Exploiting Proxy Sensing For Efficient Monitoring of Large-Scale Sensor Networks

AMITANGSHU PAL*, Temple University

KRISHNA KANT, Temple University

Large sensor networks consisting of many heterogeneous sensors are being increasingly deployed for comprehensive real-time monitoring of cyber-physical systems. From the considerations of practical deployment and manageability, such a network cannot be treated like a single monolithic system. Yet, a strict isolation across logical or physical sensing clusters is sub-optimal since considerable energy savings can be achieved by exploiting the *proxy sensing phenomenon*. Proxy sensing refers to the relationships between the data captured by sensors that respond to related signals, are located in adjoining physical spaces, or monitor similar environments. In this paper we explore how various sensors can take advantage of such correlations to adapt their sensing rates collectively so as to optimize the energy efficiency and robustness of the sensing mission. We show that semi-distributed mechanisms can accomplish this adaptation efficiently and yet provide the advantages of autonomy, relative isolation, and distributed control that is essential in a large scale network.

Additional Key Words and Phrases: Heterogeneous sensing, adaptive sampling, rate adaptation, proxy sensing.

1 INTRODUCTION

The ongoing integration of information technology in physical systems for comprehensive monitoring and intelligent control will require increasingly large and complex sensor networks. Examples of such sensor networks include those for monitoring natural hazards (tornadoes, earthquakes, flash floods, etc.), road traffic, pollution, occupancy and people movement in buildings, water/sewer systems, power grid, etc. Sensing modules with one or more of different types of sensors may be deployed over a large area such as an entire smart city or an urban area. Such networks will likely consist of many different sub-networks perhaps deployed by different vendors and even controlled by different parties. Even in the case of a single large network, treating it like a single monolithic system with centralized control will make it unwieldy. On the other hand, a strict isolation across logical or physical sensing sub-clusters is undesirable since there are considerable savings to be had by exploiting the phenomenon of "proxy sensing", i.e., ability to deduce from a sensor some information about signals that it does not measure directly. Proxy sensing is a well known phenomenon that occurs in many contexts and for at least three reasons. The first

---

*This is the corresponding author

reason is the influence of (unmeasured) property on the measured property. For example, most contaminants in the municipal drinking water can be sensed through common measures such as pH value, oxidation reduction potential (ORP), and changes in chlorine level. Other examples include correlation between people density in a public place and noise level or correlation between average temperature inside a fresh food package and spoilage symptoms. The second reason is the spatial correlation among sensors. For example, two temperature sensors in nearby rooms will likely report almost the same temperature, or their temperature difference will stay almost constant, and less frequent measurements may be adequate. Similarly, traffic density or pollution in an area may allow us to deduce it in adjoining areas. The third reason is the situational similarity between two regions which also allows for more sparse measurements. For example, we don't need densely monitor quality deterioration in similar food carried under similar cooling conditions regardless of the physical location.

While individual sensors sense a single property, it is normal to put multiple heterogeneous sensors within a sensing and communications (S&C) module (in addition to a suitable wireless communications radio). Smartphones are ubiquitous S&C modules and contain an increasing number of sensors. The sensors intended for a specific domain increasingly contain multiple sensing modalities. For example, traditionally fresh food supply chain has depended on temperature measurements, but newer sensors may also sniff VoCs (volatile organic compounds) and correlate those to quality deterioration. Similarly, plant health sensors measure temperature, moisture, soil pH, etc. Since most S&C modules are power or size constrained, their energy efficient operation is crucial. Also, in many applications, the monitoring needs are focused around certain points-of-interests (PoIs). Typically, the PoIs continue to shift, but at a far slower rate than sensing. Exploitation of proxy sensing can be very helpful in conducting the PoI related data collection optimally. In particular, the sensing rates need to be adapted dynamically so as to optimize the energy efficiency and robustness of the sensing mission.

1.1 Our Contributions

In this paper 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 context, our key contributions in this paper are as follows:

- We analyze distributed sampling rate adaptation schemes to distribute the data capturing tasks among them based on their available energy, network participation, and correlations. We propose two decomposition based distributed solutions using sub-gradient method and Nesterov's gradient descent algorithm [1] to adapt the sampling rates of the individual sensors in a multi-hop wireless sensor network.

- Through extensive simulations we show that both sub-gradient and Nesterov's method require a large number of iterations for convergence of the sampling rates and hence message exchanges. In view of this, we propose two hybrid schemes and show via extensive simulations that they converge within ∼ 10-15 iterations. We also believe that the PoI based monitoring considered here is better suited for many real applications of large sensor networks than the independent sampling rate adaptation of the nodes [2].

The paper is an extension of our prior conference publication [3], with substantial enhancements in developing the multi-sensor proxy sensing framework in multi-hop wireless sensor network environment. We have also thoroughly evaluated the applicability of such framework in multiple real application scenarios, including (a) a substation monitoring environment where the devices are

equipped with vibration, gas and temperature sensors, (b) a disaster management scenario where the devices are equipped with cameras, micro-phones and accelerometer sensors, and (c) a pipeline monitoring scenario where the devices are equipped with chlorine, ORP and pH sensors.

1.2 Paper Organization

The rest of the paper is organized as follows. Section 2 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 3 then introduces the optimization model along with the distributed approaches that are adopted, along with their pros and cons. Section 4 describes two scalable alternate schemes to make rate adaptation faster. Extensive simulations are presented in section 5. Related proposals and relevant discussions are summarized in section 6. The paper is concluded in section 7.

2 COLLABORATIVE RATE ADAPTATION

2.1 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 [4]. The monitoring needs here involve pictures, sounds, movements relevant to damage assessment and rescue at certain PoI's where the situation may be unclear. For example an accelerometer, often coupled with velocity seismometers is used to measure and record the extent of ground motion or vibration. The audio samples can also be used to track the sound of building collapsing. Videos and images can be used to build a spatial view of the damage caused by the earthquake. In this example, all three sensors measure some aspect of the *same* phenomenon of ground movement. This spatial and cross-sensor correlation can be utilized to allow the sensors to be cycled on and off to adjust their energy thriftiness. 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: The number of potential contaminants in a Water Distribution Systems (WDSs) is fairly large [5], [6], thus deploying individual sensors corresponding to each and every contaminants is costly and onerous. A more practical scheme is to use sensors that measure *indicator* or *surrogate* parameters to detect abnormal water quality for possible contamination evaluation [5]. Free chlorine is the most sensitive indicator of contamination, that shows significant changes from the base line values at concentrations often one to two orders of magnitude below the lethal concentrations. Total organic carbon (TOC) in water is another important surrogate for detecting the presence of many organic compounds. Conductivity is also observed to respond slightly to some inorganic contaminants, and some metals. Oxidation reduction potential (ORP) generally behaves similar to chlorine residual, which can be used to corroborate an observed change in the chlorine residual. pH is important to understand the water's aqueous chemistry. Turbidity or water haziness is an erratic and unreliable primary indicator of contamination.

Such multi-sensor environments can be benefited by the proposed cooperative and collective sensing framework. 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.[7]. 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: A remote forest monitoring application includes their habitants, weather, sudden forest-fires, as well as detection of poachers destroying forest proprieties. Habitant monitoring can be explored by putting multiple camera sensors in the areas of interests, whereas weather conditions can be monitored by different temperature and humidity sensors. On the other hand, forest fires motoring is extremely crucial so that proactive actions can be taken before significant destruction. In the United States, there are typically between 60,000and 80,000wildfires that occur each year, burning 3 million to 10 million acres of land [8, 9]. Some wireless devices equipped with temperature and smoke detection sensors can be deployed in the areas that are prone to wildfires. Other than fire detection, unlawful killing of wild animals or wild plants are crucial for wildlife preserve and maintenance. Especially in Africa and Asia, poaching is becoming a very serious issue with the recent increase in the cost and desire for both ivory and the black rhinoceros horn [10]. The African black rhinoceros, are critically endangered because they have decreased by 80% in the last three rhino generations [11]. The poacher detection sensors include video, audio as well as some load sensors that are placed beneath the ground. Thermal imaging can also be an effective way to detect poachers especially at night. Such monitoring activities requires multiple types of sensors placed in different places, however cumulatively detecting and reporting required parameters to maximize the overall detection coverage.

Landslide Monitoring: Landslide are short-lived, destructive phenomenon, that are caused due to steep slope angle, toe cutting, and saturated soil [12]. In India, on landslide causes an annual damage of $400 million average. The key features of landslides include soil moisture, pore pressure, soil vibration and temperature. The devices need to be buried underground to take these samples and report them to a centralized station. Soil moisture sensors are needed to measure or permittivity of the soil. As rainfall increases, rain water accumulates in the pores of the soil, exerting a negative pressure which causes the loosening of soil strength, which can be measured by vibrating wire piezometer or strain gauge type piezometer [12]. The vibrations caused by the landslides are measured using geophones, whereas the soil temperature can be measured by the temperature sensors to detect a significant anomaly. 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. Thus the multi-sensor collaborative proxy sensing can be directly applied in this context as well.

Substation Monitoring: Power distribution substations have a number of critical components such as circuit breakers and transformers that must be continuously monitored to reduce the possibility of expensive and disruptive power outages. This is becoming an increasing concern due to the aging nature of substation equipment and infrastructure [13]. Substations and distribution centers use *circuit breakers* to switch electric circuits or equipment either into or out of the broader system. These devices are typically filled with either oil or SF_6 gas. In oil-filled circuit breakers, the oil provides cooling and prevents arcing when the switch is activated. Thus a prognostic method that detects such failures by searching for relative changes in oil temperature to generate an alarm. At the same time the circuit breakers that are filled with SF_6 gas, can be checked for leakage by deploying the gas sensors on the surface. Vibration and acoustic sensors can also be placed to monitor leakage in such scenarios. The transformers in the substations also need such monitoring of surface temperatures and vibrations. This again yields an environment where both the battery life and sensor correlations are important.
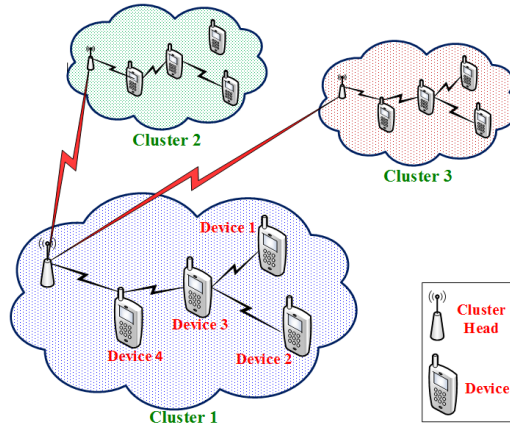
Fig. 1. Illustration of our network model.

