Collaborative Heterogeneous Sensing: An Application to Contamination Detection in Water Distribution Networks

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

Abstract- In this paper we consider sensor networks for detecting contamination in urban water distribution systems. We assume that the sensor nodes are installed at connection points only (through the manholes) and are driven by super-capacitors charged by water flow. 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. In this paper we consider such a situation and examine the problem of collaborative adaptation of heterogeneous set of sensors in order to maximize contamination detection, especially during periods of almost zero natural water flow. The paper shows, through extensive simulations, that the proposed approach can drastically reduce the contamination reporting time from $3\frac{1}{2}$ hours to $\sim$6 minutes, compared to the case without adaptation.

I. Introduction

Distributing clean and fresh water is crucial for public health, economic, environmental, and societal sustainability perspectives. In USA many of the underground pipes are 100+ years old, and so are subject to seepage, rusts, bacterial or and microbial growth etc. It is thus increasingly important to monitor and control the quality of water starting from the storage reservoirs/tanks to the residential and industrial areas. The incidents of waterborne outbreaks are numerous. In Walkerton, Canada, 2500 people were poisoned, and 7 died by e-coli in drinking water, following a resource contamination, in the year of 2000 [1]. In 2007, 8500 people were ill in Nokia, Finland due to a cross-connection of wastewater into distribution network. Traditional water quality involves manual water collection from different locations and conduct laboratory testing to characterize the water quality [2]. Such techniques cannot provide real-time, continuous water quality throughout a water distribution system. There is a great urgency to develop ICT based solutions that can detect and alert the operators in an event of a deteriorating water quality much more cheaply and continuously, than mostly manual procedures followed today.

To address these issues, in this paper we consider *water flow driven sensor networks (WDSN)* for monitoring water quality and alert the operators in case of a potential health hazard. In our recent work [3], we have considered such an architecture for leakage detection in urban water systems and introduced the idea of WDSN, that are entirely powered by water flow via a small hydro fan unit. The paper investigated the problem of low water consumption and thus low harvested energy at night time, and introduced an artificial water circulation mechanism to keep the sensor nodes running at low energy hours. The paper also developed a sampling and transmission rate adaptation scheme based on individual node's energy budget, by exploiting the highly correlated detection ability of the individual sensor nodes. The proposed network architecture assumed sensor nodes to be placed at pipe connections or valve points only, that are accessible through the manholes. This architecture is significantly different and more practical than other proposals such as PipeNet [4], MISE-PIPE [5] where the sensor nodes are assumed to be placed at any arbitrary points. This is because the water pipes are mostly buried underground, and so the manhole points are easily accessible for sensor deployment. The lower parts of the sensor nodes dip into the water for energy harvesting and measurement of contamination, velocity, etc., whereas the upper parts have a suitable wireless radio (e.g., WiFi) with antennas embedded on the exposed side of manhole covers.

This research was supported by the NSF grant CNS-1414509.

In this paper, we extend our previous work to develop a sensor network that continuously monitors various measurable parameters in water pipes and reports relevant data to a control station that can do the necessary analytics for contaminant detection to take necessary measures. 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 chemical is *inferred* via easily measurable properties of the water such as pH, conductivity, temperature or depletion profile of added chlorine. 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 assume that the wireless devices are equipped with multiple, *different* water quality sensors that measures 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 are 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. 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 water quality sensors.

The paper quantifies the advantages of our approach via extensive simulation studies using available measurement data and shows that the proposed mechanism can reduce reporting time from $3\frac{1}{2}$ hours to $\sim$6 minutes during late nights. To the best of our knowledge, this is the first work on water distribution systems with heterogeneous and collaborative contamination sensing – one that tries to achieve on-line continuous

978-1-4799-9964-4/15/$31.00 ©2015 IEEE

monitoring considering the real-world practical constraints of the water networks.

The outline of the paper is as follows. Section II describes the WDSN and the energy harvesting model. The concept of heterogeneous, adaptive sensing is introduced in section III. Section IV then gives an overview of different contaminants as well as the water quality sensors. Section V address the problem of sensing/transmission rate adaptation based on the node's energy budget. Section VI then presents the evaluation of the scheme.

