System Modeling And Analysis On Energy Optimized Data Collection

K. Indraneel [1], S. Mahesh Kumar [2]

[1]Asso.Professor, Sri Kottam Tulasi Reddy Memorial College of Engineering, Kondair, A.P, India
[2]M.Tech(CSE), Sri Kottam Tulasi Reddy Memorial College of Engineering, Kondair, A.P, India

Abstract

In wireless sensor networks converge cast is the fundamental operation and namely the collection of data from a set of sensors toward a common sink over a tree- based routing topology. In many applications, it is crucial to provide a guarantee on the delivery time as well as increase the rate of such data collection. For instance, in safety and mission-critical applications where sensor nodes are deployed to detect oil/gas leak or structural damage, the actuators and controllers need to receive data from all the sensors within a specific deadline, failure of which might lead to unpredictable and catastrophic events. This falls under the category of one-shot data collection. On the other hand, applications such as permafrost monitoring require periodic and fast data delivery over long periods of time, which falls under the category of continuous data collection.

In this paper we are concentrated on system modeling with throughput estimation and a clear study about the impact interference and routing trees.

1. Introduction

A tree base sensor network is a collection of sensors nodes, such as sink is the root of tree and leaves are the nodes. Data in such topology flows from sensor nodes (leaves) to the sink (root) of the tree. Collection of data from a set of sensors to an intermediate parent (sink) in a tree is known as converge-casting. The 'delivery-time' and 'data-rate' are application specific. As an example, in oil and gas refineries the sensor devices and controllers need to collect data from all the sensors within a specific deadline for any kind of leakage or failures. Whereas applications like weather for-casting, under-water observations needs continuous and fast data delivery for analysis, for longer periods. Here in this paper our emphasis is on such applications focusing on fast data streaming from sensor to sink node. The two common approaches for data collection are – aggregated-data converge cast where packets are aggregated at each hop, and raw-data converge cast where each data packet travel towards sink node individually. First approach is most suitable where data is highly co-related and objective is to collect maximum sensor reading and second approach is used where the reading of each sensor is equally important. Further, interference and network topology are the two prime limiting factors in wireless sensor networks. Time Division Multiple Access (TDMA) protocol is well suited to periodic data traffic to have contention free medium and to avoid collisions. The use of multiple frequency channels can allow more concurrent transmissions. Here we have shown that if multiple frequencies are employed along with TDMA, the data collection rate is affected by tree topology and not by interferences. Thus, in this paper we identify the effect of network topology on the schedule length, analyzed the performance of converge cast by using multiple frequencies as compared to those trees using a single frequency.

Periodic collection of aggregated data from sensors to a common sink over a tree topology is a fundamental operation in wireless sensor networks (WSN). In many such applications, it is of interest to maximize the rate at which the sink can receive aggregated data from the network. For instance, it has been noted that in networked structural health monitoring more than 500 samples per second are required to efficiently detect damages. Time division multiple access (TDMA) scheduling is a natural solution for such periodic data collection applications. We explore a number of techniques in order to address the basic question: "how fast can aggregated data be streamed to the sink"? These techniques provide a hierarchy of successive improvements. The simplest approach is to do some form of interference-aware minimum frame-length TDMA-scheduling that enables spatial reuse. The second step is to combine the scheduling with transmission power control. The third step is to consider the use of multiple frequency channels. We show that once multiple frequencies are employed along with spatial-reuse TDMA, the aggregated data collection rate often becomes no longer interference-limited, but rather topology-limited. Thus, the final step to enhance the rate of periodic aggregated data collection is to use an appropriate degree- constrained tree topology.

2. Related Work

Given more than a decade of research on wireless sensor networks, the number of papers written on converge cast and multi-channel protocols is too vast to enumerate comprehensively. Instead, we focus our review on key works, particularly those that have focused on real implementation and experimental validation over test beds. Although it has not been designed for high-rate performance, the de-facto state of the art routing protocol for WSN today, which offers the best, reliable delivery performance, is the Collection Tree Protocol (CTP). We therefore use it as a baseline comparison for MCC. Alternate routing approaches that also show performance improvements are Arbutus, which offers improvements in load balancing and reliability, and the queue- aware dynamic routing mechanism BCP, which shows improvement in throughput as well as robustness to external interference. Others have focused on avoiding congestion collapse and maintaining high rate delivery by using a rate control protocol on top of the routing mechanism. Examples include IFRC, and WRCP. A centralized approach to rate control, RCRT, has been shown to yield even better rate performance. Flush offers a robust, high-rate connection- oriented bulk transfer capability. All these protocols are single- channel protocols, and have been developed over CSMA, due to ease of implementation.

