ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Tree Based Data Collection from Wireless Sensor Networks

Shaik.Asif Basha
M.Tech Student,
Nimra College of Engineering and Technology.

Gousejohn
Guide,
Nimra College of Engineering and Technology.

ABSTRACT:

Fast data collection with the goal to minimize the schedule length for aggregated converge cast has been studied by us in, and also by others in, we experimentally investigated the impact of transmission power control and multiple frequency channels on the schedule length 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.

1. INTRODUCTION:

Converge cast, namely the collection of data from a set of sensors toward a common sink over a tree based routing topology, is a fundamental operation in wireless sensor networks (WSN). 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 consider such applications and focus on the following fundamental question: "How fast can data be streamed from a set of sensors to a sink over a tree based topology" In this paper study two types of data collection: (i) aggregated converge cast where packets are aggregated at each hop, and (ii) raw-data converge cast where packets are individually relayed toward the sink. Aggregated con-verge cast is applicable when a strong spatial correlation exists in the data, or the goal is to collect summarized information such as the maximum sensor reading. Raw data converge cast, on the other hand, is applicable when every sensor reading is equally important, or the correlation is minimal. We study aggregated converge cast in the context of continuous data collection, and raw data converge cast for one-shot data collection. These two types correspond to two extreme cases of data collection.

In an earlier work, the problem of applying different aggregation factors, i.e., data compression factors, was studied, and the latency of data collection was shown to be within the performance bounds of the two extreme cases of no data compression (raw-data converge cast) and full data compression (aggregated converge cast). For periodic traffic, it is well known that contention free medium access control (MAC) protocols such as TDMA (Time Division Multiple Access) are better fit for fast data collection, since they can eliminate collisions and retransmissions provide guarantee on the completion time as opposed to contention-based protocols. However, the problem of constructing conflict free (interference-free) TDMA schedules even under the simple graph-based interference model has been proved to be NP-complete. In this work, consider a TDMA framework and design polynomial-time heuristics to minimize the schedule length for both types of converge cast.

Fast data collection with the goal to minimize the schedule length for aggregated converge cast has been studied by us in, and also by others in, we experimentally investigated the impact of transmission power control and multiple frequency channels on the schedule length 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.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 etal, where nodes are assigned time slots from the bottom of the tree to the b 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 etal 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 here 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 etal. 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 are 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.

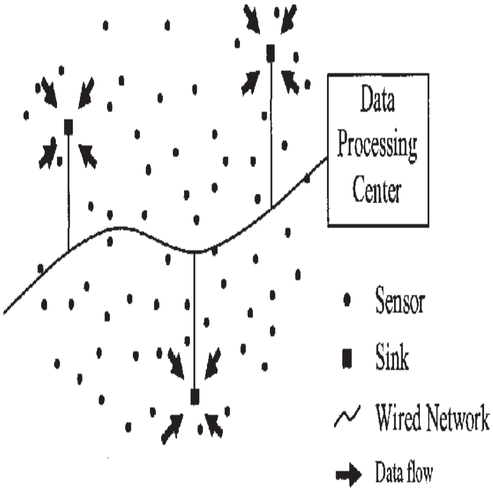
Different from their work, we consider multiple frequencies and evaluate the performance of three different channel assignment methods together with evaluating the effects of transmission power control using realistic interference and channel models, i.e., physical interference model and overlapping channels and considering the impact of routing topologies. Song etal. 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 etal. 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 etal.Maximizing the throughput of converge cast by finding a shortest-length, conflict-free schedule is studied by Lai etal.where a greedy graph coloring strategy assigns 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 sub tree is no more than half the total number of nodes in the remaining sub trees. 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, here evaluate the efficiency of three particular schemes that treat the channel assignment at different levels.

2. SYSTEM ANALYSIS SYSTEM:

a) Architecture:



System Architecture

Existing work had the objective of minimizing the completion time of converge casts. However, none of the previous work discussed the effect of multi-channel scheduling together with the comparisons of different channel assignment techniques and the impact of routing trees and none considered the problems of aggregated and raw converge casts, which represent two extreme cases of data collection.Fast data collection with the goal to minimize the schedule length for aggregated converge cast has been studied by us in, and also by others in, we experimentally investigated the impact of transmission power control and multiple frequency channels on the schedule length 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. Advantage of TMCP is that it is designed to support converge cast traffic and does not require channel switching. However, contention inside the branches is not resolved since all the nodes on the same branch communicate on the same channel.

B) Modeling And Problem Formulation:

In this model the multi-hop WSN as a graph G = (V, E), where V is the set of nodes, E = {(i, j) | i, j V} is the set of edges representing the wireless links.

