Data Preservation in Data-Intensive Sensor Networks With Spatial Correlation

Nathaniel Crary
Department of Computer Science
Azusa Pacific University
Azusa, CA
ncrary09@apu.edu

Bin Tang and Setu Taase
Department of Computer Science
California State University Dominguez Hills
Carson, CA
btang@csudh.edu,staase1@toromail.csudh.edu

ABSTRACT

Many data-intensive sensor network applications are potential big-data enabler: they are deployed in challenging environments to collect large volume of data for a long period of time. However, in the challenging environments, it is not possible to deploy base stations in or near the sensor field to collect sensory data. Therefore, the overflow data of the source nodes is first offloaded to other nodes inside the network, and is then collected when uploading opportunities become available. We call this process *data preservation in sensor networks*. In this paper, we take into account spatial correlation that exist in sensory data, and study how to minimize the total energy consumption in data preservation. We call this problem *data preservation problem with data correlation*. We show that with proper transformation, this problem is equivalent to minimum cost flow problem, which can be solved optimally and efficiently. Via simulations, we show that it outperforms an efficient greedy algorithm.

Keywords – Big Data, Data Preservation, Sensor Networks, Energy-Efficiency, Spatial Correlation

1. BACKGROUND AND MOTIVATION

Big data refers to a collection of data sets that are so large and complex, and it becomes difficult to process using traditional data processing applications. The challenges of big data include its collection, storage, search, sharing, analysis, and visualization. One of the highly anticipated key contributors of the big data is the distributed wireless sensor networks, which consists of hundreds of thousands of small sensors with sensing, commuting, and communication capacities. For example, the Internet of Things (IoT) [2] is a futuristic network of physical objects embedded with RFID (Radio-Frequency Identification Technology) sensors, to track and manage a large number of uniquely identifiable objects.

With the advances in MEMS technology and the widely used sensors including video cameras, microphones, RFID readers, telescopes and seismometers, a whole new array of data-intensive sensing applications have been researched and developed recently. These emerging sensor network applications include underwater or ocean sensor networks [15, 29], acoustic sensor networks [17], and sensor networks monitoring volcano eruption and glacial melting [19, 33]. In these challenging environments with very limited accessibility, it is not possible to deploy a base station with power outlet near or inside the sensor network to collect the data. Therefore, the generated sensory data is first stored inside the network, and then being uploaded periodically to the remote base station by data mules [12, 13], or through low rate satellite link [20].

Gathering the large volume and wide variety of the sensory data is critical in above emerging sensor networks. Each sensor node has limited storage capacity and a finite, unreplenishable battery power supply. Sensor nodes close to the event of interest constantly generate large amounts of sensory data, which can quickly exhaust their limited storage capacity. For example, a distributed acoustic monitoring and trace retrieval system, called EnviroMic [17], was designed and deployed to monitor the social behaviors of animals in the wild. In EnviroMic, an acoustic sensor with 1GB flash memory will run out of its storage in just seven hours, when it samples the entire audible spectrum. We refer to these sensor nodes with exhausted data storage as *source nodes*. Other sensor nodes that are relatively far away from the event of interest and still have available storage are referred to as *storage nodes* (sensor node whose generated data has not exceeded its storage capacity is still considered as a storage node). In order to prevent data loss, the overflow data generated at source nodes needs to be offloaded to the storage nodes before above uploading opportunities become available. The storage nodes that finally store offloaded data are referred to as *destination nodes*. We refer to this process wherein overflow data is offloaded from source nodes to destination nodes to be preserved as *data preservation in sensor networks*.

Since different sensor nodes partially monitor the same spatial region, the generated sensory data is often correlated. That is, nodes close to each other while monitoring the same event of interest often produce similar or same readings about the event, resulting in multiple observations of the same event. During data preservation, it is not necessary to offload all such redundant information for the following reasons. First, the redundant information may not contribute to the sensor network application if a smaller amount of sensor measurements is adequate to tell the event features within a certain reliability/fidelity level. Second, more im-

Mobidata'15, June 21, 2015, Hangzhou, China.

portantly, preserving any measurements inside the network costs participating nodes' battery power, which is finite and unreplenishable. Therefore, it is not desirable to offload all the generated data when a clear correlation existing among them. The goal of the problem is to minimize the total battery power consumption of data preservation while offloading only one observation of each event, considering that each event has multiple observations in the network due to spatial correlation. We refer to this problem as *data preservation problem with data correlation*.

