Planning & Acting: Optimal Markov Decision Scheduling of Aggregated Data in WSNs by Genetic Algorithm

Imane Horiya Brahmi [∓], Florian Maire [‡] , Soufiene Djahel [∓] and John Murphy [∓]

[∓] Lero, UCD School of Computer Science and Informatics, Ireland

[‡] UCD, School of Mathematical Sciences, Insight Centre for Data Analytics, Ireland

{horiya-imane.brahmi, florian.maire, soufiene.djahel, j.murphy}@ucd.ie

Abstract—Data aggregation techniques have emerged as promising solutions for extending Wireless Sensor Networks (WSNs) lifetime. However, this approach suffers from a design issue in delivering the strict requirements needed by some monitoring applications. Carefully balancing Energy, Delay and Accuracy is essential for achieving these requirements. In this work, we focus on distributed data aggregation, where a sensor estimates the network information by the exchange of readings with different priority levels. We then propose an optimal decision policy for scheduling the transmission of the aggregated data at the node level. To model the investigated problem, we first adopt Markov Decision Process (MDP) whereby we define the reward function. Then, we apply a Genetic Algorithm (GA) to find a set of optimal decisions that ensures the best trade-off between energy saving, delay and accuracy of the received data based on their priority level. The simulation results yield excellent performance and our optimization shows a significant enhancement up to 20% compared to the other policies.

Keywords – **WSNs, Data aggregation, Delay-Energy-Accuracy tradeoff, Markov Decision Process (MDP), Genetic Algorithm (GA)**

I. INTRODUCTION

WSNs [1] have attracted significant attention in recent years for their potential to replace many of the existing wired sensing technologies, as well as providing new platform solution where wired solutions are hard to deploy and maintain.

Although WSNs technology offers low cost and flexible solution for several monitoring applications, the high dependency of sensors on limited energy sources impedes its wide deployment. One basic approach to alleviate the consequences for this problem is to use in-network data aggregation and processing. This solution filters the received readings from different sources; combines them into one packet, thereby reducing the number of the transmitted packets and thus, the overall energy consumption in the network.

The critical challenge in designing the aggregation approach in WSNs is meeting the requirements imposed by various monitoring applications. Those requirements consist mainly in optimizing the delay incurred by the waiting time before transmitting the resulting packet from aggregation, as well as energy consumption in the node while insuring the highest accuracy of the transmitted data. For instance time-sensitive applications such as disaster relief, etc. require a short waiting times before forwarding the aggregated data. In contrast, periodic monitoring applications (with lower priority) are more sensitive to the accuracy of data being collected in terms of the achieved level of representativeness. An omnipresent requirement to both types of applications is energy saving for increasing the network lifetime.

We consider a scenario of distributed data aggregation, where each sensor can estimate the state of the network by exchanging its local readings with the neighbors [2]. The aggregated information is then made available to the end user at anytime in any node. In this context, each node should carefully decide whether to stop the aggregation process and transmit/forward the resulting packet, at a given medium access opportunity, or wait till the next access opportunity in order to increase the achieved reward. Moreover, in the scenarios where sensor nodes detect/forward various types of parameters with different priority levels, decisions should be taken according to the content of these different types of the message in order to fulfill the application constraints.

In this paper, we address the problem of optimal stopping at the node level, based on the received information, by:

(i) embedding the decision problem in a Markov Decision Process (MDP) [3] framework,
(ii) determining the optimal policy for the node to schedule the transmission of their aggregated packet using a Genetic Algorithm (GA) [4].

Therefore, at each access to the medium, this scheme allows the sensors to predict the future states of the network in order to act according to the best decision (to send or to postpone the aggregated packet). By construction of our model, a tradeoff between the energy consumption, the delay (i.e. waiting time) and the accuracy (i.e. the level of representativeness) is fulfilled.

The remainder of the paper is organized as follows: in Section II we briefly review some related works. Section III gives a detailed overview of our model and solve our optimization problem. In Section V, we assess our proposal and present/discuss the obtained simulation results. Finally, we conclude in Section VI.

II. RELATED WORK

MDPs framework has been widely used in uncertain environment and dynamic systems in order to optimize the network resources in WSNs. The authors of [5] reviewed in a survey different works on MDP and its applications in WSNs. In their paper the authors have discussed some MDP designs concerning data exchange and topology formation [6], resource and power optimization [7], security challenges [8], etc.