A designated node s V denotes the sink. The Euclidean distance between two nodes i and j is denoted by dij . All the nodes except s are sources, which generate packets and transmit them over a routing tree to s. We denote the spanning tree on G rooted at s by T = (V, ET), where ET E represents the tree edges. Each node is assumed to be equipped with a single half-duplex transceiver, which prevents it from sending and receiving packets simultaneously. We consider a TDMA protocol where time is divided into slots, and consecutive slots are grouped into equal sized non-overlapping frames. We use two types of interference models for our evaluation the graph-based protocol model and the SINR based physical model. In the protocol model, we assume that the interference range of a node is equal to its transmission range, i.e., two links cannot be scheduled simultaneously if the receiver of at least one link is within the range of the transmitter of the other link .In the physical model, the successful reception of a packet from i to j depends on the ratio between the received signal strength at j and the cumulative interference caused by all other concurrently transmitting nodes and the ambient noise level. Thus, a packet is received successfully at j if the signal-to-interference-plus-noise ratio, S I N Rij , is greater than a certain threshold β, i.e.,

S I N Rij =Pi • gij/_k_=i Pk • gkj + N

Where Pi is the transmitted signal power at node i, N is the ambient noise level, and gij is the propagation attenuation (link gain) between i and j. We use a simple distance dependent path-loss model to calculate the link gains as gij = d−αij, where the path-loss exponent α is a constant between 2 and 6, whose exact value depends on external conditions of the medium (humidity, obstacles, etc.), as well as the sender-receiver distance. We assume that the level of interference is static and does not change over time. For simplicity and ease of illustration, we use the protocol model in all the figures. We study aggregated converge cast in the context of periodic data collection where each source node generates a packet at the beginning of every frame, and raw data converge cast for one-shot data collection where each node has only one packet to send. We assume that the size of each packet is constant. Our goal is to deliver these packets to the sink over the routing tree as fast as possible. More specifically, we aim to schedule the edges ET of T using a minimum number of time slots while respecting the following two constraints:

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

- **Adjacency Constraint:**

Two edges (i, j) ET and (k, l) ET cannot be scheduled in the same time slot if they are adjacent to each other, i.e., if {i, j} _{k, l} _= φ. This constraint is due to the half duplex transceiver on each node which prevents it from simultaneous transmission and reception.

- **Interfering Constraint:**

The interfering constraint depends on the choice of the interference model. In the protocol model, two edges (i, j) ET and (k, l) ET cannot be scheduled simultaneously if they are at two hop distance of each other. In the physical model, an edge (i, j) ET cannot be scheduled if the S I N R at receiver j is not greater than the threshold β.

Since we consider data collection to be periodic in aggregated converge cast, each of the edges in ET is scheduled only once within each frame, and this schedule is repeated over multiple frames. Thus, a pipeline is established after a certain frame, and then onwards the sink continues to receive aggregated packets from all the source nodes once per frame.

In this system explain further details about the pipelining in the next section. On the other hand, in one-shot data collection for raw-data converge cast, the edges in ET may be scheduled multiple times and no pipelining takes place. Here use the terms link scheduling and node scheduling interchangeably as they are equivalent in our case. Note that the two other scenarios, which we do not consider in this paper due to space constraints, are one-shot aggregated converge cast and periodic raw-data converge cast. The key difference in terms of scheduling between periodic and one-shot data collection is that a node in the periodic case does not have to wait for data from its children before being scheduled. This is because a link is scheduled only once within each frame and each node generates a packet in the beginning of every frame, so a pipelining is eventually established. However, in the case of one-shot data collection, a node needs to wait for data from its children before being scheduled, which we refer to as the causality constraint.
To summarize the steps in this design, we start with tree construction and then continue with interference aware scheduling.

If the nodes can control their transmission power, scheduling phase is coupled with a transmission power control algorithm. If the nodes can change their operating frequency, channel scheduling can be coupled with time slot scheduling as it is the case with the JFTSS algorithm (Section 5.2.1) or first channels are assigned and then time slot scheduling continues as in the case of RBCA. However, the TMCP algorithm considers tree construction and channel assignment jointly and then does the scheduling of time slots.

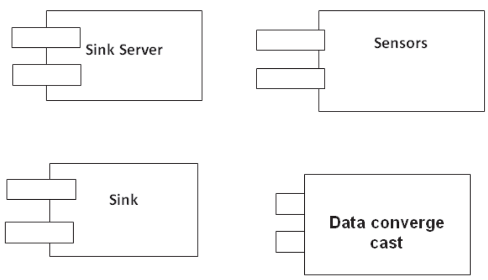
3. SYSTEM DESIGN:

a) Data Flow Diagram:

A data flow diagram (DFD) is a graphical representation of the "flow" of data through an information system, modeling its process aspects. Often they are a preliminary step used to create an overview of the system which can later be elaborated. DFDs can also be used for the visualization of data processing (structured design).

