International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459 (Online), An ISO 9001:2008 Certified Journal, Volume 3, Special Issue 1, January 2013)

International Conference on Information Systems and Computing (ICISC-2013), INDIA.

OVERVIEW OF DATA AGGREGATION BASED ROUTING PROTOCOLS IN WIRELESS SENSOR NETWORKS

V. Akila[1], Dr. T. Sheela[2]

[1] *Computer Science and Engineering, Bharath University, Chennai , India;*
[2]*Information Technology, Sri Sai Ram Engineering College,Chennai,India.*
Email: akila_be@yahoo.co.in

Abstract

Resource-constrained is a critical issue in Wireless Sensor Networks (WSNs) applications. Data aggregation is an efficient method to conserve energy by reducing packet transmissions and extend the network lifetime in WSNs. The total lifetime of the network is a major design factor for protocols used by WSNs. Routing protocols with low energy consumption are needed to prolong the lifetime of sensor network. In this paper, data aggregation based routing protocols for WSNs are analyzed. The routing protocols of structured architecture such as tree-based, cluster-based, chain-based and unstructured architecture such as structure-free are reviewed. The directions for future research are proposed in this area.

Keywords: **Wireless Sensor Networks;Mobile Sensor Networks; Data Aggregation;Routing Protocols**

I. INTRODUCTION

A wireless sensor network (WSN) is a collection of sensor nodes spatially deployed in an ad hoc fashion that performs distributed sensing tasks in a collaborative manner without relying on any underlying infrastructure support. Sensor networks are classified into two types. They are Mobile Sensor Network and Static Sensor Network. The mobile WSNs consist of a set of sensors which can move, sense, compute and transport data between themselves. The mobile sensors can trace moving target easily just by fixing them on the target and even it can move to the dangerous places by themselves but the static WSN can not have this capabilities. WSN applications include tracking wildlife migration, weather observation, pervasive computing, flood control, irrigation, Machine Health monitoring and Earthquake Detection. Sensor nodes can be deployed in geographical areas where it can be extremely difficult to recharge the in-built batteries or even replace the nodes. Hence it is the goal of every sensor network design to increase the longevity of the network. Energy consumption depends on sensing the data, processing the sensed data and transmitting or receiving them to or from its neighbor nodes. An energy-efficient solution for this is to use low duty cycling and minimizing redundant data transmission by data compression, data fusion, data aggregation and data filtering.

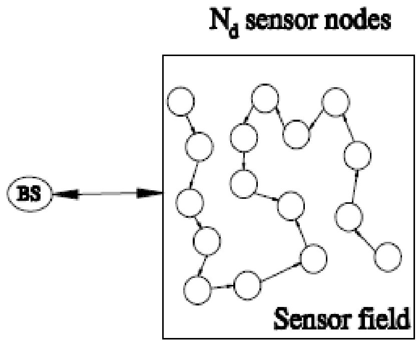

Fig.1. Description of a WSN

A WSN is assumed to have a remote base station (BS) as shown in Fig.1. A BS is an intelligent gateway node and is equipped with high computation power. To achieve long-lived WSNs, it is necessary to consider how to save energy in the conception of designed algorithms and protocols for WSNs where communication costs are usually more expensive then computing costs.

II. BASICS OF DATA AGGREGATION

Data Aggregation is the global process of gathering and routing information through a multi hop network, processing data at intermediate nodes with the objective of reducing resource consumption (in particular energy), hence it will increase the network lifetime.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459 (Online), An ISO 9001:2008 Certified Journal, Volume 3, Special Issue 1, January 2013)
International Conference on Information Systems and Computing (ICISC-2013), INDIA.