Fast data collection with the goal to minimize the schedule length for aggregated converge cast has been studied. We experimentally investigated the impact of transmission power control and multiple frequency channels on the schedule length, while the theoretical aspects were discussed, where we proposed constant factor and logarithmic approximation algorithms on geometric networks (disk graphs). Raw-data converge cast has been studied where a distributed time slot assignment scheme is proposed by Gandham et al. to minimize the TDMA schedule length for a single channel. The problem of joint scheduling and transmission power control is studied by Moscibroda for constant and uniform traffic demands. Our present work is different from the above in that we evaluate transmission power control under realistic settings and compute lower bounds on the schedule length for tree networks with algorithms to achieve these bounds. We also compare the efficiency of different channel assignment methods and interference models, and propose schemes for constructing specific routing tree topologies that enhance the data collection rate for both aggregated and raw-data converge cast. The use of orthogonal codes to eliminate interference has been studied by Annamalai et al. where nodes are assigned time slots from the bottom of the tree to the top such that a parent node

does not transmit before it receives all the packets from its children. This problem and the one addressed by Chen et al. are for one-shot raw-data converge cast. In this work, since we construct degree-constrained routing topologies to enhance the data collection rate, it may not always lead to schedules that have low latency, because the number of hops in a tree goes up as its degree goes down. Therefore, if minimizing latency is also a requirement, then further optimization, such as constructing bounded-degree, bounded-diameter trees, is needed. A study along this line with the objective to minimize the maximum latency is presented by Pan and Tseng, where they assign a beacon period to each node in a Zigbee network during which it can receive data from all its children. For raw-data converge cast, Song et al. presented a time-optimal, energy-efficient, packet scheduling algorithm with periodic traffic from all the nodes to the sink. Once interference is eliminated, their algorithm achieves the bound that we present here, however, they briefly mention a 3-coloring channel assignment scheme, and it is not clear whether the channels are frequencies, codes, or any other method to eliminate interference. Moreover, they assume a simple interference model where each node has a circular transmission range and cumulative interference from concurrent multiple senders is avoided. Different from their work, we consider multiple frequencies and evaluate the performance of three different channel assignment methods together with evaluating the effects of transmission power control using realistic interference and channel models, i.e., physical interference model and overlapping channels and considering the impact of routing topologies. Song et al. extended their work and proposed a TDMA based MAC protocol for high data rate WSNs. Tree MAC considers the differences in load at different levels of a routing tree and assigns time slots according to the depth, i.e. the hop count, of the nodes on the routing tree, such that nodes closer to the sink are assigned more slots than their children in order to mitigate congestion. However, Tree MAC operates on a single channel and achieves 1/3 of the maximum throughput similar to the bounds presented by Gandham et al. since the sink can receive every 3 time slots. The problem of minimizing the schedule length for raw-data converge cast on single channel is shown to be NP-complete on general graphs by Choi et al. Maximizing the throughput of converge cast by finding a shortest-length, conflict free schedule is studied by Lai et al. where a greedy graph coloring strategy as- signs time slots to the senders and prevent interference. They also discussed the impact of routing trees on the schedule length and proposed a routing scheme called disjoint strips to transmit data over different shortest paths. However, since the sink remains as the bottleneck,

sending data over different paths does not reduce the schedule length. As we will show in this paper, the improvement due to the routing structure comes from using capacitated minimal spanning trees for raw-data converge cast, where the number of nodes in a subtree is no more than half the total number of nodes in the remaining subtrees. The use of multiple frequencies has been studied extensively in both cellular and ad hoc networks, however, in the domain of WSN, there exist a few studies that utilize multiple channels. To this end, we evaluate the efficiency of three particular schemes that treat the channel assignment at different levels.

