SENSOR NETWORKS FOR MONITORING WATER SUPPLY AND SEWER SYSTEMS: LESSONS FROM BOSTON

Ivan Stoianov¹, Lama Nachman², Andrew Whittle³, Sam Madden⁴, Ralph Kling⁵

¹ Research Associate, Imperial College London, UK, ivan.stoianov@imperial.ac.uk
² Senior Research Scientist, Sensor Network Operations, Intel Research, Santa Clara, CA, USA
³ Professor, Department of Civil and Environmental Engineering, MIT, Cambridge, MA, USA
⁴ Assistant Professor, Department of Electrical Engineering and Computer Science, MIT, Cambridge, MA, USA
⁵ Director, Sensor Network Operations, Intel Research, Santa Clara, CA, USA

Abstract

Recent developments in wireless sensor networks (WSN) promise to have significant impact on a broad range of applications relating to environmental monitoring, structural health monitoring, security and water safety. The convergence of the Internet, telecommunications, and novel information technologies with techniques for miniaturisation now provides vast opportunities for the application of low-cost monitoring solutions which could drastically increase the spatial and temporal resolution of environmental data.

The paper describes the development of a prototype monitoring system which bridges advances in wireless sensor networks with advances in hydraulic and water quality modeling. The prototype monitoring system was deployed at Boston Water and Sewer Commission (BWSC) in December 2004, and it has been successfully collecting and charting near-real time hydraulic and water quality data as well as water levels in combined sewer outflows (CSO). The remote monitoring system has unique functionalities in terms of sampling rates (up to 1000 S/s), time synchronization (up to 1 ms) and in-network processing. These features create novel opportunities for wirelessly collecting data for applications such as hydraulic pressure transients, remote acoustic leak detection together with low-duty cycle applications such as monitoring water quality parameters and water levels in CSOs.

The trial with BWSC has been tremendously useful to prototype hardware and software tools, and to identify deployment and operational challenges in using sensor networks for monitoring and management of large scale water supply systems.

Keywords
Industrial application of sensor networks, Water supply systems, Real-time monitoring and embedded systems

1. INTRODUCTION

Monitoring large scale urban infrastructure such as water supply and sewer networks for detecting leaks, changes in water quality and preventing water contamination caused by sewer overflows has the potential to save municipalities millions of dollars a year and bring significant social benefits by reducing public health hazards. In the US alone, there are approximately 160,000 public drinking water systems that comprise around 700,000 miles of water distribution mains (EPA 2005b). A recent study carried out by the US Environmental Protection Agency estimates that community water systems need $277 billion over the next 20 years (2003-2023) to install, upgrade, and replace infrastructure (EPA 2005a). Transmission and distribution projects represent the largest category of this estimate with $184 billion in needs. The problems of aging and failing infrastructure have been further exacerbated with the threat of contaminant
intrusion due to leaking pipes (Friedman et al. 2005) or malicious human actions. These operational challenges and public health threats are major incentives encouraging the development of new technologies for in-line monitoring systems that can optimize operation of the large scale supply networks, prolong service life, evaluate performance and improve the security of water supply to customers. The integration of near real-time data with accurate analytical models can be used in a variety of applications ranging from optimization of pump scheduling (efficient power management), to the detection and quantification of leaks, and the implementation of an early warning system for contaminant intrusion. To implement these critical applications, the water utilities require a large number of spatially distributed measurement points to represent accurately the complex, highly non-linear temporal and spatial processes that occur in water supply and sewer systems.

In current practice, continuous data collection is limited to a small number of high-risk, high-cost measurement locations that collect hydraulic data, and control the status of pumps and valves. The acquired flow data are used for billing purposes (e.g. data from Automatic Meter Readers, AMR) and are hardly ever integrated into monitoring systems which dynamically model the stochastic processes that occur in water supply and sewer systems. Water quality sampling in distribution networks is generally done through grab samples (i.e., single point in time) that are either analyzed on-site or returned to a laboratory. Water quality can also be monitored remotely through recent developments of multi-parameter sensors which measure surrogate parameters such as pH, redox, conductivity and DO (e.g., Censar; http://www.censar.com/; and Hach, http://www.hach.com/) which are interfaced to a logger with wireless communication capabilities. The main reason for the limited level of continuous monitoring is the prohibitively high price of traditional telemetry systems and in some case the high cost of ownership.