Related ideas in the literature have used this framework for different purposes, either for ensuring the tradeoff between energy saving and transmission delay or energy, throughput and delay. For example in [2] the authors study the fundamental trade-off between energy saving and transmission delay of the packets in distributed data aggregation. The main objective was to maximize a reward function involving an aggregation gain and a discount due to the waiting time. The authors proposed two learning-based distributed approximation algorithms that perform close to the optimal solution. In this work the representativeness of the received data is not considered in the decision-making. The authors of [9] used MDP in order to select messages forwarding based on the importance
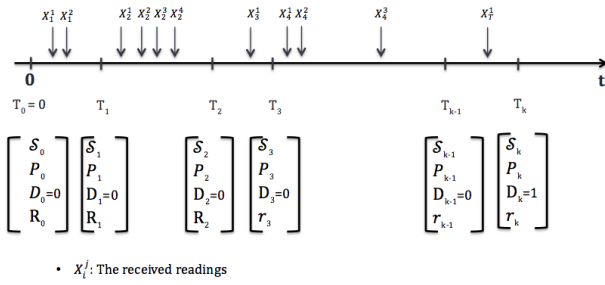
Figure 1: An illustration of the decision process

of the message and the energy level of the nodes. In this work the authors consider the scenario where sensors act as a relay of different priority packets. The authors used the MDP in order to determine which packets need to be sent and which one will be discarded based on the priority level of the messages. In [10] an application-oriented dynamic tuning methodology for WSNs based on MDP was proposed. The main objective in this work is to optimize a reward function that combine power consumption, throughput and delay. All the above related methodologies, involve the so-called Bellman equations which are, most of the time, intractable. As a result, they resort to numerical approximations which represent a computational burden. In addition, designing an aggregation scheme considering different types of packets and ensuring a tradeoff between energy delay and accuracy was not addressed in previous works. Our approach addresses this in the next subsections.

III. PROBLEM FORMULATION

We consider the scenario of structure-free distributed data aggregation in WSNs. Sensor nodes can estimate the information of the whole network through data exchange and aggregation with their neighboring nodes. We assume that we have no sink and the estimated information on the state of the network is available at any sensor node in the network, at any time. The exchange of the information is asynchronous and the arrival of samples at any node is random. The arrival of samples at each node can be either from its local sensing activities or neighboring nodes. The aggregation process starts at each sensor node upon reception of the first arrival sample. Once the channel is sensed as free, the aggregated packet can be sent. It is worth to mention that, in this work, a random channel access scheme is assumed due to the popularity of such scheme in WSNs [11].

The objective of the paper is to determine an optimal decision policy that achieves Energy-Delay-Accuracy tradeoff at each node.

We define T_k as the time period from when a node gets the first sample to the kth availability of the multiaccess channel for transmission as shown in the Figure 1.

Let us denote

- $\{T_k, k > 0\}$ the sequence of random variables corresponding to the times where the node has access to the medium
- $\{N_k, k > 0\}$ the number of received readings from the time T_0 to the k-th access to the medium, T_k
- X_i^j the random variable which represents the j-th received reading by the node during the interval (T_{i-1}, T_i)

At each access to the medium a node evaluate its state. At $t = T_k$ we have the following observations:

- $(X_1^1, .., X_1^{N_1})$ between (T_0, T_1)
- $(X_2^1, .., X_2^{N_2 - N_1})$ between (T_1, T_2)

$$\vdots$$

- $(X_k^1, .., X_k^{N_k - N_{k-1}})$ between (T_{k-1}, T_k)

Each node evaluates the received readings and decides whether it should stop waiting for more readings, allowing to save energy and

to reduce the delays, or if it should postpone the decision until the next interval to receive more readings, hence improving the accuracy. We formulate the decision problem with a Markov Decision Process (MDP) model characterized by the following tuple (S, D, P, R), where:

- S is the set of states
- $D = \{0, 1\}$ is the set of possible decisions
- P is a transition probability, it's the measure of the next state given the current state and the current decision
- R is the reward function

A. States of the node

At each available access to the medium T_k, a node evaluates its state S_k. This is modeled as a vector containing all the available information to make decision at time T_k and is defined by the following four parameters:

$$S_k = (E_k, W_k, X_k, A_k), \qquad (1)$$

where:

- E_k: Energy level at time T_k
- W_k: Delay, defined as the waiting time since the last transmission
- $X_k = (X_k^1, .., X_k^{N_k - N_{k-1}})$: Set of all the received observations from T_{k-1} to T_k with their priority level
- A_k: Accuracy at time T_k

Following [9], E_k is defined as the available energy at time T_k and depends on the decision D_{k-1}

$$E_k = E_{k-1} - D_{k-1}C_{1,k-1} - (1 - D_{k-1})C_{0,k-1}, \qquad (2)$$

where $C_{1,k-1}$ is the energy consumed when the node decides to stop the aggregation and transmit the resulting packet, and $C_{0,k}$ is the energy consumed for staying alive until T_k (i.e. sensing/receiving information while waiting for more packets).