Data aggregation techniques may require some form of synchronization among nodes. The best strategy at a given node is not always to send data as soon as it is available. Waiting for information from neighboring nodes may lead to better data aggregation opportunities and, in turn, improved performance. The main timing strategies [8] proposed are Periodic simple aggregation, Periodic per-hop aggregation and Periodic per-hop adjusted aggregation. The efficiency of data aggregation algorithms depends on the correlation among the data generated by different information sources (sensor units).Such a correlation can be *spatial*, when the values generated by close-by sensors are related, *temporal*, when the sensor readings change slowly over time. *Aggregation gain* is defined as the measure of reduction in the communication traffic due to the aggregation. To achieve a higher aggregation ratio, one category focuses on establishing a proper routing scheme and the other focuses efficient timing control. The most important ingredient for in-network aggregation is a well designed routing protocol. *Classic routing protocols* typically forward data along the shortest path to the destination. *Data centric routing protocols* forward data based on the packet content and choose the next hop in order to promote in-network aggregation and minimize energy expenditure.

III. Routing Protocols And Hierarchies For Data Aggregation

Data Aggregation deals with the problem of forwarding packets and to enhance existing routing algorithms. In the following, we focus on each class of routing protocols separately (structure-free, tree-based, cluster-based and chain-based) by reviewing the main concepts and briefly commenting the pros and cons of each scheme.

3.1 Tree-based Approaches

It is usually based on a hierarchical organization of the nodes in the network. According to the tree-based approach as shown in Fig. 2, a spanning tree rooted at the sink is constructed first. Subsequently, such a structure is exploited in answering queries generated by the sink. This is done by performing in network aggregation along the aggregation tree by proceeding level by level from its leaves to its root. A key advantage of using a tree topology is no approximation error. The disadvantage of using a tree topology is communication error.

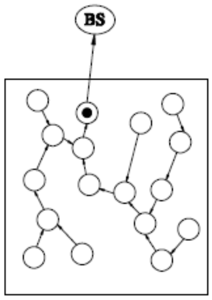

Fig.2. A Tree-based Routing Topology

SRTSD-DA [1] – SRTSD (Spanning Routing Tree protocol based on State Diffusion) and SRTSD-DA (SRTSD with Data Aggregation) protocols to ensure the connectivity and to save energy in mobile WSNs. *Routing tree construction* consists of startup phase, diffusion phase and schedule creation phase. In the startup phase, all sensors set their state to be interrupted, but the state of sink is always connected. The sink node broadcasts message to interrupted nodes within the communication radius of the sink. Nodes receive the location of sink node from the message. If the distance between the node and the sink is less than the reliable communication distance then the node ID and hop count are added into its connected nodes list (CNL). In *Diffusion phase*, the nodes traverses the CNL to find the best next hop node with smallest network cost *Wi* ,which is defined as

$$W_i = h_i \,/\, \left(E_i / E_o\right)^3 \qquad (1)$$

Where h_i is the hop count from node i to sink, E_i is residual energy of node i and each node is assumed to have the same initial energy E_o and changes its state to be connected and broadcasts the message. The diffusion phase continues until all nodes become connected. After diffusion phase, all the *connected* nodes form a tree whose root is sink node. In *Sensed Data Transmission Phase*, the connected sensors transmit sensed data to its next hop. The intermediate nodes receive data and forward the data directly. The SRTSD-DA algorithm is similar to SRTSD algorithm. We can change SRTSD to SRTSD-DA by doing small changes of diffusion phase and sensed data transmission phase of SRTSD. In the *diffusion phase* of SRTSD-DA, the network cost.

$$W_i = d_{AB} \,/\, \left(E_i / E_o\right) \qquad (2)$$

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459 (Online), An ISO 9001:2008 Certified Journal, Volume 3, Special Issue 1, January 2013)
International Conference on Information Systems and Computing (ICISC-2013), INDIA.

Where node *A* is a connected sensor in the CNL of sensor *B*, and d_{AB} is the distance between node *A* and node *B*. In the *sensed data transmission phase*, the intermediate nodes can aggregate the data and transmit aggregated data to its next hop.