A DFD shows what kinds of information will be input to and output from the system, where the data will come from and go to, and where the data will be stored. It does not show information about the timing of processes, or information about whether processes will operate in sequence or in parallel (which is shown on a flowchart).

The DFD is also called as bubble chart. It is a simple graphical formalism that can be used to represent a system in terms of the input data to the system, various processing carried out on these data, and the output data is generated by the system.

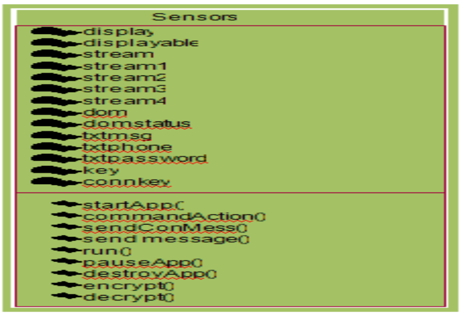


Component Diagram

ISSN No: 2348-4845
International Journal & Magazine of Engineering,
Technology, Management and Research
A Peer Reviewed Open Access International Journal

b) Class Diagram:

Class-based Modeling, or more commonly class-orientation, refers to the style of object-oriented programming in which inheritance is achieved by defining classes of objects; as opposed to the objects themselves (compare Prototype-based programming).



4. Module Description:

a) Periodic Aggregated Converge cast:

Data aggregation is a commonly used technique in WSN that can eliminate redundancy and minimize the number of transmissions, thus saving energy and improving network lifetime. Aggregation can be performed in many ways, such as by suppressing duplicate messages; using data compression and packet merging techniques; or taking advantage of the correlation in the sensor readings.

b) Transmission Power Control:

We evaluate the impact of transmission power control, multiple channels, and routing trees on the scheduling performance for both aggregated and raw-data converge cast.. Although the techniques of transmission power control and multi-channel scheduling have been well studied for eliminating interference in general wireless networks, their performances for bounding the completion of data collection in WSNs have not been explored in detail in the previous studies.

The fundamental novelty of our approach lies in the extensive exploration of the efficiency of transmission power control and multichannel communication on achieving fast converge cast operations in WSNs.

c) Aggregated Data Collection:

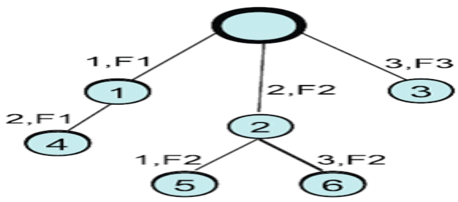
We augment their scheme with a new set of rules and grow the tree hop by hop outwards from the sink. We assume that the nodes know their minimum-hop counts to sink.

d) Raw Data Collection:

The data collection rate often no longer remains limited by interference but by the topology of the network. Thus, in the final step, we construct network topologies with specific properties that help in further enhancing the rate. Our primary conclusion is that, combining these different techniques can provide an order of magnitude improvement for aggregated converge cast, and a factor of two improvement for raw-data converge cast, compared to single-channel TDMA scheduling on minimum-hop routing trees.

e)Tree-Based Multi-Channel Protocol (TMCP):

TMCP is a greedy, tree-based, multi-channel protocol for data collection applications.



Schedule generated with TMCP

It partitions the network into multiple sub trees and minimizes the intra tree interference by assigning different channels to the nodes residing on different branches starting from the top to the bottom of the tree. Figure shows the same tree given in Fig. which is scheduled according to TMCP for aggregated data collection.

Here, the nodes on the leftmost branch is assigned frequency F1, second branch is assigned frequency F2 and the last branch is assigned frequency F3 and after the channel assignments, time slots are assigned to the nodes with the BFSTime Slot Assignment algorithm.