II. Water Flow Driven Sensor Network (WDSN)

We consider a large water distribution network such as the one in Fig 1 (taken from Philadelphia's water network and showing only a tiny part of the entire network). Only a small part of the network is shown; the overall network is rather irregular with more densely connected sections serving more populated areas and layout closely following the street layouts. The colors represent water pressure, red being highest and blue lowest. The graph is cyclic but the flows are affected by the valve settings - certain valves may allow only 1-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, etc., possibly allowing for seepage of contaminants or vulnerability to deliberate infrastructure attacks.

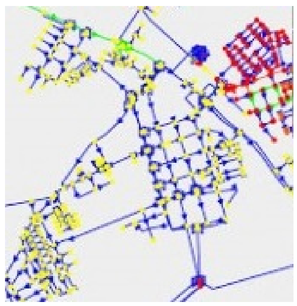

Fig. 1. A sample water distribution system.

Typical water distribution system consists of *main* lines running from the reservoirs, and further divided into *sub-mains* and *branch* lines from where service connections are given to the customers. Water distribution networks are normally divided up into *District Metering Areas* or DMAs, with ability to not only measure relevant parameters such as inflows, outflows, flow head (pressure), etc., but also to control them. In dense urban areas, a DMA may consist of couple of city blocks, and we consider that as the object of study here.

A. Network Operation

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 synchronized and use the basic *Low Power Listening (LPL)* [6] principle to conserve energy. 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, power availability, and security 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 be required to cover the entire DMA, and can be enabled via LPL mechanism, we limit ourselves to single hop communication in this paper. The single hop limitation may require deploying multiple sink nodes in a DMA, but it suffices to pretend that there is a single virtual sink node for the entire DMA.

B. Energy Harvesting model:

The energy harvesting model is described in [3], where we assumed that each node is equipped with a suitable fan based harvester. The kinetic energy of the streaming fluid rotates the blade and generates electricity. The basic equations governing this energy conversion are well established [7]. 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 (1000kg/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

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

where η_e denotes the overall efficiency. In reality, η_e depends on a large variety of factors, and is the domain of mechanical/electrical design of the harvesting unit.

III. Heterogeneous Sensor Networks with Adaptive Sampling

A. Heterogeneous and Collaborative Sensing

Heterogeneous sensor networks are defined as networks where the wireless nodes are equipped with multiple, different types of sensors, such as audio, video, acceleration etc. Recently, wireless devices are increasingly being equipped with multi-modal sensors. Cellphones and Smart-phones are perfect examples of wireless nodes with multiple sensors. These multi-modal sensor nodes are increasingly being used for disaster recovery applications [8], [9] earthquake monitoring [10], [11], infectious disease surveillance [12] etc. 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 vibration. The audio samples can also be used to track the sound of building collapsing. Videos and images can be used to build a spatial view of the damage caused by the earthquake. In this example, all three sensors measure some aspect of the *same* phenomenon of ground movement. 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 building collapsing can happen due to other disasters, such as cyclones, hurricanes etc.

In this paper 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 Fig. 2 that 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, and 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.

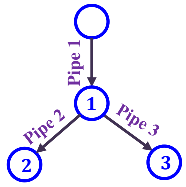

Fig. 2. Contamination detection in a water pipe network.

B. Adaptive Sampling

In branch pipes, the water flow rate may vary significantly and drop to near zero late at night. Thus it is challenging to keep the network up and running in these long lull hours. 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, the 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. The key challenge in such a WDSN with small super-capacitors is that, (a) when the water usage is high, the harvested energy produced is significantly higher than the limited storage ability if the super-capacitors, whereas (b) at the time of lull hours, the energy availability is too low to run the sensor nodes. In such a situation 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.

IV. Various Water Quality Indicators

EPA has defined 12 classes of potential water contaminants, which are reported in [13], [14]. 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 [13]. It is reported in [13] that 10 of the 12 classes of contaminants can be detected by measuring three common surrogate parameters, [14]: chlorine residual, conductivity, and total organic carbon (TOC).

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 $\sim$0.5 mg/L to more than 1 mg/L, depending on baseline levels and variability. To measure the TOC in water, absorbence of ultraviolet light at 254-nanometer wavelength is sometimes used [13]. 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 [13].

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, that 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 $\mathcal{E} = c.l$ 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 paper 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 paper.

V. 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 [15] 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 T 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 time scale. The notations used for the problem formulations are listed in Table I.

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.