2.2 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 entire network consists of several disjoint clusters, each having its own cluster-head (CH). This division may be natural – dictated by different physical/logical clusters, possibly managed by different entities. The division could also be artificial, as in a large network broken up into multiple clusters for the purposes of adaptation discussed here. In any case, the nodes in each cluster forward their traffic to their cluster-head. We assume, for simplicity, that the cluster-heads can directly communicate with each others and are not energy constrained. The overall network model is depicted in Fig. 1. Such network model is applicable in many WSN application scenarios as described in section 2.1. For example, a disaster management scenario may deploy some access points (sinks) in a large geographic area, each one of them forms 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.

The PoIs will generally evolve over time, but so long as this is a very slow process it does not affect the issues discussed here. 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) [14] schemes like X-Mac principle [15] 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 a 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. Furthermore, we 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's direct communication, 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.

2.3 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 $\mathbf{A}^{\mathbf{T}}.\mathbf{A}$ as follows:

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

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

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

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

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

$$\nabla^2 f(x) \geq \sigma.I \quad \forall x \in \text{dom}(f) \tag{5}$$

where I is the identity matrix.

Table 1. 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
r_i^t	$\triangleq$	Sampling rate of sensor t of node i
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
v_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

3 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.

3.1 Basic Problem Formulation

We now formulate the basic 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 1. 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 $w_i^{tk} \leq 1$. Basically w_i^{tk} brings the notion of proxy-sensing in the model, i.e. more than one sensor can sense the POIs with some weight.

Considering the above two factors, the effective rate with which the k-th POI is monitored by the t type sensors is given by $\xi^{tk} = \sum_{i=1}^{S} w_i^{tk}.r_i^t$. Thus the fair event reporting ability is ensured by modeling the utility of reporting the k-th POI by the t sensors as $U^k \left(\sum_{t=1}^{T} \xi^{tk} \right) = \log \left(\sum_{t=1}^{T} \xi^{tk} \right)$.

Our objective is to maximize the overall event monitoring capability, i.e. $\sum_{k=1}^{\mathcal{E}} U^k \left(\sum_{t=1}^{T} \xi^{tk} \right)$, after satisfying the energy budget of the individual nodes. Thus the overall optimization problem can be written as

Original Problem (OP):

$$\text{Maximize} \quad \sum_{k=1}^{\mathcal{E}} \log \left(\sum_{i=1}^{S} \sum_{t=1}^{T} w_i^{tk}.r_i^t \right)$$

$$\text{subject to} \quad \sum_t r_i^t \left(e^t + c^t \right) + \sum_t \sum_{j \in \mathbb{D}_i} \alpha_{ji}.r_j^t.c^t \leq E_i \quad \forall i \tag{6}$$

$$R_m \leq r_i^t \leq R_M \quad \forall i, \forall j, \forall t$$

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 in our scheme each node 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 easily using standard methods. However, such a solution amounts to collecting all of information involved in the above formulation in one central node for performing the optimization. This may be difficult in a large network for several reasons including difficulty in obtaining consistent global state at a central point, information visibility/transfer restriction across subnetworks owned or operated by different parties, longer path delays due to wireless channel effects (e.g., fading, interference), and higher likelihood of information unavailability to the central node due to failures, drained battery, shadowing, and attacks at intermediate nodes.

Thus our objective is to adopt an iterative, *distributed* approach to solve this problem. Notice that all of the issues mentioned here can arise with the distributed approach as well, but will become less severe as the amount of required interaction between nodes and the lengths of paths over which the information must travel decreases. However, in all cases – from fully distributed to fully centralized – the information exchange algorithms must cope with lack of availability of required information. This involves the following: (a) a timeout mechanism so as to limit the amount of time for which algorithm must wait to receive information, (b) approximation of the missing information via a time series prediction technique based on prior values, and (c) handling of stale information

(e.g., messages that arrive following the timeout). Since the techniques to do this are well known and well practiced, we do not dwell on those further.

3.2 Distributed Problem Formulation

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 $\sum_{i=1}^{S} \sum_{t=1}^{T} w_i^{tk}.r_i^t$. Second, the function is non-separable with respect to each node i. The first problem can be resolved by adding an *augmented* variable [16], [17] 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 [18] and is described as follows.

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

$$\log\left(\sum_{i=1}^{S}\sum_{t=1}^{T} w_i^{tk}.r_i^t\right) \geq \sum_{i=1}^{S}\sum_{t=1}^{T} \theta_i^{tk}\log\left(\frac{w_i^{tk}.r_i^t}{\theta_i^{tk}}\right) \quad \forall i, \forall t, \forall k$$

$$\text{where} \quad \theta_i^{tk} = \frac{w_i^{tk}.r_i^t}{\sum_{i=1}^{S}\sum_{t=1}^{T} w_i^{tk}.r_i^t} \quad \forall i, \forall t, \forall k \tag{7}$$

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

Modified Optimization Problem (MOP):

$$\text{Maximize} \quad U = \sum_{k=1}^{\mathcal{E}}\sum_{i=1}^{S}\sum_{t=1}^{T} \theta_i^{tk}\log\left(\frac{w_i^{tk}.r_i^t}{\theta_i^{tk}}\right)$$

$$\text{subject to} \quad \sum_{t} r_i^t\left(e^t + c^t\right) + \sum_{t}\sum_{j\in\mathbb{D}_i} \alpha_{ji}.r_j^t.c^t \leq E_i \quad \forall i \tag{8}$$

$$R_m \leq r_i^t \leq R_M \quad \forall i, \forall j, \forall t$$

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 [1]. We will show that neither scheme works well when applied in a fully distributed manner; however, they set the stage for the partially distributed scheme discussed in section 4.

3.3 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* [16, 19, 20]. We first assume that

$$\Lambda_i^t = \sum_{k=1}^{\mathcal{E}} \theta_i^{tk}\log\left(\frac{w_i^{tk}.r_i^t}{\theta_i^{tk}}\right) \quad \text{and} \quad \Theta_i^t = r_i^t\left(e^t + c^t\right) \tag{9}$$

Then the Lagrangian can be defined as:

$$\mathbb{L}(r, \nu) = \sum_{i=1}^{S}\sum_{t=1}^{T} \Lambda_i^t - \sum_{i=1}^{S} \nu_i\left(\sum_{t} \Theta_i^t + \sum_{t}\sum_{j\in\mathbb{D}_i} \alpha_{ji}.r_j^t.c^t - E_i\right) \tag{10}$$

The objective function of the dual problem can be written as:

$$D(v) = \text{Max } \mathbb{L}(r, v) = \text{Max} \sum_{i=1}^{S} \sum_{t=1}^{\mathcal{T}} \Lambda_i^t - \sum_{i=1}^{S} v_i \left(\sum_t \Theta_i^t + \sum_t \sum_{j \in \mathbb{D}_i} \alpha_{ji}.r_j^t.c^t - E_i \right)$$

$$= \text{Max} \sum_{i=1}^{S} \sum_{t=1}^{\mathcal{T}} \Lambda_i^t - \sum_{i=1}^{S} \sum_{t=1}^{\mathcal{T}} v_i.\Theta_i^t - \sum_{i=1}^{S} \sum_{t=1}^{\mathcal{T}} \sum_{j \in \mathbb{D}_i} v_i.\alpha_{ji}.r_j^t.c^t + \sum_{i=1}^{S} v_i.E_i$$

$$= \text{Max} \sum_{i=1}^{S} \sum_{t=1}^{\mathcal{T}} \Lambda_i^t - \sum_{i=1}^{S} \sum_{t=1}^{\mathcal{T}} v_i.\Theta_i^t - \sum_{i=1}^{S} \sum_{t=1}^{\mathcal{T}} \sum_{j \in \mathbb{A}_i} v_j.\alpha_{ij}.r_i^t.c^t + \sum_{i=1}^{S} v_i.E_i$$

$$= \text{Max} \sum_{i=1}^{S} \sum_{t=1}^{\mathcal{T}} \left(\Lambda_i^t - v_i.\Theta_i^t - \sum_{j \in \mathbb{A}_i} v_j.\alpha_{ij}.r_i^t.c^t \right) + \sum_{i=1}^{S} v_i.E_i$$

The dual problem is to $\min_{v_i \geq 0} D(v)$. v is the Lagrange multiplier vector, which can be iteratively updated using

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

$$= \left[v_i^\ell + \gamma \left(\sum_t r_i^t.e^t + \left(\sum_t r_i^t + \sum_t \sum_{j \in \mathbb{D}_i} \alpha_{ji}.r_j^t \right) c^t - E_i \right) \right]^+ \tag{11}$$

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.

$$\text{Max} \sum_{k=1}^{\mathcal{E}} \theta_i^{tk} \log \left(\frac{w_i^{tk}.r_i^t}{\theta_i^{tk}} \right) - v_i.r_i^t \left(e^t + c^t \right) - \sum_{j \in \mathbb{A}_i} v_j.\alpha_{ij}.r_i^t.c^t$$

$$\Rightarrow r_i^t = \left[\frac{\sum_{k=1}^{\mathcal{E}} \theta_i^{tk}}{v_i(e^t + c^t) + \sum_{j \in \mathbb{A}_i} v_j.\alpha_{ij}.c^t} \right]_{R_m}^{R_M} \tag{12}$$

THEOREM 3.1. *The proposed distributed version of **MOP** converges to the optimal solution of the original problem **OP**.*

PROOF. Let us define (r^*, v^*, θ^*) are the optimal solution of **MOP**. It can be shown that (r^*, v^*) also satisfies the KKT condition of the original problem **OP**. The KKT condition of the **MOP** is given by

$$\frac{\partial}{\partial r_i^{tk}} \left(\sum_{k=1}^{\mathcal{E}} \sum_{i=1}^{S} \sum_{t=1}^{\mathcal{T}} \theta_i^{tk} \log \left(\frac{w_i^{tk}.r_i^t}{\theta_i^{tk}} \right) \right) \Bigg|_{r^*} - \sum_{i=1}^{S} v_i^* \left(\sum_t \left(e^t + c^t \right) + \sum_t \sum_{j \in \mathbb{A}_i} \alpha_{ij}.c^t \right) = 0$$

$$v_i^*.r_i^{t*} \left(e^t + c^t \right) - \sum_{i=1}^{S} \sum_{t=1}^{\mathcal{T}} \sum_{j \in \mathbb{D}_i} v_i^*.\alpha_{ji}.r_j^t.c^t + \sum_{i=1}^{S} v_i^*.E_i = 0 \tag{13}$$

$$v_i^* \geq 0$$

Notice that the last two KKT conditions of **OP** and **MOP** are identical. Now as

$$\frac{\partial}{\partial r_i^{tk}} \left(\sum_{k=1}^{\mathcal{E}} \sum_{i=1}^{S} \sum_{t=1}^{\mathcal{T}} \theta_i^{tk} \log \left(\frac{w_i^{tk}.r_i^t}{\theta_i^{tk}} \right) \right) \Bigg|_{r^*} = \frac{\partial}{\partial r_i^{tk}} \left(\sum_{k=1}^{\mathcal{E}} \log \left(\sum_{i=1}^{S} \sum_{t=1}^{\mathcal{T}} w_i^{tk}.r_i^t \right) \right) \Bigg|_{r^*} \tag{14}$$

the first condition of **OP** and **MOP** are also identical at point (r^*, v^*). 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 (11) and equation (12) 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

$$
T_i = \overbrace{\sum_t r_i^t}^{\text{Self traffic}} + \overbrace{\sum_t \sum_{j \in \mathbb{D}_i} \alpha_{ji}.r_j^t}^{\text{Traffic from descendants}} = \sum_t r_i^t + \sum_{i \in \mathbb{P}_j} \alpha_{ji}.T_j \quad \text{(for non-leaf nodes)}
$$
$$
= \sum_t r_i^t \quad \text{(for leaf nodes)}
$$
(15)