In this paper we present the development and field validation of a generic monitoring solution tailored to the specific needs of the water industry which builds upon recent advances in wireless sensor networks. We have chosen two major applications for the development and evaluation of our monitoring solution. These are (i) hydraulic and water quality monitoring of water transmission and distribution systems which also includes capturing fast pressure transient events; and (ii) monitoring the water level in sewer collectors and combined sewer outflows. Recent laboratory experiments carried out at MIT further extended the list of applications by demonstrating how the system could be used for remote acoustic leak detection (Tokmouline, 2006). A key challenge was the integration of these different applications in terms of bandwidth and sampling regimes within a generic wireless data collection network. To the best knowledge of the authors, this is the first deployment of a monitoring solution in the water industry that is capable of remotely capturing hydraulic pressure transients and displaying raw high-frequency data in near real-time.

We formulate the technical requirements and develop a complete solution which includes sensors, wireless data collection system, middleware and back-end applications for data analysis. As this was a proof-of-concept, we were primarily interested in the following objectives:

- Outline of the requirements for the wireless data collection system. The applications which we cover are the most demanding ones within the water industry in terms of sampling rates, bandwidth and operational environment. Therefore, we are confident that the presented system can satisfy wide scope of monitoring needs within the water industry and beyond;
- Evaluate the cost of deployment, maintenance and ownership;
- Assess the reliability and robustness of sensors and wireless sensor nodes under extreme environmental conditions (e.g. in sewer collectors);
- Evaluate the performance of the deployed platform in terms of reliable data transfer, processor speed and network bandwidth in dense urban environment;
- Learn from operating the data collection network over a long period of time (15 months) in collaboration with Boston Water and Sewer Commission (BWSC) under real-life conditions.

The remainder of this paper is organized as following. Section 2 presents an overview of current data acquisition practice within the water industry. Section 3 outlines the applications and the motivation for applying wireless sensor networks. Section 4 outlines the architecture of the monitoring system and
describes the critical components. In Section 5, we describe the deployment at Boston Water and Sewer Commission and present the results of our preliminary data analysis. In Section 6, we summarize the lessons learned.

2. CURRENT DATA ACQUISITION PRACTICE

There is a variety of telemetry solutions which the water utilities are using and frequently these solutions are integrated into SCADA systems (Supervisory Control and Data Acquisition). The schematic of a typical SCADA system is shown in Figure 1, and it has four major components that are interconnected via a network: (i) remote telemetry and automation devices, such as outstations, data-loggers and PLCs (Programmable Logic Controller); (ii) data gatherers which acquire and manage the telemetry data; (iii) data server providing telemetry data for users and other applications; and, (iv) workstations which provide a user interface.

![Figure 1: Schematic of a SCADA system](image)

The outstations are connected to the data gatherers via a range of different media including telephone lines (PSTN lines, cell phone modem), leased lines, radio (UHV, VHF, spread spectrum), private networks, fieldbuses (e.g. Profibus), and satellite. The workstations communicate with the data gatherers via local and wide area networks as appropriate, such as X25, Ethernet or asynchronous links. Communications interfaces between workstations, data gatherers and corporate systems are provided through Industry Standard Protocols such as TCP/IP or OSI standards. The data gatherers (DGs) provide the data collection service at the heart of the system by scheduling and executing telemetry polling, managing and distributing the real time database, and serving the workstations. The outstations are grouped into sets (clusters) and then each set is interfaced to two DGs: a primary and a secondary, to minimize the risk of failure. During normal operation, the primary DG polls the set’s outstations and collects the corresponding data. This data is forwarded to the secondary DG where a further copy of the set’s database is maintained.

Current SCADA systems are expensive and their deployment within the water industry is limited to critical sites. Many SCADA protocols are vendor specific and proprietary as the legacy of the early low-bandwidth protocols remains. In general, SCADA systems serve low-data rate applications (e.g. collecting data once every 15 minutes; or only when the DG polls the outstation) and provide little flexibility in terms of changing sampling regimes, adaptive sampling, local processing and remote (over-the-air) software update. Many SCADA protocols now contain extensions to operate over TCP/IP, although many utilities prefer not to connect SCADA systems to the Internet for security reasons.

There is a growing need for monitoring solutions that can be deployed at much lower cost and faster using in-house expertise while providing much higher spatial and temporal density. These novel monitoring solutions are expected to complement traditional telemetry and SCADA systems while
generating a high level of monitoring redundancy. As an illustration of this trend, many sensor vendors have started to offer embedded cellular connectivity in their products. For example, ABB provides GSM/SMS connectivity to its FieldIT AquaMaster water meters, enabling information to be remotely collected via SMS messaging (www.abb.com). The AquaMaster flowmeters can be remotely configured by sending an SMS message. Data are being recorded at predefined intervals of 15 minutes; with an option for a high resolution one minute sampling rate. The data are transmitted once every 24 hours for battery operated units which have a projected battery life of 5 years. Data can also be pulled on demand by dialing an individual sensor and collecting sensor data (flow rate, pressure, total water consumption, alarms) and status information such as battery level. The battery life is reduced to two months for a data collection and communication of once every 5 minutes; and to approximately 6 months for data collection and communication every 15 minutes. Using commercially available SMS Gateway solutions, the automated SMS meter readings can then be received, decoded and exported to an existing billing application or database to provide near real-time usage information via Internet. Wireless communication is also supported via packet switched (e.g. 1xRTT) or circuit switched (e.g. CDMA) cellular (http://www.telog.com/).