MSPIN [2] – Modified SPIN achieves energy savings by discarding packet transmission to the opposite direction of sink node. This is called selective transmission. In *distance discovery* phase, sink node broadcasts startup packet in the network, sensor node stores the hop value as its hop distance from the sink node. The sensor node re-broadcast the startup packet after increasing the hop value by 1. If a sensor node receives multiple startup packets from different intermediate nodes then the minimum hop distance from multiple startup packets is the hop value for that sensor node. In *Negotiation* phase, the source node sends an ADV message. If hop distance of the receiving node is less than the hop distance sending node, then the receiving nodes send REQ message to the sending node for current data. In *Data transmission* phase, after request is received by the sending node, data is immediately sent to the requesting node. If the requesting nodes are intermediate nodes other than the sink node Negotiation phase repeats. This process continues till data reaches the sink node.

3.2. Cluster-based Approaches

It consists of a hierarchical organization of the network as shown in Fig.3. Nodes are subdivided into clusters and special nodes (cluster-heads) are elected in order to aggregate data locally and transmit the result of such an aggregation to the sink. The advantages and disadvantages of cluster-based schemes are very similar to those of tree-based approaches.

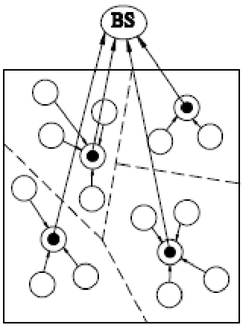

Fig.3.A Cluster-based Routing Topology

BED [3] -Biased Energy Distribution is to ensure well balanced energy consumption in order to maximize network lifetime. The initial configuration of the network commences when the sink floods the entire network with discovery packets. Each node determines the number of hops it takes to reach the sink and locates the nearest neighboring nodes for forwarding data to the sink from the discovery packets. In *aggregator selection algorithm*, sensor nodes can become an aggregator with a specified probability *p*. Each node generates a random number between 0 and 1 and if this number is less than the probability *p*, the node will request the amount of energy of its neighboring nodes by sending them energy request packets. The neighboring nodes then respond with energy reply packets that contain their IDs and amount of energy. Based on the energy levels, higher and lower energy nodes are determined. Nodes with more energy are thus selected as aggregators. The aggregator selection algorithm is repeated at specified intervals. The *cluster configuration algorithm* is achieved through the broadcasting of aggregator notify packets. When a node receives an aggregator notify packet, it stores the ID of the transmitting node in its routing table in order to know the route to the aggregator for forwarding data through the discovered route. If a node receives aggregator notify packets from more than one aggregator, the node will join the cluster that is closer to the aggregator. The *data forwarding algorithm* use the information in their routing table to find paths to the sink. After receiving the data from its cluster members, the aggregator compresses the data to reduce the amount of packets to be forwarded to the sink. The aggregator transmits the data to the sink through multi-hop transmission.

LEACH-MF [4] - Low Energy Adaptive Clustering Hierarchy-Multi- layer clustering achieves the goal of extending the network survival time and eliminates redundant information. The *cluster building phase* is similar to LEACH [9]. In the stable *data transfer phase*, the gathered data packages delivered to sink node by cluster-heads. When there are too many sensors in a large area, the number of cluster-heads is also tremendous. So, we can make these cluster-heads join to the new cluster-heads (super cluster-heads).Now the super cluster-heads are used to send data packages to the sink node. This is called *multi-layer clustering*. Data fusion is performed by cluster heads, after the members send the gathered packages to cluster-heads.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459 (Online), An ISO 9001:2008 Certified Journal, Volume 3, Special Issue 1, January 2013)
International Conference on Information Systems and Computing (ICISC-2013), INDIA.

The redundant information is eliminated by creating a data structure. The data structure is used to store packages gathered last time in the node and set a threshold. If the current package minus last package is greater than the threshold, the data structure discards the last package and stores the current package in it, finally sends it to the cluster-head. Otherwise, discards the current package.

