Water Flow Driven Sensor Networks for Leakage and Contamination Monitoring in Distribution Pipelines

AMITANGSHU PAL and KRISHNA KANT, Temple University, USA

In this article, we introduce the concept of Water Flow Driven Sensor Networks for leakage and contamination monitoring in urban water distribution systems. The unique aspect of our work is that the sensor network can be deployed in the underground water network with only access to connection points (through manholes) and driven only by water harvested energy without the need for AC power or frequent battery changes. Although water systems may be affected by a large variety of contaminants, only a few sensors can be practically deployed. Thus, many types of contaminants are sensed via "proxy sensing," which may not be 100% reliable. The main problems addressed are (a) adaptation of the network to the available energy to maximize leak/contamination detection and (b) minimal artificial water circulation or leakage to improve detectability during periods of almost zero natural water flow. The article shows, through extensive simulations, that the proposed approach can drastically reduce the leakage/contamination reporting time (from 3.5h up to ~6min), and the adaptation can reduce this circulation by ~33% and yet enhance the collected/transmitted data by 30%.

CCS Concepts: • **Networks** → **Sensor networks**;

Additional Key Words and Phrases: Water distribution sensor networks, heterogeneous sensing, water quality monitoring, leak detection, adaptive sampling

ACM Reference format:
Amitangshu Pal and Krishna Kant. 2019. Water Flow Driven Sensor Networks for Leakage and Contamination Monitoring in Distribution Pipelines. *ACM Trans. Sen. Netw.* 15, 4, Article 37 (August 2019), 43 pages.
https://doi.org/10.1145/3342513

1 INTRODUCTION

Water Distribution Systems (WDS) carry fresh water from supply sources and storage reservoirs/tanks to industrial, commercial, and residential areas through a complex web of pipeline systems. However, fresh water supplies continue to dwindle, and by 2025, two-thirds of the world will experience water stress and about 25% will experience abject water scarcity [1]. While the stress on urban water systems continues to increase due to movement of population to urban areas, most of these systems are in poor shape and subject to significant amounts of water leaks, seepage, and contamination [2]. Traditional water quality monitoring is largely manual [3] and inadequate for large, stressed water distribution systems. Thus, there is a great urgency to develop Information and Communications Technology-based solutions that can detect and localize leaks and contamination much more cheaply than mostly manual procedures followed today. A quick

Authors' addresses: A. Pal (corresponding author) and K. Kant, Temple University, 1925 N. 12th Street, Philadelphia, PA, 19121, USA; emails: {amitangshu.pal, kkant}@temple.edu.

1550-4859/2019/08-ART37 $15.00
https://doi.org/10.1145/3342513

detection also helps to increase the working lifetime of these systems and is immensely valuable to cash-strapped water distribution utilities.

The main objective of this article is to develop a sensor network that continuously monitors water leaks and contamination in water pipes and reports relevant data to a control station that can do the necessary analytics for detection and localization. Though conceptually straightforward, effective solutions to this problem are extremely challenging due to the following limitations of the environment: (a) most pipes are mostly buried underground, and only accessible at connection points through manholes, (b) manholes usually do not have access to AC power, (c) RF communications through the water, pipe material, soil, rock, and so on, face high attenuation and thus require high power, (d) it is not possible to harvest solar energy for powering the sensing and communications, and (e) changing batteries regularly in manholes can be quite expensive.

To address these issues, in this article, we consider *water flow driven sensor networks (WDSN)* [4, 5] that are entirely powered by water flow via a small hydro fan unit. We use a small super-capacitor for storing the harvested energy, primarily because of the long cycle life and high charge-discharge efficiency of current super-capacitors [6]. The sensor node is assumed to be at pipe connection or valve points only, installed through the manholes. Its lower part dips into the water for energy harvesting and measurement of contamination, velocity, and so on, and the upper part sports the energy storage, voltage booster, regulator and computing/communications unit. We assume that the upper part has a suitable wireless radio (e.g., WiFi) with antennas embedded on the exposed side of "smart" manhole covers that are already available [7]. In practice, the sensor modules may be deployed only at certain connection point, for example, the connections of only larger diameter pipes or those that very old or otherwise vulnerable to leakage or contamination. However, for simplicity, we will assume deployment at every connection point.

Due to the varying flow rate in the pipes (driven by water consumption), the availability of harvested energy varies both in spatial and temporal domains. In branch pipes, the flow rate may vary significantly and drop to near zero late at night and in other special circumstances such as very few people being at home during normal working hours. Thus, it is essential to adapt the data collection and transmission to the available energy profile. By exploiting the highly correlated detection ability of the individual sensor nodes, we develop a dynamic sampling and transmission rate adaptation scheme based on individual node's energy budget. Since most water distribution systems have pumps that can be controlled centrally from a control room, it is possible to circulate water into the system artificially by pumping water from a reservoir back to the same or another reservoir. We study the role of such circulation to keep the network alive during low natural flow rates and thereby improve the contamination/leakage detection capabilities.

Since the water can be contaminated in numerous ways, sensing for individual contaminant is impractical, and it is essential collaboratively use a set of sensors deployed at different junctions so that they can collectively provide a good coverage of all major contaminants. In this article, we develop a collaborative and adaptive sampling rate adaptation scheme, based on the individual node's harvested energy as well as the correlated sensing abilities of various quality sensors. Thus, the key contributions of this article are as follows:

- We develop a sensor network architecture where fixed sensors are deployed through manholes and are powered from the harvested energy from the flowing water.
- Since the flow rate in the pipes may at times be inadequate to harvest adequate energy to keep the sensor network alive, we propose an optimal sampling and transmission rate adaptation scheme based on the nodes energy budget.
- In systems where an automated artificial water circulation is feasible, we examine the problem of minimal circulation to keep the sensor network alive.

- We exploit the correlated sensing abilities of multiple quality sensors to effectively sense for several classes of contaminants by forming coalitions of sensors.
- We quantify the advantages of our approach via extensive simulation studies using available measurement data and shows that the proposed mechanism can reduce reporting time from 3.5h up to ~6min during late nights with only very small artificial circulation when needed.

To the best of our knowledge, this is the first work on water distribution systems that tries to achieve optimal monitoring under the real-world practical constraints of the water networks.

As in any cyberphysical system, security becomes an important issue in WDSN as well, but a comprehensive treatment of security issues is beyond the scope of this article. For the most part, WDSN security is subject to same challenges and can use the same approaches as other wireless sensor networks (e.g., see, References [8, 9]). We have developed a lightweight integrity mechanism for critical smart grid communications and this can be used for WDSN as well [10].

The outline of the article is as follows. Section 2 provides an overview of a WDS and discusses the motivation of the leakage and contamination detection problem. Section 3 describes the WDSN and the artificial water circulation mechanism. Section 4 then develops a WDSN charging model to provide adequate energy to the sensor network during low water flow periods. Sections 5 and 6 address the problem of sensing rate adaptation based on the node's energy budget. Section 7 then presents the evaluation of the scheme. The related works are reported in Section 8. The article is concluded in Section 9.

2 BACKGROUND

2.1 Problem of Water Leakage and Contamination

The problem of leakage in urban water distribution systems pervades throughout the world, including, significantly, the most advanced countries as well. For example, in the U.S., most water systems are 100+ years old, particularly in large cities on the east coast. For example, a 2010 audit in Philadelphia revealed 26% water loss due to leakage and another 8% due to metering inaccuracies, water theft, and data handling and management issues [12]. A comprehensive survey in Reference [11] shows loss percentages ranging from 15% to 35% over 36 cities in the U.S. as summarized in Table 1. Europe loses more than 25% of its water to leaks, with some countries reaching the 50% mark [13].

The EPA report [14] provides a comprehensive look at the water loss in the U.S. drinking water distribution systems, including details on water metering, meter types, water auditing, leak detection and location, pipe replacement/rehabilitation, and maintenance and preventive measures. A recent paper by Kunkel of the Philadelphia Water Department [15] and his earlier book [16] provide comprehensive discussion of the state of water systems and leak monitoring/fixing. Reference [17] provides a detailed review of practical leak management methodologies for wpipes.

Contamination goes hand in hand with leakage due to seepage through leaks, rusted pipes, internal build ups, operational mistakes, and so on. The incidents of waterborne outbreaks are numerous. In Walkerton, Canada, 2,500 people were poisoned, and 7 died by *Eschericia coli* in drinking water, following a resource contamination, in the year 2000 [18]. In 2007, 8,500 people were ill in Nokia, Finland due to a cross-connection of wastewater into the distribution network.

As the potential contaminants or chemicals are numerous, sensing each of the specific chemicals can be very expensive and slow, since they often require taking water samples that are treated with suitable reactants, measured for specific byproducts, and then discarded. This difficulty forces *proxy sensing* techniques where the presence of certain chemicals is *inferred* via easily measurable properties of the water such as pH, conductivity, temperature, or depletion profile of added chlorine.

Table 1. Water Loss in Distribution Systems [11]

Water department	Loss	Water department	Loss	Water department	Loss
Philadelphia Water Department (PA)	31.18%	Cleveland Division of Water (OH)	28.72%	Memphis Light, Gas & Water (TN)	15.20%
Jefferson Parish Water Department (LA)	24.12%	Portland Water District (ME)	18.06%	Ann Arbor Utilities Department (MI)	25.78%
North Penn Water Authority (PA)	16.25%	Waterloo Water Works (IA)	15.58%	Lorain Utilities Department (OH)	20.00%
Madison County Water Department (AL)	26.77%	Elmira Water Board (NY)	25.27%	Lebanon Authority (PA)	21.08%
Renton (WA)	18.66%	Williamsport Municipal Water Authority (PA)	35.13%	Albany (OR)	24.91%
Lake County East Utilities (OH)	15.72%	Paradise Irrigation District (CA)	16.57%	Cordele (GA)	15.19%
Piqua Municipal Water System (OH)	21.10%	Fredericksburg (VA)	25.00%	Clearfield Municipal Authority (PA)	23.61%
Miami Utility Department (OK)	26.61%	Glens Falls Water Department (NY)	24.48%	City of Converse Public Works (TX)	29.85%
Anson County Water System (NC)	24.87%	Berea College Utilities (KY)	18.10%	Crossett Water Commission (AR)	16.52%
Cincinnati Water Works (OH)	17.65%	Duluth/ Public Works & Utilities/ Water (MN)	16.23%	Selmer Utility Division (TN)	25.00%
Shoshone Municipal Pipeline (WY)	15.19%	Bellingham DPW (MA)	23.43%	Spencer Municipal Utilities (IA)	15.90%
Warren County Utility District (TN)	16.67%	Cleveland Division of Water (OH)	28.72%	Eastpointe Water and Sewer (MI)	25.88%

Such proxy sensing is not intended for sensing any specific type of contaminant but can sense some aspects of the contaminant's property and thus provide either corroboration or sensing of the event at some degraded level of reliability. We will assume that the sensing modules are equipped with multiple, *different* water quality sensors that measure the common surrogate parameters and report them to a centralized station. This introduces the notion of a heterogeneous water distribution sensor network. The purpose of bringing the heterogeneity in sensing is that deploying a large number of water quality sensors at every junction points is expensive and redundant for cash-strapped water distribution utilities. Thus, a subset of different types of sensors are introduced at the junction points, which together can give a cumulative spatial water quality measurement of a distribution area.

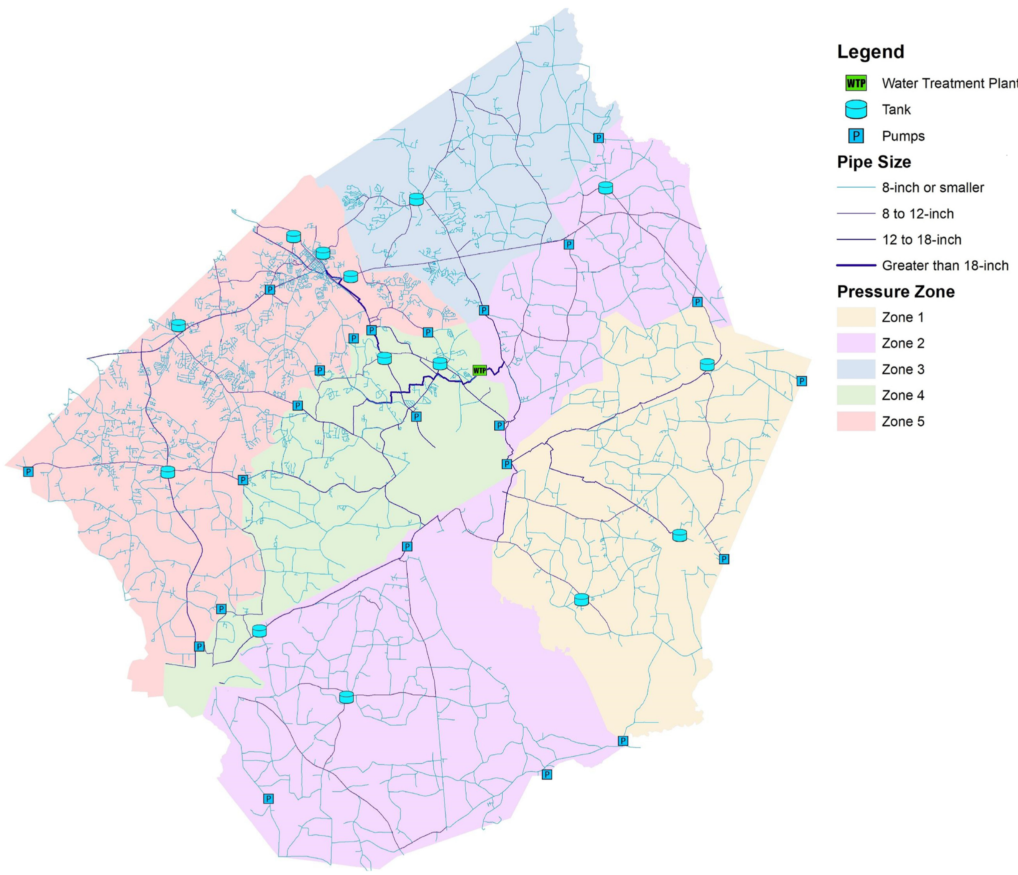


Fig. 1. A sample water distribution system showing different pipe sizes and pressure zones.

2.2 Water Distribution Networks

A water distribution system consists of a number of water reservoir and *main* lines running from them, further divided into *sub-mains* and *branch* lines from where service connections are given to the customers. The network has a number of pumping stations that can generate the flows as needed for the water distribution both for consumption and for storage into tanks. Increasingly, the network has a sophisticated Supervisory Control and Data Acquisition system that can be accessed through a control room, thus enabling operation of pumps and controlling the pumped water volume remotely [19]. We will exploit this capability to generate an *occasional* artificial water circulation for keeping the proposed sensor network alive.

Figure 1 shows an example of a water distribution network. The overall network is pretty irregular, with more densely connected sections serving more populated areas and layout closely following the street layouts. The colors represent different water pressure zones, zone-1 being lowest and zone-5 highest. The graph is cyclic, but the flows are affected by the valve settings—certain valves may allow only one-way flows. However, loops in water flow do occur and are normal. The node degree is uniformly small and rarely exceeds 4. The pipes and connections, on account of their age, workmanship, operating environment, and materials used may be in varying physical conditions including deposits, rust/corrosion, cracks/holes, weakened portions, poor/leaky fittings, and so on, possibly allowing for seepage of contaminants or vulnerability to deliberate infrastructure attacks.

Water distribution networks are normally divided up into *District Metering Areas* or DMAs. A DMA typically spans a neighborhood with only a few incoming and outgoing main lines and can be regarded as a subnetwork that the utility would like to be able to isolate or regulate. Thus, all incoming and outgoing pipes from the DMA have valves and flow/pressure meters installed. At the very least, sensing/communications should be installed at the granularity of DMAs, but denser deployment can be highly desirable.

2.3 Sensing of Water Leakage and Contamination

Numerous sensing techniques have been developed for monitoring of water infrastructure, and they span many different technologies, including magnetics, sound, ultrasound, thermal sensing, video imaging, and so on. Depending on the type, the sensors may be installed in the vicinity of water pipes, on the pipe surface, or inserted inside the pipes. The U.S. EPA report in Reference [14] provides a detailed discussion of a variety of sensors. Ultrasonic utilizes time-of-flight measurements of wave propagation (Doppler shift) of an applied ultrasonic signal to determine the fluid velocity. Magnetic Induction produces voltage proportional to the flow velocity and is relatively accurate across a wide range of flow rates. Both devices measure the water flow rate, and relating that to small leaks in a noisy environment with many pressure transients can be very challenging.

Acoustics is often used for a direct leak detection by determining the location and extent of leak using vibrations generated by leaking water [20]. The frequency of vibrations generally goes down as leak size increases; therefore, the method is more useful for smaller leaks. However, smaller leaks generate lower volume. A current method on the market (Sahara leak detection) inserts a tethered acoustic device in the pipe for leak detection, and has been used to detect water leaks. The efficacy of acoustic devices varies widely depending on several factors, such as the type of leak, the opening size, pipe material, and soil conditions [21]. In some cases, it may be impossible to detect leaks with acoustic devices, such as when there is a background leak, when the pressure caused by the leak is very low, or when the soil is already waterlogged at the time of sounding. Geophones can be used to detect/monitor ground-level sounds. However, overall, such methods are crude at best and require significant operator expertise/training. Many other techniques have been devised but they all suffer from one problem or another. Chemical tracers are not very accurate in determining leak volume or leak location. Ground-penetrating radar (GPR) can detect and locate the leaks accurately [22–24] but requires heavy human involvement along with a map of the distribution system. Methods based on change in solid properties due to leak can be useful but not very accurate. Artificially created pressure transients (e.g., by opening/closing valves) can be used for leak detection and location [25] but require painstaking calibration and cannot handle long pipes where the transients would fade quickly. Leak detection using in-pipe inspection robots are discussed in References [26, 27].

Numerous water quality sensors are commonly used for monitoring routine water quality parameters, such as pH, chlorine, total organic carbon (TOC), oxidation reduction potential (ORP), and conductivity and temperature [28, 29]. Several studies are devoted towards contamination event detection using the data obtained from these sensors. In Reference [30] the authors have studied the responses of different contaminants on off-the-shelf commercial products for monitoring standard drinking water parameters (such as pH, free chlorine, oxidation reduction potential (ORP), dissolved oxygen, specific conductance, turbidity, total organic carbon (TOC), chloride, ammonia, and nitrate). Most sensors respond to a large number of tested contaminants. In Reference [31] the authors have explored a real-time event adaptive detection, identification and warning (READiw) system using 11 chemical and biological contaminants. In Reference [32] the author has studied signals from five separate water quality measurements (pH, conductivity, turbidity, chlorine residual, TOC) to trigger a contamination event. In Reference [33] the authors have presented a contamination detection methodology in drinking water system using Dempster-Shafer evidence theory. Data driven estimation model for water contamination detection is reported in References [34, 35].

