Partially Observable Markov Decision Processes (POMDPs) and Wireless Body Area Networks (WBAN): A Survey

Yahaya O. Mohammed and Uthman A. Baroudi *
King Fahd University of Petroleum and Minerals
Dhahran, Saudi Arabia
[e-mail: {mdonimisi,ubaroudi}@kfupm.edu.sa]
*Corresponding author

Received December 1, 2012; revised February 21, 2013; accepted March 19, 2013; published May 31, 2013

Abstract

Wireless body area network (WBAN) is a promising candidate for future health monitoring system. Nevertheless, the path to mature solutions is still facing a lot of challenges that need to be overcome. Energy efficient scheduling is one of these challenges given the scarcity of available energy of biosensors and the lack of portability. Therefore, researchers from academia, industry and health sectors are working together to realize practical solutions for these challenges. The main difficulty in WBAN is the uncertainty in the state of the monitored system. Intelligent learning approaches such as a Markov Decision Process (MDP) were proposed to tackle this issue. A Markov Decision Process (MDP) is a form of Markov Chain in which the transition matrix depends on the action taken by the decision maker (agent) at each time step. The agent receives a reward, which depends on the action and the state. The goal is to find a function, called a policy, which specifies which action to take in each state, so as to maximize some utility functions (e.g., the mean or expected discounted sum) of the sequence of rewards. A partially Observable Markov Decision Processes (POMDP) is a generalization of Markov decision processes that allows for the incomplete information regarding the state of the system. In this case, the state is not visible to the agent. This has many applications in operations research and artificial intelligence. Due to incomplete knowledge of the system, this uncertainty makes formulating and solving POMDP models mathematically complex and computationally expensive. Limited progress has been made in terms of applying POMPD to real applications. In this paper, we surveyed the existing methods and algorithms for solving POMDP in the general domain and in particular in Wireless body area network (WBAN). In addition, the papers discussed recent real implementation of POMDP on practical problems of WBAN. We believe that this work will provide valuable insights for the newcomers who would like to pursue related research in the domain of WBAN.

Keywords: Markov Decision Process, Biosensors, Wireless Sensor Networks, Wireless body area network (WBAN).

http://dx.doi.org/10.3837/tiis.2013.05.006

1. Introduction

The recent advancements in hardware miniaturization facilitate the realization of many applications especially in the medical and health fields. Recently, Wireless Body Area Networks (WBANs) attracted significant attention. It consists of small, intelligent devices attached on or implanted in the body which are capable of establishing a wireless communication link. These devices provide continuous health monitoring and real-time feedback to the user or medical personnel [1]. These networks typically have two types of devices: sensors and actuators. Sensors are used to acquire or measure certain conditions of the human body externally such as body temperature, or internally such as glucose level in the blood. On the other hand, the actuators are concerned in the next step to be taken after examining the recorded measurements. These two devices sometimes work in conjunction with each other. For example, for a diabetic patient who requires insulin injection, when the sensor measures the glucose level, the actuator can be equipped with a built-in reservoir and pump which administers the correct dose of insulin to be given based on the glucose level measurements by the sensor [1]. WBAN has numerous possible applications; some are shown in **Fig. 1**.

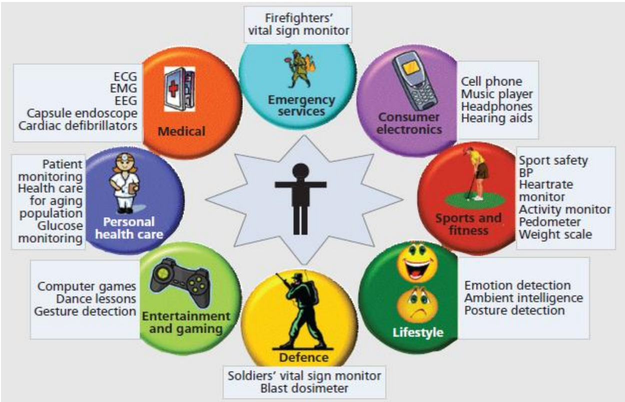

Fig. 1. WBAN Applications [9]

The Wireless Body Area Networks can be differentiated from the regular Wireless sensor networks by the following distinguished characteristics [1]. Firstly, the devices used are very small (less than 1cm^3) and have limited energy resources. Secondly, although the wanted lifetime is expected to extend to several years or even decades for implanted devices, for most devices, it is impossible to recharge or change the batteries. Thirdly, all devices are equally important and no redundant devices are available. Fourthly, to cope with health concerns and minimize interference, an extremely low transmit power per node is needed. Fifthly, the surrounding environment (i.e. human body) is a (very) lossy medium [2] and consequently, the waves are attenuated considerably before they reach the receiver. In addition, because the devices are located on the human body that can be in motion, the deployed WBANs should be robust against frequent changes in the network topology.

More importantly, WBAN should be very dependable [1][3] where high reliability, low delay and stringent security mechanisms are required. Finally, the devices operating in WBANs are often very heterogeneous and they may have very different demands or may require different resources of the network in terms of data rates, power consumption and reliability [1].

Having stated the above challenges, the literature is rich with works trying to find solutions for these problems. Actually, several papers have surveyed these works [4]-[9] The authors in [5] surveyed the existing and emerging wireless communication technologies and critical parameters for the system design. While the authors in [6] focused on surveying the recent energy-efficient medium access control (MAC) protocols for wireless body area networks (WBANs) along with a number of open research challenges with regard to prospects of medium access techniques and other issues. In [7], a survey of wireless Body Area Network (WBAN) was preseneted and the authors introduced the notion of a virtual doctor server (VDS) in existing WBAN architecture. This archticture supports various patient health care services and it keeps the historical data about the patient, generate the daily tips and advices for him, call the doctor or emergency squad if required and can provide first aid assistance instructions on patient or any of his close relative's PDA's. In [8][9], the authors surveyed the expected challenges in realizing WBANs for diverse applications.

However, according to the authors' knowledge, there is no published work that surveys the recent literature for new techniques or methods that solve the related optimization problems. One of the main challenges that face the WBAN designer is the high uncertainty in the status of the monitored system which leads to major difficulties in decision making process related to optimizing the activities of the implanted sensors, for example. Partially observable Markov decision process (POMDP) is a generalization of Markov decision processes (MDPs) that provide a mathematical framework for modeling decision-making in situations where outcomes are partly random and partly under the control of a decision maker. It provides a rich framework for agent planning under uncertainty [10][11]. Implantation of devices like biosensors in human body is a critical operational issue and application that requires efficient and accurate monitoring and management. MDPs have been applied in different facets of life such as medicine, education, quality control, scheduling, data communication, scheduling and so on [12]. In most of these domains, the full and exact state of the system is known. Specifically, amongst the many applications are optimizations of software in mobile phones [13]. The authors formulate the optimization of resources such as battery life in a mobile phone as a decision processes and proposed a dynamic programming techniques to solve the problem. In [14], the authors presented an optimal power and rate allocation control in multiple input and multiple output (MIMO) wireless system over Markovian fading channels. Earlier works for the use sensors for patient health monitoring include; lifeGuard [15] codeBlue [17], and Ubimon [18]. However, energy management and the environmental constraints are challenges facing these systems. To tackle this energy resource management problem, MDPs was introduced in the effort to design optimally performing system. In [19], MDPs was used to design MDP controllers for optimal energy utilization and information gathering in a health monitoring sensor network. However, the author concluded that POMDPs has greater representational power in the quest to manage this sort of wireless networks. Due to arrays of possible applications, various algorithms, methods and techniques have been used or proposed to solve this type of Markov decision processes.