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 (11). For calculating the individual sampling rates of the nodes from equation (12), we assume $F_i = \sum_{j \in \mathbb{A}_i} v_j.\alpha_{ij}$. We refer to v_i as node i's *local price*, whereas F_i as its *aggregate price*. Each node distributes the local and aggregate prices to its children, which then use these to calculate their own aggregate prices. For example, node i calculates its aggregate price F_i by using the (v_j, F_j)s of its PPs as follows:

$$
F_i = \sum_{j \in \mathbb{A}_i} v_j.\alpha_{ij} = \sum_{j \in \mathbb{P}_i} \alpha_{ij} \left(\overbrace{v_j}^{\text{Local price of } j} + \overbrace{F_j}^{\text{Aggregate price of } j} \right) \quad \text{(for non-CH nodes)}
$$
$$
= 0 \quad \text{(for CH nodes)}
$$
(16)

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 v_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

$$
\alpha_{ij} = \frac{1/(\text{P-ETX}_{ij})}{\sum_{x \in \mathbb{P}_i} 1/(\text{P-ETX}_{ix})}
$$
(17)

This is based on the intuition that the nodes choose their routes in proportion to their route qualities. All the nodes then calculate their (v_i, F_i, T_i, r_i^t) based on their local information as follows. The CHs broadcast their aggregate price F_i to be zero to their children. The immediate children of the CHs calculate their aggregate price F_i as well as their sampling rates using equation (16) and (12) and broadcast. In the first round, the θ_i^{tk} and v_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 v_i using equation (11). The subsequent PP nodes update their traffic using equation (15) as well as their v_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. The conceptual flow diagram of SDS is illustrated in Fig. 2.

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
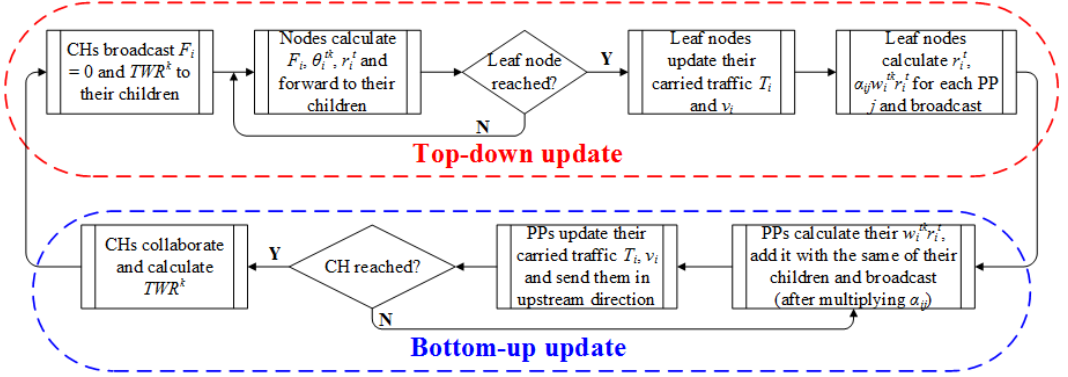
Fig. 2. Conceptual flow diagram of SDS.

varying channel conditions. We will explore the effects of these practical deployment issues on the convergence of this iterative scheme in future.

3.4 Nesterov's Gradient Descent based Distributed Scheme (NDS)

We propose another scheme using Nesterov's method [1] 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 3.5.

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 budget constraint of problem (6) is written as $\mathbf{A}.\mathbf{R} \leq \mathbf{E}$. For example, let us consider a hypothetical scenario of Fig. 3 with three devices, each equipped with three sensors. In this figure devices 1 and

Fig. 3. Illustration of NDS.

2 sense and transmit their packets to device 3, and thus the energy budget constraint of problem (6) can be written as:
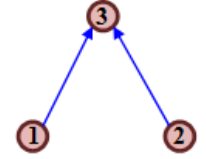
$$
\overbrace{\begin{bmatrix} e^1 + c^1 & e^2 + c^2 & e^3 + c^3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & e^1 + c^1 & e^2 + c^2 & e^3 + c^3 & 0 & 0 & 0 \\ \alpha_{13}.r_1^1.c^1 & \alpha_{13}.r_1^2.c^2 & \alpha_{13}.r_1^3.c^3 & \alpha_{23}.r_2^1.c^1 & \alpha_{23}.r_2^2.c^2 & \alpha_{23}.r_2^3.c^3 & e^1 + c^1 & e^2 + c^2 & e^3 + c^3 \end{bmatrix}}^{\mathbf{A}}
\overbrace{\begin{bmatrix} r_1^1 \\ r_1^2 \\ r_1^3 \\ r_2^1 \\ r_2^2 \\ r_2^3 \\ r_3^1 \\ r_3^2 \\ r_3^3 \end{bmatrix}}^{\mathbf{R}}
\leq
\overbrace{\begin{bmatrix} E_1 \\ E_2 \\ E_3 \end{bmatrix}}^{\mathbf{E}}
$$

$$(18)$$

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 $\Omega = \max\{c^t\} \, \forall \, t$, $\Gamma = \sum_t \left(e^t + c^t\right)$, and $\Delta = \sum_t c^t$. With these we propose the following theorems.

THEOREM 3.2. *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 [21] for the proof. □

THEOREM 3.3. *The dual of problem (6) is Lipschitz continuous with constants* $L_1 = \frac{S.\Upsilon^2.(\Gamma+(S-1).\Delta).(\Upsilon+(S-1).\Omega)}{\mathcal{E}.W_m}$, *and* $L_2 = \frac{S.\mathcal{T}.R_M^2.(\Gamma+(S-1).\Delta).(\Upsilon+(S-1).\Omega)}{\mathcal{E}.W_m}$, *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:

$$\sigma = \min\left(-\frac{\partial^2 U}{\partial r_i^{t\,2}}\right) = \min\left(\frac{\sum_k \theta_i^{tk}}{r_i^{t\,2}}\right) = \min\left(\frac{\sum_k w_i^{tk}.r_i^t}{r_i^{t\,2}.\sum_{i=1}^{S}\sum_{t=1}^{\mathcal{T}} w_i^{tk}.r_i^t}\right)$$
$$= \min\left(\frac{\sum_k w_i^{tk}}{r_i^t.\sum_{i=1}^{S}\sum_{t=1}^{\mathcal{T}} w_i^{tk}.r_i^t}\right) \geq \min\left(\frac{\sum_k w_i^{tk}}{r_i^t.\sum_{i=1}^{S}\sum_{t=1}^{\mathcal{T}} r_i^t}\right) \geq \frac{\mathcal{E}.W_m}{S.\Upsilon^2} \tag{19}$$

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

$$\sigma \geq \min\left(\frac{\sum_k w_i^{tk}}{r_i^t.\sum_{i=1}^{S}\sum_{t=1}^{\mathcal{T}} r_i^t}\right) \geq \frac{\mathcal{E}.W_m}{S.\mathcal{T}.R_M^2} \tag{20}$$

We next calculate the spectral radius of $\mathbf{A^T.A}$ as follows. While deriving $\rho(\mathbf{A^T.A})$, 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 $\rho(\mathbf{A^T.A})$ and σ, by using Theorem 3.2.

$$\|\mathbf{A}\|_\infty \leq \sum_t \left(e^t + c^t\right) + (S-1)\sum_t c^t = \Gamma + (S-1).\Delta$$
$$\|\mathbf{A}\|_1 \leq \max_t\{e^t + c^t\} + (S-1)\max_t\{c^t\} = \Psi + (S-1).\Omega$$
$$\therefore \rho(\mathbf{A^T A}) \leq \|\mathbf{A}\|_1.\|\mathbf{A}\|_\infty \leq (\Gamma + (S-1).\Delta).(\Psi + (S-1).\Omega)$$
$$L_1 = \frac{S.\Upsilon^2.(\Gamma + (S-1).\Delta).(\Psi + (S-1).\Omega)}{\mathcal{E}.W_m} \text{ from equation(19)}$$
$$L_2 = \frac{S.\mathcal{T}.R_M^2.(\Gamma + (S-1).\Delta).(\Psi + (S-1).\Omega)}{\mathcal{E}.W_m} \text{ from equation(20)} \tag{21}$$