A comprehensive survey of multi-channel mechanisms for wireless networks (mostly for 802.11-based ad-hoc networks) was made as research. MMSN is the first multichannel MAC protocol especially designed for WSNs with devices having half-duplex single transceiver. There is a common broadcast channel, and nodes contend to access the channel on different unicast frequencies. The authors propose a multi-channel protocol with dynamic channel allocation by clustering a WSN. In each cluster, the header node needs to collect information and schedule for its member nodes. TMCP is a tree-based channel allocation mechanism, in which the tree is partitioned into separate trees, each of which is allocated a separate channel (minimizing the need for synchronization). Other approaches to multi-channel MAC design include the cluster-based dynamic control-theoretic approach, and MC-LMAC, which offers a distributed joint time and frequency scheduling mechanism. These schemes have all been evaluated primarily through simulations alone, or with limited test bed experiments (as in the case of TMCP). They are also not guaranteed to offer maximum throughput converge cast. We now turn to practically implemented mechanisms for multichannel collection in wireless sensor networks that have been evaluated through extensive test bed experiments. Y- MAC presents a multi-channel protocol for wireless sensor networks that is based on lightweight channel hopping. Nodes on a link hop to a new channel when traffic bursts occur, following a predetermined sequence. The focus of Y- MAC is to improve energy efficiency by reducing contention, rather than high rate performance, and it does not offer any guarantees in this regard. WRAP uses multiple frequency channels with time synchronization. Data collection in WRAP is implemented using a token-passing scheme. Designed for highly dense deployments, it allows only one node in the network to be actively transmitting at any time. For this reason, while it does cut down drastically on interference and congestion, WRAP does not guarantee maximum rate performance in an arbitrary network.

In its basic form, it allows the sink to establish a connection with a particular sensor node and download data from that node at the highest rate possible. By keeping the sink occupied half the time (it must remain idle whenever the node one hop from it is in receive mode) it can achieve at least 50% of the maximum throughput. The authors of this work point out that to fully occupy the sink, with sufficient channels, it is possible to schedule two concurrent flows, keeping the sink fully occupied and thus achieving the maximum possible throughput, which is also our goal in this work. A key distinction, however, is that the MCC protocol we describe in this work is able to collect data fairly from all nodes in the network, rather than establishing connections only to one or two at a time. Thus, it is more suitable for real-time, fair data gathering applications. Also, MCC incorporates a balanced routing mechanism to ensure that the maximum rate is achievable. Moreover, we carefully evaluate the maximum possible network throughput with respect to packet size, net- work topology, use of ACK, etc., and demonstrate conclusively that this rate can be achieved in practice by our protocol.

3. System Modeling

Let $G = (V, E)$ is a multi-hop WSN graph, where V is the set of sensor nodes, and $E = (I, j)$: $(I, j) \in V$ is the set of wireless links. Let s is the sink node such that $s \in V$. The distance between two nodes i and j is denoted by d_{ij}. All the nodes other than s generate and transmit data packets through a network path to sink s. Let, $T = (V, ET)$ is a spanning tree on G where ET E and represents the tree edges. It is assumed that each node has half duplex transceiver; therefore it cannot simultaneously send and receive data. We have used equal sized timeframe TDMA protocol and two types of interference models for analysis namely: SINR based physical model and graph based model. The interference range of a node is equal to its transmission range which means two links cannot be scheduled at the same time if receiver of one link is within the transmitter range of the other link. In SINR model the successful reception of a packet from i to j depends on cumulative interference caused by all concurrent transmitting nodes and the ratio between the received signal strength at j. The size of each data packet is assumed to be same. For fast data routing we aim to schedule the edges ET of T using a minimum number of time slots with two constraints:

Adjacency constraint: it states that two edges in ET cannot be scheduled in same time slot if they are adjacent to each other. This is because of half duplex transceiver available on nodes.

Interfering constraint: The interfering constraint depends on the choice of the interference model.

For a periodic data collection in aggregated converge cast each edge in ET is scheduled in a pipeline manner. The sink receives packets from the pipeline one after another. On the other hand, for raw data converge cast the edge in ET is scheduled multiple times hence no pipeline is there.

4. Throughput Estimation

Although the maximum throughput may naively be considered to be simply the link rate of the underlying radio protocol, in practice the throughput may be even lower due to hardware limitations. In this section, we systematically study one widely used WSN hardware platform, the Tmote Sky, and understand its achievable rate performance with respect to different key parameters.