TABLE I. Table of Notations

		Indices
i, j	$\triangleq$	Index for the sensors (1, ..., $\mathcal{S}$)
k	$\triangleq$	Index for contamination events (1, ..., $\mathcal{E}$)
m	$\triangleq$	Index for nodes (1, ..., $\mathcal{N}$)
		Binary input variables
$x_i^m \in 0, 1$	$\triangleq$	Whether or not sensor i is attached to node m
		Other variables
p_i^{km}	$\triangleq$	Probability that a contamination event k can be inferred from the readings of sensor i of node m
$\mathbb{P}_i^m$	$\triangleq$	Sensing + transmission power consumption for 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

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. On the other hand, 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 $e_k = \sum_{i=1}^{\mathcal{S}}\sum_{m=1}^{\mathcal{N}} p_i^{km}.r_i^m$. 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

$$\begin{aligned} \text{Maximize} \quad & \sum_{k=1}^{\mathcal{E}} \log\left(\sum_{i=1}^{\mathcal{S}}\sum_{m=1}^{\mathcal{N}} p_i^{km}.r_i^m\right) \\ \text{subject to} \quad & \sum_{i=1}^{\mathcal{S}} r_i^m.\mathbb{P}_i^m.T \le E^m - \mathcal{A}^m - \mathcal{O}^m - \tau \;\; \forall m \\ & 0 \le r_i^m \le \mathbb{M}.x_i^m \;\; \forall i, \forall m \\ & 0 \le r_i^m \le \mathbb{R} \;\; \forall i, \forall m \end{aligned} \tag{3}$$

where $\mathcal{A}^m$ and $\mathcal{O}^m$ are estimated energy arrival and power consumption 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 it's energy budget. All the nodes try to maintain a minimum energy threshold, which is assumed to be τ. The second set of constraints says that r_i^m is one only if node m is equipped with sensor i. The third set of constraints bound the maximum sampling rate of a sensor to be $\mathbb{R}$. As log is a concave function, the above problem is a convex optimization problem, which can be computed centrally by solving the Lagrangian and KKT conditions.

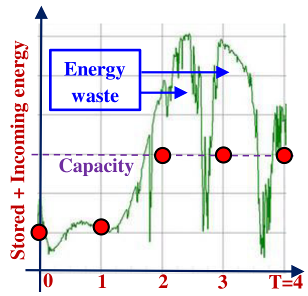

Fig. 3. Energy wastage due to lack of storage capacity. Red dots are the actual energy.

Advanced HCS: Notice that the optimization problem formulation (4) 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 Fig. 3. Thus the optimization problem (4) overestimates the sampling rates of the sensor. If this loss factor is not taken into account, 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 (4) 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:

$$
\begin{aligned}
\text{Max} \quad & \sum_{t=1}^{T}\sum_{k=1}^{\mathcal{E}} \log\left(\sum_{i=1}^{\mathcal{S}}\sum_{m=1}^{\mathcal{N}} p_i^{km}.r_{ti}^m\right) \\
\text{s.t.} \quad & L_0^m = E^m \ \ \forall m \\
& L_t^m = \min\left\{\mathbb{C}^m, L_{t-1}^m + \mathcal{A}_t^m - \sum_{i=1}^{\mathcal{S}} \mathbb{P}_i^m.r_{ti}^m.\Delta t - \mathcal{O}_t^m\right\} \ \ \forall m, \forall t \\
& L_t^m \geq \tau \ \ \forall m, \forall t \\
& 0 \leq r_i^m \leq \mathbb{M}.x_i^m \ \ \forall i, \forall m \\
& 0 \leq r_i^m \leq \mathbb{R} \ \ \forall i, \forall m
\end{aligned}
$$

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 Δt 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.

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 [3]. This paper addresses only 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 is not the scope of this paper.

VI. Simulation Results

Ideally, the evaluation of the scheme should be done with a real water distribution network testbed, however, 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 paper is largely based on simulations that account for the water flow physics [18] and use parameters obtained from characterization of real water distribution systems.

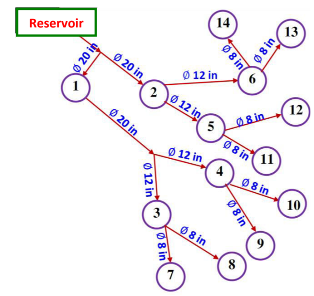

Fig. 4. Simulation topology, red arrows show the direction of water-flow [3].

We study the proposed rate adaptation scheme in *Castalia* [19], 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 Fig. 4. 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.