The second parameter of the node state is the waiting time W_k. At time T_k, denoting by $T_k^{(last)}$ the time of last transmission, we have:

$$W_k = T_k - T_k^{(last)} = W_{k-1} + T_k - T_{k-1}. \qquad (3)$$

We choose an exponential distribution to model the time between two consecutive medium access for a given node[2]. Therefore, $T_k - T_{k-1}$ is an exponential random variable with parameter $\delta > 0$. Finally, conditionally on $W_{k-1} = w_{k-1}$, W_k is a shifted random variable:

$$W_k = w_{k-1} + \tau, \qquad \tau \sim \exp(\delta), \qquad (4)$$

For $i \in \{1, 2, \ldots, k\}$, $X_i^1, X_i^2, \ldots, X_i^{N_i - N_{i-1}}$ is the set of received readings by a node. We assume that the random variable N_k, denoting the number of messages between T_{k-1} and T_k follows a Poisson distribution with parameter λ i.e.

$$\mathbb{P}(N_k - N_{k-1} = n \mid T_k - T_{k-1} = \tau) = \frac{(\lambda \tau)^n}{n!} \exp(-\lambda \tau), \quad (5)$$

The state of the node depends also on the accuracy of the received observations. We define the accuracy based on the representativeness of the received data. However, there are other parameters which can also be considered in defining the accuracy of the received observation such as: the instant of samples generation, number of samples or the value of the received observations. In this work, we focus on the representativeness of the data in a finite set of sensors and define it as follows:

$$A_k = \begin{cases} \frac{N_k - N_{k-1}}{Nbr_{tot}} & if \quad D_{k-1} = 1 \\ A_{k-1} + \frac{N_k - N_{k-1}}{Nbr_{tot}} & if \quad D_{k-1} = 0 \end{cases} \qquad (6)$$

where, Nbr_{tot} is the total number of expected packets from the neighbors.

B. Decision Epochs

At time T_k, the sensor node must take the decision D_k either to send ($D_k = 1$) or to continue waiting ($D_k = 0$). The stopping policy $\Pi = \{D_1, D_2, ..., D_k, ...\}$ represents a sequence of decision rules, and depends on the state vector i.e.

$$D_k = d_k(s_k) = d_k(X_k, E_k, T_k, A_k). \quad (7)$$

Our ultimate goal is to find the set of decisions $\{D_1, D_2, ...\}$ that maximize an expected reward function (see Section III-D).

C. State Dynamics

Following the model introduced in the previous sections, we define the probability of any transition from state S_{k-1} to S_k. Let $\mathbb{P}(S_k|S_{k-1}, D_{k-1})$ represent the probability of reaching a state S_k knowing the state S_{k-1} and the decision D_{k-1}:

$$\mathbb{P}(S_k|S_{k-1}, D_{k-1}) = \mathbb{P}(X_k, A_k, W_k, E_k|S_{k-1}, D_{k-1}). \quad (8)$$

We assume that δ and λ are two known real positive constants. Since the parameters are non independent, we consider the following hierarchical model:

- E_k depends on E_{k-1}, D_{k-1}
- W_k depends on W_{k-1}, D_{k-1}
- A_k depends on A_{k-1}, D_{k-1}, W_k, W_{k-1}
- X_k depends on A_k, A_{k-1}, D_{k-1}

Using the total law of probability, this writes

$$\begin{aligned}
\mathbb{P}(S_k|S_{k-1}, D_{k-1}) = {} & \mathbb{P}(X_k|D_{k-1}, S_{k-1}, A_k, W_k, E_k) \times \\
& \mathbb{P}(A_k|A_{k-1}, D_{k-1}, S_{k-1}) \times \\
& \mathbb{P}(W_k|W_{k-1}, D_{k-1}) \times \\
& \mathbb{P}(E_k|E_{k-1}, D_{k-1}). \quad (9)
\end{aligned}$$

