On the Feasibility of Distributed Sampling Rate Adaptation in Heterogeneous and Collaborative Wireless Sensor Networks

Amitangshu Pal and Krishna Kant
Computer and Information Sciences, Temple University, Philadelphia, PA 19122
E-mail:{amitangshu.pal,kkant}@temple.edu

Abstract- In this paper we develop a general framework for multi-sensor, heterogeneous sensing in collaborative wireless sensor networks (WSNs) that can be used in a variety of large scale monitoring applications. In order to achieve better tolerance against unstable wireless links and nodes with inadequate battery, it is important to consider distributed approaches for sampling rate adaptation. We show that the fully distributed mechanisms suffer from high convergence time, which make them difficult to implement in large-scale WSNs. To overcome this limitation, we next propose two alternate approaches. We perform extensive simulations to compare these schemes and argue their scalability and applicability in real world monitoring scenarios.

I. Introduction

Heterogeneous sensor networks are defined as networks where the wireless nodes are equipped with multiple, different types of sensors, such as audio, video, acceleration etc. Devices with multiple such sensors are becoming ubiquitous, particularly because of proliferation of smartphones. Such sensors can be exploited for a variety of applications including disaster recovery [1], earthquake monitoring [2], infectious disease surveillance [3] etc. In many such applications the monitoring is of most interest at certain points-of-interests (PoIs), and many sensing devices can do the monitoring (e.g., most cameras in the vicinity of PoI can capture picture and most microphones can capture the sound). Naturally, the sensor readings in these cases are highly *correlated*. In a heterogeneous sensing environment, this correlation among the sensing devices can result from two factors. First one is the *spatial* correlation, which results from the vicinity of the sensing devices. The second factor is the *cross-sensor* correlation, which is the inherent dependencies among different types of sensors. For example in an earthquake monitoring scenario, the accelerometer, video or audio samples will have a high degree of correlation. Since the sensor nodes are generally constrained by limited battery power, energy efficient sampling and communication that exploits the correlations is essential for longest possible monitoring. In particular, the correlations among sensors amount to redundant information collection, which needs to be suitably controlled. We specifically consider scenarios where the sensors form an ad-hoc network among themselves with only certain distinguished nodes connecting to the external world via technologies such as cellular or satellite. In such cases, a sensor node may act both as a data source and a data relay. Both of these need to be considered in adaptation for available energy.

In this paper, we analyze distributed sampling rate adaptation schemes in the *multi-sensor equipped wireless devices* to distribute the data capturing tasks among them based on their available energy, network participation, and correlations. The primary motivation for distributed schemes is their better resilience in the face of temporary failures (e.g., fading, interference, signal drop in wireless links) and permanent failures (e.g., a node moving out of range or stop participating) as compared to centralized schemes. To address these issues, we propose two decomposition based distributed solutions using sub-gradient method and Nesterov's gradient descent algorithm [4] to adapt the sampling rates of the individual sensors in a multi-hop wireless sensor network. Contrary to the prior works that have focused on independent sampling rate adaptation of the nodes [5], we introduce the notion of *Points of Interests (PoIs)* in a geographic area that are collectively monitored by the network sensors. We also incorporate the effects of two types of correlations (spatial and cross-sensor correlation) among the sensing devices in our optimization framework which is sparse in the literature.

This research was supported by the NSF grant CNS-1461932.

Unfortunately, distributed solutions to optimization problems may involve a large number of iterations and hence message exchanges. This limits the scalability of the schemes in a practical energy constrained sensor networking environment. In order to address the high cost of fully distributed solutions in this context, we propose two hybrid schemes for faster convergence. Extensive simulations show that the proposed solutions ensure convergence within $\sim$ 10-15 iterations, that achieve faster convergence by balancing the amount of control overhead against resilience.

The rest of the paper is organized as follows. Section II discusses a number of collaborative sensing applications that will benefit from the methods described in this paper. This section also includes some basic definitions and notations. Section III then introduces the optimization model along with the distributed approaches that are adopted, along with their pros and cons. Section IV describes two scalable alternate schemes to make rate adaptation faster. Extensive simulations are presented in section V. Related proposals and relevant discussions are summarized in section VI. The paper is concluded in section VII.

II. Collaborative Rate Adaptation

A. Potential Application Areas

Collaborative sensing and rate adaptation of a set of heterogeneous sensors for optimal monitoring of certain points of interest (PoI's) is applicable to a wide variety of sensing applications, of which we mention a few in the following:

Disaster Monitoring: Disasters of various forms such as earthquakes, hurricanes, electromagnetic storms, etc. can substantially disrupt the cellular communications infrastructure thereby benefiting substantially from ad-hoc networks involving smartphones and the especially deployed emergency communications infrastructure [1]. The monitoring needs here involve pictures, sounds, movements relevant to damage assessment and rescue at certain PoI's where the situation may

978-1-5090-2279-3/16/$31.00 ©2016 IEEE

be unclear. The solutions deployed must be aware of limited battery lifetime of smartphones and the correlations that exist between the data gathered by various sensors.

Urban Water Contaminant Detection: Because of the large number of contaminants in the water, the detection is often done by a limited set of "proxy sensors" such as chlorine, organic carbon, conductivity, pH, turbidity, etc.[6]. Limited battery life and correlations are again important in this application, and the monitoring is of most interest at certain potential sources of contamination.

Forest Monitoring: Two significant applications in this space include (a) monitoring of poaching activity in certain areas via sensing of movement, thermal profile, track marks, disturbance to vegetation, etc., and (b) monitoring of potential wildfire via sensing of temperature, smoke, flames, flammable liquids, etc. at certain vulnerable points (i.e., those with lot of dry vegetation). Both applications require specialized sensors which are battery limited and mutually correlated.

Landslide Monitoring: Landslides are quite common in several parts of the world and can be predicted with varying degree of success based on topography, composition of soil layers, moisture, pore pressure, soil vibration/movement, etc. at most vulnerable points[7]. Since the monitoring devices need to be buried, their battery life is crucial. Also, many sensor readings are correlated which can be exploited for energy adaptation.

Substation Monitoring: Power distribution substations have a number of critical components such as circuit breakers and transformers that must be continuously monitored for overheating, leakage of coolant (oil or SF_6 gas), vibrations, and physical damage for equipments that are very old or most vulnerable [8]. This again yields an environment where both the battery life and sensor correlations are important.

B. Network Model

We next describe the network model for our scheme. We assume that some wireless devices are equipped with multiple sensors to sense different physical parameters and report them to a centralized place. For the sake of clarity, we define these individual wireless devices as *nodes*, whereas the word *sensor* is used to describe various sensors attached to that node. For example a smartphone can be considered as a node, which is equipped with different sensors such as accelerometers, temperature and audio sensors etc.

We assume that the whole network is divided into multiple disjoint clusters, each one having its own cluster-head (CH). Nodes in each cluster forward their traffic to their cluster-head. We assume that the cluster-heads can directly communicate with each others and are not energy constrained. Thus there is no centralized sink in our model. Such network model is applicable in many WSN application scenarios as described in section II-A. For example, consider a disaster management scenario where some access points (sinks) are deployed in a large geographic area, each one of them form a cluster of some nodes. These access points are equipped with cellular or satellite antennas and thus can communicate with themselves. Nodes broadcast periodic beacons to exchange various control parameters. The nodes discover their neighbors and construct their routes to their CH by using a *Collection Tree Protocol (CTP)*.

To estimate the quality of a route, we use a path metric that is obtained as the sum of the expected number of transmissions (ETX) on each of its links, which is the same principle applied in CTP. An ETX for a link is the expected number of transmission attempts required to deliver a packet successfully over the link. In CTP, path selection is performed based on maximizing a path quality metric or minimizing the path-ETX, which is the sum of link ETXs along the path. We also define min-ETX of a node as the path-ETX of the best quality route towards their CH.

In our model we define *potential parents (PPs)* of a node as the set of neighbors whose ETX are less than that of the node. Along with the ETX, the CHs also broadcast some points of interests (PoI) that need to be monitored by the nodes, in their beacon messages. In disaster management applications, these points may be the areas that are largely damaged by the disaster, or the areas whose level of damage is unknown to the management. In case of water contaminant detection, these are the points where the WDS wants to monitor the level of contaminants and water qualities. These PoIs can be evolved over time by the CHs. These PoIs are used for adapting their sampling rates, as mentioned later on.