As discussed earlier, deploying the suitable sensing and communications in the underground water network infrastructure is very challenging, which makes many approaches impractical. For example, dense deployment of sensors on the pipe surface has been proposed in many solutions, including PipeNet [36], NAWMS [37], MISE-PIPE [38], and PipeTECT [39], but is unlikely to be practical. Other proposals such as SPAMMS [40] requires RFID tags painted inside of the pipe, which again is impractical. Flowing sensors such as the flowing acoustic sensors proposed in Kadri [41] are more realistic but require carefully planned manual insertion and removal from the pipes and could get stuck inside the pipe (e.g., in the growing algae) or around the valves. Self maneuvering robotic flowing sensors such as TriopusNet [42] can be more dextrous but more suited to special investigations rather than routine use. Other proposals on fixed sensors include [43–46] and have similar limitations.

Contrary to these approaches, we focus on techniques that only involve monitoring of water flow characteristics and contamination at the connection points that are accessible for sensing. The communications are somewhat challenging, since the only practical way to deploy an antenna outside the manhole is to integrate it on top of the smart manhole cover (e.g., a loop or dipole antenna that is sturdy and sits flat on the top of manhole cover so that driving over it does not affect its operation. The leakage detection still remains challenging, since the leak will likely occur somewhere in the middle of the pipe and the only parameters that can be measured at connection points are flow rate, pressure, and possibly flow pattern. For larger leaks, a historical recording of flow rate and pressure and comparing changes over time can be used to determine the potential location of the leaks to the extent of pipe segments. A more accurate localization can be done by running a co-simulation of the network using a package such as WaterGEMS. However, this article is not focused on exploring such methods in details; the article is about keeping the sensing network operational and working efficiently so that it can do the best job that the deployed sensors can support.

3 WATER FLOW DRIVEN SENSOR NETWORK (WDSN)

We assume a software-based leak/contamination detection scheme, where the whole DMA is modeled in a simulator, such as Water-GEMS [47]. The sensor nodes placed at different sections of the DMA record and report different contamination monitoring parameters along with pressure, temperature, velocity, and so on. These sensor readings are then compared against the simulated values at those points. If the sensor data from a node shows a wide variation from the simulator's output, then a leak/contamination is suspected in the nearby regions of the sensor node. Thus, using the deployed sensors to continuously monitor the hydraulics of the DMA system, anomalies (such as substantial changes in pressure or flow demand) that might indicate leaks can be found as they happen, which is illustrated in Figure 2(a). After such a leak/contamination is suspected, traditional direct and manual inspection methods [48, 49] can be exercised to pinpoint the leak location or contamination origin.

3.1 Network Operation

As stated earlier, we assume that the sensing and communication nodes are deployed only at connection points and harvest energy from flowing water. We assume that the nodes are not time

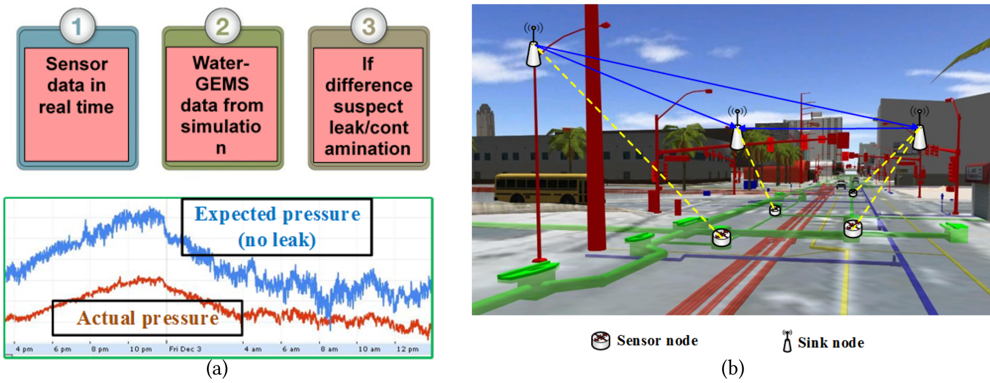


Fig. 2. (a) Indirect leak/contamination detection in a WDSN. (b) Architecture of the proposed WDSN.

synchronized and use the basic *Low Power Listening (LPL)* [50] principle to conserve energy.[1] In LPL, idle receivers run on a suitable sleep/awake duty cycle, and the senders always prepend their message with a sufficiently long preamble to ensure communication with a receiver caught sleeping. In addition, we also assume a set of strategically deployed sink nodes for data collection. These sink nodes are assumed to have a steady source of power (e.g., AC or long lasting batteries) and have a second communication interface (likely wired) to the central control node for the DMA. We assume that the sink nodes are deployed separate from the limitation of manhole locations—based on accessibility and power availability considerations. All non-sink nodes collect, store and forward their sensing data and remaining energy to their nearest sink node using *single-hop, direct WiFi/long-range Zigbee communication*. While, in general, multi-hop communication may also be enabled via tree-based forwarding mechanisms [51–53], we limit ourselves to single hop communication in this article. The single hop limitation may require deploying multiple sink nodes in a DMA, which are energy sufficient and always active. They can communicate with each others and with the control center via long-range WiFi/3G/4G/LTE technologies. Although not necessary, for simplicity, we assume that one such sink node is special in that all other send their data to it for final transmittal to the control center. The entire WDSN architecture is shown in Figure 2(b).

The energy harvested by the normal sensor nodes in a WDSN depends on the water flow rate. We assume that each node is equipped with a suitable fan-based harvester, where kinetic energy of the streaming fluid rotates the blade and generates electricity as shown in Figure 3(a). The basic equations governing this energy conversion are well established [54]. The kinetic power (in Watts) of the moving fluid at velocity v (m/s), passing through the fan of area A (m^2) is given by

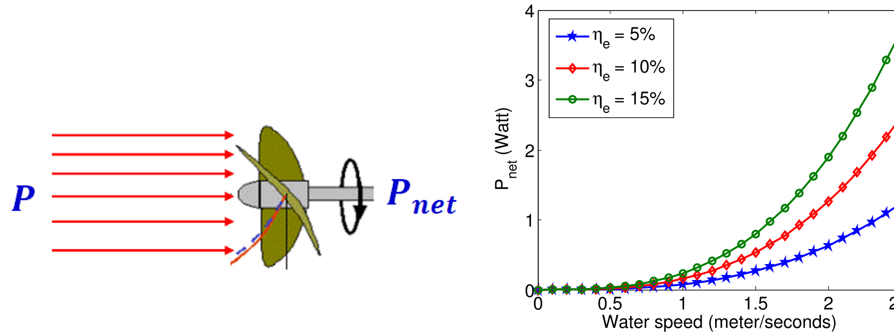
$$P = \frac{\partial}{\partial t}\left(\frac{1}{2}mv^2\right) = \frac{1}{2}v^2\frac{\partial m}{\partial t} = \frac{1}{2}v^2\rho A v = \frac{1}{2}\rho A v^3, \tag{1}$$

where m is the mass of the fluid and ρ is the density (1,000kg/m^3 for water). Let η_{fd} denote the fluid-dynamic efficiency of the rotating body, η_{em} the electro-mechanical conversion efficiency, and η_{sc} the charging efficiency of super-capacitor. Then the net electrical power for super-capacitor charging is

$$P_{\text{net}} = \eta_{fd}\eta_{em}\eta_{sc} = \eta_e \cdot P, \tag{2}$$

where η_e denotes the overall efficiency. Needless to say, η_e depends on a large variety of factors, and is the domain of mechanical/electrical design. Here, we only assume a plausible range for η_e—usually not much better than $\tilde{1}0\%$. Figure 3(b) shows the final harvested power P_{net} as a function of

[1]The purpose of using asynchronous MAC is to avoid the overhead of time synchronization in a large WDSN.



(a) (b)

Fig. 3. (a) A fan-based energy harvester. (b) P_{net} with different water velocities.

water flow rate with 2.5" diameter fan and $\eta_e = 5–15\%$. The most interesting aspect of this graph is that at very low water velocities, the harvested energy is effectively zero, thus requiring effective management to ensure operation during low flow periods, particularly late at night when the flow rate may stay low for hours.

3.2 Keeping Network Alive

The simplest approach to keep the WDSN alive is to simply choose adequate capacity super-capacitors to get through long lull periods. However, this approach not only makes the solution very expensive but also ignores an important aspect of water distribution networks: if the flow rate is very low, then the leakage rate and contamination spread rates will also be very low. Thus, a better idea is to *adapt* the sampling and transmission rates to the charging rate and thereby provide effective coverage without needing large energy storage. In fact, there is a sort of inherent compensation mechanism here: if a large leak develops during the lull period, the relevant sensors will automatically get charged up and become operational. Similarly, a very leaky system may always provide adequate harvestable energy, and large capacitors are wasteful. Nevertheless, it may undesirable to let the contamination monitoring frequency go down drastically during long lull periods. For this, we propose an artificial water circulation mechanism within the DMA to replenish super-capacitors.

We exploit this capability to circulate water artificially for the purposes of generating flows. Such water circulation does not entail any water loss—it is simply a circular movement among reservoirs as shown in Figure 4. Here the water injected or taken out of the system from nodes 1 and 2, and would also result in additional water flows in other nearby loops as shown. Obviously, the impact of an isolate circulation will go down rapidly as we move away from circulation area. In other words, if we want a significant artificial water flow in segments that are multiple hops away from the reservoirs and pumps (e.g., segment 5–6 in Figure 4), we would need substantial artificial flow rate, which may not be possible or desirable. In those cases, we can deploy another trick—an artificial drainage of water at certain points (e.g., at node 5 or 6). This capability—operable from the control room—is also becoming increasingly available in water systems, mostly for the purposes of flushing the pipes. For obvious reasons, we want to minimize such artificial *leakage*. Artificial circulation and leakage mechanisms may also be useful without automated control, if manual action is very infrequent.

Obviously, there is a tradeoff between the capacity (and hence cost/size) of super-capacitors and the frequency and magnitude of the artificial flows created. Large super-capacitors can be charged by 1–2 significant flows during the night, but smaller ones will require many smaller flows. It is

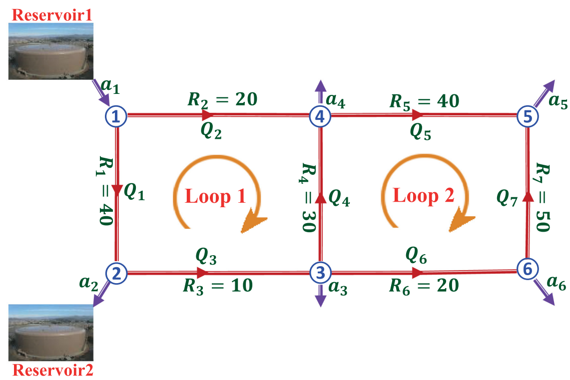


Fig. 4. Graphical representation of a simple water distribution network.

possible to define an optimization problem that determines super-capacitor sizing based on all these factors, but we do not delve into that issue for lack of space. Instead, we discuss a model for calculating the required flow rate at different connection points to provide sufficient harvesting energy for all the sensor nodes. This aspect is naturally coupled with the basic rate adaptation mechanism, which is required to minimize need for circulation.

4 WDSN ARTIFICIAL WATER CIRCULATION MODEL

We assume that all sensor nodes in the DMA are charged for a short *charging time* τ, whenever the voltage of certain number of sensor nodes drops below a threshold V_{thresh}. Let V_{ini} and V_{target} denote the initial and required target voltages, and $\mathscr{C}$ the capacitance of the super-capacitor. Then the energy stored in the super-capacitor by charging is given by [55]

$$P_{\text{net}}.\tau \geq \frac{1}{2}\mathscr{C}\left(V_{\text{target}}^2 - V_{\text{ini}}^2\right), \tag{3}$$

where P_{net} is the charging power in Equation (2). It follows that the required water velocity is given by

$$v \geq \left[\mathscr{C}\left(V_{\text{target}}^2 - V_{\text{ini}}^2\right)/\tau \cdot \rho \cdot A \cdot \eta_e\right]^{\frac{1}{3}} = \mathbb{V} \text{ (assume)}. \tag{4}$$

Thus, after charging, all nodes will have voltages of at least V_{target}. (Note that the nodes that are almost fully charged already may not take much additional charge.) The artificial water flow for charging can be generated only at *pumping points* $\mathbb{PP}$, which are the pumps associated with reservoirs/tanks. From $\mathbb{PP}$s, water can be pumped in or taken out of the system at certain regulated rates. In reality there is a maximum limit of water flow-rate that the pumping points can generate or the pipes can tolerate. We now describe three optimization problems that differ based on their design objectives.

MIN_DIFF: As the pumping points are limited, both in number as well as their pumping rates, the minimum velocity requirements of all the sensor nodes (Equation (4)) may not be met. Thus, *the objective of MIN_DIFF is to minimize the sum of the differences between the required velocity and the achieved velocity at all the pipe-sections where the sensor nodes are placed.* Let a_i, Q_j, R_j denote the nodal flows at node i, pipe discharges, and pipe resistances of pipe j, respectively. Let v_i and A_i denote the water velocity and area of the ith pipe, respectively, and $\mathbb{V}_j$ the minimum required velocity of the sensor node at pipe j. Let $\mathscr{V}_j^{\max}$ denote the maximum water velocity supported at pipe segment j, and let $a_i^{\max}$ denote the maximum flow supported at pumping point i. Let $\mathbb{C}$ and $\mathbb{L}$

denote the number of connection points and loops in the distribution system, respectively, $\mathbb{P}$ the set of all pipes, and $\mathbb{S}$ the subset where sensor nodes are placed. Then,

$$
\begin{aligned}
\textbf{Minimize} \quad & \sum_{j \in \mathbb{S}} \max(0, \mathbb{V}_j - v_j) \\
\text{subject to} \quad & v_j = \frac{|Q_j|}{A_j} \leq \mathcal{V}_j^{\max}, \quad \forall j \in \mathbb{P} \\
& \pm a_i \pm \sum_{\substack{\text{pipe } j \\ \text{connected to } i}} Q_j = 0, \quad \forall i = 1, 2, \ldots, \mathbb{C} \\
& \pm \sum_{\substack{\text{pipe } j \\ \in \mathbb{L}_l}} R_j Q_j^2 = 0, \quad \forall L_l = L_1, L_2, \ldots, L_{\mathbb{L}} \\
& |a_i| \leq a_i^{\max} \quad \forall i \in \mathbb{PP}.
\end{aligned} \tag{5}
$$

We have used the Darcy-Weisbach formula in Equation (5) to calculate the frictional head loss. Among the non-pumping points, if there exists some background flow at some connection point i, a_i is assigned to that background flow. Otherwise, a_i is assumed to be zero, at the non-pumping points. These background flows can be estimated from the water flow measurements reported by the sensor nodes.

In Equation (5), the first set of constraints ensures that the water velocity through a pipe segment j (which is equal to its volumetric flow rate/discharge divided by the cross-sectional area of the pipe segment) is less than the maximum water velocity $\mathcal{V}_j^{\max}$ that the pipe segment can support. The second set of constraints are the *node-flow continuity relationships* that ensure that the sum of the inflows and outflows at all connection points are zero. The third set of constraints are the *loop-head loss relationships* that state that the sum of head losses in pipes forming a loop is zero [56]. The ± sign is used in these two constraints to take into account the direction of the water flow assumed. The fourth constraint states that the water-flow rate at all the pumping point i is less than some maximum threshold $a_i^{\max}$. Note that when the objective value of **MIN_DIFF** is zero, the pumping points can satisfy the velocity requirements of all the sensor nodes.

MIN_PUMPING: If the velocity requirements of all sensor nodes can be satisfied, then we would like to do it in such way that the amount of water pumped is minimized. Note that MIN_PUMPING is solved only if the result of MIN_DIFF gives zero objective value. Given the notations and explanation of MIN_DIFF, the following formulation should be clear and is not explained further:

$$
\begin{aligned}
\textbf{Minimize} \quad & \sum_{i \in \mathbb{PP}} |a_i| \\
\text{subject to} \quad & v_j = \frac{|Q_j|}{A_j} \leq \mathcal{V}_j^{\max}, \quad \forall j \in \mathbb{P} \\
& v_k \geq \mathbb{V}_k, \quad \forall k \in \mathbb{S} \\
& \pm a_i \pm \sum_{\substack{\text{pipe } j \\ \text{connected to } i}} Q_j = 0, \quad \forall i = 1, 2, \ldots, \mathbb{C}
\end{aligned}
$$

$$\pm \sum_{\substack{\text{pipe } j \\ \in \mathbb{L}_l}} R_j Q_j^2 = 0, \quad \forall L_l = L_1, L_2, \ldots, L_{\mathbb{L}}$$
$$|a_i| \leq a_i^{\max} \quad \forall i \in \mathbb{PP}. \tag{6}$$

MIN_DISCHARGE: This problem is a variant of MIN_PUMPING problem where we assume the existence of a set of *discharge* (or deliberate leakage) points $\mathbb{LP}$ as well, from where water can be deliberately leaked at different regulated rates. As the number of pumping points is limited, the idea is to discharge/leak some amount of water (only for the charging time) to generate certain water-flow at few pipe-sections that keeps the sensor nodes running. Since the discharge wastes water, we want to minimize it. This can be modeled as follows:

$$\begin{aligned}
\textbf{Minimize} \quad & \sum_{i \in \mathbb{LP}} |a_i| \\
\text{subject to} \quad & v_j = \frac{|Q_j|}{A_j} \leq \mathscr{V}_j^{\max}, \quad \forall j \in \mathbb{P} \\
& v_k \geq \mathbb{V}_k, \quad \forall k \in \mathbb{S} \\
& \pm a_i \pm \sum_{\substack{\text{pipe } j \\ \text{connected to } i}} Q_j = 0, \quad \forall i = 1, 2, \ldots, \mathbb{C} \\
& \pm \sum_{\substack{\text{pipe } j \\ \in \mathbb{L}_l}} R_j Q_j^2 = 0, \quad \forall L_l = L_1, L_2, \ldots, L_{\mathbb{L}} \\
& |a_i| \leq a_i^{\max} \quad \forall i \in \mathbb{PP} \cup \mathbb{LP}.
\end{aligned} \tag{7}$$

In the above formulation, the zero objective value means that without any water discharge (waste), the pumping points can satisfy the node's demands. In that case the secondary objective of maintaining a minimum flow-rate at the pumping points, can be achieved by again solving the **MIN_PUMPING** problem as before.

