Bessy M Kuriakose, M. Princy / International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 www.ijera.com
Vol. 3, Issue 2, March -April 2013, pp.1784-1789

Efficient Routing of Correlated Data in Energy Constrained Wireless Sensor Networks

Bessy M Kuriakose*, M. Princy**
*(PG Scholar, Department of Information Technology, Karunya University, Coimbatore-14)
** (Lecturer, Department of Information Technology, Karunya University, Coimbatore-14)

ABSTRACT

A routing protocol in a wireless sensor network is usually used to find a route to the destination so that the nodes can report the event to the sink in an energy efficient manner. The battery limitations of the sensor nodes and the characteristics of the environment where the nodes are deployed, make the routing problem very challenging. The Sensor data from different nodes in a dense region may also be highly correlated. Such data when routed across a wireless sensor network leads to the redundancy of data at various nodes, thereby consuming a vast amount of energy. This paper discusses the issues faced in a correlation and interference-aware wireless sensor network with a single base station and thereby proposes a technique called the MM-ICAR that provides energy efficient routing for correlated data with multiple intermediate base stations that deliver data to the sink in wireless sensor networks.

Keywords - Wireless Sensor Networks, Correlation, Data Aggregation, Energy Consumption, Base station

I. INTRODUCTION

Wireless sensor networks are large scale networks that consist of small sensor nodes that are densely deployed in an ad hoc fashion. Each sensor node consists of one or more microcontrollers that provide processing capability, multiple types of memory, an RF transceiver, a power source, various sensors and actuators. Sensors generate a data stream by periodically measuring the physical environment around them. This data is sent to the base station which acts as a gateway that relays the data through a wireless network to the users. Wireless Sensor Networks are tremendously adaptable and can be deployed to support a large diversity of applications in many different conditions, whether they are poised of fixed or mobile sensor nodes. Once the sensor nodes are positioned, they self-organize into a self-governing wireless ad hoc network, which requires very little or no maintenance. Sensor nodes then cooperate with each other to carry out the responsibilities of the application for which they are deployed. The main task of sensor nodes is to sense and collect data from an intended environment, process the data, and transmit it back to the base station where the fundamental application resides. Accomplishing this task efficiently requires the advance of an energy-efficient routing protocol to set up paths between sensor nodes and the data sink. The path selection must be in such a way that the network lifetime is maximized. The distinctiveness of the environment within which sensor nodes typically operate, tied with rigorous resource and energy insufficiency, make the routing problem very challenging.

Energy, however, is a key concern in WSNs, which must achieve an extended life span while working on limited battery reserves. Correlated data induces data redundancy in the destination node thereby consuming large amount of energy and decreasing the network lifetime. Interference of data from other nodes also affects the accuracy of data being sent to the base station. This paper discusses the drawbacks of having a single base station to receive the data from all the sensors and proposes a technique of employing multiple numbers of base stations. To the best of our knowledge, such a technique has not been employed in a routing technique that considers energy conservation, interference and the correlation of data among the different nodes. Simulation results show that by this technique energy consumption has been reduced to a considerable level.

II. EXISTING TECHNIQUES

A.Directed Diffusion

Directed Diffusion [1][2] is a data centric protocol. It establishes low latency paths between the source and the destination which would in turn become the aggregation tree with the sink at the root. Opportunistic data aggregation takes place at the nodes whenever similar data happens to arrive. It consists of several elements such as interests, data messages, gradients, and reinforcements. Initially, sink node requests for data by sending interest messages. An interest message is a query message or an exploratory message, which specifies the needs of the user to its neighbors for a named data. The data is named using attribute-value pairs. This named data corresponds to the collected or processed information of an event that matches the interest of a user. The interests are broadcasted over the entire network by the sink node. Whenever a

node receives an interest, it will check whether it already exists or is a new one. If it is a new interest message, the sensor node will set up a gradient toward the sink node which issued the interest message to receive the data that matches the interest. Every node establishes a gradient to its neighboring node. After this stage of establishing gradients, the source node then begins to send the related data that matches the interest to the sink. The data are generally broadcasted to all its gradient neighbors.