Paper Organization. The rest of the paper is organized as follows. In Section 2, we discuss the related work in the field of big-data and sensor networks. In Section 3, we identify and formulate data preservation problem with data correlation, and show that with proper transformations, it can be modeled as a minimum cost flow problem. We present and analyze the simulation results in Section 4. Section 5 concludes the paper and points out some future work.

2. RELATED WORK

In the last few years, "Data-Intensive Sensor Networks" have been used in literature to represent one trend of sensor network development [6, 24]. That is, sensor networks are beginning to generate large amount of data. Recently, it further takes a "big-data" perspective and rethinks some of the research problems in traditional sensor network [26]. Takaishi et al. [26] observed that although the data generated by an individual sensor may not appear to be significant, the overall data generated across numerous sensors in the densely distributed sensor network can produce a significant portion of the big data.

Inspired by above work, we envision a data-intensive sensor network wherein a large volume of sensory data are generated therefore energy-efficiency becomes a even more critical problem in emerging sensor networks. In this paper, we consider a sensor network that is deployed in remote area or inhospitable region, therefore there is no base station available in the sensor network. How to effectively preserve the generated sensory data inside the network becomes a new challenge. The first line of work is a sequence of system research in disconnection-tolerant storage sensor networks [17, 18, 32, 36]. The authors in these papers design acoustic sensor networks, which monitor the social behavior of animals in the wild. Since no base station is available, they design cooperative distributed storage systems specifically for disconnected operations of sensor networks, to improve the utilization of the network's data storage capacity. The other line of research instead takes an algorithmic approach by focusing on the hardness of the problems and the optimality of their solutions [11, 25, 27, 28, 34]. Tang et al. [25, 28] address the energy-efficient data redistribution problem in data-intensive sensor networks, and propose efficient centralized and distributed algorithms. Hou et al. [11] and Takahashi et al. [25] study how to maximize the minimum remaining energy of the nodes that finally store the data, in order to store the data for long period of time. Xue et al. [34] consider that sensory data from different source nodes have different importance, and study how to preserve data with highest importance. All above work, however, does not consider spatial correlation of sensory data, which could be utilized to achieve energy efficient data preservation. By introducing spatial correlation, we generalize the energy-efficient data redistribution problem studied in [28].

There has been very active research in data spatial correlation in sensor networks. Vuran et al. [30] establish a theoretical framework that captures and exploits the data correlation in order to develop efficient communication protocols in sensor networks. Jindal et al. [14] create a mathematical model to generate synthetic and spatially correlated data. In terms of specific research thrusts, data spatial correlation has been used in data aggregation and routing algorithms [5, 23], MAC protocol design [31], data storage and querying [16], and data encoding and compression [23, 37]. All above work assume existence of a base station so that the correlated data can be transmitted from sensor nodes to the base station directly. Our work does not assume the existence of base stations (due to the challenging environments in which the sensor network applications are deployed), rather it focuses on how to offload data from storage-depleted source nodes into the sensor network. Therefore, the spatial correlation addressed in this paper adopts a totally different network model compared to all the existing spatial correlation research in sensor networks.

Network flow algorithms (including maximum flow [3], minimum cost flow [9, 22, 28], and multi-commodity flow [35]) have been used to model and solve several problems in sensor network research. Bodlaender et al. [3] study the integer maximum flow problem in wireless sensor networks with energy constraint. They show that despite the efficiency of traditional maximum flow problems, the integer maximum flow in sensor network is indeed strongly NP-complete and in fact APX-hard. Patel et al. [22] study minimizing the energy cost of sending data packets from sensor nodes to base stations while satisfying the capacity limits of wireless links. They propose a routing protocol based on minimum cost flow algorithm. Ha et al. [9] adopt minimum cost algorithm with the aim of increasing the monitoring coverage and the operational lifetime of mesh-based sensor networks. Tang et al. [28] formulate the energy-efficient data redistribution problem in data-intensive sensor networks as a minimum cost flow problem. Xue et al. [35] model energy efficient routing for data aggregation in sensor networks as a multicommodity flow problem, where a commodity represents the data generated from a sensor node. However, none of above work addresses data spatial correlation in sensor networks. The data aggregation model in [35] assumes that information from different source nodes can be assembled at relay nodes when they are transmitted to the base station. However, no explicit spatial correlation model is given and therefore it is unclear the role of spatial correlation in the data aggregation process. To the best of our knowledge, this work is the first one to formulate data spatial correlation in sensor networks as a minimum cost flow algorithm and solve it optimally.