We illustrate these problems using the simple example in Figure 4. Let us assume that the water is pumped in from reservoir 1 and pumped out from reservoir 2. Let nodes 3–6 be discharge points, with discharge rates of a_3–a_6. Thus, the *node-flow continuity relationships* and the *loop-head loss relationships* are as follows:

$$\left.\begin{array}{ll}
a_1 - Q_1 - Q_2 = 0 & Q_1 - a_2 - Q_3 = 0 \\
Q_3 - a_3 - Q_4 - Q_6 = 0 & Q_4 + Q_2 - a_4 - Q_5 = 0 \\
Q_5 + Q_7 - a_5 = 0 & Q_6 - a_6 - Q_7 = 0
\end{array}\right\} \quad \text{Node-flow continuity relationships,} \tag{8}$$

$$\left.\begin{array}{l}
R_2 Q_2^2 - R_4 Q_4^2 - R_3 Q_3^2 - R_1 Q_1^2 = 0 \\
R_5 Q_5^2 - R_7 Q_7^2 - R_6 Q_6^2 + R_4 Q_4^2 = 0
\end{array}\right\} \quad \text{Loop-head loss relationships.} \tag{9}$$

Equation (8) is simply the flow conservation law at each node in Figure 4. Equation (9) states that the loop-head loss in any loop must be zero.

We solve the above optimization problems for this case using **AMPL**, which is a modeling language for solving large-scale optimization problems [57]. The necessary parameters are listed in Table 2. The maximum pumping capacity $a^{\max}$ (assumed to be 0.069m^3/s in Table 2) corresponds to a flow rate of 7ft/s in a pipe with 8in diameter.

Figure 5 shows the variation of $\sum_{j \in \mathbb{P}} \max\left(0, \mathbb{V}_j - v_j\right)$ with different charging times. As expected, the objective value decreases with the increase in charging time and with reduced target

Table 2. Parameters Used

Var	Values	Var	Values	Var	Values
$a_1..a_6$	0	Fan diameter	2.5in	Pipe diameter	8in
$a^{\max}$	0.069m^3/s	$\mathscr{V}_j^{\max}$	7ft/s	$\mathscr{C}$	25F
V_{ini}	0.9V	V_{target}	1.5V	η_e	10%

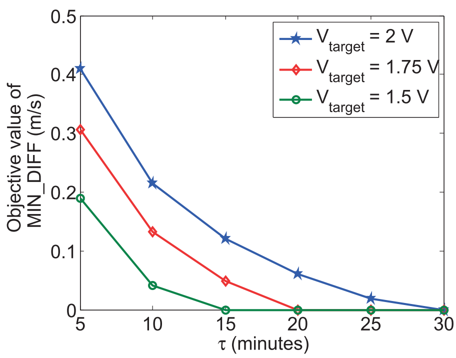


Fig. 5. Objective value of MIN_DIFF vs. charging time τ.

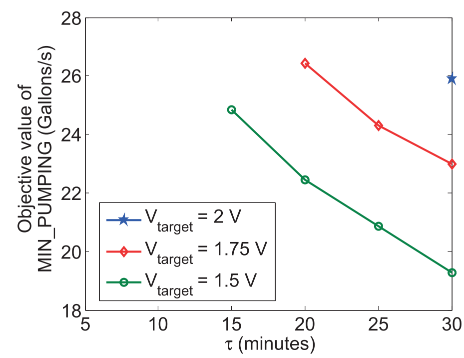


Fig. 6. Objective value of MIN_PUMPING vs. charging time τ.

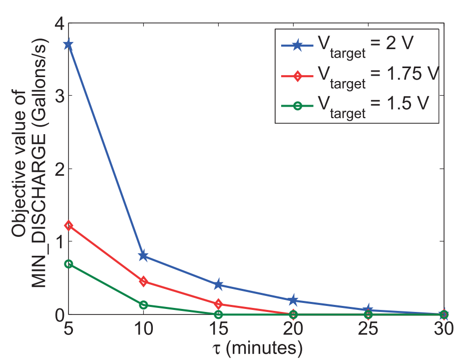


Fig. 7. Objective value of MIN_DISCHARGE vs. charging time τ.

voltage V_{target}. Figure 6 shows the total amount of water pumped in and out through the $\mathbb{PP}$s, with the variation of charging time. In Figure 6, initially the optimization problem MIN_PUMPING is infeasible, so there are no points in the graph. When the objective value of MIN_DIFF is zero, MIN_PUMPING starts giving feasible solutions, which is also a decreasing function of charging time.

Figure 7 shows the total amount of discharge with different charging times, where the maximum discharge rate is assumed to be same as the maximum pumping rates. Note that Figures 5 and 7 show a strong similarity, this is because whenever there is a non-zero difference between the required and achieved velocity in MIN_DIFF, there needs to be some non-zero discharges in MIN_DISCHARGE. Also the total discharges decrease with increasing charging time, as the sensor nodes get more time to replenish their super-capacitors.

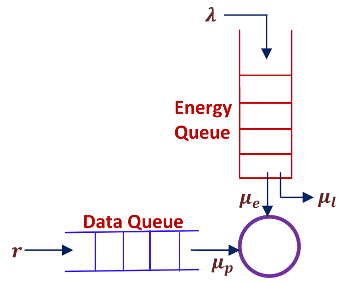


Fig. 8. Block diagram of an individual node's data queue and energy queue.

5 SAMPLING RATE ADAPTATION FOR LEAKAGE MONITORING

As discussed earlier, adaptation of the measurement/transmission activity to current state of the charge in super-capacitors is crucial for maintaining maximal coverage of the leakage/contamination detection activity. This is true even with artificial water circulation, since sensor nodes in certain segments of the water network may be difficult to charge effectively.

In general, a set of sensor nodes in a vicinity may have significant dependency with respect to water flows and hence their chargeability and leakage/contamination detection performance. We can consider these nodes as forming a *coalition* in the game theoretic sense, which can be exploited for improved performance. Coalition can be formed by simulating leaks/contaminations at different pipe sections, using any commercial simulator such as Water-GEMS [47] and by looking at the inter-dependencies among the detection abilities of the individual nodes, i.e., if there is a leak/contamination at any pipe section in a coalition, at least few sensor nodes are able to detect it. The coalition members can collaboratively adapt their sampling rates based on the individual node's energy availability, i.e., the low sampling rate of the sensor nodes with low harvested energy is compensated by the higher sampling rate of the nodes with higher energy. Such a mechanism is quite different from the individual node-based rate/energy allocation schemes discussed in the literature [58, 59] and is discussed in the following.

We assume that time is divided into *intervals* of T time units. The sampling rates are updated periodically in every interval as follows. All nodes keep track of their average harvested energy in each interval. Based on their historical energy profiles, they predict their expected harvested energy for the next interval, which is then used by the sensor nodes to calculate their maximum possible sampling rate as described in Section 5.2, which are used for the pressure or water-flow sensors. Unlike the leakage detection sensors (like pressure or water-flow sensors), the number of potential contaminants are numerous, as mentioned before. In such scenarios, the sensing devices are equipped with heterogeneous and multiple sensors. The collaborative adaptations of such heterogeneous, multi-sensor devices are discussed in Section 6. The maximum possible sampling rate is broadcast using *beacon* messages. The sink uses these rates to compute the optimal sampling rates of individual nodes and broadcasts by sending beacons, as described in Section 5.2.

5.1 Predicting Energy Harvesting

Figure 8 shows the conceptual model for energy harvesting. Available energy is stored in the *energy queue (EQ)*, which is a super-capacitor in our case. The sampled values are stored in a volatile RAM, which we call *data queue (DQ)*. While transmitting packets, a sensor node takes ℓ items from the RAM with $\ell_{\min} \leq \ell \leq \ell_{\max}$. Here, $\ell_{\min}$ is the minimum number of samples that a node will accumulate before transmitting if it has enough energy to do immediate transmission.

However, if the node is low on energy, then it will continue sampling and storing samples in the RAM (if possible). When the next energy burst arrives, it will transmit all accumulated samples up to the limit of $\ell_{\max}$.

Each sensor node estimates the energy arrival in its super-capacitor in periodic intervals of T using a normalized least mean square (NLMS) adaptive filter. In NLMS filter, historical harvested energy profile is stored in the vector E_{t-1}. Based on this profile, predicted harvested energy for the next interval λ_t is calculated by a dot product between E_{t-1} and the coefficients of the adaptive filter $\mathbb{W}_{t-1}$ using $\lambda_t = E_{t-1}\mathbb{W}_{t-1}$ and the error e_t is recorded. The filter coefficient is then modified as

$$\mathbb{W}_t = \mathbb{W}_{t-1} + \frac{s \cdot e_t \cdot E_{t-1}}{1 + |E_{t-1}|^2}, \tag{10}$$

where s is the step size of the filter. The super-capacitor leakage power and average power consumption due to different operations (sensing/transmission/reception etc) are assumed to be μ_l and μ_e, respectively. The average power consumption μ_e needs to be adapted based on the energy availability to maintain the energy conservation, i.e.,

$$A_e + \lambda_t - (\mu_e + \mu_l)T \geq 0 \quad \therefore \mu_e \leq \frac{A_e + \lambda_t}{T} - \mu_l, \tag{11}$$

where A_e is the amount of available energy at the beginning of that interval t.

The sampled values are stored in the DQ with an arrival rate of r, while the packet transmission rate is μ_p. Note that λ_t and μ_e are expressed in units of energy, whereas r is expressed in number of samples. We calculate the maximum sampling rate that the sensor node can support in the next time interval, without DQ buffer overflow. Assume that at the time of computing the maximum sampling rate, the number of packets waiting in the DQ is $\mathcal{N}$. The DQ capacity is assumed to be C. To maintain the energy budget, $\mu_e = \mathcal{A} \cdot r + \mathcal{B} \cdot \mu_p + \mathcal{C}$, where $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ are constants that capture the power consumption due to sensing, transmission and other operations (beacon transmission/reception, processing, etc.), respectively. To avoid DQ buffer overflow,

$$\begin{aligned} &\mathcal{N} + (r - \ell_m \cdot \mu_p) \cdot T \leq C \\ \therefore r \leq\ &\frac{C - \mathcal{N}}{T} + \ell_m \cdot \mu_p = \frac{C - \mathcal{N}}{T} + \ell_m \frac{\mu_e - \mathcal{A} \cdot r - \mathcal{C}}{\mathcal{B}} \\ \therefore r \leq\ &\frac{\frac{C-\mathcal{N}}{T} + \ell_m \frac{\mu_e - \mathcal{C}}{\mathcal{B}}}{1 + \ell_m \cdot \frac{\mathcal{A}}{\mathcal{B}}} = \mathbb{R} \quad \text{(assume)}, \end{aligned} \tag{12}$$

which gives the upper limit on the sampling rate. All sensor nodes periodically calculate their maximum sampling rate $\mathbb{R}$ and broadcast them in their beacon messages, which is used by the sink to determine the sampling rates of all the individual sensor nodes.

5.2 Computing Optimal Sampling Rate

Upon receiving the maximum sampling rate $\mathbb{R}$ from all the sensor nodes, the sink formulates the sampling rate adaptation problem to maximize a certain utility function, under the required energy constraints. Suppose that there are N nodes in a coalition. As the detection abilities of the sensor nodes in a coalition are highly correlated, the sensor nodes in a coalition can share the data sampling task among themselves for reduced energy consumption, based on their available harvested energy. We define the utility of a node i by considering two factors

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

- In a WDSN, main lines are generally more important than branch lines, as water from the main lines are distributed to different sub-mains and branches. Thus, a sensor node placed in a main line is considered to be more important than sensors placed in branches. Thus, we assign a relative weight α_i to the sampled data of node i, based on its position in the WDSN.

Beyond the distributional hierarchy, there may be other considerations in assigning the weights α_i, as determined by water system personnel. For example, the water pressure often varies significantly within a DMA, and nodes in higher pressure area can be given higher weights because of greater water loss and more potential damage due to leaks there. For contamination monitoring, one can assign weights to the nodes that are close to the reservoirs, because any contamination close to the reservoir needs to be detected sooner, to avoid its spread. Higher weights can also be assigned to older and more damage prone pipes. Also in a coalition, the detection abilities of certain sensor nodes may be higher compared to others, thus those nodes can be assigned higher weights.

Considering these factors, the **weighted proportional fairness** within a coalition can be achieved by modeling the utility function of node i as $U_i(r_i) = \alpha_i \cdot \log(r_i)$, where α_i is the normalized weight. Our objective is to maximize the overall utility of the coalition, i.e., $\sum_{i=1}^{N} U_i(r_i)$, after satisfying the energy budget of individual nodes. We also assume that the sink places an upper limit of M samples/interval from a coalition, to avoid redundant sampling, i.e., $\sum_{i=1}^{N} r_i \cdot T \leq M$, or $\sum_{i=1}^{N} r_i \leq \frac{M}{T} = \mathbb{M}$. Intuitively, we can think that the sensor nodes in a coalition work as a *single virtual sensor node* that senses and reports at a maximum rate of M samples/interval. M is basically a controlling parameter that controls the overall sampling rate of the coalition, i.e., if the sink wants to receive the samples more frequently, then it increases M and vice versa. Thus, the optimization problem can be written as

$$\begin{aligned} &\textbf{Maximize} \quad \sum_{i=1}^{N} U_i(r_i), \\ &\text{subject to} \quad \sum_{i=1}^{N} r_i \leq \mathbb{M}, \qquad r_i \leq \mathbb{R}_i, \forall i, \qquad r_i \geq 0, \forall i, \end{aligned} \tag{13}$$

where $r_i \leq \mathbb{R}_i$ is the maximum sampling rate constraint (MSRC) obtained from Equation (12). As log is a concave function, this problem is a convex optimization problem, that can be solved centrally by solving the corresponding Lagrangian and Karush-Kuhn-Tucker (KKT) conditions. We propose an algorithm to solve this problem, which is presented in the Section 5.3.

5.3 Proposed Rate Adaptation Scheme CARA

Based on the steps described above, we now describe our proposed *Collaborative and Adaptive Rate Allocation (CARA)* scheme, as shown in Algorithm 1. In this algorithm, the sink maintains two sets of nodes: unassigned U and assigned A. Initially, all nodes belong to set U, but are transferred to set A as rates are assigned for them. The sink first assigns the sampling rates to each sensor i as $r_i = \frac{\alpha_i}{\sum_{i \in U} \alpha_i} \mathbb{M}$ (lines 4–7). The difference between assigned sampling rate r_i and maximum sampling rate $\mathbb{R}_i$ is stored in diff[i]. After this sampling rate assignment, if the MSRC (obtained from Equation (12)) is violated for any node j, then diff[j] < 0. For those nodes, the sink assigns their rates as their maximum rate $\mathbb{R}_j$ (line 11) and divides the diff[j] fairly among other nodes (lines 14–17). This process is continued until the MSRC is satisfied for all the nodes. The calculated sampling rates are broadcasted to all the sensor nodes.

ALGORITHM 1: Collaborative Adaptive Rate Allocation scheme (CARA)

INPUT : Maximum sampling rate $\mathbb{R}_i$, utility weights α_i and $\mathbb{M}$.
OUTPUT : Sampling rates r_i $\forall i \in \{1, 2, \ldots, N\}$.
A = $\{\phi\}$; U = $\{1, \ldots, N\}$;
for each node $i = \{1, 2, \ldots, N\}$ **do**
    $r_i = \frac{\alpha_i}{\sum_{i \in U} \alpha_i} \mathbb{M}$;
    diff$[i] = \mathbb{R}_i - r_i$;
end for
for each node $k = \{1, 2, \ldots, N\}$ **do**
    Sort node $\in U$ in increasing order of diff$[k]$;
    Put them in order in list L;
    $j = L[0]$;
    if diff$[j] < 0$ **then**
        $r_j = \mathbb{R}_j$; $A = A \cup j$; $U = U \backslash j$;
        for each node $i = \{1, 2, \ldots, N\}$ AND $i \in U$ **do**
            $r_i = r_i + \frac{\alpha_i}{\sum_{i \in U} \alpha_i} \cdot$ abs(diff$[j]$);
            diff$[i] = \mathbb{R}_i - r_i$;
        end for
        diff$[j] = 0$;
    else
        EXIT
    end if
end for
return r_i $\forall i$

Optimality of the proposed scheme: The proposed scheme assigns the sampling rate fairly to all the nodes based on their weighted utilities considering the energy constraints. We prove that the proposed scheme is optimal under the given assumptions. We first propose and prove Lemma 1 and Lemma 2 as follows.

LEMMA 5.1. *The solution of the optimization problem*

$$\textbf{Maximize} \sum_{i=1}^{N} U_i(r_i) = \sum_{i=1}^{N} \alpha_i \cdot \log(r_i) \quad \text{subject to} \sum_{i=1}^{N} r_i \leq \mathbb{M},\ r_i \geq 0, \forall i \tag{14}$$

is $r_i = \frac{\alpha_i}{\sum_i \alpha_i} \mathbb{M}$, $\forall\ i = \{1, 2, \ldots, N\}$.

PROOF. Clearly r_i cannot be zero for any i. This is because making $r_i = 0$ makes the objective value $-\infty$. Thus, the last constraint is inactive. Then the Lagrangian and KKT conditions of problem Equation (14) are

$$L = \sum_{i=1}^{N} \alpha_i \cdot log(r_i) - \lambda \left(\sum_{i=1}^{N} r_i - \mathbb{M} \right), \tag{15}$$

$$\frac{\partial L}{\partial r_i} = \frac{\alpha_i}{r_i} - \lambda = 0, \tag{16}$$

$$\lambda \left(\sum_{i=1}^{N} r_i - \mathbb{M} \right) = 0. \tag{17}$$

Equation (16) gives $r_i = \frac{\alpha_i}{\lambda}$ and $\lambda \neq 0$. Putting this in Equation (17), we get $\lambda = \frac{\sum_i \alpha_i}{\mathbb{M}}$, which makes $r_i = \frac{\alpha_i}{\sum_i \alpha_i} \mathbb{M}$. □

Lemma 5.2. *If the sampling rate of a node is reduced by an amount δ from the optimal rate in problem Equation (14), and divided among others proportionately, then the overall objective function is a decreasing function of δ.*

Proof. Let us assume that for any node j, we assign an amount $\frac{\alpha_j}{\sum_i \alpha_i}\mathbb{M} - \delta$ and divide δ among all others proportionately, so that all nodes $i \neq j$ are assigned a rate of $\frac{\alpha_i}{\sum_i \alpha_i}\mathbb{M} + \frac{\alpha_i}{\sum_{i\neq j} \alpha_i}\delta$. Assume $\Delta = \sum_{i=1}^{N} \alpha_i$ and $\Gamma = \sum_{i\neq j} \alpha_i$. If the new objective function if $F(\cdot)$, then we can show that

$$\begin{aligned} F &= \alpha_j \log\left(\frac{\alpha_j}{\Gamma}\mathbb{M} - \delta\right) + \sum_{i\neq j} \alpha_i \log\left(\frac{\alpha_i}{\Gamma}\mathbb{M} + \frac{\alpha_i}{\Delta}\delta\right), \\ \frac{\partial F}{\partial \delta} &= -\frac{\alpha_j}{\frac{\alpha_j}{\Gamma}\mathbb{M} - \delta} + \sum_{i\neq j} \frac{\alpha_i}{\Delta\left(\frac{\mathbb{M}}{\Gamma} + \frac{\delta}{\Delta}\right)} < 0. \end{aligned} \tag{18}$$

Thus, $F(\cdot)$ is a strictly decreasing function of δ. □

Theorem 5.3. *The proposed CARA algorithm gives optimal solution for problem Equation (13).*

Proof. From Lemma 1, we get $r_i = \frac{\alpha_i}{\Gamma}\mathbb{M}$ for the optimization problem Equation (14). Now, we introduce the maximum sampling rate constraint (MSRC) $r_i \leq \mathbb{R}_i$ in problem Equation (14). Suppose $r_j = \frac{\alpha_j}{\Gamma}\mathbb{M}$ violates the MSRC of node j, i.e., $r_j > \mathbb{R}_j$ and diff$[j] = \mathbb{R}_j - r_j = \mathbb{R}_j - \frac{\alpha_j}{\Gamma}\mathbb{M}$.