The wide-scale adoption of these communication solutions illustrates that the water industry is actively looking for novel low-cost monitoring solutions. The bandwidth however remains limited as the data are primarily used for billing with limited use for near real-time monitoring. Capturing high-frequency data is exclusively done via manual data collection.

3. APPLICATIONS

The applications which we selected for the development and evaluation of our monitoring solution include (i) hydraulic and water quality monitoring of water transmission and distribution systems (this also includes capturing fast pressure transient events); (ii) remote acoustic leak detection including remote cross-correlation; and, (iii) monitoring the water level in sewer collectors and combined sewer outflows. The reasons we concentrated on these applications are as following:

- the monitored infrastructure, water transmission, distribution and sewer pipes, share the same spatial distribution. Figure 2 shows the complexity and density of underground infrastructures for a street in London, UK. All large cities worldwide face the same challenges in using the streets as conduits for utilities to transport water, gas, electricity and telecommunication services;
- water distribution and sewer pipes are frequently located within close proximity. Figure 3 shows a broken sewer pipe located on top of a leaking water pipe. Under certain hydraulic (pressure) conditions, the leak can become an entry point for the intrusion of contaminants which might introduce significant public health hazards;
- the pipeline infrastructure is generally operated by a single company (water utility) and having the opportunity to address all these applications within a generic monitoring system provides excellent opportunities to enable an integrated modelling and management approach while keeping the cost for the monitoring solution low; and,
- the lead author has significant expertise in modelling and operational control of water supply systems. A number of problems were identified through extensive field deployments using custom built data loggers for time synchronized data collection of hydraulic transients (Stoianov et. al. 2003a). Extensive research was carried out to address these problems. The practical implementation of these solutions, however, was hindered by the costly manual data collection and the technological limitations of current telemetry solutions.

Near real-time hydraulic and water quality monitoring in water supply and transmission systems are essential for detecting failures (such as leaks and bursts), optimizing operational control, pump scheduling, chlorination and chlorine residual and implementing an early warning system for contaminant intrusion. The monitoring process is highly dependent upon the density of the measurements and the accuracy of the simulation model. The hydraulic model which solves a system of non-linear equations approximates the network behaviour by calculating pipe flows, velocities, head-losses, pressures and heads, reservoir levels and reservoir inflows. State estimation techniques are well suited for the purpose of on-line monitoring as they allow tracking the time varying flows and pressures. These techniques are frequently used in the electrical and gas industries, but the scarce number of monitoring points precludes their use in the water industry. State estimation is defined as the computation of the minimum set of values necessary to completely describe all other pertinent variables in a given system from some measurement data. The state estimator algorithm maps the available new information from measurements into a state-space using an over-determined set of equations. This is typically formulated as a projection resulting in a minimization problem. The principles of hydraulic modelling and state estimation can be extended for modelling water quality parameters. Figure 4 outlines a monitoring system that is designed to use near real-time data coupled with accurate hydraulic and water quality models for detecting and tracing a contamination event. In this example, simulations were carried out to demonstrate how hydraulic data from pressure sensors and flow meters can be combined with water quality data obtained from multi-parameter water quality sensors such as pH, dissolved oxygen, conductivity and free chlorine. The developed model simulated the spatial spread of an introduce contaminant at time 0, 2 hours, 4 hours and 24 hours. The data are then projected over the GIS (Geographical Information System) and used for minimizing the effects of contamination.

The sampling regimes are split into a **continuous** (periodic) mode and **burst** mode. The sampling regimes and rates that were defined for this application can be summarized as following:

- **Continuous mode**: Collect for $A$ seconds (e.g. 5, 10, 15 seconds specified remotely by the user) every $B$ minutes (e.g. 1, 5, 15 minutes specified remotely by the user) with a SR of $C$ S/s (e.g. 1, 10, 100, 1000, 2000 S/s specified remotely by the user). The outputs include average, minimum, maximum, and standard deviation;
- The acquired data are communicate to gateway once every $D$ minutes (e.g. as collected, 1, 5, 15, 30, 60 minutes, 6 hours, 12 hours, 24 hours or when a threshold is exceeded – these options are specified remotely by the user and can be changed in near real-time);
• Complete remote control (bi-directional) to change the collection regime, sampling rate and communication frequency;