□

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 [1, 22]. 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 (11)) 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 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 $v_i^{\ell+1}$ 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^t $\forall i \in \{1, 2, \ldots, N\}$.
3: $\text{tmp}_i = 0$;
4: **for** each iteration $\ell = \{1, 2, \ldots, \mathcal{L}\}$ **do**
5:      Update rate r_i^t and θ_i^{tk};
6:      $u_i^{\ell+1} = \left[v^\ell + \frac{\Delta f_i}{L} \right]^+$;
7:      $\text{tmp}_i = \text{tmp}_i + \frac{\ell+1}{2} . \Delta f_i$;
8:      $v_i^{\ell+1} = \left[\frac{\text{tmp}_i}{L} \right]^+$;
9:      $v_i^{\ell+1} = \frac{\ell+1}{\ell+3} . u_i^{\ell+1} + \frac{2}{\ell+3} . v_i^{\ell+1}$;
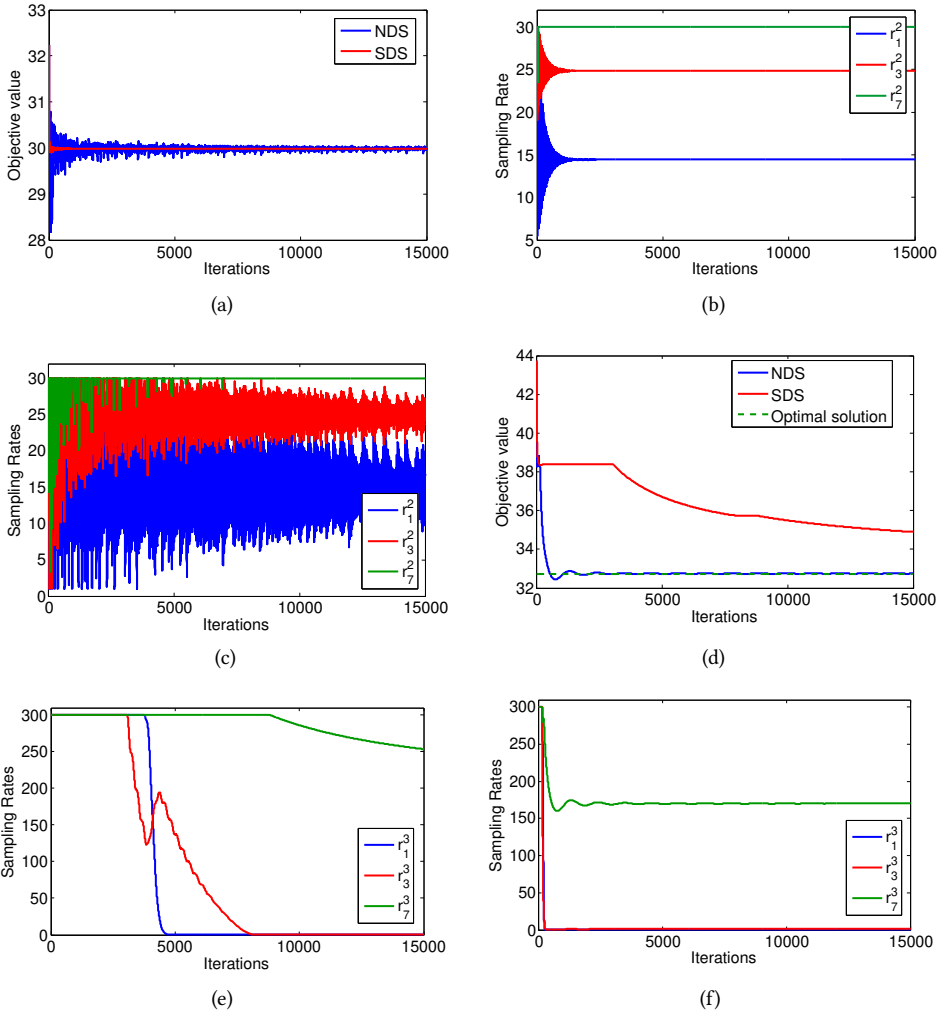10: **end for**

Fig. 4. 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).

3.5   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 [13]. 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 ∼140 ms. The nodes consume ∼20 mA [23] at the transmit mode, which is the representative of Crossbow's MICAz wireless sensor nodes, which are equipped with the Atmel ATmega128L processor running at 8MHz, 2.4 GHz Chipcon CC2420 radio, 128KB program memory, 512KB measurement flash, and 4KB EEPROM [23]. 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 [13].

We compare the proposed distributed rate adaption scheme of **MOP** along with the solution of **OP** obtained from AMPL solver [24]. The result is shown in Fig. 4 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. 4(a)-(c)) R_M is assumed to be 30/hr. whereas it is assumed to be 300/hr. in the second case (Fig. 4(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(12) 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. 4(d)-(f). In both cases we observe that the distributed version of **MOP** closely matches (Fig. 4(a) does not show the optimal solution for clarity) with the original problem of (6), which validates the claim of Theorem 1.

Fig. 4(b)-(c) and Fig. 4(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 higher due to large number of iterations (in the order of few hundreds to thousands) which limits their usefulness in large-scale WSN applications.

3.6   Effect of the number of hops

There is an evident tradeoff between the number of hops from the nodes to the CHs and the convergence time. When the number of hops from the nodes to the CHs increases, the convergence time increases as well. Fig. 5 shows the effects of different hop-counts on the convergence time. We have changed the binary tree topology according to the maximum tree-heights, while keeping the number of nodes identical. From Fig. 5 we can observe that the convergence time reduces as the tree height decreases from 3 to 1. From Fig. 5(c) shows the sampling rate of node 3, which also verifies faster convergence with smaller tree height. However, to reduce the maximum tree height or reducing the number of hop-counts, higher number of CHs need to be deployed in the area of interest, which increases the deployment and maintenance cost. Thus the tradeoff is between the deployment cost and the speed of convergence.
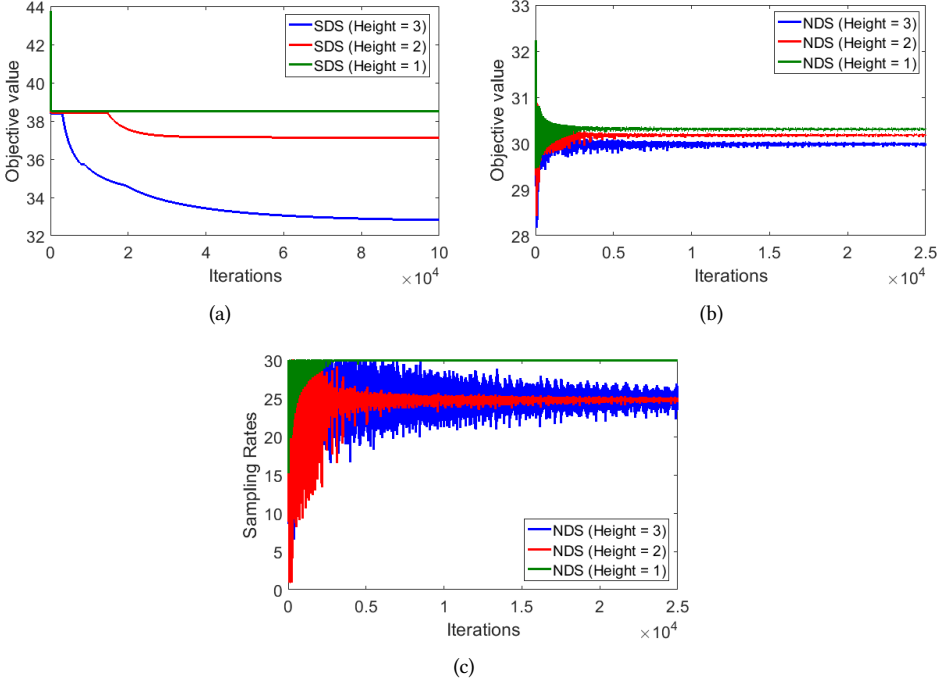
(a)

(b)

(c)

Fig. 5. Effect of multi-hopping. (a) Convergence of SDS with (R_m, R_M) = (1, 300). (b) Convergence of NDS with (R_m, R_M) = (1, 30) and the (c) sampling rate of node 3.

4 SCALABLE SEMI-DISTRIBUTED 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 the following reasons:

- For implementing these fully distributed schemes, too many control messages need to be exchanged in between the nodes and the CHs, which will drain their batteries.
- On the other hand the convergence time of these schemes 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.

4.1 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

---

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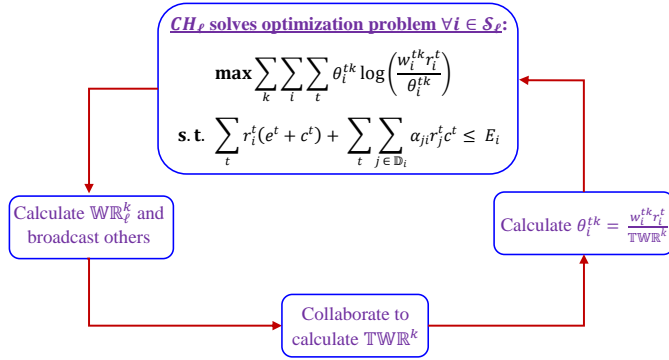10: **end while**

Fig. 6. Illustration of the SDRA scheme.

is shown in Algorithm 2. We assume that there are C cluster-heads, where CH_ℓ is associated with a set ζ_ℓ of $\mathcal{S}_\ell$ nodes. Thus the CHs solve 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 $i \in \zeta_\ell$) using any random θ_i^{tk} (at the first iteration), which can be done using any existing solvers like AMPL [24], GLPK [25], CVX [26] etc. It then calculates its local weighted rate $\mathbb{WR}_\ell^k = \sum_{i \in \zeta_\ell} \sum_t w_i^{tk} . r_i^t$ $\forall k$ 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. The overall scheme is shown in Fig. 6.

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.

4.2 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 schedules. 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 becomes

ADRA Problem:

$$
\begin{aligned}
\text{Maximize} \quad & \sum_{k=1}^{\mathcal{E}} \log \left(\sum_{i=1}^{S} \sum_{t=1}^{\mathcal{T}} w_i^{tk}.r_i^t \right) \\
\text{subject to} \quad & \sum_t r_i^t \left(e^t + c^t \right) \le E_i \quad \forall i \\
& R_m \le r_i^t \le R_M \quad \forall i, \forall j, \forall t
\end{aligned}
\tag{22}
$$

By using the Jensen's inequality similar to equation (7) we obtain:

Modified ADRA Problem (MADRA):

$$
\begin{aligned}
\text{Maximize} \quad & U = \sum_{k=1}^{\mathcal{E}} \sum_{i=1}^{S} \sum_{t=1}^{\mathcal{T}} \theta_i^{tk} \log \left(\frac{w_i^{tk}.r_i^t}{\theta_i^{tk}} \right) \\
\text{subject to} \quad & \sum_t r_i^t \left(e^t + c^t \right) \le E_i \quad \forall i \\
& R_m \le r_i^t \le R_M \quad \forall i, \forall j, \forall t
\end{aligned}
\tag{23}
$$

By using equation (7) we can prove that solving **MADRA** in equation (22) is equivalent to solving **ADRA** in equation (23). The **MADRA** is strictly concave, for a given θ_i^{tk} and thus can be solved using Algorithm 3. In this scheme, the nodes first assign the sampling rates to each sensor i as $r_i^t = \frac{E_i \cdot \sum_k \theta_i^{tk}}{\sum_i \sum_k \theta_i^{tk}(e^i + c^i)}$ (line 3). If the sampling rates are less or more than the specified thresholds R_m and R_M, then the node divides the Δ fairly among other sensors (line 5-24). The nodes first initialize an empty set V (line 4). If the sampling rate of a sensor is less R_m, it changes its sampling rate to R_m, include that sensor into V and divides the Δ fairly among the sensors that are not in V. This process is repeated for all the sensors (lines 5-14). After that the same procedure is applied when the sampling rates are more then R_M for any sensor (lines 15-24).