We consider this problem in two steps. In the first step, we divide a total sampling rate of $\tilde{\mathbb{M}} = \frac{\Gamma}{\alpha_j}\mathbb{R}_j$ over N nodes. Then using Lemma 1, $r_j = \frac{\alpha_j}{\Gamma} \cdot \frac{\Gamma}{\alpha_j}\mathbb{R}_j = \mathbb{R}_j$. Thus, node j's MSRC is satisfied. For any other node i, $r_i = \frac{\alpha_i}{\Gamma} \cdot \frac{\Gamma}{\alpha_j} \cdot \mathbb{R}_i = \frac{\alpha_i}{\alpha_j}\mathbb{R}_j$. At this point, node j's utility cannot be improved any further by changing r_j (reducing r_j results in degradation of the overall objective as shown in Lemma 2).

Now, we introduce $\mathbb{M} - \tilde{\mathbb{M}}$ amount of additional sampling rates to this system. Clearly node j cannot be assigned more rates. Thus, we assign $\mathbb{M} - \tilde{\mathbb{M}}$ among all $i \neq j$ fairly. Using Lemma 1, the new rates of all $i \neq j$ are

$$\begin{aligned} r_i^{\text{new}} &= r_i + \frac{\alpha_i}{\Delta} \cdot (\mathbb{M} - \tilde{\mathbb{M}}) = \frac{\alpha_i}{\alpha_j}\mathbb{R}_j + \frac{\alpha_i}{\Delta} \cdot \mathbb{M} - \frac{\alpha_i}{\Delta} \cdot \frac{\Gamma}{\alpha_j} \cdot \mathbb{R}_j \\ &= \frac{\alpha_i}{\Gamma} \cdot \mathbb{M} + \frac{\alpha_i}{\Delta}\left(\frac{\alpha_j}{\Gamma} \cdot \mathbb{M} - \mathbb{R}_j\right) = r_i + \frac{\alpha_i}{\sum_{i\neq j} \alpha_i} \cdot \text{abs}\left(\text{diff}[j]\right), \end{aligned} \tag{19}$$

which is the same as the rate assigned according to Algorithm 1 (line 15). At this stage, if any other nodes violate the MSRC constraint, then we do similar operation till the MSRC constraint is fulfilled for all the nodes. Thus, the optimal sampling rates are obtained for all the nodes. □

6 HETEROGENEOUS MULTI-SENSOR CONTAMINATION SENSING AND ADAPTATION

Heterogeneous sensor networks are defined as networks where the wireless nodes are equipped with multiple, different types of sensors, such as audio, video, acceleration, and so on. Recently, wireless devices are increasingly being equipped with multi-modal sensors. Cellphones and Smartphones are perfect examples of wireless nodes with multiple sensors. These multi-modal sensor nodes are increasingly being used for disaster recovery applications [60, 61], earthquake monitoring [62, 63], infectious disease surveillance [64], and so on. In several such applications, the sensor readings are correlated—spatially, temporally, and across different types of sensors. One such example is an earthquake monitoring scenario, where the necessary sensors are acceleration, audio, cameras, etc., multiple of them can be integrated in a wireless device. An accelerometer, often coupled with velocity seismometers is used to measure and record the extent of ground motion or

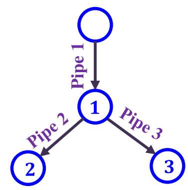


Fig. 9. Contamination detection in a water pipe network.

vibration. The audio samples can also be used to track the sound of a 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. Notice that in this example, the acceleration sensor is the *direct* sensor for detecting the earthquake, whereas the audio and cameras are proxy sensors. Such proxy sensors are less reliable, as the sound and pictures of a building collapsing can happen due to other disasters, such as cyclones, hurricanes, and so on.

In this article, we extend the concept of heterogeneous, multi-modal sensing in the context of contamination detection in water distribution networks. Our main objective is to explore the idea of the above *correlated* detection abilities of different sensors, to efficiently and collaboratively report contamination events, especially at the time of low energy hours at night. In a WDS, correlation among the sensors can result from two factors. First one is the *spatial* correlation, which results from the water distribution distribution topology and the water flows. As an example, let us consider Figure 9, which shows a Y junction, where the water comes through in pipe 1 and gets distributed to pipes 2 and 3. Assume that there are some sensors (e.g., rust, chlorine, pH, etc.) at each node. A contamination at pipe 1 is propagated to pipes 2 and 3, which obviates the need for sensors at node 1, due to spatial correlation created by water flow. The second factor is the *cross-sensor* correlation, which is the inherent dependencies among different types of sensors. As an example, chlorine and pH levels are correlated (or "compatible"), and for short durations it is possible to deduce the concentration of one from the other. As energy thriftiness is crucial in most sensing applications, the spatial and cross-sensor correlation would allow the sensors to be cycled on and off so that it is still possible to do the sensing effectively.

In such a multi-sensor environment, the collaborative sampling is useful in a heterogeneous WDSN in two ways. First, the cross-sensor correlation among the sensors need to be utilized to turn-off some of the sensor module, while still maintaining a reasonable coverage. Second, spatial and correlated detection abilities of different wireless devices at different junction points can be used to reduce the sampling rates of the energy critical nodes, which are then compensated by increasing the sampling rate of the high energy nodes.

6.1 Various Water Quality Indicators

EPA has defined 12 classes of potential water contaminants, which are reported in References [67, 68]. Table 3 depicts these contaminant classes along with their examples. As the number of potential contaminants is fairly large, 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 [67]. It is reported in Reference [67] that 10 of the 12 classes of contaminants can be detected by measuring three common surrogate parameters [68]: chlorine residual, conductivity, and total organic carbon (TOC).

Table 3. Different Contaminant Classes [65, 66]

Contaminant Class	Examples
Toxic Industrial Chemical	Cyanide
Toxic Inorganics	Arsenite
Pesticides	Oxamyl
Odorless Pesticides	Aldicarb
Chemical Warfare Agents	VX, G-type nerve agent, Potassium cyanide
Radionuclides	Alpha, Beta, and Gamma emitters
Bacterial Toxins	Botulinum toxins
Plant Toxins	Ricin
Waterborne Pathogens	*Vibrio cholerae, Salmonella typhi*
Bioterrorism Agents	*Bacillus anthracis, Escherichia coli*
Hydrocarbons	Gasoline

The above studies found that 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. These studies also indicate that the total organic carbon (TOC) in water is an important surrogate for detecting the presence of many organic compounds, with a sensitivity ranging from ~0.5mg/L to more than 1mg/L, depending on baseline levels and variability. To measure the TOC in water, absorbence of ultraviolet light at 254-nanometer wavelength is sometimes used [67]. This is because organic contents absorb ultraviolet light, and so measuring the ultraviolet absorbance provides an indication of organic concentration in water.

Other than chlorine and TOC, conductivity is also observed to respond slightly to some inorganic contaminants, and some metals, although the response is relatively weak compared to free chlorine residual and TOC. However, conductivity sensors have the potential of detecting some contaminants that do not trigger chlorine or TOC. Generally, the conductivity sensors respond to the contaminants at higher concentrations.

Beyond free chlorine residual, TOC, and conductivity, other water quality parameters are also sensitive to various contaminant classes. 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. However, it may be useful in understanding water contamination along with other measured parameters. These six parameters constitute the most common set of surrogates typically included in an water quality monitoring sensor network [67].

The two classes of contaminants that are not sensed by the above six parameters are chemical warfare agents and plant toxins. Volatile organic carbon (VOC) analyzers can be used to detect and identify specific compounds of such contaminants.

Definitions and assumptions: We assume that few wireless devices, equipped with multiple heterogeneous sensors are deployed in a WDS for the monitoring purpose. As multiple sensors are installed in a wireless device, we term the entire device as a *node*, whereas the word *sensor* is used to describe various sensors (chlorine, OPR, pH, etc.) attached to that node. We assume that the contaminants can enter into the system due to some leakage, or by deliberate means, through different pipe sections. A contamination *event* corresponds to a specific type of contaminant, which enters into the system through a particular pipe section. Thus, if c types of contaminants are considered in a l pipe WDS, then a total of $\mathcal{E} = c \cdot l$ contamination events is assumed in this

WDS. We consider a loop-free WDS where a contamination event propagate at the downstream connection points, based on the direction of the water flow. We neglect the response time of the individual sensors at different contaminant concentrations, which will be our future research focus.

In this article, we only consider contaminants that are largely *benign*, and are resulted due to slow seepage or build up in the pipes. Contaminants that are seriously life-threatening, such as highly toxic substances injected due to terrorism or accidental toxic chemical spills need to be monitored with sensors specifically designed for such contaminants, which sample at regular frequency, without considering the energy conservation. Determination of such highly toxic substances usually require stopping water supply and flushing out the entire system, which we do not consider in this article.

6.2 Proposed Adaptive Schemes

In general, a set of sensor nodes in a vicinity may have significant dependency with respect to their contamination detection capabilities. We can consider these nodes as forming a *coalition* in the game theoretic sense, which can be exploited for improved performance. Coalition can be formed by simulating contaminations at different pipe sections, using any commercial simulator such as Water-GEMS [47] and by looking at the inter-dependencies among the detection abilities of the individual nodes, i.e., if there is a contamination at any pipe section in a coalition, at least few sensor nodes are able to detect it. Also, any contamination within a coalition needs to be quickly propagated in the downstream direction, so that the downstream sensors can quickly respond to the contamination and report. The coalition members can collaboratively adapt their sampling rates, or can switch off some of their sensors, based on the individual node's energy availability, i.e., the low sampling rate of the sensor nodes with low harvested energy is compensated by the higher sampling rate of the nodes with higher energy.

We assume that time is divided into *intervals* of Φ time units. The sampling rates and the sensor's activities are updated periodically in every interval as follows. All nodes keep track of their average harvested energy in each interval. Based on their historical energy profiles, they predict their expected harvested energy for the next interval, using a normalized least mean square (NLMS) adaptive filter. The predicted energy availability for an interval, as well as the stored energy of the super-capacitor are broadcast by individual nodes using *beacon* messages. This is then used by the sink to adapt individual sensor's activity or their sampling rates. Below, we propose two versions of the sampling rate adaptation schemes. The first scheme, named *Heterogeneous Collaborative Sampling (HCS)*, tries to match the sampling rates of the nodes to match their predicted energy budget within an interval. We also propose another version of HCS, named *Advanced HCS (AHCS)*, that takes into account the super-capacitor storage capacity and the energy loss due to the lack of storage, in a smaller timescale. The notations used for the problem formulations are listed in Table 4.

6.3 Heterogeneous Collaborative Sampling (HCS)

Upon receiving the energy availability information from all the sensor nodes, the sink formulates the sampling rate adaptation problem to maximize a certain utility function, under the required energy constraints. Suppose that there are $\mathcal{N}$ nodes in a coalition, and a total of $\mathcal{S}$ various types of sensors. Notice that the detection abilities of different sensors may be correlated. As an example, in presence of Glyphosate, the chlorine, pH, and ORP sensor readings change simultaneously. Figure 10 shows a conceptual block diagram of the proxy sensing, where the vertices P_1, P_2, P_3 denote the types of contaminants (like Glyphosate, Dimethyl sulfoxide), and S_1, S_2, S_3, S_4 denote different sensors (like chlorine, pH, and ORP sensors) that sense the contaminant properties. The edges in between them are the *weights* that reflects how accurately a sensor senses a contaminants.

Table 4. Table of Notations

		Indices
i, j	$\triangleq$	Index for the sensors $(1, \ldots, \mathcal{S})$
k	$\triangleq$	Index for contamination events $(1, \ldots, \mathcal{E})$
m	$\triangleq$	Index for nodes $(1, \ldots, \mathcal{N})$
t	$\triangleq$	Index for sub-intervals $(1, \ldots, T)$
		Binary input variables
$x_i^m \in 0, 1$	$\triangleq$	Whether or not sensor i is attached to node m
		Other variables
Φ	$\triangleq$	Duration of an interval
δ_t	$\triangleq$	Duration of sub-interval t
c^i	$\triangleq$	Energy expenditure for transmitting a sample point by sensor i
e^i	$\triangleq$	Energy expenditure for sensing a sample point of sensor i
$\mathcal{A}^m$	$\triangleq$	Estimated energy arrival within an interval of node m
$\mathcal{A}_t^m$	$\triangleq$	Estimated energy arrival at node m within sub-interval t
O^m	$\triangleq$	Energy expenditure within an interval of node m
O_t^m	$\triangleq$	Energy expenditure at node m within sub-interval t
E^m	$\triangleq$	Stored energy at the super-capacitor of node m
τ	$\triangleq$	Minimum energy threshold
Υ_t^m	$\triangleq$	Energy allocation Υ_t^m of sensor node m at sub-interval t
$\mathcal{S}_t^m$	$\triangleq$	Remaining energy of the super-capacitor of node m at sub-interval t
Υ_t^m	$\triangleq$	Energy budget of node m at sub-interval t
$\mathbb{C}^m$	$\triangleq$	Maximum capacity of the super-capacitor
p_i^{km}	$\triangleq$	Probability that a contamination event k can be inferred from the readings of sensor i of node m
r_i^m	$\triangleq$	Sampling rate of sensor i of node m
r_{ti}^m	$\triangleq$	Sampling rate of sensor i of node m at sub-interval t
L_t^m	$\triangleq$	Remaining energy in the super-capacitor of node m at the end of sub-interval t
R_m, R_M	$\triangleq$	Minimum and maximum sampling rate allowed in any sensor

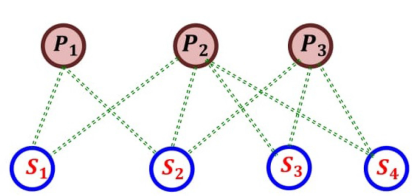


Fig. 10. A conceptual overview of proxy sensing.

However, some of the nodes within a coalition, in the downstream direction of a contamination event can respond due to the contamination propagation. As the detection abilities of the sensor nodes in a coalition are highly correlated, the sensor nodes in a coalition can share the data sampling task among themselves for reduced energy consumption, based on their available harvested energy. We define the overall utility of reporting a contamination event k by considering the following factors:

- The rate at which the sensors sample and report, i.e., $r_i^m\ \forall\ i, m$. As the sensing rate increases, the overall detection ability increases.
- Their corresponding contamination detection probabilities, which is represented as $p_i^{km}\ \forall\ i, m, k$. If a sensor shows significant deviation due to a contamination event k, then increasing its sampling rate enhances the contribution to the utility function of event k. p_i^{km} is considered to be zero if (a) node m is not in the downstream of event k, or (b) the approximate water propagation time from the contamination point to node m is more than some threshold. Condition (b) is important, because a contamination event needs to be detected within a reasonable amount of time.

Considering these two factors, the effective rate at which an event k is reported by the sensors is given by $e_k = \sum_{i=1}^{\mathcal{S}} \sum_{m=1}^{\mathcal{N}} p_i^{km} \cdot r_i^m$. Thus, the fair event reporting ability is ensured by modeling the utility of event k as $U_k(e_k) = \log(e_k)$. Our objective is to maximize the overall event reporting capability, i.e., $\sum_{k=1}^{\mathcal{E}} U_k(e_k)$, after satisfying the energy budget of the individual nodes. Thus, the overall optimization problem can be written as

$$\begin{aligned} \textbf{Maximize} \quad & \sum_{k=1}^{\mathcal{E}} \log\left(\sum_{i=1}^{\mathcal{S}} \sum_{m=1}^{\mathcal{N}} p_i^{km} \cdot r_i^m\right), \\ \text{subject to} \quad & \sum_{i=1}^{\mathcal{S}} r_i^m (e^i + c^i) \leq \frac{E^m + \mathcal{A}^m - O^m - \tau}{\Phi} = \mathbb{E}^m \quad \forall m \\ & R_m \cdot x_i^m \leq r_i^m \leq R_M \cdot x_i^m \quad \forall i, \forall m, \end{aligned} \tag{20}$$

where $\mathcal{A}^m$ and O^m are estimated energy arrival and energy expenditure within an interval of node m, respectively. E^m is the stored energy at the super-capacitor of node m at the beginning of the interval. The first set of constraints state that the power consumption for event reporting at any node is less than its energy budget. All the nodes try to maintain a minimum energy threshold, which is assumed to be τ. Φ denotes the duration of an interval. The second set of constraints says that r_i^m is non-zero only if node m is equipped with sensor i. This set of constraints also bound the minimum and maximum sampling rate of a sensor to be R_m and R_M, respectively.

6.4 Proposed Rate Adaptation Scheme for HCS

However, solving problem Equation (20) is challenging because of the non-concavity of the objective function. Although log is a strictly concave function with respect to the variables $p_i^{km} \cdot r_i^t$, the objective function is non-strictly concave function because of the term $\sum_{i=1}^{\mathcal{S}} \sum_{t=1}^{\mathcal{T}} p_i^{km} \cdot r_i^m$. To cope with this, we adopt the scheme similar to Reference [69] and is described as follows.

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

$$\begin{aligned} & \log\left(\sum_{i=1}^{\mathcal{S}} \sum_{m=1}^{\mathcal{N}} p_i^{km} \cdot r_i^m\right) \geq \sum_{i=1}^{\mathcal{S}} \sum_{m=1}^{\mathcal{N}} \theta_i^{km} \log\left(\frac{p_i^{km} \cdot r_i^m}{\theta_i^{km}}\right) \quad \forall i, \forall m, \forall k, \\ & \text{where} \quad \theta_i^{km} = \frac{p_i^{km} \cdot r_i^m}{\sum_{i=1}^{\mathcal{S}} \sum_{m=1}^{\mathcal{N}} p_i^{km} \cdot r_i^m} \quad \forall i, \forall m, \forall k. \end{aligned} \tag{21}$$

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

Modified Optimization Problem (MOP):

$$
\begin{aligned}
\text{Maximize} \quad & U = \sum_{k=1}^{\mathcal{E}} \sum_{i=1}^{\mathcal{S}} \sum_{m=1}^{\mathcal{N}} \theta_i^{km} \log\left(\frac{p_i^{km} \cdot r_i^m}{\theta_i^{km}}\right), \\
\text{subject to} \quad & \sum_i r_i^m (e^i + c^i) \leq \mathbb{E}^m \quad \forall m \\
& R_m \cdot x_i^m \leq r_i^m \leq R_M \cdot x_i^m \quad \forall i, \forall j, \forall m.
\end{aligned}
\tag{22}
$$

Later on in Theorem 6.2, we prove that solving **MOP** in Equation (22) is equivalent to solving **OP** in Equation (20).