3. DATA PRESERVATION PROBLEM WITH DATA CORRELATION

Network Model. The sensor network is represented as an undirected graph $G(V, E)$, where $V = \{1, 2, ..., N\}$ is the set of N nodes, and E is the set of edges. There are a sequence of events occurred inside the network, each is sensed and observed by multiple sensor nodes (therefore referred to as *source nodes*). Thus each event corresponds to multiple observations, each is stored at a different source node. The sensory data is therefore modeled as a sequence of observations, each of which has the same size of k bits. Without loss of generality, let $V_1 = \{1, 2, ..., p\}$ be the p source nodes (with depleted storage space), and $V_2 = \{p+1, p+2, ..., N\}$ be the set of storage nodes ($V_1 \cap V_2 = \emptyset$, $V_1 \cup V_2 = V$). Let d_i be the total number of observations source node i needs to offload. Let $a = \sum_{i=1}^{p} d_i$ denote the total number of observations to be offloaded in the network. Let m_i be the available free storage space at storage node $i \in V_2$, measured in number of observations (that is, storage node i can store m_i more observations). If $i \in V_1$, then $m_i = 0$, implying that a source node has zero available storage space. If $i \in V_2$, then $m_i \geq 0$, implying that storage node i can store another m_i observations. We assume that the total size of the observations to be preserved is less than or equal to the size of the total available storage space in the network, that is, $a \leq \sum_{i=p+1}^{N} m_i$.

Energy Model. We adopt the first order radio model [10] wherein for a k-bit observation data sent over distance l meters, the *transmission energy* (on the sender side) is $E_t(k,l) = \epsilon_{elec} * k + \epsilon_{amp} * k * l^2$, the *receiving energy* (on the receiver side) is $E_r(k) = \epsilon_{elec} * k$. Here $\epsilon_{elec} = 100nJ/bit$ is the energy consumption per bit on the transmitter circuit and receiver circuit, and $\epsilon_{amp} = 100pJ/bit/m^2$ calculates the energy consumption per bit on the transmit amplifier. Let $w_{u,v}$ denote the total energy consumption when node u sends a k-bit observation data to its one hop neighbor v over their distance $l_{u,v}$, then $w_{u,v} = E_t(k, l_{u,v}) + E_r(k)$. Note that receiving energy is independent of distance between sender and receiver, and that $w_{v,u} = w_{u,v}$. Now for any arbitrary two nodes i and j in the network that are multiple hops away from each other, let $c_{i,j}$ be the minimum energy consumption of sending one observation from i to j along path $SP_{i,j}$. Here $SP_{i,j}$ is referred to as the *minimum energy consumption path* between i and j. Then $c_{i,j} = \sum_{(u,v) \in SP_{i,j}} w_{u,v}$, wherein both $c_{i,j}$ and $SP_{i,j}$ can be easily obtained using Dijsktra's shortest path algorithm by assigning weight $w_{u,v}$ to edge $(u,v) \in E$.

Problem Formulation. We first present the spatial correlation model, then formulate the problem. It should be noted that data preservation with temporal correlation can be treated in a much simpler way, considering that all the temporally-correlated observations are generated at the same source node. A source node can therefore choose one copy from the set of temporally-correlated observations to offload. For research of temporal correlation in sensor networks with base stations, please refer to [30].

Data Spatial Correlation Model. In our spatial correlation model, the observations from different source nodes for the same event is identical. Thus each event corresponds to multiple identical observations, each is from a different source node. To conserve energy, we would like to preserve only one observation for each event, to indicate the existence of such event. Specifically, out of the total a observations, there are only $b < a$ unique events $D = \{D_1, D_2, ..., D_b\}$, each with k bits.[1] Let n_j denote the number of observations event D_j has, and $D_{j,k}$ denote the k^{th} observation of D_j. Let $s(j,k) \in V_1$ denote the source node of $D_{j,k}$, where $1 \leq j \leq b$ and $1 \leq k \leq n_j$. We have $\sum_{j=1}^{b} n_j = a$. Therefore, it needs

[1]Our work generalizes the work of [28], which assumes $b = a$ therefore data spatial correlation does not exist.