THEOREM 4.1. *For a given θ_i^{tk}, Algorithm 3 gives optimal rate allocation of the node sensors.*

PROOF. Line 3 can be derived by solving the Lagrangian and KKT conditions of problem equation (8) (ignoring the last set of constraints) which are as follows

$$
\mathbb{L} = \sum_{k=1}^{\mathcal{E}} \sum_{i=1}^{S} \sum_{t=1}^{\mathcal{T}} \theta_i^{tk} \log \left(\frac{p_i^{tk}.r_i^t}{\theta_i^{tk}} \right) - \sum_{i=1}^{S} \lambda_i \left(\sum_t r_i^t \left(e^t + c^t \right) - E_i \right)
\tag{24}
$$

$$
\frac{\partial \mathbb{L}}{\partial r_i^t} = \frac{\sum_{k=1}^{\mathcal{E}} \theta_i^{tk}}{r_i^t} - \lambda_i (e^t + c^t) = 0
\tag{25}
$$

$$
\lambda_i \left(\sum_t r_i^t \left(e^t + c^t \right) - E_i \right) = 0
\tag{26}
$$

Equation (25) gives $r_i^t = \frac{\sum_{k=1}^{\mathcal{E}} \theta_i^{tk}}{\lambda_i (e^t + c^t)}$ and $\lambda_i \ne 0$. Putting this in equation (26), we get $\lambda_i = \frac{\sum_{t=1}^{\mathcal{T}} \sum_{k=1}^{\mathcal{E}} \theta_i^{tk}}{E_i}$, which makes $r_i^t = \frac{E_i \cdot \sum_k \theta_i^{tk}}{\sum_{t=1}^{\mathcal{T}} \sum_{k=1}^{\mathcal{E}} \theta_i^{tk}(e^t + c^t)}$. $\qquad \square$

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

---

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 $\forall i$.
3: $r_i^t = \frac{E_i \cdot \sum_k \theta_i^{tk}}{\sum_t \sum_k \theta_i^{tk}(e^i+c^i)} \forall t$;
4: $V = \{\phi\}$;
5: **for** each sensor $t = \{1, 2, \ldots, \mathcal{T}\}$ **do**
6:     **if** $r_i^t < R_m$ **then**
7:         Assign $r_i^t = R_m$;
8:         $V = V \cup t$;
9:         $\Delta = E_i - \sum_i r_i^t(e^i + c^i)$;
10:         **for** each sensor $m \notin V$ **do**
11:             $r_i^m = r_i^m + \frac{\sum_k \theta_j^{mk}}{\sum_{m \notin V} \sum_k \theta_i^{mk}} \cdot \frac{\Delta}{(e^m+c^m)}$;
12:         **end for**
13:     **end if**
14: **end for**
15: **for** each sensor $t = \{1, 2, \ldots, \mathcal{T}\}$ **do**
16:     **if** $r_i^t > R_M$ **then**
17:         Assign $r_i^t = R_M$;
18:         $V = V \cup t$;
19:         $\Delta = E_i - \sum_i r_i^t(e^t + c^t)$;
20:         **for** each sensor $m \notin V$ **do**
21:             $r_i^m = r_i^m + \frac{\sum_k \theta_j^{mk}}{\sum_{m \notin V} \sum_k \theta_i^{mk}} \cdot \frac{\Delta}{(e^m+c^m)}$;
22:         **end for**
23:     **end if**
24: **end for**
25: return r_i^t $\forall t$

---

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.
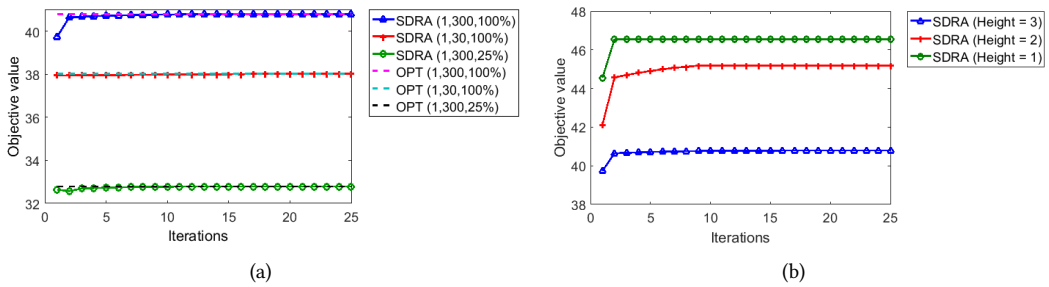

Fig. 7. (a) Convergence of SDRA with different R_m, R_M and battery charges. The numbers within the braces are R_m, R_M and battery charge levels respectively. OPT denotes the "optimal solution". (b) The effects of tree heights on convergence time.

4.3 Validation of SDRA and ADRA

To validate the convergence of SDRA and ADRA, we assume a scenario of 5 clusters, each having 14 nodes placed in a binary tree fashion of height 3. Fig. 7 and Fig. 8 show the convergence of these two schemes with different remaining battery capacities and R_M. From these figures we can observe
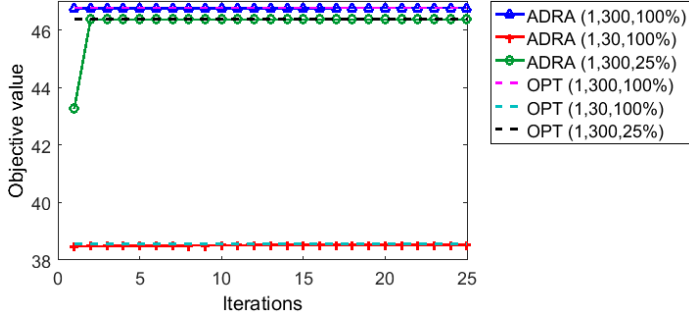
Fig. 8. Convergence of ADRA with different R_m, R_M and battery charges. The numbers within the braces are R_m, R_M and battery charge levels respectively. OPT denotes the "optimal solution".

that in both schemes, the objective values match with the optimal solution obtained from AMPL. We can also observe that the schemes converge pretty fast (within ~10-15 iterations), which make them suitable as a scalable solution in energy-constrained WSNs. Also notice that the objective values increase with the remaining battery capacity and R_M, which happens due to the increase in sampling rate of the individual sensors.

Fig. 7(b) shows the effect of the tree heights on SDRA convergence, when R_M is assumed to be 300, and the battery capacity is assumed to be of 100%. This figure shows that the effect of multi-hopping on convergence time is greatly reduced in the case of SRDA, which fulfills the primary requirement of using this scheme. We can also observe that the objective value increases with smaller tree heights, because of the smaller number of hops from the nodes to the CHs. In case of ADRA the effects of multi-hopping is ignored, and thus the scheme is not dependent on the tree heights.

5 PERFORMANCE EVALUATION OF COLLABORATIVE SENSING

We evaluate the adaptation schemes in three different application scenarios. First we evaluate the schemes on a substation monitoring scenario where sensor nodes are deployed on different electrical equipment. We next evaluate the scheme in a disaster monitoring scenario where multiple smartphones are deployed to monitor certain PoIs for situation monitoring. Finally we evaluate the adaption effects on a pipeline monitoring scenario where he monitoring devices are equipped with multiple, heterogeneous sensors for sensing different water contaminants. For all these experiments, we assume scheduled transmission, and thus the preamble length is neglected. The purpose of building such a setup is to demonstrate the effect of collaborative sensing in presence of multi-sensor equipped devices. Different sensors used in these scenarios along with their current/power consumption is shown in Table 2.

5.1 Effects of adaptation in substation monitoring

5.1.1 Experimental Setting. We first study the effectiveness of the proposed sampling rate adaptation schemes in a substation monitoring scenario, where each device is equipped with a vibration sensor, a SF_6 density sensor and a temperature sensor. These sensors consume 9.5, 150, 7.5 mA respectively with a sampling time of 7000, 400 and 112 milliseconds, which lead to the power consumption as reported in Table 2. We assume MicaZ sensor nodes which consume ~20 mA at the transmit mode [23]. These nodes use CC2420 radios, which are packet level radios with a maximum packet length of 127 bytes along with a data rate of 250 kbps [30], which results in a transmission time of ~4 milliseconds. The devices are uniformly deployed in a geographic area of 200×200 sq. m. The fresh battery capacity of the nodes are assumed to be 5000 mAHr [31, 32]. The remaining

Table 2. Different Sensor Types, and Their Current/Power Consumption, Data Obtained From [13, 27–29]

Application	Sensor Type	Current/Power Consumption
Substation Monitoring	Vibration/sound	9.5 mA
	SF_6 gas sensor	150 mA
	Ambient temperature	7.5 mA
Disaster Monitoring	Video camera	1258 mW
	Microphone	329 mW
	Accelerometer	96 mW
Pipeline Monitoring	Chlorine sensor	4 mA
	ORP	20.2 mA
	pH	25.5 mA

battery capacities of the nodes are uniformly randomly chosen among (25-100%). Depending on their remaining battery capacities, these devices are divided into 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 nodes are expected to remain active for 12 months.

5.1.2 Experimental Results. Fig. 9(a) shows the mean energy budget of the nodes at different tiers, whereas Fig. 9(b) shows their corresponding usages, which is defined as the cumulative energy expenditure of the devices due to sampling and forwarding of the packets. From Fig. 9(b) we can observe that the energy usages are adapted proportional to the individual node's power budgets. Fig. 9(c)-(e) show the mean sampling rates of sound sensor, SF_6 gas sensor and temperature sensors at various tiers. We can observe that the sampling rate of the first tier nodes are much higher compared to the other tiers, due to their higher battery charges. We can also observe that the average number of samples captured by 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. 9(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 proposed scheme is adaptive to 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 if the cumulative network traffic grows with the number of nodes, the average number of data forwarding by the individual nodes are similar.

5.2 Effects of adaptation in disaster monitoring

5.2.1 Experimental Setting. We now study the usage of the sampling rate adaptation schemes in a disaster monitoring scenario where few smartphones collectively monitor the situation using the sensors attached to them. The simulated system topology consists of 75 devices including 5 access points (APs), placed in an area of 100×1000 sq. m. Each AP is associated with a cluster of 14 smartphones, that are placed in a binary tree fashion with a height of 3. A fully charged phone is assumed to have a battery capacity of 11.78 Whr, which is typically the capacities for current smartphones. The phones are expected to remain active for 24 hours, and the power budgets for
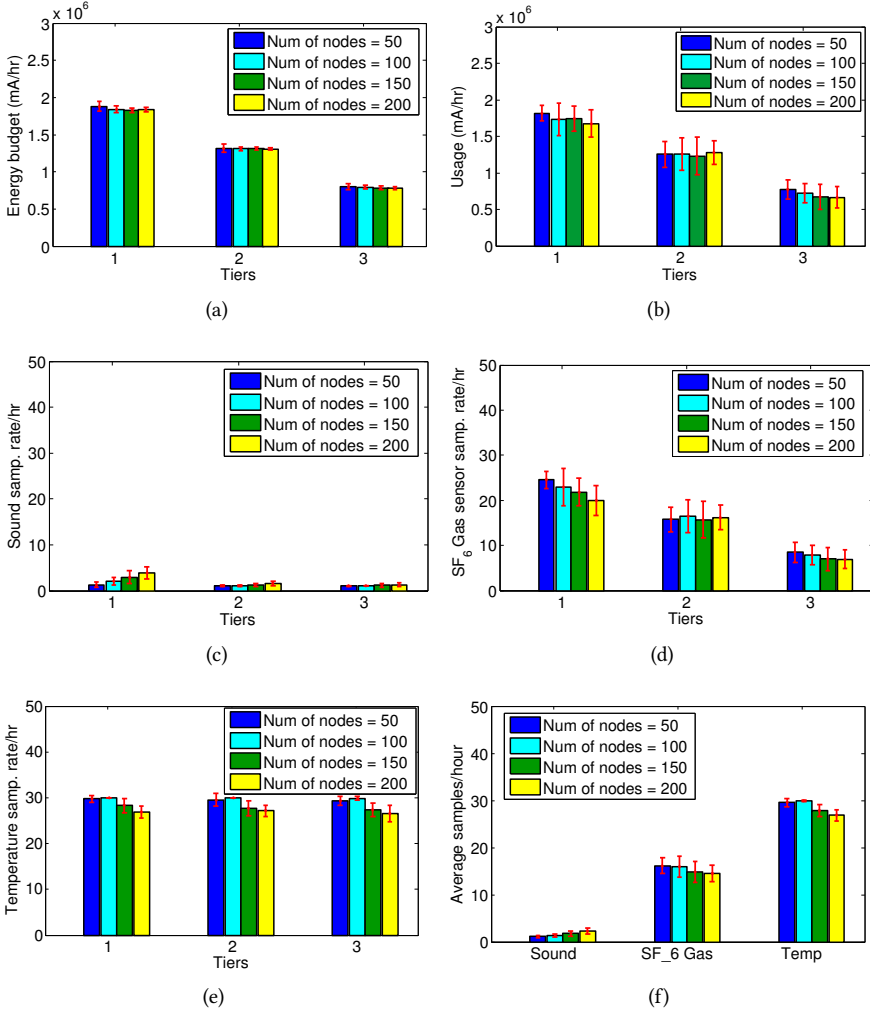
Fig. 9. (a) Average energy budget and (b) usage profiles of the nodes. Sampling rates of (c) sound sensor, (d) SF_6 gas sensor, and (e) temperature sensor of the different tier nodes. (f) Average sampling rates of different sensors.