The **MOP** is strictly concave, for a given θ_i^{km} and thus can be solved using Algorithm 2. In this scheme, the nodes first assign the sampling rates to each sensor i as $r_i^m = \frac{\mathbb{E}^m \cdot \sum_k \theta_i^{km}}{\sum_i \sum_k \theta_i^{km}(e^i + c^i)}$ (line 3). If the sampling rates are less or more than the specified thresholds R_m and R_M, then the node divides the remaining energy Δ fairly among other sensors (line 5–25). For doing this the nodes initialize an empty set V (line 4). If the sampling rate of a sensor is less than R_m, then it changes its sampling rate to R_m, includes 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. The same procedure is applied when the sampling rates are more then R_M for any sensor.

ALGORITHM 2: Proposed Rate Adaptation scheme for problem Equation (22)

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

THEOREM 6.1. *For a given θ_i^{km}, Algorithm 2 gives optimal rate allocation of the node sensors.*

Proof. Line 3 can be derived by solving the Lagrangian and KKT conditions of problem Equation (22) (ignoring the last set of constraints), which are as follows:

$$L = \sum_{k=1}^{\mathcal{E}} \sum_{i=1}^{\mathcal{S}} \sum_{m=1}^{\mathcal{N}} \theta_i^{km} \log\left(\frac{p_i^{km} \cdot r_i^m}{\theta_i^{km}}\right) - \sum_{m=1}^{\mathcal{N}} \lambda_m \left(\sum_i r_i^m (e^i + c^i) - \mathbb{E}^m\right),$$

$$\frac{\partial L}{\partial r_i^m} = \frac{\sum_{k=1}^{\mathcal{E}} \theta_i^{km}}{r_i^m} - \lambda_m (e^i + c^i) = 0, \tag{23}$$

$$\lambda_m \left(\sum_i r_i^m (e^i + c^i) - \mathbb{E}^m\right) = 0. \tag{24}$$

Equation (23) gives $r_i^m = \frac{\sum_{k=1}^{\mathcal{E}} \theta_i^{km}}{\lambda_m (e^i + c^i)}$ and $\lambda_m \neq 0$. Putting this in Equation (24), we get $\lambda_m = \frac{\sum_{i=1}^{\mathcal{S}} \sum_{k=1}^{\mathcal{E}} \theta_i^{km}}{\mathbb{E}^m}$, which makes $r_i^m = \frac{\mathbb{E}^m \cdot \sum_k \theta_i^{km}}{\sum_{i=1}^{\mathcal{S}} \sum_{k=1}^{\mathcal{E}} \theta_i^{km} (e^i + c^i)}$. □

The overall scheme flows as follows. The sink node first solves problem Equation (22) using any random θ_i^{tk} (at the first iteration) as described in Algorithm 2. It then calculates (a) the total weighted rate $\mathbb{TWR}^k = \sum_\ell \mathbb{WR}_\ell^k = \sum_i \sum_m p_i^{km} \cdot r_i^m$, (b) then calculates $\theta_i^{km} = \frac{p_i^{km} \cdot r_i^m}{\mathbb{TWR}^k}$ (Equation (21)), and (c) solves its optimization problem Equation (22) using the new θ_i^{km}. This process goes on until the solution converges. Upon convergence the calculated sampling rates are sent to the individual sensor nodes. The overall scheme is shown in Algorithm 3.

ALGORITHM 3: Heterogeneous Collaborative Sampling (HCS)

```
1: INPUT : x_i^m, p_i^{km}, P_i^m, R_m, R_M.
2: OUTPUT : Sampling rates r_i^{km} ∀i ∈ {1, 2, ..., S}.
3: while not converged do
4:     for each sensor i do
5:         Update the sampling rate r_i^m and θ_i^{km};
6:         Calculate WR_ℓ^k = Σ_i Σ_t p_i^{km} · r_i^m ∀k;
7:     end for
8:     Calculates TWR^k = Σ_ℓ WR_ℓ^k;
9: end while
```

Theorem 6.2. *The proposed version of* ***MOP*** *converges to the optimal solution of the original problem* ***OP****.*

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

$$\begin{aligned} &\frac{\partial}{\partial r_i^m}\left(\sum_{k=1}^{\mathcal{E}} \sum_{i=1}^{\mathcal{S}} \sum_{m=1}^{\mathcal{N}} \theta_i^{km} \log\left(\frac{p_i^{km} \cdot r_i^m}{\theta_i^{km}}\right)\right)\Bigg|_{\boldsymbol{r}^*} - \lambda_m^* (e^i + c^i) = 0, \\ &\lambda_m^* \left(\sum_{i=1}^{\mathcal{S}} r_i^{m^*} (e^i + c^i) - \mathbb{E}^m\right) = 0. \\ &\lambda_m^* \geq 0 \end{aligned} \tag{25}$$

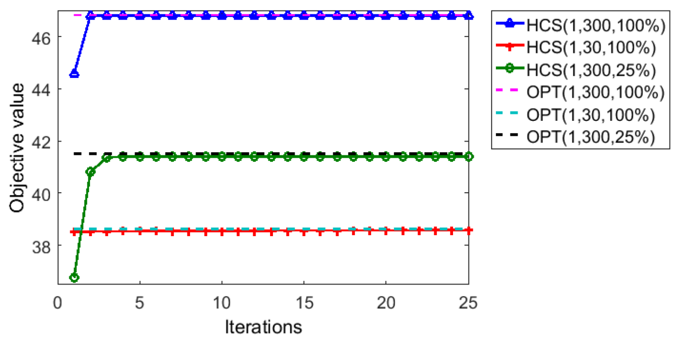


Fig. 11. Convergence of HCS with different R_m, R_M and super-capacitor charges. The numbers within the braces are R_m, R_M and supercapacitor charge levels, respectively.

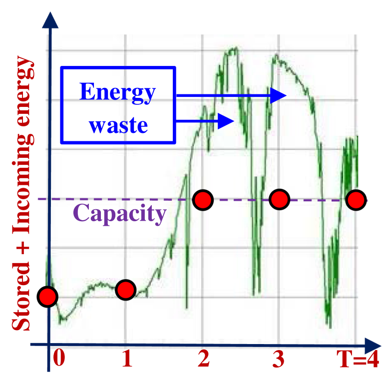


Fig. 12. Energy wastage due to lack of storage capacity. Red dots are the actual energy.

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

$$\frac{\partial}{\partial r_i^m}\left(\sum_{k=1}^{\mathcal{E}}\sum_{i=1}^{\mathcal{S}}\sum_{t=1}^{\mathcal{T}}\theta_i^{km}\log\left(\frac{p_i^{km}\cdot r_i^m}{\theta_i^{km}}\right)\right)\Bigg|_{r^*} = \frac{\partial}{\partial r_i^m}\left(\sum_{k=1}^{\mathcal{E}}\log\left(\sum_{i=1}^{\mathcal{S}}\sum_{m=1}^{\mathcal{N}}p_i^{km}\cdot r_i^m\right)\right)\Bigg|_{r^*}, \tag{26}$$

the first condition of **OP** and **MOP** are also identical at point $(\boldsymbol{r}^*, \boldsymbol{\lambda}^*)$. Thus, the proof follows. □

Figure 11 shows the convergence of Algorithm 3, where we assume that the super-capacitor capacity of the devices are assumed to be 5000mAh. The nodes consume 2.8mJ for transmission. Assume that all the nodes use three sensors, with current consumption of 9.5, 150, and 7.5mA, respectively, and sampling time of 7,000, 400, and 112ms. The nodes are expected to remain active for 12 months, and the power budgets for sensing and forwarding are calculated accordingly. $\mathcal{E}$ is assumed to be 5 and $p_i^{km} = 0.5\ \forall i, k, m$. We assume 75 devices for Figure 11. From Figure 11, we can observe that the objective values obtained from Algorithm 3 match with the optimal solution obtained from AMPL solver [57].

6.5 Advanced HCS

Notice that the optimization problem formulation Equation (20) does not consider the super-capacitor storage capacity into account. In a situation where the cumulative sum of the stored and incoming harvested energy is more than the maximum capacity $\mathbb{C}^m$ of the super-capacitor, the energy is not stored and are lost, as shown in Figure 12. Thus, the optimization problem

Equation (20) overestimates the sampling rates of the sensor. If this loss factor is not taken into account, then the nodes will die faster based on the assigned sampling rates, which drastically deteriorates their event reporting capabilities at the low energy hours. To alleviate this problem, we improve the formulation of problem Equation (20) by dividing an interval into smaller sub-intervals, and taking into consideration the energy arrival in each sub-interval. The improved version of this problem formulation, called AHCS, can be written as follows:

$$\begin{aligned} \textbf{Maximize} \quad & \sum_{t=1}^{T}\sum_{k=1}^{\mathcal{E}} \log\left(\sum_{i=1}^{\mathcal{S}}\sum_{m=1}^{\mathcal{N}} p_i^{km} \cdot r_{ti}^m\right), \\ \text{subject to} \quad & L_0^m = E^m \quad \forall m, \\ & L_t^m = \min\left\{\mathbb{C}^m, L_{t-1}^m + \mathcal{A}_t^m - \sum_{i=1}^{\mathcal{S}} r_{ti}^m (e^i + c^i)\delta_t - O_t^m\right\} \quad \forall m, \forall t, \\ & L_t^m \geq \tau \quad \forall m, \forall t, \\ & R_m \cdot x_i^m \leq r_i^m \leq R_M \cdot x_i^m \quad \forall i, \forall m, \end{aligned} \tag{27}$$

where r_{ti}^m is the sampling rate of sensor i of node m at time sub-interval t. Each sub-intervals is assumed to be of δ_t time units. The first constraint states that the initial energy at any node, at the beginning of an interval is E^m. The second set of constraints ensures that the remaining energy after every sub-interval does not go beyond the super-capacitor capacity $\mathbb{C}^m$. The third constraint ensures that the remaining energy at the end of any sub-interval is more than a threshold τ.

For solving Problem Equation (27), we need to understand two situations that we want to avoid. First is the scenario where a sensor node exhausts all its energy at any sub-interval and die. This may be a result of aggressive energy expenditure at the previous sub-intervals, which needs to be avoided. Second is the scenario where the super-capacitor of the sensor node reaches its maximum level at any sub-interval and thus miss the recharging opportunity. This is a result of conservative energy usage at the previous sub-intervals. To resolve these two situations, we adopt the energy allocation scheme proposed in Reference [70], which allocates the energy budgets of the sub-intervals.

Assume that π_t^m is the average energy arrival in T sub-intervals, i.e., $\pi_t^m = \sum_{t=1}^{T} \mathcal{A}_t^m / T$. Then the energy allocation Υ_t^m of sensor node m at sub-interval t can be expressed as

$$\Upsilon_t^m = (1 - \Omega^m)\pi_t^m + \Omega^m \mathcal{A}_t^m, \tag{28}$$

where $\Omega^m \in (0, 1)$ is the weight to regulate the energy allocation at any sub-interval. Ω^m can be obtained from Algorithm 4 where $\mathcal{K} > 0$ is assumed to be a small constant. In Algorithm 4 assume that $\mathcal{S}_t^m$ is the remaining energy of the super-capacitor of node m at sub-interval t. Thus, $\mathcal{S}_{t+1}^m = \min(\mathcal{S}_t^m + \mathcal{A}_t^m - \Upsilon_t^m, \mathbb{C}^m)$. In Algorithm 4 the value of Ω^m is updated repeatedly until the stopping criteria is reached (line 12). Algorithm 4 is designed based on the following intuition: if the capacity of the super-capacitor is sufficient, $\max_{\forall t}\{\frac{O_t^m}{\mathcal{A}_t^m}\}$ is negative for all sub-intervals, which results in $\Omega^m = 0$. This is the second stopping criteria (line 12), which states that if the super-capacitor is large enough to store all the harvested energy at any sub-interval, then the optimal energy budget is equal to π_t^m at any sub-interval. However, when the capacity is deficient, $\max_{\forall t}\{\frac{O_t^m}{\mathcal{A}_t^m}\}$ will be positive, which will increase Ω^m and thus the loop continues (lines 5–12). $\max_{\forall t}\{\frac{O_t^m}{\mathcal{A}_t^m}\}$ is zero when the super-capacitor capacity is just large enough so that there is neither an energy excess nor any deficiency at some sub-intervals, which is the first stopping criteria (line 12).

After finding Ω^m from Algorithm 4, we then can calculate the energy budget at the sub-intervals Υ_t^m from the obtained Ω^m. When then use this energy budget for solving problem Equation (20) at any sub-interval.

ALGORITHM 4: Finding Ω^m in Advanced HCS

1: INPUT : $\mathcal{A}_t^m, E^m, \mathbb{C}^m$.
2: OUTPUT : $\Omega^m \; \forall m \in \{1, 2, \ldots, \mathcal{N}\}$.
3: $\mathcal{A}_1^m = \mathcal{A}_1^m + E^m$; //Accounting initial energy at the first sub-interval
4: $\pi_t^m = \sum_{t=1}^{T} \mathcal{A}_t^m / T$;
5: **repeat**
6: **for** $t = 1, 2, \ldots, T$ **do**
7: $\Upsilon_t^m = \left[(1 - \Omega^m)\pi_t^m + \Omega^m \mathcal{A}_t^m\right]_0^{\mathcal{S}_t^m + \mathcal{A}_t^m}$;
8: $O_t^m = \mathcal{S}_t^m + \mathcal{A}_t^m - \Upsilon_t^m - \mathbb{C}^m - \tau$;
9: $\mathcal{S}_{t+1}^m = \min\left(\mathcal{S}_t^m + \mathcal{A}_t^m - \Upsilon_t^m, \mathbb{C}^m\right)$;
10: **end for**
11: $\Omega^m = \left[\Omega^m + \mathcal{K} \times \max_{\forall t}\left\{\frac{O_t^m}{\mathcal{A}_t^m}\right\}\right]_0^1$;
12: **until** $\left(\max_{\forall t} O_t^m == 0\right)$ OR $\left(\max_{\forall t} O_t^m < 0 \text{ while } \Omega^m = 0\right)$;
13: return Ω^m;

Post-detection measures: After detecting the presence of any contaminants, the WDS administrator may direct all the nodes to start sampling at higher rates to know the level of that contamination throughout the network and can take necessary measures based on type and spread of the contamination. The energy due to extra sampling can then be compensated by some artificial water flow mechanisms as proposed in Section 4. In this article, we only address the collaborative event reporting scheme considering the energy budgets of the nodes, whereas the specific post-detection steps should be taken by the WDS operators, and so it is not within the scope of this article.

7 SIMULATION RESULTS

Ideally, the evaluation of the scheme should be done with a real water distribution network, however, this is simply not possible in practice. Water distribution companies are generally not even willing to share the data they already collect, much less providing access to their distribution systems. There are real challenges in putting together a realistic network in the lab (e.g., access to large volume water supply, reservoirs, energy harvesters, etc.). As a result, the evaluation in this article is largely based on simulations that account for the water flow physics [56] and use parameters obtained from characterization of real water distribution systems.

We study the proposed rate adaptation scheme in *Castalia* [71], which is an application-level simulator for wireless sensor network based on OMNeT++. The simulated system topology along with the pipe diameters are shown in Figure 13. Water from the reservoir comes to nodes 1 and 2 (first-level nodes), distributed to nodes 3–6 (second level nodes), and then to 7–14 (third-level nodes). Each node has the fan for energy harvesting, a super-capacitor, water sensors, a small computer, and WiFi radio. The cross-sectional area of the fans are chosen as $\frac{1}{16}$th of the pipe cross section, to avoid blocking the normal water flow. For simulations, τ is assumed to be 1min, which is much less than in Section 4. The difference can be attributed to the fact that we now have successively smaller pipe diameters (going from levels 1 to 2 to 3), which increases water velocity and helps with respect to charging of nodes. Due to this structure, it is also reasonable to assume that all nodes fall into a single coalition.

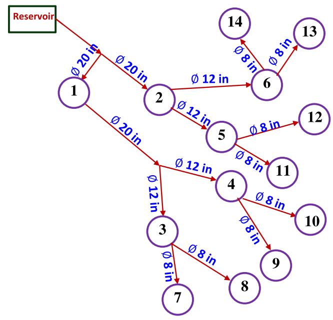


Fig. 13. Simulation topology; red arrows show the direction of water-flow.

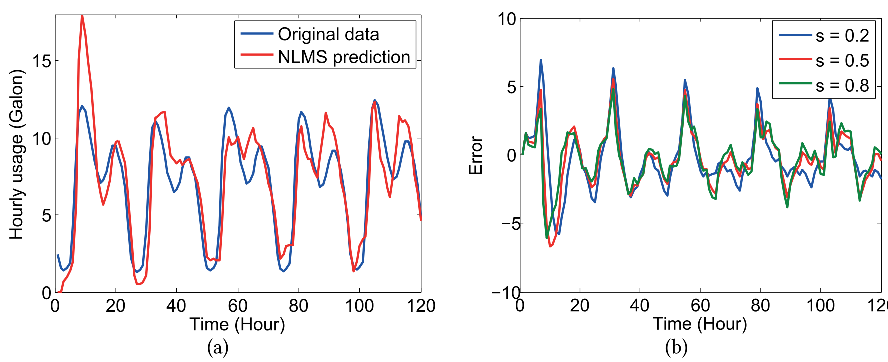


Fig. 14. (a) Hourly water usage of a single-family home for five days ($s = 0.8$). (b) Convergence of the forward predictor with different s.

We model the harvested energy arrival from water-flow based on the average water usage pattern, taken from Reference [72], and shown in Figure 14(a) for a typical single-family home over five days. The total daily usage is 169 ± 10.6 gallons. Reference [73] reports the maximum water velocity in real systems as 7.5ft/s. We conservatively assume that for the third level nodes have a water-velocity of 5.0ft/s at peak hours and compute those for other two layers using the flow continuity relationships. We also calculate water velocities and the *available energy* at other times based on the usage pattern and variation. Figure 14(b) shows the NLMS (normalized least mean square) filter predictor of the *available energy* for five days. We use s as 0.8.