We model the harvested energy arrival from water-flow based on the average water usage pattern, taken from [20]. The total daily usage is 169 ± 10.6 gallons. Reference [21] reports the maximum water velocity in real systems as 7.5 ft/sec. We conservatively assume that for the third level nodes have a water-velocity of 5.0 ft/sec at peak hours and compute those for other two layers using the flow continuity relationships.

The sink node broadcasts the assigned rates every $T = 6$ hr (interval time). The intervals are chosen to be 11PM - 5 AM, 5 AM - 11 AM, 11 AM - 5 PM, 5 PM - 11 PM, such that the harvested energy profile within the intervals are similar. The beacon interval of the sensor nodes is assumed to be 30 minutes. 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:

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

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 4 in our application.

We assume harvesting efficiency $\eta_e = 10\%$. 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 2.7V for all nodes. The super-capacitor leakage power is calculated as $P_0.\exp(a.V_c)$ [22], 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% [22]. The sampling/transmission is stopped, whenever the capacitor voltage goes below τ = 1.2 V. The super-capacitor stops supplying power to the sensor node below 0.9 V, which is considered as very low voltage.

We model both our scheme, 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) [23]. 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 [24], 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 [23]. 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 II reports some of the commercial water quality sensors to measure the corresponding 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. Parameters used for simulations are listed in Table III.

TABLE II. DIFFERENT SENSOR SPECIFICATIONS

Parameter	Sensor	Voltage tange	Current draw
Chlorine	Chlorine sensor Type 8232 [25]	12-30 V	4 mA
ORP	WQ600 [26]	10-36 V	0.2 mA + sensor output
pH	WQ201 [26]	10-30 V	5.5 mA + sensor output

TABLE III. SIMULATION PARAMETERS

Var	Values	Var	Values	Var	Values	Var	Values
P_{Bt}	1000 mW	T_{Bt}	280 ms	P_{Br}	200 mW	T_{Br}	280 ms
P_{Dt}	1000 mW	T_{Dt}	280 ms	P_{Dr}	200 mW	T_{Dr}	280 ms
P_P	200 mW	T_P	3 ms	P_S	Cl: 48 mW ORP: 202 mW pH: 255 mW	T_S	400 ms

Fig. 5 shows the mean energy profile of nodes from normal water flow over 24 hrs 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.

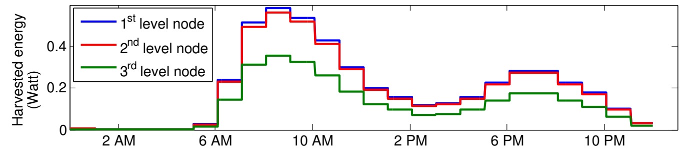

Fig. 5. Mean energy harvested over time for different nodes [3].

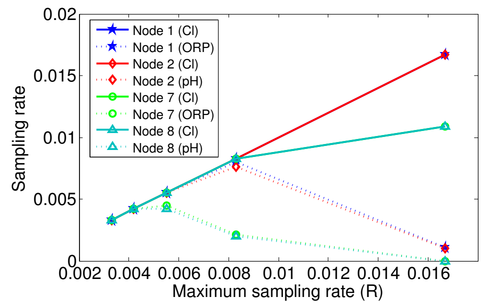

Fig. 6. Comparison of sampling rates of different sensors.

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 [27]. Our main objective is to show the effect of adaptation at the low energy hours, i.e. from 11 PM to 5 AM. Fig. 6 shows the variation of the assigned sampling rates with different $\mathbb{R}$ for 11 PM to 5 AM. From Fig. 6 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. Fig. 7 shows the overall utility of the coalition, which is an increasing function of $\mathbb{R}$. As $\mathbb{R}$ increases, the sampling rates of the individual nodes increase, which enhances the net coalition utility or the overall event reporting ability.

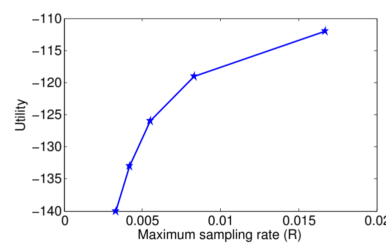

Fig. 7. The net utility function of HCS.