1.1 A Motivating Problem

In this section, we present an example of WBAN problem that can be formulated and solved using POMDP. Consider the problem of implanting a subcutaneous biosensor in a human body. This biosensor in required to measure some physiological and metabolic vital signals such as temperature, pulse rate and glucose level. The operation and monitoring of these biosensors need to be accurate and efficient due to the critical nature of the application. The biosensor communicates wirelessly with a base station outside the body. The state of the system can be model as three variables vector (temperature, Energy and channel state which changes over time. The objective is to minimize the expected total energy consumption and maximize the number of measurements (or samples) subject to constraints such as the maximum temperature increase and quality of service. In order to study the above system, it is modeled as a POMDP since its state cannot be completely observed. The formulated model is solved using available techniques to obtain optimal policy, then these policies will be characterized and new heuristics are proposed based on reinforcement learning [20].

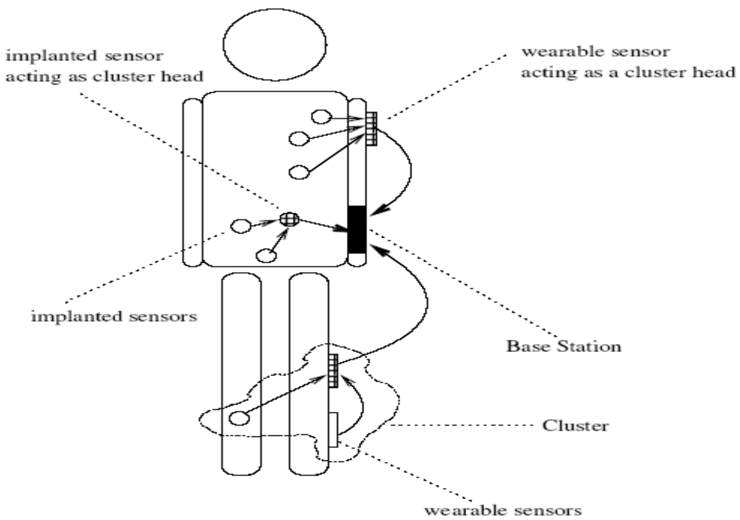

Fig. 2. Implanted and wearable biosensor in human body

There are limited research works on planning and management of biosensors, considering the increase in research for the design and development of biosensors and the possible increase in area of related applications. Therefore, there is a need for a comprehensive research on the optimization approaches of this class of wireless sensors. Most of the existing research efforts are dedicated to traditional wireless sensor network. Biosensors, in addition, to the general features of wireless sensor network, they have additional external factors from the environment such as temperature that can affect its performance. The background of this work is found in [22]. In this work, the author modeled the problem as Markov Decision Process (MDP) where the state of the system is assumed to be deterministic, but this is not the case in real life situations where decisions are made based on the observation of the current state which unknown apriori. The most suitable model is to use partial observable markov decision process.

The remaining part of this article is organized as follows; Section 2 described in detail, the POMDPs and some solutions techniques. In Section 3, the use of POMDPs framework in wireless sensor network is presented. Section 4 presented two recent real systems that employed POMDP for optimizing the perfromance of WBANs. An overview of applications of POMDPs in other fields of human endeavor was provided in section 5, while the challenges facing this framework is presented in section 6 and finally section 7 concluded the article.

2. Partially Observable Markov Decision Process (POMDPs)

A partially observable Markov decision process (POMDP) is a generalization of Markov decision processes that provide a mathematical framework for modeling decision-making in situations where outcomes are partly random and partly under the control of a decision maker. It provides a rich framework for agent planning under uncertainty [10][11].

A POMDP can be used to model an agent making a real world sequential decision after a set of actions under uncertainty to achieve a goal [11]. It is a model for deciding how to act in ``an accessible, stochastic environment with a known transition model" [23]. A discrete-time POMDP models the relationship between an agent and its environment. A POMDP is described by the following set of parameters:

- a set of states $S = \{s_1, s_2, s_3 \ldots, s_{|s|}\}$
- a set of actions $A = \{a_1, a_2, a_3 \ldots a_{|A|}\}$
- a set of observations $O = \{o_1, o_2, o_3, \ldots, o_{|o|}\}$
- a set of transition probabilities $T = (s_i, a, s_j) = P(s_j | s_i, a)$
- a set of observation probabilities $\beta(si, a, o) = p(oi | sj, a)$.
- a set of rewards $R: S \text{ X } A \rightarrow R$
- a discount factor $\gamma \in [0, 1]$
- an initial belief $b_o(s)$

Formally, a POMDP is specified as a tuple (S, A, O, T, β, R), where S is a set of states, A is a set of actions, and O is a set of observations. In each time step, the agent lies in some state s in S; it takes some action a in A and moves from s_i to a new state s_j. Due to the uncertainty in action, the end state s_i is modeled as a conditional probability function $T(s_i, a, s_j) = P(s_j | s_i, a)$ which models the effect of action and gives the probability that the agent transits to s_j, after taking action a from state s. The agent then relates an observation to information on its state. Due to the uncertainty in observation, the observation result o in O is also modeled as a conditional probability function $\beta(s, a, o) = p(o | s, a)$. At each step, the agent receives a real-value reward model $\beta(s, a)$, if it takes action a from state s, and the agent's goal is to maximize its expected total reward by choosing a suitable sequence of actions. A reward function R is defined as R: S x A ➔ R . This function is used to influence the agent's characteristics. In POMDP, the objective is to plan, planning means computing an optimal policy that maximizes the expected total discounted or long term reward over an infinite horizon. Since the state of the agent's state is not completely visible, that is, partially observable, this leads us to rely on belief b, which is a probability distribution over S. The initial belief is denoted as $b_o(s)$.

A POMDP policy π maps a belief b to a prescribed action a *in* A. The expected reward for policy π starting from belief b is defined as

$$J^{\pi}(b) = E[\sum_{t=0}^{\infty} \gamma^t r(s_t, a_t) | b, \pi] \quad (1)$$

Where $\gamma < 1$ is the discount factor. The optimal policy π^* is obtained by optimizing the long-term reward.

$$\pi^* = arg \max_{\pi} J^{\pi} (b_0) \quad (2)$$

Where b_0 is the initial belief. The optimal policy, noted π^* yields the highest expected reward value for each belief state, compactly represented by the optimal value function, noted V^*. The value function V^*(b) is a solution to Bellman optimality equation as follows.

$$V^*(b) = \max_{a \epsilon A}[r(b, a) + \gamma \sum_{o \epsilon O} \beta(o|b, a) V^*(\tau(b, ao))] \quad (3)$$

Exact solutions to POMDP problems are intractable, and various approximate solutions techniques were developed.