sensing and forwarding are calculated accordingly. The phones use three sensors: video-camera, micro-phone and accelerometer. The power consumption of these sensors are 1258, 329, 96 mW [29] respectively with a sampling time of 5 seconds (approximated from [33]). The data payload size of the packets is assumed to be of 1 KB. We assume a radio transmission rate of 10 Mbps. We assume that these phones use their WiFi radios for packets transmission, which consumes approximately 1000 mW during the transmitting state [34].

5.2.2 Experimental Results. We next discuss the performance of the sampling rate adaptations of the phones, depending on their remaining battery capacities. The remaining battery capacities of the phones are uniformly randomly chosen among (25-100%). Depending on their battery capacities, we
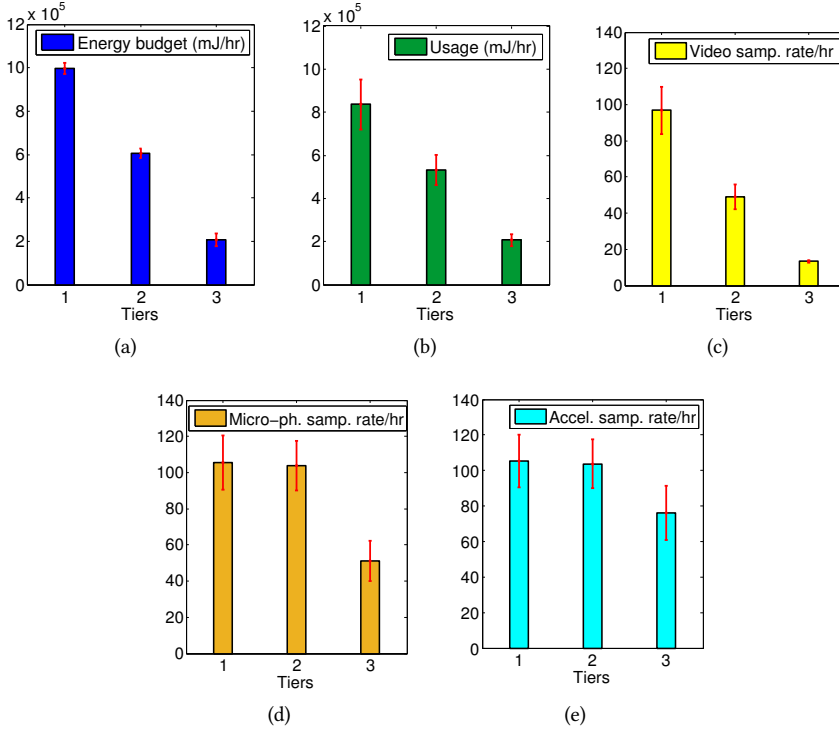
Fig. 10. (a) Energy budget and (b) usage profiles of the phones. Sampling rates of (c) video-camera, (d) micro-phone, (e) accelerometer sensor of the smartphones.

divide the phones in three *tiers*. The first, second, and third tier phones have battery charges of (75-100%), (50-75%), (25-50%) respectively. R_m and R_M are assumed to be 12/hr. and 120/hr. respectively. Fig. 10(a) shows the mean energy budget of the phones at different tiers, Fig. 10(b) shows their corresponding usages. The usage is defined as the cumulative energy expenditure due to sampling and forwarding of the packets. From Fig. 10(b) we can observe that the phone usages are adapted proportional to their power budgets. Fig. 10(c)-(e) show the mean sampling rates of video-camera, micro-phone and accelerometers at various levels. From these figures we can observe that overall the first tier phones have a sampling rate much higher than other tiers, due to their higher battery capacities. We can also observe that the average number of samples of the video-cameras are much lesser (∼ 2-6 times) compared to the other two sensors, especially for the second and third tier phones, which are more energy constrained. The reason is due to the higher power consumption of the video-cameras while data capturing. Fig. 11 shows the overall sampling rates of various sensors, which also confirms the fact that adaptation tries to conserve energy by mostly switching off sensors with higher power consumption.

Fig. 11. Average sampling rates of different sensors.

This shows the adaptive ability of the sampling rates based on the individual phone's power budget as well as the power consumption of the smartphone sensors.
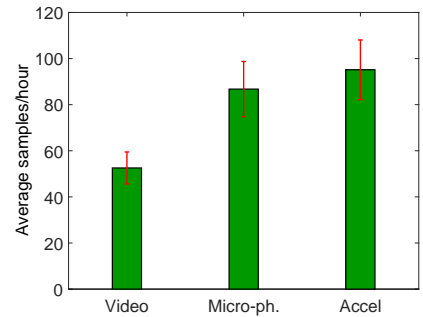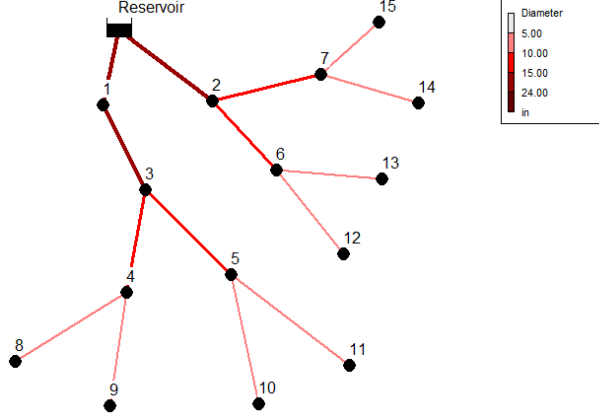
Fig. 12. Simulation topology of the water distribution network.

5.3 Effects of adaptations in pipeline contamination monitoring

5.3.1 Experimental Setting. We next simulate the rate adaptation schemes on a pipeline monitoring system for the purpose of water contamination detection. The simulated system topology along with the pipe diameters are shown in Fig. 12. Water from the reservoir first comes to nodes 1 to 3, which we denote as the first level nodes. The first level nodes distribute the water to nodes 4-7 (second level nodes), and from there to 7-14 (third level nodes). Each node is equipped with a fan that harvests energy and store in a super-capacitor. other than that the nodes are also equipped with sensors for contaminant monitoring. The cross-sectional area of the fans are chosen to be $\frac{1}{16}$-th of the pipe cross section, so that the normal water flow is not blocked by the fan movement.
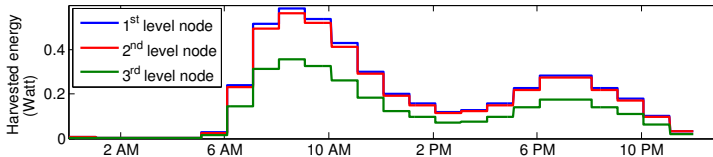

Fig. 13. Mean energy harvested over time for different nodes [35].

To model the harvested energy arrival at the sensor nodes we observe that the maximum water velocity in typical pipes are around 7.5 ft/sec [36]. We thus conservatively assume a water-velocity of 5.0 ft/sec for the third level nodes at peak hours and then compute those for the other two layers using the flow continuity relationships. The harvesting efficiency is assumed to be 10%. We assume that the super-capacitors of all the nodes are of 25Farad with an initial voltage of 2.7Vs. With this the mean energy profile of nodes at the three levels is shown in Fig. 13. Notice that the harvested energy at the super-capacitors depends on the water flow velocity and the fan diameter. In our example, the water flow-velocity increases at the lower levels, whereas the fan diameter decreases, which results in the behaviour shown in Fig. 13.

To design a contaminant monitoring scenario, we assume two types of water contaminants: Glyphosate and Dimethyl sulfoxide (DMSO) [37]. We also assume that all nodes are equipped with chlorine (Cl) sensors. To introduce the notion of heterogeneous sensing, we assume that the odd-numbered nodes are equipped with ORP sensors, whereas others are equipped with pH sensors. Both of the contaminants are detected by the chlorine sensor, whereas ORP and pH only respond
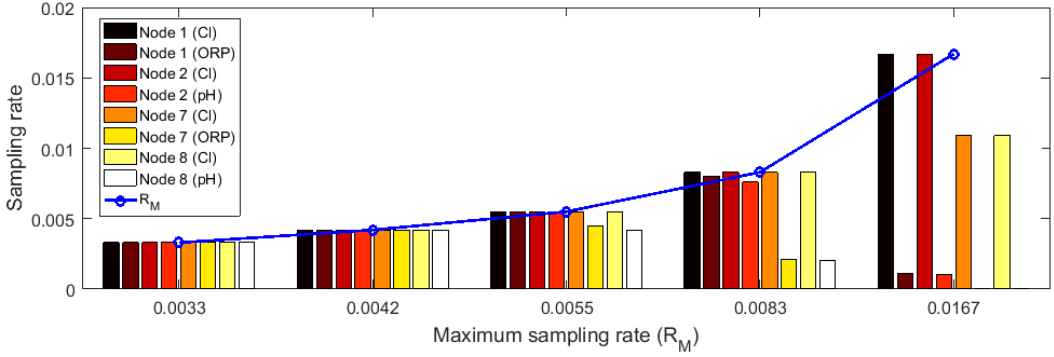
Fig. 14. Comparison of sampling rates of different sensors.

to Glyphosate[37, 38]. We also assume that a sensor detects a contaminant with a probability of 100% if the sensor responds to a contaminant and zero otherwise. The chlorine sensors consume 4 mA @ 12 V [27], whereas ORP and pH sensors consume 20.2 mA @ 10 V and 25.5 mA @ 10 V [28] respectively. The sampling time for these sensors are chosen to be 400 milliseconds. We assumed that the contaminants propagate at all the nodes in the downstream direction of the water flow. We use nodes 1, 3 and 7 to show the characteristics of first, second and third level nodes respectively. For the packet transmission, we assume MicaZ sensor nodes which consume ~20 mA at the transmit mode, with a transmission time of ~4 milliseconds.

