

Efficient delay-constraint data collection in wireless sensor networks

Long Tran-Thanh / János Levendovszky

Received 2011-09-13

Abstract

This paper investigates a relaxed version of the delay-constraint (i.e. time dependent) data collection in wireless sensor networks, namely maximizing the number of collected data with given delay-constraint. In particular, we aim to maximize the total amount of data that can be successfully delivered to the base station within a time delay constraint T . In addition, this problem is studied in networks with rechargeable nodes, where the battery of each sensor node can be recharged, by using an external energy source (e.g. the sun). Given all this, we aim to devise an efficient routing algorithm that takes all of the aforementioned aspects into account, exploiting the advantages of the relaxation and the ability of energy recharging. In this paper, we propose a decentralized algorithm for the relaxed problem, which, under certain assumptions, is optimal in terms of achieving maximal collected data. We prove that this algorithm has polynomial time complexity. In addition, by using extensive simulation results, we show that the algorithm has low communication overhead on average.

Keywords

wireless sensor networks · data collection · delay-constraint

Long Tran-Thanh

Department of Technology, BME, H-1117 Budapest, Magyar tudósok krt. 2., Hungary
e-mail: ttl@cs.bme.hu

János Levendovszky

Department of Technology, BME, H-1117 Budapest, Magyar tudósok krt. 2., Hungary
e-mail: levendov@hit.bme.hu

1 Introduction

Due to the recent advances in electronics and wireless communications, the development of low-cost, low-energy, multi-functional wireless sensors have received increasing attention [7], and are becoming extensively used in a wide variety of applications ranging from environmental, habitat and traffic monitoring, to object tracking and military field observations (for more details see [10, 15]). Since one of the major goals of the wireless sensor networks (WSNs) is to collect information from the surrounding environment [16], it is important to develop efficient data collection schemes within the WSN domain. To this end, this paper focuses on time dependent data collection (TDDC), whereby data has to be collected with respect to given delay constraints. This is especially important in applications, such as object tracking and area surveillance systems, where real-time data transmission plays an important role. To date, within the TDDC domain, most research studies are concerned with collecting all the data in the network, such that the delay of each data packet cannot exceed a given threshold. In the literature, this problem is referred to as the Hop-bounded Minimum Spanning Tree (HBMST) problem, which is proven to be NP-hard [2]. However, due to the emerging real-time applications, there is a need to relax this problem so that it can lend itself to solutions in polynomial complexity.

Furthermore, most of the proposed algorithms focuses on energy-awareness, in order to maximize the life span of the WSNs. That is, it is necessary to efficiently manage the energy consumption of the sensor nodes. Otherwise, rapid battery depletion may lead to insufficient data collection from the network. Recently, however, a number of research works have proposed *energy harvesting sensors*, which have the capability of scavenging ambient energy from their surrounding environment, using solar, vibration, temperature, and radioactive sources [5, 17]. For example, the ZebraNet project used solar harvesting nodes mounted to the zebras to monitor the animals' behavior [19]. Within such applications, agents typically seek to comply with the concept of *energy-neutrality*, in which the energy consumption of an agent should be equal to the harvested energy. The advantage of such approaches is that agents can indefinitely extend

their life span, which is especially important when data collection has to operate over a prolonged period of time. By exploiting this capability, designers can focus on novel approaches of data collection, which are *energy-neutral*.

With these approaches, WSNs can achieve better performance compared to their energy-aware counterparts, since they are not constrained by energy limits. To date, however, most of previous work on efficient delay-constraint data collection in WSNs do not exploit the advantages of energy harvesting sensors. A notable exception is the work of Tran-Thanh and Levendovszky [18], which, similarly to this work, investigates relaxed delay-constraint schemes for data collection in WSNs with energy harvesting. However, they ignored the energy consumption of receiving, that is they assumed that to receive a single data packet, each sensor did not have to consume energy. In fact, real world sensors also need to spend a certain amount of energy in order to accept messages from other nodes [13]. Given this, our work can be regarded as the extension of Tran-Thanh and Levendovszky [18], in which we also take the energy cost of data receiving into account, and thus, make the model more realistic.