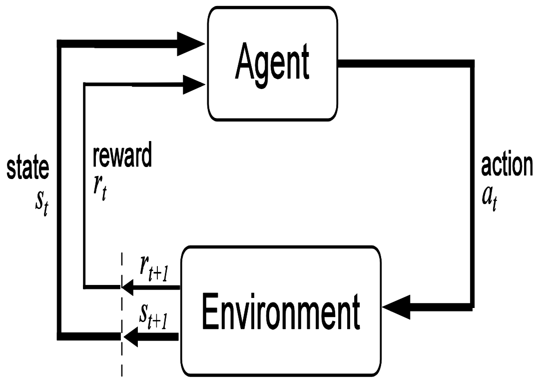

Fig. 3. The general interaction between agent and environment [15][21]

2.1 A review of Solution Techniques for POMDP

In this section, we present an overview of classical algorithms for solving POMDP and investigate the improvements so far in both exact and approximate methods. Solving a POMDP exactly is an intractable problem. Due to computational explosion of results, most efforts are tailored toward the management of the constraints to avoid the exponential complexity increase. The developed methods and algorithms can be classified under various categories such as exact value iteration, policy iteration, witness and greedy solution. Some of the approximate algorithms are either optimal or suboptimal. The entire algorithms would be based on the above classification. Also, most POMDP algorithms depend on dynamic programming; some efforts have been exerted in designing algorithms for updating the dynamic programming.

Classical methods for solving POMDP are value iteration [24], approximate (heuristic) solution techniques in POMDP aimed at reducing the complexity [25]. Also, a fast-point based algorithm for POMDP was proposed in [26], it ensures that in each iteration of value function, the new value functions are upper bounded. Another promising technique is approximate Linear programming for solving hybrid factored MDPs [27]. The first exact method for solving POMDP is called One pass [27], here, the algorithm chooses an arbitrary belief point and defines a set of constraints and it constructs vectors for each point. In this case, the constraint in the believe space is guaranteed to dominate. The region defines the intersection of three regions. A vector is constructed for that region and the strategy is best for the belief point and some nearby belief point.

In [12], an approximate algorithm called Enumeration algorithm that concentrates on how to generate sets of points by generating all possible sets of vectors by selecting action for each observation in the vectors. In this case, a large number of vectors will be generated and some might be useless. The useless ones can be eliminated to decrease the computation time. An extension to the enumeration algorithm is Incremental pruning [27] using interleaved generation and redundancy testing breaking up the dynamic programming update of the value function and the maximizing set of alpha vectors is determined from bottom to top leading to iterative purging of smaller set of alpha vectors. Grid based Algorithm [28] is another example of approximation algorithm in which the value function is determined for sets of points and interpolated to evaluate the optimal action that would be encountered for other belief state that are not in the grid point. Linear Support [29], is a modification of [10] idea by choosing less strict constraints and not focusing on actions and future course of action . It simply picks a point, generate vectors for the point and check if the region of the vectors is correct for all corners. Corners mean vertices of the region. In [30], the author presented a different idea called Witness algorithm by focusing on the point that the vector is not dominated, it concentrates on the best value function. Instead of finding V and prune it down. It focused on finding V* (optimal policy) directly.

In Point based Algorithms [31], a set of belief points are first sampled from the belief simplex (by letting the agent interact with the environment) and then planning is performed on these points only. In [32], a randomized version of point-based value iteration was presented. Here, the key idea is that, in each value iteration step, it can improve the value of all points in the belief set by only updating the value and its gradient of a subset of the points, and we select these belief points in a randomized greedy manner. In [33], similar to [26], it selects some representative points in the belief space and iteratively applies value update to those points.

2.2 Some Variations and Improvements of POMDPs Solution Methods

Most of the existing algorithms are based on traditional value function iterations algorithms that try to minimize each point in the belief space search. Researchers, in recent years in the quest for optimal solution continued to explore various approaches to prune down the belief space in order to reduce the computational cost. In this section, we present several of those enhancements and modifications tailored towards specific applications.

In [34], the authors, presented a Monte Carlo based approximation methods to POMDP. Their method does not require analytical tractability; the system is modular and can be treated as plug and play. It is inherently non-myopic allowing tradeoff between short terms for long term reward. This approximation was adopted for adaptive sensing. In [35], Region-based Dynamic programming reduced the linear programs in cross sum pruning, and hence reducing the computational complexity. This was achieved by dividing the belief space into smaller regions and perform independent pruning in each region.

In Parallel roll out [36], It is a novel approach proposed to generalize the roll out algorithm [37], by rolling out a set of policies rather than single policy. This aimed at the class of problems and forms a system path through which multiple heuristics policies are automatically combined to a new system policy to enhance performance. POMDPs algorithms operate with limited amount of memory which led to weak theoretical guaranty. In [38], they provided approach that addresses the space requirements and guaranty that optimal results can be maintained. Policy search algorithms [39][40][41][42] have also been devised to concentrate on the computation of belief states, these have succeeded to some large extent but the problem with policy search algorithm is that it could be data inefficient; many policy search algorithms have difficulty using data from its previous computed policies. But for policy compression [43][44][45] focuses on relevant belief space but also allows the use of relevant training data to estimate the effect of any policy.

3. POMDP and Wireless Body Area Network

Generally, sensors deployments require high cost of maintenance, energy conservation; hence, there is a need to optimize power to increase the total number of measurement over a given period of time. In most of these applications, the states are not fully observable, sometimes even are not observable at all. There are many factors affect the performance of sensors such as energy, efficiency, power etc. These factors by nature constitute random processes. Depending on application or user defined performance measure, each state may have different definition. For example, in measuring coverage, one can look at a sensor state and whether it is active or inactive, good or bad, on or off, etc. POMDP provides a rich mathematical structure for planning these types of problems under uncertainty. The cardinal issue in sensor scheduling is how to attain optimal tradeoff between the cost of sensor usage and the sensor performance and the goal of the system is to maximize user defined goals [46].

In order to minimize energy consumption, sensor networks require a scheduling scheme to reduce the number of active sensors at any given time while monitoring the quality of network [46]. In general, this is an optimization problem [47]. In [48], the sensors are partitioned into disjoint set, with each set expected to cover the target. They provide a heuristic to solve the mixed linear programming problem that arises from their model. But the challenges of these techniques were that it cannot be applied dynamically. Similarly, in [49], Voronoi diagrams were used to determine the hull points to place the sensors in order to maximize coverage and enhance battery life time. But in [41], self-scheduling of duty eligible rules was applied, where a node can turn itself on or off by observing it nearest neighbor. It turns itself off, if its nearest neighbor is active and provides an optimal coverage. In [50], the authors defined a threshold called k-cover (predefined constant) where a set of sensors less than *k* is determined through neighbor communication. In this approach, a central controller is assigned the duty of monitoring the coverage.

3.1 Biosensors

Biosensors can be described as a sensor that integrates a biological element with physiochemical transducers to produce an electronic signal proportional to a single analyte (substance or chemical constituent that is of interest such as glucose) which is then conveyed to a detector [51]. Figure 4 shows the block diagram for a biosensor. A biosensor can be implanted or worn. (See figure 5). This class of sensors has a wide range of potential applications such as in health care both (clinical and laboratory,) for measurement of metabolite, diabetes insulin therapy, drug administration and artificial pancreas in implantable glucose sensors. Industrial process control in the monitoring of bioreactor, and military applications are another domain of applications. For detail discussion on the applications of biosensor see [52].

