2016 American Control Conference (ACC)
Boston Marriott Copley Place
July 6-8, 2016. Boston, MA, USA

Distributed Energy-Aware Mobile Sensor Coverage: A Game Theoretic Approach

Evripidis Paraskevas, Dipankar Maity and John S. Baras

***Abstract*— Mobile Sensor Networks (MSN) are used to monitor large areas and collect measurements, e.g. temperature, humidity, or pressure data. Mobile sensors try to optimize their benefit from sensing a particular area, while keeping their energy consumption at the lowest possible level. We formulate this energy-aware sensor coverage problem as a potential game, where the mobile sensor nodes are considered as agents. The utility function of the game captures the trade-off between the benefit from coverage and the energy consumption. We also propose a distributed learning strategy for this potential game. The algorithm enables a bit-valued information exchange between the agents. Finally, it is proved that the learning rule converges to a Nash Equilibrium.**

I. Introduction

Efficient monitoring of large areas has attracted great interest from the research community in the recent past. Monitoring incorporates the collection and process of measurements related to temperature, humidity and other quantities of interest. For this purpose, mobile sensor networks (MSN) are being used to cover large areas under surveillance. Examples include coverage via large video cameras, environmental monitoring, monitoring for threats, monitoring for transportation congestion and efficiency, monitoring for medical and other emergencies. Sensor networks are also used in smart grid technologies. Mobile sensors have limited memory and energy and this needs to be taken into consideration while deploying such a sensor network.

In this paper, we investigate the mobile sensor coverage problem, while addressing the energy limitations of such types of networks. Mobile nodes aim to collectively optimize a global objective by making optimal local decisions. The objective is to maximize the benefits from coverage in a particular area and simultaneously minimize energy consumption due to sensing.

The energy-aware coverage problem is modeled using a game theoretic approach, where all the mobile sensor nodes participate in a game. Similar approaches, which use game theory for coverage problems have been presented in [1], [2], [3]. The goal of this approach is to introduce an efficient learning rule that ensures the existence of a pure Nash Equilibrium (NE) [4] and guarantees the convergence to NE.

We propose a distributed learning rule, based on [5], which enables a bit-valued information exchange over a communication graph. Our learning strategy focuses on optimizing the welfare of the game (i.e. sum of individual agents' utilities) for any arbitrary design of utility functions. We show that our algorithm induces a perturbed Markov chain and moreover, we prove convergence to a pure NE.

In our paper we present three main contributions in this problem. Firstly, we design a suitable utility function to capture the trade-off between sensing/processing and energy consumption. Secondly, we propose a distributed learning rule that enables bit-valued communication between the agents. Finally, we analyze and prove the convergence to a NE. Our proposed distributed learning algorithm also supports the case of continuous action space, which is essential for the mobile sensor coverage problem and it remained unaddressed in the vast majority of the past work.

The rest of the paper is organized as follows. Section II provides a concise literature survey. Section III describes the relevant preliminaries on game theory and perturbed Markov chains for the energy-aware coverage problem. Section IV describes the problem statement and gives an introduction to our approach. Section V contains the description of the distributed learning algorithm, the analysis of the proposed algorithm, and the proof of convergence. Finally, we conclude our work in Section VI.

II. Related Work

Several approaches have been proposed for the sensor coverage problem including some recent advancements [2], [3]. In [6], an optimization problem is defined to maximize sensors' coverage while taking into account the communication cost. Another method for sensor coverage has been introduced in [7], where the authors propose an estimation model consisting of a summation of n distributions and the estimation algorithm adjusts the weighting functions of these distributions. However, the proposed estimation scheme does not scale satisfactorily with n.

Game theoretic approaches have also been widely adopted to optimize sensor coverage problems. The approach was introduced in [8] and [9] and has been used for solving decision making problems in a multi-agent setup. Martinez et al. [2] propose a game-theoretic approach for distributed coverage using a mobile sensor network and they introduce a novel utility function that captures the trade-off between efficient coverage and energy consumption. [3] studies a sensor coverage potential game, using reinforcement learning, with a utility function that takes into consideration energy consumption due to sensing and movement.

* The first two authors contributed equally to the work.

Work partially supported by NSF grant CNS-135655, DARPA through ARO grant W911NF1410384 and Vencore (ACS) through grant FA8750C150038.

The authors are with the Department of Electrical and Computer Engineering and Institute for Systems Research, University of Maryland, College Park, USA. (email: {evripar, dmaity, baras}@umd.edu)

A variety of decentralized learning rules have been proposed for optimal action selection. Some of them are independent of the utility design, but are still proved to lead to a pure NE. We propose a similar learning rule in our work, which is optimal for a broad class of utility functions. In [10], Marden et al. proposed a decentralized learning algorithm to address the issue of unknown payoff structure. The proposed algorithm allows agents to learn actions that lead to welfare maximization without any knowledge of the functional form of the utilities. The algorithm can be used for optimization of complex systems with many distributed components, such as the routing of data packets in networks and the design and control of wind farms. However, the convergence of this algorithm is guaranteed only under an assumption on the form of the utilities called "interdependence". To overcome this, new decentralized learning rules were proposed by Menon et al. in [5] and [11]. The learning rule in [5] uses a bit-valued inter-agent communication and is proved to converge to a NE, without the interdependence property for the utility functions. In our paper, we modify this distributed learning algorithm to model the energy-aware mobile sensor coverage game and derive new (relaxed) conditions for convergence to NE. We have introduced a variation of this algorithm for continuous action space, which is then partitioned into a finite set of states.