Directed Diffusion is a technique that establishes paths using path reinforcement where the node has the ability to decide whether to accept data from its neighboring nodes. The neighboring nodes then deliver the data with the lowest latency based on the decision. But it is found that this technique is not desirable as data cannot be aggregated where they originate.

B.Greedy Incremental Tree

A shortest path is established for only the first source to the sink whereas each of the other sources is incrementally connected at the closest point on the existing tree in order to create a Greedy incremental Tree[3]. In this approach, each interest message contains an *energy cost*. In addition to this, each source on the established path also generates an *incremental cost message* which corresponds to each new interest message received. The incremental cost message contains the incremental energy cost required for delivering the corresponding interest sample to the existing tree which is only transmitted and updated along the aggregation tree toward the sink. The incremental energy-cost field can be updated only by closer nodes in order to determine the closest point.

In this approach, the most preferred neighbor to reinforce is a neighbor which has delivered the interest message sample or its corresponding incremental cost message at the lowest energy cost. Intermediate nodes either process or keep the received data for a period of time before aggregating multiple messages into a single aggregate. The nodes compute associated energy for the aggregate which can be used for path pruning. The challenge is to find the set of incoming aggregates which cover the data items at the smallest cost. Due to dynamic nature of the network, there are chances for multiple paths to be reinforced. The unnecessary paths are pruned or negatively reinforced. This technique achieves significant energy reduction in a high density network.

C. Low Energy Adaptive Clustering Hierarchy

Low Energy Adaptive Clustering Hierarchy (LEACH) is a hierarchical clustering algorithm proposed in [4][5] for wireless sensor networks. It is a cluster-based protocol which randomly selects a few sensor nodes as cluster heads and rotates this role in order to evenly distribute the energy consumption among the sensors in the network. The cluster head nodes compresses the data arriving from nodes in the cluster and send the aggregated packet to the sink in order to reduce the amount of information that is transmitted to the sink. LEACH uses a TDMA/CDMA MAC to reduce inter-cluster and intra-cluster collisions. Data collection in LEACH is centralized and is periodically performed. After a particular amount of time, a randomized rotation of the role of the Cluster Head is done so that uniform energy consumption by the sensor nodes is achieved.

This routing technique is categorized into two phases, the setup phase and the steady state phase. In the setup phase, the clusters are organized and Cluster Heads are selected. During the setup phase, a predetermined fraction of nodes elect themselves as Cluster Heads. A sensor node chooses a pseudo random number which is between 0 and 1. If this number is less than a threshold value, T(n), the node becomes a cluster-head for the current round. Each elected Cluster Head then advertises itself as the new cluster head to the rest of the nodes in the network. The non-cluster head nodes decide on the cluster to which they want to belong to, after receiving this advertisement and inform the appropriate cluster-heads that they will be a member of the cluster. This decision is based on the signal strength of the advertisement. After receiving the messages from all the nodes that would like to be included in the cluster and based on the number of nodes in the cluster, the cluster-head node creates a TDMA schedule and assigns each node a time slot when it can transmit. This schedule is broadcasted to all the nodes in the cluster. The actual data transfer to the sink takes place in the steady state phase.

During the steady state phase, the sensor nodes can do their job of sensing and transmitting data to the cluster-heads. The cluster-head node, after receiving all the data, aggregates it before sending it to the base-station. After a predetermined amount of time, the network goes back into the setup phase again and enters another round of selecting a new Cluster Head. Each cluster communicates using different CDMA codes to reduce the interference from nodes belonging to other clusters. The duration of the steady state phase is longer than the duration of the setup phase in order to minimize overhead. However in this technique, data aggregation can be performed only at the cluster head node and hence there are chances for data redundancy to occur at the neighboring nodes within the clusters.

D.Power Efficient Gathering Sensor Information Systems

PEGASIS[6] being the extension of the LEACH protocol forms chains of the sensor nodes so that each node can transmit and receive from a neighbor and only one node is selected from that

chain to transmit to the sink. The data is gathered and moves from one node to another where it is aggregated and eventually sent to the sink. The chain is constructed based on the greedy algorithm.

Unlike LEACH, PEGASIS has no cluster formation and it employs only one node to transmit the data to the sink. The fundamental idea of the protocol is that in order to extend the network lifetime, nodes need to communicate only with their nearest neighbors and they take turns in transmitting to the sink node. When a round of all nodes communicating with the base-station is complete, a new round will start. This reduces the energy required to transmit the data per round as the energy dissipation is spread uniformly over all nodes.