The following paragaphs reviews the management of biosensor considering the environmental constraints. In [53], thermal aware routing protocol is proposed. This protocol is to balance the temperature amongst the biosensors, to avoid what the authors called “hot spot”. That is, if the temperature of a biosensor node exceeds a specified threshold. But the protocol, did not take into consideration the effect of the wireless channels and energy.

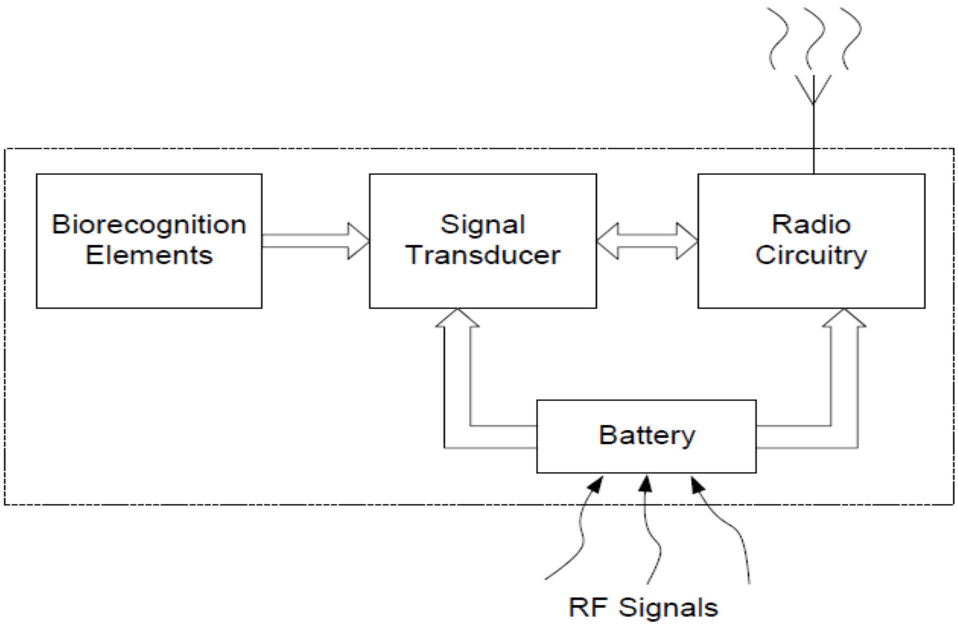

Fig. 4. Component of a biosensor **[22].**

In [54], the authors exploit the notion of leadership clusters to calculate the optimal temperature in an implanted biosensor network. They used the Penne's heat equation and finite domain time difference method to calculate the temperature increase and specific absorption rate. The scheme is referred to as Temperature Increase Potential (TIP). Genetic algorithm and TIP are combined for the computation of temperature increase. However, they did not incorporate the effect of the wireless channel and the energy in the model.

The management problem of rechargeable biosensor in a temperature sensitive environment like human body was considered in [55]. Optimal number of sample was considered with the constraint of effect of temperature on tissue and the heat generated while recharging the biosensor. The problem was formulated as Markov Decision Problem (MDP) with numerical and simulation result presented. Though, an interesting result was obtained, however, the formulation of the problem with MDP, with assumption that all states are deterministic does not address the real life situation of noisy sensors and channels. In [56], the authors modeled the dynamic scheduling of biosensors to manage the heat generated. The scheduling problem was modeled as MDP. Though, heat effect was considered in this work, but modeling the scheduling problem as MDP does not incorporate the noisy behavior of the sensor and the channels.

3.2 Implantable Wireless Body Sensor Networks

Wireless body sensor is a miniscule, subcutaneous sensors implanted around the tissue or organ to detect changes and communicate to a wireless device to alert the patient [53]. These sensors are implanted in human body to monitor biophysical activity (see Fig. 4). The implantation requirement brought about additional challenges in addition to the general requirement of wireless sensor networks. Each sensor node is limited in size, memory-storage constraint, power, communication capabilities. Furthermore, temperature and the properties of the environment that human body metabolic dynamics such as blood flow, heartbeat, and pulse rate will affect the performance of this type of sensors. Therefore, there is a need for optimizing its functionalities such as optimal scheduling for its activities, energy utilization, coverage, response time, temporal accuracy, and effective sample rate etc [58]. Also, there is a need to optimize related hardware issues such as size, power consumption, etc. In addition, implantable biosensor has additional requirements of simplicity, low detection limit, speed, precision, ruggedness, stability, room for improvement, utility and sensitivity. This field of research is still in its infancy with a lot of challenges and open problems [57].

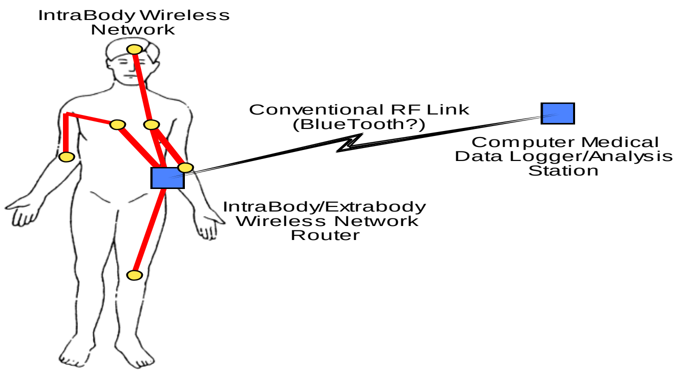

Fig. 5. Implanted biosensor in human body **[22].**

3.3 Problem Specific POMDP Models for Wireless Sensor Scheduling

In [46], the authors considered the problem of scheduling sensors deployment in an area to detect moving objects. In order to avoid waste of energy and maintain high coverage, only k-sensors out of n sensors are selected to work at a particular time. The object movements are modeled as a finite state Markov chain and assumed states can only be observed through noisy channels. Also, to be able to detect moving objects, their distances are reported to a central controller to do data fusion and trilateration. The following subsections describe in detail the POMDP modeling process. In each model, we need to define the states and transitions model, the observation model and the reward model.

3.3.1 The POMDP System Model

The model is presented in [46] as follows. The objective is to find an optimal policy that selects k sensors out of n deployed sensors while maximizing the coverage and the probability of object detection.

3.3.2 States and Transitions Model

The deployed sensors are modeled by *n* variables s_1, s_2, ...s_n. Each variable is assumed to have two states [*ON*,*OFF*], where *ON* refers to the working state and *OFF* refers to an idle state for energy saving. The initial state of a sensor is *ON*. The location of tracking object is represented by variable *X*. *X* is assumed to be independent of the other variables, as the moving object's path is not influenced by the state of the sensor network. Due to lack of knowledge about the car's intended path, it is assumed to follow a random waypoint motion pattern, in which the car either stay in its current position, or move to neighboring ones. The space is divided into grids and allows only one object to stay in a cell at each time. The location of the moving object is represented by the states of variable *X*. The action space is simply composed of the C_n^k possible assignments for selecting *k* sensors from *n* deployed sensors without order. The movement of object and sensor selection process is modeled as two Markov chains with one-step transition probability matrix:

$$B_i^a = \left(\emptyset_{pq}\right)_{p,q \in I_i} = pr(e_i^{t+1} = q | e_i^t = p) \tag{4}$$

where *a* stands for an action, *p* and *q* are two states belonging to the state space I_i.