Table 1: Parameters for the Curve-Fit of Throughput Estimation Model

	α (ms)	β (ms/byte)
Sink, no ACK	1.79	0.062
Relay, no ACK	3.50	0.079
Leaf, no ACK	3.35	0.079
Sink, ACK	3.95	0.078
Relay, ACK	5.81	0.085
Leaf, ACK	5.52	0.079

A. Basic Model

We show in Fig. 1 a high level view of a node's architecture (as our focus is on understanding throughput performance for the communication stack, we omit the sensor itself in this figure). The node has a micro-controller, a radio with 2 separate buffers (one for TX, another for RX), and a bus to copy data between them. The total time for transmission or reception of a single packet, TPacket, can be written as a linear combination of a constant term and a term proportional to the packet size.

$$T_{Packet} = \alpha + \beta \cdot Packet\_Size$$
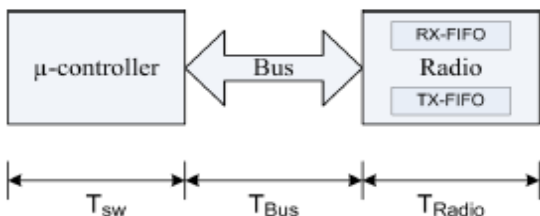
Figure 1: A High Level View of Sensor Node Architecture

The α pertains to per-packet software and radio latencies that are independent of the packet size. The rate term β depends on the CPU rate, the bus

transfer rate, and the radio rate, and the degree of pipelining achieved between successive packets. While α is very software-specific, based on hardware specs for the Tmote Sky device, we can estimate that β lies between 0.04 and 0.09 ms / byte. The corresponding throughput can then be calculated as:

$$Throughput = \frac{Packet\_Size}{T_{Packet}}$$
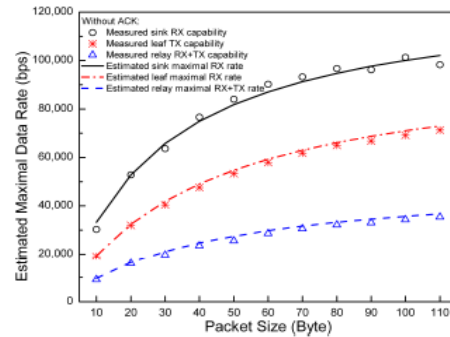
Figure 2: Measured versus estimated throughput of Tmote Sky device for ACK-disabled case.

B. Estimating the Maximum Converge cast Throughput

In practice, the measurement of Tpacket and the corresponding throughput will be different depending on a nodes role and where it is being measured. Since we are concerned with a converge cast application, the three key node roles are sink, relay, and leaf. We therefore conduct experiments to estimate three kinds of throughput values: the receive rate of a sink node, the transmit rate of a leaf node, and the (transmit + receive) rate of a relay node. In these experiments we allocate time slots to nodes and use synchronized transmissions.
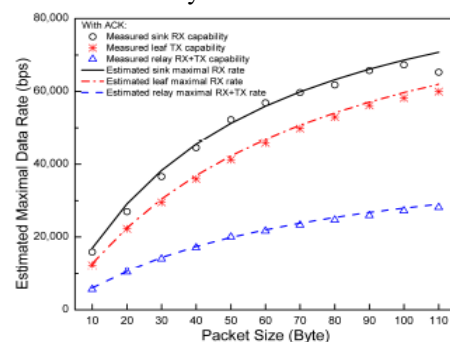
Figure 3: Measured versus estimated throughput of Tmote Sky device for ACK-enabled case

To measure sink receive rate, we use a star-topology with the sink receiving packets from multiple nodes one hop away from it. To measure the leaf transmit rate, we use a simple two-node link with the transmitter continuously sending packets to the receiver. To measure the relay rate, we consider a linear topology with each node

receiving on a different channel to avoid interference. We measure the relay rate of intermediate nodes as they alternate between sending and receiving packets. We vary the packet size from 10 bytes to 110 bytes and conduct the experiments with and without link-layer acknowledgement packets. For each setting we vary the slot length of packet transmissions to determine the maximum achievable rate and plot only those maximum rate points. The results we obtain are shown in Fig. 2 and Fig. 3.