First, let us provide the transition probability of the energy E_k knowing E_{k-1} and D_{k-1}. For any $e > 0$ and given E_{k-1} and D_{k-1}, the probability of $E_k = e$ is

$$\mathbb{P}(E_k = e|E_{k-1}, D_{k-1}) = \begin{cases} \mathbb{P}(E_k = e|E_{k-1}, D_{k-1} = 0) \\ \mathbb{P}(E_k = e|E_{k-1}, D_{k-1} = 1) \end{cases} \quad (10)$$

$$\mathbb{P}(E_k = e|E_{k-1}, D_{k-1} = 0) = \begin{cases} 1 & \text{if } e = E_{k-1} - C_{0,k} \\ 0 & \text{otherwise} \end{cases} \quad (11)$$

$$\mathbb{P}(E_k = e|E_{k-1}, D_{k-1} = 1) = \begin{cases} 1 & \text{if } e = E_{k-1} - C_{1,k} \\ 0 & \text{otherwise} \end{cases} \quad (12)$$

We now focus on the transition of the waiting time W_k, given W_{k-1} and D_{k-1}. If $D_{k-1} = 1$, then the waiting time has been reset at time T_{k-1} and a new receiving period has started. In this case, for any $w_k > 0$, we have:

$$\mathbb{P}(W_k = w_k | W_{k-1} = w_{k-1}, D_{k-1} = 1) = \delta \exp(-\delta w_k). \quad (13)$$

Otherwise, if $D_{k-1} = 0$, the waiting time is incremented with a draw from an exponential distribution with parameter δ:

$$\begin{aligned}
& \mathbb{P}(W_k = w_k | W_{k-1} = w_{k-1}, D_{k-1} = 0) \\
& = \begin{cases} \delta \exp(-\delta(w_k - w_{k-1})) & \text{if} & w_k \geq w_{k-1} \\ 0 & \text{otherwise} \end{cases} \quad (14)
\end{aligned}$$

In our data model, each reading X_i^j is a random variable which takes its values in $\{1, 2\}$:

- $X_k^i = 1$ refers to applications sensitive to the delay
- $X_k^i = 2$ refers to applications sensitive to the accuracy

We assume that they are independent from each other and that *a priori*, a message of class 1 (resp. of class 2) is received with probability ρ (resp. with probability $1 - \rho$). This yields:

$$\begin{aligned}
& \mathbb{P}\left\{X_k = \left(X_k^{(1)}, X_k^{(2)}, ..., X_k^{(N_k - N_{k-1})}\right) |N_k, N_{K-1}\right\} = \\
& \rho^{\sum_{i=1}^{N_k} \mathbb{1}_{\{X_k^i = 1\}}} \times (1 - \rho)^{\sum_{i=1}^{N_k} \mathbb{1}_{\{X_k^i = 2\}}}. \quad (15)
\end{aligned}$$

Using the law of total probability, we provide the transition for the accuracy A_k:

$$\begin{aligned}
& \mathbb{P}(A_k = a|A_{k-1}, W_k, W_{k-1}, D_{k-1}) = \\
& \quad \sum_{n=0}^{\infty} \mathbb{P}_r(A_k = a \cap N_k = n|A_{k-1}, W_k, W_{k-1}, D_{k-1}) \\
& = \sum_{n=0}^{\infty} \mathbb{P}_r(A_k = a|N_k = n, A_{k-1}, W_k, W_{k-1}, D_{k-1}) \times \\
& \qquad \mathbb{P}_r(N_k = n|W_k, W_{k-1}, D_{k-1}) \\
& = \sum_{n=0}^{\infty} \mathbb{1}_{(A_{k-1} + \frac{N_k - N_{k-1}}{Nbr_{tot}})} \times \frac{(\lambda \tau)^n}{n!} \exp(-\lambda \tau)
\end{aligned}$$

D. Reward Function

The reward function is a deterministic function and depends on the energy consumption of the node, the delay (waiting time) and the accuracy of the received packets:

$$R_k = F(E_k, W_k, A_k). \quad (16)$$

This reward function $R_k(S_k, D_k)$ is determined at each access opportunity to the medium, given the node state S_k and the decision taken D_k. We define it as follows:

$$\begin{aligned}
R_k(S_k, D_k) = {} & \omega_E F_E(S_k, D_k) + \omega_W F_W(S_k, D_k) + \\
& \omega_A F_A(S_k, D_k), \quad (17)
\end{aligned}$$

where $F_E(S_k, D_k)$, $F_W(S_k, D_k)$, $F_A(S_k, D_k)$ represent the energy, delay and accuracy reward functions respectively, and ω_E, ω_W and ω_A are their corresponding weight factors.

The energy and the accuracy reward functions are defined in the equation (2) and (6). Therefore, we define the delay function as cited in [10]:

$$F_W(S_k, D_k) = \begin{cases} 1, & 0 < W_k \leq L_D \\ \frac{(U_D - W_k)}{(U_D - L_D)}, & L_D < W_k < U_D \\ 0, & W_k \geq U_D \end{cases} \quad (18)$$

where W_k is the waiting at time T_k and L_D and U_D are two constants defining the lower and upper tolerated delay respectively. These two delays can be set based on the application requirement.