We assume that nodes are not time synchronized and they apply some Low Power Listening (LPL) [9] schemes like X-Mac principle [10] to conserve energy. In X-Mac the sender sends a number of strobe packets that span the complete length of a sleep-wake cycle to ensure that the receiving node detects it regardless of when it wakes up. The strobe packets contain the address of the receiver, thus other neighbors or overhearers can refrain from keeping their radios on when hearing a strobe for another node. Whenever the receiver receives a strobe packet, it sends an acknowledgment, upon hearing that the sender immediately sends the data packet. Such scheme reduces the duration and power consumption of receiving as well as overhearing. Thus the primary source of energy consumption is to sense different parameters and forward them to the CHs in a multi-hop fashion.

We assume that there are $\mathcal{E}$ POIs that need to be monitored by the nodes. Along with the POIs, the CHs also broadcast the areas around the POIs that are more important to be monitored, which are represented as a radius of ϱ^k around the k-th POI. Assume that there are $\mathcal{S}$ number of nodes in the whole area and $\mathcal{T}$ is the number of different types of sensors that a node has. Assume that p_i^{tk} is the weight of covering a POI k by a sensor of type t for the i-th node. If a node lies within the radius of a POI its weight is 1, and beyond the radius its weight drops exponentially with distance. Thus

$$p_i^{tk}(d_i^k) = 1 \quad d_i^k < \varrho^k \quad \text{and} \quad e^{-\eta^t(d_i^k - \varrho^k)} \quad d_i^k > \varrho^k \tag{1}$$

where d_i^k the distance between the i-th node and the k-th POI. η^t is the decay rate beyond ϱ^k, which is different for different sensors. We assume that the nodes are localized and use their position information to measure their distance from the POIs.

C. Definitions and Notation

We now briefly define some notations, terminologies and basics that are related to our derivations presented later on.

Matrix norms: We define the l_1, l_2 and l_∞ norm of a $m \times n$ matrix $\mathbf{A}$ as $||\mathbf{A}||_1$, $||\mathbf{A}||_2$ and $||\mathbf{A}||_\infty$ respectively. $||\mathbf{A}||_1 = \max_{1 \le j \le n} \sum_{i=1}^{m} |a_{ij}|$, which is the maximum absolute column sum of matrix $\mathbf{A}$. $||\mathbf{A}||_\infty = \max_{1 \le i \le m} \sum_{j=1}^{n} |a_{ij}|$, which is the maximum absolute row sum of $\mathbf{A}$. The l_2 norm is related to the *spectral radius* of matrix $\mathbf{A^T}.\mathbf{A}$ as follows:

$$\rho\left(\mathbf{A^T}.\mathbf{A}\right) = ||\mathbf{A}||_2^2 \le ||\mathbf{A}||_1.||\mathbf{A}||_\infty \tag{2}$$

where $\rho\left(\mathbf{A^T}.\mathbf{A}\right)$ is the spectral radius of $\mathbf{A^T}.\mathbf{A}$.

Lipschitz continuity: We next define the definition of *Lipschitz continuity* of a function, which measures the change of the function values versus the change in the independent variable $x \in \mathbb{I}$ for a general function $f(x)$. If x_1 and x_2 are two numbers, then $|x_2 - x_1|$ is the change in the input and $|f(x_2) - f(x_1)|$ is the corresponding change in the output. We say that f is Lipschitz continuous with Lipschitz constant L, if there is a positive constant L such that

$$|f(x_1) - f(x_2)| \le L.|x_1 - x_2| \quad \forall x_1, x_2 \in \mathbb{I} \tag{3}$$

Notice that the Lipschitz constant is the *upper estimate* on how much the function f changes and the actual change might be much smaller than indicated by the constant.

Strongly convex function: A convex function f is σ-strongly convex if

$$f(y) \le f(x) + \bigtriangledown f(x)^T.(y - x) + \frac{\sigma}{2}||y - x||_2^2 \quad \forall x, y \in \mathrm{dom}(f) \tag{4}$$

If f is twice differentiable, then m-strong convexity is equivalent to

$$\bigtriangledown^2 f(x) \succeq \sigma.I \quad \forall x \in \mathrm{dom}(f) \tag{5}$$

where I is the identity matrix.

TABLE I. TABLE OF NOTATIONS

		Indices
i, j	$\triangleq$	Index for the nodes (1, ..., $\mathcal{S}$)
t	$\triangleq$	Index for sensors (1, ..., $\mathcal{T}$)
k	$\triangleq$	Index for the POIs (1, ..., $\mathcal{E}$)
		Variables
c^t	$\triangleq$	Energy expenditure for transmitting a sample point by sensor t
e^t	$\triangleq$	Energy expenditure for sensing a sample point of sensor t
E_i	$\triangleq$	Available energy budget per unit time for the i-th node
w_i^{tk}	$\triangleq$	Weight of covering the k-th POI based on the t-th sensor readings of node i
α_{ij}	$\triangleq$	Fraction of node i-th traffic that passes through node j
$\mathbb{A}_i$	$\triangleq$	Set of ancestors of node i
$\mathbb{D}_i$	$\triangleq$	Set of descendants of node i
$\mathbb{P}_i$	$\triangleq$	Set of PPs of node i
R_m, R_M	$\triangleq$	Minimum and maximum sampling rate allowed in any sensor
W_m	$\triangleq$	Minimum weight for covering the POIs
ν_i^ℓ	$\triangleq$	Lagrange multiplier of node i at iteration round ℓ
γ	$\triangleq$	Constant step size used in the sub-gradient method
L	$\triangleq$	Lipschitz constant
$\rho(M)$	$\triangleq$	Spectral radius of matrix M
$\|M\|_1$	$\triangleq$	l_1-norm of matrix M
$\|M\|_\infty$	$\triangleq$	l_∞-norm of matrix M

III. ADAPTIVE SAMPLING AND TRANSMISSION

In a multi-sensor, collaborative WSN the nodes need to adapt their transmission and sensing rates of individual sensor to match their battery powers, so that they can cumulatively share their data sampling and forwarding tasks depending on their power budgets. We next formulate the adaptive sampling and transmission rate adaptation scheme, with the objective of maximizing the overall coverage of the POIs, and to ensure that the nodes do not run out of batteries. The notations used for the problem formulations are listed in Table I. We define the utility of sensing t by a node i by considering two factors

- The sensing rate r_i^t. As r_i^t increases the number of sampled points increases and so does the utility.

- Their corresponding weight w_i^{tk} while covering the POIs. w_i^{tk} is the product of two factors: (a) p_i^{tk} which is dependent on how far a node is from the PoIs, and (b) the relative weights a_i^{tk} among the sensors (photos may be considered more important then audio samples etc). If an overall weight of a sensor is more it's contribution to the utility function increases. Without any loss of generality we assume that $w_i^{tk} \le 1$.

Considering the above two factors, the effective rate with which the k-th POI is monitored by the t type sensors is given by $\xi^{tk} = \sum_{i=1}^{\mathcal{S}} w_i^{tk}.r_i^t$. Thus the fair event reporting ability is ensured by modeling the utility of reporting the k-th POI by the t sensors as $U^k\left(\sum_{t=1}^{\mathcal{T}} \xi^{tk}\right) = \log\left(\sum_{t=1}^{\mathcal{T}} \xi^{tk}\right)$. Our objective is to maximize the overall event monitoring capability, i.e. $\sum_{k=1}^{\mathcal{E}} U^k\left(\sum_{t=1}^{\mathcal{T}} \xi^{tk}\right)$, after satisfying the energy budget of the individual nodes. Thus the overall optimization problem can be written as

Original Problem (OP):

$$\begin{aligned} \text{Maximize} \quad & \sum_{k=1}^{\mathcal{E}} \log\left(\sum_{i=1}^{\mathcal{S}} \sum_{t=1}^{\mathcal{T}} w_i^{tk}.r_i^t\right) \\ \text{subject to} \quad & \sum_t r_i^t\left(e^t + c^t\right) + \sum_t \sum_{j \in \mathbb{D}_i} \alpha_{ji}.r_j^t.c^t \le E_i \ \ \forall i \\ & R_m \le r_i^t \le R_M \ \ \forall i, \forall j, \forall t \end{aligned} \tag{6}$$

We assume that α_{ji} is the fraction of traffic of node j that passes through it's PP i. Later on we show that our scheme only needs the α_{ji} of their PPs. Thus a node j can approximate the α_{ji} of its PP i from its number of PPs as well as their route costs. The first constraint is the *energy budget constraint* that states that the energy spent for sensing and transmission is less than some threshold E_i. R_m and R_M are the minimum and maximum sampling rates of the node-sensors, that are adjusted based on (a) the remaining battery charges of the nodes, (b) the availability of the sensors, and (c) the data collecting requirements of the scheme. The model automatically takes into account the effect of *spatial* correlation among the nearby nodes by incorporating the term w_i^{tk} in this model. Also the *inherent* correlation among the sensors are taken into account by putting their cumulative effect inside the $\log$ function.

