

## Analysis of Fast Data Collection in Tree-Based Wireless Sensor Networks

CH.Harsha Manjari <sup>#1</sup>, CH.Satyananda Reddy <sup>\*2</sup>

Department of Computer Science & Systems Engineering,  
Andhra University College Of Engineering,  
Visakhapatnam,AP,India.

### Abstract

Fast and energy efficient data collection in an energy constraint wireless sensor network is always a challenging issue. Using different techniques under practical simulation models for the many-to-one communication example known as converge cast. We first consider time scheduling on a single frequency channel with the aim of minimizing the number of time slots required to complete a converge cast. Then, we combine scheduling with transmission power control to improve the effects of intrusion, and show that power control is reducing the schedule length under a single frequency, scheduling transmission using multiple frequencies is more competent. The energy efficient data collection schedule length used for a propose algorithms that achieve these bounds on the network. We also evaluate the performance of various channel assignment methods and find most incredible that for moderate size networks nodes, the use of multifrequency scheduling can suffice to eliminate most of the interference. For the study we have used some realistic simulation models under many-to-one communication paradigm called convergecast, a single frequency channel and TDMA technique to have minimum time slots for convergecasting.

**Keywords:** Convergecast, TDMA scheduling, multiple channels, power-control, routing trees, Topology.

### 1. Introduction

CONVERGECAST, 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) [1]. 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 [2], 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 [3] require periodic and fast data delivery over long periods of time, which falls under the category of continuous data collection.

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 [4] for any kind of leakage or failures. Whereas applications like weather forecasting, 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 [6] are – **aggregated-data convergecast** where packets are aggregated at each hop, and **raw-data convergecast** where each data packet travel towards sink node individually. First approach is most suitable where data is highly correlated 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) [5] 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, and analyzed the performance of convergecast by using multiple frequencies as compared to those trees using a single frequency.

## 2. Background Theory

Gandham et al. [5] proposed a distributed time slot assignment scheme, for a single channel in TDMA schedule length. Fast data collection with minimum schedule length for aggregated convergecast is discussed by Chen et al. in [6]. Annamalai et al. [7] uses the concept of orthogonal codes to remove interferences, where each node has been assigned time slots from bottom to the top of the tree such that a parent has to wait till it receives all the data packet from its children. Pan and Tseng [8] described a beacon period, assigned to every sensor node in Zigbee network, scheme to reduce latency. A node can receive data only in the allotted beacon period. Song et al.[9] described a time optimal energy efficient packet scheduling algorithm for raw-data convergecast with periodic traffic. They assumed a simple interference model in which every node has a circular transmission range and interferences from concurrent multiple senders is neglected. Song et al [9] further extended their work and proposed a TDMA- based MAC protocol for high-data-rate [10]. Choi et al. [11] shows that for a single channel the minimum schedule length for raw-data convergecast is NP complete on general graphs. Lai et al. [12] uses a greedy graph coloring approach to find the shortest path to the senders for throughput optimization. They also focused on impact of routing trees on schedule length and devised a new approach called disjoint strips to transmit data over different shortest paths. The use

of multiple frequencies is widely described in [13] and [14].

## 3. System Modeling and Discussion

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,  $T = (V, E_T)$  is a spanning tree on  $G$  where  $E_T \subseteq E$  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  $E_T$  of  $T$  using a minimum number of time slots with two constraints:

### Adjacency constraint:

It states that two edges in  $E_T$  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 convergecast each edge in  $E_T$  is scheduled in a pipeline manner. The sink receives packets from the pipeline one after another. On the other hand, for raw data convergecast the edge in  $E_T$  is scheduled multiple times hence no pipeline is there.

