A Wireless Sensor Network Based Closed-Loop System for Subsurface Contaminant Plume Monitoring *

Qi Han
Dept. of Math and Computer Sciences
Colorado School of Mines
qhan@mines.edu

Anura P. Jayasumana
Dept. of Electrical and Computer Engineering
Colorado State University, Fort Collins, CO 80523
Anura.Jayasumana@Colostate.edu

Tissa Illangaskare, Toshihiro Sakaki
Division of Environmental Science and Engineering
Colorado School of Mines, Golden, CO 80401
{tissa,tsakaki}@mines.edu

Abstract

A closed-loop contaminant plume monitoring system is being developed that integrates wireless sensor network based monitoring with numerical models for subsurface plumes. The system is based on a novel virtual sensor network architecture that supports the formation, usage, adaptation, and maintenance of dynamic subsets of nodes. This automated monitoring system is intended to capture transient plumes to assess the source, track plumes in real-time, and predict future plumes behavior using numerical models that can be continuously re-calibrated by sensor data. This paper presents recent progress made in (1) developing a proof-of-concept study using a porous media test bed with sensors deployed; and (2) developing distributed algorithms for virtual sensor networking.

1 Introduction

Toxic chemicals and biological agents are released into the subsurface as a result of accidental spills, improper disposal, or intentional damage. These releases can occur at a single or multiple locations, some known and some unknown. They cause migrating plumes with concentrations that are spatially distributed and transient. The plume invades previously unaffected or cleaner parts of the environment, posing potential risk to humans and the ecological environment, so temporal and spatial monitoring of the plume concentrations are needed to assess risk, make decisions and take remedial action.

The current monitoring technologies involve the use of pre-determined sampling locations where the plume already is present or is expected to arrive. This ad-hoc placement of sampling ports to monitor three-dimensional spatially complex and transient plumes is often inefficient, expensive and sometimes ineffective.

With recent advances in sensor networking technology and the miniaturization of sensors for sampling contaminants, it has now become possible to significantly expand real-time fine-scale monitoring capabilities. Wireless sensor network (WSN) based monitoring will be able to capture transient plumes to assess the source and predict future plume behavior using numerical models that can be continuously re-calibrated. The primary goal of this project is to develop a real-time closed-loop system integrating wireless sensor data acquisition to numerical flow transport models, with the real-time data from the WSN calibrating the transport models dynamically, and the models controlling the spatial and temporal parameters for data acquisition in the WSN.

2 A Closed-Loop System Architecture

Tight coupling of collected sensor data with numerical models for plume detection and prediction raises several research questions, such as how the WSN should be operated with the knowledge of plumes behavior, how the numerical models should be calibrated with the continuous input of sensor readings, and what the implications of the interaction between WSN and the numerical simulator are. In light of these issues, we propose to develop a unified system architecture (Figure 1). It integrates in a closed loop the sensors that monitor subsurface contaminants, the commu-

*This project is supported in part by NSF CSR grant CNS-0720875.

nication protocols for energy-efficient data acquisition, the inverse models for identification of plume parameters, and the forward models for prediction of plume behavior. More specifically, sensor data are used in an inverse model to estimate the flow and transport parameters in the subsurface of interest. A forward model is fed with the new estimated parameters to make a new prediction of plume movement. The forward and inverse models required for tracking the plumes are extremely complex, requiring high-performance servers to be present in the closed loop. Predictions of direction of plumes are used in the form of configuration instructions to the sensor network to guide more energy-efficient detection and tracking of plumes.

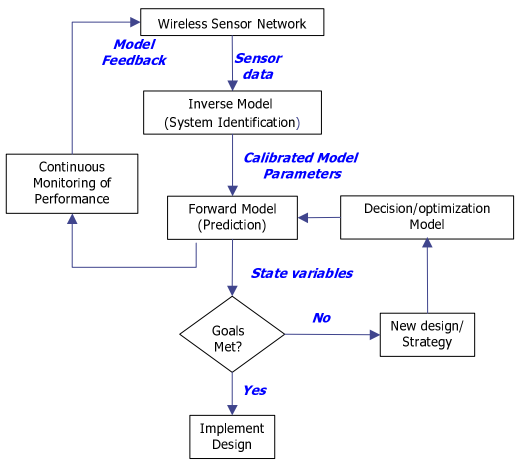

Figure 1. Closed loop operation of WSN and numerical simulator for plumes detection, movement prediction and remediation