We assumed in our work that the sensors in the network collect two different types of data with different priority levels. Therefore, we define two different sets of weights ω_I^r for $I \in \{E, W, A\}$ where $r \in \{1, 2\}$. ω_I^r depends on the message type, so that the delay reward function will have a higher weight value when $X_k^i = 1$ and the accuracy reward function will have a higher weight value when $X_k^i = 2$. We keep the same weight for the energy reward function for both messages. Finally, we weight our function according to the number of the received messages for each category, as it is shown in the following equation:

$$\begin{aligned}
R_k(S_k, D_k) = \sum_{r=1}^{2} & \frac{\#\{X_k^i = r\}}{Nbr_k} \times \Big[\omega_E^r F_E(S_k, D_k) + \omega_W^r \\
& F_W(S_k, D_k) + \omega_A^r F_A(S_k, D_k) \Big], \quad (19)
\end{aligned}$$

E. Optimality Equation

For a given state S_k, a sensor node makes a decision $D_k \in \{0, 1\}$. The set of all decision is called the policy $\pi \in \{0, 1\}^{\infty}$. Therefore, the reward received by the node at time k is $R_k(S_k, D_k)$. The expected total reward v_k over the decisions making horizon $\pi_k = (D_k, \ldots, D_{\infty})$ is equal to:

$$v_k(\pi_k) = R_k(S_k, D_k) + \underbrace{\mathbb{E}\left\{ \sum_{t=k+1}^{\infty} R_t(S_t, D_t) \middle| \pi_k \right\}}_{\psi(\pi_k, S_k)}. \quad (20)$$

We assume a finite decision making horizon H for the expected total reward $v_k(\pi_k)$, such that:

$$\psi(\pi_k, S_k) = \int \cdots \int \sum_{t=k+1}^{H} R_t(S_t, D_t)$$
$$f(S_{t+1}, S_{t+2}, \ldots, S_H)|S_k, \pi_k)\mathrm{d}S_{t+1} \ldots \mathrm{d}S_H, \quad (21)$$

where f denotes the probability density function of the future states $S_{t+1}, \ldots, S_H$ given the current state S_k and the decision making horizon $\pi_k = (D_k, D_{k+1}, \ldots, D_H)$.

Our objective is to find the policy $\pi_k \in \{0, 1\}^{H-k+1}$, that maximizes $v_k(\pi_k)$

$$\pi_k^* = \arg \max_{\pi_k \in \{0,1\}^{H-k+1}} v_k(\pi_k), \quad (22)$$

The total expected reward involves the actual reward $R_k(S_k, D_k)$, defined in (19) and the expected reward $\psi(\pi_k, S_k)$. However, because of the underlying MDP model, an analytical expression of ψ is not available. We resort to a pointwise Monte-Carlo estimation of ψ to overcome this issue and for a given (π_k, S_k), we approximate $\psi(\pi_k, S_k)$ by

$$\hat{\psi}(\pi_k, S_k) \approx \frac{1}{N_{sc}} \sum_{j=1}^{N_{sc}} \left\{ \sum_{t=k+1}^{H} R_t(S_t^j, D_t) \right\}, \quad (23)$$

where N_{sc} defines the number of scenarios in the Monte-Carlo estimation. This number is set in the simulation part.

IV. SOLVING THE OPTIMALITY EQUATION USING A GA

In this section, we solve our optimality equation and find a set of a finite optimal policy $\hat{\pi}_t^*$ that optimizes the

$$\hat{v}_k(\pi_k) = R_k(S_k, D_k) + \hat{\psi}(\pi_k, S_k). \quad (24)$$

We propose to use a Genetic Algorithm (GA) [4] in order to perform this optimization problem. The GA provides a smart metaheuristic for solving combinational optimization problems [12] and is well suited to discrete optimization problems, that we are aiming to solve in (22). By analogy to evolutionary theories which predict that in a random population only the most adapted individuals to the environment will survive, the GA looks for the features that correspond to the fittest individual.

The GA starts by selecting successive generations from an initial random population, which corresponds to the a set of decisions in our context. At convergence, the GA will provide the fittest combinations from the last generation with respect to the environment (i.e. the best combination of decisions which maximize the reward function (19) after many generations). The Algorithm 1 summarizes the different steps performed at each iteration of the GA.