III. Preliminary Background

A. Game Theory Background

In this section, we provide some basic definitions from game theory [2], [4], [12] that we use for our model. Based on these principles, we formulate the energy-aware optimal coverage problem as an exact potential non-cooperative game among the sensor nodes.

Definition 3.1: A strategic game $\Gamma =< \mathcal{V}, \mathcal{A}, \mathcal{U} >$ consists of:

1) A set $\mathcal{V}$ of heterogeneous players, where $i \in \mathcal{V} = \{1, \ldots, N\}$.
2) An action set $\mathcal{A} := \prod_{i=1}^{N} \mathcal{A}_i$, the space of all actions, where $\alpha_i \in \mathcal{A}_i$ is the action of player i and an (multiplayer) action $\alpha \in \mathcal{A}$ has components $\alpha_1, \ldots, \alpha_N$.
3) The utility function $U_i : \mathcal{A} \to \mathbb{R}$, which models the payoff of player i over action profiles.

Definition 3.2: Let α_{-i} be the action profile of all the other players except i and $\mathcal{A}_{-i} = \prod_{j \neq i} \mathcal{A}_j$.

The notion of NE [12] is crucial in non-cooperative game theory setup and is defined as follows:

Definition 3.3: Consider the strategic game Γ. An action profile $\alpha^* := (\alpha_i^*, \alpha_{-i}^*)$ is a NE of the game Γ, if for all $i \in \mathcal{V}$ and for all $\alpha_i \in \mathcal{A}_i$ it holds that $U_i(\alpha^*) \geq U_i(\alpha_i, \alpha_{-i}^*)$.

An action profile corresponding to a NE indicates an action in which no player has benefit to deviate. Potential games constitute of an important class of strategic games, where the change in a player's utility caused by a deviation can be exactly measured by a chosen potential function.

Definition 3.4: The strategic game Γ is an exact potential game with potential function $\Phi : \mathcal{A} \to \mathbb{R}$, if for every $i \in \mathcal{V}$, for every $\alpha_{-i} \in \mathcal{A}_{-i}$, and $\forall\ \alpha_i, \alpha_i' \in \mathcal{A}_i$, it holds that

$$\Phi(\alpha_i, \alpha_{-i}) - \Phi(\alpha_i', \alpha_{-i}) = U(\alpha_i, \alpha_{-i}) - U(\alpha_i', \alpha_{-i}) \quad (1)$$

The objective of the multi-agent system is to collaboratively maximize the welfare function $W^* = max_{\alpha \in \mathcal{A}} W(\alpha)$, where $W(\alpha) = \sum_{i=1}^{N} U_i(\alpha)$.

B. Perturbed Markov Chains

In this section, we describe the definitions and the theory of perturbed Markov chains [5], [13]. Let $P(0)$ be the 1-step transition probability matrix of a Markov chain on a finite state space S. We refer to this chain as the *unperturbed chain*.

Definition 3.5: A regular perturbation of $P(0)$ consists of a stochastic matrix valued function $P(\epsilon)$ on some non-degenerate interval $(0, a]$ that satisfies, for all $x, y \in S$,

1) $P(\epsilon)$ is irreducible and aperiodic for all $\epsilon \in (0, a]$,
2) $\lim_{\epsilon \to 0} P_{x,y}(\epsilon) = P_{x,y}(0)$ and
3) if $P_{x,y}(\epsilon) > 0$ for some ϵ, then $\exists\ r(x, y) \geq 0$ such that $0 < \lim_{\epsilon \to 0} \epsilon^{-r(x,y)} P_{x,y}(\epsilon) < \infty$.

From the first condition in Def. (3.5) we conclude that there exists a unique stationary distribution $\mu(\epsilon)$, which satisfies $\mu(\epsilon)P(\epsilon) = \mu(\epsilon)$ for each $\epsilon \in (0, a]$. The other two conditions indicate how the perturbed chain converges to the unperturbed one as $\epsilon \to 0$.

Let $\mathcal{L} = \{f \in \mathcal{C}^\infty |\ f(\epsilon) \geq 0, f(\epsilon) = \sum_{i=1}^{L} a_i \epsilon^{b_i}$ for some $a_i \in \mathbb{R}, b_i \geq 0,\ Dom(f) = (0, \infty)\}$ for some large enough but fixed $L \in \mathbb{N}$, where $\mathcal{C}^\infty$ is the space of smooth functions.

We introduce some notation that will be helpful while stating the main result regarding perturbed Markov chains. The parameter $r(x, y)$ is called the *1-step transition resistance* from state x to y. Notice that $r(x, y) = 0$ holds only for the one step transitions $x \to y$ allowed under $P(0)$. A *path* $h(a \to b)$ from a state $a \in S$ to $b \in S$ is an ordered set $\{a = x_1, x_2, \ldots, x_n = b\} \subseteq S$, such that every transition $x_k \to x_{k+1}$ in the sequence has positive 1-step probability according to $P(\epsilon)$. The resistance of the path is define as

$$r(h) = \sum_{k=1}^{n-1} r(x_k, x_{k+1}) \quad (2)$$

Definition 3.6: For any two states x and y, the *resistance* from x to y is defined by $\rho(x, y) = \min\{r(h) |\ h(x \to y)$ is a path$\}$.