On the basis of this background, we propose a energy-neutral data collection scheme for data collection in WSNs, that takes receiving cost into account. In so doing, we first regard a WSN with energy harvesting nodes. Then we address TDDC from the following perspective: we aim to achieve maximal data collection with time limit, subject to a given delay constraint, instead of building up a minimum spanning trees. Then we propose an optimal decentralized algorithm to solve this relaxed problem, proving that the complexity of the algorithm is polynomial in time. finally, by extensive simulation results, we show that the proposed algorithm has low communication overhead in average.

The paper is organized as follows. In section 2, we give an overview of the related work. In section 3, we introduce our models of the radio links and the WSN. The data collection with time limit problem is discussed in section 4, and an optimal decentralized solution for this problem is also proposed within the section. finally, we conclude this work in section 5.

2 Related Work

In the literature, approaches to TDDC include: (i) real-time minimal-delay routing techniques; and (ii) delay-constraint data collection (i.e. data collection with time limit).

From the perspective of minimal-delay routing, most of the related works focuses on energy-awareness. For instance, Akkaya and Younis [1] suggested an energy-aware routing protocol to deliver the data within a bounded delay in the WSNs. In the proposed mechanism, energy-aware paths, that guarantees a certain end-to-end delay, are chosen for data forwarding. In addition, Ammari and Das [4] advocated an approach that characterizes the trade-off between energy and delay. The authors divided the transmission range of sensors into concentric circular bands (CCBs), in order to help a node to express its degree

of interest in minimizing two conflicting metrics, namely energy consumption and delay. More related to our work, He [11] proposed a DFS tree based delay-minimum energy-aware routing protocol (DERP) for real-time communication. This algorithm minimizes the worst-case delay in the network.

In the domain of delay-constraint data collection, Althaus et al. [3] proposed an algorithm to compute a feasible-hop spanning tree with expected cost $O(\log n)$ times of the optimal case, where n was the number of the vertex in the graph. Cheng et al. [6] studied the delay-degree-bounded data collection problem, and presented three heuristic algorithms. But the most related to our work is the one proposed by Xu et al. [20]. There, a load-aware power-increased topology control algorithm (LPTC) was proposed in order to heuristically solve the TDDC, taking the total number of relaying packets into account.

Although the number of proposed approaches is numerous, to our best knowledge, they are all energy-concerned, not exploiting the energy harvesting capability of the (new generation) nodes. In the contrast, Tran-Thanh and Levendovszky [18] developed two data collection schemes that takes energy harvesting into account. However, as mentioned earlier, their model is not realistic, since they ignored the receiving cost of data forwarding.

3 Model Description

WSN is typically a set of wireless sensor nodes, including a base station, that collects the data from the nodes by multi-hop packet forwarding via certain paths. Let $N = \{1, 2, \dots, n\}$ be the set of nodes, and let BS denote the base station. We assume that all nodes are static and their battery are rechargeable. Consequently, the main goal here in our case is not to minimize the energy consumption of the nodes, since the energy can be refilled from time to time.

However, due to the limited capacity of the battery and the fact that energy harvesting does not happen at every single time slot, we assume that at every time slot, each node i has a bound for its available energy called "energy capacity" for both transmission and receiving, denoted with B_i^{Tx} , and B_i^{Rx} , respectively. That is, during that time slot, the transmission and receiving energy consumption of node i cannot exceed B_i^{Tx} , and B_i^{Rx} , respectively. For the sake of simplicity, we assume that both B_i^{Tx} and B_i^{Rx} are fixed overtime.

Now, let P_i^{Tx} denote the transmission power level of node i . Let τ_i^b denote the time needed to transmit a single bit from node i . If $l_{i,j}(t)$ denotes the size of total data transmitted via link (i, j) at time slot t , then

$$P_i^{Tx} \tau_i^b \sum_{j \in O_i} l_{i,j}(t) \leq B_i^{Tx}, \forall i \in N, \quad (1)$$

where O_i is the set of node i 's neighbors, which can receive data from node i . That is, the energy consumption of total transmitted data at time slot t cannot exceed the transmitting energy capacity.