The sink node broadcasts the assigned rates every $T = 1$h (interval time), chosen such that the harvested energy does not change significantly within the interval time. The beacon interval of the sensor nodes is assumed to be 30min. We assume that the nodes use asynchronous Low Power Listening that makes them sleep most of the time and wake-up periodically to check the channel activity. The power consumption in each node is represented as [74]

$$P_{\text{node}} = \frac{P_{Bt}T_{Bt}}{T_B} + \mathscr{M} \cdot P_{Dt}T_{Dt} + \mathscr{N} \cdot P_{Br}T_{Br} + \mathscr{S} \cdot P_sT_s + \mathscr{P} \cdot P_PT_P, \tag{29}$$

where P_x and T_x represent the power consumption and the duration, respectively, of the event x; and T_B represents the beacon interval. Transmission/reception of beacons is denoted by B_t/B_r, data transmit/receive is denoted by D_t/D_r, and processing and sensing are denoted as P and S, respectively. $\mathscr{M}$, $\mathscr{N}$, and $\mathscr{S}$ are the number of data transmission, beacon reception, and data sampling, respectively. $\mathscr{P}$ represents the number of times that a node wakes-up per second to check if the channel is busy, and is set to 8 in our application. Because of this reason the preamble

Table 5. Simulation Parameters [75–78]

Var	Values	Var	Values	Var	Values	Var	Values
P_{Bt}	1,000mW	T_{Bt}	140ms	P_{Br}	200mW	T_{Br}	140ms
P_{Dt}	1,000mW	T_{Dt}	140ms	P_{Dr}	200mW	T_{Dr}	140ms
P_P	200mW	T_P	3ms	P_S	500mW	T_S	400ms

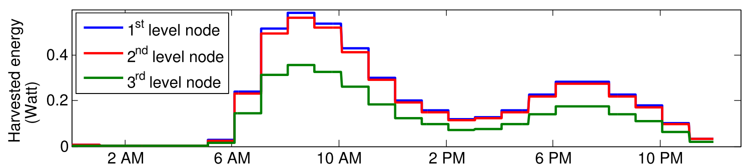


Fig. 15. Mean energy harvested over time for different nodes.

length will be atleast $\frac{1}{\mathscr{P}}$ = 125ms while using LPL. Based on this preamble length, the values of T_{Bt}, T_{Br}, T_{Dt} are set to ~140ms (i.e., preamble length + the transmission time for data/beacon packet). For the radio communication from the sensor nodes to the sink, we assume using low-power WiFi [75] or long-range Zigbee (such as XBee-PRO [76]) having power consumption of ~1000 mW in transmit mode and ~200mW in receive mode.

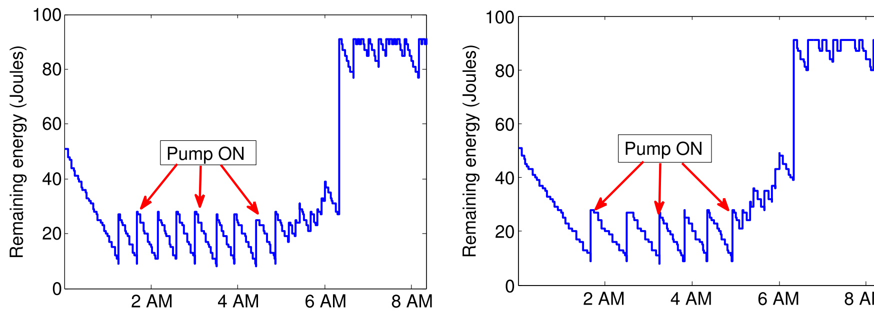
We assume DQ capacity as 20 samples, $\ell_{\min}$, $\ell_{\max}$ as 3, 5, respectively, $\mathbb{M}$ as 600pkts/h, and harvesting efficiency η_e = 10%. In reality η_e itself is dependent on flow velocity and load, but for simplicity, we keep it fixed at 10% for our simulations. The super-capacitor is assumed to be of 25Farad @2.7V with an initial voltage of 2.0V for all nodes. The super-capacitor leakage power is calculated as $P_0 \cdot \exp(a \cdot V_c)$ [6], where V_c is the super-capacitor voltage and P_0 and a are constants obtained from best-fitting the experimentally obtained results, and are $P_0 = 2.572e^{-17}$ and $a = 11.982$, respectively. The DC-DC converter efficiency (in between the super-capacitor and the sensor node) is assumed to be 75% [6].

We assume α at levels 1, 2, and 3 in the ratio 4, 2, and 1 to reflect the fact that detection at higher levels of the distribution network is much more important than at lower levels. We use the following policy for water circulation: if 50% of the nodes go below the threshold voltage of V_{thresh} = 0.9V, water is pumped in through the pipes of nodes 1–2 to boost the node voltages to V_{target} = 1.5V (or higher) at all nodes. We model both our scheme, i.e., CARA, and the simpler non-adaptive scheme called *equal rate allocation (ERA)*, which assigns the same sampling rate to all nodes. We use nodes 1, 3, and 7 to show the characteristics of first, second and third level nodes, respectively. We run the simulation for 24h. Parameters used for simulations are listed in Table 5.

Figure 15 shows the mean energy profile of nodes from normal water flow over 24h at levels 1, 2, and 3. To model fine-grain harvesting variations due to water flow turbulence, the actual energy arrival is modeled as uniformly distributed around the mean. The energy harvested depends on the water velocity and fan diameters. In this example, the water velocity increases at lower levels but the fan diameter decreases, thereby resulting in the behavior shown.

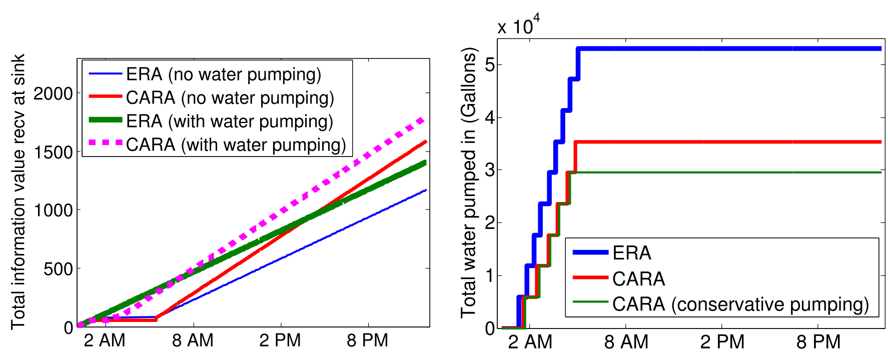
7.1 Benefits of Energy Adaptation

We compare our rate adaptation scheme with ERA in Figures 16(a) and 16(b), which show the remaining energy of node 7 during the crucial night hours. It is seen that without adaptation, the pump is on frequently, because the sensor node continues to sample fast and dies more often. However, with adaptation, both the sampling rate and hence the pumping rate slow down. In particular, CARA effectively reduces the pumping frequency by 33% as shown in Figure 17(b).



(a) (b)

Fig. 16. Remaining energy over time for (a) ERA scheme, (b) CARA scheme.



(a) (b)

Fig. 17. (a) Comparison of total effective information received at the sink. (b) Comparison of total water pumped into the system.

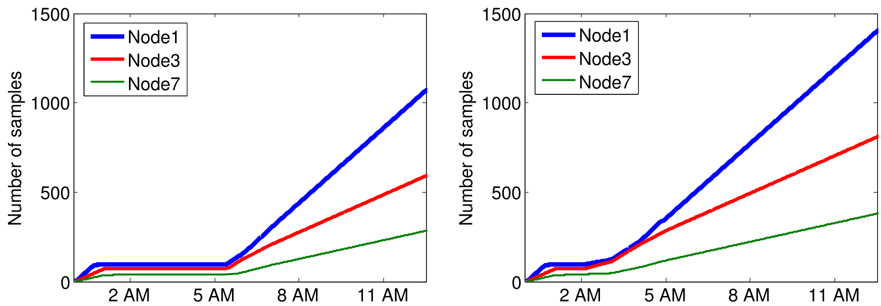
Also by dynamically adapting the sampling rates to maximize the system utility, CARA achieves 35% of higher *information* measure (defined as the product of the number of samples and their weights) without any water circulation and 30% in presence of artificial circulation, compared to ERA as seen from Figure 17(a).

CARA can further reduce the pumping frequency by exploiting the mutually inter-dependent detection abilities among the members of a coalition. In particular, Figure 17(b) also shows the pumped water amount for CARA under the policy that pumping is done only when all sensor nodes within a coalition die. This policy reduces pumping frequency by another 20%.

7.2 Benefits of Artificial Water Circulation

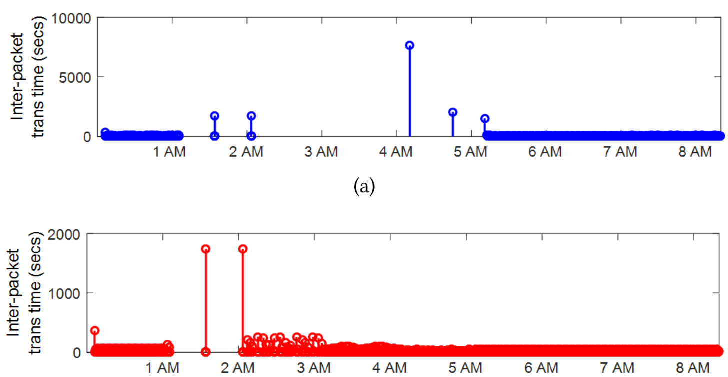
Figures 18(a) and 18(b) show the sampling schedules of nodes 1, 3, and 7 without and with artificial water circulation. It is seen that the artificial circulation at the night drastically improves the monitoring. It is also seen that the sampling rate of the higher level nodes are higher because of higher energy availability and higher chosen utility weight (i.e., α). This clearly shows the adaptive nature of CARA based on the individual nodes relative weights and energy availability.

From Figure 19(a), we can observe that without water circulation, it takes 3h for the system to notify the WDS administrator of small leaks/contamination spread. However, a small amount of artificial circulation reduces this to about 20min as seen from Figure 19(b). By providing the extra sampling capability at night, the water circulation procedure improves the effective *information*



(a) (b)

Fig. 18. Number of samples with time, (a) without water circulation and (b) with water circulation.



(a)

(b)

Fig. 19. Comparison of event reporting time (a) without and (b) with water circulation.

measure by 14% as observed from Figure 17(a). Also in this example, the amount of extra water circulation during night hours is only a small fraction (<0.7%) of the total water-flow throughout the day as derived from Figure 17(b).

7.3 Effects of Heterogeneous, Multi-sensor Contamination Sensing

We model both our schemes, i.e., HCS and AHCS, along with a simple non-adaptive scheme called *equal rate allocation (ERA)*, which assigns same sampling rate to all nodes. We assume two types of contaminants: Glyphosate and Dimethyl sulfoxide (DMSO) [66]. All the nodes are equipped with chlorine (Cl) sensors. Along with that the node with odd numbers are equipped with ORP sensors, whereas other are equipped with pH sensors. This brings the notion of heterogeneous sensing in a WDS system. Both contaminants are detected by the chlorine sensor, whereas ORP and pH only respond to Glyphosate. This brings the notion of collaborative sensing, considering the correlated detection ability among the sensors. In presence of Glyphosate in water [79], ORP increases, while chlorine and pH decrease. This is because Glyphosate is slightly acidic and has some oxidizing ability. DMSO reduces the chlorine concentration, whereas ORP and pH shows minor fluctuations [66]. A sensor's detection probability is assumed to be 100% if the sensor responds to a contaminant and zero otherwise. The probabilistic reliability modeling depends on a sensor's level of accuracy as well as the contaminant concentration at the sensing point, which we keep as part of our future work. Table 6 reports some of the commercial water quality sensors to measure the corresponding

Table 6. Different Sensor Specifications

Parameter	Sensor	Voltage range	Current draw
Chlorine	Chlorine sensor Type 8232 [80]	12–30V	4mA
ORP	WQ600 [81]	10–36V	0.2mA + sensor output (4–20mA)
pH	WQ201 [81]	10–30V	5.5mA + sensor output (4–20mA)

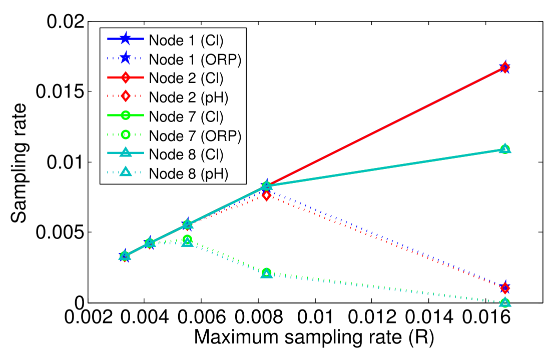


Fig. 20. Comparison of sampling rates of different sensors.

physical parameters, as well as their voltage requirements and current consumptions. 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.

Effect of maximum sampling rate $\mathbb{R}$*:* We solve the optimization problem HCS using **AMPL**, which is a modeling language for solving large-scale optimization problems [57]. Our main objective is to show the effect of adaptation at the low energy hours, i.e., from 11 PM to 5 AM. Figure 20 shows the variation of the assigned sampling rates with different $\mathbb{R}$ for 11 PM to 5 AM. From Figure 20, we can observe that at low $\mathbb{R}$, all the sensors sample at their maximum sampling rates. As $\mathbb{R}$ increases, some of the sensors start reducing their sampling frequencies. At higher $\mathbb{R}$, the first level nodes have higher sampling rates compared to the third level nodes. This is because of the higher harvested energy availability of the higher level nodes, which clearly shows the adaptive nature of HCS based on the energy availability of the individual nodes. Another interesting thing to notice is that for higher $\mathbb{R}$, the Chlorine sensor is used more often compared to others. This is because of the fact that the Chlorine sensor has least power consumption, which makes them more suitable to use frequently at low energy hours.

Effect of adaptation on event reporting time: Figure 21 shows effects of the collaborative adaptation on the event reporting time of the coalition. We assume $\mathbb{R}$ to be 0.0167 (1 sample/min) for Figures 21 and 22. From Figure 21(a), we can observe that without any adaptation, it takes more than $3\frac{1}{2}$ hours for the system to report the WDS administer of a contamination event. However, with the adaptation scheme HCS, the event reporting time is reduced to about 1 hour 45 minutes. We notice that the reporting time is still high. This is because of the energy loss due to the limited storage capacity of the super-capacitors, which is not considered in HCS modeling.

For AHCS, we divide the interval to smaller sub-intervals of 1h and adapt the sampling rates by considering the energy wastage due to lack of storage, as mentioned in Section 5. This reduces the reporting time to about 6min as seen from Figure 21(c). Figure 22 shows the sampling rate of the nodes at different time of night. From this figure, we can observe that at relatively high

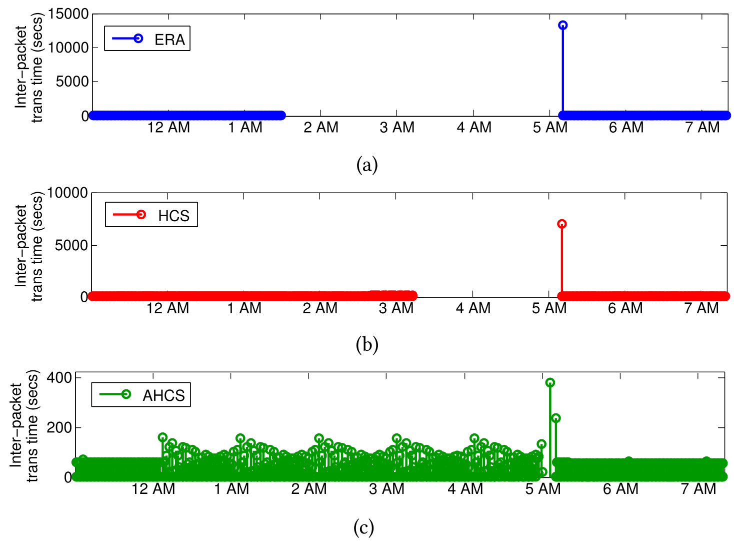


Fig. 21. Comparison of event reporting time (a) without adaptation (ERA), (b) with adaptation (HCS), and (c) with adaptation (AHCS).

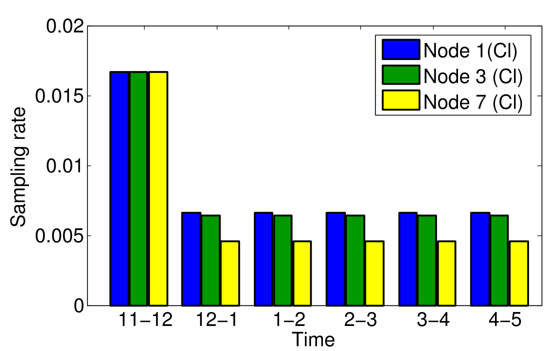


Fig. 22. Comparison of sampling rates at different timescales at lull hours.

energy hours (11 AM–12 PM), the sampling rate of all the nodes are high. This is because at 11 PM, all the nodes have fully charged super-capacitors, and so the incoming harvested energy plus the stored energy is sufficient for sampling at higher frequency, which would otherwise be wasted due to limited super-capacitor capacity. From 12 PM onwards, the harvested energy is much lower and so the nodes start reducing their sampling rates. In this case also, we notice that the higher level nodes are more active in sensing due to the adaptive nature of the AHCS, which drastically improves the contamination reporting time of a coalition.

7.4 Results from WaterNetGen

We also evaluated our rate adaptation scheme from a WDS simulator, named WaterNetGen [82], which is an EPANET [83] extension to automatically build a WDS system. We generate a synthetic WDS coalition consisting of 20 nodes (or junctions), 27 pipes, and a tank as shown in Figure 23(a). We assume that this coalition serves 100 inhabitants with a per capita demand of 200 liters per day. We use PVC10 pipes of 0.6Mpa for our simulations. We assume that each junction is equipped with a device with three sensors (chlorine, ORP, and pH) for reporting the contamination events.

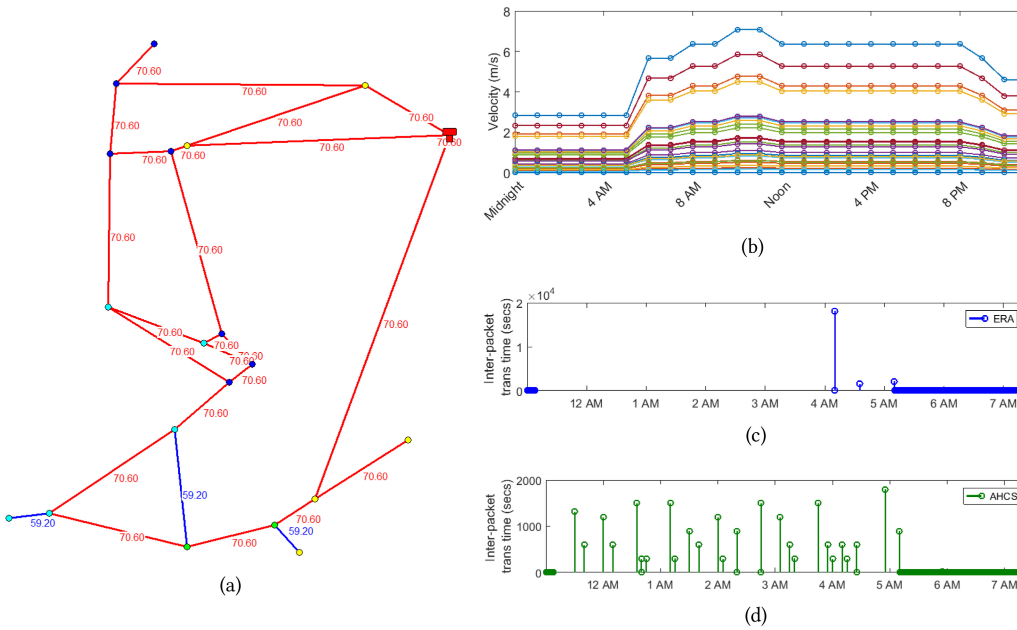


Fig. 23. (a) Sample pipeline topology with one pump, 20 nodes and 27 pipes. The values in the pipes represents the pipe diameters (in mm). (b) Water velocities through the pipes at different times of the day. Comparison of event reporting time for (c) ERA and (d) AHCS.