The system architecture is supported by two design goals. The first is to develop, evaluate and demonstrate algorithms, protocols and software to support *Virtual Sensor Networks* (VSNs) over Wireless Sensor Networks. A VSN is a subset of sensor nodes of a wireless sensor network, dedicated to a certain task or an application at a given time [4]. The membership of a VSN may change over time. As the nodes in a VSN may be distributed over the physical network, they may not be able to communicate directly with each other. Thus VSN depends on the remaining nodes providing VSN support functionality to create, maintain and operate VSNs. By making available the functionality that would allow a subset of nodes in a WSN to behave as if they were on their own independent sensor network, even though the communication among these nodes occurs via nodes not belonging to the subset, implementation of sensor network applications will be significantly simplified. The VSN concept will seamlessly support tracking of multiple chemical plumes, as a separate VSN can track each different plume, and easily adapt to tracking chemical plumes that merge or split as they migrate.

The second goal is to develop a novel, efficient, and integrated subsurface chemical plume monitoring system, capable of capturing transient plumes in real-time to assess the source and predict future plume behavior. The data gathered by WSNs will continuously re-calibrate numerical models, while the models control the monitoring process. With the proposed system, the distribution of tasks associated with the characterization, pattern recognition, risk assessment, remediation design and decision making poses many opportunities and design alternatives. It is essential to take into account the limitations and capabilities of nodes, and at the same time maintain the resolution and accuracy required for real-time monitoring.

The two goals are synergistic. The plume tracking application demands the capability for tracking multiple plumes, imposes restrictions on physical communication topology, which in turn makes it necessary for subsets of sensor nodes to communicate via remaining nodes, and requires capabilities such as migration, division, and merging of VSNs. In short, it demands a comprehensive set of VSN capabilities. It also provides a practical deployment platform with associated environmental hazards on which the VSN algorithms and protocols can be evaluated.

3 Current Status

We next discuss progress made so far and how our current work will be extended to achieve the stated goals.

3.1 A Proof-of-Concept Study in an Intermediate-scale Tank

We have conducted a proof of concept demonstration study using a physical aquifer test bed constructed in an intermediate scale tank (8 ft x 4 inches x 16 inches) [5, 6]. The test system includes a set of ten conductivity probes from EchoTE individually connected to wireless sensor boards (TelosB). Figure 2 is a picture of the tank with motes and sensors deployed. The probes are placed within the tank and wired to the breakout boxes specifically built for a better sensor-to-mote interface. The motes are enclosed in a small transparent plastic box to prevent water and sand from splashing on them, and double-taped to the breakout boxes which are attached to a five feet tall rack above the tank. We used six motes in this experiment that are placed about one feet apart, and each has one or two probes attached to them (for a total of 10 probes placed in the tank). The motes are powered via USB connections to a regular power supply. The tank is packed using five well-characterized silica sands to represent a heterogeneous aquifer.

A tracer test has been conducted with a sodium bromide solution with a known concentration [5]. In the test, the

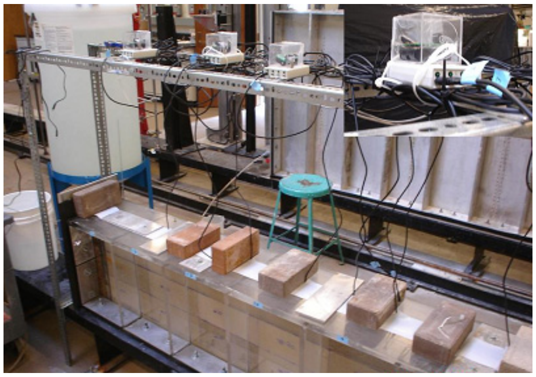

Figure 2. An intermediate scale tank instrumented with motes and conductivity probes

tracer was continuously injected into a steady flow field for several minutes, and concentration at different points in the tank was measured with the calibrated soil moisture/electrical conductivity sensors via the motes. The accuracy of the sensor-measured concentrations was compared to traditional grab samples analyzed using an ion chromatograph (see the comparison for sensor 1 in Figure 3). In

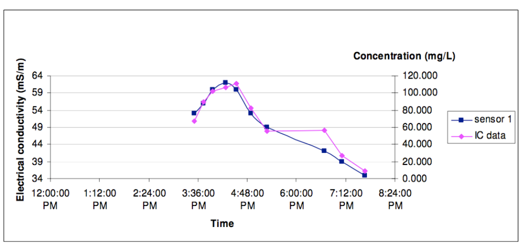

Figure 3. Electronic conductivity and concentration values of sensor 1

general, sensor data captures a trend similar to the IC data, hence quantitatively useful for tracking a plume.