• Adaptive sampling for a particular parameter. If the data exceeds a pre-determined threshold then sampling rate is increased while the communication intervals are decreased. For example, if pH goes above 9.5, then collect data every \( E \) minutes (e.g. 1 minute) and communicate data to a data gatherer every \( F \) minutes (e.g. 5 minutes);

• **Burst Mode**: sampling rate of 1000 S/s burst mode over a period of 5 minutes. This will be performed under burst demand request from the server at a pre-defined Start Time. Minimum 15 minutes will be allowed for the server to send the Start Time for the burst data collection mode to the sensor nodes. The acquired data are used for a sophisticated analysis and modelling of hydraulic transients for detecting failures in air valves and large bursts, and fine tuning of control valves in large diameter transmission pipelines (Stoianov et al. 2003b);

• The acquired high-frequency data are compressed in near real-time using lossless data compression algorithms to reduce communication time and power consumption as the sensor nodes are battery operated;

• The acquired burst mode data will be time synchronized between sensor nodes located in separate clusters to 1ms; and,

• The remote sensor nodes are re-programmed remotely (software/firmware update over the air).

![Figure 4: Simulation of the spread of a contaminant over time](image)

### 3.2. Remote Acoustic Leak Detection. Sampling requirements.

Acoustic Emission and vibration signals have been widely used as a non-destructive testing (NDT) technique for detecting and locating leaks in pipes. Generally, a leak generates noise due to the rapid release of energy which results into a transient elastic wave. To perform leak detection, vibration or acoustic signals are manually acquired at two access points using sensors such as accelerometers or hydrophones on either side of the location of a suspected leak (Figure 5).

If a leak exists, a distinct peak may be found in the cross-correlation of the two signals \( s_1(t) \) and \( s_2(t) \). This gives the time delay \( r_{\text{peak}} \) that corresponds to the difference in arrival times between the signals at each sensor. The location of the leak relative to one of the measurement points, \( d_i \), can be calculated using a relationship between the time delay \( r_{\text{peak}} \), the distance \( d \) between the access points, and the propagation wavespeed \( c \) in the buried pipe.
If $s_1(t)$ and $s_2(t)$ are two stationary random signals with zero mean, the cross-correlation function is defined by (Gao et al. 2006; Oppenheim et al. 1986)

$$R_{s_1s_2}(\tau) = E[s_1(t)s_2(t + \tau)]$$

where $\tau$ is the time lag and $E$ is the expectation operator. The value of $\tau$ that maximizes the equation provides an estimate $\tau_{\text{peak}}$ of the time delay. A procedure to calculate the cross-correlation function using sampled data is illustrated in Figure 6. The cross-correlation estimator can be obtained from the inverse Fourier transform of $X_1^*(\omega)X_2(\omega)$ and scaled appropriately for normalization, $X_1(\omega)$ and $X_2(\omega)$ are the Fourier transforms of $s_1(t)$ and $s_2(t)$ (and $*$ denotes complex conjugation).

Commercial products such as MLOG offered by Flow Metrix Inc (http://www.flowmetrix.com/) and Phocus2 offered by Primayer (www.primayer.co.uk) provide functionalities for remote and drive-by data acquisition. Various processing algorithms are used to locally analyze the noise characteristics to provide status information which is defined as leak, possible leak, and no-leak. While these products facilitate unattended night time data collection (during hours of low background noise) and approximate identification of leaking areas, they still require manual intervention for accurately pinpointing leaks (the cross-correlation).

The data collection system presented in this paper can provides functionalities that go beyond the listed commercial systems by enabling both local processing of status information and centralized pair-wise
data processing of high-frequency time synchronized data. The developed data collection and processing system can provide significant benefits for monitoring high-consequence pipelines and areas. The sampling regime requirements for the remote acoustic leak detection can be summarized as following:

- **Burst data** collection is carried out $G$ times per 24 hours (e.g. $G = 4$, specified remotely by user);
- Collect data for $H$ minutes (e.g. $H = 5$ specified by user) with a SR of 1000 S/s (2000 S/s is supported in the new version of the hardware);
- Process data to identify a status (Leak, Possible Leak, No Leak, DoNotKnow). The local data analysis includes time-frequency algorithms together with a classification algorithm. The processing algorithms perform real-time processing on the sensor node (mote). This requires the development of middleware for plugging computational routines which can be remotely queried and updated (over-the-air software update);
- Time-synchronize (time stamp) acquired data with accuracy of 1 ms; and,
- Communicate status information. If status information differs from NoLeak status, then transfer high-frequency data to a central server and carry out pair-wise cross-correlation.