Table 7. Event Reporting Time of Different WDS Typologies

	Number of nodes	Number of pipes	Worst reporting time (ERA)	Worst reporting time (AHCS)
Topology-1	10	13	4h 38min	35min
Topology-2	15	19	1h 18min	15min
Topology-3	20	25	5h 2min	25min

We also assume that any contamination can be detected at all the junctions within a coalition. Figure 23(b) shows the variation of average flow velocities at different pipes on different time of the day, which is based on the diurnal patterns of water usage. This figure clearly shows the spatial and temporal variation of water flow in different pipes, thus strengthens the need for adaptations.

Figures 23(c) and 23(d) show the effects of the AHCS scheme in comparison to ERA to reduce the contamination reporting time at the night hours. We assume $\mathbb{R}$ to be 0.333 samples/s. From Figure 23, we can observe that without any adaptations the event reporting time is more than 5h, whereas in case of AHCS the worst event reporting time is reduced to almost 0.5h. We also conduct the same experiment on three other randomly generated synthetic typologies (each forming a coalition) with different number of nodes and pipes, as shown in Figure 24. Table 7 reports the number of nodes/pipes of these topologies along with their worst event reporting time with ERA and AHCS. Table 7 also confirms that AHCS reduces the event reporting time from several hours in case of ERA to few minutes in the low energy hours.

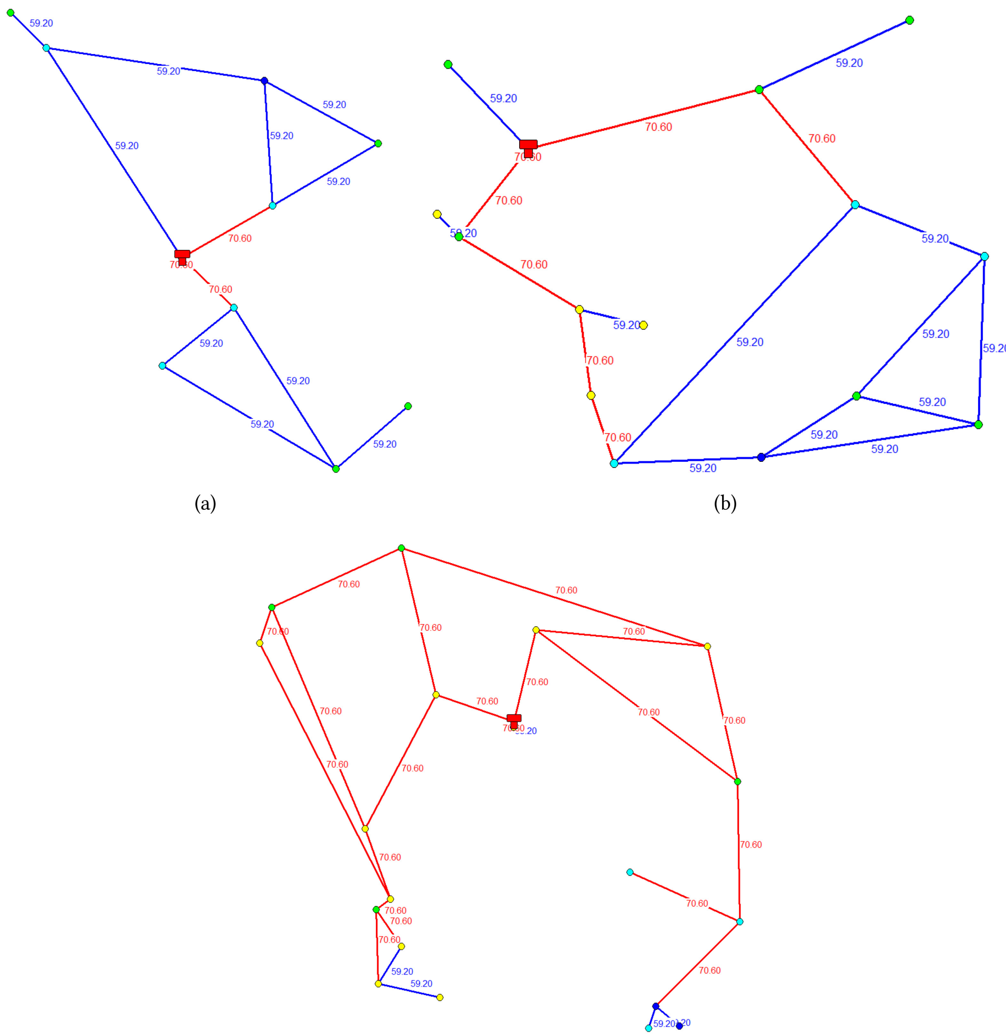


Fig. 24. Different WDS pipeline topologies (a)–(c) with various number of nodes and pipes. The values in the pipes represents the pipe diameters (in mm).

7.5 Cost analysis

Table 8 shows the cost estimation of different sensing/communication modules. The water-flow meters and sensing device's costs range from few hundreds dollars to thousand dollars. These are per-unit retail costs; any large scale deployment will likely procure these items at a fraction of the costs listed here. Also, in many instances, the flow meters likely already exist at the junction point. The zigbee/WiFi modules to communicate between the sensing devices to the sink nodes are typically cheap and much less expensive than the meters and other sensors. For example, the XBee-PRO (802.15.4) has an advertised communication range of 4000 meters and data rate of 250Kb/s [85]. This is quite adequate for our application, although even larger range (2 mile) modules are available at a slighly higher cost. A DMA is usually a few KM across, and the connection

Table 8. System Cost [76, 84]

Category	Type	Price
Water flow meter	WMP101-100 1″ Plastic Bodied In-Line Magmeter	$961
Water flow meter	WMP101-100 2″ Plastic Bodied In-Line Magmeter	$1,041
Water flow meter	WMP101-100 3″ Plastic Bodied In-Line Magmeter	$1,256
Water quality sensor	CL205 Chlorine Electrode Module	$129
Water quality sensor	WQ600 ORP Sensor	$755
Water quality sensor	WQ201 pH Sensor	$793
Communication module	XBee-PRO	$20
Harness for sensor/comm node	custom development	$50
Smart manhole cover w/ Antenna	custom development	$50

points usually closer; therefore, a single sink node per DMA should be enough in most cases. In a smart city context, the communication modules (deployed at each pipe connection point) could also be used for communicating other data such as urban pollution monitoring data.

The initial deployment of the sensing and communications infrastructure would involve the cost of not only the sensors and electronics, but perhaps more significantly miscellaneous other costs. One such item is the proper harness for the sensing/communication modules that is secure and can protect the electronics from heat, freezing, vibrations, dust, and so on. Another item is the antenna integrated with the manhole cover. Smart manhole covers already exist and are being deployed by several cities including Austin, Green Bay, and so on, for monitoring of storm-water [7]. We were unable to find accurate cost of these, and have simply assumed them to be $50 each in the table. Other difficult to quantify costs include the personnel costs of physical and electronic aspects of deployment, connecting sink nodes with the control center (e.g., via cellular links), testing, certification, and so on.

Given the difficulty in an accurate cost estimation, we conservatively assume a total cost of $2,500 per connection point, such cost can be easily recovered due to reduced water loss. Consider, for example, the city of Philadelphia [86]. The city serves 250MGD (million gallons/day) of treated water at about 2 cents/gallon, which amounts to $5M/day. Suppose that the network can cut down the water leakage from the reported 31% to 11%, which amounts to $1M/day savings. The city has about 91K valves on mainline pipes (of 6″ or larger diameter). As stated earlier, it is perhaps not necessary to instrument every valve. Suppose that we instrument about 1 in 3 valves, chosen carefully so that there is almost equal coverage of all the 3,000 miles of piping in the city. Thus, the deployment cost is approximately $2,500×30K, or $75M, or equivalent of 2.5 months worth of water saved.

8 RELATED WORK

8.1 Modeling of Water Distribution Networks

Modeling of water distribution networks in terms of supplies, demands, head (pressure) loss/gain, mass conservation, and so on, is routinely used for designing and configuring water distribution systems. For example, WaterGEMS is a commercial package that incorporates many features [87]. Such models have been used for leak modeling as well by considering them as demands that depend on the pressure [88]. Reference [89] thoroughly surveys of mathematical models used in water industry. It discusses both steady state analysis of water systems and those subject to transients, and also considers model uncertainties. Reference [90] provides a modeling framework for faults in water supply networks by using a combination of deterministic modeling and machine learning techniques. The latter consists of classification of patterns of demand and determining anomalies from such patterns. The pattern classification is based on neural networks and the patterns could be either crisp or fuzzy. Reference [91] presents a decision support system for operational monitoring and control of water distribution systems based on a similar fuzzy neural network technique. Reference [92] develops a probabilistic method for prediction of pipe failures and leak occurrence based on Bayesian analysis of continuously collected vibration data.

Different mathematical models have been explored for leak modeling as well by considering them as demands that depend on the pressure [93, 94]. Similar mathematical models are also explored for contamination detection in References [79, 95, 96]. The WaterGEMS modeling package incorporates many such features for designing and configuring water distribution systems [47].

8.2 Network Utility Maximization

Network utility maximization has received a significant attention in the last decade ever since the seminal framework the seminal work in References [97, 98]. In these works the users utility is assumed to be strictly concave function of users rate, and the resource constraints are set to be linear. The users distributedly maximize their aggregate utility under their resource constraints. Various types of fairness-based utilities are discussed in Reference [99]. Multi-path utility maximizations are addressed in References [69, 100–103], where the utility function is non-strictly concave with respect to the individual users rate due to multi-path routing. To convexify the utility function, proximal approach are proposed in References [100–103], whereas in Reference [69] the authors have proposed a modified strictly concave utility function and proposed a successive approximation method.

8.3 Energy Management in WSNs

Energy management in sensor network is a well researched area. Control of sleep/wakeup cycle is a standard technique that is explored in several MAC proposals [104]. Other techniques for reducing energy consumption include data compression and source coding [105] transmit power control [53, 106, 107], multiple channel assignment [52, 74, 108, 109], and so on. While these proposals are mainly motivated towards maximizing the life-time of the sensor network, our objective in this article is to schedule the operations according to the energy harvesting opportunities and adapt them to the energy availability that varies dynamically while maximizing the collection of most useful samples. In this regard, some relevant papers are References [110, 111], where the authors propose fair rate adaptation for interference or congestion control; however, they do not consider adaptation for meeting individual node's energy budget. Authors of References [58, 59] propose energy aware rate adaptation schemes using dual decomposition in a distributed manner, that can incur high control overhead and long running time, which make their schemes impractical especially in the context of resource constrained sensor networks [112]. In contrast, our technique is a *collaborative* rate adaptation that exploits correlated detection of a "coalition" of sensor nodes and is computed in a *centralized* manner to avoid the overhead of distributed computations. Moreover, we consider the presence of multiple and different types of sensors per node, as well as their interdependencies in the event reporting process, which is novel compared to the existing works. Such a scheme can be used in many other energy harvesting environments where the sensor nodes have correlated event detection or sensing capabilities.

9 CONCLUSIONS

In this article, we have explored the pervasive problem of monitoring leakage and contamination in urban water distribution systems, which in most instances are quite old and in poor shape with a lot of leakage and possibility of contamination. Unfortunately, most utilities are under severe financial constraints and unable to overhaul the system. The proposed sensor network can be deployed relatively inexpensively, and it would more than pay for itself due to significantly lower water loss, less need for system flushing due to potential contamination, and quick repairs to prevent further damage. Our proposal is to make the sensors driven by the flowing water and thereby avoid any need for either AC power or battery change. This brings challenges when the water flow rate is low, and we show that a minimal artificial water circulation in the system can be exploited to keep the network alive at all times. We also reduce energy consumption by dividing sensors into coalitions along with collaborative sampling within a coalition. In the future, we will consider an optimal sparse deployment of sensors by considering various constraints in terms of number and type of sensors deployed and the effectiveness of the leak/contamination detection.

ABBREVIATIONS

WDS	Water Distribution Systems
WDSN	Water flow driven sensor networks
DMA	District Metering Area
LPL	Low Power Listening
EQ	Energy Queue
DQ	Data Queue
NLMS	Normalized Least Mean Square
MSRC	Maximum Sampling Rate Constraint
CARA	Collaborative and Adaptive Rate Allocation
KKT	Karush-Kuhn-Tucker
EPA	Environmental Protection Agency
TOC	Total Organic Carbon
ORP	Oxidation Reduction Potential
VOC	Volatile organic carbon
NLMS	normalized least mean square
HCS	Heterogeneous Collaborative Sampling
AHCS	Advanced HCS
ERA	Equal rate allocation
DMSO	Dimethyl sulfoxide
GPR	Ground penetrating radar

REFERENCES

[1] D. Seckler, R. Barker, and U. Amarasingh. 2010. Water scarcity in the 21st Century. *Int. J. Water Resour. Dev.* 15, 1 (2010), 29–42.

[2] Giuseppe Anastasi, Marco Conti, and Mario Di Francesco. 2009. Extending the lifetime of wireless sensor networks through adaptive sleep. *IEEE Trans. Industr. Informat.* 5, 3 (2009), 351–365.

[3] Theofanis P. Lambrou, Christos C. Anastasiou, Christos G. Panayiotou, and Marios M. Polycarpou. 2014. A low-cost sensor network for real-time monitoring and contamination detection in drinking water distribution systems. *IEEE Sensors J.* 14, 8 (2014), 2765–2772.

[4] Amitangshu Pal and Krishna Kant. 2015. Water flow driven sensor networks for leakage and contamination monitoring. In *Proceedings of the IEEE WoWMoM.* 1–9.

[5] Amitangshu Pal and Krishna Kant. 2015. Collaborative heterogeneous sensing: An application to contamination detection in water distribution networks. In *Proceedings of the ICCCN.* 1–8.

[6] Christian Renner, Jurgen Jessen, and Volker Turau. 2009. Lifetime prediction for supercapacitor-powered wireless sensor nodes. In *Proceedings of the FGSN.*

[7] Anus W. Stocking. 2017. Smart manholes help nation's largest cities cut down on costs. Retrieved from https://www.concreteconstruction.net/projects/infrastructure/smart-manholes-help-nations-largest-cities-cut-down-on-costs_o.

[8] Elaine Shi and Adrian Perrig. 2004. Designing secure sensor networks. *IEEE Wireless Commun.* 11, 6 (2004), 38–43.
[9] Adrian Perrig, John A. Stankovic, and David Wagner. 2004. Security in wireless sensor networks. *Commun. ACM* 47, 6 (2004), 53–57.
[10] Alireza Jolfaei and Krishna Kant. 2018. A lightweight integrity protection scheme for low latency smart grid applications. *Comput. Secur. J.* https://www.sciencedirect.com/science/article/pii/S016740481831126X.
[11] Environmental Protection Agency. 2010. Control and mitigation of drinking water losses in distribution systems. Retrieved from water.epa.gov/type/drink/pws/smallsystems/upload/Water_Loss_Control_508_FINALDEc.pdf.
[12] Overview of Philadelphia's water infrastructure. [n.d.] Retrieved from www.phillywatersheds.org/watershedissues/infrastructuremanagement.
[13] Tariq Al-Kadi, Ziyad Al-Tuwaijri, and Abdullah Al-Omran. 2013. Wireless sensor networks for leakage detection in underground pipelines: A survey paper. In *Proceedings of the AASNET.* 491–498.
[14] U.S. Environmental Protection Agency. 2010. *Control and Mitigation of Drinking Water Losses in Distribution Systems.* Technical Report. U.S. EPA, Department of Water.
[15] George Kunkel and Reinhard Sturm. 2011. Piloting proactive, advanced leakage management technologies. *J. Amer. Water Works Assoc.* 103, 2 (2011), 62–75.
[16] Julian Thornton and Reinhard Sturm. 2008. *Water Loss Control.* McGraw-Hill.
[17] R. Puust, Z. Kapelanb, D. A. Savicb, and T. Koppela. 2010. A review of methods for leakage management in pipe networks. In *Urban Water Journal*, vol. 7. 25–45.
[18] Cyrille Lemoine. 2012. Smart sensors network for water quality monitoring. A way to re-invent water supply management? Retrieved from https://www.researchgate.net/publication/273452875_Smart_water_sensors_network_a_way_to_re-invent_water_supply_management.
[19] Orbit Communications. [n.d.]. Wireless Pump Control. Retrieved from http://glenvaletechnology.com/Frameset/Content.taf?Item_uid=879.
[20] Majid Ahadi and Mehrdad Sharif Bakhtiar. 2010. Leak detection in water-filled plastic pipes through the application of tuned wavelet transforms to acoustic emission signals. *Appl. Acoust.* 71, 7 (2010), 634–639.
[21] D. Butler and Palmer Environmental. 2000. *Leakage Detection and Management: A Comprehensive Guide to Technology and Practice in the Water Supply Industry.* Palmer Environmental.
[22] Matthaios Bimpas, Angelos Amditis, and Nikolaos Uzunoglu. 2010. Detection of water leaks in supply pipes using continuous wave sensor operating at 2.45GHz. *J. Appl. Geophys.* 70, 3 (2010), 226–236.
[23] A. De Coster, J.L. PÃ©rez Medina, M. Nottebaere, K. Alkhalifeh, X. Neyt, J. Vanderdonckt, and S. Lambot. 2019. Towards an improvement of GPR-based detection of pipes and leaks in water distribution networks. *J. Appl. Geophys.* 162 (2019), 138–151.
[24] Tengku Sarah Tengku Amran, Mohamad Pauzi Ismail, Mohamad Ridzuan Ahmad, Mohamad Syafiq Mohd Amin, Suhairy Sani, Noor Azreen Masenwat, Mohd Azmi Ismail, and Shu-Hazri Abdul Hamid. 2017. Detection of underground water distribution piping system and leakages using ground penetrating radar (GPR). *AIP Conf. Proc.* 1799, 1 (2017), 030004.
[25] Andrew F. Colombo, Pedro Lee, and Bryan W. Karney. 2009. A selective literature review of transient-based leak detection methods. *J. Hydro-environ. Res.* 2, 4 (2009), 212–227.
[26] D. Waleed, S. H. Mustafa, S. Mukhopadhyay, M. F. Abdel-Hafez, M. A. K. Jaradat, K. R. Dias, F. Arif, and J. I. Ahmed. 2019. An in-pipe leak detection robot with a neural-network-based leak verification system. *IEEE Sensors J.* 19, 3 (2019), 1153–1165.
[27] Muhammad Zahid Abbas, Kamarulnizam Abu Baker, Muhammad Ayaz, Hafiz Mohamed, Moeenuddin Tariq, Adnan Ahmed, and Muhammad Faheem. 2018. Key factors involved in pipeline monitoring techniques using robots and WSNs: Comprehensive survey. *J. Pipeline Syst. Engineer. Pract.* 9, 2 (2018), 04018001.
[28] H. Che and S. Liu. 2014. Contaminant detection using multiple conventional water quality sensors in an early warning system. *Procedia Eng.* 89 (2014), 479–487.
[29] Kofi Sarpong Adu-Manu, Cristiano Tapparello, Wendi B. Heinzelman, Ferdinand Apietu Katsriku, and Jamal-Deen Abdulai. 2017. Water quality monitoring using wireless sensor networks: Current trends and future research directions. *Trans. Sensor Netw.* 13, 1 (2017), 4:1–4:41.
[30] John Hall, Alan D. Zaffiro, Randall B. Marx, Paul C. Kefauver, E. Radha Krishnan, Roy C. Haught, and Jonathan G. Herrmann. 2007. On–line water quality parameters as indicators of distribution system contamination. *Amer. Water Works Assoc.* 99, 1 (2007), 66–77.
[31] Y. Jeffrey Yang, Roy C. Haught, and James A. Goodrich. 2009. Real-time contaminant detection and classification in a drinking water pipe using conventional water quality sensors: Techniques and experimental results. *J. Environ. Manage.* 90, 8 (2009), 2494–2506.
[32] D. J. Kroll. 2006. *Securing Our Water Supply: Protecting a Vulnerable Resource.* PennWell Corporation. Retrieved from https://books.google.com/books?id=37Omyz1Hy5wC.