Definition 3.7: Given a subset $A \subset S$, its *co-radius* is given by $CR(A) = \max_{x \in S \setminus A} \min_{y \in A} \rho(x, y)$.

Hence, $\rho(x, y)$ can be defined as the minimum resistance over all possible paths starting at state x and ending at state y. The co-radius indicates the maximum resistance that must be overcome in order to enter it from outside. We extend the definition of resistance to include resistance between two subsets $S_1, S_2 \subset S$:

$$\rho(S_1, S_2) = \min_{x \in S_1, y \in S_2} \rho(x, y). \quad (3)$$

Since $P(\epsilon)$ is irreducible for $\epsilon > 0$, $\rho(S_1, S_2) < \infty$ for all $S_1, S_2 \subset S$.

Definition 3.8: A *recurrence or communication class* of a Markov chain is a non-empty subset of states $E \subseteq S$ such that for any $x, y \in E$, $\exists\ h(x \to y)$ and for any $x \in E$ and $y \in S \setminus E$, $\nexists\ h(x \to y)$.

Let us denote the recurrence classes of the unperturbed chain $P(0)$ as $E_1, ..., E_M$ and $E = \{E_1, E_2, \cdots E_M\}$. Let us consider a directed graph $\mathcal{G}_{RC}$ on the vertex set $\{1, ..., M\}$ with each vertex corresponding to a recurrence class. Let a *j-tree* be a spanning subtree in $\mathcal{G}_{RC}$ that contains a unique directed path from each vertex in $\{1, ..., M\} \setminus \{j\}$ to j and denote the set of all j-trees in $\mathcal{G}_{RC}$ by $\mathcal{T}^j_{RC}$.

Definition 3.9: The *stochastic potential* of a recurrence class E_i is

$$\gamma(E_i) = \min_{T \in \mathcal{T}^i_{RC}} \sum_{(j,k) \in T} \rho(E_j, E_k). \tag{4}$$

Let $\gamma^* = \min_{E_i} \gamma(E_i)$. Finally, we can state the main result regarding perturbed Markov chains, based on [13].

Theorem 3.10 ([13]): Let $E_1, ..., E_M$ denote the recurrence classes of the Markov chain $P(0)$ on a finite state space S. Let $P(\epsilon)$ be a regular perturbation of $P(0)$ and let $\mu(\epsilon)$ denote its unique stationary distribution. Then,

1) As $\epsilon \to 0$, $\mu(\epsilon) \to \mu(0)$, where $\mu(0)$ is a stationary distribution of $P(0)$ and
2) A state is stochastically stable i.e. $\mu_x(0) > 0 \Leftrightarrow x \in E_i$ such that $\gamma(E_i) = \gamma^*$.

IV. Problem Statement

In this paper, we analyze a scenario of a Mobile Sensor Network (MSN), where sensors are randomly deployed in an area and their task is to monitor this area. The tasks of the sensors are to optimally move in the area since not all parts of the area are equally valuable at all time. In addition, sensors are equipped with limited battery power that should be spent judicially throughout the monitoring procedure. Hence, mobile sensors aim at maximizing the payoff from sensing specific portions of the area, while minimizing the overall energy consumption. Since the ultimate task is global for all the sensors, they unanimously have to decide their actions, in order to perform the above mentioned trade-off. This multi-agent trade-off problem can be formalized as a multi-player game [2].

We define the action profile (in game theoretic sense) at time t by a_t and the corresponding action for the the i-th agent is $(a_t)_i$ in our game. The action in this scenario consists of selecting the position o_t and the radius of sensing r_t. We treat the sensors to be points in the space and they can sense a circular region centering itself for some prespecified radius. We consider a two-dimensional area, which is discretized into a lattice. Each square of the lattice has unit dimensions and is labeled with the coordinate of its center $p = (p^x, p^y)$. The collection of all squares of the lattice is denoted by P. The sensors can place themselves only at these lattice centers and they have the privilege to choose a sensing radius around them. As can be predicted the higher the radius the higher the energy expenditure to sense the region. The location of the sensor (agent) in our scenario is denoted by $(o_t)_i = ((p^x_t)_i, (p^y_t)_i) \in P$. The sensing area is defined by a disc with radius $(r_t)_i$ that takes values within a range $[r_{min}, r_{max}] \subset \mathbb{R}$. Each agent's action can be modeled as a tuple of the position and the radius. Hence, for agent i the action is denoted as $(a_t)_i := ((o_t)_i, (r_t)_i) \in (\mathcal{A}_{t-1})_i$, where $(\mathcal{A}_{t-1})_i$ is the available action set for agent i at time t. The action profile for all agents is $a_t = ((a_t)_1, \ldots, (a_t)_N) \in \mathcal{A} = \prod_{i=1}^N (\mathcal{A}_{t-1})_i$. The current available action set contains the time index $t-1$, since this set may be constrained based on the action chosen at time $t-1$, i.e. formally speaking $\mathcal{A}_{t-1} = \mathcal{A}(a_{t-1})$.

V. Game Theoretic Approach

In this section, the utility function for the game is formulated and we also describe a distributed algorithm, which can be used to achieve a NE for the underlying game.

