Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 2, Issue. 4, April 2013, pg.559 – 567

RESEARCH ARTICLE

Optimized Rapid Data Collection in Tree Based WSN

DEVASAHAYAM[1], K.P. KALIYAMURTHIE[2]
[1]Department of Information Technology, Bharath University, India
[2]Department of Information Technology, Bharath University, India

Abstract— Fast and energy efficient data collection in an energy constraint ad-hoc sensor network is always a challenging issue. The network topology and interferences causes significant effects on data collection and hence on sensors' energy usage. Various approaches using single channel, multichannel and convergecasting had already been proposed. Here in this paper we have shown data collection performance using multi-frequency in channel assignment, and effect of network topology, for moderate size networks of about 50-100 nodes. 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.

Key Terms: - Convergecast; Multi-channel; Topology; Energy; TDMA

I. 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 [1] 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 [3] 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 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) [2] 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.

The rest of the paper is organized as follows: in Section II, we discuss related works. In Section III, we describe system modeling and some discussions. In Section IV, we have shown multichannel scheduling for interference elimination. In Section V, we focus on impact of network topology on data forwarding. Section VI gives the evaluation work based on previous discussions. Finally Section VII concludes the paper.

II. RELATED WORK

Gandham et al. [2] 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 [3]. Annamalai et al. [4] 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 [5] 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.[6] 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. [6] further extended their work and proposed a TDMA- based MAC protocol for high-data-rate [7]. Choi et al. [8] shows that for a single channel the minimum schedule length for raw-data convergecast is NP complete on general graphs. Lai et al. [9] 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 [10] and [11].

III. SYSTEM MODELLING 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) Є V is the set of wireless links. Let *s* is the sink node such that *s* Є 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 the 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.

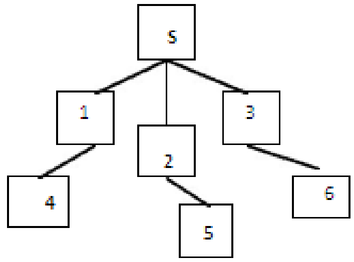


Figure 1: Tree Topology

Algorithm 1: Local Time Slot Scheduling

---

1. Initialize node[buffer]=FULL
2. Pick a node (N)
3. If (N = Sink) then

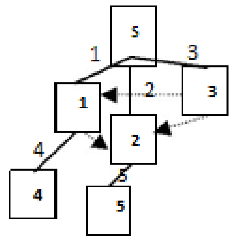
   Among available sub-tree, select one with largest number of remaining packets (i).

4. Plan a link (root(i), 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- sub trees are eligible, the sink does not receive any packet during that time slot. Inside each top-sub tree, nodes are scheduled according to the rules in lines 5-8. We define a sub tree to be active if there are still packets left in the sub tree to be relayed. If a node's buffer is empty and the sub tree 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 [12]. 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.



(a)

	Frame 1				
	s1	s2	s3	s4	s5
S	1	2	3		
1				4	
2					5
	Frame 2				
	s1	s2	s3	s4	s5
S	{1, 4}	{2, 5}	3		
1				4	
2					5

(b)

Figure 2: (a), (b) Aggregated Convergecast Pipeline

In Fig. 2(a) and 2(b) aggregated convergecast 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

IV. MULTICHANNEL SCHEDULING FOR INTERFERENCE ELIMINATION

Multichannel communication allows concurrent transmissions by using different frequencies [13], 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 [14]. 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.

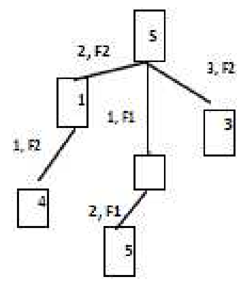


Figure 3: JFTSS scheduling

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) [10] 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.

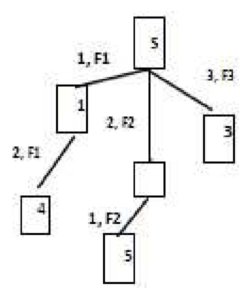


Figure 4: TMCP Scheduling

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.

V. 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 [16] 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.

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

N_h = all unconnected nodes at hop distance *h;* Connect node N'_h having single parent: T=T U N'_h ;

4. Update $N_h = N_h \setminus N'_h$;
5. Sort N_h;
6. for all i in N_h 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

The rules associated to the algorithm are:

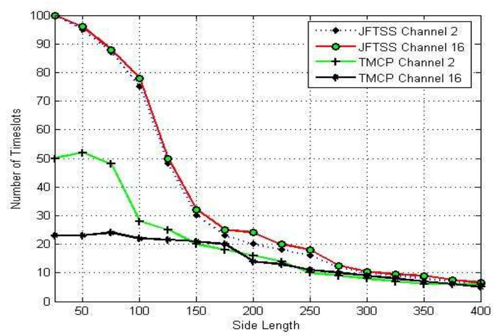
Rule 1. Nodes having single parents are connected first.

Rule 2. Node with multiple parents, a Reservoir Set (RS) is created and selects one from it.