3.3.3 Observation Model

GPS is used in wireless sensor network, to determine the location of object directly. By using the received signal strength (RSSI) a sensor can determine the distance to the object. The controller utilized the estimated distance to find the location of object using trilateration or other data fusion methods. The errors and noise in the detection process is as a result of the uncertainty in the positioning. In this case, they mapped such a positioning process to the observation model of the POMDP. It is also assumed that the observation space *O* has the same size as the state space *X*, where each observation $o \in O$ indicates an observation on the location of the moving object. For the detection probability, based on the detection techniques many detection coverage models have been proposed for different event scenarios. In this work, the author adopts the probabilistic coverage model. They considered a general signal propagation model where the signal parameter (e.g., the sound pressure of a sound source) fades along with the signal propagation Depending on the hypothesis of whether the target is present (H_1) or not (H_0), the readings at the sensor s_i are given by

$$H_0 : x_i = n_i$$

$$H_1 : xi = \frac{\theta}{d_i^{\propto}} + n_i \tag{2}$$

Where θ is the fading exponent, d_k^2. = $d\ \alpha$ (*si, x*) is the Euclidean distance between the sensor s_i and the space point. *x*, and n_i is the measurement noise. It is often assumed that the noise follows a Gaussian distribution with zero mean and variance δ_k^2. The overall detection probability by a set of sensors for position x *was computed* as follows using decision fusion:

$$p_x^i = 1 - \prod_{j=1}^{k}\left(1 - p_x^j\right) \tag{3}$$

Note that P_x^j depends on the distance between a sensor and the space point. The threshold can be set and defined such that a point is covered by K sensors if its overall detection probability is not less than a threshold. At this point, the probabilistic coverage model can be converting to a Boolean coverage model [46].

3.3.4 Reward Model

In [46], they considered two optimization objectives, detection probability and coverage. Detection probability refers to the concept that k sensors were chosen to minimize the uncertainty regarding the moving object's true location. Coverage means that in each simulation time the k active sensors can give a reasonable detection probability for the moving objects. Because of the probabilistic coverage model, in order to avoid an extremely small detection probability, a low boundary of detection probability was set to 0.1. The moving object will be considered as lost when the detection probability is lower than 0.1 and corresponding to the low boundary; the sensor range in the Boolean coverage model is set to 11 m. In order to maximize the coverage, a reward function that assigns a constant positive reward 1 when the current location of moving object is inside the sensing range of any selected k active sensors. Such a reward function will guide the POMDP to avoid losing the observation of moving object. By dividing the area into grids, the reward was determined as follows:

$$R_{cov} = \sum f(x)$$
$$f(x) = \begin{cases} 0, & p_x^i < \omega \\ 1, & otherwise \end{cases} \tag{4}$$

Where ω is the threshold of detection probability. When there are sufficient sensors covering the moving object, only k sensors having the highest detection probability will be selected to save the energy. A positive reward was defined as $R_{det}(x, \{si\})$ for a (grid, sensor) tuple $(x, \{si\})$ as following:

$$R_{det} = \sum\left(1 - \prod_{i=1}^{k}\left(1 - p_x^k\right)\right) \tag{5}$$

3.3.5 Discussions

In this section, we analyze the issues that arise from the formulation of the above sensor scheduling problem discussed above. For the sake of simplicity, the authors assumed the discrete case which is represented using Gaussian random variables. The space also, is divided into grids and only allows object to stay in a cell at each time. This assumption will minimize the search space but it affects the accuracy of the developed policies. Similarly, many the researchers opt to make the problem tractable by limiting the dynamics of these assumptions are made to reduce the general problem of POMDP to a simpler tractable problem of MDP.

4. POMDP in Real Implementation of WBAN

Having discussed the therotical basis for POMDP in the above sections, we shall present here two real expermintations [59][61] of employing POMDP to optimize the performance of WBANs.

4.1 The KNOWME WBAN [59][61]

The KNOWME network targets applications in pediatric obesity. The system is a multimodal system which is developed to track an individual's level of stress, physical activity, as well as other vital signs simultaneously. Such data must also be anchorable to context, such as time of day and geographical location, which provides a greater understanding of how the external environment impacts health. The KNOWME network is a fruit of a multiyear collaboration between communications engineers, computer systems designers, and preventive health researchers.

The KNOWME network is composed of a three-tier architecture as shown in Fig. 6. The WBAN layer or sensor layer is the first tier, which wirelessly provides physiological signals. The second tier is the mobile phone, and it acts as a data collection hub for the external sensor data. Initial processing is executed in the mobile phone and it provides simple feedback to the user instantly. A back-end server is the last tier and it can provide additional processing as well as data storage.

The problem is modeled as a partially observable Markov decision process (POMDP) to capture the system's sequential nature. POMDPs model the dynamics of physical state change via a Markov chain (Fig. 7b). Dynamic programming (DP) and greedy search strategies are employed to optimize the trade-off between classification error and energy cost, and significant energy gains are obtained. The POMDP framework is using realistic assumptions and the monitored time-evolving activity is known only through noisy observations. In addition, this system assumes that the fusion center (mobile phone) is energy constrained, and not the sensor nodes, and thus, the sensor selection process is optimized to meet this criterion. However, due to DP high complexity, a new approximate time-sharing sensor selection scheme based on cost functional properties was devised. The developed system was implemented and evaluated on real-world data collected from an implanted WBAN.

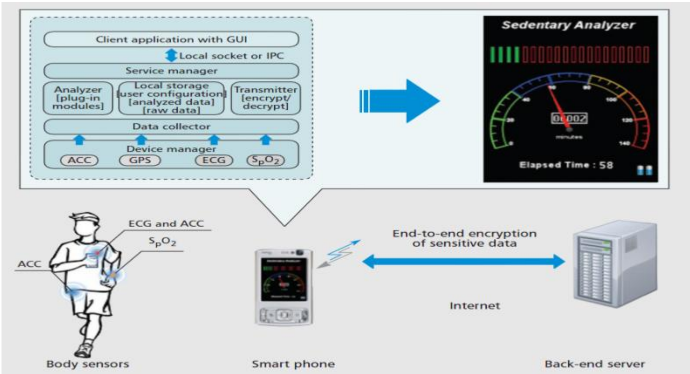

Fig. 6. The KNOWME system architecture flow diagram and screen shot from the sedentary behavior analyzer implemented on the mobile platform [61].

4.2 CARER [60]

CARER (Continuous Activity Recognition with Embedded Reasoning) is a software sensing support system that supports dynamic sensor scheduling algorithms by exploiting knowledge of contexts. CARER system makes real-time dynamic decisions as to whether to obtain additional sensor measurements for state classification.