Algorithm 1

Input: A fitness function $v_k(\pi_k)$, the population size $PopSize$ and the number of generations $GenNbr$

(1) *Initialisation:* start with a random set of population $\{\pi_1^m\}_{m=1}^{PopSize}, \{\pi_1^m\} \in \{0, 1\}^{H-k+1}$

(2) *Evaluation:* measure the fitness of the population by computing $v_k(\pi_1^m)$ and keep the $PopSize$ fittest individuals

(3) *Selection:* select $PopSize$ fittest couples

(4) *Reproduction:* each couple provides an individual, the shared parent decisions are preserved while the remaining decisions are selected randomly.

Go to step (2) unless $GenNbr$ generations have been completed.

V. PERFORMANCE EVALUATION

In this section we analyze the performance of our proposed scheme in different scenarios through simulations using Matlab and show the substantial gain achieved through the optimization scheme we propose. We first start by describing some common features of the experimental setup, followed up by the description of the obtained results.

- The random number of sample arrivals is assumed to have a Poisson distribution with $\lambda = 10$ for the generation of readings during two consecutive access opportunities to the medium. The readings have different priority levels: we generate them with different probability $\rho = 0.3$ for $X_k^i = 1$ and $1 - \rho$ for $X_k^i = 2$. The weight is set to $\omega^1 = [0.4, 0.5, 0.1]$ and $\omega^2 = [0.4, 0.2, 0.4]$. This allows to privilege the delay when $X_k^i = 1$ and the accuracy when $X_k^i = 2$. We have set the total number of the neighbor each node to $N_{tot} = 20$.

- We use an exponential model with $\delta = 6$ for the random inter-arrival time of decision epochs (i.e. access to the medium). We use a simulation horizon of 3 time units.

- We have used a simple deterministic energy model fixed three constant and known parameters: for all $k > 0$, the energy spent on idle state $C_{0,k} = .01$, the energy spent on transmitting a single message $C_{1,k} = .1$ and $E_{init} = 1$ the initial energy.

- For the delay function, we have set the upper to $U_D = 1$ and the lower delays to $L_D = 0.1$.

- In our estimation, we have set the number of Monte Carlo simulations to $N_{sc} = 100$.

- We have solved the optimization problem throughout the GA available in the Matlab Global Optimization Toolbox. We set the population size $PopSize = 50$ and a generation number $GenNbr = 5$.

A. Results:

We have considered four schemes of policy design for the decision problem in distributed data aggregation: (1) our policy scheme using GA (Optimal Decision), (2) random decision (Random Decision), (3) decision policy which privileges delay transmission (i.e. the node send the resulting packet every time it gets access to the medium, $\pi_{WaitingTime}$) and (4) decision policy which privileges the accuracy of the data (i.e. the node send the resulting packet only when it receives all the expected packets, $\pi_{Accuracy}$).

First, we show the impact of the different policy decisions on the reward function over the decision epochs for a finite horizon H. Figure 2 shows the gain achieved by our optimization compared to the three other schemes. We notice the general decreasing trend of the reward function over time. This is explained (i) since the energy
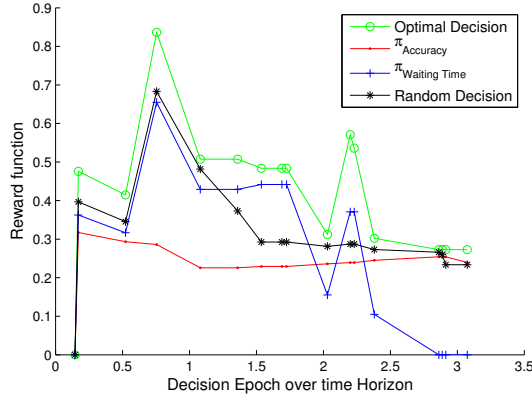
Figure 2: The impact of the different policy decisions on the Reward Function over the decision epoch
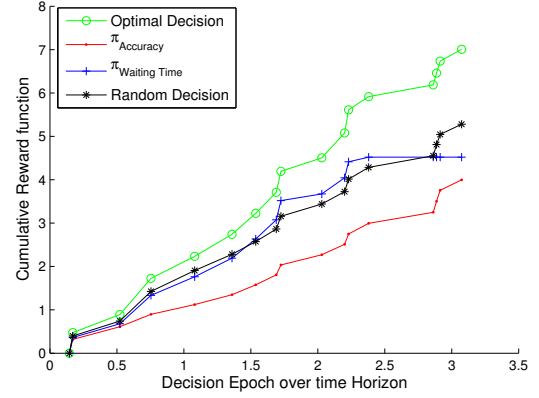

Figure 3: Cumulative Reward Function over time

decreases every time a sensor node receives, waits or transmits data, (ii) because the delay penalizes the reward when it exceeds the upper bound. On the other side, the accuracy is likely to increase the reward function since more and more readings are available as time goes. The crucial feature of our solution is that it allows a tradeoff between the expected gain in accuracy and the expected losses due to the energy and the delay. In Figure 3, we show the impact of the cumulative reward function over time horizon: this allows to assess the significant gain offered by our scheme at a global scale.