PEGASIS has two main objectives: to increase the lifetime of each node, as a result of which the network lifetime will be increased and to allow only local coordination between the nodes that are close together so that the bandwidth consumption is reduced. To trace the nearest neighbor node in PEGASIS, each node uses the signal strength to compute the distance to all neighboring nodes and then regulate the signal strength so that only one node can be heard. The chain in PEGASIS will consist of those nodes that are closest to each other and form a path to the sink node. The aggregated data will be sent to the sink by any node in the chain and the nodes in the chain will take turns in sending to the base-station in order to evenly distribute the energy consumption among all the nodes. [6][7] shows that PEGASIS is able to increase the network lifetime twice as much as that of LEACH protocol. This performance gain is achieved through the elimination of the overhead caused by dynamic cluster formation in LEACH and by decreasing the number of transmissions and receptions by using data aggregation. Even though the clustering overhead is avoided, PEGASIS still requires dynamic topology adjustment as each sensor node needs to know about energy status of its neighbors in order to have knowledge of where to route its data. This technique assumes that all nodes maintain a complete database about the location of all other nodes in the network. It also assumes that all sensor nodes have the same level of energy and they are likely to die at the same time. This technique introduces an excessive delay for distant node on the chain. To reduce the latency of data gathering, multi level chaining is used.

E.Slepian Wolf Coding

Slepian-Wolf coding [8] is a distributed source coding technique. In this technique, all sources are coded with a total rate that is equal to the joint entropy. This is done without explicit communication between each other, as long as their individual rates are at least equal to their respective conditional entropies. Assume that $N1$, $N2\ldots NN$ are the source nodes and $H(N1,N2, \ldots, NN)$ be the joint entropy of the data from these nodes. The samples taken at nodes are spatially correlated. It is assumed that each random variable is taken from a discrete-time random process which is independent and identically distributed over time and has a countable discrete alphabet. If node $N1$ codes its data at a rate according to its unconditional entropy, $H(N1)$, and node $N2$ codes its data at a rate according to its entropy conditioned on $N1$'s, $H(N2N1)$, the sink can obtain the joint entropy $H(N1, N2)$ from $H(N1)$ and $H(N2N1)$. As the transmitted data from $N1$ and $N2$ have completely eliminated redundancy after the coding, shortest paths are optimal routes for $N1$ and $N2$ to deliver the data to the sink. The Slepian Wolf algorithm forms the routing structure based on Shortest Path Tree algorithm from source nodes to the sink. On the Shortest Path Tree, node $X1$ denotes the closest node to the sink and node XN is the furthest.

The algorithm involves the following steps, the first being the Ordering of the total weights on the SPT from the nodes to the sink: each node needs its index in the ordered sequence of nodes in order to determine on which other nodes to condition when computing its rate assignment. For instance, it may happen that the distance on the graph between nodes and is large. Thus, closeness in the ordering on the SPT does not mean necessarily proximity in distance on the graph.

The next step is computation of the rate assignment. For each node, we need to know locally all the distances among the nodes, in order to be able to compute the rate assignment as it involves a conditional entropy including all these nodes. This means that, for a distributed algorithm, global knowledge should be available at nodes, which might not be the case in a practical situation. That is, the closest node to the sink is coded with a rate equal to its unconditioned entropy; each remaining node is coded with a rate equal to its respective entropy conditioned on all nodes closer to the sink than itself.

If Slepian–Wolf coding is used in a network-correlated data gathering scenarios, then the optimization is separated as an optimal transmission structure that needs to be determined and the optimal rate allocation that has to be found for this transmission structure.