### 3.3. Monitoring Combined Sewer Outflows. Sampling requirements.

Combined sewer systems are sewers that are designed to collect rainwater runoff, domestic sewage, and industrial wastewater in the same pipe. Most of the time, combined sewer systems transport all of their wastewater to a sewage treatment plant, where it is treated and discharged to a water body. During periods of heavy rainfall, however, the wastewater volume in a combined sewer system can exceed the capacity of the sewer system or treatment plant. For this reason, combined sewer systems were designed to overflow and discharge excess wastewater directly to nearby streams, rivers, or other water bodies. These overflows, called combined sewer overflows (CSOs), are among the major sources for water quality impairments as the discharge contains not only storm water but also untreated human and industrial waste, toxic materials, and debris. They are a major water pollution concern for 772 large cities in the U.S. (typically older communities) that have combined sewer systems (EPA 2006).

Combined sewer systems could greatly benefit from real-time control (RTC) which is a custom-designed computer-assisted management system that is activated during a wet-weather flow event. Though uses of RTC systems had started in the mid 60s (EPA 1974), recent developments in wireless sensor networks, telecommunication, instrumentation, and automation are turning RTC into a viable solution. RTC management provides a cost-effective solution in comparison to construction projects designed to separate combined sewers in urban areas. RTC systems are designed to perform a variety of management functions in a given sewerage system such as routing flows to a treatment plant, or other designated points; control flooding, overflows, or surcharges; maximize storage space; optimize treatment plant capacity; prevent operational problems; and, protect receiving waters. Field (2000) defines the basic components of RTC systems as sensors, automated gates and strategies. The RTC equipment includes measurement devices for water level, flow, rainfall intensity and sometimes pollutant concentration, and regulators for pumps, gates and weirs. The reliability of the RTC equipment, calibration and maintenance present significant challenges as the monitoring equipment is subjected to extreme fouling, corrosion and frequently placed in not easily accessible locations. Furthermore, the equipment needs to be intrinsically safe as it could potentially cause ignition of gases in the sewer atmosphere.

In this study, we only demonstrate measuring reliably water level in sewer collectors. The equipment is generic to allow the interface of additional sensors. The sampling regime requirements for monitoring combined sewer outflows can be summarized as following:

- Use multiple sensors to create hardware redundancy for reliable monitoring and sensor fault identification;
- Periodic mode of data collection: Collect data for $I$ seconds (e.g. $I = 10s$ specified remotely by user) every $J$ minutes (e.g. $J = 5mins$ specified by user) with SR of 1 S/s (outputs include average, minimum, maximum, and standard deviation);
Communicate acquired data to gateway every $K$ minutes (e.g. $K = 15$ mins);
Adaptive sampling: If the collected data exceeds a user-specified threshold, then start collecting data once every minute and communicate the data at 5 minutes intervals until the level drops below the threshold;
Use radar-measured precipitation and/or data from rain gauges to change the sampling regime;
All data collection and communication parameters are specified remotely by the user.

4. WIRELESS MONITORING SYSTEM: SYSTEM ARCHITECTURE

The application specifications listed in Section 3, place demanding requirements for the wireless monitoring system in terms of bandwidth, long-distance communication and accurate time synchronization across a wide spread monitoring system, and local data processing. Furthermore, many of the sensor locations do not have access to power, and rely on battery operation. Therefore, the major challenge in developing the wireless monitoring system is how to balance the conflict between long-distance communication, bandwidth, local data processing and the constraints for low-power consumption.

To better address these challenges, a prototype hierarchical wireless monitoring system was developed. The schematic of the system with its main components (sensors, communication, middleware and back-end) are presented in Figure 7. The system consists of a three-tier (subsystems) communication structure which utilizes a cluster-based power management protocol, and a reliable bulk transport. In this way, the subsystems work together to coordinate periodic and burst data collection across a large number of sensing points while maximizing sleep time. The first tier contains energy-constrained sensor nodes with low transmission range which form clusters. The data from the sensor nodes are transmitted to local data gatherers which compose the second tier. The data gatherers are not energy constrained as these can be installed at street lights, illuminated street signs and bollards, or equipped with solar panels. This setup eliminates the need of digging up the pavements and maintaining power cables which is costly and risky particularly in dense urban environment. The data gatherers combine cluster head nodes which control the sleep schedule of each sensor node and a gateway (an industrial single board computer – SBC). The gateway initiates and controls the long-range communication to a central server via TCP/IP over GPRS (General Packet Radio Service). Secure Shell network protocol SSH-2, (Barrett et al. 2005) is used to establish a secure tunnel between the gateway and the server for bi-directional communication. The SSH protocol guarantees confidentiality and integrity of the data exchanged between the gateway and the server using public-key cryptography and message authentication codes. A data control center on the back-end stores and process data on a server at MIT, and displays the acquired data via a web browser (http://db.csail.mit.edu/dcnui/). The application is built on open source web technology, deploying Linux/Apache/PostgreSQL/PHP stack in a client-server model. Google Maps Programmer’s toolkit (API) was used to build geospatial viewing tools for the deployed monitoring locations (Figure 8). This open source framework facilitates a rapid application development at low cost. As the user interface is just a common browser window, it runs on any computer and on a hand-held device which enables quick data interrogation and validation by office and field engineers. A basic set of additional functionalities were added to the control center such as charting near real-time data and status information, alarm notification via email and SMS messages based on thresholds, executing pre-defined queries on historical data and account management to authorize users to view acquired data, and interface between PostgreSQL database and Matlab.
The following sub-sections provide a brief overview of the system components and operation of the first two tiers of the data collection network.