We have used different finite decision horizons $H \in \{0, 1, 3\}$ with $N_{sc} = 100$ in the Figure 4. The case $H = 0$ corresponds to a maximization of the sole R_k in (24) . As expected, the larger the horizon H, the better the prediction and the greater the reward. Meanwhile, a large horizon H generates an heavy computational burden: indeed, the Monte-Carlo simulation forecasts the network state for a time window of length H. As shown in Figure 4, the benefit provided by increasing H is decreasing over time (the precision of the network state prediction indeed shrinks for distant states). This yields a tradeoff between the reward gain and the computational burden. In the present simulation, $H = 1$ provides a satisfactory tradeoff.

We show in Figure 5 how the different decisions influence the energy function for both type of messages since they have the same ω. The energy function decreases when the sensor nodes transmits at any medium access: this represents the worst case (see $\pi_{WaitingTime}$). On the other hand, more energy is saved when there is less transmission ($\pi_{Accuracy}$). We notice from this figure that the energy function observed with the optimal decision is close the best solution ($\pi_{Accuracy}$). However, the decrease of all energy functions is due to the packets send and received.

The last set of results in Figures 6 and 7 present the delay and accuracy functions for the different types of message. In this set of results, we compare our scheme with the best schemes: $\pi_{WaitingTime}$ for the delays plots and $\pi_{Accuracy}$ for the accuracy plots. This shows how our results are close to these schemes, and how our scheme gives priority to the accuracy when the readings are more sensitive to the accuracy, and to the delay when they are sensitive to the waiting time.

In Figure 6, we show how the reading sensitive to the delay ($X_k^i = 1$) achieves a lower delay function. Those results are quite similar to those achieved in the scenario where the sensors forward the resulted packet every time they have access to the medium $\pi_{WaitingTime}$. Recall that when the waiting time is small, the delay function is equal to 1 (see equation (18)).

We show in Figure 7 how the accuracy function of the messages sensitive to the accuracy is almost equal to $\pi_{Accuracy}$.
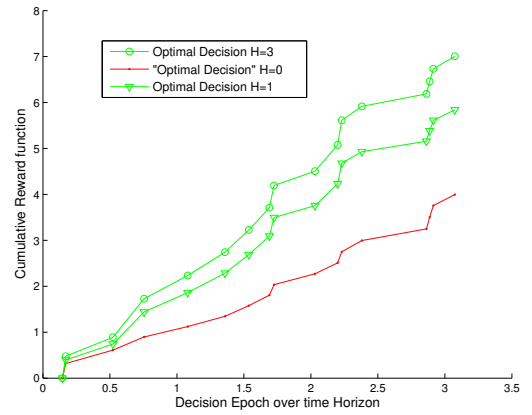

Figure 4: The impact of different finite decision horizons H on the Cumulative Reward Function over time
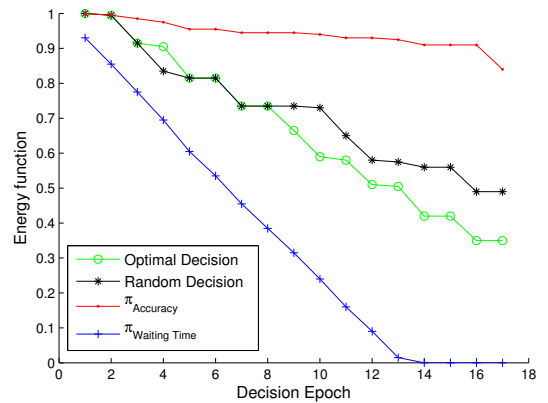

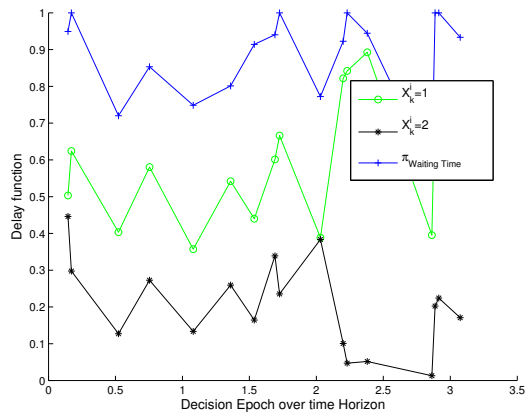Figure 5: The impact of the different policy decisions on the Energy Function