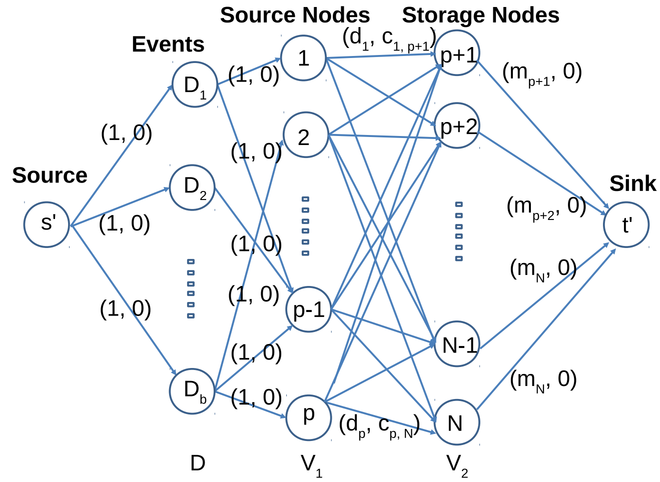


Figure 1: Data preservation problem with data correlation is equivalent to minimum cost flow Problem. The first number in each parenthesis is edge capacity, the second is edge cost.

to decide:

1) out of n_j observations each unique event D_j, which one is selected to offload,

2) a *preservation function* $r : D \to V_2$, indicating one observation of D_j is offloaded to destination node $r(j)$ via the minimum energy consumption path between $s(j,k)$ and $r(j)$, assuming the k^{th} copy of D_j is selected to offload.

Our goal is to offload one observation for each event into the network to be preserved, such that the *total preservation cost* τ is minimized, where

$$\tau = \sum_{j=1}^{b} \min_{1 \leq k \leq n_j} c_{s(j,k), r(j)}, \tag{1}$$

under the storage capacity constraint that the total number of observations offloaded to node $i \in V_2$ is less than or equal to i's available storage capacity, i.e.,

$$|\{j | r(j) = i, 1 \leq j \leq b\}| \leq m_i, \qquad \text{for all } i \in V_2.$$

Minimum Cost Flow Solution. We show that the problem of data preservation with correlation is equivalent to the minimum cost flow problem [1, 21], which is stated as below. Given a graph in which each edge has a capacity and a cost, some nodes are supply nodes and some are demand nodes, the goal is to find flows from supply nodes to demand nodes with minimum cost such that the capacity constraint of each edge is satisfied.

THEOREM 1. *The problem of data preservation with data correlation is equivalent to the minimum cost flow problem.*

Proof: We first transform sensor network graph $G(V,E)$ into a new graph $G'(V', E')$ as follows, shown in Fig. 1:

1. $V' = V \cup \{s'\} \cup \{t'\} \cup D$, where s' is the new source node, t' is the new sink node.

2. E' include following edges (from left to right in Fig. 1):

   - $\{(s', D_j), D_j \in D\}$: capacity and cost of each edge are 1 and 0, respectively.
   - $\{(D_j, i), D_j \in D$ and $i \in V_1\}$, if i is a source node of D_j: capacity and cost of each edge are 1 and 0, respectively. Note that not every pair of D_j and i are connected by an edge.
   - $\{(i, j), i \in V_1$ and $j \in V_2\}$: capacity and cost of edge (i, j) are d_i and $c_{i,j}$, respectively.
   - $\{(i, t'), i \in V_2\}$: capacity and cost of edge (i, t') are m_i and 0, respectively.