4.1. Tier I: Sensor Nodes

In the proposed two-tier monitoring setup, the transmission range requirement of the sensor units is within 10-100m distance as they communicate with the data gatherer. At this initial stage of our development, we decided to use Bluetooth (2.4 GHz license-free ISM band) as a choice for short-range communication within a cluster because of the RF-method of FHSS (Frequency Hopping Spread Spectrum) which makes it more robust in outdoor environment, the high transfer data rate (1 Mbps), low cost and the application/cluster requirements. The cluster formation includes small number of nodes with one to two hops exchanging periodic or burst data over relatively short periods of time. The choice of the radio is particularly important as it impacts not only energy consumption, range and reliability of data transmission (quality of service) but also the software design (e.g., network self-assembly, multi-hop routing, time synchronization and in-network processing).

The applications under consideration relied on the implementation of computationally intensive real-time processing of high-frequency data which required more advanced microprocessor architectures for the sensor nodes while maintaining low power consumption. These requirements were successfully addressed.
by a novel sensor node platform developed by Intel Research (Kling 2003). The first version of the Intel Mote which we deployed in the Boston Water trial is built on a 3x3cm circuit board that integrates a wireless microcontroller module (32-bit ARM7TDMI processor running at 12MHz, 64kB of Ram, 512 kB of FLASH and a CMOS Bluetooth radio) and various digital I/O options using stackable connectors. The connectors expose two UART ports (up to 960 kb/s) which support very high sampling rates, e.g. 16 bit data at 20kHz, USB client, GPIOs and power. The radio’s range is approximately 30 meters with the built in antenna, however, we were able to extend the range up to a 100 meters using a custom-built external antenna. We used TinyOS (http://www.tinyos.net/) as an open source operating system for the Intel Mote. The OS and radio stack leave about 11 KB of free SRAM to be used by the application. Another advantage of the Intel Mote is its modularity which allows custom sensor boards, interface boards and debug boards to be attached to the system in a flexible manner.

Key components in the tiered communication structure are the cluster-based power management protocol and the reliable bulk transport of high-bandwidth data. These elements work together to coordinate periodic data collection across the nodes within a cluster while minimizing power consumption and utilizing Bluetooth master/slave and piconet/scatternet operation. Subsequently, the network is self-organizing on start-up by employing a distributed node discovery and connection procedure (Nachman et. al 2005). After establishing the basic network, routing information is exchanged between the nodes to permit automatic network repair in the event of node or link failures, while a low power mode maintains network connectivity. The nodes in a cluster wake up based on a Wake_UP parameter communicated by the cluster head at the end of a previous period. Once the cluster nodes are awake, the cluster head initiates metric-based single-destination-DSDV routing (Yarvis, et.al, 2002) to allow all nodes to find a path to the cluster head. Next, each node sends periodic TraceRoute packets to the cluster head, allowing the cluster head to discover the nodes in its clusters. The cluster head waits a predefined period, to allow all nodes to report. Once discovery is complete, the cluster head sends a data capture and transfer request to each node. The resulting data is transferred using the bulk transfer protocol (Nachman et. al 2005). Once data collection is complete, the cluster head sends beacons for time synchronization using PPS (pulse-per-second) signal provided by an embedded GPS. These functions generally could be performed by many single board computers (SBC). For our field trial we used a research platform developed by Intel called Stargate (http://platformx.sourceforge.net/). The Stargate platform is a 400MHz, PXA55 XScale processor, 64 MB SDRAM, 32 MB Flash with Ethernet, Serial, JTAG, USB, PCMCIA, Compact Flash connectors, Bluetooth and 802.11 (through the CF or PCMCIA slot) running Linux OS Kernel 2.4.19. We used General Packer Radio Service (GPRS) for long-range communication which is available worldwide via GSM cellular networks. For this purpose, we interfaced a Sierra Wireless A750 GPRS modem with the Stargate platform. Furthermore, we added 802.11b (WiFi) connectivity via Netgear MA401 card (CF slot) so that we can locally access the gateway for drive-by data collection and software upgrade. We used Motorola GPS engine M12+ specifically optimized for timing applications.

