) 963–984.

[37] B. J. Oommen, T. D. Roberts, Continuous learning automata solutions to the capacity assignment problem, IEEE Transactions on Computers C-49 (2000) 608–620.

[38] A. F. Atlassis, N. H. Loukas, A. V. Vasilakos, The use of learning algorithms in ATM networks call admission control problem: A methodology, Computer Networks 34 (2000) 341–353.

[39] A. F. Atlassis, A. V. Vasilakos, The use of reinforcement learning algorithms in traffic control of high speed networks, Advances in Computational Intelligence and Learning (2002) 353–369.

[40] A. V. Vasilakos, M. P. Saltouros, A. F. Atlassis, W. Pedrycz, Optimizing QoS routing in hierarchical ATM networks using computational intelligence techniques, IEEE Transactions on Systems, Man and Cybernetics: Part C 33 (2003) 297–312.

[41] M. Barzohar, D. B. Cooper, Automatic finding of main roads in aerial images by using geometric-stochastic models and estimation, IEEE Transactions on Pattern Analysis and Machine Intelligence 7 (1996) 707–722.

[42] R. L. Cook, Stochastic sampling in computer graphics, ACM Trans. Graph. 5 (1986) 51–72.

[43] A. Yazidi, O.-C. Granmo, B. J. Oommen, Service selection in stochastic environments: a learning-automaton based solution, Applied Intelligence 36 (3) (2012) 617–637.

[44] S. Misra, P. V. Krishna, K. Kalaiselvan, V. Saritha, M. S. Obaidat, Learning automata-based qos framework for cloud iaas, IEEE Transactions on Network and Service Management 11 (1) (2014) 15–24.

[45] A. Yazidi, O.-C. Granmo, B. J. Oommen, M. Gerdes, F. Reichert, A user-centric approach for personalized service provisioning in pervasive environments, Wireless Personal Communications 61 (3) (2011) 543–566.

[46] S. M. Vahidipour, M. R. Meybodi, M. Esnaashari, Learning automata-based adaptive petri net and its application to priority assignment in queuing systems with unknown parameters, IEEE Transactions on Systems, Man, and Cybernetics: Systems 45 (10) (2015) 1373–1384.

[47] P. Nicopolitidis, G. I. Papadimitriou, A. S. Pomportsis, Learning automata-based polling protocols for wireless lans, IEEE Transactions on Communications 51 (3) (2003) 453-463.

[48] P. Nicopolitidis, G. I. Papadimitriou, A. S. Pomportsis, P. Sarigiannidis, M. S. Obaidat, Adaptive wireless networks using learning automata, IEEE Wireless Communications 18 (2) (2011) 75–81.

[49] M. S. Obaidat, G. I. Papadimitriou, A. S. Pomportsis, H. S. Laskaridis, Learning automata-based bus arbitration for shared-edium ATM switches, IEEE Transactions on Systems, Man, and Cybernetics: Part B 32 (2002) 815–820.

[50] O.-C. Granmo, B. J. Oommen, Solving stochastic nonlinear resource allocation problems using a hierarchy of twofold resource allocation automata, IEEE Transactions on Computers 59 (4) (2010) 545–560.

[51] P. Sastry, V. Phansalkar, M. Thathachar, Decentralized learning of nash equilibria in multi-person stochastic games with incomplete information, IEEE Transactions on systems, man, and cybernetics 24 (5) (1994) 769–777.

Anis Yazidi received the M.Sc. and Ph.D. degrees from the University of Agder, Grimstad, Norway, in 2008 and 2012, respectively. He was a Researcher with Teknova AS, Grimstad. He is currently a Full Professor with the Department of Computer Science, Oslo Metropolitan University (OsloMet), Oslo, Norway, where he is leading the research group in applied artificial intelligence. His current research interests include machine learning, learning automata, stochastic optimization, and autonomous computing.

Hugo Hammer received the M.Sc. and Ph.D. degrees from the Norwegian University of Science and Technology, Trondheim, Norway, in 2003 and 2008, respectively. He holds a position of an Associate Professor in statistics with the Department of Compute Science, Oslo Metropolitan University (OsloMet), Norway. His current research interests include computer intensive statistical methods, Bayesian statistics, and learning systems.

Konstantin Samouylov received his PhD in probability theory from the Moscow State University, in 1985, and a Full Doctor of Sciences degree in telecommunications from the Moscow Technical University of Communications and Informatics, in 2005. During 19851996 he held several positions at the Faculty of Science of the Peoples Friendship University of Russia (RUDN University) where he became a head of Telecommunications System Department in 1996. Since 2014 he is a head of the Applied Probability and Informatics Department, and since 2017 he also holds the position of Director of Applied Mathematics and Communications Technology Institute (IAM and CT) at the RUDN University. He was visiting professor/professor-research at Lappeenranta University of Technology and Helsinki University of Technology (Aalto), Finland; Moscow Technical University of Telecommunications and Informatics, Russia; Moscow International Higher Business School (Mirbis), Russia; University of Pisa, Italy. He was a member of the ITU-T SG11 and IFIP TC6 WG 6.7. He has worked in a number of Research and Development projects within different frameworks, e.g., COST IRACON, projects of Russian Foundation for Basic Research (RFBR), TEKES (Finland) and companies including Nokia, Telecom Finland, VTT, Rostelecom, etc. He is a member of editorial boards and reviewer of several scientific magazines, he is co-chair and TPC member of several international conferences. His current research interests include various aspects of probability theory, stochastic processes and queuing systems, teletraffic theory, performance analysis of 4G/5G networks, resource allocation in heterogeneous wireless networks, social networks and big data analysis. He has authored and co-authored over 150 scientific and conference papers and six books. Prof. Samouylovs honors include the 2018 IEEE GLOBECOM Conference Best Paper Award.

Enrique Herrera-Viedma received the M.Sc. and Ph.D. degrees in computer science from the University of Granada, Granada, Spain, in 1993 and 1996, respectively. He is currently a Professor of computer science and the Vice-President for Research and Knowledge Transfer with the University of Granada. He has been identified as one of the world's most influential researchers by the Shanghai Center and Thomson Reuters/Clarivate Analytics in both computer science and engineering in the years 20142018. His current research interests include group decision making, consensus models, linguistic modeling, aggregation of information, information retrieval, bibliometric, digital libraries, web quality evaluation, recommender systems, and social media. Dr. Herrera-Viedma is the Vice-President for Publications in SMC Society and an Associate Editor in several journals, such as the IEEE Transactions on Fuzzy Systems, the IEEE Transactions on Systems, Man, and Cybernetics: Systems, Information Sciences, Applied Soft Computing, Soft Computing, Fuzzy Optimization and Decision Making, International Journal of Fuzzy Systems, Journal of Intelligent & Fuzzy Systems, Engineering Applications of Artificial Intelligence, Journal of Ambient Intelligence and Humanized Computing, International Journal of Machine Learning and Cybernetics, and Knowledge-Based Systems.