The concentration values at different points were then used to calibrate the groundwater flow and transport model and estimate other physical parameters such as dispersivity and porosity in the tank. Observations made during this empirical study helped us to identify several issues that must be addressed before meaningful model calibration. For example, timing between sensed observations and simulated ones in the model should be synchronized. Missing data or completely abnormal readings severely impacts the accuracy of model calibration, so we are in the process of designing a fault detection algorithm for event driven sensor systems.

3.2 Distributed Algorithms for Virtual Sensor Networking

Clustering is a technique commonly employed for enhancing communication efficiency of sensor networks. Network clustering is the process of organizing the network into connected groups of nodes called *clusters* with one node selected to be the leader of the cluster or the *cluster-head*.

Cluster Formation and Organization: Two key components for efficient cluster based communication in sensor networks correspond to cluster formation and cluster organization. Consider a two-dimensional homogeneous sensor field. From the point of view of efficient communication, it is desirable to have a set of circular clusters, with the cluster head in the center, and the radius approximately equal to the communication range of the transmitter. However, with distributed clustering schemes, the shape of clusters formed becomes arbitrary, and the sizes of clusters varies widely. Furthermore, it is desirable to organize clusters according to a certain desired attribute, e.g., a tree of clusters with n branches per node (cluster head). We are in the process of developing efficient strategies for cluster formation and organization.

In [1] we provide a tunable scheme for controlled top-down cluster formation, hierarchical hop-ahead clustering (HHC) to achieve different cluster and cluster tree characteristics. Figure 4 compares results of HHC with Simple Hierarchical Clustering scheme (SHC) of 802.15.4, an emerging standard for communication among devices such as wireless sensor nodes. Results show that the HHC scheme is able to form more uniform clusters, and can be tuned to achieve certain desirable cluster tree characteristics. Our ongoing work involves the development of a labeling scheme that would assign nodes an ID based on its position in tree as well as its relative geographical location. Such a labeling scheme will facilitate the development of an energy efficient routing scheme for routing between different physical clusters of the sensor network.

Phenomena-aware Clustering: In addition to the clustering concept, another key insight exploited to reduce the cost of data collection is to realize that not all data is equally important and that some readings are more interesting than others. It is desirable then that data collection be specifically event-driven rather than continuous and network wide. In such an event-driven model, nodes only send data if it is deemed interesting according to some criterion and do not report data otherwise. We have considered both networking clustering and event-driven data reporting in tandem.

To model a wireless sensor network, we use a connected, weighted, undirected graph $G(N, E)$ where the set of vertices is equivalent to the set of sensor nodes N. An edge exists between two vertices if and only if they can communicate directly. The requirements of our clustering algorithm are tied to two classic NP-Complete problems, related to traditional clustering and event-driven data collection respectively. First, the choice of clusterheads is typically modeled as a minimum k-hop dominating set. The