The optimization problem (6) is a convex optimization problem and so can be solved *centrally* by solving the KKT conditions, which can be done at one of the sink nodes that serves as a *leader* to others. However such a centralized protocol is vulnerable to a single point of failure or attack. Also in a WSN some nodes and links may occasionally go down while conveying the topology information to the

leader, which hampers the calculated sampling rates at the leader. Thus our objective is to adopt an iterative, *distributed* approach to solve this problem. Notice that such node or link failures can happen while running the distributed approach as well, however such failures can be tackled by adopting some interpolation mechanisms. For example in some iteration if a node i cannot successfully send its update to its neighbors, the neighbors can adopt some interpolation mechanisms to estimate the updated value of i, from its recent history.

However solving the problem in a distributed manner brings two key challenges. First, although $\log$ is a strictly concave function with respect to the variables $w_i^{tk}.r_i^t$, the objective function is non-strictly concave function because of the term $\sum_{i=1}^{\mathcal{S}}\sum_{t=1}^{\mathcal{T}} w_i^{tk}.r_i^t$. Second, the function is non-separable with respect to each node i. The first problem can be resolved by adding an *augmented* variable [11], [12] to the objective function to make it strictly-concave. However the new strictly concave function still remains non-separable with respect to i. To cope with this, we adopt the scheme similar to [13] and is described as follows.

As $\log$ is a concave function, by using Jensen's inequality we can obtain

$$\begin{aligned}&\log\left(\sum_{i=1}^{\mathcal{S}}\sum_{t=1}^{\mathcal{T}} w_i^{tk}.r_i^t\right) \geq \sum_{i=1}^{\mathcal{S}}\sum_{t=1}^{\mathcal{T}} \theta_i^{tk}\log\left(\frac{w_i^{tk}.r_i^t}{\theta_i^{tk}}\right) \quad \forall i,\forall t,\forall k\\ &\text{where} \quad \theta_i^{tk} = \frac{w_i^{tk}.r_i^t}{\sum_{i=1}^{\mathcal{S}}\sum_{t=1}^{\mathcal{T}} w_i^{tk}.r_i^t} \quad \forall i,\forall t,\forall k\end{aligned} \tag{7}$$

Using the modified objective function the new optimization problem **MOP** is modeled as follows:

Modified Optimization Problem (MOP):

$$\begin{aligned}\text{Maximize} \quad & U = \sum_{k=1}^{\mathcal{E}}\sum_{i=1}^{\mathcal{S}}\sum_{t=1}^{\mathcal{T}} \theta_i^{tk}\log\left(\frac{w_i^{tk}.r_i^t}{\theta_i^{tk}}\right)\\ \text{subject to} \quad & \sum_t r_i^t\left(e^t + c^t\right) + \sum_t\sum_{j\in\mathbb{D}_i} \alpha_{ji}.r_j^t.c^t \leq E_i \quad \forall i\\ & R_m \leq r_i^t \leq R_M \quad \forall i,\forall j,\forall t\end{aligned} \tag{8}$$

In the following we propose two distributed schemes to solve problem(8). The first one is a sub-gradient based distributed solution, whereas the other one is based on Nesterov's method [4].

A. A Sub-gradient based Distributed Scheme (SDS)

The **MOP** is strictly concave as well as separable in i, for a given θ_i^{tk}. Thus we can now solve this distributively using dual-decomposition as done in the previous *utility maximization problems* [14], [15], [11]. The dual function can be written as follows, where we assume that $\Lambda_i^t = \sum_{k=1}^{\mathcal{E}} \theta_i^{tk}\log\left(\frac{w_i^{tk}.r_i^t}{\theta_i^{tk}}\right)$ and $\Theta_i^t = r_i^t\left(e^t + c^t\right)$.

$$\begin{aligned}D(\nu) &= \text{Max}\sum_{i=1}^{\mathcal{S}}\sum_{t=1}^{\mathcal{T}}\Lambda_i^t - \sum_{i=1}^{\mathcal{S}}\nu_i\left(\sum_t\Theta_i^t + \sum_t\sum_{j\in\mathbb{D}_i}\alpha_{ji}.r_j^t.c^t - E_i\right)\\ &= \text{Max}\sum_{i=1}^{\mathcal{S}}\sum_{t=1}^{\mathcal{T}}\Lambda_i^t - \sum_{i=1}^{\mathcal{S}}\sum_{t=1}^{\mathcal{T}}\nu_i.\Theta_i^t - \sum_{i=1}^{\mathcal{S}}\sum_{t=1}^{\mathcal{T}}\sum_{j\in\mathbb{D}_i}\nu_i.\alpha_{ji}.r_j^t.c^t + \sum_{i=1}^{\mathcal{S}}\nu_i.E_i\\ &= \text{Max}\sum_{i=1}^{\mathcal{S}}\sum_{t=1}^{\mathcal{T}}\Lambda_i^t - \sum_{i=1}^{\mathcal{S}}\sum_{t=1}^{\mathcal{T}}\nu_i.\Theta_i^t - \sum_{i=1}^{\mathcal{S}}\sum_{t=1}^{\mathcal{T}}\sum_{j\in\mathbb{A}_i}\nu_j.\alpha_{ij}.r_i^t.c^t + \sum_{i=1}^{\mathcal{S}}\nu_i.E_i\\ &= \text{Max}\sum_{i=1}^{\mathcal{S}}\sum_{t=1}^{\mathcal{T}}\left(\Lambda_i^t - \nu_i.\Theta_i^t - \sum_{j\in\mathbb{A}_i}\nu_j.\alpha_{ij}.r_i^t.c^t\right) + \sum_{i=1}^{\mathcal{S}}\nu_i.E_i\end{aligned}$$

and the dual is to $\min_{\nu_i\geq 0} D(\nu)$. ν is the Lagrange multiplier vector, which can be iteratively updated using

$$\begin{aligned}\nu_i^{\ell+1} &= \left[\nu_i^\ell + \gamma\left(\sum_t r_i^t\left(e^t + c^t\right) + \sum_t\sum_{j\in\mathbb{D}_i}\alpha_{ji}.r_j^t.c^t - E_i\right)\right]^+\\ &= \left[\nu_i^\ell + \gamma\left(\sum_t r_i^t.e^t + \left(\sum_t r_i^t + \sum_t\sum_{j\in\mathbb{D}_i}\alpha_{ji}.r_j^t\right)c^t - E_i\right)\right]^+\end{aligned} \tag{9}$$

where $[x]^+ = \max(x,0)$ and γ is the step-size of the sub-gradient method. With the updated Lagrange multipliers, the following optimization problem is solved to update the sampling rates of individual nodes.

$$\begin{aligned}&\text{Max}\sum_{k=1}^{\mathcal{E}}\theta_i^{tk}\log\left(\frac{w_i^{tk}.r_i^t}{\theta_i^{tk}}\right) - \nu_i.r_i^t\left(e^t + c^t\right) - \sum_{j\in\mathbb{A}_i}\nu_j.\alpha_{ij}.r_i^t.c^t\\ &\Rightarrow r_i^t = \left[\frac{\sum_{k=1}^{\mathcal{E}}\theta_i^{tk}}{\nu_i(e^t + c^t) + \sum_{j\in\mathbb{A}_i}\nu_j.\alpha_{ij}.c^t}\right]_{R_m}^{R_M}\end{aligned} \tag{10}$$

Theorem 1: The proposed distributed version of **MOP** converges to the optimal solution of the original problem **OP**.