A. Utility Design

As described before, the utility function should capture the trade-off between the effectiveness of sensor coverage and the energy consumption caused by sensing. At this point, suppose that we want to optimize a stochastic process $f : X \to \mathbb{R}$, distributed over the space X where the sensors are placed, with a given probability density function μ over the space X. The lattice structure P is constructed over this space X. In order to find the covered area by a sensor, we define a disk $D_i((o_t)_i, (r_t)_i)$ centered at $(o_t)_i$ with radius $(r_t)_i$ around the sensor-agent. Hence, the total area which is sensed by the sensors can be expressed as $X_{covered}(a_t) = (\bigcup_{i=1}^N D_i((o_t)_i, (r_t)_i) \bigcap P)$.

The coverage gain is expressed with the function $F(a_t)$ that is defined as

$$F(a_t) = \int_{X_{covered}(a_t)} f(\xi) \cdot \mu(\xi) d\xi \tag{5}$$

We also need to define the energy consumption caused by sensing, transmitting and receiving packets from other agents. The energy consumption due to sensing and reception is proportional to the covered region. This can be modeled by $E^{cons}_i = C_i((r_t)_i)^2$, where $C_i > 0$ is a constant that depend on the sensor.

The trade-off between the coverage gain and energy consumption for agent i is captured in U_i in the following way

$$U_i(\alpha_t) = F((\alpha_t)_i) - E^{cons}_i((\alpha_t)_i) \tag{6}$$

It can be shown that our game is indeed an exact potential game by introducing a potential function $\Phi(\alpha_t)$. Potential games have a key property that the extrema values of the potential function Φ correspond to a pure NE for the game.

Lemma 5.1: The formulated game is an exact potential game with potential function defined as

$$\Phi(\alpha_t) = \sum_{i=1}^N F((\alpha_t)_i) - \sum_{i=1}^N E^{cons}_i((\alpha_t)_i) \tag{7}$$

Proof: As stated in [3] and [14], a state-based potential game should satisfy

$$\Phi((\alpha_t')_i, (\alpha_t)_{-i}) - \Phi(\alpha_t) = U((\alpha_t')_i, (\alpha_t)_{-i}) - U_i(\alpha_t) \quad (8)$$

In order to verify the above property of potential games, we used the definition of Φ and U from Eq. (6) and (7)

$$\begin{aligned}
&\Phi((\alpha_t')_i, (\alpha_t)_{-i}) - \Phi(\alpha_t) = \\
&\sum_{j=1, j\neq i}^{N} [F((\alpha_t)_j, (\alpha_t)_{-j}) - C_j((r_t)_i)^2] + [F((\alpha_t')_i, (\alpha_t)_{-i}) \\
&- C_i((r_t')_i)^2] + \sum_{k=1}^{N} [-F((\alpha_t)_k, (\alpha_t)_{-k}) + C_k((r_t)_k)^2] \\
&= F((\alpha_t')_i, (\alpha_t)_{-i}) - F((\alpha_t)_i, (\alpha_t)_{-i}) - C_i[((r_t')_i)^2 - ((r_t)_i)^2] \\
&= U_i((\alpha_t')_i, (\alpha_t)_{-i}) - U_i((\alpha_t)_i, (\alpha_t)_{-i}) \quad (9)
\end{aligned}$$

where all terms are being eliminated except from agent i that changes action from $(\alpha_t)_i$ to $(\alpha_t')_i$. Hence, our game is proved to be an exact potential game. ■

Therefore, in our case, the objective is to maximize $W = \sum_{i=1}^{N} U_i$ and hence we seek for actions:

$$\mathcal{A}^* = \{\arg\max_{a \in \mathcal{A}} W(a)\}$$

In the following section, we state the distributed learning strategy for our energy-aware coverage game. However, it should be noted that the algorithm defined and analyzed in the following section is general for all similar multi-agent games and we use the sensor coverage problem as an direct application of the algorithm.

B. Distributed Learning Strategy

In this section, we describe the distributed learning strategy that is used to reach a NE. The distributed algorithm is based on the recent work [5], [10]. However several modifications were needed for our purpose. The algorithm incorporates a bit-valued agent interaction though a simple directed interaction graph. The interaction graph $\mathcal{G}_I$, as defined in [5], is a directed graph and consists of the set of agents, whose actions can affect the payoff of other agents. Our framework also consists of the communication graph $\mathcal{G}_c$, which is a directed graph representing the explicit information exchange between the agents. The directed edge (i, j) in $\mathcal{G}_c(a_t)$ indicates that agent i is able to send a message to agent j at time t, when the joint action a_t is chosen. We define the neighbors of agent i at time t for the action profile a_t in the communication graph as $\mathcal{N}_i(a_t)$.

We partition the continuous action space into finite number of states such that $R = \{r_i | i = 1, 2, .., k\}$ where r_i's are disjoint intervals within $[r_{\min}, r_{\max}]$ satisfying $\cup_{i=1}^{k} r_i = [r_{\min}, r_{\max}]$. Each agent selects the radius by a Gibbs distribution, given in Eq. (10). Let $(\mathcal{A}_{t-1})_i$ be the set of feasible actions for agent i at time t, $(\mathcal{A}_{t-1})_i^r$ denote the feasible components corresponding to the radius and $(\mathcal{A}_{t-1})_i^o$ correspond to the position of the mobile sensor.