OEERP [5] -Optimized Energy Efficient Routing Protocol effectively saves the energy of each node by reducing the number of broadcasting messages to the sink node. In the *Cluster Formation Phase*, some of the sensor nodes are randomly selected as a cluster-heads and these cluster-heads broadcast the ADV_CH packet to all the sensor nodes within the transmission range of different cluster heads. The sensor nodes join under the nearest cluster-head group by sending the JOIN_REQ packet to cluster-head to form a cluster. These randomly selected cluster-heads process the information up to certain time period (time-slot) until new cluster heads are selected. After new cluster-heads are elected to form clusters and process the received information, old cluster-heads remain as normal nodes. This process of changing the cluster head is repeated for every time-slot thereby every node gets an opportunity of becoming cluster head. This kind of approach could lead to uniform drain of node energy. In the *Information Processing Phase*, sensor node involves in sensing and sending the information to its cluster head, where the data is aggregated. In *the Data Dissemination Phase* the cluster head disseminates the aggregated information to the sink node.

3.3. Chain-based Approaches

In cluster-based sensor networks, sensors transmit data to the cluster head. However, if the cluster head is far away from the sensors, they might expend excessive energy in communication. The key idea behind chain-based data aggregation is that each sensor transmits only to its closest neighbor as shown in Fig. 4.

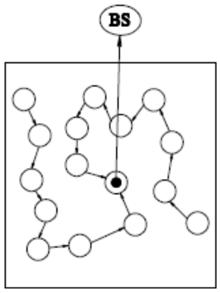

Fig.4. A Chain-based Routing Topology

CREEC [6] - Chain Routing with Even Energy Consumption achieves a longer average life time using maximizing the fairness of energy distribution and minimizing the total energy consumption. In *Throwing Schedule*, BS calculates the number of throwing to be assigned to any node for a super round. Every node can unify the throwing energy, if the number of throwing is assigned as a real number in inverse-proportion to the one-time throwing energy. In *Chain Establishment*, the BS generates a chain with three steps in every super round. In first step, the BS calculates the cumulative forwarding energy and sorts all nodes according to the decreasing order of cumulative forwarding energy and classifies them to three node levels. The two most depleted nodes in the sorted list belong to level-1, the next numbers of depleted nodes are selected as level-2 and the rest are level-3 nodes. Level-1 nodes are assigned as two leaf nodes in the chain. A level-1 or level-2 node is given a chance to preoccupy a very short adjacent link. In this way, forwarding energy consumption is compensated and balanced. The Second Step generates a chain using the constrained Kruskal's MST algorithm. It picks up the shortest link and the selected link is added to the current working chain under construction one by one until no more selection is possible. In third step, all non optimum greedy algorithms have a bad property to have long links at a few of last choices. The chain length is reduced by a link trimming algorithm called link swap. CREEC updates chains in every super round and builds new chain to save energy at depleted nodes.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459 (Online), An ISO 9001:2008 Certified Journal, Volume 3, Special Issue 1, January 2013)
International Conference on Information Systems and Computing (ICISC-2013), INDIA.

3.4. Structure-free Approaches

Structure-free approaches achieve data aggregation using local information, so they do not spend extra energy to build a structure. Structure free mechanisms perform well in a dynamic environment and periodic schemes are not applicable.

RDAG [7] –Real-time Data Aggregation Protocol involves the mechanisms of spatial aggregation and temporal aggregation. Real-time Data-aware Routing (RDR) supported by efficient data aggregation. To achieve a higher aggregation ratio, RDR makes a dynamic forwarding decision by sending the packets to the node which is spatially convergent. Fitness value is computed by considering the number of same-type packets at a node. If the forwarder has some packets of the same type than a node have a higher priority to be selected as the next hop If the forwarder does not have any packets of the same type then decision will be based on the real time routing policy to assign the priority. A node will be picked out of the neighbor node if the priority is the highest otherwise the back pressure rerouting mechanism is aimed instead of packet dropping. In the Judiciously Waiting policy for each packet along the way to the sink node to introduce artificial delays and increase temporal convergence for effective aggregation while meeting the time constraint of the data. Packet's slack time is proportional to the remaining hop count to the sink node. The remaining hop count to the sink node is formally calculated by dividing the distance from the current node to the sink and to the next hop forwarding node. By decreasing the remaining hop count a lower part of slack time is used for data aggregation as the packet moves closer to the sink. Finally, the packet uses its entire remaining deadline as a slack time in the current node if the next hop node is the sink node.