Proof: Let us define $(\boldsymbol{r^*}, \boldsymbol{\nu^*}, \boldsymbol{\theta^*})$ are the optimal solution of **MOP**. It can be shown that $(\boldsymbol{r^*}, \boldsymbol{\nu^*})$ also satisfies the KKT condition of the original problem **OP**. The KKT condition of the **MOP** is given by

$$\begin{aligned}&\frac{\partial}{\partial r_i^{tk}}\left(\sum_{k=1}^{\mathcal{E}}\sum_{i=1}^{\mathcal{S}}\sum_{t=1}^{\mathcal{T}}\theta_i^{tk}\log\left(\frac{w_i^{tk}.r_i^t}{\theta_i^{tk}}\right)\right)\Bigg|_{\boldsymbol{r^*}}\\ &\qquad - \sum_{i=1}^{\mathcal{S}}\nu_i^*\left(\sum_t\left(e^t + c^t\right) + \sum_t\sum_{j\in\mathbb{A}_i}\alpha_{ij}.c^t\right) = 0\\ &\nu_i^*.r_i^{t^*}\left(e^t + c^t\right) - \sum_{i=1}^{\mathcal{S}}\sum_{t=1}^{\mathcal{T}}\sum_{j\in\mathbb{D}_i}\nu_i^*.\alpha_{ji}.r_j^{t^*}.c^t + \sum_{i=1}^{\mathcal{S}}\nu_i^*.E_i = 0\\ &\nu_i^* \geq 0\end{aligned} \tag{11}$$

Notice that the last two KKT conditions of **OP** and **MOP** are identical. Now as $\frac{\partial}{\partial r_i^{tk}}\left(\sum_{k=1}^{\mathcal{E}}\sum_{i=1}^{\mathcal{S}}\sum_{t=1}^{\mathcal{T}}\theta_i^{tk}\log\left(\frac{w_i^{tk}.r_i^t}{\theta_i^{tk}}\right)\right)\Big|_{\boldsymbol{r^*}} = \frac{\partial}{\partial r_i^{tk}}\left(\sum_{k=1}^{\mathcal{E}}\log\left(\sum_{i=1}^{\mathcal{S}}\sum_{t=1}^{\mathcal{T}} w_i^{tk}.r_i^t\right)\right)\Big|_{\boldsymbol{r^*}}$, the first condition of **OP** and **MOP** are also identical at point $(\boldsymbol{r^*}, \boldsymbol{\nu^*})$. Thus the proof follows. ■

We first need to make sure that the information needed to perform this iterative scheme can be obtained from the neighborhood information. To do that we need to modify equation (9) and equation (10) so that the nodes only need to exchange *local* information in between each others, i.e. information exchange in between the PPs and their children are only needed. We assume that the total transmitted traffic by the node i is given by

$$\begin{aligned}T_i &= \sum_t r_i^t + \sum_t\sum_{j\in\mathbb{D}_i}\alpha_{ji}.r_j^t = \sum_t r_i^t + \sum_{i\in\mathbb{P}_j}\alpha_{ji}.T_j \ \text{ for non-leaf nodes}\\ &= \sum_t r_i^t \ \text{ for leaf nodes}\end{aligned} \tag{12}$$

Thus the nodes can calculate their individual traffic T_i by collecting the carried traffic by their children, which is used to calculate the Lagrange multipliers from equation (9). Similarly the individual sampling rates of the nodes can be obtained from equation (10) by assuming $F_i = \sum_{j\in\mathbb{A}_i}\nu_j.\alpha_{ij}$ and obtaining the (ν_j, F_j)s of the PPs as follows:

$$\begin{aligned}F_i &= \sum_{j\in\mathbb{A}_i}\nu_j.\alpha_{ij} = \sum_{j\in\mathbb{P}_i}\alpha_{ij}\left(\nu_j + F_j\right) \ \text{ for non-CH nodes}\\ &= 0 \ \text{ for CH nodes}\end{aligned} \tag{13}$$

With these we now propose the distributed mechanism to calculate the optimal sampling rates for the node sensors. In the proposed scheme the node's traffic T_i and ν_i are updated in a bottom-up manner, whereas the (F_i, r_i^t) are updated in a top-down fashion. The nodes broadcast their α to their neighbors periodically. If a node i has $|\mathbb{P}_i|$ PPs with path-ETX of P-ETX$_{ij}$, $\forall j \in \mathbb{P}_i$ if it sends its packets through j then $\alpha_{ij} = \frac{\frac{1}{\text{P-ETX}_{ij}}}{\sum_{x\in\mathbb{P}_i}\frac{1}{\text{P-ETX}_{ix}}}$. This is based on the intuition that the nodes choose their routes in proportion to their route qualities. All the nodes then calculate their (ν_i, F_i, T_i, r_i^t) based on their local information as follows. The CHs broadcast their F_i to be zero to their children. The immediate children of the CHs calculate their F_i as well as their sampling rates using equation(13) and (10) and broadcast. In the first round, the θ_i^{tk} and ν_i are assumed to be some arbitrary value for all the nodes. This process goes on until the leaf nodes are reached. The leaf nodes then update their carried traffic T_i and broadcast. With the updated T_i, they also calculate their ν_i using equation(9). The subsequent PP nodes update their traffic using equation (12) as well as their ν_i, and this process goes on until it reaches the CHs. While calculating the sampling rates, each leaf node i also calculates the $\alpha_{ij}.w_i^{tk}.r_i^t$ $\forall k$, corresponding to each PP j, and broadcast to their PPs. The PPs also calculate their $w_i^{tk}.r_i^t$, add it with the same of their children and broadcast (after multiplying α_{ij}), until it reaches their CH. In the next iteration, the CHs collaborate to calculate the value of $\mathbb{TWR}^k = \sum_i \sum_t w_i^{tk}.r_i^t$ $\forall k$ and propagate it in their own cluster in the top-down fashion, which the nodes use while calculating their θ_i^{tk} as shown in equation(7). This top-down and bottom-up process continues iteratively.

We assume that the required message exchanges for the top-down and bottom-up operations are successfully delivered to the corresponding nodes. However in reality there may be packet loss due to neighboring interference. In such situations, some interpolation mechanisms can be adopted based on the historical values sent by the corresponding nodes. We assume that the ETX of the links remain constant throughout the iterative process. We also assume that the parent-child relations remain the same throughput the iterative process which in reality will vary due to the varying channel conditions. We will explore the effects of these practical deployment issues on the convergence of this iterative scheme in future.

B. Nesterov's Gradient Descent based Distributed Scheme (NDS)

We propose another scheme using Nesterov's method [4] by utilizing the Lipschitz continuity property of the dual objective function. The resulted scheme usually achieves faster convergence with respect to the traditional sub-gradient method, however in specific scenarios the scheme may perform worse than SDS as reported in section III-C.

We first develop a Lipschitz constant for the optimization problem (6). We consider a $\mathcal{S}\times\mathcal{S}.\mathcal{T}$ matrix $\mathbf{A}$, where the rows represent the set of nodes and columns represent the set of all sensors. The entries in the A_{ij} represent the power consumption of a node i for the sampling or transmission of sensor j's readings. $\mathbf{R}$ and $\mathbf{E}$ are the $\mathcal{S}.\mathcal{T}\times 1$ and $\mathcal{S}\times 1$ vector that represent the set of the sensors sampling rates and the available energy of the nodes respectively. Using the above notations the energy conservation constraint of problem (6) is written as $\mathbf{A}.\mathbf{R} \leq \mathbf{E}$. Clearly any entry of A_{ij} has a maximum value of $\Psi = \max(e^t + c^t), \forall t$. Also assume that $\mathbb{E} = \max\{E_i\}$ $\forall\, i$, and $\varepsilon = \min\{e^t\}$ $\forall\, t$, then the maximum sampling rate that a node can achieve for a particular sensor is given by $\Upsilon = \frac{\mathbb{E}}{\varepsilon}$. We also assume that $\Omega = \max\{c^t\}$ $\forall\, t$, $\Gamma = \sum_t (e^t + c^t)$, and $\Delta = \sum_t c^t$. With these we propose the following theorems.

Theorem 2: The dual of problem (6) is Lipschitz continuous with constant given by $L = \frac{\rho(\mathbf{A^T A})}{\sigma}$, where the objective function U is σ-strongly concave.

Proof: Please refer to [16] for the proof. ■

Theorem 3: The dual of problem (6) is Lipschitz continuous with constants $L_1 = \frac{\mathcal{S}.\Upsilon^2.(\Gamma+(\mathcal{S}-1).\Delta).(\Upsilon+(\mathcal{S}-1).\Omega)}{\mathcal{E}.W_m}$, and $L_2 = \frac{\mathcal{S}.\mathcal{T}.R_M^2.(\Gamma+(\mathcal{S}-1).\Delta).(\Upsilon+(\mathcal{S}-1).\Omega)}{\mathcal{E}.W_m}$, where $W_m = \min\{w_i^{tk}\}$ $\forall\, i,t,k$.