F.Minimum Energy Gathering Algorithm

If data aggregation by conditional coding is achievable only when side information is explicitly available, the optimal data gathering problem is NP-complete. The two classes of source coding are further classified with explicit side information: self-coding and foreign coding. In self-coding technique, data are only allowed to be encoded at the source node and only in the presence of side information from at least one other node. In contrast, foreign coding is a technique that allows a

node to encode raw data originating from another node using its own data as it is routed toward the sink via itself. MEGA (Minimum Energy Gathering Algorithm) [9][10] is a foreign coding technique that can be illustrated as follows: Consider a sensor network as a graph $G = (V, E)$. The weight $w(i)$ is defined as the cost of transmitting one bit of data on edge e to the sink node. The raw data packet from node vi is denoted by pi with size si. Likewise, the raw data packet from node vj is denoted by pj. If pi is encoded with side information pj at node vj, the encoded packet is denoted pij, and its equivalent size is sij. The compression rate depends on the data correlation between the nodes vi and vj, denoted by the correlation co-efficient rij which can be calculated as $rij = 1 – sij/si$. Initially, MEGA computes for each node via corresponding encoding node vj. To achieve this, MEGA assumes a complete directed graph $G' = (V, E')$. The weight $w'(e)$ for a directed edge $e = (vi, vj)$ in E' is defined by the expression as in [9]

$$w'(e) = si(s(vi, vj) + s(vj, t)(1 – ij)),$$

where $s(vi, vj)$ denotes the weight of the shortest path from vi to vj in G. The weight of an edge in G' therefore corresponds to the total energy consumption in order to route a data packet pi to the sink using node vj as the encoding relay node. After this, a directed minimum spanning tree is constructed rooted at sink t, where edge (vi, vj) denotes that vj is the best encoding node for vi. The raw data is then delivered on the shortest path from node vi to its encoding relay node vj. After compression, the encoded data is then sent through the shortest path from vj to t which is the destination.

G.Low Energy Gathering Algorithm

Self-coding nodes can only encode their own raw data in the presence of other raw data that is routed through them. An algorithm called Low Energy Gathering Algorithm (LEGA) which is based on a shallow light tree (SLT), is proposed as a source coding scheme in [9][10].

Shallow Light Tree is a spanning tree that approximates both minimum spanning tree (MST) and SPT for a given node most likely the sink node. Initially, the SLT spanning tree is formed with the sink node t as the root. The sink then broadcasts its raw data packet to all of its one-hop neighbors in the Tree. Upon receiving a raw data packet from a neighboring node, node vi encodes its locally measured data using the data of the neighboring node, and transmits the encoded packet *to* the sink t via the path given by the SLT tree. Then node then broadcasts its packet to all its one-hop neighbors except the previous neighboring node. The sink t has its own data available in the vicinity or it can even use the data of one of its first-hop neighbors and hence can also carry out recursive decoding of the gathered data, based on the encoded data it has received from all other nodes in the network. LEGA is established to be an approximation algorithm for a self-coding scheme with an approximation ratio of 2(1 + □ 2) as in [10].

Although both MEGA and LEGA routing techniques can achieve a near-optimal performance under a source coding model with explicit side information, their performance is subjected to high discrepancies in dense networks where the adjacent data has high redundancy. The reason is that source data can only be encoded once, and its data redundancy with other nodes except the one providing side information, cannot be eliminated.

H.Interference and Correlation Aware Routing

Energy efficient routing algorithms are necessary to decide which set of sensor nodes form a route to the sink from a given node so that the energy consumption at each node becomes affordable. In Wireless Sensor Networks, the sensed data from the different sensor nodes in a location may be correlated and transmitting all this information across the network to the destination can increase the traffic and thereby the data redundancy at the destination nodes. This results in inefficient energy consumption and hence reduces the throughput of the entire network. Hence, these routing algorithms must be correlation aware that checks the data at each node using the Maximum Correlated Data Aggregation algorithm which determines the data rate of the transmitted data at each node. The entropy of the data is calculated after ensuring if any side information is available or not. The energy efficiency and the network lifetime may also be impacted by the interference of data that an intermediate node might cause to its neighboring nodes. Hence, the effect of interference must also be considered as per [11] in addition to energy conservation as well as aggregation of correlated data in terms of route selection strategies.