Figure 6: The impact of the different policy decisions on the Delay Function
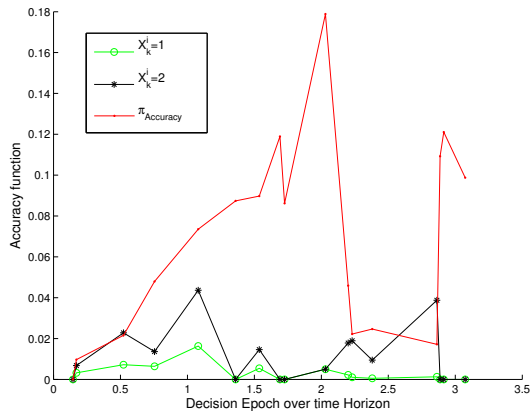

Figure 7: The impact of the different policy decisions on the Accuracy Function

VI. CONCLUSION

A common requirement in WSNs technology is the demand to operate under device constraints (i.e. sensor lifetime, delay, etc). However, in the scenarios where the nodes forward messages with different priority levels, finding the best trade-off between the energy consumption, the transmission delay and the achieved accuracy is essential for meeting the application requirements. This paper has introduced a new stochastic decision framework to find an optimal policy that ensures Energy-Delay-Accuracy tradeoff in distributed data aggregation in WSNs for multipurpose applications. We first applied Markov Decision Process (MDP) framework for analyzing the decision problem and proposed to use a GA to determine the current and future optimal times for scheduling the aggregated data. This ensures a trade-off between the above three metrics. The performance evaluation results highlight the effectiveness and have shown that our algorithm achieves optimal performance in balancing energy, delay and accuracy.

VII. ACKNOWLEDGEMENT

REFERENCES

[1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: a survey," *Computer networks*, vol. 38, no. 4, pp. 393–422, 2002.

[2] Z. Ye, A. A. Abouzeid, and J. Ai, "Optimal stochastic policies for distributed data aggregation in wireless sensor networks," *Networking, IEEE/ACM Transactions on*, vol. 17, no. 5, pp. 1494–1507, 2009.

[3] M. L. Puterman, *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons, 2014.

[4] D. E. Goldberg and J. H. Holland, "Genetic algorithms and machine learning," *Machine learning*, vol. 3, no. 2, pp. 95–99, 1988.

[5] M. A. Alsheikh, D. T. Hoang, D. Niyato, H.-P. Tan, and S. Lin, "Markov decision processes with applications in wireless sensor networks: A survey," *arXiv preprint arXiv:1501.00644*, 2015.

[6] J. Hao, Z. Yao, K. Huang, B. Zhang, and C. Li, "An energy-efficient routing protocol with controllable expected delay in duty-cycled wireless sensor networks," in *Communications (ICC), 2013 IEEE International Conference on*. IEEE, 2013, pp. 6215–6219.

[7] Z. Lin and M. van der Schaar, "Autonomic and distributed joint routing and power control for delay-sensitive applications in multi-hop wireless networks," *Wireless Communications, IEEE Transactions on*, vol. 10, no. 1, pp. 102–113, 2011.

[8] X. Li, Y. Zhu, and B. Li, "Optimal anti-jamming strategy in sensor networks," in *Communications (ICC), 2012 IEEE International Conference on*. IEEE, 2012, pp. 178–182.

[9] R. Arroyo-Valles, A. G. Marques, and J. Cid-Sueiro, "Optimal selective forwarding for energy saving in wireless sensor networks," *Wireless Communications, IEEE Transactions on*, vol. 10, no. 1, pp. 164–175, 2011.

[10] A. Munir and A. Gordon-Ross, "An mdp-based application oriented optimal policy for wireless sensor networks," in *Proceedings of the 7th IEEE/ACM International Conference on Hardware/Software Codesign and System Synthesis*, ser. CODES+ISSS '09, 2009, pp. 183–192.

[11] A. Bachir, M. Dohler, T. Watteyne, and K. K. Leung, "Mac essentials for wireless sensor networks," *Communications Surveys & Tutorials, IEEE*, vol. 12, no. 2, pp. 222–248, 2010.

[12] R. V. Kulkarni, A. Forster, and G. K. Venayagamoorthy, "Computational intelligence in wireless sensor networks: A survey," *Communications Surveys & Tutorials, IEEE*, vol. 13, no. 1, pp. 68–96, 2011.