Proof: We first derive the σ of U, based on the strongly-concavity property of equation(5) as follows:

$$\begin{aligned}\sigma &= \min\left(-\frac{\partial^2 U}{\partial {r_i^t}^2}\right) = \min\left(\frac{\sum_k \theta_i^{tk}}{{r_i^t}^2}\right) = \min\left(\frac{\sum_k w_i^{tk}.r_i^t}{{r_i^t}^2.\sum_{i=1}^{\mathcal{S}}\sum_{t=1}^{\mathcal{T}} w_i^{tk}.r_i^t}\right)\\ &= \min\left(\frac{\sum_k w_i^{tk}}{r_i^t.\sum_{i=1}^{\mathcal{S}}\sum_{t=1}^{\mathcal{T}} w_i^{tk}.r_i^t}\right) \geq \min\left(\frac{\sum_k w_i^{tk}}{r_i^t.\sum_{i=1}^{\mathcal{S}}\sum_{t=1}^{\mathcal{T}} r_i^t}\right)\\ &\geq \frac{\mathcal{E}.W_m}{\mathcal{S}.\Upsilon^2}\end{aligned} \tag{14}$$

σ can also be expressed in terms of R_M as follows:

$$\sigma \geq \min\left(\frac{\sum_k w_i^{tk}}{r_i^t.\sum_{i=1}^{\mathcal{S}}\sum_{t=1}^{\mathcal{T}} r_i^t}\right) \geq \frac{\mathcal{E}.W_m}{\mathcal{S}.\mathcal{T}.R_M^2} \tag{15}$$

We next calculate the spectral radius of $\mathbf{A^T}.\mathbf{A}$ as follows. While deriving $\rho(\mathbf{A^T}.\mathbf{A})$, we require the values of $\|\mathbf{A}\|_\infty$ and $\|\mathbf{A}\|_1$ (refer to equation(2)), which are the maximum sum of the rows and columns of matrix $\mathbf{A}$ respectively. We next calculate L by using the values of $\rho(\mathbf{A^T}.\mathbf{A})$ and σ, by using Theorem 2.

$$\begin{aligned}&\|\mathbf{A}\|_\infty \leq \sum_t \left(e^t + c^t\right) + (\mathcal{S}-1)\sum_t c^t = \Gamma + (\mathcal{S}-1).\Delta\\ &\|\mathbf{A}\|_1 \leq \max_t\{e^t + c^t\} + (\mathcal{S}-1)\max_t\{c^t\} = \Psi + (\mathcal{S}-1).\Omega\\ &\rho(\mathbf{A^T A}) \leq \|\mathbf{A}\|_1.\|\mathbf{A}\|_\infty \leq (\Gamma + (\mathcal{S}-1).\Delta).(\Psi + (\mathcal{S}-1).\Omega)\\ &L_1 = \frac{\mathcal{S}.\Upsilon^2.(\Gamma + (\mathcal{S}-1).\Delta).(\Psi + (\mathcal{S}-1).\Omega)}{\mathcal{E}.W_m} \text{ from equation(14)}\\ &L_2 = \frac{\mathcal{S}.\mathcal{T}.R_M^2.(\Gamma + (\mathcal{S}-1).\Delta).(\Psi + (\mathcal{S}-1).\Omega)}{\mathcal{E}.W_m} \text{ from equation(15)}\end{aligned} \tag{16}$$

■

We assume a Lipshitz constant $L = \min\{L_1, L_2\}$. We assume that the CHs calculate the L and broadcast to their cluster nodes. We utilize the Lipschitz gradient property of the dual objective function of problem(8) to develop a gradient descent scheme using Nesterov's algorithm [4], [17]. The scheme, presented in Algorithm 1, can be implemented in a distributed fashion by each node i. In Algorithm 1, Δf of a node i is assumed to be $\Delta f_i = \sum_t r_i^t.e^t + \left(\sum_t r_i^t + \sum_t \sum_{j\in\mathbb{D}_i} \alpha_{ji}.r_j^t\right)c^t - E_i$. Each node solves the objective function to get its sampling rate in line 5. In line 6, $u_i^{\ell+1}$ is the solution of the standard gradient descent (identical to equation (9)) with step-size $\frac{1}{L}$ at the k-th iteration, which encodes the *current* gradient. On the other hand in line 8, $v_i^{\ell+1}$ is the solution of the gradient descent step that proceeds

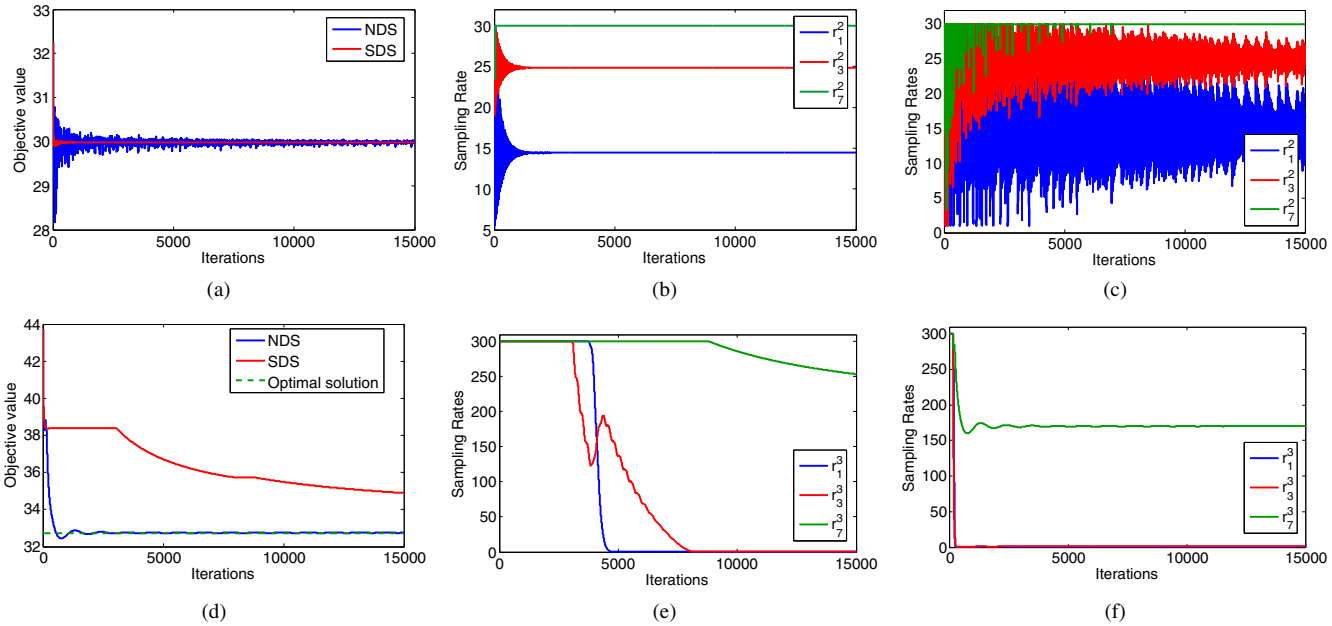

Fig. 1. Comparison of (a) convergence, (b) sampling rates of SDS, (c) sampling rates of NDS with (R_m, R_M) = (1, 30). Similar comparison of (d) convergence, (e) sampling rates of SDS, (f) sampling rates of NDS with (R_m, R_M) = (1, 300).

along a direction determined by the weighted sum of the negative gradients in all the previous iteration rounds, thus this encodes the *historical* gradients. The later iteration rounds have larger weights than the earlier ones. Finally $\nu_i^{\ell+1}$ calculates the weighted sum of the current gradient and the historical gradients in line 9.