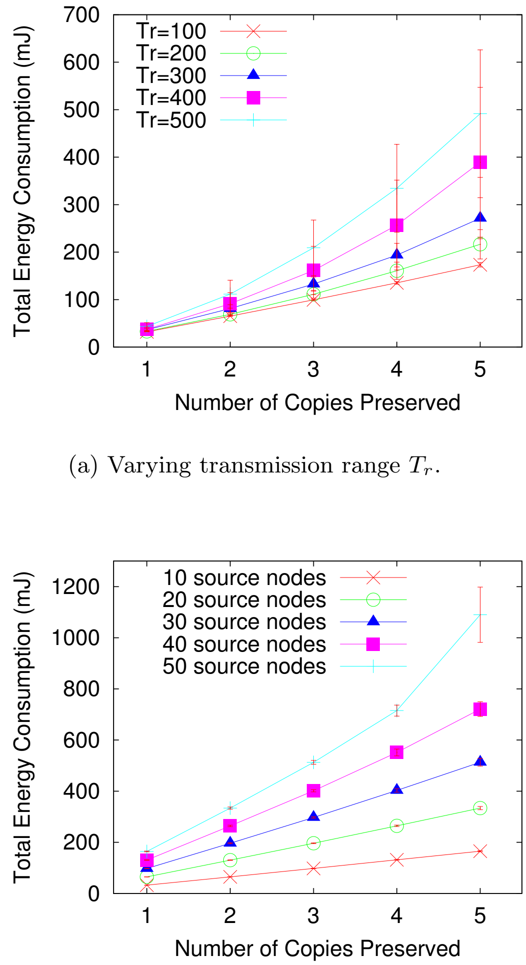
3. Set both the supply at s' and the demand at t' as b, the number of unique events. The supply of other nodes in V' is set as 0.

Now b amount of valid flow from s' to t' must include 1 amount on edge (s', D_1), 1 amount on (s', D_2), ..., and 1 amount on (s', D_b). This achieves our goal that only one observation of each event is offloaded. b is the amount of maximum possible flow in the network. Solving the minimum cost flow problem on $G'(V', E')$ gives the minimum preservation cost in the data preservation problem with data correlation in $G(V, E)$. ∎

<u>Minimum Cost Flow Algorithm.</u> Minimum cost flow problem can be solved efficiently and optimally [1]. We adopt minimum cost flow algorithm proposed in [7, 8]. It is based on scaling push-relabel method and its implementation works well over a wide range of problem classes. This algorithm has the time complexity of $O(n^2 m \log(nc))$, where n, m, and c are the number of nodes, the number of edges, and the maximum capacity of an edge in graph G'. With above transformation, $n = 2 + b + N$, $m <= b + b \times p + p \times (N - p) + N - p = (p+1) \times (N - p + b)$, $c = \max(\max_i\{d_i\}, \max_i\{m_i\})$. Therefore the time complexity of the minimum cost flow algorithm in G' is $p(b+N)^2(N-p+b)\log((b+N)c) = O\big(p \times (b+N)^3 \times \log((b+N)c)\big)$.

For the transformation part, we adopt Djkstra's shortest path algorithm with Fibonacci heap, which takes $O(|E| + N\log N)$ [4]. Therefore finding the minimum energy consumption paths of all pairs of source and storage nodes takes $O(p \times (N-p) \times (|E| + N\log N))$, which is still less than the time complexity of minimum cost flow algorithm.

Greedy Algorithm. Next we present a simple but more efficient greedy algorithm (Algorithm 1) and compare it with optimal minimum cost flow algorithm. For each unique event, it finds the closest storage node (with available space) to one of the source nodes of that event, and offloads its observation from this source node to this storage node. It stops until one observation of all the b events are offloaded. Finding the minimum energy consumption paths of all pairs of source and storage nodes (line 1) takes $O(p \times (N-p) \times (|E| + N\log N))$, the rest of Algorithm 1 takes $O((N-p) \times a)$, where a is the total number of observations. Therefore, the time complexity of Algorithm 1 is $O(p \times (N-p) \times (|E| + N\log N) + (N-p) \times a)$.



(a) Varying transmission range T_r.

(b) Varying number of source nodes.

Figure 2: Performance Study of the Optimal Algorithm.

ALGORITHM 1. *Greedy Data Preservation Algorithm.*
Input: *Sensor Network Graph* $G(V, E)$
Output: *Total preservation cost* τ
0. **Notations**:
   min_j: *minimum energy cost of offloading* D_j;
1. *Calculate* $c_{i,j}$ *for all* $i \in V_1$, $j \in V_2$;
2. $\tau = 0$;
3. **for** $(1 \le j \le b)$
4. $min_j = \infty$;
5. **for** $(1 \le k \le n_j)$
6. **for** $(p + 1 \le l \le N)$;
7. **if** $(c_{s(j,k),l} < min_j)$
8. $min_j = c_{s(j,k),l}$;
9. **end for;**
10. **end for;**
11. $\tau = \tau + min_j$;
12. **end for;**
13. **RETURN** τ.