The conditional probability for agent k choosing a radius from set r_j given the center of the sensor to be at o and the immediate past join action to be a_j is considered to be

$$p_k^t(r \in r_i | o, a_j) = \frac{1}{m(r_i)} \frac{\int_{r \in r_i} \epsilon_t^{-U_k(r, o, a_{-j})} dr}{\sum_{r_l \in (\mathcal{A}_{t-1})_k^r} \int_{r \in r_l} \epsilon_t^{-U_k(r, o, a_{-j})} dr} \quad (10)$$

where $m(r_i)$ is the measure of the set r_i i.e. the length of the corresponding interval in this case.

Using Mean-value-Theorem for integrals, each of the integrals in (10) can be represented as $\int_{r \in r_i} \epsilon^{U_k(r, o, a_{-j})} dr = m(r_i) \epsilon^{U_k(\hat{r}_i, o, a_{-j})}$; where $\hat{r}_i$ is an interior point of the interval r_i determined by the mean value theorem. Let us denote a finite set $\hat{R} = \{\hat{r}_1, \hat{r}_2, \cdots \hat{r}_k\}$. With slight abuse of notation, representing $U_k(a, b, c) = U_k^a$ for fixed b and c, we can write (10) as follows

$$p_k^t(r \in r_i | o, a_j) = \frac{\epsilon_t^{-U_k^{\hat{r}_i}}}{\sum_{r_j \in (\mathcal{A}_{t-1})_k^r} m(r_j) \epsilon_t^{-U_i^{\hat{r}_j}}} \quad (11)$$

Let $\hat{r}_* \in (\mathcal{A}_{t-1})_k^r \subseteq \hat{R}$ such that $U_i^{\hat{r}_*} = \max_{r_j \in (\mathcal{A}_{t-1})_k^r} U_i^{\hat{r}_j}$. It can be easily verified that,

$$\lim_{\epsilon \to 0} \frac{p_k^t(r \in r_i | o, a_j)}{\epsilon^{(U_i^{\hat{r}_*} - U_i^{\hat{r}_i})}} = \frac{1}{m(\hat{r}_*)} \quad (12)$$

This gives us the resistance between actions [Def. 3.5]. Since the game is a potential game, we can denote: $U_k^{\hat{r}_*} - U_k^{\hat{r}_i} = \Phi(\hat{r}_*) - \Phi(\hat{r}_i)$. Let us denote the $\max_{r_k \in (\mathcal{A}_{t-1})_i^r} \Phi(r_k) = \Phi^*$ and hence $U_i^{\hat{r}_*} - U_i^{\hat{r}_i} = \Phi^* - \Phi(\hat{r}_i)$. Note that Φ^* and $\Phi(\hat{r}_i)$ both depends on a_{-j} as well but to maintain brevity we suppress this information.

Each agent i is endowed with a state $(x_t)_i = [(a_t)_i, (m_t)_i]$ at time t, where $(a_t)_i$ corresponds to the action taken and $(m_t)_i$ is a $\{0, 1\}$-valued *mood* of the agent i at time t. As described in [5], $(m)_i = 1$ is defined as the *content* state and $(m)_i = 0$ is defined as the *discontent* state of the agent i. The collection of the states of all agents at time t is represented as $x_t = [a_t, m_t]$.

For a given state x, we denote the joint action by a^x and joint mood by m^x; and similarly the action and the mood of i-th agent is denoted by $(a^x)_i$ and $(m^x)_i$ respectively.

Let $\{\epsilon_t\}_{t \in \mathbb{N}}$ with $\lim_{t \to \infty} \epsilon_t = 0$ and constant $l > 0$, are pre-specified. The agent i performs the following rules sequentially to update its action when the joint action in the last step was a_{t-1}. The performance does not depend on the initialization of the algorithm and it can be initialized randomly.

Algorithm 5.2:
Start

Step 1: Receive $(m_{t-1})_j$ from all $j \in \mathcal{N}_i(t-1)$ i.e. the neighbors of i in $\mathcal{G}_c(t-1)$. Calculate the temporary *mood* $\tilde{m}_i$ as follows:

1) If $(m_{t-1})_j = 1 \ \forall j \in \{i\} \cup \mathcal{N}_i(t-1)$ set $\tilde{m}_i = 1$;

2) else set $\tilde{m}_i = 0$.

Step 2: Pick $(a_t)_i = (o_i, r_i)$ as follows:

$p_i^t(o, r|a_{t-1}) = p_i^t(o|a_{t-1})p_i^t(r|o, a_{t-1})$, and $p_i^t(r|o, a_{t-1})$ as given in (11). The choice of o is independent of a_{t-1} i.e. $p_i^t(o|a_{t-1}) = p_i^t(o)$.

1) If $\tilde{m}_i = 1$, pick o_i from $(\mathcal{A}_{t-1})_i^o$ according to the following rules:

$$p(o) = \begin{cases} 1 - \epsilon_t^l & \text{if } o = (o_{t-1})_i \\ \frac{\epsilon_t^l}{|(\mathcal{A}_{t-1})_i^o| - 1} & \text{otherwise} \end{cases} \tag{13}$$

2) Else if $\tilde{m}_i = 0$, pick o_i uniformly from $(\mathcal{A}_{t-1})_i^o$ i.e.

$$p(o) = \frac{1}{|(\mathcal{A}_{t-1})_i^o|} \tag{14}$$

Step 3: Measure the payoff

$$(U_t^{mes})_i = U_i((a_t)_i, (a_{t-1})_{-i}) \tag{15}$$

and we define $U_i^* = \max_{(a_t)_i}(U_t^{mes})_i$.