4.2. Tier II: Data Gather and Gateway

The second tier acts as a cluster head, data gatherer and a gateway which manages the cluster, controls the long-range communication with the remote server and sends time beacons for time synchronization using PPS (pulse-per-second) signal provided by an embedded GPS. These functions generally could be performed by many single board computers (SBC). For our field trial we used a research platform developed by Intel called Stargate (http://platformx.sourceforge.net/). The Stargate platform is a 400MHz, PXA55 XScale processor, 64 MB SDRAM, 32 MB Flash with Ethernet, Serial, JTAG, USB, PCMCIA, Compact Flash connectors, Bluetooth and 802.11 (through the CF or PCMCIA slot) running Linux OS Kernel 2.4.19. We used General Packer Radio Service (GPRS) for long-range communication which is available worldwide via GSM cellular networks. For this purpose, we interfaced a Sierra Wireless A750 GPRS modem with the Stargate platform. Furthermore, we added 802.11b (WiFi) connectivity via Netgear MA401 card (CF slot) so that we can locally access the gateway for drive-by data collection and software upgrade. We used Motorola GPS engine M12+ specifically optimized for timing applications.
5. BOSTON WATER DEPLOYMENT

In December 2004, in collaboration with Boston Water and Sewer Commission we deployed three monitoring clusters as a proof-of-concept (http://db.csail.mit.edu/dcnui/PhotoAlbum/index.html presents a series of photos detailing the installation). The trial which is still running aims to answer a wide range of technical and economic questions such as ease of deployment, cost of installation and maintenance, reliability of data communication, reliability of sensors, and packaging.

5.1. Installation

The three monitoring clusters were selected to represent the applications listed in Section 3. For all three clusters, the gateways were installed at neighbouring lamp posts which provide direct access to power (110V ac power lines). In addition, the gateways have back-up battery with re-charging circuitry which allows one week operation in the case of a power failure. A number of additional design challenges had to be overcome which included packaging, temperature control via heat sinks and water proof ventilation (temperature measured in the gateway enclosure on a sunny hot day reached 140°F), sensors, and antenna design and its installation in the road surface.

**Cluster 1** includes monitoring pressure and pH in a 12” cast-iron pipe which supplies potable water. Data are collected at 5 minutes interval for a period of 30 seconds. The pH probe is warmed up (powered) for a period of 15 seconds, and then readings are taken once per second for 15s. The data are communicated to the data gatherer and the server every 5 minutes. We used a pH glass electrode with Ag/AgCl reference cell for which an immersion apparatus was developed to lower the probe in the pipe through a 1” access point. We also developed a signal conditioning circuitry to condition the output signal to 0.5-4.5 Vdc which corresponds to a pH range of 3-11. The signal conditioning circuitry for the pH probe consumed less than 10mW of power. Significant amount of time and effort were spent on the selection and modification of the pressure sensor. We needed a low-cost sensor (less than 200 USD) with good accuracy (+/-0.3%FS) and long term stability. The most critical parameters however were the start-
up time, the dynamic response for capturing pressure transients and the sensor performance under aggressive power cycling. In order to address this challenge, we used an OEM piezoresistive silicon sensor for which an advanced ASIC compensation technology was developed to achieve accuracy better than +/-0.2%FS including effects of non-linearity, hysteresis and repeatability; start-up time of less than 20ms; fast dynamic response and power consumption of less than 10mW. Pressure data are collected at 5 minutes intervals for a period of 30s with a SR of 600S/s. The raw data are communicated to the data gatherer every 5mins where the data are compressed and send to the server.

Figure 9: Cluster 1: Installation of pH probe; Antenna embedded in the road surface

**Cluster 2** (Figure 10) includes monitoring pressure in a 8” cast iron pipe. Data are collected in a similar way to Cluster 1;

**Cluster 3** (Figure 11) includes monitoring the water level in a combined sewer outflow collector. As this is an aggressive environment, we decided to use hardware redundancy and implement a voting algorithm which identifies sensor failures or drifts. This information will optimize maintenance and increase the reliability of data. For this purpose, we implemented three sensors, two pressure transducers at the bottom of the collector and an ultrasonic sensor on the top. The pressure sensors are low-power devices consuming less than 10mW while the ultrasonic sensor is a high-power device consuming around 500-600 mW. Therefore, we used the pressure sensors for continuous (periodic) monitoring while the ultrasonic sensor was only used to verify the readings from the pressure sensors when their difference exceeded a threshold or when the water level exceeded the weir height. Data from the pressure sensors are collected at 5 minutes interval for a period of 30 seconds. Sensors are powered for 10s before readings are taken with a sampling rate of 1S/s. Both raw data and average data are transmitted to the data gatherer after every data collection.