Effect of adaptation on event reporting time: Fig. 8 shows effects of the collaborative adaptation on the event reporting time of the coalition. We assume $\mathbb{R}$ to be 0.0167 (1 sample/minute) for Fig. 8-Fig. 9. From Fig. 8(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

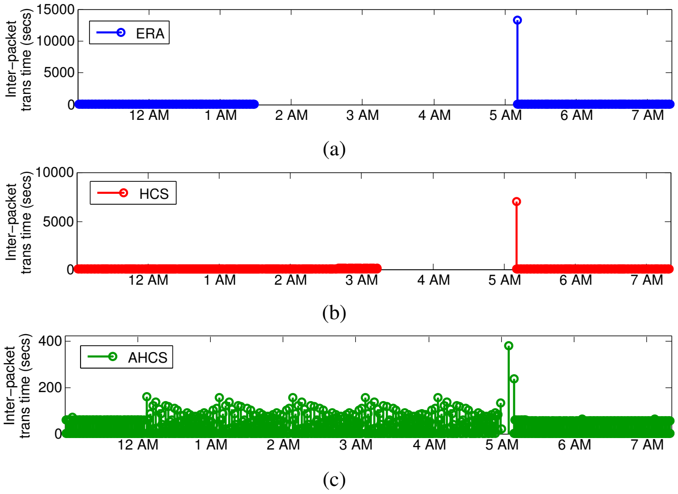

Fig. 8. Comparison of event reporting time (a) without adaptation (ERA), (b) with adaptation (HCS), and (c) with adaptation (AHCS).

reporting time is reduced to about 1 hour 45 minutes. We notice that the reporting time is still high. Thus 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 one hour, and adapt the sampling rates by considering the energy wastage due to lack of storage, as mentioned in section V. This reduces the reporting time to about 6 minutes as seen from Fig. 8(c). Fig. 9 shows the sampling rate of the nodes at different time of night. From this figure, we can observe that at relatively high 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.

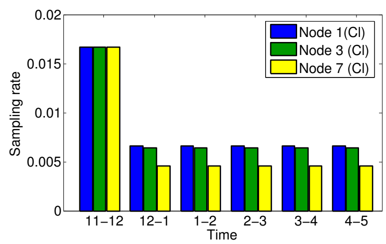

Fig. 9. Comparison of sampling rates at different time scales at lull hours.

VII. Related Work

Wireless sensor networks for pipeline monitoring is well-mined, but most of them require manual access to the pipes and thus not very useful for much of water distribution system. For example, PipeNet [4], MISE-PIPE [5] involve sensor deployment along pipe length on the outside whereas SPAMMS [28] requires RFID tags painted inside of the pipe. In [29] the authors propose a self maneuvering robot going through the pipes and monitors the leaks. Contrary to these approaches, we focus on techniques that do not require any ground digging or even following of buried pipelines from above-ground, which itself may be impractical. Different mathematical models are explored for contamination events detection in [24], [30], [31].

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 [32], [33], [34], [35], [36], [37]. Other techniques for reducing energy consumption include data compression and source coding [38], [39], transmit power control [40], [41], [42], [43], multiple channel assignment [44], [45], [46], [47], [48], [49] etc. While these proposals are mainly motivated towards maximizing the life-time of the sensor network, our objective in this paper 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 [50] and [51], 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 in [16] and [17] 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. 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. Some centralized and collaborative rate adaptation schemes are reported in [3], [52]. Contrary to these literatures, in this paper we consider the presence of multiple and different types of sensors per node, as well as their inter-dependencies 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.

VIII. Conclusions

In this paper, we have explored water flow driven sensor network that monitors and identifies abnormal water quality and contamination, to generate an alert. We showed that the scheme can significantly reduce the contamination detection time during periods of low water flow. We also motivated the advantage of collaborative sampling within a coalition, and we plan to explore this aspect along with the optimal coalition formation in more detail in the future. We also want to bring the contamination propagation latency into account, with an objective of adapting the sampling rates to minimize the event reporting time at the event of a contamination. Usually a real WDS consists of a large number of pipes interconnected with each other following the street layouts, which forms a large number of pipe loops, as shown in Fig. 1. We want to study the contamination propagation characteristics in such a loop-based pipe network. Another important concern of WDSN is its security [53], [54], even if the sensor nodes are considered to be physically secured. To make the WDSN dysfunctional, an adversary can eavesdrop to acquire secure network information or can inject interfering or jamming signals, which impedes the wireless communications and at the same time depletes the receiver's super-capacitors. Making the resource constrained WDSN resistant to these security concerns is one of our future research endeavors. We also plan to design a proof-of-concept experimental setup to demonstrate the feasibility of our assumed energy harvesting model along with the adaptation benefits.