Step 4: Update the mood $(m_t)_i$ as follows:

1) if $\tilde{m}_i = 1$ and $((a_t)_i, (U_t^{mes})_i) = ((a_{t-1})_i, (U_{t-1})_i)$, set $(m_t)_i = \text{Ber}(1 - \epsilon_t^l)$;
2) if $\tilde{m}_i = 0$ or ($\tilde{m}_i = 1$ and $((a_t)_i, (U_t^{mes})_i) \neq ((a_{t-1})_i, (U_{t-1})_i)$) set $(m_t)_i = \text{Ber}(\epsilon_t^{U_i^* - (U_t^{mes})_i})$. where Ber$(\cdot)$ is the Bernoulli distribution.

Step 5: update $(U_t)_i \leftarrow (U_t^{mes})_i$

Step 6: Broadcast $(m_t)_i$ to the neighbors in $\mathcal{G}_c(t)$.

Stop

Thus the learning strategy induces a non-homogeneous perturbed Markov chain $P(\epsilon_t)$ with state space in $\mathcal{A} \times \{0, 1\}^N$. Let us denote $U_i^* - (U_t^{mes})_i$ by β_3^i which in general depends on the joint action a and hence on the state x. Sometimes, we will refer the same as $\beta_3^i(x)$ to explicitly show the dependence of β_3^i on x.

C. Algorithm Analysis

Let $\mathcal{E} = \left\{\frac{n(\epsilon)}{d(\epsilon)} |\ n(\epsilon), d(\epsilon) \in \mathcal{L} \text{ and } deg(n(\epsilon)) \geq deg(d(\epsilon))\right\}$, where $deg(f(\epsilon))$ is the lowest exponent of ϵ present in $f(\epsilon)$.

Proposition 5.3: The distributed learning Alg. (Alg. 5.2) induces a perturbed Markov chain.

Proof: Firstly, it is trivial to check that $\forall x, y \in S$ $\lim_{\epsilon \to 0} P_{x,y}(\epsilon) = P_{x,y}(0)$. This is a direct consequence of the fact that $P_{x,y}(\epsilon) \in \mathcal{E}$ for all $x, y \in S$.

Secondly, consider any state $x_{t-1} = [a_{t-1}, m_{t-1}]$ at time $t-1$; for an agent i, irrespective of the modes of itself and others, it can choose the same action $(a_t)_i = (a_{t-1})_i = [o_i, r_i]$ at time t with a probability at least $\min\{(1 - \epsilon_t^l), 1/|(\mathcal{A}_{t-1})_i^o|\}p(r_i|o_i, a_{t-1})$ [Alg. 5.2, Step 2]. Similarly the agent can choose any other action at time t with some probability strictly great than 0 (the exact lower bound on this probability can be calculated from step 2 of Alg. 5.2). The mood $(m_t)_i$ can be changed to 1 with probability at least $\min\{(1 - \epsilon_t^l), \epsilon_t^{\beta_3^i}\}$, and can be set to 0 with probability greater than $\min\{\epsilon_t^l, 1 - \epsilon_t^{\beta_3^i}\}$. Hence the chain is irreducible and aperiodic at the same time.

From the structure of the probabilities defined in (11) and the steps 1.3, 2.1 and 4 in the Alg. 5.2, it is clear that for every state $x, y \in S$, $P_{x,y}(\epsilon) \in \mathcal{E}$. Let, $P_{x,y}(\epsilon) = \frac{n(\epsilon)}{d(\epsilon)}$, where $n(\epsilon) = \sum_{i=0}^{L_n} \alpha_i^n \epsilon^{b_i^n}$; $\alpha_i^n \in \mathbb{R}$, $b_{i+1}^n > b_i^n \geq 0$, $L_n \in \mathbb{N}$ and similarly $d(\epsilon) = \sum_{i=0}^{L_d} \alpha_i^d \epsilon^{b_i^d}$; $\alpha_i^d \in \mathbb{R}$, $b_{i+1}^d > b_i^d \geq 0$ and $L_d \in \mathbb{N}$. Therefore $deg(n(\epsilon)) = b_0^n$ and $deg(d(\epsilon)) = b_0^d (\leq b_0^n)$. Hence, $\lim_{\epsilon \to 0} \epsilon^{deg(d(\epsilon)) - deg(n(\epsilon))} P_{x,y}(\epsilon) = \frac{\alpha_0^n}{\alpha_0^d}$. Therefore, $P_{x,y}(\epsilon)$ satisfies all the three properties of a perturbed Markov chain enlisted in definition 3.5. ■

Remark 5.4: A direct consequence of Proposition 5.3 is that $P(\epsilon)$ is a regular perturbation of $P(0)$ and $P(\epsilon)$ has a stationary distribution $\mu(\epsilon)$ that converges to $\mu(0)$ (a stationary distribution of $P(0)$) as $\epsilon \to 0$ [Theorem 3.10].

Definition 5.5: Let, $C^0 = \{x \in S|\ m^x = \mathbf{1},\ (a^x)_i = (o, r)\ s.t.\ r = \hat{r}_*(o, a^x)\}$ and $D^0 = \{x \in S|\ m^x = \mathbf{0},\ (a^x)_i = (o, r)\ s.t.\ r = \hat{r}_*(o, a^x)\}$ where $\hat{r}_*(o, a^x) = \arg\min_{r \in \hat{R}} p(r, o, (a^x)_{-i})$ and $p(r, o, (a^x)_{-i}) = \frac{\epsilon_t^{U_i(r, o, (a^x)_{-i})}}{\sum_{\hat{r}_n \in \hat{R}} m(r_n) \epsilon_t^{U_i(\hat{r}_n, o, (a^x)_{-i})}}$. $r_n \in R$ is the interval such that $\hat{r}_n \in r_n$.