C. Lessons

The modeling and throughput experiments summarized in Fig. 2 and Fig. 3 yield three key observations that inform our design of a high-throughput collection protocol:

- The throughput performance is very much a concave function of the packet size. To our knowledge, ours is the first work to systematically quantify using real WSN hardware how the throughput varies with the packet size.
- The throughput is higher without ACK than with ACK packets. This is intuitive, as there is additional over- head incurred in generating and waiting for acknowledgements. However, using ACK feedback will improve reliability.
- The throughput performance is clearly affected significantly by the role of a node. We find that the maximum sink receive rate is slightly higher than the maximum leaf transmission rate.

This is due to the possibility of greater pipelining at the sink which is receiving data from multiple transmitters. Thus, in a linear topology, the throughput bottleneck would be the relay node's maximum throughput. Note though that if we have even three different branches, assuming all transmissions are scheduled to avoid collisions, the sum rate of these branches exceeds the maximum sink receive rate. Thus the maximum collection rate of converge cast is bounded by the maximum sink receive rate. We will therefore compare the throughput achieved by our protocol with the maximum estimated sink throughput.

5. Impact of Interference and Routing Trees

So far, we have focused on computing spatial-reuse TDMA schedules where transmissions take place on the same frequency at a constant transmission power. In this section, we focus on different methods to mitigate the effects of interference on the schedule length. First, we discuss the benefits of using transmission power control and explain the basics of a possible algorithm. Then we discuss the advantages of using

multiple channels by considering 3 different channel assignment schemes.

- Transmission Power Control
- Multi-Channel Scheduling
- Joint Frequency Time Slot Scheduling (JFTSS)
- Tree Based Multi-Channel Protocol (TBMCP)
- Receiver Based Channel Assignment (RBCA)

Besides transmission power control and multiple channels, the network topology and the degree of connectivity also affect the scheduling performance. In this section, we describe schemes to construct topologies with specific properties that help to reduce the schedule length.

- Aggregated Data Collection
- Raw Data Collection

To illustrate the gains of degree-constrained trees, consider the case when all the N nodes are in range of each other and that of the sink. If the nodes select their parents according to minimum-hop without a degree constraint, then all of them will select the sink, and this will give a schedule length of N. However, if we limit the number of children per node to 2, then this will result in two subtrees rooted at the sink, and if there are enough frequencies to eliminate interference, the network can be scheduled using only 2 time slots, thus achieving a factor of N/2 reduction in the schedule length.

Algorithm 1 BFS-TimeSlotAssignment

1. Input: $T = (V, E_T)$
2. while $E_T \neq \phi$ do
3. $e \leftarrow$ next edge from E_T in BFS order
4.   Assign minimum time slot t to edge e respecting adjacency and interfering constraints
5. $E_T \leftarrow E_T \setminus \{e\}$
6. end while

Algorithm 2 Local-TimeSlotAssignment

1. $node.\text{buffer} = full$
2. if $\{node$ is sink$\}$ then
3.   Among the eligible top-subtrees, choose the one with the largest number of total (remaining) packets, say top-subtree i
4.   Schedule link $(root(i), s)$ respecting interfering constraint
5. else
6.   if $\{node.\text{buffer} == empty\}$ then
7.   Choose a random child c of $node$ whose buffer is $full$
8.   Schedule link $(c, node)$ respecting interfering constraint
9. $c.\text{buffer} = empty$
10. $node.\text{buffer} = full$
11.   end if
12. end if

Algorithm 3 DEGREE-CONSTRAINED TREES

1. **Input**: $G(V, E)$, s, max_degree
2. $T \leftarrow \{s\}$
3. **for all** $i \in V$ **do**
4. $\quad C(i) \leftarrow 0$; $HC(i) \leftarrow \infty$
5. **end for**
6. $HC(s) \leftarrow 0$
7. **while** $|T| \neq |V|$ **do**
8. $\quad$ Choose $i' \notin T$ such that:
9. $\quad\quad$ (a) $(i, i') \in E$, for some $i \in T$ with $C(i) < max\_degree - 1$
10. $\quad\quad$ (b) $HC(i')$ is minimized
11. $\quad T \leftarrow T \cup \{i'\}$
12. $\quad HC(i') = HC(i') + 1$
13. $\quad C(i) \leftarrow C(i) + 1$
14. $\quad$ **if** $\forall i \in V$, $C(i) = max\_degree$ **then**
15. $\quad\quad$ **break**
16. $\quad$ **end if**
17. **end while**