References

[1] "Smart sensors network for water quality monitoring. a way to re-invent water supply management?" 2012.

[2] T. P. Lambrou, C. C. Anastasiou, C. G. Panayiotou, and M. M. Polycarpou, "A low-cost sensor network for real-time monitoring and contamination detection in drinking water distribution systems," *IEEE Sensors Journal*, vol. 14, no. 8, pp. 2765–2772, 2014.

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

[4] I. Stoianov, L. Nachman, S. Madden, and T. Tokmouline, "Pipenet: A wireless sensor network for pipeline monitoring," in *IPSN*, 2007, pp. 264–273.

[5] Z. Sun, P. Wang, M. C. Vuran, M. Al-Rodhaan, A. Al-Dhelaan, and I. F. Akyildiz, "MISE-PIPE: magnetic induction-based wireless sensor networks for underground pipeline monitoring," *Ad Hoc Networks*, vol. 9, no. 3, pp. 218–227, 2011.

[6] D. Moss, J. Hui, and K. Klues, "Low Power Listening, Core Working Group, TEP 105."

[7] M. Khan, M. Iqbal, and J. Quaicoe, "River current energy conversion systems: Progress, prospects and challenges," *Renewable and Sustainable Energy Reviews*, vol. 12, no. 8, pp. 2177–2193, 2008.

[8] S. M. George, W. Zhou, H. Chenji, M. Won, Y.-O. Lee, A. Pazarloglou, R. Stoleru, and P. Barooah, "Distressnet: a wireless ad hoc and sensor network architecture for situation management in disaster response," *IEEE Communications Magazine*, vol. 48, no. 3, pp. 128–136, 2010.

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

[10] P. W. Gething and A. J. Tatem, "Can mobile phone data improve emergency response to natural disasters?" *PLoS Medicine*, vol. 8, 2011.

[11] W. Shan, J. Feng, J. Chang, F. Yang, and Z. Li, "Collecting earthquake disaster area information using smart phone," in *IEEE ICSSE*, 2012, pp. 310–314.

[12] C. Yang, J. Yang, X. Luo, and P. Gong, "Use of mobile phones in an emergency reporting system for infectious disease surveillance after the sichuan earthquake in china," *Bulletin of the World Health Organization*, vol. 87, pp. 619–623, 2009.

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

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

[15] www.bentley.com/en-US/Products/WaterGEMS/Features-List.htm.

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

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

[18] P. R. Bhave and R. Gupta, "Analysis of water distribution networks," 2006.

[19] "Castalia: A Simulator for WSN," castalia.npc.nicta.com.au/.

[20] "Embedded Energy in Water Study 3: End-use Water Demand Profile Final Research Plan," uc-ciee.org/downloads/Eeiswtudy3.pdf, 2009.

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

[22] C. Renner, J. Jessen, and V. Turau, "Lifetime Prediction for Supercapacitor-Powered Wireless Sensor Nodes."

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

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

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

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

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

[28] J. Kim, G. Sharma, N. Boudriga, and S. S. Iyengar, "SPAMMS: A sensor-based pipeline autonomous monitoring and maintenance system," in *COMSNETS*, 2010, pp. 1–10.

[29] C. Choi and K. Youcef-Toumi, "Robot design for high flow liquid pipe networks," in *IROS*, 2013, pp. 246–251.

[30] H. Che, S. Liu, and K. Smith, "Performance evaluation for a contamination detection method using multiple water quality sensors in an early warning system," *Water Quality Control and Management*, vol. 7, no. 4, pp. 1422–1436, 2015.

[31] D. Hou, H. He, P. Huang, G. Zhang, and H. Loaiciga, "Detection of water-quality contamination events based on multi-sensor fusion using an extented dempstershafer method," *Measurement Science and Technology*, vol. 24, no. 5, 2013.

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

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

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

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

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

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

[38] C. Tang and C. Raghavendra, "Compression techniques for wireless sensor networks," pp. 207–231, 2004.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

[53] E. Shi and A. Perrig, "Designing secure sensor networks," *IEEE Wireless Commun.*, vol. 11, no. 6, pp. 38–43, 2004.

[54] A. Perrig, J. A. Stankovic, and D. Wagner, "Security in wireless sensor networks," *Commun. ACM*, vol. 47, no. 6, pp. 53–57, 2004.