Similarly, let P_i^{Rx} denote the receiving power level of node i . Let σ_i^b denote the time needed to receive a single bit at node i . Given this, we have:

$$P_i^{Rx} \sigma_i^b \sum_{j \in I_i} l_{j,i}(t) \leq B_i^{Rx}, \forall i \in N, \quad (2)$$

where I_i is the set of node i 's neighbors, which can transmit data to node i .

We assume that each node has a limited memory, that is, it can receive a limited size of data at each time slot (however, the BS does not have memory limit). Let Q_i denote this limit ($Q_{BS} = \infty$). Thus, one can write:

$$Re_i(t-1) + \sum_{j \in I_i} l_{j,i}(t) - \sum_{j \in O_i} l_{i,j}(t) \leq Q_i, \forall i \in N, \quad (3)$$

where $Re_i(t-1)$ is the amount of residual data (i.e. data left in the memory) of node i after time slot $(t-1)$. Note that

$$Re_i(t) = Re_i(t-1) + \sum_{j \in I_i} l_{j,i}(t) - \sum_{j \in O_i} l_{i,j}(t) \quad (4)$$

In real-world, data loss may occur, due to signal interference and channel fading. When data loss occurs, re-transmission is needed in order to deliver data to the destination. Thus, data loss has a large impact on the value of effective bandwidth (i.e. effective capacity of a communication link). As mentioned earlier, other parameters such as the channel coding technique also affects the value of the effective bandwidth. Here, we assume that both the sender and the receiver nodes of the communication link know this value. Let $c_{i,j}$ denote the effective bandwidth of the communication link between nodes i and j . If there is no communication link between these nodes, then $c_{i,j} = 0$. Thus, we have

$$l_{i,j}(t) \leq c_{i,j} \tau, \forall i \in N, \forall j \in N \cup \{BS\} \quad (5)$$

where τ is the length of a single time slot (we assume that every time slot has the same length). That is, the transmitted data over link (i, j) at each time slot is limited by the link's effective bandwidth.

Finally, we assume that data received at each node i at time slot t can only be forwarded from slot $(t+1)$. This assumption is reasonable, since without it, data could be delivered to the BS instantaneously.

4 Maximal Data Collection with Time Limit

In this section, we study data collection with time limit in the WSN domain. In particular, we concentrate on applications in which the goal is to maximize the total collected data from the network, given a fixed time limit. Here we use the notations defined in Section 3. Let M denote the set of nodes which want to send data to the BS. If $i \in M$ then let m_i denote the size of data (in bytes) node i want to send. Given time limit T , our goal here is to maximize the following formula:

$$\max \sum_{t=1}^T \sum_{j \in I(BS)} l_{j,BS}(t) \quad (6)$$

s.t.

$$\sum_{t=1}^T \sum_{p \in O_i} l_{i,p}(t) \leq m_i + \sum_{t=1}^T \sum_{q \in I_i} l_{q,i}(t), \forall i \in M \quad (7)$$

$$\sum_{t=1}^T \sum_{p \in O_j} l_{j,p}(t) \leq \sum_{t=1}^T \sum_{q \in I_j} l_{q,j}(t), \forall j \in N/M \quad (8)$$

These two constraints guarantee that the nodes cannot send more than the amount of received data (including the amount of collected data as well). In addition, we also take the constraints mentioned in Section 3 (equations 1, 2, 3 and 5).

Given this, we transform the problem into a max flow problem in order to find the optimal solution for the problem defined in formula (6). It can be done as follows. Let G be the graph with vertices r_i^t, v_i^t and u_i^t , for all $i \in N$ and $t \in [0, T]$. Here, each tuple of vertices (r_i^t, v_i^t, u_i^t) represents the node i at time slot t . If nodes i and j in the WSN network can communicate to each other (i.e. there is a communication link between them), then in the graph G , for all $t \in [0, T-1]$ we connect u_i^t with r_j^{t+1} by a link with capacity $c_{i,j} \tau$.