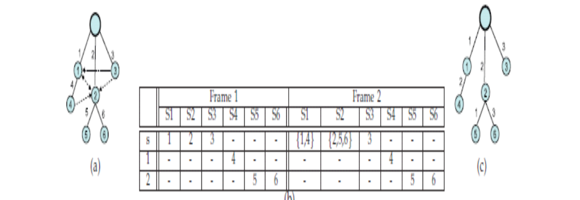
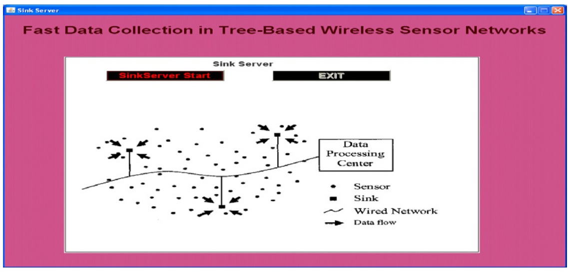
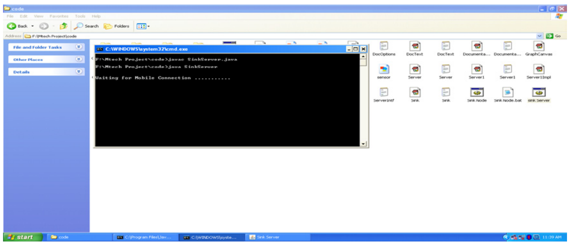


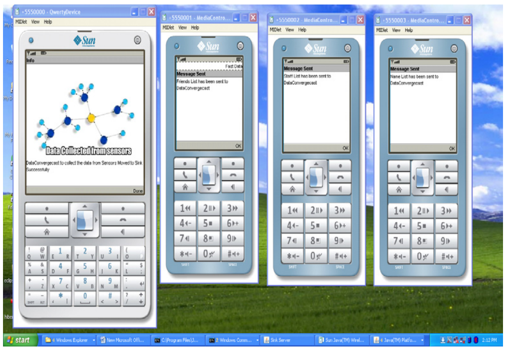
Fig. Aggregated convergecast and pipelining: (a) Schedule length of 6 in the presence of interfering links. (b) Node ids from which (aggregated) packets are received by their corresponding parents in each time slot over different frames. (c) Schedule length of 3 using BFS-TIMESLOTASSIGNMENT when all the interfering links are eliminated.



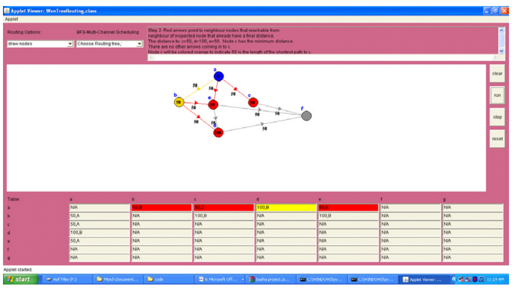
Home page



Mobile connection



Data Collected from Sensors



Data Converge Cast

CONCLUSION:

In this paper, we studied fast converge cast in WSN where nodes communicate using a TDMA protocol to minimize the schedule length. We 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, multiple channels are more effective. We 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).

Once interference is completely eliminated, we proved that with half-duplex radios the achievable schedule length is lower-bounded by the maximum degree in the routing tree for aggregated converge cast, and by max (2nk – 1, N) for raw-data converge cast. Using optimal converge cast scheduling algorithms, we showed that the lower bounds are achievable once a suitable routing scheme is used. Through extensive simulations, we demonstrated up to an order of magnitude reduction in the schedule length for aggregated, and a 50% reduction for raw-data converge cast. In future, we will explore scenarios with variable amounts of data and implement and evaluate the combination of the schemes considered.

REFERENCES:

[1] S. Gandham, Y. Zhang, and Q. Huang, “Distributed time optimal scheduling for converge cast in wireless sensor networks,” Computer Networks, vol. 52, no. 3, pp. 610–629, 2008.

[2] K. K. Chintalapudi and L. Venkatraman, “On the design of Mac protocols for low-latency hard real-time discrete control applications over 802.15.4 hardware,” in IPSN ’08, pp. 356–367.

[3] I. Talzi, A. Hasler, S. Gruber, and C. Tschudin, “Permasense: investigating permafrost with a wsn in the swiss alps,” in EmNets ’07, Cork, Ireland, pp. 8–12.

[4] S. Upadhyayula and S. Gupta, "Spanning tree based algorithms for low latency and energy efficient data aggregation enhanced converge cast (dac) in wireless sensor networks," Ad Hoc Networks, vol. 5, no. 5, pp. 626–648, 2007.

[5] T. Moscibroda, "The worst-case capacity of wireless sensor networks," in IPSN '07, Cambridge, MA, USA, pp. 1–10.

[6] T. ElBatt and A. Ephremides, "Joint scheduling and power control for wireless ad-hoc networks," in INFO-COM '02, Jun, pp. 976–984.

[7] O. D. Incel and B. Krishnamachari, "Enhancing the data collection rate of tree-based aggregation in wireless sensor networks," in SECON '08, San Francisco, CA, USA, pp. 569–577.