Algorithm 1 Nesterov's Gradient Descent based Distributed Scheme (NDS)

1: INPUT : R_m, R_M, W_m.
2: OUTPUT : Sampling rates r_i^{tk} $\forall i \in \{1, 2, ..., N\}$.
3: tmp$_i$ = 0;
4: **for** each iteration $\ell = \{1, 2, ..., \mathcal{L}\}$ **do**
5: Update rate r_i^{tk} and θ_i^{tk};
6: $u_i^{\ell+1} = \left[\nu^{\ell} + \frac{\Delta f_i}{L}\right]_{-}^{+}$;
7: tmp$_i$ = tmp$_i$ + $\frac{\ell+1}{2}.\Delta f_i$;
8: $v_i^{\ell+1} = \left[\frac{\text{tmp}_i}{L}\right]_{-}^{+}$;
9: $\nu_i^{\ell+1} = \frac{\ell+1}{\ell+3}.u_i^{\ell+1} + \frac{2}{\ell+3}.v_i^{\ell+1}$;
10: **end for**

C. Validation of SDS and NDS

We have assumed a hypothetical scenario to monitor the health of power equipments in a substation. The sensor nodes run on batteries that are assumed to have a battery capacity of 5000 mAHr [8]. Unless otherwise mentioned, we assume that the nodes use asynchronous Low Power Listening (X-Mac) that makes them sleep most of the time and wake-up periodically to check the channel activity. The nodes wakes-up 8 times per second to check if the channel is busy, which makes the radio on time for transmission $\sim$140 ms. The nodes consume $\sim$20 mA [18] at the transmit mode, which is the representative of Crossbow's MICAz wireless sensor nodes [18]. The nodes are expected to remain active for 12 months, and the power budgets for sensing and forwarding are calculated accordingly. The nodes use there sensors: sound, SF_6 gas and temperature. The power consumption of these sensors are 9.5, 150, 7.5 mA respectively with a sampling time of 7000, 400 and 112 milliseconds [8].

We compare the proposed distributed rate adaption scheme of **MOP** along with the solution of **OP** obtained from AMPL solver [19]. The result is shown in Fig. 1 where 14 nodes with fully charged batteries are deployed in a binary tree structure with a height of 3, rooted at a CH. R_m and w_i^{tk} are assumed to be 1/hr. and 0.5 respectively. We assume the step-size γ of the SDS to be $\frac{1}{L}$. We consider two cases depending on the value of R_M; in the first case (Fig. 1(a)-(c)) R_M is assumed to be 30/hr. whereas it is assumed to be 300/hr. in the second case (Fig. 1(d)-(f)). Notice that in the first case NDS performs poorly compared to SDS in terms of convergence rate. We surmise that because of small R_M the calculation of r_i^t from equation(10) frequently becomes R_M and so keeping track of the historical gradients does not make the scheme faster. On the other hand in the second case NDS outperforms SDS in terms of convergence speed as shown in Fig. 1(d)-(f). In both cases we observe that the distributed version of **MOP** closely matches (Fig. 1(a) does not show the optimal solution for clarity) with the original problem of (6), which validates the claim of Theorem 1.

Fig. 1(b)-(c) and Fig. 1(e)-(f) show the sampling rates of nodes 1, 3 and 7 which are the representatives of the first, second and third level nodes of the binary tree. From these figures we can observe that with identical battery charges, the first level nodes have sampling rates higher than the others. This is because of the fact that assigning more sampling tasks to the first level nodes reduces the effect of forwarding traffic, which improves the overall utility. We can also observe that in both SDS and NDS schemes, the convergence time is significantly high due to large number of iterations (in the order of few hundreds to thousands) which limits their usefulness in large-scale WSN applications.

IV. Scalable Alternatives

As the convergence time of the fully distributed versions are too high for a general multi-hop network, solving this problem at the node levels is simply not practical because of (a) too many control messages need to be exchanged in between

the nodes and the CHs, which will drain their batteries, (b) the convergence time is significantly high due to the large number of nodes in a geographic area. The main reason for the high convergence time is multi-hopping, as the individual nodes and their ancestors need to gradually adapt their sampling rates so that the higher level nodes (ancestors) do not run out of battery while forwarding traffic from their descendants, and at the same time the overall utility is maximized. Thus the above mentioned distributed schemes are not scalable especially for large WSNs. We next propose two alternatives to overcome this limitation. The first scheme is a semi-distributed approach where the CHs collaborate and decide the sampling rates of the nodes, after collecting the topology information and battery profiles of the nodes. The second alternative is based on the assumption that the packet transmissions in between the nodes follow some scheduled mechanisms and so the power consumption due to data forwarding is negligible.

Algorithm 2 Semi-Distributed rate adaptation (SDRA)

1: INPUT : Node's battery and connectivity profiles, R_m, R_M.
2: OUTPUT : Sampling rates r_i^{tk} $\forall i \in \{1, 2, ..., \mathcal{S}\}$.
3: **while** not converged **do**
4: **for** each CH_ℓ **do**
5: Update the sampling rate r_i^{tk} and θ_i^{tk};
6: Calculate $\mathbb{WR}_\ell^k = \sum_{i \in \boldsymbol{\zeta_\ell}} \sum_t w_i^{tk}.r_i^t$ $\forall k$;
7: Transmit $\mathbb{WR}_\ell^k$ to the CH;
8: **end for**
9: CHs collaboratively calculate the $\mathbb{TWR}^k = \sum_\ell \mathbb{WR}_\ell^k$;
10: **end while**

A. A Semi-distributed Collaborative Approach

We propose a *Semi-Distributed Rate Adaptation (SDRA)* scheme where the cluster-heads collect the topology information and the remaining battery power of the nodes in its own cluster and then collaborate with each other to decide the sampling rates of the nodes iteratively. The overall scheme is shown in Algorithm 2. We assume that there are $\mathcal{C}$ cluster-heads, where CH_ℓ is associated with a set $\boldsymbol{\zeta_\ell}$ of $\mathcal{S}_\ell$ nodes. Thus the CHs solve $\mathcal{C}$ subproblems in a distributed manner, that are linked with the parameter θ_i^{tk}. At first the CHs collect the required connectivity information and remaining battery capacities of the nodes. Each CH_ℓ solves problem (8) locally (using $\mathcal{S} = \mathcal{S}_\ell$ and $i \in \boldsymbol{\zeta_\ell}$) using any random θ_i^{tk} (at the first iteration), which can be done using any existing solvers like AMPL [19], GLPK [20], CVX [21] etc. It then calculates its local weighted rate $\mathbb{WR}_\ell^k = \sum_{i \in \boldsymbol{\zeta_\ell}} \sum_t w_i^{tk}.r_i^t$ $\forall k$ and broadcast it to the other CHs. The CHs collaboratively calculate the total weighted rate $\mathbb{TWR}^k = \sum_\ell \mathbb{WR}_\ell^k = \sum_i \sum_t w_i^{tk}.r_i^t$. The CHs then calculate $\theta_i^{tk} = \frac{w_i^{tk}.r_i^t}{\mathbb{TWR}^k}$ (equation (7)) and solve its local optimization problem using the new θ_i^{tk}. This process goes on until the solution converges. Upon convergence the calculated sampling rates are sent to the nodes by their CHs.

Notice that there is an inherent trade-off in between the number of clusters, the number of message exchanges in between them, and resiliency. Less number of cluster reduces the number of message exchanges in between them, but less resilient to node/link failure (with respect to some fixed error tolerance) or attack, whereas more clusters improve the resilience at the cost of higher control overhead.

B. An Approximation Scheme for Scheduled WSNs

We also propose an *Approximation scheme for Distributed Rate Adaptation (ADRA)*, which is applicable only in a scheduled WSN. In a scheduled WSN the nodes wake up in their own schedule. The nodes know the schedule of their neighbors, thus they wait for the schedule of their parents, switch on their transmitters for that particular time and transmit. In this scenario, the nodes do not need to send multiple strobe packets as done in X-Mac, and so the transmission power consumption is very low. In this case the second term in the energy budget equation $\left(\sum_t \sum_{j \in \mathbb{D}_i} \alpha_{ji}.r_j^t.c^t \text{ in equation(8)}\right)$ can be neglected. Thus the multi-hopping term of problem(8) is ignored, and then the problem can be solved in a distributed fashion fairly quickly using Algorithm 3. In this scheme, the nodes maintain two sets of sensors: unassigned U and assigned A. Initially, all sensors belong to set U, but are transferred to set A as rates are assigned for them. The node first assigns the sampling rates to each sensor t as $r_i^t = R_m + \frac{\Delta_i . \sum_k \theta_i^{tk}}{\sum_t \sum_k \theta_i^{tk}(e^t + c^t)}$ (line 4-6). The difference between assigned sampling rate r_i^t and maximum sampling rate R_M is stored in diff[t]. After this sampling rate assignment, if for any sensor ℓ, diff[ℓ] $<$ 0, the node assigns its sampling rates as its maximum rate R_M (line 13) and divides the Δ_i fairly among other sensors (line 14-17). This process goes on until all the sensors are assigned their sampling rates. After calculating the sampling rates, the nodes calculate $\alpha_{ij}.w_i^{tk}.r_i^t$ $\forall k$ and broadcast to their PPs, which is used for calculate $\mathbb{TWR}^k = \sum_i \sum_t w_i^{tk}.r_i^t$ $\forall k$ at the CHs as mentioned in SDS. Using the new $\mathbb{TWR}^k$ the nodes next calculate the updated sampling rates of the sensors by putting the new θ_i^{tk} in ADRA, and this process goes on until the solution converges.