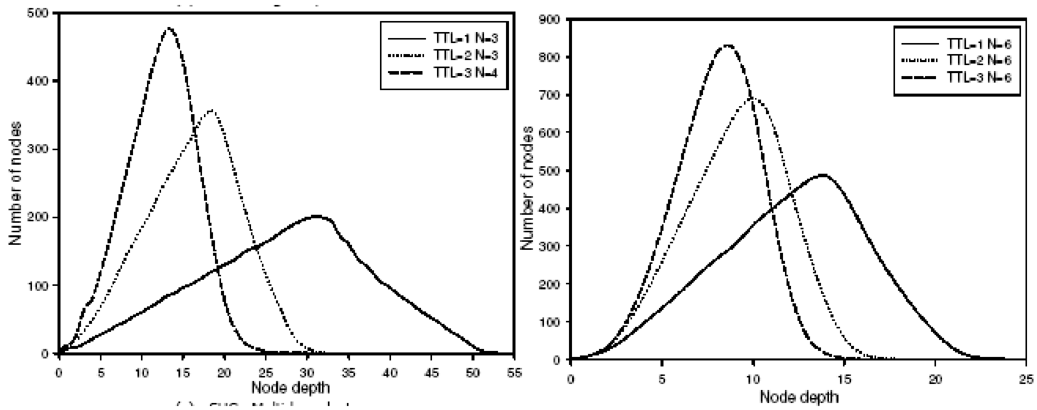

Figure 4. Number of nodes at different levels of the cluster tree for breadth first tree formation. Left: Simple hierarchical clustering scheme (SHC); and Right: Hierarchical hop-ahead clustering (HHC).

second problem emerges from attempting to route data to a base station from only a subset of nodes in the graph as mandated by event-driven data collection, called the Steiner tree problem. Therefore, our problem is defined formally as follows: Let there be an undirected graph $G(V, E)$ with weighted edges. Let there be a time variable subset of vertices $S(t) \subseteq V$. Find another time variable subset of vertices $D(t) \subset V$ such that every member of $S(t)$ is connected to exactly one member of set $D(t)$ at time t through a series of shortest paths, each of no more than k hops. This set D must have the property that it minimizes the objective function $f(W, |D|)$ where W is the sum of the weights of all the edges that exist on shortest paths between a member of S to a member of D and $|D|$ is the size of set D. Moreover D must be the set of minimal size among all sets that minimize f. We have proved that this problem is NP-Complete.

We have designed PHARE (PHenomena AwaRE clustering in wireless sensor networks), a distributed general-purpose clustering algorithm for WSNs that does not consider all nodes to be equal in importance [3]. PHARE possesses two essential properties. First, clusters have a locally determined, dynamic lifetime. Second, PHARE seeks to form clusters around groups of nodes that report events with high frequency. The first property is achieved by making clusterheads monitor the frequency of data reports from its members. The cluster's lifetime is made inversely proportional to this frequency, so that areas of lower reporting frequency recluster less often, thus saving energy. The second property is achieved by allowing clusterheads to average the position of member nodes weighted by their reporting frequencies, and thus obtain a long term center of activity for the nodes. The next time the network is reclustered, nodes are selected that are in proximity to this center, so that nodes with similar reports are naturally grouped together, thus improving aggregation.

When a cluster is active and monitoring the environment, the average reporting frequency and center of activity are periodically recalculated at the clusterheads. The remaining cluster lifetime is made either longer or shorter to take advantage of any shifts in activity. The center of activity is also shifted with periodic recalculation. To achieve fault tolerance, clusterheads send out heartbeat messages to their members and nearby members of surrounding clusters periodically. These heartbeats contain the center of phenomena, so that if the clusterhead fails, no heartbeat is sent out and after a while, the members detect this and use the activity center to position a new cluster. The reclustering procedure itself starts by giving each node an initial probability proportional to both its remaining energy and its proximity to the local activity center. These probabilities are continually doubled through a series of iterations until one of them hits 1. Nodes that do so announce themselves as clusterheads. Nodes who overhear this message, join as members through the path of lowest energy consumption.

We have theoretically proved that PHARE completes in constant time; PHARE always forms connected clusters; PHARE is able to detect and recover from dead or asymmetric links; and PHARE detects and recovers from node failures.

Detection and Tracking of Amorphous Events with Dynamic Signatures: A synopsis of the most relevant research reveals that the shortfall in current detection and tracking protocols is the universal assumption that events never combine into a large whole nor disintegrate into several smaller phenomena. These assumptions do not hold for subsurface contaminant plumes. We have designed a detection and tracking algorithm called DRAGON which can handle events that may be created or destroyed through event splits and merges [2].

Two central concepts define how fully generalized events are detected and tracked. They are the notions of *event center of mass* and *node momentum*. An event's center of mass delineates the event's location and serves as a reference point for tracking its movement. The measure of a node's momentum is a function of a node's readings and position relative to the event's center of mass. Thus defined, the node momentum is the decision variable that controls whether two events should remain logically distinct or instead be folded into one entity. The possible outcomes of the aforementioned decision control the event splitting and merging powers unique to our solution.