Now, to model the receiving capacity of each node i , we do the following. Since the energy limit for receiving at node i is B_i^{RX} , the maximal size of data node i can receive at each time slot t is $\frac{B_i^{RX}}{P_i^{RX} \sigma_i^b}$. Given this, let connect vertices r_i^t and v_j^t by a link with capacity $\frac{B_i^{RX}}{P_i^{RX} \sigma_i^b}$.

Similarly, to model the transmission capacity of each node i , for each i , we connect v_i^t with u_i^t by a link with capacity $\frac{B_i^{TX}}{P_i^{TX} \tau_i^b}$, where $t \in [0, T-1]$.

Furthermore, to model that the memory limit for each node i is Q_i , we connect v_i^t with v_i^{t+1} by a link with capacity Q_i .

In addition, we also add a vertex S to G , and if node $i \in M$, then we connect S with v_i^0 in G by a link with capacity m_i (i.e. the amount of data to be delivered on node i). Finally, we add a vertex BS to G as well, this vertex represents the BS in the WSN. If node i can communicate directly with the BS in the WSN, then we connect r_i^t with BS via a link with capacity $c_{i,BS} \tau$, for each $t \in [1, T]$. Given this directed acyclic graph (DAG) G , our goal is to find the maximal flow from S to BS , with respect to the given link capacities.

Here, the similarity between this problem and our original, maximal data collection with given time limit problem is clear. Thus, we focus on solving this maximal flow problem. In the literature, a number of polynomial time decentralized algorithms have been proposed for the maximal flow problem, such as the pre-flow-push approach of Goldberg and Tarjan [9], and the augmenting path method introduced by Ford-Fulkerson [8]. In this paper, we rely on the method proposed by Pham et al. [14]. This distributed algorithm is based on the aforementioned preflow-push technique, which can be briefly described as follows.

Let V and E denote the set of vertices and edges of the flow network, respectively. Let $c(e)$ denote the capacity of edge $e \in E$. finally, let s and t denote the source and sink nodes of the network, respectively. Each node $v \in V$ maintains a height func-

tion $h(v)$ such that it satisfies: (i) for the sink node t , $h(t) = 0$; and (ii) $h(u) \geq h(v) + 1$ if $(u, v) \in E$. In fact, $h(v)$ is a lower bound on the shortest path (on number of hops) from node v to the sink t . Furthermore, let $excess(v)$ denote the difference between the total incoming flow and the total outgoing flow of node v , that is:

$$excess(v) = \sum_{u \in I_v} f(u, v) - \sum_{w \in O_v} f(v, w) \quad (9)$$

where $f(i, j)$ is the flow function of edge $(i, j) \in E$. Based on this, we can define the pre-flow as follows. A preflow f is a function $f : E \rightarrow \mathbb{R}^+$ with:

- $0 \leq f(e) \leq c(e), \forall e \in E$
- $excess(v) \geq 0, \forall v \in V \setminus \{s, t\}$

Let us note that our original problem also defines a pre-flow (see equations 7 and 8).

The idea of the algorithm is to *push* the excess of a node in a pre-flow towards the sink t , from a vertex with higher height to one of its neighboring vertices with lower height. In order to do this in a decentralized environment, each of the nodes executes asynchronously a program that coordinates the moves of the algorithm. For more details, see [14].

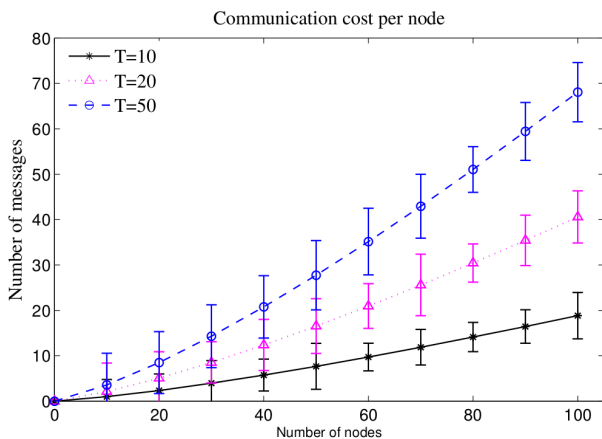


Fig. 1. The communication cost of the proposed algorithm.