4. PERFORMANCE EVALUATION

We study the performances of algorithms for the data preservation with data correlation, viz. the minimum cost flow algorithm (referred to as Optimal) and the greedy algorithm (referred to as Greedy). For fair comparison, both algorithms take the same input file, which specifies network topology, set of source nodes, number of events, number of observations at each source node, and storage capacity of each storage node. Each data point below is an average over five runs. In all plots, the error bars indicate 95% confidence interval.

Simulation Setup. We wrote our own simulator in Java. We generate a 100-node sensor network in a $1000m \times 1000m$ field, using a recursive graph construction technique. First a sensor node is randomly placed, then a random angle (between $0°$ and $360°$) and a random distance (less than or equal to transmission range) to that node are picked, at which the second node is placed. Each iteration one node is generated to branch off the existing graph. This is repeated until all 100 nodes are generated. In contrast to the traditional sensor network generation, wherein a number of sensor nodes are randomly generated inside the network, above graph construction technique guarantees that the sensor network constructed is always connected. Along the graph construction, nodes are randomly picked to be source nodes. Afterwards, observations of each event are randomly generated and distributed among the source nodes. There are 50 unique events, each has 5 copies of observations due to spatial correlation. Each observation data is of size 400 Bytes. The storage capacity of each storage node is 10 KByte (the size of 25 observation data).

Performance of Optimal. Figure 2 shows the performance of Optimal by varying the transmission ranges and the number of observations preserved for each event.[2] Figure 2 (a) shows that when transmission range is small, the total energy consumption linearly increases with the number of observation copies offloaded; while when transmission range gets larger, the total energy consumption increases faster. This is because when transmission range is small, the observation data offloading takes place in a multiple hop manner while when transmission range is large, more nodes are directly connected and thus can communicate with with each other directly. It has been shown in wireless communication that it costs more energy to send from one node to another directly than by way of multiple hops, due to the fact that energy consumption is proportional to the square of the communication distance. Figure 2 (b) shows the performance comparison when varying number of source nodes in the network from 10 to 50. It shows that energy consumption increases more dramatically with the increase of number of preserved copies, when number of source nodes increase. This is because with more source nodes, it gets more difficult to offload the observation copies since there are less number of storage nodes are available.

Comparing Optimal with Greedy. Finally we compare Optimal with Greedy. Figure 3 shows that when number of source nodes are small, both Optimal and Greedy perform similarly. However, when number of source nodes are large, Optimal evidently performs much better than Greedy, with less energy consumption incurred for data preservation. This is because when small number of source nodes are located randomly inside the network, there are ample number of storage nodes around each source node, thus the Greedy can always find a close storage node to preserve each observation, thus performing as well as the Optimal. However, when increasing the number of source nodes, the situations get more stressful as more number of observation copies need to offload while there are relatively less number of storage nodes. As a result, the Greedy becomes less effective as it only focuses on offloading observations from each source node currently considered in that iteration, and does not take into account how source nodes are coordinated to offload data. In contrast, minimum cost flow algorithm takes into account all the factors among source nodes and gives a cooperative offloading strategies such that the total energy cost is minimum.

[2]Minimum cost flow transformation (Fig. 1) can be easily modified to represent that multiple observations of each event are preserved.

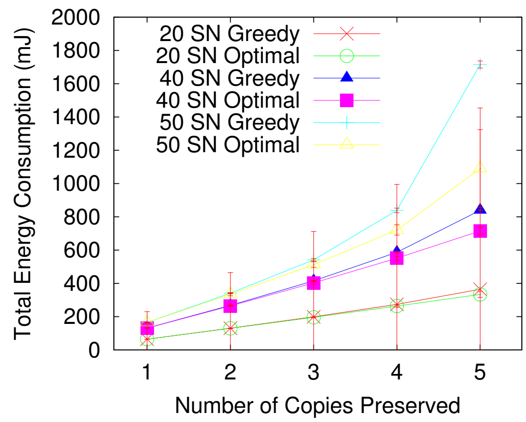


Figure 3: Performance comparison between Optimal and Greedy Algorithms.