5.2. Performance

The performance of the data collection network is being evaluated on four criteria: (i) the ability to collect and deliver data to the gateway; (ii) the ability to transfer the data from the gateway to the server via the GPRS link; (iii) the ability to recover from loss or errors; (iv) the long-term performance of the deployed sensors.

During the initial stage (December, 2004 – July, 2005) we observed a series of problems with the gateways ranging from strange GPRS modem power modes to corruption of the Linux kernel. Detail
analysis of these problems identified design faults with voltage regulators and the watchdog timer on the Stargate platform. An external watchdog and automated reset feature were added to the gateway nodes to monitor gateway performance. The gateway is rebooted if the application software halts. In addition, the external watchdog timer reboots the gateway once every 24 hours. Adding these features eliminated the observed problems and reduced the risk of unforeseen problems in the gateway software that would require manual intervention by an operator.

The modified gateways were installed in July 2005 and they have been operating since which is nearly one year of continuous trouble-free operation. The Intel motes have been operating successfully without hardware failures so far.

The communication performance is variable. The packet reception rate for the cluster ranges between 65-85%. The GPRS packet reception rate is within 78-90%, however, all collected data are...
transmitted from the gateway to the server as the gateway archives the data if a connection to the server cannot be established. The sensor node however does not currently have the functionalities to separate data acquisition from communication and if a connection to the cluster head cannot be established then data are not acquired. A newer version of the hardware (Intel Mote v2) has already addressed this limitation. We are also in the process of logging weather conditions to correlate humidity and rainfall to the packet reception rate. Surprisingly, the packet reception rate was high (82%) in January 2005 which was the snowiest month on record in Boston with snow accumulation of greater than 1.5 m.

The battery life (6V 12Ah battery) has been consistent with a duration of around 50-62 days. The Intel mote consumes 2 mA in sleep mode; 16 mA for Intel mote plus pressure sensor and A/D board; around 30 mA for Intel Mote plus radio, sensor and A/D board. This short battery life is due to the very aggressive data acquisition and communication cycles. Separating the acquisition from communication and adopting communication intervals of 15 mins with adaptive data acquisition and storage will increase the battery life beyond one year.

The performance of the pressure sensors exceeded our expectations. The sensors have been operating since December 2004 under extreme environmental conditions. The pH sensor however has required frequent maintenance and replacement ranging from a couple of weeks to six months. The replacement of the pH sensor is under consideration with a micro non-glass ISFET probe.

The monitoring system was successful in accurately capturing several critical events such as the emergency failure of the power supply for the Deer Island Sewage Treatment Plant in Boston on the 15th of October (Figure 12) when approximately 25 million gallons of untreated sewage were released into Quincy Bay (Boston Globe, 17th October, 2005: Untreated Sewage released into bay).

![Figure 12: Sewage release on the 15th of October, 2005](image)

The availability of a larger number of monitoring stations such as the one we deployed could have provided near real time information for utilising the spare buffer capacity of the system thus significantly reducing the discharge volume.

6. CONCLUSIONS

In this paper, we demonstrated how advances in wireless sensor networks, communication and sensing technologies could provide much needed increase in spatial and temporal resolution of hydraulic and
water quality data for better understanding and monitoring large scale water supply and sewer systems. The developed prototype enables us to remotely acquire, view and process both high and low-frequency time-synchronized data from large scale water supply systems. The field trial with Boston Water and Sewer Commission has provided invaluable information about the performance of sensors, sensor nodes, data collection network, radio, hardware and software tools. This information is critical for the current upgrade of the monitoring system in terms of radio, network protocols and application layer.

Finally, we demonstrated the use of a sensor network to meet almost one year of continuous operation requirement with a minimum technical support of replacing batteries every 60 days under extreme outdoor conditions. Several techniques including careful protocol design, external watchdogs and periodic resets of system state enabled sufficient reliability for completely unattended operation. The data collection in this trial was primarily focused on the proof of concept for the communication, reliability of hardware and sensors. We are in the process of extending the trial so that we can acquire data of sufficient quality and quantity to demonstrate the advantages of using the data with much enhanced analytical models.

ACKNOWLEDGEMENTS

This research was carried out while the led Author was a Post-Doctoral researcher at MIT. The work was supported by the Cambridge-MIT Institute (CMI) and by Intel Research. The authors are especially grateful to John Sullivan (Chief Engineer, BWSC) for authorizing the field trial and to Paul Canavan and Tom Bernier (BWSC) for their tremendous help during the installation. Timur Tokmouline provided invaluable advice, time and effort for the web visualisation tools.

REFERENCES