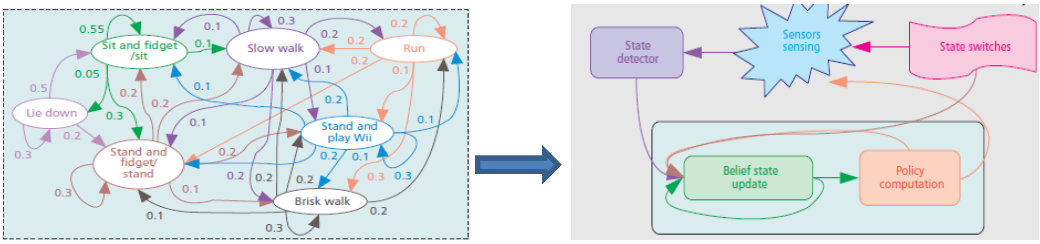

Fig. 7. a) System diagram of sensor selection algorithm; b) Markov chain of seven physical activities improved by using the output policy of POMDP [61]

To achieve this objectives, CARER employs Partially Observable Markov Decision Process (POMDP) to implement the sensors scheduling algorithm. POMDP provides a direct method for reducing energy demand while ensuring classification accuracy. The algorithm stochastically determines when a new measurement is necessary to estimate the activity in the next episode, and adjusts sensor usage accordingly.

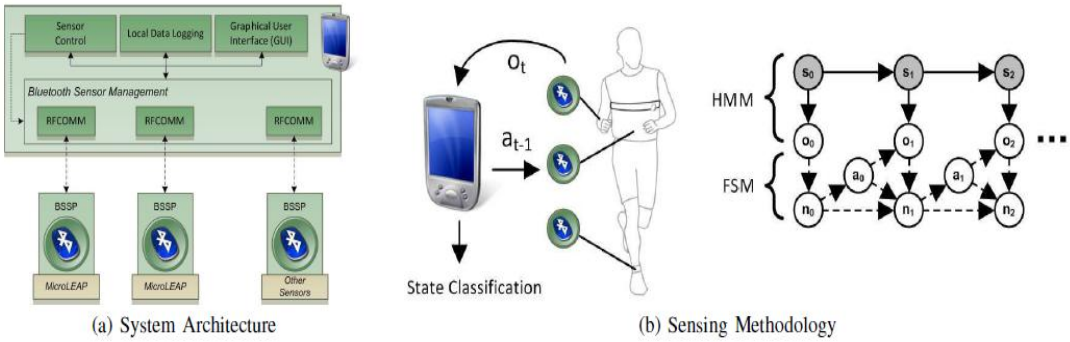

Fig. 8. CARER System **[60].**

4.2.1 POMDP Formulation

The process consists of a hidden Markov model (HMM) that models the activity transitions (e.g. sitting, walking) and a control policy based on finite-state machine (FSM) that controls the availability of Bluetooth sensors (see Fig. 8). The objective is to accurately detect activity patterns in different parts of the body given some resource constraint.

In POMDP formulation, the following spaces are defined: the State Space |S|, the Observation Space |Ω| and the Action Space, |A|. The State Space |S| contains the set of the activities of interest. The Observation Space, |Ω| contains the set of observation symbols extracted from sensor measurements. The Action Space, |A|, consists of the set of all available sensor controls. (e.g., activate and de-activate Bluetooth sensors). Sensor measurements are made at each discrete time step (or sensing epoch) t, sensor, given that the sensors are active (a_{t-1}= *activate*). At the end of the epoch, an observation symbol o_t is generated, and it is used to compute the hidden state s_t. Simultaneously, the FSM maintains an internal state n_t and generates a new sensor decision at that determines whether sensors are to be activated next. POMDP estimates the posterior state probabilities based on the corresponding observation symbol. The control policy is formulated as a constrained POMDP where the constraint is the average energy consumption per epoch as shown in Eq 6. The corresponding POMDP model parameters are as follows:
S = {Static,RightArm,RightLeg,WholeBody}
A = {Deactivate, Activate}
Ω = {0, 1, ..., |S| − 1}
π = {N, ψ,η }

$$P(S_{t+1}|S_t) = \frac{count(S_t, S_{t+1})}{count(S_t)}$$

$$P(O_{t+1}|S_{t+1}, a_t) = \begin{cases} P(O_{t+1}|S_{t+1}); if\ a_t = Activate \\ \frac{1}{\Omega}; otherwise \end{cases} \tag{6}$$

Where N is the set of *nodes* n ∈ N in the FSM, $\eta = P(n_{t+1}|n_t, a_t, o_{t+1})$, and $\psi = P(a_t|n_t)$. η is responsible for maintaining the internal state information regarding the control policy, while ψ determines the relative frequency of different sensor controls a ∈ A.

A real experiment was conducted for three days (2 days for training and one day for testing) using three MicroLEAPs worn on the right wrist, waist, and right ankle to capture the activities of interest. The mobile device serves as the data aggregator, and continuously logs all sensor data for both training and testing. The experinetal results have shown that the implemented POMDP approach improves the overall classification accuracy for a given energy consumption budget by using the state dependencies as additional information for classification.

5. Challenges and Open Issues

In this section, we discuss some of the major challenges facing the application of POMDP in real life WBAN applications. Fully observable MDP have been applied in so many areas as we earlier mentioned in [12], but not so for POMDP due to numerous challenges, most of the research ended up as published papers in reference journals. Though most real life problems are not fully observable which made POMDP a good candidate but because of some of these challenges, they have not been largely harnessed. The complexity of interaction between the state and information makes the modeling problem a difficult task. POMDP has difficulty in handling some problems with inherited peculiar characteristics.

While most of the models assume finite decision point approach, most of real life problems are not. POMDP applications are data intensive and new approaches need to be developed to minimize the collected information as well as retrieving this information. Data mining techniques would help in this domain. In addition, artificial intelligent techniques such as reinforcement learning would help in striking a balance between the accuracy required by the application and the quality of collected information.

MDP has been used a lot for solving and approximating the solution of POMDP. However, it ends with simplistic cases which are far away from the real life problems. Hence, new directions should be explored to enhance the formulation models and representation.

On the other hand, the limited number of problems with known solutions along with lack of efficient algorithms to solve POMDP problem make the practitioners less enthusiastic in adopting these approaches which also hinder the development is this area.

The future WBANs should be dependable where a conjunction of reliability, security and availability is provided [2]. Furthermore, building WBANs according to hierarchical configuration [60] where sensors can be partitioned into different layers, with each layer providing different granularities can help in achieving better coordination and determination of the state of the monitored system.

6. Conclusion and Future Work

Wireless body area network (WBAN) is gaining continuous attraction from industry and academia for its potential health applications. However, the realization of this type of network is facing a lot of challenges especially in managing the power supply as well as power consumption. The Partially Observable Markov Decision Processes (POMDPs) could provide the means for achieving these objectives. In this paper, we have presented the traditional approaches for solving POMDP and several enhancements along with recently developed methods and techniques in the quest to provide optimal solution techniques for POMDP. Also, we addressed some real implementations of POMDP for health applications. Finally, we described the challenges facing the usage of POMDP. We hope this survey will pave the path for advanced research in this important and critical fieldireless body area network (WBAN) is gaining attraction from industry and academia for its potentail health applications.