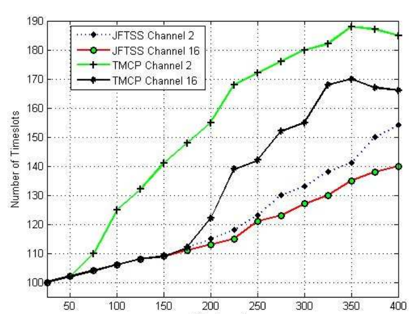
Rule 3. After selecting a node from RS a search set S is constructed to decide which particular branch the node should be added to. S therefore consists of nodes that are not yet connected but are neighbors of a node with high hop-count.

VI. PERFORMANCE EVALUATION

In this section, we evaluate performance of multiple channels and network tree topology on scheduling for both aggregated and raw-data convergecast. We deploy sensor nodes randomly in a region with dimensions varying between 30x30 m^2 and 400x400 m^2 to have different network density. The number of nodes is fixed to 100 and for different parameters; we average each point over 1500 runs. An exponential path-loss model for signal propagation along with path-loss exponent varying between 3 and 4 is used. We have simulated the behavior of CC2420 radios used on TmoteSky motes and are able to operate on 16 different frequencies. The transmission power can be adjusted between -24 and 0 dBm over eight different levels and the SINR threshold is set to β= -3dB. Firstly, the schedule length of single-channel TDMA is determined, secondly its improvement using multiple channels and routing trees is evaluated. All the nodes transmit at maximum power and uses minimum hop tree. In TMCP time slots are assigned according to Algorithm 1 for raw data convergecast and Algorithm 2 for aggregated convergecast. The path loss exponent is 3.5. The results are shown in Fig. 5(a) and 5(b). It is evident from Fig. 5(a) that with just two frequencies interference limitations are eliminated and the performance gains are limited by the connectivity structure. With multichannel communication a 40 percent reduction in schedule length is observed as compared to transmitting on a single channel with maximum power. Further, JFTSS can optimally schedule the network using 16 channels as shown in graph of fig.5. In dense deployments, TMCP performs better due to construction of different routing trees i.e., when L = 20, JFTSS construct a star topology, whereas TMCP constructs a 2-branch tree with two channels and a 16-branch tree with 16 channels.



(a)



(b)

Figure 5: Multiple Channel Scheduling: (a) Aggregated convergecast
(b) Raw-data convergecast

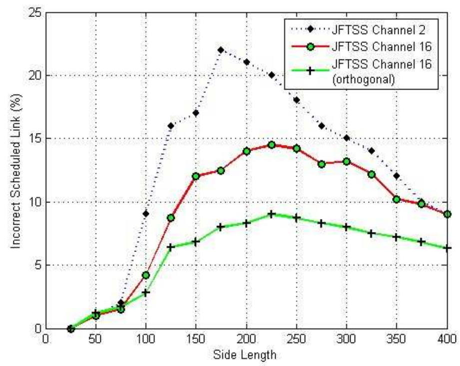


Figure 6: JFTSS performance when node are not correctly scheduled

From fig. 5(b) we observe that none of the methods can eliminate interference completely with two channels and TMCP needs 16 channels to reach a performance similar to that achieved by JFTSS with only two channels. This is because in JFTSS when a node is receiving from its children, its parent can transmit simultaneously on a different channel, which is not possible due to intra branch interference in TMCP. The results also verify that JFTSS can achieve a schedule length which is bounded by $Max(2n_k\text{-}1,N)$. The impact of interference and incorrect slot assignment is shown in Fig. 6. It shows JFTSS performance where nodes are not correctly scheduled. The top two lines show the errors for two and 16 channels with both the assumptions, whereas the bottom line shows the errors only for the orthogonality assumption. The errors are higher in sparser deployments because an individual sender is not capable enough to jam concurrent transmissions. While in dense deployments an individual can create jam because of shorter inter node distances. A short distance makes high availability of channels for concurrent transmissions, moreover channels are not orthogonal.

VII. 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 convergecasting 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.

REFERENCES

[1] [1] 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.

[2] 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.

[3] 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.

[4] V. Annamalai, S.K.S. Gupta, and L. Schwiebert, “On Tree-Based Convergecasting in Wireless Sensor Networks,” Proceeding IEEE WirelessCommunication and Networking Conf.(WCNC’03),vol. 3, pp. 1942-1947, 2003.

[5] 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.

[6] 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.

[7] 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 Cosmputing, vol. 5, no. 6, pp.

750-765, 2009.

[8] H. Choi, J. Wang and E. Hughes, "Scheduling for Information Gathering on Sensor Network," Wireless Networks, vol. 15, pp. 127-140, 2009.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

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

[15] C.H. Papadimitriou, "The Complexity of the Capacitated Tree Problem," Networks, vol. 8, no. 3, pp. 217-230, 1978.

[16] H. Dai and R. Han, "A Node-Centric Load Balancing Algorithm for Wireless Sensor Networks," Proc. IEEE Conf. Global Telecomm. (GlobeCom '03), pp. 548-552, 2003.