5. CONCLUSION AND FUTURE WORK

When large amounts of data are generated in data-intensive sensor networks, energy efficiency becomes a more critical issue. In this paper, we take a “big-data” perspective of sensor networks and focus a new paradigm in emerging sensor network applications, wherein large volumes of data must be preserved inside the network due to the absence of base stations. Preserving large amount of data under the storage space and battery energy constraints of sensor nodes is a challenging problem. We exploit spatial correlation that commonly exist among sensory data, and propose a minimum cost flow approach to solve data preservation problem optimally. This paper serves as the first step to study big data preservation while utilizing data spatial correlation. Currently, our spatial correlation model assumes that one event has multiple identical observations. As future work, we will consider more realistic spatial correlation model wherein data similarity varies according to distances among sensor nodes.

Acknowledgment

This work was supported in part by NSF Grants CNS-1116849, HRD-1302873, and the Chancellor’s Office of the California State University. We thank Douglass Kinnes, Nicholas

Storm, Ian Marriott, and Thomas Mifflin for their input.

6. REFERENCES

[1] Ravindra K. Ahuja, Thomas L. Magnanti, and James B. Orlin. *Network Flows: Theory, Algorithms, and Applications*. Prentice Hall, 1993.

[2] Luigi Atzori, Antonio Iera, and Giacomo Morabito. The internet of things: A survey. *Comput. Netw.*, 54(15):2787–2805, October 2010.

[3] Hans L. Bodlaender, Richard B. Tan, Thomas C. Dijk, and Jan Leeuwen. Integer maximum flow in wireless sensor networks with energy constraint. In *Proc. of the 11th Scandinavian workshop on Algorithm Theory, SWAT 08*, pages 102–113.

[4] Thomas Corman, Charles Leiserson, Ronald Rivest, and Clifford Stein. *Introduction to Algorithms*. MIT Press, 2009.

[5] R. Cristescu, B. Beferull-Lozano, M. Vetterli, and R. Wattenhofer. Network correlated data gathering with explicit communication: Np-completeness and algorithms. *IEEE/ACM Trans. Netw.*, 14(1):41–54, 2006.

[6] Fatme El-Moukaddem, Eric Torng, and Guoliang Xing. Mobile relay configuration in data-intensive wireless sensor networks. *IEEE Transactions on Mobile Computing*, 12(2):261–273, February 2013.

[7] A. V. Goldberg. Andrew Goldberg's network optimization library. http://www.avglab.com/andrew/soft.html.

[8] A. V. Goldberg. An efficient implementation of a scaling minimum-cost flow algorithm. *Journal of Algorithms*, 22(1):1–29, 1997.

[9] Rick W. Ha, Pin-Han Ho, X. Sherman Shen, and Junshan Zhang. Sleep scheduling for wireless sensor networks via network flow model. *Comput. Commun.*, 29:2469–2481, August.

[10] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan. Energy-efficient communication protocol for wireless microsensor networks. In *Proc. of HICSS 2000*.

[11] Xiang Hou, Zane Sumpter, Burson Lucas, and Bin Tang. Maximizing data preservation in intermittently connected sensor networks. In *Proc. of the IEEE International Conference on Mobile Ad-hoc and Sensor Systems (MASS 2012), short paper*.

[12] Sushant Jain, Rahul Shah, Waylon Brunette, Gaetano Borriello, and Sumit Roy. Exploiting mobility for energy efficient data collection in wireless sensor networks. *MONET*, 11(3):327–339, 2006.

[13] D. Jea, A. A. Somasundara, and M. B. Srivastava. Multiple controlled mobile elements (data mules) for data collection in sensor networks. In *Proc. of the IEEE DCOSS*, pages 244–257, 2005.

[14] A. Jindal and K. Psounis. Modeling spatially correlated data in sensor networks. *ACM Trans. Sensor Networks (TOSN)*, 2(4):466–499, 2006.

[15] S. Li, Y. Liu, and X. Li. Capacity of large scale wireless networks under gaussian channel model. In *Proc. of MOBICOM 2008*, pages 140–151.

[16] Changlei Liu and Guohong Cao. Distributed monitoring and aggregation in wireless sensor networks. In *Proc. of Infocom 2010*.

[17] L. Luo, Q. Cao, C. Huang, L. Wang, T. Abdelzaher, and J. Stankovic. Design, implementation, and evaluation of enviromic: A storage-centric audio sensor network. *ACM Transactions on Sensor Networks*, 5(3):1–35, 2009.

[18] L. Luo, C. Huang, T. Abdelzaher, and J. Stankovic. Envirostore: A cooperative storage system for disconnected operation in sensor networks. In *Proc. of INFOCOM 2007*.