5.3.2 Experimental Results. Fig. 14 shows the effect of sampling rate adaptation at the low energy hours, i.e. from 11 PM to 5 AM, with different R_M. Fig. 14 shows that at low R_M, all the sensors sample at their maximum sampling rates. As R_M increases, some of the sensors especially with higher current consumption, start reducing their sampling rates. We can also observe that at higher R_M, the first level nodes (i.e. 1-2) have higher sampling rates than the third level nodes (i.e. 7-8), because of more harvested energy availability at higher levels. Also notice that at high R_M, the Chlorine sensor is used more often as compared to others. This is because the Chlorine sensor consumes less power compared to OPR and pH, thus the devices use it more often at low energy hours.

6 RELATED WORKS

6.1 Network Utility Maximization

Utility maximization while sharing the network resources is a well researched area. In seminal works in [19, 20], the user's utility function is assumed to be strictly concave function of user's rate, whereas the resource constraints are linear. The users solve a distributed optimization problem to maximize their aggregate utility under their resource constraints. In [39] the authors have discussed the concept of fairness in utility maximization. In [40] the authors have proposed a Newton method based update procedure for speeding up the algorithm. Authors in [41] have used matrix splitting techniques to further improve the convergence time. In [42] the authors have proposed an alternating direction method of multipliers (ADMM) approach to improve the convergence speed. However, ADMM requires the objective function to fulfill the properties of strong concavity and strong smoothness, which may fail to hold in general. As opposed to the above literature our objective function is non-strictly concave, which makes the problem challenging to solve in a distributed manner.

References [16–18, 43, 44] have addressed the multi-path utility maximization problem, with non-strictly concave utility function with respect to the individual users rate. To convexify the utility function, the authors in [16, 17, 43, 44] have discussed a proximal approach, whereas in [18] the authors have presented a modified strictly concave utility function and proposed a successive approximation method for solving the optimization problem. The network utility maximization problem is used in different application areas including solving the flow-control problems [45, 46], charging plug-in electric vehicles system [47], provisioning local public goods in networks [48], solving power allocation and spectrum sharing problem [49] etc. In this paper we have solved the non-strictly concave objective function in a large-scale sensor network scenario, where the existence of multiple hops in between the sensing devices makes the problem even more challenging, whereas slower the convergence speed.

6.2 Energy Management, Rate Adaptation and Collaborative Sensing in WSNs

Energy management in sensor network is a well mined area. A number of studies have reported energy aware topology control [50–53]; power aware routing [54–57]; and sleep management, where a subset of wireless nodes are turned off to conserve energy [58–60]. Several MAC protocols [61–66] are introduced for low-power duty cycling to conserve energy. Multi-channel MAC protocols such as MMSN [62], TMMAC [63], MMAC [64] are also studied in the literature. Data compression and source coding based techniques are introduced in [67, 68]. Adaptive transmit power control for developing energy aware protocols are proposed in [31, 69–71]. In [70, 72] the authors have applied power control to reduce interference effects and thereby improve the communication performance, whereas the authors in [31, 71, 73] have used transmit power control for reducing the effect of network overhearing in WSNs. The use of multiple channels for alleviating the network overhearing minimization, and thereby reducing power consumption is discussed in [32, 74–76] etc. In [32, 77], the authors have proposed tree-based multi-channel protocol, whereas control theory and game theory based strategies are studied in [78, 79].

Adapting the energy consumption of the sensing devices can be achieved by adapting the rate at which data packets are generated for transmission. For exact reconstruction of the signal variations (e.g. for estimation of variations of temperature, vibrations, gas-density, etc.) the traditional method is to use periodic sampling, i.e. sampling at the *Nyquist* rate or higher. Adapting the sampling rate leads to non-uniform sampling, in which a key problem is achieving perfect reconstruction of the signal waveform from its samples. A number of non-uniform sampling schemes have been reported in the literature that can provide asymptotically zero reconstruction error with appropriate reconstruction methods, when applied to non-stationary signals [80]. *Level crossing sampling* (*LCS*) is a subclass of non-uniform sampling, which resolves this issue by sampling the signal when the signal crosses a set of predefined levels [81–83]. Concepts similar to LCS have been applied for various other signal-dependent sampling schemes such as zero-crossing sampling, Lebesgue sampling, and reference signal crossing sampling [84–86]. A review of different LCS schemes may be found in [87]. However, these approaches discuss independent rate adaptation by the sensor nodes, rather than collaborative and cooperative adaptations.

Sampling rate adaptation based techniques for balancing the workload and fairness are also widely studied in the context of sensor networks. In [88, 89] the authors have proposed a fair rate adaptation scheme for interference or congestion control in WSNs. In [90] the authors have proposed a rate adaptation scheme that iteratively uses linear programming for finding the lexicographic rate assignment for WSNs. Several energy aware sampling rate adaptation schemes are discussed in [2, 91]. Collaborative and distributed sampling rate adaptation schemes for energy harvesting sensor networks are discussed in [35, 92]. In [7] the authors have introduced a collaborative and

heterogeneous sensing scheme, where the effects of multiple, different types of sensors per device is considered, along with their inter-dependencies in the event reporting process. As opposed to these contributions, we have developed a general framework for collaborative proxy sensing, with multi-sensor equipped devices, while maximizing the overall utility of the network. We have also analyzed the limitations of typical distributed sampling rate adaptation schemes in such environments, and then develop alternate approaches to cope with the long convergence time.

7   CONCLUSIONS

In this paper we utilize the rich multi-modal sensory capability of the modern wireless devices to build a general framework for collaborative proxy sensing and sampling rate adaptation in a multi-sensor WSN environments. We have explore several distributed, collaborative sampling rate adaptation schemes that allow spatially correlated devices to share their data capturing tasks based on their energy availability and network participation for improved energy efficiency. We utilized the general sub-gradient method and the Nesterov's gradient descent method to solve this distributed rate adaptation problem, which result in slow convergence. To overcome the slow convergence of the fully distributed algorithms, we also proposed two semi-distributed solutions. We have also shown the performance of such collaborative, rate adaptation schemes in several practical scenarios. The paper has studied the distributed schemes with some key assumptions like static channel conditions and parent-child relations, no packet loss etc., the effects of which can be largely approximated through interpolation mechanisms in real-world scenarios.
.

REFERENCES

[1] I. Necoara and J. A. K. Suykens, "Application of a smoothing technique to decomposition in convex optimization," *IEEE Trans. Automat. Contr.*, vol. 53, no. 11, pp. 2674–2679, 2008.

[2] R. Liu, P. Sinha, and C. E. Koksal, "Joint energy management and resource allocation in rechargeable sensor networks," in *INFOCOM*, 2010, pp. 902–910.

[3] A. Pal and K. Kant, "On the feasibility of distributed sampling rate adaptation in heterogeneous and collaborative wireless sensor networks," in *ICCCN*, 2016, pp. 1–9.

[4] M. Raj, K. Kant, and S. K. Das, "E-DARWIN: energy aware disaster recovery network using wifi tethering," in *IEEE ICCCN*, 2014, pp. 1–8.

[5] http://www.ch2m.com/sites/default/files/content/article/attachments/CH2M-HILL-Quality-Monitoring.pdf.

[6] "Water Security Initiative: Interim Guidance on Planning for Contamination Warning System Deployment ," 2007.

[7] A. Pal and K. Kant, "Collaborative heterogeneous sensing: An application to contamination detection in water distribution networks," in *IEEE ICCCN*, 2015.

[8] http://www.nifc.gov/fireInfo/fireInfo_stats_totalFires.html.

[9] M. Ángeles Serna, R. Casado, A. Bermúdez, N. Pereira, and S. Tennina, "Distributed forest fire monitoring using wireless sensor networks," *International Journal of Distributed Sensor Networks*, 2015.

[10] M. E. Cambron, C. Brode, P. Butler, and G. Olszewkski, "Poacher detection at fence crossing," in *IEEE SoutheastCon*, 2015.

[11] "Iucn red list of threatened species," 2008.

[12] 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.

[13] A. Nasipuri, R. Cox, J. Conrad, L. V. der Zel, B. Rodriguez, and R. McKosky, "Design considerations for a large-scale wireless sensor network for substation monitoring," in *LCN*, 2010, pp. 866–873.

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

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

[16] W. Wang, M. Palaniswami, and S. H. Low, "Optimal flow control and routing in multi-path networks," *Performance Evaluation*, vol. 52, no. 2-3, pp. 119–132, 2003.

[17] X. Lin and N. B. Shroff, "Utility maximization for communication networks with multipath routing," *IEEE Trans. Automat. Contr.*, vol. 51, no. 5, pp. 766–781, 2006.

[18] P. L. Vo, T. A. Le, S. Lee, C. S. Hong, B. Kim, and H. Song, "Multi-path utility maximization and multi-path TCP design," *J. Parallel Distrib. Comput.*, vol. 74, no. 1, pp. 1848–1857, 2014.

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

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

[21] A. Beck, A. Nedic, A. Ozdaglar, and M. Teboulle, "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.

[22] P. Tsiaflakis, I. Necoara, J. A. K. Suykens, and M. Moonen, "Improved dual decomposition based optimization for DSL dynamic spectrum management," *IEEE Transactions on Signal Processing*, vol. 58, no. 4, pp. 2230–2245, 2010.

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

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

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

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

[27] http://www.burkert.com/en/type/8232.

[28] http://www.globalw.com/products/waterquality.html.

[29] F. B. Abdesslem, A. Phillips, and T. Henderson, "Less is more: energy-efficient mobile sensing with senseless," in *ACM MobiHeld*, 2009, pp. 61–62.

[30] M. Johnson, M. Healy, P. van de Ven, M. J. Hayes, J. Nelson, T. Newe, and E. Lewis, "A comparative review of wireless sensor network mote technologies," in *IEEE Sensors*, 2009, pp. 1439–1442.

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

[32] A. Pal and A. Nasipuri, "Distributed routing and channel selection for multi-channel wireless sensor networks," *J. Sensor and Actuator Networks*, vol. 6, no. 3, p. 10, 2017.

[33] Y. Wang, J. Lin, M. Annavaram, Q. Jacobson, J. I. Hong, B. Krishnamachari, and N. M. Sadeh, "A framework of energy efficient mobile sensing for automatic user state recognition," in *MobiSys*, 2009, pp. 179–192.