Lemma 5.6 ([5]): If for every $a \in \mathcal{A}$, $\mathcal{G}_c(a) \cup \mathcal{G}_I(a)$ is strongly connected, the recurrence classes of the unperturbed chain $P(0)$ are D^0 and the singletons $z \in C^0$.

Proof: Setting $\epsilon_t = 0$ in the Alg. 5.2, we can easily notice that $m_{t-1} = \mathbf{0}$ implies $m_t = \mathbf{0}$. So D^0 is a recurrence class according to $P(0)$. Similarly, $m_{t-1} = \mathbf{1}$ implies all $(m_t)_i = 1$ by step 1 of the algorithm. The step 2.1 of Alg. 5.2 along with (11) ensures all the agents select their previous actions. Hence each element of C^0 is a separate recurrence class. ■

Lemma 5.7: Under the same assumption as in Lemma 5.6, for any $x, x' \in C^0$, $y, y' \in D^0$, and $z \in S \backslash (C^0 \cup D^0)$:

$$\rho(x, y) = kl, \tag{16}$$

$$\rho(y, x) = \sum_{i=1}^{N} \beta_3^i(x), \tag{17}$$

$$\rho(x, x') = l|\eta|, \text{s.t. } \eta = \{i : (o^x)_i \neq (o^{x'})_i\}, \tag{18}$$

$$\rho(y, y') = 0, \tag{19}$$

$$\rho(z, y) = 0, \tag{20}$$

Proof: Let k be the smallest number such that one can choose a set $I \subset \{1, 2, .., N\}$ of k agents in a way that $I \cup (\cup_{i \in I} \mathcal{N}_i)$ is the whole set of agents $\mathcal{V} = \{1, 2, \cdots, N\}$. To change from a state in C^0 to a state in D^0, the agents $i \in I$ should change their moods using either step 4.1 or the combination of steps 2.1 (changing action) and 4.2. Both of these changes incur the same resistance l. $\forall j \in \mathcal{N}_i$, $\tilde{m}_j = 0$ as soon as $m_i = 0$. Mood m_j can be changed to 0 via a zero resistance path by step 4.2. Therefore k such agents need to change their moods so that all the agents can change their moods and hence the new state belongs to D^0. Note that, the

change of the action $a = [o, r]$ under $\tilde{m}_i = 0$ can be done with zero resistance using step 2.2 and (11). This proves Eq. (16) and obviously $k \leq N$.

For a change of state from D^0 to any state in C^0, the actions can be selected via a zero resistance path as in step 2.2. Since all $\tilde{m}_i = 0$, m_i has to be made equal to 1 via step 4.2 with a cumulative resistance of $\sum_{i \in \mathcal{V}} \beta_3^i(x)$ and hence Eq. (17) is obtained.

For a change from $x \in C^0$ to $x' \in C^0$, if any agent i has its center o_i different from its previous value, it can make such a change in action via a path of resistance l by step 2.1 or it can change its mood with resistance l and then choose the action $a^{x'}$ with a zero resistance path using step 2.2, and finally change its mood with resistance β_3^i by step 3.2. However, for the latter case, since agent i's change of mood will affect $\tilde{m}_j \ \forall j \in \mathcal{N}_i$. Thus, the neighbors need a change from $\tilde{m}_j = 0$ to $m_j = 1$ by resistance β_3^j. Therefore, the minimum resistance for such a change will be used to adopt the former strategy, i.e. changing action using step 2.1, incurring a resistance l. By denoting $\eta = \{i : (o^x)_i \neq (o^{x'})_i\}$ and the cardinality of η by $|\eta|$, we arrive at Eq. (18), where $(o^x)_i$ is the center of the i-th agent at state x.

All the states in D^0 are accessible from one another under the unperturbed Markov chain $P(0)$ and Eq. (19) holds.

Note that the $\mathcal{G}_c(a) \cup \mathcal{G}_I(a)$ is strongly connected and we divide the agents into two groups $\mathcal{V}_0 = \{i |\ \tilde{m}_i = 0\}$ and $\mathcal{V}_1 = \{i |\ \tilde{m}_i = 1\}$. Due to the strong connectivity assumption, for all $i \in \mathcal{V}_1$, $\exists j \in \mathcal{V}_0$ such that there is a path from j to i. Therefore, agents in $\mathcal{V}_0$ can change their actions with 0 resistance (step 2.2) in a way that affects the utility of some $i \in \mathcal{V}_1$ and as a consequence $m_i = 0$ with zero resistance (step 4.2). Thus finally for all $i \in \mathcal{V}$, $m_i = 0$. This fact along with Eq. (19) implies (20). ■

Lemma 5.8: The stochastically stable set of states is $\{x_i \in C^0 |\ W(a^{x_i}) = W^*\}$.

Proof: The proof follows the similar line of thoughts as done in [5] by constructing the $j-trees$ (Def. 3.9) rooted at $\{x_i \in C^0 |\ W(a_i^x) = W^*\}$ and comparing it to the other $j - trees$ rooted at other nodes. However few difference should be noted here that:

1) An outward edge from D_0 to x_i has a resistance of 0 (Alg. 5.2 step 4.2). In [5], it was W^*.
2) The above fact required $l > W^*$ in [5] but we do not require any such constraint on l.