References

[1] M. Chen, S. Gonzalez, A. Vasilakos, H. Cao, V.C.M. Leung, "Body Area Networks: A Survey," *Springer Journal of Mobile Network Applications*, April 2011, Vol. 16, No. 2, pp 171-193. **Article (CrossRef Link)**

[2] Reusens, E., Josephm, W., Latre, B., Braem, B., Vermeeren, G., Tanghe, E., Martens, L., Moerman, I. and Blondia, C., "Characterization of On-Body Communication Channel and Energy Efficient Topology Design for Wireless Body Area Networks," *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 6, Nov. 2009, pp. 933-945. **Article (CrossRef Link).**

[3] Hovakeemian, Y Naik, K. and Nayak, A. ,"A survey on dependability in Body Area Networks," in *Proc. 5th International Symposium on Medical Information & Communication Technology (ISMICT)*, 27-30 March 2011, pp. 10 – 14. **Article (CrossRef Link).**

[4] Jiang Xing and Yunru Zhu, "A Survey on Body Area Network," in *Proc. 5th International Conference on Wireless Communications, Networking and Mobile Computing (WiCom '09)*, 24-26 Sept. 2009, pp. 1 – 4. **Article (CrossRef Link).**

[5] Jovanov, Emil, "A survey of power efficient technologies for Wireless Body Area Networks," in *Proc. IEEE 30th Annual International Conference on Engineering in Medicine and Biology Society (EMBS'08)*, 20-25 Aug. 2008, pp. 3628. **Article (CrossRef Link).**

[6] Gopalan, S.A. and Jong-Tae Park, "Energy-efficient MAC protocols for wireless body area networks: Survey," in *Proc. of International Congress on Ultra Modern Telecommunications and Control Systems and Workshops* (ICUMT), 18-20 Oct. 2010, pp. 739 – 744. **Article (CrossRef Link).**

[7] Barakah, D.M. and Ammad-uddin, M., "A Survey of Challenges and Applications of Wireless Body Area Network (WBAN) and Role of a Virtual Doctor Server in Existing Architecture Intelligent Systems," in *Proc. Third International Conference on Modelling and Simulation (ISMS)*, 8-10 Feb. 2012, pp. 214 – 219. **Article (CrossRef Link).**

[8] Patel, M. and W. Jianfeng, "Applications, challenges, and prospective in emerging body area networking technologies," *IEEE Wireless Communications Magazine*, vol. 17, no. 1, Feb. 2010, pp. 80 – 88. **Article (CrossRef Link)**

[9] M.A. Hanson, H.C. Powell Jr., A.T. Barth, K. Ringgenberg, B.H. Calhoun, J.H. Aylor, J.Lach, "Body Area Sensor Networks: Challenges and Opportunities," *IEEE Computer* Magazine, Vol. 42(1), Jan. 2009, pp. 58-65. **Article (CrossRef Link).**
[10] R. D. Smallwood and E. J. Sondik, The optimal control of partially observable Markov decision processes, 1971, Stanford University.
[11] L. P. Kaelbling, M.L. Littman, A.R. Cassandra, "Planning and acting in partially observable stochastic domains," *Journal of Artificial Intelligence*, 1998. 101: p. 99-134.
[12] G. Monahan, "A survey of partially observable Markove decision processes," *J. Management Science*, 1982(28): p. 1-16.
[13] Tang, L.C., kari, O., Frank, M., Xin, L. Venkatesh, A., "Markov Decision Process (MDP) framework for Optimizing Software on Mobile Phones," in *Proc. of 7th ACM international conference on Embedded software (EMSOFT '09)*, pp. 11-20, 2009. **Article (CrossRef Link).**
[14] Dejan, V.D., Vikram, K., "MIMO transmission control in fading channels-A constrained Markov Decision process Formulation with Monotone Randomized Policies," *IEEE Transactions on Signal processing*, 55(10), 2007. **Article (CrossRef Link).**
[15] Kottapalli, V. A., et al., "Two-tiered wireless sensor network architecture for structural health Monitoring," in *Proc. of 10th Annual International Symposium on Smart Structures and Materials*, 2003.
[16] Montgomery, K. et al., "LifeGuard-A Personal Physiological Monitor for Extreme Environments," in *Proc. of Medicine Meets Virtual Reality (MMVR04)*, 2004. **Article (CrossRef Link).**
[17] Malan, D., et al., "Wireless sensor networks for habitat monitoring," in *Proc. of International Workshop on Wearable and Implantable Body Sensor Network*, 2004. **Article (CrossRef Link).**
[18] Ng, J. , et al., "Ubiquitous monitoring environment for wearable and implantable sensors (UbiMon)," in *Proc. of International Conference on Ubiquitous Computing (Ubicomp)*, 2004.
[19] Anand, P., et al., "Markov Decision Processes for Control of a Sensor Network-based Health Monitoring System" in *Proc. of Seventh Conference on Innovative Applications of Artificial intelligence(IAAI)*, pp. 1529-1534, 2005.
[20] Goetschalck, R.; Sanner, S.; Driessens, K., "Reinforcement Learning with the Use of Costly Features," Recent Advances in Reinforcement Learning, Lecture Notes in Computer Science, vol. 5323, pp 124-135, 2008. **Article (CrossRef Link).**
[21] Kwame, O.H., "Application of Markov Decision processes to a simplified Model of Robotic fighter," *African Institute for Mathematical Sciences*, 2009.
[22] Osais, Y., "New optimization Models for directional and Biological Wireless Sensor Networks, PhD. Dissertation, Department of System and Computer Engineering. Carleton University, Ottawa, Canada, 2010.
[23] Russell, S. and P. Norvig, eds. Artificial Intelligence: A Modern Approach, ed. P. Hall. 1995.
[24] Hauskrecht, M. and H. Fraser, Planning treatment of ischemic heart disease with partially observable Markov decision processes. Artificial Intelligence in Medicine, 2000(18): p. 221–244.
[25] Hauskrecht, M., "Value function approximation for partially observable Markov decision processes," *Journal of Artificial Intelligence research*, 2000.13: p. 33-95.
[26] Nikos, V., Mathhhijs, T.J.S., "A fast point-based algorithm for POMDPs," in *Proc. of Belgian-Dutch Conference on Machine Learning*, Brussels, Belgium, pp. 170–176, Jan. 2004.
[27] Milos, H., Branislav, K., "Approximate Linear programming for Solving Hybrid factored MDPs," in *Proc. of 9th International Symposium on Artificial Intelligence and Mathematics*, 2006.