### A. Raw-Data Convergecast

In it data of each sensor is equally important, therefore aggregation is not desirable. Each packet is individually scheduled to reach sink node. The problem of minimizing the scheduling length for raw-data convergecast is proved to be NP-complete. Fig. 1, shows a basic tree structure where  $\{s, 1, 4\}$ ,  $\{s, 2, 5\}$ ,  $\{s, 3, 6\}$  are branches of tree and  $\{1, 4\}$ ,  $\{2, 5\}$ ,  $\{3, 6\}$  are sub-trees. We can deduce a local time slot allotment algorithm for each node with an objective to schedule parallel transmissions and allow sink to collect data packets continuously. We assume that sink knows the number of available nodes in each top sub-trees. Each node maintains buffer and state of full or empty if it has data packet available or not. The algorithm for raw data convergecast slot allotment is shown in Algorithm 1.

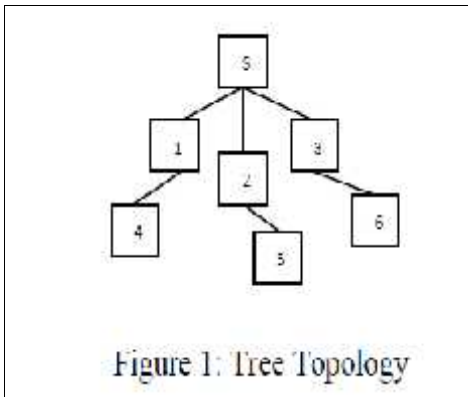


Figure 1: Tree Topology

Algorithm 1: Local Time Slot Scheduling

```

1. Initialize link[buffer] = FULL;
2. Pick a node (N);
3. If (N = Sink) then
   Among available sub tree, select one with largest
   number of remaining packets (s)
4. Plan a link (root(s), S)
5. Else If (N != Sink and node[buffer] = EMPTY)
   then
6. Select a child (C) at random whose buffer is full
7. Plan a link (C, node)
8. C[buffer] = EMPTY;
9. End If
10. End If

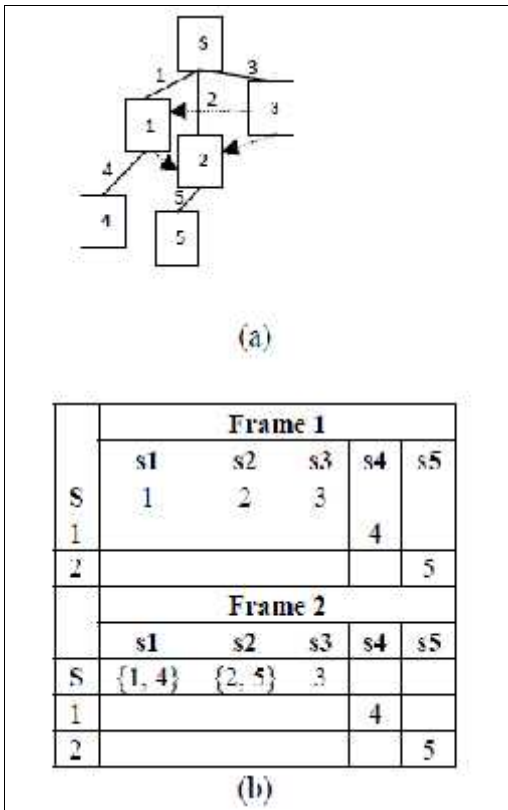
```

In Algorithm 1 lines 3-4 gives scheduling rules between sink and root of sub trees. A top subtree is eligible to be elected for transmission if it has at least one packet for transmission. If none of the top- subtrees are eligible, the sink does not receive any packet during that time slot. Inside each top-subtree, nodes are scheduled according to the rules in lines 5-8. We define a subtree to be active if there are still packets left in the subtree to be relayed. If a node's buffer is empty and the subtree rooted at this node is active, we schedule one of its children at random whose buffer is not empty.

### B. Aggregated Data Convergecast

For continuous monitoring applications data aggregation technique is most suitable. It helps in removing data redundancy and minimizes count of transmission, thus saves energy [15]. Aggregation can be achieved by different techniques such as by data compression, suppressing duplicate messages or by packet merging technology etc. The size of aggregated data transmitted is same and does not depend on to the size of data on individual sensor. The examples of such aggregation functions are MIN, MAX, MEDIAN, AVERAGE, etc.