However, in our case, each physical node i has $3(T + 1)$ virtual instances ($r_i^t, v_i^t, u_i^t, t \in [0, T]$) in the network flow. That is, we have to modify the algorithm, since real physical communication is allowed only between real nodes. Given this, we make two different modifications as follows.

- Each node i runs $3(T + 1)$ instances of the program, representing virtual vertices r_i^t, v_i^t, u_i^t .
- To make the communication of virtual vertices possible, each physical message must contain the ID of the sender and receiver vertices. For instance if u_i^t wants to communicate with r_j^{t+1} , then the message is sent via the communication link (i, j) with sender ID $\langle u_i^t \rangle$ and receiver ID $\langle r_j^{t+1} \rangle$.

In the cited work [14], the authors proved that the algorithm has $O(n^2m)$ communication complexity (overhead) and $O(n^2)$

time complexity, where n is the number of vertices and m is the number of edges in the flow network. Here, since each node i of the WSN has $3(T + 1)$ instances of vertices in G , thus, $n = 3N(T + 1)$. To calculate m , we do the following. Each real communication link is represented T times in G , thus we have $m_{WSN}T$ links that represent the communication links, where m_{WSN} is the number of communication links in the WSN. Furthermore, each r_i^t is connected with v_i^t, v_i^t with u_i^t , and v_i^t with v_i^{t+1} . That is, in addition, we have another $(3T - 1)N$ links. Thus, $m = m_{WSN}T + (3T - 1)N$. Therefore, the complexity of the algorithm is $O(9N^2T^3m_{WSN} + 9N^3T^2(3T - 1))$ in communication and $O(9N^2(T + 1)^2)$ in time in our case.

Since the communication complexity is $O(9N^2T^3m_{WSN} + 9N^3T^2(3T - 1))$, this indicates that the algorithm needs approximately $9N^2T^3m_{WSN}$ messages on average in order to achieve optimal routing scheme (m_{WSN} typically dominates N , that is, m_{WSN} is $O(N^2)$). However, this complexity is typically significant, compared to the communication complexity of other data collection algorithms. Nevertheless, by using extensive simulations, we demonstrate that the communication overhead of the algorithm is typically low in WSNs, compared to the number of the nodes. In so doing, we set up a simulation environment, where all the parameters, such as number of nodes, energy consumption values, memory limit, and number of packets to send, are all randomly set. We group the simulation results into groups by the number of nodes. Thus, our main focus was to measure the average number of messages per node needed in the algorithm. The results are depicted in Fig. 1. According to the results, the number of messages per node in the worst case is linear, compared to the total network size (number of nodes in the network). This is much better than the previously determined theoretical communication complexity

5 Conclusions

In this paper, we have introduced a relaxation of the TDDC problem, namely: the maximal data collection with given time limit problem. This problem is relevant for WSNs, in which nodes are capable of energy recharging, and the communication links are limited in bandwidth. For this relaxed problem, we proposed a polynomial time algorithm, that achieves optimal solution. We have also demonstrated that the proposed algorithm has low communication overhead as well.

References

- 1 Akkaya K, Younis M, *Energy-aware delay-constrained data in wireless sensor networks.*, Journal of Communication Systems **17** (2004), no. 6, DOI 10.1002/dac.v17:6.
- 2 Alfandari L, Paschos V T, *Approximating minimum spanning tree of depth 2*, International Transactions in Operational Research **6** (1999), 222-607, DOI 10.1016/S0969-6016(99)00018-0.
- 3 Althaus E, Funke S, Harpeled S, *Approximating k-hop Minimumspanning trees*, Operations Research Letters **33** (2005), no. 2, 115-120, DOI 10.1016/j.orl.2004.05.005.