Algorithm 3 Approximation scheme for Distributed Rate Adaptation (ADRA)

1: INPUT : θ_i^{tk}, R_m, R_M.
2: OUTPUT : Sampling rates r_i^t $\forall t$.
3: A=$\{\phi\}$; U=$\{1, ..., \mathcal{T}\}$;
4: Assign $r_i^t = R_m$ $\forall t$
5: $\Delta_i = E_i - \sum_t r_i^t$;
6: $r_i^t = r_i^t + \frac{\Delta_i . \sum_k \theta_i^{tk}}{\sum_t \sum_k \theta_i^{tk}(e^t + c^t)} \forall t$;
7: diff[t] = $R_M - r_i^t \forall t$;
8: **for** each sensor $x = \{1, 2, ..., \mathcal{T}\}$ **do**
9: Sort node $\in U$ in increasing order of diff[x];
10: Put them in order in list L;
11: ℓ = $L[0]$;
12: **if** diff[ℓ] < 0 **then**
13: $r_i^\ell = R_M$; $A = A \cup t$; $U = U \backslash t$; Δ_i = abs(diff[ℓ]) $\times (e^\ell + c^\ell)$
14: **for** each sensor $t = \{1, 2, ..., \mathcal{T}\}$ AND $t \in U$ **do**
15: $r_i^t = r_i^t + \frac{\sum_k \theta_i^{tk}}{\sum_{i \in U} \sum_k \sum_t \theta_i^{tk}} . \frac{\Delta_i}{(e^t + c^t)}$;
16: diff[t] = $R_M - r_i^t$;
17: **end for**
18: diff[ℓ] = 0;
19: **else**
20: EXIT
21: **end if**
22: **end for**
23: return r_i^t $\forall t$

Theorem 4: For a given θ_i^{tk}, Algorithm 3 gives optimal rate allocation of the node sensors.

Proof: Line 7 can be derived by solving the KKT conditions of the problem(8), ignoring the multi-hop term. The rest of the proof is similar to [22] and is omitted here. ■

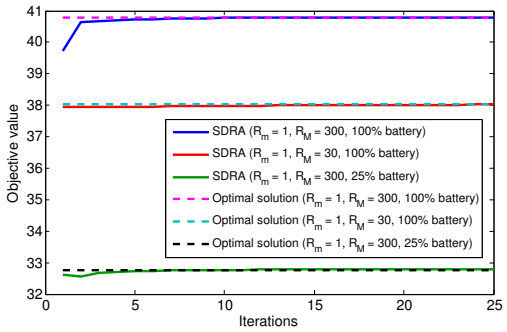

Fig. 2. Convergence of SDRA with different R_m, R_M and battery charges.

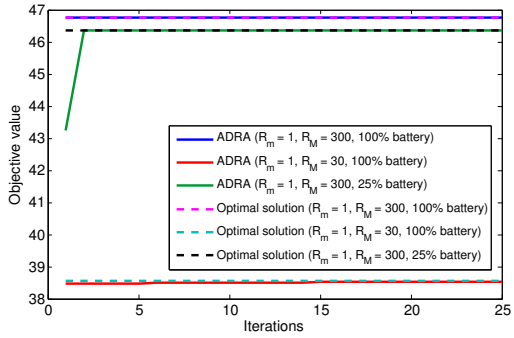

Fig. 3. Convergence of ADRA with different R_m, R_M and battery charges.

C. Validation of SDRA and ADRA

Fig. 2 and Fig. 3 show the convergence of SDRA and ADRA with different remaining capacities and R_M. For these figures we assume 5 clusters each one having 14 nodes, placed in a binary tree fashion with a height of 3. In both schemes, the objective values match with the optimal solution obtained from AMPL. We can also observe that the schemes converge within $\sim$10-15 iterations, which make them suitable as a scalable solution in energy-constrained WSNs.

V. Performance Evaluation

We study the effectiveness of the proposed sampling rate adaptations of the nodes, depending on their energy budgets. We uniformly deploy the wireless devices in a geographic area of 200$\times$200 sq. m. The fresh battery capacity of the nodes are assumed to be 5000 mAHr. The remaining battery capacities of the nodes are uniformly randomly chosen among (25-100%). Depending on their remaining battery capacities, we divide the nodes in three *tiers*. The first, second, and third tier nodes have battery charges of (75-100%), (50-75%), (25-50%) respectively. R_m and R_M are assumed to be 1/hr. and 30/hr. respectively. The types of sensors and their corresponding power consumption parameters are kept identical to those in section III-C.

Fig. 4(a) shows the mean energy budget of the nodes at different tiers, Fig. 4(b) shows their corresponding usages. The usage is defined as the cumulative energy expenditure due to sampling and forwarding of the packets. From Fig. 4(b) we can observe that the energy usages are adapted proportional to the individual node's power budgets. Fig. 4(c)-(e) show the mean sampling rates of sound sensor, SF_6 gas sensor and temperature sensors at various tiers. From these figures we can observe that overall the first tier nodes have a sampling rate much higher than other tiers, due to their higher battery charges. We can also observe that the average number of samples of the sound sensors are much less ($\sim$ 14-25 times) compared to the other two sensors. The reason is due to the higher power consumption of the sound sensors while data capturing. Fig. 4(f) shows the overall sampling rates of various sensors, which also confirms the fact that the proposed scheme tries to conserve energy by mostly switching off sensors with higher power consumption. This shows the adaptive ability of the scheme based on the individual node's power budget as well as the power consumption of the individual sensors.

We can also observe that the sampling rates of the individual sensors do not change significantly with the number of nodes, even if the total traffic of the entire network are higher with more nodes. This is because of the uniform distribution of the nodes, which ensures that even of the cumulative network traffic grows with the number of nodes, the average number of data forwarding by the individual nodes are similar.

VI. Related Works

A. Network Utility Maximization

Network utility maximization has received a significant attention in the last decade ever since the seminal framework the seminal work in [14], [15]. In these works the users utility is assumed to be strictly concave function of users rate, and the resource constraints are set to be linear. The users distributedly maximize their aggregate utility under their resource constraints. Various types of fairness based utilities are discussed in [23]. Multi-path utility maximizations are addressed in [11], [12], [13] where the utility function is non-strictly concave with respect to the individual users rate due to multi-path routing. To convexify the utility function, proximal approach are proposed in [11], [12] whereas in [13] the authors have proposed a modified strictly concave utility function and proposed a successive approximation method.

B. Collaborative Sensing, Rate Adaptation and Energy Awareness in WSNs

Energy management in sensor network is a well researched area. Control of sleep/wakeup cycle is a standard technique that is explored in several MAC proposals [24], [25]. Other techniques for reducing energy consumption include data compression and source coding [26], transmit power control [27], [28], multiple channel assignment [29], [30] etc. Sampling rate adaptation for wireless sensor networks has been widely studied. The authors in [31], [32] have proposed a fair rate adaptation for interference or congestion control, whereas several energy aware rate allocation schemes are discussed in [33], [5]. In [22], [34], the authors have proposed collaborative and distributed sampling rate adaptation schemes for energy harvesting sensor networks. Collaborative heterogeneous sensing is introduced in [6], where multiple and different types of sensors per node is considered, along with their inter-dependencies in the event reporting process.