In Fig. 2(a) and 2(b) aggregated converge cast pipeline is shown for five nodes. The solid lines represent tree edges, and the dotted lines represent interfering links. The numbers beside the links represent the time slots at which the links are scheduled to transmit. The numbers inside the circles denote node ids. The table shows the list of senders and receiver in each time slot. Here at the end of *Frame 1* the sink has no data packets from node 5, as the schedule is repeated, it receives aggregated packets from 2 and 5 in next frame slot. The entries  $\{1, 4\}$  and  $\{2, 5\}$  in the table shows single packets comprising aggregated data from nodes 1 and 4, and from nodes 2 and 5 respectively. Therefore after *Frame 2* a pipeline is generated and sink receives aggregated packets from all the nodes. Now a time slot allotment algorithm can be generated and is shown in Algorithm 2.



Algorithm 2. Aggregation Tree Algorithm

- 1 Start
- 2 Let  $T = (V, E_T)$
- 3 while  $E_T$  is not EMPTY do
  - Select edge (e) from  $E_T$  using Breadth First Search Manner
- 4 Allocate minimum time slot  $t$  to the selected edge e
- 5 Move to next edge in  $E_T$
- 6 End

4. Multi Channel Scheduling for Interference Elimination

Multichannel communication allows concurrent transmissions by using different

frequencies [16], hence eliminate interference. Although typical WSN radios operate on a limited bandwidth, their operating frequencies can be adjusted. It enables multiple concurrent transmissions and more data delivery. By assuming fixed bandwidth channels, we explain two channel assignment methods and study their pros and cons for both type of convergecast. These methods are link level (JFTSS) and cluster level (TMCP). Joint Frequency Time Slot Scheduling (JFTSS) enables a greedy joint solution for maximal time schedule. A maximal schedule is that which meets the adjacency and interfering constraints, and no further links can be scheduled for concurrent transmissions on any time slot. A comparative study of single channel and multichannel system is discussed in [17]. JFTSS scheduling in a network starts from the link having highest number of packets for transmission. If the link loads are equal, the most constrained link is opted first. Initially algorithm has an empty schedule and links are sorted as per loads. The links having adjacency constraint with scheduled link are excluded from the list of link to be scheduled in a given time slot. Only the link having non interfering constraint with scheduled link can be scheduled in the same slot and those having interfering constraint can be scheduled on different channels. If no more links are possible to be scheduled for a given slot, the scheduler continues with scheduling in the next slot.

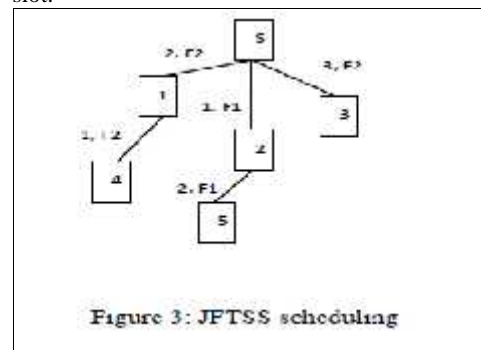


Fig. 3 shows the same tree in Fig. 2(a) which is scheduled according to JFTSS to collect aggregated data. JFTSS starts with link (2, s) on frequency 1 (F1) and then schedules link (4, 1) on the first slot on frequency 2 (F2). Then, links (5, 2) on frequency 1 (F1) and (1, s) on frequency 2 (F2) are scheduled on the second slot and (3, s) on

frequency 2 (F2) are scheduled on the last slot. An advantage of JFTSS is that it is easy to incorporate the physical interference model; however, it is hard to have a distributed solution since the interference relationship between all the links must be known.

Tree Multi-Channel Protocol (TMCP) [13] is a greedy tree-based multichannel protocol. It divides a complete tree into number of sub-trees and reduces intra tree interference by using different channels to the nodes.