III. MM-ICAR

One of the major issues faced by a wireless sensor network is the energy conservation at each node and although lots of techniques have been proposed by different researchers, it still remains a serious crisis. The different sensor network models considered by most of the researchers have only a single base station that has a constant location and they usually appear as the root of the tree topology. In most of the traditional routing schemes, the sensor nodes in the network gather data from the respective environment and find routes to deliver the gathered data to the sink node. The energy consumed in delivering a message (E) from any sensor node i to the sink node is directly proportional to the number of hops (H) the message has to travel.

$$E_i \propto H_i$$

This implies that wherever the sensor node may be located, it has to deliver the data to the sink be it far or near. In cases where the base station is located far from the source nodes, the number of hops increases and as the number of hops increases, the energy consumption at each node also increases. Hence to effectively reduce the consumption of energy for various routing activities at each node, the idea of employing multiple intermediate moving agents called data mules [12] can be merged with the ICAR technique.

Data mules are intermediate agents which are fitted with transceivers. It is assumed that the mules have sufficient energy which can also be recharged when it comes in contact with the base station. Hence, the energy in mules is renewable. They also have large memory storage capacity in order to collect the data from a maximum of four sensors. When in proximity, MULEs pick up data from the sensors, buffer it, and deliver it to wired access points. This can lead to considerable power savings at the sensors as they only have to transmit over a short-range mostly in a single hop. But in case of a multi hop transmission, the sensor nodes that are one-hop away from its nearest data mule are drained of their energy faster than other nodes in the sensor network. It is obvious that the sensor nodes which are one hop away from data mule need to forward messages originating from many other nodes, in addition to delivering their own messages. In doing so, these sensor nodes drain their energy faster and become inefficient. Hence, it affects the communication of data from many source nodes with the mule agent and the network becomes inactive. To avoid this problem, the data mules are chosen to be moving agents like animals, vehicles, humans, etc… that roam around periodically to different locations.

The working of this technique goes as follows: The entire network lifetime is divided into equal intervals of time called slots. In the beginning of each slot, the location of the mule agent is assumed to be moved to another location so that there is a rotation for the nearest one hop neighbor sensor node and hence, there is no opportunity for the same nodes to drain their energy trying to forward their own data as well as the data from other nodes. During each time unit, it is assumed that each node gathers equal amount of data and the energy dissipated for transmission or the reception of the data by each node is a constant value. The data mules gather data from at most four sensors which are at its proximity. It is required that the sensors must complete their transmission before the agent changes its location. In case of multi hop transmission, the sensor nodes that are quite far from the intermediate agents can transmit their data through intermediate nodes which are in the range of the agents.

The number of intermediate agents used to collect data depends merely on the number of sensors. Hence, the number of intermediate agents used can be calculated using the formula,

n=N/6

Where, N is the Number of sensors in the entire network. The number 6 is used as a threshold value that can divide the network into virtual classifications. The data collected are aggregated at the intermediate agents so that redundant data is not transmitted thereby increasing the throughput of data across the network towards the sink node. Data aggregation is performed at the end of each time slot.

Simulation results have shown that as the sensors have to transmit data only to the agents that come to its proximity, energy can be conserved sufficiently than the ICAR technique which considers energy efficient routing for correlated data in wireless sensor networks. Similarly, throughput has increased considerably when compared to the previous technique.