VII. Conclusions

In this paper we propose a novel framework for collaborative sampling rate adaptation in a multi-sensor equipped heterogeneous WSN. We propose effective mechanisms to utilize the rich multi-modal sensory capability of wireless devices to collectively monitor few interest points in a geographic area, and discuss several distributed and collaborative sampling rate adaptation schemes that allows spatially correlated devices to share the data capturing tasks among themselves based on

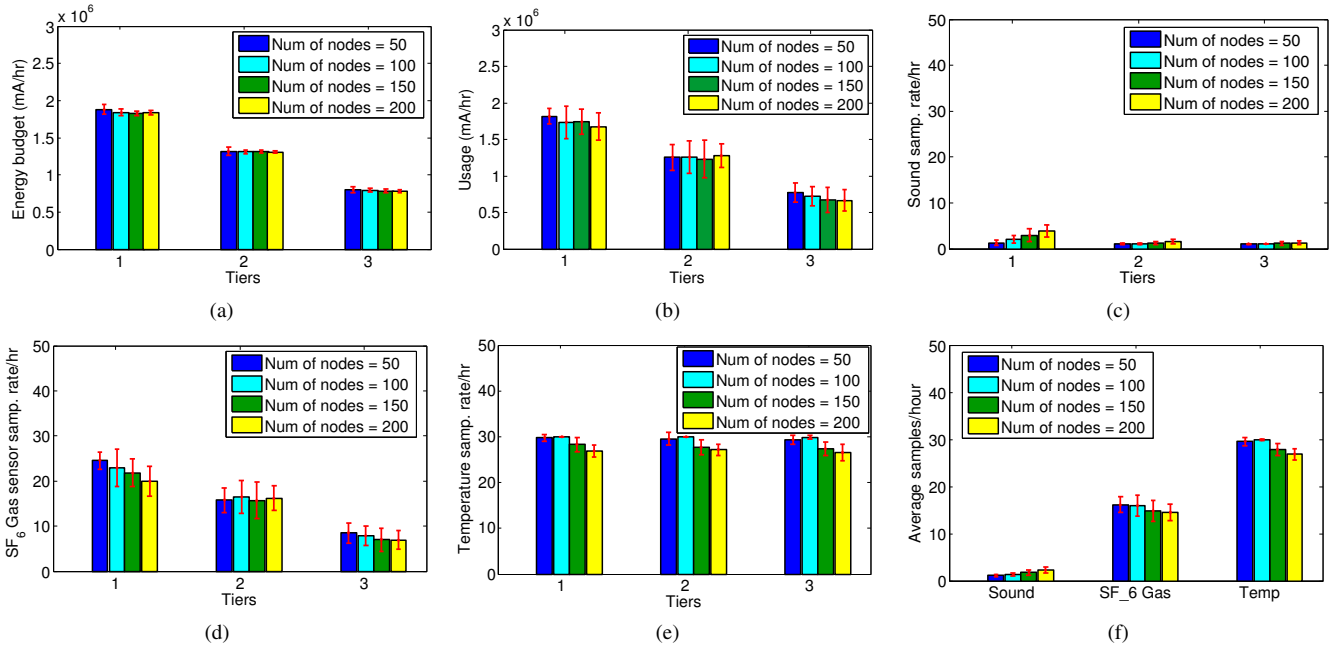

Fig. 4. (a) Average energy budget and (b) usage profiles of the nodes. Sampling rates of (c) sound sensor, (d) SF_6 gas sensor, and (e) temperature sensor of the different tier nodes. (f) Average sampling rates of different sensors.

their available energy and network participation for improved energy efficiency. We utilized general sub-gradient method and Nesterov's gradient descent method to solve this distributed rate adaptation problem. We also proposed two hybrid solutions to overcome the slow convergence rate of the fully distributed solutions. The paper has done a preliminary study of the feasibility of the distributed schemes with a number of assumptions like static channel conditions and parent-child relations, no packet loss etc. In reality all these practical issues bring inaccuracies into the model, many of which can be approximated through interpolation mechanisms. In future we want to study the proposed distributed rate adaptation schemes considering the practical deployment issues with realistic channel models and interference effects.

REFERENCES

[1] M.Raj *et al.*, "E-DARWIN: energy aware disaster recovery network using wifi tethering," in *IEEE ICCCN*, 2014, pp. 1–8.

[2] P. W.Gething *et al.*, "Can mobile phone data improve emergency response to natural disasters?" *PLoS Medicine*, vol. 8, 2011.

[3] C.Yang *et al.*, "Use of mobile phones in an emergency reporting system for infectious disease surveillance after the sichuan earthquake in china," *Bulletin of the World Health Organization*, vol. 87, pp. 619–623, 2009.

[4] I.Necoara *et al.*, "Application of a smoothing technique to decomposition in convex optimization," *IEEE Trans. Automat. Contr.*, vol. 53, no. 11, pp. 2674–2679, 2008.

[5] R.Liu *et al.*, "Joint energy management and resource allocation in rechargeable sensor networks," in *INFOCOM*, 2010, pp. 902–910.

[6] A.Pal *et al.*, "Collaborative heterogeneous sensing: An application to contamination detection in water distribution networks," in *IEEE ICCCN*, 2015.

[7] M. V.Ramesh, "Design, development, and deployment of a wireless sensor network for detection of landslides," *Ad Hoc Networks*, vol. 13, pp. 2–18, 2014.

[8] A.Nasipuri *et al.*, "Design considerations for a large-scale wireless sensor network for substation monitoring," in *LCN*, 2010, pp. 866–873.

[9] http://www.tinyos.net/tinyos-2.x/doc/html/tep105.html.

[10] https://github.com/contiki-os/contiki/wiki/Radio-duty-cycling.

[11] W.Wang *et al.*, "Optimal flow control and routing in multi-path networks," *Performance Evaluation*, vol. 52, no. 2-3, pp. 119–132, 2003.

[12] X.Lin *et al.*, "Utility maximization for communication networks with multipath routing," *IEEE Trans. Automat. Contr.*, vol. 51, no. 5, pp. 766–781, 2006.

[13] P. L.Vo *et al.*, "Multi-path utility maximization and multi-path TCP design," *J. Parallel Distrib. Comput.*, vol. 74, no. 1, pp. 1848–1857, 2014.

[14] F.Kelly, "Charging and rate control for elastic traffic," *European Transactions on Telecommunications*, vol. 8, pp. 33–37, 1997.

[15] S. H.Low *et al.*, "Optimization flow control-i: basic algorithm and convergence," *IEEE/ACM Transactions on Networking*, vol. 7, no. 6, pp. 861–874, 1999.

[16] A.Beck *et al.*, "An o(1/k) gradient method for network resource allocation problems," *IEEE Transactions on Control of Network Systems*, vol. 1, no. 1, pp. 64–73, 2014.

[17] P.Tsiaflakis *et al.*, "Improved dual decomposition based optimization for DSL dynamic spectrum management," *IEEE Transactions on Signal Processing*, vol. 58, no. 4, pp. 2230–2245, 2010.

[18] http://www.memsic.com/userfiles/files/Datasheets/WSN/micaz_datasheet-t.pdf.

[19] http://ampl.com/products/solvers/.

[20] https://www.gnu.org/software/glpk/.

[21] http://cvxr.com/cvx/doc/solver.html.

[22] A.Pal *et al.*, "Water flow driven sensor networks for leakage and contamination monitoring," in *IEEE WoWMoM*, 2015.

[23] J.Mo *et al.*, "Fair end-to-end window-based congestion control," *IEEE/ACM Trans. Netw.*, vol. 8, no. 5, pp. 556–567, 2000.

[24] Y.Kim *et al.*, "Y-MAC: An energy-efficient multi-channel mac protocol for dense wireless sensor networks," in *IPSN*, 2008, pp. 53–63.

[25] J.Zhang *et al.*, "TMMAC: An energy efficient multi-channel mac protocol for ad hoc networks," in *ICC*, 2007, pp. 3554–3561.

[26] C.Tang *et al.*, "Compression techniques for wireless sensor networks," pp. 207–231, 2004.

[27] A.Pal *et al.*, "A distributed power control and routing scheme for rechargeable sensor networks," in *IEEE SoutheastCon*, 2013.

[28] A.Pal *et al.*, "PCOR: A joint power control and routing scheme for rechargeable sensor networks," in *IEEE WCNC*, 2014, pp. 2230–2235.

[29] A.Pal *et al.*, "A distributed channel selection scheme for multi-channelwireless sensor networks," in *MobiHoc*, 2012, pp. 263–264.

[30] A.Pal *et al.*, "DRCS: A distributed routing and channel selection scheme for multi-channel wireless sensor networks," in *IEEE PerSeNS*, 2013, pp. 602–608.

[31] S.Rangwala *et al.*, "Interference-aware fair rate control in wireless sensor networks," in *SIGCOMM*, 2006, pp. 63–74.

[32] C. T.Ee *et al.*, "Congestion control and fairness for many-to-one routing in sensor networks," in *SenSys*, 2004, pp. 148–161.

[33] L.Su *et al.*, "Towards optimal rate allocation for data aggregation in wireless sensor networks," in *MobiHoc*, 2011, p. 19.

[34] B.Zhang *et al.*, "Maximum utility rate allocation for energy harvesting wireless sensor networks," in *MSWiM*, 2011, pp. 7–16.