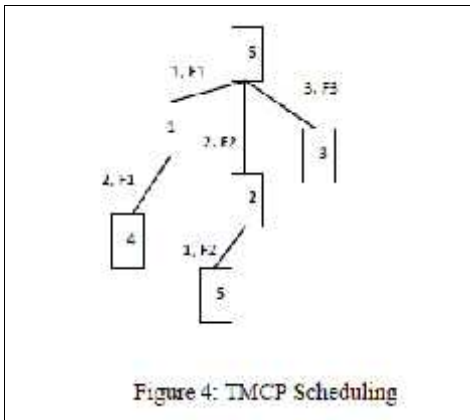


Figure 4: TMCP Scheduling

Fig. 4 shows the same tree given in Fig. 2a, scheduled according to TMCP. Here, the nodes on the leftmost branch are 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 algorithm 2. The advantage of TMCP is that it is designed to support convergecast traffic and does not require channel switching. Since all the nodes communicate on same channel, the contention inside branches is not resolved.

### 5. Impact of Network Topology

Besides multiple channels, the network topology and the degree of connectivity makes impact on scheduling performance. As described in [8], network trees that have more parallel transmissions do not necessarily result in small schedule lengths. As an example in star topology network with N nodes the schedule length is N, whereas it is (2N-1) for a bus topology once

interference is eliminated. In this section we construct a spanning tree with constraint  $n_k < (N+1)/2$ , where  $n_k$  are number of branches and N is the number of nodes. A balanced tree satisfying this constraint is a variant of a capacitated minimal spanning tree (CMST) [15]. The CMST algorithm can determine a minimum-hop spanning tree in a vertex weighted graph, such that the weight of every subtree linked to the root does not exceed a prescribed capacity Here we have taken weight of each link as 1, and prescribed capacity is (N+1)/2. Here, we propose a method, described in Algorithm 3, based on greedy scheme presented by Dai and Han to solves a variant of the CMST problem. By using it, searches for routing trees with an equal number of nodes on each branch. It is assumed that every node know their minimum-hop counts to sink node.

The rules associated to the algorithm are:

- Rule 1.** Nodes having single parents are connected first.
- Rule 2.** Node with multiple parents, a Reservoir Set (RS) is created and select one from it.
- Rule 3.** After selecting a node from RS a search set S is constructed to decide which particular branch the node should be added to. S therefore consist of nodes that are not yet connected but are neighbors of a node with high hop-count.

---

Algorithm 3: CMST Tree Creation

```

1. Given G(V, E) with sink node S
2. Let P is roots of top subtrees and T={s} U P, k=2,
   RS(i)=unconnected neighbors of i, S(i)=0,
3. while k != Maximum_hop_count do
   Nk = all unconnected nodes at hop distance k;
   Connect node Nk having single parent T=T U Nk;
4. Update Nk = Nk \ Nk;
5. Sort Nk;
6. for all i in Nk do
   for all j in P to which i can connect do
       Link (i, j);
   End for
   T=T U {i} U Link(i, j);
   Update RS(i);
End for
7. k=k+1;
8. End while
    
```

---

## 6. Conclusion

In this paper, we have discussed fast convergecast 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. Through extensive simulations, we demonstrated up to certain extent reduction in schedule length for aggregated data converge casting and approximately 50 percent reduction for raw-data convergecast. As a future work we will explore aspects related to variable amount of data and evaluate the various schemes considered.