■

Theorem 5.9 (Main Result): *Let for every action $a \in \mathcal{A}$, $\mathcal{G}_c(a) \cup \mathcal{G}_I(a)$ be strongly connected. Let $x_t = [a_t, m_t]$ denotes the state of all agents at time t, then*

$$\lim_{t \to \infty} P(a_t \in \mathcal{A}^*) = 1$$

Proof: This Theorem is similar to Theorem 1 in [5]. Only difference in our theorem is that we have relaxed the condition $\sum_{t=1}^{\infty} \epsilon_t^{\kappa} = \infty$ where $\kappa = \min_{E_i \in E} CR(E_i)$ and E is the set of recurrence classes of $P(0)$. By careful observation, we can say that $\kappa = 0$. To show this, we proceed by finding the co-radius of $x_i \in C^0$ such that $W(a^{x_i}) = W^*$. Let us take $v \in D^0$, then $\rho(v, v') = 0$ by Lemma 5.7 for all $v' \in D_0$. Let us choose $v' = (a^{v'}, m^{v'})$ such that $a^{v'} = a^{x_i}$. Therefore, clearly $\rho(v', x_i) = 0$ [Alg. 5.2 step 4.2] and hence $\rho(v, x_i) = \rho(v, v') + \rho(v', x) = 0$. Now, if $v \in S \backslash (C^0 \cup D^0)$, then $\rho(v, v') = 0$ for all $v' \in D_0$ and since we already have proved that $\rho(v', x_i) = 0$ for all $v' \in D_0$, we can conclude $\rho(v, x_i) = 0$ for all $v \in S \backslash (C^0 \cup D^0)$. Therefore $CR(x_i) = 0$ and that implies $\kappa = 0$. ■

VI. Conclusion

In this paper, we present a game theoretic methodology to solve the energy-aware coverage problem for mobile sensor networks (MSN) in a decentralized fashion. The utility function captures the trade-off between the efficient coverage and the energy consumption due to sensing, receiving packets and localization. The decentralized learning algorithm incorporates the exchange of certain bit-valued information between the agents over a directed communication graph. Finally, we prove that this algorithm converges to a NE. However, unlike the previous work [5], the convergence of ϵ_t is not constrained and consequently the convergence to NE.

References

[1] J. Marden and A. Wierman, "Distributed welfare games with applications to sensor coverage," in *Proc. of the 47th IEEE Conference on Decision and Control (CDC)*, Dec. 2008, pp. 1708–1713.

[2] M. Zhu and S. Martinez, "Distributed coverage games for energy-aware mobile sensor networks," *SIAM Journal on Control and Optimization*, vol. 51, no. 1, pp. 1–27, 2013.

[3] S. Rahili and W. Ren, "Game theory control solution for sensor coverage problem in unknown environment," in *Proc. of the 53rd IEEE Annual Conference on Decision and Control (CDC)*, Dec. 2014, pp. 1173–1178.

[4] D. Fudenberg and D. K. Levine, *The Theory of Learning in Games*, ser. MIT Press Books. The MIT Press, June 1998, vol. 1, no. 0262061945.

[5] A. Menon and J. Baras, "A distributed learning algorithm with bit-valued communications for multi-agent welfare optimization," in *Proc. of the 52nd IEEE Annual Conference on Decision and Control (CDC)*, Dec. 2013, pp. 2406–2411.

[6] W. Li and C. Cassandras, "Distributed cooperative coverage control of sensor networks," in *Proc. of the 44th IEEE Conference on Decision and Control, and European Control Conference (CDC-ECC)*, Dec. 2005, pp. 2542–2547.

[7] M. Schwager, D. Rus, and J.-J. Slotine, "Decentralized, adaptive coverage control for networked robots," *Int. J. Rob. Res.*, vol. 28, no. 3, pp. 357–375, 2009.

[8] R. Gopalakrishnan, J. R. Marden, and A. Wierman, "An architectural view of game theoretic control," *SIGMETRICS Perform. Eval. Rev.*, vol. 38, no. 3, pp. 31–36, Jan. 2011.

[9] T. Alpcan, L. Pavel, and N. Stefanovic, "A control theoretic approach to noncooperative game design," in *Proc. of the 48th IEEE Conference on Decision and Control held jointly with the 28th Chinese Control Conference (CDC/CCC)*, Dec. 2009, pp. 8575–8580.

[10] J. Marden, H. Young, and L. Pao, "Achieving pareto optimality through distributed learning," in *Proc. of the 51st IEEE Annual Conference on Decision and Control (CDC)*, Dec. 2012, pp. 7419–7424.

[11] A. Menon and J. S. Baras, "Convergence guarantees for a decentralized algorithm achieving Pareto optimality," in *Proc. of the 2013 American Control Conference (ACC)*, June 2013, pp. 1932–1937.

[12] D. Fudenberg and J. Tirole, *Game Theory*. Cambridge, MA: MIT Press, 1991.

[13] H. P. Young, "The evolution of conventions," *Econometrica: Journal of the Econometric Society*, pp. 57–84, 1993.

[14] J. R. Marden, "State based potential games," *Automatica*, vol. 48, no. 12, pp. 3075–3088, Dec. 2012.