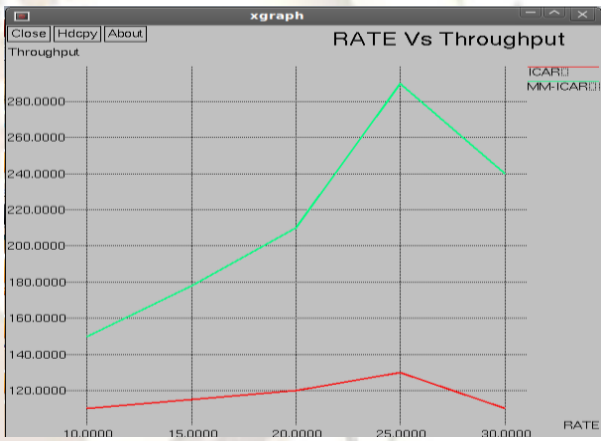

Fig 1: Graph showing the Throughput increase for the technique MM-ICAR

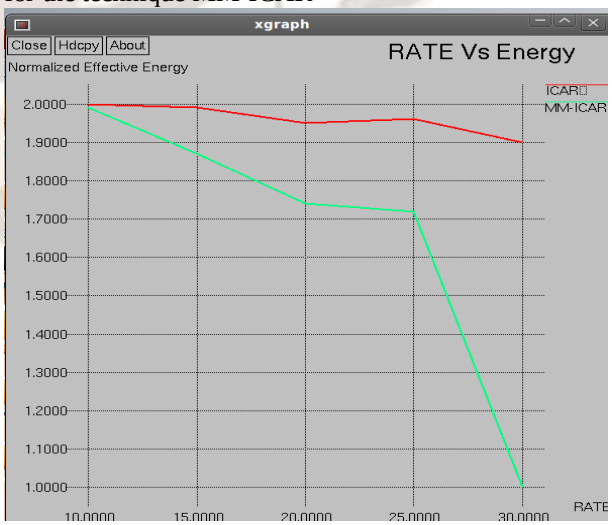

Fig 2: Graph showing the decrease in energy consumption for the technique MM-ICAR

IV. CONCLUSION

Designing an efficient transmission pattern in a wireless sensor network where all sensor nodes aggregate correlated data over intermediate nodes on the route to the sink in addition to making the sensor nodes aware of the interference of data from other nodes, have been addressed. Analysis of the impact of data aggregation in establishing routing paths towards the sink for the energy minimization problem is also done. These researches on several techniques that are related to the efficient routing of correlated data in a wireless sensor network have shown that power conservation remains a very crucial research area for the viability of wireless services. As a solution, we have added the idea of using multiple and mobile base stations to the energy efficient and interference aware routing of correlated data in wireless sensor networks. This replaces the existing techniques of using a single base station to which data from several sensor nodes can be sent.

REFERENCES

[1] Shousheng Zhao, Fengqi Yu, Baohua Zhao," *An Energy Efficient Directed Diffusion Routing Protocol*", International Conference on Computational Intelligence and Security 2007.

[2] Linliang Zhao, Gaoqiang Liu, Jie Chen, Zhiwei hang" *Flooding and Directed Diffusion Routing Algorithm in Wireless Sensor Networks*",Ninth International Conference on Hybrid Intelligent Systems,2009.

[3] C. Intanagonwiwat *et al.*, "*Impact of Network Density on Data Aggregation in Wireless Sensor Networks*," Proc. 22nd Int'l. Conf. Distrib. Comp. Sys., Vienna, Austria, July 2002.

[4] W.R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "*Energy-efficient Communication Protocol for Wireless Microsensor Networks*", in IEEE Computer Society Proceedings of the Thirty Third Hawaii International Conference on System Sciences (HICSS '00), Washington, DC, USA, Jan. 2000, vol. 8, pp. 8020.

[5] W.R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "*An Application-Specific Protocol Architecture for Wireless Microsensor Networks*" in IEEE Tmnsactions on Wireless Communications (October 2002), vol. 1(4), pp. 660-670.

[6] S. Lindsey and C. S. Raghavendra, "*Pegasis: Power-Efficient Gathering in Sensor Information Systems*," Proc. IEEE Aerospace Conf., Big Sky, MT, Mar. 2002.

[7] Kemal Akkaya and Mohamed Younis, "*A Survey on Routing Protocols for Wireless Sensor Networks*", *Ad hoc Networks*, vol. 3, no. 3, May 2005, pp. 325-349.

[8] R. Cristescu, B. Beferull-Lozano, and M. Vetterli, "*Networked Slepian-Wolf: Theory, Algorithms, and Scaling Laws*," *IEEE Trans. Info. Theory*, vol. 51, no.12, Dec. 2005, pp. 4057–73.

[9] P.V. Rickenbach and R. Wattenhofer, "*Gathering Correlated Data in Sensor Networks*," *Proc. ACM Joint Wksp. Foundations of Mobile Comp.*, Philadelphia, PA, Oct. 2004.

[10] R. Cristescu *et al.*, "*Network Correlated Data Gathering with Explicit Communication: NP-Completeness and Algorithms*," *IEEE/ACM Trans. Net.*, vol. 14, no. 1, Feb. 2006, pp. 41–54.

[11] Engin Zeydan et al, "*Efficient Routing for Correlated data in Wireless Sensor Networks*" IEEE 2008.

[12] R. Shah, S. Roy, S. Jain, and W. Brunette, *Data mules: Modeling a three-tier architecture for sparse sensor networks*," Proc. IEEE Workshop on Sensor Network Protocols and Applications, 2003.