[28] Zhang, N.L. and W. Liu, "Planning in Stochastic domains: problem Characteristics and approximations," Hong Kong University of Science and Technology, 1996.
[29] Lovejoy, W., "Computationally Feasible Bounds for Partially Observable Markov Decision Processes," *J. Operations Research*, p. 162-175, 1991. **Article (CrossRef Link).**
[30] Cheng, H.T., "Algorithms for Partially Observable Markov Decision Processes," Ph.D. Thesis University of British Columbia, Vancouver, BC (1988).
[31] Littman, M.L., "The Witness Algorithm: Solving Partially Observable Markov Decision Processes," Technical Report: CS-94-40, 1994, Brown University, department of Computer Science.
[32] Pineau, J., et al, "Point based value iteration: anytime algorithm for POMDPs," in *Proc. of Int. joint conference on artificial intelligence*, 2003. Mexico.
[33] Matthijs T. J., et al, "Randomized Point-based Value Iteration for POMDPs," *Journal of Artificial Intelligence Research*, p. 195-220, 2005.
[34] Edwin, K.P., et al., "Monte Carlo Basex Partially Observable Markov Decision Process Approximation for Adaptive sensing," in *Proc. of the 9th International Workshop on Discrete Event Systems*, Goteborg, Sweden, pp. 173-180, 2008. **Article (CrossRef Link)**
[35] Zhengzhu, F., "Region-Based Dynamic programming for partially Observable Markov Decision Processes," 2006, Department of Computer science University of Massachusetts: Amherst.
[36] Hyeong, S.O., et al., "Parallel Rollout for Online Solution of Partially Observable Decision Process," Kluwer Academic Publishers, pp. 309-341, 2004. **Article (CrossRef Link).**
[37] Berstekas, D.P. and D.A. Castanon, "Rollout Algorithms for Stochastic Scheduling Problems," *Journal of Heuristics*, pp. 89-108, 1999. **Article (CrossRef Link).**
[38] Amato, C., et al., "Solving POMDPs Using Quadratically Constrained Linear Programs," AAMAS, Hokkaido, Japan, 2006.
[39] Ng, A., P. Ron, and D. Koller, "Policy Search via Density Estimation," *Advances in Neural Information Processing Systems 12*, Denver, Colorado, pp. 1022-1028, 2000.
[40] Ng, A.Y. and M. Jordan, "PEGASUS: A Policy Search Method for Large MDPs and POMDPs," in *Proc. of Uncertainty in Artificial Intelligence*. 2000.
[41] Satia, J.K. and R.E. Lave, "Markovian Decision Processes With Probabilistic Observation of States," Management Science, 1973.20(1). **Article (CrossRef Link).**
[42] Smith, T. and R. Simmons, "Heuristic search value iteration for POMDPs," *Uncertainty in Artificial Intelligence*. 2004.
[43] Poupart, P. and C. Boutilier, "Value-directed compression of POMDPs," *Advances in Neural Information Processing Systems 15 MIT press*, 2003.
[44] Roy, N., et al., "Finding Approximate POMDP Solutions Through Belief Compression," *Journal of Artificial Intelligence Research*, 2005.23: p. 1-40. **Article (CrossRef Link)**
[45] Roy, N. and G. Gordon, "Exponential Family PCA for Belief Compression in POMDPs," *Advances in Neural Information Processing Systems* 15, 2003.
[46] Xin, F., et al., "A POMDP Based k-Coverage Dynamic Scheduling Protocol for Wireless Sensor Network," in *Proc. of IEEE Globecom*, 2010. **Article (CrossRef Link).**
[47] Puterman, M.L., Markov Decision Processes: Discrete Stochastic Dynamic Programming, Publisher: Wiley-Interscience, 1994.
[48] Cadei, M. and D.Z. Du, "Improving Wireless sensor network lifetime through power aware organization," *Wireless Networks*, 2005. 11(3): pp. 333-340. **Article (CrossRef Link).**
[49] Shih, E., et al., "Physical Layer Driven Protocol And Algorithm Design For Energy Efficient Wireless Sensor Network," in *Proc. of the ACM 7th annual International conference on mobile computing and networking*. 2001. New York, USA. **Article (CrossRef Link).**

[50] Tian, D. and N.D. Georganas, "A coverage-Preserving node scheduling scheme for large wireless sensor networks," in *Proc. of 1st ACM International Workshop on Wireless Sensor Networks and Applications*. 2002. New York, USA. **Article (CrossRef Link)**

[51] Huang, C.F. and Y.C. Tsengm, "The Coverage Problem in A Wireless Sensor Network," in *Proc. 2nd ACM International Conference on Wireless Sensor Networks and Applications*," 2003. New York, USA. **Article (CrossRef Link).**

[52] Bayardo, R., et al., "Infosleuth: An agent-based semantic integration of information in open and dynamic environments," in *Proc. of the ACM SIGMOD International Conference on Management of Data*. 1997. **Article (CrossRef Link).**

[53] Yahya, O., et al., "Thermal Management of Biosensor Networks," in *Proc.* 7th *IEEE Consumer Communications and Networking Conference (CCNC)*, pp. 1-5. 9-12, Jan. 2010. **Article (CrossRef Link).**

[54] Tang, Q., et al., "Thermal Aware Routing Algorithm For Implanted Sensor Networks," Distributed Computing in Sensor Systems, Springer, pp. 206-217, 2005. **Article (CrossRef Link).**

[55] Tang, Q., et al., "Communication Scheduling to Minimize Thermal Effects of Implanted Biosensor Networks in Homogenous Tissue," *IEEE Transactions on Biomedical Engineering*, 52(7): pp. 1285-1294, 2005.

[56] Osais, Y., et al., "Optimal Management of Rechargeable Biosensors in Temperature Sensitive Environment," *Proc. 72nd IEEE Vehicular Technology Conference (VTC 2010-Fall),* 6-9 Sept. , pp. 1-5, 2010. **Article (CrossRef Link).**

[57] White, I., "Optimal Diagnostic Questionnaires Which Allow Less Than Truthful Responses," *Information and Control*, Vol. 32, pp. 61-74, 1976. **Article (CrossRef Link).**

[58] White, I., "Bounds on Optimal Cost for a Replacement Problem with Partial Observations, *Naval Research Logistics Quarterly*, p. 415-422, 1979. **Article (CrossRef Link).**

[59] Zois, D.-S., et al., "A POMDP Framework for Heterogeneous Sensor Selection in Wireless Body Area Networks," in *Proc. of IEEE INFOCOM* 2012, pp. 2611-2615. **Article (CrossRef Link).**

[60] Mitra, U, et al, "KNOWME: A Case Study in Wireless Body Area Sensor Network Design," *IEEE Communications Magazine*, pp. 116-125, May 2102. **Article (CrossRef Link).**

[61] Au, L., et al., "CARER: Efficient Dynamic Sensing for Continuous Activity Monitoring," in *Proc. 33rd Annual International Conference of the IEEE EMBS*, Boston, Massachusetts USA, August 30 - September 3, pp. 2228-2232, 2011. **Article (CrossRef Link).**

Mohammed Onimisi Yahaya received the B.Tech and M.Sc degree in Computer Science from Abubakar Tafawa Balewa University, Bauchi, Nigeria, in 2004 and 2010 respectively. He is currently a Ph.D candidate in Computer Science and Engineering program at King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia.

Uthman Baroudi is currently an Assistant Professor in the Department of Computer Engineering at KFUPM. Before joining KFUPM, he worked for Nortel Networks, Ottawa, Canada in R&D for next generation wireless networks. He has extensive teaching and research experience. He taught and developed several graduate and undergraduate COE courses in the area wireless and computer networking and computer systems performance evaluation. His research interest includes radio resource management (RRM) and QoS provisioning for the next generation wireless networks, wireless sensor and actuator networks. He has over 40 publications in reputable international journals and conference proceedings, and 2 US patent.