[34] L. Zhang, B. Tiwana, Z. Qian, Z. Wang, R. P. Dick, Z. M. Mao, and L. Yang, "Accurate online power estimation and automatic battery behavior based power model generation for smartphones," in *CODES+ISSS*, 2010, pp. 105–114.

[35] A. Pal and K. Kant, "Water flow driven sensor networks for leakage and contamination monitoring," in *IEEE WoWMoM*, 2015.

[36] "Design Criteria For Water Distribution Systems ," www.wmwd.com/DocumentCenter/Home/View/239, 2011.

[37] "Distribution SystemWater Quality Monitoring:Sensor Technology Evaluation Methodology and Results ," 2009.

[38] S. Liu, H. Che, K. Smith, and L. Chen, "Contamination event detection using multiple types of conventional water quality sensors in source water," *Environmental Science Processes and Impacts*, pp. 2028 – 2038, 2014.

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

[40] D. Dolev, A. Zymnis, S. P. Boyd, D. Bickson, and Y. Tock, "Distributed large scale network utility maximization," in *IEEE ISIT*, 2009, pp. 829–833.

[41] E. Wei, A. E. Ozdaglar, and A. Jadbabaie, "A distributed newton method for network utility maximization," in *IEEE CDC*, 2010, pp. 1816–1821.

[42] R. Gupta, L. Vandenberghe, and M. Gerla, "Centralized network utility maximization over aggregate flows," in *WiOpt*, 2016, pp. 227–234.

[43] H. Han, S. Shakkottai, C. V. Hollot, R. Srikant, and D. F. Towsley, "Multi-path TCP: a joint congestion control and routing scheme to exploit path diversity in the internet," *IEEE/ACM Trans. Netw.*, vol. 16, no. 6, pp. 1260–1271, 2006.

[44] T. Voice, "Stability of multi-path dual congestion control algorithms," *IEEE/ACM Trans. Netw.*, vol. 15, no. 6, pp. 1231–1239, 2007.

[45] R. Jain and J. C. Walrand, "An efficient nash-implementation mechanism for network resource allocation," *Automatica*, vol. 46, no. 8, pp. 1276–1283, 2010.

[46] F. Farhadi, S. J. Golestani, and D. Teneketzis, "A surrogate optimization-based mechanism for resource allocation and routing in networks with strategic agents," *IEEE Trans. Automat. Contr.*, vol. 64, no. 2, pp. 464–479, 2019.

[47] S. Bhattacharya, K. Kar, J. H. Chow, and A. Gupta, "Extended second price auctions with elastic supply for PEV charging in the smart grid," *IEEE Trans. Smart Grid*, vol. 7, no. 4, pp. 2082–2093, 2016.

[48] S. Sharma and D. Teneketzis, "Local public good provisioning in networks: A nash implementation mechanism," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 11, pp. 2105–2116, 2012.

[49] A. Kakhbod and D. Teneketzis, "Power allocation and spectrum sharing in multi-user, multi-channel systems with strategic users," *IEEE Trans. Automat. Contr.*, vol. 57, no. 9, pp. 2338–2342, 2012.

[50] R. Ramanathan and R. Rosales-Hain, "Topology control of multihop wireless networks using transmit power adjustment," in *IEEE INFOCOM*, 2000, pp. 404–413.

[51] R. Rajaraman, "Topology control and routing in ad hoc networks: a survey," *SIGACT News*, vol. 33, pp. 60–73, 2002.

[52] L. Li, J. Y. Halpern, P. Bahl, Y.-M. Wang, and R. Wattenhofer, "Analysis of a cone-based distributed topology control algorithm for wireless multi-hop networks," in *ACM PODC*, 2001, pp. 264–273.

[53] R. Wattenhofer, L. Li, P. Bahl, and Y. min Wang, "Distributed topology control for power efficient operation in multihop wireless ad hoc networks," in *IEEE INFOCOM*, 2001, pp. 1388–1397.

[54] S. Singh, M. Woo, and C. S. Raghavendra, "Power-aware routing in mobile ad hoc networks," in *ACM MobiCom*, 1998, pp. 181–190.

[55] S. Doshi, S. Bhandare, and T. X. Brown, "An on-demand minimum energy routing protocol for a wireless ad hoc network," *SIGMOBILE Mob. Comput. Commun. Rev.*, vol. 6, pp. 50–66, 2002.

[56] S. Doshi and T. X. Brown, "Minimum energy routing schemes for a wireless ad hoc network," in *IEEE INFOCOM*, 2002.

[57] J.-H. Chang and L. Tassiulas, "Energy conserving routing in wireless ad hoc networks," in *IEEE INFOCOM*, 2000.

[58] B. Chen, K. Jamieson, H. Balakrishnan, and R. Morris, "Span: an energy-efficient coordination algorithm for topology maintenance in ad hoc wireless networks," *Wirel. Netw.*, vol. 8, pp. 481–494, 2002.

[59] A. Cerpa and D. Estrin, "ASCENT: Adaptive self-configuring sensor networks topologies," in *IEEE INFOCOM*, 2004.

[60] X. Wang, G. Xing, Y. Zhang, C. Lu, R. Pless, and C. Gill, "Integrated coverage and connectivity configuration in wireless sensor networks," in *SenSys*, 2003, pp. 28–39.

[61] Y. Kim, H. Shin, and H. Cha, "Y-MAC: An energy-efficient multi-channel mac protocol for dense wireless sensor networks," in *IPSN*, 2008, pp. 53–63.

[62] G. Zhou, C. Huang, T. Yan, T. He, J. A. Stankovic, and T. F. Abdelzaher, "MMSN: Multi-frequency media access control for wireless sensor networks," in *INFOCOM*, 2006.

[63] J. Zhang, G. Zhou, C. Huang, S. H. Son, and J. A. Stankovic, "TMMAC: An energy efficient multi-channel mac protocol for ad hoc networks," in *ICC*, 2007, pp. 3554–3561.

[64] X. Chen, P. Han, Q.-S. He, S. liang Tu, and Z. long Chen, "A multi-channel mac protocol for wireless sensor networks," in *CIT*, 2006.

[65] M. D. Jovanovic and G. L. Djordjevic, "TFMAC: Multi-channel mac protocol for wireless sensor networks," in *TELSIKS*, 2007.

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

[67] C. Tang and C. S. Raghavendra, "Compression techniques for wireless sensor networks," in *Wireless Sensor Networks*. Springer US, 2004, pp. 207–231.

[68] Z. Xiong, A. Liveris, and S. Chen, "Distributed source coding for sensor networks," *IEEE Signal Processing Magazine*, vol. 21, no. 5, pp. 80–94, 2004.

[69] S. Sengupta, M. Chatterjee, and K. A. Kwiat, "A game theoretic framework for power control in wireless sensor networks," *IEEE Trans. Computers*, vol. 59, no. 2, pp. 231–242, 2010.

[70] S. Lin, J. Zhang, G. Zhou, L. Gu, J. A. Stankovic, and T. He, "ATPC: adaptive transmission power control for wireless sensor networks," in *SenSys*, 2006, pp. 223–236.

[71] A. Pal, B. Soibam, and A. Nasipuri, "A distributed power control and routing scheme for rechargeable sensor networks," in *IEEE SoutheastCon*, 2013.

[72] Y. Fu, M. Sha, G. Hackmann, and C. Lu, "Practical control of transmission power for wireless sensor networks," in *ICNP*, 2012.

[73] A. Pal and A. Nasipuri, "Joint power control and routing for rechargeable wireless sensor networks," *IEEE Access*, vol. 7, pp. 123 992–124 007, 2019.

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

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

[76] A. Pal and A. Nasipuri, "Lifetime of asynchronous wireless sensor networks with multiple channels and power control," in *IEEE WCNC*, 2014, pp. 2874–2879.

[77] Y. Wu, J. A. Stankovic, T. He, and S. Lin, "Realistic and efficient multi-channel communications in wireless sensor networks," in *INFOCOM*, 2008, pp. 1193–1201.

[78] H. K. Le, D. Henriksson, and T. F. Abdelzaher, "A control theory approach to throughput optimization in multi-channel collection sensor networks," in *IPSN*, 2007, pp. 31–40.

[79] Q. Yu, J. Chen, Y. Fan, X. Shen, and Y. Sun, "Multi-channel assignment in wireless sensor networks: A game theoretic approach," in *INFOCOM*, 2010, pp. 1127–1135.

[80] F. Marvasti, *Nonuniform Sampling*. Kluwer Academic, 2001.

[81] M. Miskovicz, "Efficiency of level-crossing sampling for bandlimited gaussian random process," in *Proceedings of IEEE International Workshop on Factory Communication Systems*, June 2006, pp. 137–142.

[82] K. Guan and A. Singer, "A level-crossing sampling scheme for non-bandlimited signals," in *Proceedings of International Conference on Acoustic, Speech and Signal processing*, 2006, pp. 381–83.

[83] K. Guan and A. Singer, "Opportunistic sampling of bursty signals by level-crossing: an information theoretical approach," in *Proceeding of Conference on Information Science and Systems*, March 2007, pp. 701–707.

[84] F. Bond and C. Cahn, "On the sampling the zeros of bandwidth limited signals," *Information Theory, IRE Transactions on*, vol. 4, no. 3, pp. 110–113, 1958.

[85] K. Astrom and B. Bernhardsson, "Comparison of riemann and lebesgue sampling for first order stochastic systems," in *Decision and Control, 2002, Proceedings of the 41st IEEE Conference on*, vol. 2, 2002, pp. 2011 – 2016 vol.2.

[86] I. Bilinskis, *Digital Alias Free Signal Processing*. John Wiley & Sons, New Yor, 2007.

[87] S. M. Qaisar, L. Fesquet, and M. Renaudin, "Adaptive rate sampling and filtering based on level crossing sampling," *EURASIP J. Adv. Signal Process*, vol. 2009, pp. 32:1–32:12, January 2009. [Online]. Available: http://dx.doi.org/10.1155/2009/971656

[88] S. Rangwala, R. Gummadi, R. Govindan, and K. Psounis, "Interference-aware fair rate control in wireless sensor networks," in *SIGCOMM*, 2006, pp. 63–74.

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

[90] S. Chen, Y. Fang, and Y. Xia, "Lexicographic maxmin fairness for data collection in wireless sensor networks," *IEEE Trans. Mob. Comput.*, vol. 6, no. 7, pp. 762–776, 2007.

[91] L. Su, Y. Gao, Y. Yang, and G. Cao, "Towards optimal rate allocation for data aggregation in wireless sensor networks," in *MobiHoc*, 2011, p. 19.

[92] B. Zhang, R. Simon, and H. Aydin, "Maximum utility rate allocation for energy harvesting wireless sensor networks," in *MSWiM*, 2011, pp. 7–16.