## 7. References

- [1] S. Gandham, Y. Zhang, and Q. Huang, "Distributed time-optimal scheduling for convergecast in wireless sensor networks," *Computer Networks*, vol. 52, no. 3.
- [2] I. Talzi, A. Hasler, S. Gruber, and C. Tschudin, "Permasens: investigating permafrost with a wsn in the swiss Alps," in *EmNets '07*, Cork, Ireland.
- [3] S. Upadhyayula and S. Gupta, "Spanning tree based algorithms for low latency and energy efficient data aggregation enhanced convergecast (dac) in wireless sensor networks," *Ad Hoc Networks*, vol. 5.
- [4] 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," *Proceeding of International Conference on Information Processing in Sensor Networks (IPSN '08)*, pp. 356-367, 2008.
- [5] S. Gandham, Y. Zhang, and Q. Huang, "Distributed Time-Optimal Scheduling for Convergecast in Wireless Sensor Networks," *Computer Networks*, vol. 52, no. 3, pp. 610-629, 2008.
- [6] X. Chen, X. Hu, and J. Zhu, "Minimum Data Aggregation Time Problem in Wireless Sensor Networks," *Proc. Int'l Conf. Mobile Ad-Hoc and Sensor Networks (MSN '05)*, pp. 133-142, 2005.
- [7] V. Annamalai, S.K.S. Gupta, and L. Schwiebert, "On Tree-Based Convergecasting in Wireless Sensor Networks," *Proceeding IEEE Wireless Communication and Networking Conf. (WCNC '03)*, vol. 3, pp. 1942-1947, 2003.
- [8] M. Pan and Y. Tseng, "Quick Convergecast in ZigBee Beacon-Enabled Tree-Based Wireless Sensor Networks," *Computer Comm.*, vol. 31, no. 5, pp. 999-1011, 2008.
- [9] W. Song, F. Yuan and R. LaHusen, "Time-Optimum Packet Scheduling for Many-to-One Routing in Wireless Sensor Networks," *Proc. IEEE Int'l Conf. Mobile Ad-Hoc and Sensor Systems (MASS '06)*, pp. 81-90, 2006.
- [10] W. Song, H. Renjie, B. Shirazi, and R. LaHusen, "TreeMAC: Localized TDMA MAC Protocol for Real Time High Data Rate Sensor Networks," *Journal of Pervasive and Mobile Computing*, vol. 5, no. 6, pp. 750-765, 2009.
- [11] H. Choi, J. Wang and E. Hughes, "Scheduling for Information Gathering on Sensor Network," *Wireless Networks*, vol. 15, pp. 127-140, 2009.
- [12] N. Lai, C. King, and C. Lin, "On Maximizing the Throughput of Convergecast in Wireless Sensor Networks," *Proc. Int'l Conf. Advances in Grid and Pervasive Computing (GPC '08)*, pp. 396-408, 2008.
- [13] Y. Wu, J.A. Stankovic, T. He, and S. Lin, "Realistic and Efficient Multi Channel Communication in Wireless Sensor Networks," *Proc. IEEE INFOCOM*, pp. 1193-1201, 2008.
- [14] Y. Kim, H. Shin, and H. Cha, "Y-MAC: An Energy-Efficient Multi Channel MAC Protocol for Dense Wireless Sensor Networks," *Proc. Int'l Conf. Information Processing in Sensor Networks (IPSN '08)*, pp. 53-63, Apr. 2008.
- [15] B. Krishnamachari, D. Estrin, and S.B. Wicker, "The Impact of Data Aggregation in Wireless Sensor Networks," *Proc. Int'l Conf. Distributed Computing Systems Workshops (ICDCSW '02)*, pp. 575-578, 2002.
- [16] P. Kyasanur and N.H. Vaidya, "Capacity of Multi-Channel Wireless Networks: Impact of Number of Channels and Interfaces," *Proc. ACM MobiCom*, pp. 43-57, 2005.

[17]G. Sharma, R.R. Mazumdar and N.B. Shroff, "On the Complexity of Scheduling in Wireless Networks," Proc. ACM MobiCom, pp. 227-238, 2006.

## 8. About the Authors



CH. Harsha Manjari is currently pursuing her 5 Years Integrated M.Tech (IT) in Computer Science and Systems Engineering at Andhra University College of Engineering, Visakhapatnam.

Her area of interests includes Mobile Computing.



CH. Satyananda Reddy is currently working as a Sr.Assistant Professor in Computer Science and Systems Engineering department, Andhra University College of Engineering, Visakhapatnam. He is registered for Ph.D.

(CSSE) in Andhra University. His research interests include Software Engineering, DBMS, and Operating Systems.