6. Conclusion

In this paper, we studied fast converge cast in WSN, addressed the fundamental limitations due to interference and half-duplex transceivers on the nodes and explored techniques to overcome the same. We found that while transmission power control helps in reducing the schedule length, and also observed that node-based (RBCA) and link-based (JFTSS) channel assignment schemes are more efficient in terms of eliminating interference as compared to assigning different channels on different branches of the tree (TMCP).

Algorithm 4 CAPACITATED-MINIMALSPANNINGTREE

1. **Input**: $G(V, E)$, s
2. **Initialize**:
3. $\quad B \leftarrow$ roots of top subtrees $\quad$ // the branches
4. $\quad T \leftarrow \{s\} \cup B$
5. $\quad \forall i \in V$, $GS(i) \leftarrow$ unconnected neighbors of i at further hops
6. $\quad \forall b \in B$, $W(b) \leftarrow 1$
7. $\quad h \leftarrow 2$
8. **while** $h \neq max\_hop\_count$ **do**
9. $\quad N_h \leftarrow$ unconnected nodes at hop distance h
10. $\quad$ Connect nodes N'_h that have a single potential parent: $T \leftarrow T \bigcup N'_h$
11. $\quad$ Update $N_h \leftarrow N_h \setminus N'_h$
12. $\quad$ Sort N_h in non-increasing order of $|GS|$
13. $\quad$ **for all** $i \in N_h$ **do**
14. $\quad\quad$ **for all** $b \in B$ to which i can connect **do**
15. $\quad\quad\quad$ Construct $SS(i, b)$
16. $\quad\quad$ **end for**
17. $\quad\quad$ Connect i to b for which $W(b) + |SS(i, b)|$ is minimum
18. $\quad\quad$ Update $GS(i)$ and $W(b)$
19. $\quad\quad T \leftarrow T \bigcup \{i\} \bigcup SS(i, b)$
20. $\quad$ **end for**
21. $\quad h \leftarrow h + 1$
22. **end while**

In this paper, we have discussed fast converge cast methods in wireless sensor network, where nodes communicate using TDMA protocol so as to minimize the scheduling length. We have focused on fundamental shortcoming because of interference and half duplex transceivers available on the nodes. We observed that multiple channel method is helpful in reducing schedule length. We also determined that link-based (JFTSS) channel assignment schemes are more energy efficient in removing interference, if compared to TMCP scheduling schemes.

We have explored a number of techniques to enhance the aggregated data collection over a tree topology in WSN. We have proposed and implemented MCC, a new high-rate, multi-channel time-scheduled protocol for converge cast data collection in wireless sensor networks.

7. References

[1]. Ozlem Durmaz Incel, Amitabha Ghosh, Bhaskar Krishnamachari, and Krishnakant Chintalapudi ¨ - "Fast Data Collection in Tree-Based Wireless Sensor Networks", IEEE Transaction on Mobile Computing, Vol. 11, No. 1, Jan 2012.

[2]. Sharad, Shailendra Mishra, Ashok Kumar Sharma Durg Singh Chauhan – "Analysis on Energy Optimized Data Collection in Tree Based Ad-Hoc Sensor Network", I.J.Modern Education and Computer Science, 2012, 3, 42-49.

[3]. Ozlem Durmaz Incel, Bhaskar Krishnamachari D – "Enhancing the Data Collection Rate of Tree-Based Aggregation in Wireless Sensor Networks", IEEE Communications Society subject matter experts for publication in the IEEE SECON 2008 proceedings.

[4]. Ying Chen and Bhaskar Krishnamachari – "MCC: a High-Throughput Multi-Channel Data Collection Protocol for Wireless Sensor Networks", University of Southern California, Los Angeles, CA, 90089.

[5]. Devasahayam, K.P. Kaliyamurthie – "Optimized Rapid Data Collection in Tree Based WSN", IJCSMC, Vol. 2, Issue. 4, April 2013, pg.559 – 567.