[33] Dibo Hou, Huimei He, Pingjie Huang, Guangxin Zhang, and Hugo Loaiciga. 2013. Detection of water-quality contamination events based on multi-sensor fusion using an extented Dempster–Shafer method. *Measure. Sci. Technol.* 24, 5 (2013), 055801.

[34] Lina Perelman, Jonathan Arad, Mashor Housh, and Avi Ostfeld. 2012. Event detection in water distribution systems from multivariate water quality time series. *Environ. Sci. Technol.* 46, 15 (2012), 8212–8219.

[35] Jonathan Arad, Mashor Housh, Lina Perelman, and Avi Ostfeld. 2013. A dynamic thresholds scheme for contaminant event detection in water distribution systems. *Water Res.* 47, 5 (2013), 1899–1908.

[36] Ivan Stoianov, Lama Nachman, Sam Madden, Timur Tokmouline, and M Csail. 2007. PIPENET: A wireless sensor network for pipeline monitoring. In *Proceedings of the ACM/IEEE IPSN*. 264–273.

[37] Younghun Kim, Thomas Schmid, Zainul Charbiwala, Jonathan Friedman, and Mani B. Srivastava. 2008. NAWMS: Nonintrusive autonomous water monitoring system. In *Proceedings of the ACM SenSys*. 309–322.

[38] Zhi Sun, Pu Wang, Mehmet C. Vuran, Mznah Al-Rodhaan, Abdullah Al-Dhelaan, and Ian F. Akyildiz. 2011. MISE-PIPE: Magnetic induction-based wireless sensor networks for underground pipeline monitoring. *Ad Hoc Netw.* 9, 3 (2011), 218–227.

[39] Masanobu Shinozuka, Pai H. Chou, Sehwan Kim, Hong Rok Kim, E. Yoon, Hadil Mustafa, and Debdut Karmakar. 2010. Nondestructive monitoring of a pipe network using a MEMS-based wireless network. In *Proceedings of the SPIE SS/NDE*.

[40] Jong-Hoon Kim, Gokarna Sharma, Noureddine Boudriga, and S. Sitharama Iyengar. SPAMMS: A sensor-based pipeline autonomous monitoring and maintenance system. In *Proceedings of the IEEE COMSNETS*. 1–10.

[41] A. Kadri, A. Abu-Dayya, D. Trinchero, and R. Stefanelli. 2011. Autonomous sensing for leakage detection in underground water pipelines. In *Proceedings of the ICST*. 639–643.

[42] Ted Tsung-Te Lai, Wei-Ju Chen, Kuei-Han Li, Polly Huang, and Hao-Hua Chu. 2012. TriopusNet: Automating wireless sensor network deployment and replacement in pipeline monitoring. In *Proceedings of the ACM/IEEE IPSN*. 61–72.

[43] Avi Ostfeld, James G Uber, Elad Salomons, Jonathan W Berry, William E Hart, Cindy A Phillips, Jean-Paul Watson, Gianluca Dorini, Philip Jonkergouw, Zoran Kapelan, et al. 2008. The battle of the water sensor networks (BWSN): A design challenge for engineers and algorithms. *J. Water Resour. Plan. Manage.* 134, 6 (2008), 556–568.

[44] William E Hart and Regan Murray. 2010. Review of sensor placement strategies for contamination warning systems in drinking water distribution systems. *J. Water Resour. Plan. Manage.* 136, 6 (2010), 611–619.

[45] M Prasanna Kumar and MS Kumar. 2009. Comparative study of three types of controllers for water distribution networks.*J. Amer. Water Works Assoc.* 101, 1 (2009).

[46] U. Manohar and M. Kumar. 2014. Modeling equitable distribution of water: Dynamic inversion-based controller approach. *J. Water Resour. Plan. Manage.* 140, 5 (2014), 607–619.

[47] Bentley. [n.d.] Retrieved from www.bentley.com/en-US/Products/WaterGEMS/Features-List.html.

[48] H. Al-Qahtani, Y.N. Al-Nassar, M.A. Hawwa, and K. Youcef-Toumi. 2012. Acoustic probe for leak detection in water pipelines. Retrieved from https://www.google.com/patents/US20120285221 US Patent App. 13/412,718.

[49] Philadelphia Water Department. 2007. Pure technologies, PWD. Retrieved from https://www.environmental-expert.com/articles/case-study-the-philadelphia-water-department-pwd-677191.

[50] M. Sha, G. Hackmann, and C. Lu. 2013. Energy-efficient Low Power Listening for wireless sensor networks in noisy environments. In *Proceedings of the ACM/IEEE IPSN*. 277–288.

[51] Omprakash Gnawali, Rodrigo Fonseca, Kyle Jamieson, David Moss, and Philip Levis. 2009. Collection tree protocol. In *Proceedings of the SenSys*. 1–14.

[52] Amitangshu Pal and Asis Nasipuri. 2017. Distributed routing and channel selection for multi-channel wireless sensor networks. *J. Sensor Actuat. Netw.* 6, 3 (2017), 10.

[53] Amitangshu Pal and Asis Nasipuri. 2014. PCOR: A joint power control and routing scheme for rechargeable sensor networks. In *Proceedings of the IEEE WCNC*. 2230–2235.

[54] M. J. Khan, M. T. Iqbal, and J. E. Quaicoe. 2008. River current energy conversion systems: Progress, prospects and challenges. *Renew. Sustain. Energy Rev.* 12, 8 (2008), 2177–2193.

[55] James W. Nilsson and Susan Riedel. 2007. *Electric Circuits*, 8th ed. Prentice-Hall, Upper Saddle River, NJ.

[56] Pramod R. Bhave and Rajesh Gupta. 2006. Analysis of water distribution networks. (2006).

[57] AMPL. [n.d.] Retrieved from http://ampl.com/products/solvers/.

[58] Lu Su, Yan Gao, Yong Yang, and Guohong Cao. 2011. Towards optimal rate allocation for data aggregation in wireless sensor networks. In *Proceedings of the MobiHoc*.

[59] Ren-Shiou Liu, Prasun Sinha, and Can Emre Koksal. 2010. Joint energy management and resource allocation in rechargeable sensor networks. In *Proceedings of the INFOCOM*. 902–910.

[60] Stephen M. George, Wei Zhou, Harshavardhan Chenji, Myounggyu Won, Yong-Oh Lee, Andria Pazarloglou, Radu Stoleru, and Prabir Barooah. 2010. DistressNet: A wireless ad hoc and sensor network architecture for situation management in disaster response. *IEEE Commun. Mag.* 48, 3 (2010), 128–136.

[61] Mayank Raj, Krishna Kant, and Sajal K. Das. 2014. E-DARWIN: Energy aware disaster recovery network using WiFi tethering. In *Proceedings of the IEEE ICCCN*. 1–8.
[62] Peter W. Gething and Andrew J. Tatem. 2011. Can mobile phone data improve emergency response to natural disasters?*PLoS Med.* 8, 8 (2011), e1001085. DOI: 10.1371/journal.pmed.1001085
[63] Weifeng Shan, Jilin Feng, Jianjun Chang, Fan Yang, and Zhonghua Li. 2012. Collecting earthquake disaster area information using smart phone. In *Proceedings of the IEEE ICSSE*. 310–314.
[64] Changhong Yang, Jun Yang, Xiangshu Luo, and Peng Gong. 2009. Use of mobile phones in an emergency reporting system for infectious disease surveillance after the Sichuan earthquake in China. *Bull. World Health Org.* 87 (2009), 619–623.
[65] United States Environmental Protection Agency. 2015. Water quality surveillance and response system primer. Retrieved from https://www.epa.gov/sites/production/files/2015-06/documents/water_quality_sureveillance_and_response_system_primer.pdf.
[66] United States Environmental Protection Agency. 2009. Distribution SystemWater Quality Monitoring: Sensor Technology Evaluation Methodology and Results. Retrieved from https://www.epa.gov/sites/production/files/2015-06/documents/distribution_system_water_quality_monitoring_sensor_technology_evaluation_methodology_results.pdf.
[67] Philadelphia Water Department and CH2M HILL. 2013. Selection of Online Water Quality Monitoring Technologies and Station Design. Retrieved from https://www.epa.gov/sites/production/files/2017-10/documents/pwd_report_selection_of_owqm_technologies_station_design.pdf.
[68] United States Environmental Protection Agency. 2007. Water Security Initiative: Interim Guidance on Planning for Contamination Warning System Deployment. Retrieved from https://www.hsdl.org/?view&did=34212.
[69] Phuong Luu Vo, Tuan Anh Le, Sungwon Lee, Choong Seon Hong, Byeongsik Kim, and Hoyoung Song. 2014. Multi-path utility maximization and multi-path TCP design. *J. Parallel Distrib. Comput.* 74, 1 (2014), 1848–1857.
[70] Yongmin Zhang, Shibo He, Jiming Chen, Youxian Sun, and Xuemin (Sherman) Shen. 2013. Distributed sampling rate control for rechargeable sensor nodes with limited battery capacity. *IEEE Trans. Wireless Commun.* 12, 6 (2013), 3096–3106.
[71] Castalia: A Simulator for WSN. [n.d.] Retrieved from castalia.npc.nicta.com.au/.
[72] Aquacraft Inc. 2009. Embedded Energy in Water Study 3: End-use Water Demand Profile Final Research Plan. Retrieved from uc-ciee.org/downloads/Eeiswtudy3.pdf.
[73] Western Municipal Water District. 2011. Design Criteria for Water Distribution Systems. Retrieved from www.wmwd.com/DocumentCenter/Home/View/239.
[74] Amitangshu Pal. 2013. *Dynamic routing with cross-layer adaptations for multi-hop wireless networks*. Ph.D. Dissertation. Charlotte, NC.
[75] Mahmoud Shuker Mahmoud and Auday A. H. Mohamad. 2016. A study of efficient power consumption wireless communication techniques/modules for Internet of Things (IoT) applications. *Adv. Internet Things* 6 (2016), 19–29.
[76] Digi International Inc. [n.d.] Retrieved from https://www.sparkfun.com/datasheets/Wireless/Zigbee/XBee-Datasheet.pdf.
[77] Global Water. [n.d.] Retrieved from http://www.globalw.com/products/wmx101.html.
[78] Global Water. [n.d.] Retrieved from http://www.globalw.com/products/wmp101.html.
[79] Shuming Liu, Han Che, Kate Smith, and Lei Chen. 2014. Contamination event detection using multiple types of conventional water quality sensors in source water. *Environ. Sci. Processes Impacts* 16, 8 (2014), 2028–2038.
[80] Burkert Fluid Control Systems. [n.d.] Retrieved from http://www.burkert.com/en/type/8232.
[81] Global Water. [n.d.] Retrieved from http://www.globalw.com/products/waterquality.html.
[82] Joao Muranho, Ana Ferreira, Joaquim Sousa, Abel Gomes, and Alfeu Sa Marques. 2012. WaterNetGen: An EPANET extension for automatic water distribution network models generation and pipe sizing. *Water Supply* 12, 1 (2012), 117–123.
[83] EPANET: Application for modeling drinking water distribution systems. [n.d.] Retrieved from https://www.epa.gov/water-research/epanet.
[84] Global Water. [n.d.] Retrieved from http://www.globalw.com/PriceList/GWI_Price_List.pdf.
[85] Mouser Electronics. [n.d.] Retrieved from https://www.mouser.com/ProductDetail/Digi-International/XB24CZ7WIT-004?qs=sGAEpiMZZMtJacPDJcUJYw9DFeLukfx0ucfi4vTugkhUEsZRIRtPdw%3D%3D.
[86] Water rates. 2016. Retrieved from https://www.phila.gov/water/rateboard/PDF/RateFilingPresentation_BV-16.02.22_FINAL.PDF.
[87] Water distribution analysis and design software. [n.d.] Retrieved from http://www.bentley.com/en-US/Products/WaterGEMS/Features-List.html.
[88] Z. Y. Wu, P. Sage, and D. Turtle. 2009. Pressure-dependent leak detection model and its application to a district water system. In *Journal of Water Resources Planning and Management*, vol. 136.

[89] J. Izquierdo, R. Perez, and P. Iglesias. 2004. Mathematical models and methods in the water industry. In *Mathematical and Computer Modelling*, vol. 39. 1353–1374.
[90] J. Izquierdo, P. A. Lopez, F. J. Martinez, and R. Perez. 2007. Fault detection in water supply systems using hybrid (theory and data-driven) modelling. In *Mathematical and Computer Modelling*, vol. 46. 341–350.
[91] C. Arsene, B. Gabrys, and D. Al-Dabass. 2012. Decision support system for water distribution systems based on neural networks and graphs theory for leakage detection. In *Expert Systems with Applications*, vol. 39. 13214–13224.
[92] W. Li and W. Ling. 2011. Development of systems for detection, early warning, and control of pipeline leakage in drinking water distribution: A case study. In *Journal of Environmental Sciences*, vol. 23. 1816–822.
[93] J. Izquierdo, R. Perez, and P. L. Iglesias. 2004. Mathematical models and methods in the water industry. *Math. Comput. Model.* 39, 11–12 (2004), 1353–1374.
[94] Zheng Yi Wu, Paul Sage, and David Turtle. 2010. Pressure-dependent leak detection model and its application to a district water system. *J. Water Resour. Plan. Manage.* 136, 1 (2010), 116–128.
[95] Han Che, Shuming Liu, and Kate Smith. 2015. Performance evaluation for a contamination detection method using multiple water quality sensors in an early warning system. *Water Quality Control Manage.* 7, 4 (2015), 1422–1436.
[96] Dibo Hou, Huimei He, Pingjie Huang, Guangxin Zhang, and Hugo Loaiciga. 2013. Detection of water-quality contamination events based on multi-sensor fusion using an extented Dempster–Shafer method. *Measure. Sci. Technol.* 24, 5 (2013).
[97] Frank Kelly. 1997. Charging and rate control for elastic traffic. *Eur. Trans. on Telecommun.* 8 (1997), 33–37.
[98] Steven H. Low and David E. Lapsley. 1999. Optimization flow control-I: Basic algorithm and convergence. *IEEE/ACM Trans. on Netw.* 7, 6 (1999), 861–874.
[99] Jeonghoon Mo and Jean C. Walrand. 2000. Fair end-to-end window-based congestion control. *IEEE/ACM Trans. Netw.* 8, 5 (2000), 556–567.
[100] Wei-Hua Wang, Marimuthu Palaniswami, and Steven H. Low. 2003. Optimal flow control and routing in multi-path networks. *Perform. Eval.* 52, 2–3 (2003), 119–132.
[101] Xiaojun Lin and Ness B. Shroff. 2006. Utility maximization for communication networks with multipath routing. *IEEE Trans. Automat. Contr.* 51, 5 (2006), 766–781.
[102] Huaizhong Han, Srinivas Shakkottai, C. V. Hollot, R. Srikant, and Donald F. Towsley. 2006. Multi-path TCP: A joint congestion control and routing scheme to exploit path diversity in the internet. *IEEE/ACM Trans. Netw.* 16, 6 (2006), 1260–1271.
[103] Thomas Voice. 2007. Stability of multi-path dual congestion control algorithms. *IEEE/ACM Trans. Netw.* 15, 6 (2007), 1231–1239.
[104] Abdelmalik Bachir, Mischa Dohler, Thomas Watteyne, and Kin K. Leung. 2010. MAC essentials for wireless sensor networks. *IEEE Commun. Surveys Tutor.* 12, 2 (2010), 222–248.
[105] Caimu Tang and Cauligi S. Raghavendra. 2004. Compression techniques for wireless sensor networks. (2004), 207–231. Retrieved from https://link.springer.com/chapter/10.1007/978-1-4020-7884-2_10.
[106] Amitangshu Pal, Bonee Soibam, and Asis Nasipuri. 2013. A distributed power control and routing scheme for rechargeable sensor networks. In *Proceedings of the IEEE SoutheastCon.*
[107] Amitangshu Pal and Asis Nasipuri. 2014. Lifetime of asynchronous wireless sensor networks with multiple channels and power control. In *Proceedings of the IEEE WCNC.* 2874–2879.
[108] Amitangshu Pal and Asis Nasipuri. 2012. A distributed channel selection scheme for multi-channelwireless sensor networks. In *Proceedings of the MobiHoc.* 263–264.
[109] Amitangshu Pal and Asis Nasipuri. 2013. DRCS: A distributed routing and channel selection scheme for multi-channel wireless sensor networks. In *Proceedings of the IEEE PerSeNS.* 602–608.
[110] Sumit Rangwala, Ramakrishna Gummadi, Ramesh Govindan, and Konstantinos Psounis. 2006. Interference-aware fair rate control in wireless sensor networks. In *Proceedings of the SIGCOMM.* 63–74.
[111] Cheng Tien Ee and Ruzena Bajcsy. 2004. Congestion control and fairness for many-to-one routing in sensor networks. In *Proceedings of the SenSys.* 148–161.
[112] Amitangshu Pal and Krishna Kant. 2016. On the feasibility of distributed sampling rate adaptation in heterogeneous and collaborative wireless sensor networks. In *Proceedings of the ICCCN.* 1–9.

Received November 2018; revised March 2019; accepted May 2019