[19] K. Martinez, R. Ong, and J.K. Hart. Glacsweb: a sensor network for hostile environments. In *Proc. of SECON 2004*, pages 81–87.

[20] Ioannis Mathioudakis, Neil M. White, and Nick R. Harris. Wireless sensor networks: Applications utilizing satellite links. In *Proc. of the IEEE 18th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2007)*, pages 1–5, 2007.

[21] C.H. Papadimitriou and K. Steiglitz. Combinatorial optimization: Algorithms and complexities. *Prentice Hall*, 1982.

[22] M. Patel, S. Venkatesan, and R. Chandrasekaran. Efficient minimum-cost bandwidth-constrained routing in wireless sensor networks. *pecial Issue on "Wireless Networks and Pervasive Computing, Journal of Pervasive Computing and Communications (JPCC)*, 2(2), 2006.

[23] S. Pattem, B. Krishnamachari, and R. Govindan. The impact of spatial correlation on routing with compression in wireless sensor networks. *ACM Trans. Sensor Networks(TOSN)*, 4(8):24–33, 2008.

[24] Divyashee Sharma. *Efficient Information Access in Data-Intensive Sensor Networks*. Ph.D. Thesis, University of Pittsburg, 2010.

[25] Masaaki Takahashi, Bin Tang, and Neeraj Jaggi. Energy-efficient data preservation in intermittently connected sensor networks. In *Proc. of the International Workshop on Wireless Sensor, Actuator and Robot Networks (WiSARN), in conjunction with IEEE INFOCOM 2011.*

[26] D. Takaishi, H. Nishiyama, N. Kato, and R. Miura. Toward energy efficient big data gathering in densely distributed sensor networks. *IEEE Transactions on Emerging Topics in Computing*, 2(3):388–397, 2014.

[27] Bin Tang, Neeraj Jaggi, Haijie Wu, and Rohini Kurkal. Energy efficient data redistribution in sensor networks. In *Proc. of IEEE MASS 2010*, pages 352–361.

[28] Bin Tang, Neeraj Jaggi, Haijie Wu, and Rohini Kurkal. Energy efficient data redistribution in sensor networks. *ACM Transactions on Sensor Networks*, 9(2):1–28, May 2013.

[29] I. Vasilescu, K. Kotay, D. Rus, M. Dunbabin, and P. Corke. Data collection, storage, and retrieval with an underwater sensor network. In *Proc. of SenSys 2005*, pages 154–165.

[30] M. Vuran, O. Akan, and I. Akyildiz. Spatio-temporal correlation: theory and applications for wireless sensor networks. *Computer Networks*, 45:245–259, 2004.

[31] M. Vuran and I. Akyildiz. Spatial correlation-based collaborative medium access control in wireless sensor networks. *IEEE/ACM Trans. Netw.*, 14(2):316–329, april 2006.

[32] Lili Wang, Yong Yang, Dong Kun Noh, Hieu Le, Tarek Abdelzaher, Michael Ward, and Jie Liu. Adaptsens: An adaptive data collection and storage service for solar-powered sensor networks. In *Proc. of the 30th IEEE Real-Time Systems Symposium (RTSS 2009)*.

[33] Geoff Werner-Allen, Konrad Lorincz, Jeff Johnson, Jonathan Lees, and Matt Welsh. Fidelity and yield in a volcano monitoring sensor network. In *Proc. of OSDI 2006*, pages 381–396.

[34] Xinyu Xue, Xiang Hou, Bin Tang, and Rajiv Bagai. Data preservation in intermittently connected sensor networks with data priorities. In *Proc. of IEEE SECON 2013*, pages 65–73.

[35] Yuan Xue, Yi Cui, and Klara Nahrstedt. Maximizing lifetime for data aggregation in wireless sensor networks. *Mob. Netw. Appl.*, 10(6):853–864, December 2005.

[36] Yong Yang, Lili Wang, Dong Kun Noh, Hieu Khac Le, and Tarek F. Abdelzaher. Solarstore: enhancing data reliability in solar-powered storage-centric sensor networks. In *Proc. of MobiSys 2009*, pages 333–346, 2009.

[37] K. Yuan, B. Li, and B. Liang. A distributed framework for correlated data gathering in sensor networks. *IEEE Trans. Veh. Technol.*, 57(1):578–593, 2008.