- 4 **Ammari H M, Das S K**, *A trade-off between energy and delay in data dissemination for wireless sensor networks using transmission range slicing*, *Computer Communications* **31** (2008), no. 9, 1687-1704, DOI 10.1016/j.comcom.2007.11.006.
- 5 **Beeby S P, Tudor M J, White N M**, *Energy harvesting vibration sources for microsystems applications*, *Measurement Science and Technology* **9** (2006), 175-195, DOI 10.1088/0957-0233/17/12/R01.
- 6 **Cheng H, Liu Q, Jia X**, *Heuristic algorithms for real-time data aggregation in wireless sensor networks*, *International Conference on Communications and Mobile Computing*, posted on 2006, DOI 10.1145/1143549.1143774., (to appear in print).
- 7 **Chong C Y, Kumar S P**, *Sensor networks: Evolution, opportunities and challenges*, *Proceedings of IEEE* **91** (2003), no. 8, 1247-1256, available at doi:10.1.1.74.2558.
- 8 **Ford L R, Fulkerson D R**, *Flows in networks*, Princeton University Press, Princeton, 1962.
- 9 **Goldberg A V, Tarjan R E**, *A New Approach to the Maximum flow Problem*, *Journal of ACM* **35** (1988), no. 4, 921-940, available at doi:10.1145/12130.12144.
- 10 **Goldsmith A, Wicker S**, *Design challenges for energy-constrained ad hoc wireless networks*, *IEEE Wireless Communications Magazine* **9** (2002), 8-27.
- 11 **He L**, *Delay-minimum energy-aware routing protocol (DERP) for wireless sensor networks*, *Eighth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel Distributed Computing* **3** (2007), 155-160.
- 12 **Lewis F L**, *Wireless sensor networks*, *Smart Environments: Technologies, Protocols, and Applications*, 2004.
- 13 **Mathur G, Desnoyers P, Ganesan D, Shenoy P**, *Ultra-low power data storage for sensor networks*, *Proceedings of the fifth international conference on Information processing in sensor networks*, posted on 2006, 374-381, DOI 10.1109/MWC.2002.1028874, (to appear in print).
- 14 **Pham T L, Bui M, Lavallee I, Do S H**, *A Distributed Preflow-Push for the Maximum flow Problem*, *Innovative Internet Community Systems*, posted on 2006, 195-206, DOI 10.1007/11749776.17, (to appear in print).
- 15 **Rogers A, Corkill D D, Jennings N R**, *Agent technologies for sensor networks*, *IEEE Intelligent Systems* **24** (2009), no. 2, 13-17, available at doi:10.1109/MIS.2009.22.
- 16 **Stankovic J A**, *Research challenges for wireless sensor networks*, *ACM SIGBED Review* **1** (2004), no. 2, 9-12, DOI 10.1145/1121776.1121780.
- 17 **Torah R, Glynne-Jones P, Tudor M, O'Donnell T, Roy S, Beeby S**, *Self-powered autonomous wireless sensor node using vibration energy harvesting*, *Measurement Science and Technology*, posted on 2008, 125202.1-125202.8, DOI 10.1088/0957-0233/19/12/125202, (to appear in print).
- 18 **Tran-Thanh L, Leventovszky J**, *Optimal Time Dependent Data Collection Schemes in Wireless Sensor Networks*, *Proceedings of the Eleventh IEEE Wireless Communications and Networking Conference* (2010), to appear.
- 19 **Zhang P, Sadler C M, Lyon S A, Martonosi M**, *Hardware design experiences in zebrant*, *Proceedings of the Second international conference on Embedded networked sensor systems* **19** (2004), 227-238, DOI 10.1145/1031495.1031522.
- 20 **Xu H, Huang L, Liu W, Wang G, Wang Y**, *Topology control for delay-constraint data collection in wireless sensor networks*, *Computer Communications* **32** (2009), no. 17, 1820-1828, DOI 10.1016/j.comcom.2009.07.002.