Key phenomena for triggering DRAGON are when nodes suddenly start to detect or cease to detect an event. A node begins to detect an event if its mass value is greater than 0 for the first time. These nodes either spawn new events or join existing ones. A simple solution is to allow a newly detecting node to join an existing event if the mass functions are identical and the node preserves individual stability for the event in question. If there are no neighboring events, or all events would be made unstable with the addition of node n, then n forms a new event. On the reverse side, the algorithm can be triggered when a node's mass drops to 0 and is removed from the event. Other possible triggers can be local mass shifts. Each cluster may maintain a local center of mass and total mass and call the algorithm when they change beyond certain amounts. When considering merges between two events, it is desirable to have both of their algorithm executions occur in sync. To achieve this, an event which triggers an execution of DRAGON could also trigger execution for neighboring events. Regular data collection periods are another possible trigger. One particularly intriguing option left for future work is to allow DRAGON to work with an event prediction algorithm, such as a flow transport model in this project.

The abilities and needs of DRAGON motivate three distinct phases of execution: Summary, Split, and Merge (Figure 5). The necessary decision predicates require two aggregates: center of mass and total mass. The Summary phase computes and distributes these aggregates to all clusters containing the event. This phase precedes the others and is the starting point of DRAGON. The Split and Merge phases are charged with checking and performing event splits and merges respectively and they both come after the summary phase. We have analytically proved that if two events are jointly stable then the event which results from the union is itself individually stable. Therefore, the product of a merge does not need to check for a split. This means that the merge phase can come after the split phase and there will be no cyclical dependency between the two phases, thus avoiding the thrashing problem. We have theoretically proved that DRAGON will terminate and it also scales well with event size, speed, and density.

The work presented here forms the basis of VSN formation and maintenance. The next step is to use the output of application models (such as flow transport model in this project) to more efficiently drive the operation of reclustering and event tracking.

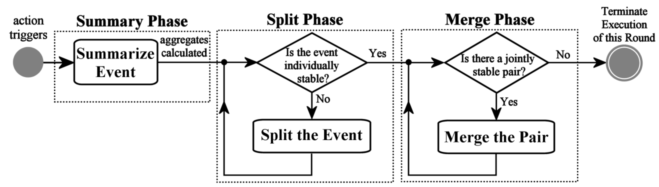

Figure 5. Algorithm phases of DRAGON

4 Conclusions

Plans are underway to integrate the algorithms developed here into a full-fledged system that uses the output of the flow transport model for more efficient event tracking, which in turn drives more targeted sensor data acquisition for more dynamic calibration of the flow transport model.

Acknowledgements: We would like to thank the following students who have contributed to this project: Dilum Bandara, Kevin Barnhart, Nick Hubbell, Denney Liptak, Philip Loden, Lisa Porta, and Inigo Urteaga.

References

[1] H. M. N. D. Bandara and A. P. Jayasumana. An enhanced top-down cluster and cluster tree formation algorithm for wireless sensor networks,. In *Proceedings of the Second International Conference on Industrial and Information Systems (ICIIS)*, pages 37–42, August 2007.

[2] N. Hubbell and Q. Han. Dragon: Detection and tracking of amorphous events with dynamic signatures in wireless sensor networks. Technical report, Colorado Shool of Mines, 2008.

[3] N. Hubbell, I. Urteaga, Q. Han, and A. P. Jayasumana. Phare: Phenomena aware clustering. Technical report, Colorado Shool of Mines, 2008.

[4] A. P. Jayasumana, Q. Han, and T. Illangasekare. Virtual sensor networks - a resource efficient approach for concurrent applications. In *Proceedings of The 4th IEEE International Conference on Information Technology - Next Generations (ITNG)*, 2007.

[5] L. Porta. Continuous plume monitoring and model calibration using a wireless sensor network: Proof of concept in intermediate scale tank. Master's thesis, Colorado School of Mines, Golden, CO, USA, 2007.

[6] L. Porta, T. Illangasekare, P. Loden, Q. Han, D. Liptak, and A. P. Jayasumana. Continuous plume monitoring and model calibration using a wireless sensor network: Proof of concept in intermediate-scale tank test. Poster, American Geophysical Union Fall Meeting (AGU), 2006.