IV. CONCLUSIONS

The area of wireless sensor networking has been receiving attention among researchers in recent years. Data aggregation is one of the key techniques to solve the resource-constrained problem in WSN. A variety of new data aggregation based routing protocols targeted specifically at the wireless sensor networking environment have been proposed. All of them focus on optimizing important performance measures such as network lifetime and energy consumption.

The advantages and disadvantages of each data aggregation based routing protocols have been described in this paper.

REFERENCES

[1] Wei Xing, Kuntai Li, Yongchao Wang, Wen Zhou, "A Spanning Routing Tree Protocol Based on State Diffusion for Mobile Sensor Networks," in: IEEE Wireless Communications and Signal Processing (WCSP), 2011 International Conference ,pp.1-5.

[2] Zeenat Rehena, Sarbani Roy, Nandini Mukherjee, "A Modified SPIN for Wireless Sensor Networks,"in:IEEE Communication Systems and Networks (COMSNETS), 2011 International Conference ,pp.1-4.

[3] A. F. Salami, F. Anwar, H. Bello-Salau, A. M. Aibinu, "A Novel Biased Energy Distribution (BED) Scheme for Clustering Sensor Networks," in: IEEE Mechatronics (ICOM), 2011 International Conference ,pp.1-4.

[4] Jan-Feng Yan, Yuan-Liu Liu,"Improved LEACH Routing Protocol For Large Scale Wireless Sensor Networks Routing," in: IEEE Electronics, Communications and Control (ICECC), 2011 International Conference,pp. 3754 – 3757.

[5] K. Kishan Chand, P Vijaya Bharati and B. Seetha Ramanjaneyulu, "Optimized Energy Efficient Routing Protocol for Life-Time Improvement in Wireless Sensor Networks," in: IEEE Advances in Engineering, Science and Management (ICAESM), 2012 International Conference ,pp. 345 - 349 .

[6] Jisoo Shin and Changjin Suh **"**CREEC: Chain Routing with Even Energy Consumption," in: IEEE Communications and Networks, 2011,pp.17-25.

[7] Mohammad Hossein Yeganeh, Hamed Yousefi, Naser Alinaghipour,Ali Movaghar, "RDAG: A Structure-free Real-time Data Aggregation Protocol for Wireless Sensor Networks",in:Embedded and Real-Time Computing Systems and Applications (RTCSA), 2011 IEEE 17th International Conference ,Vol 1,pp.51-60.

[8] E. Fasolo, M. Rossi, J. Widmer, and M. Zorzi. "In-network aggregation techniques for wireless sensor networks: A survey",*IEEE Wireless Communications*, 14(2):70–87, 2007.

[9] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan,"Energy- Efficient Communication Protocol for Wireless Microsensor Networks," The 33rd Annual Hawaii International Conference on System Sciences, pp. 3005-3014, 2000.

[10] S. Lindsey, C. Raghavendra, and K.M. Sivalingam, "Data gathering algorithms in sensor networks using energy metrics," IEEE Trans.Parallel and Distributed Systems, vol. 13, no. 9, September 2002, pp.924-935.

[11] H. O. Tan and I. Korpeoglu, "Power efficient data gathering and aggregation in wireless sensor networks," SIGMOD Record, vol. 32, no.4, pp.66-71, December 2003.

[12] Do-Seong Kim and Yeong-Jee Chung, "Self-Organization Routing Protocol Supporting Mobile Nodes for Wireless Sensor Network," In Proc. Intl. Multi-Symposiums on Computer and Computational Sciences, IMSCCS 2006.

EXPLORING RESEARCH AND INNOVATIONS

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459 (Online), An ISO 9001:2008 Certified Journal, Volume 3, Special Issue 1, January 2013)

International Conference on Information Systems and Computing (ICISC-2013), INDIA.

[13] J. Zhang, Q. Wu, F. Ren, T. He, and C. Lin., "Effective data aggregation supported by dynamic routing in wireless sensor networks," In Proceedings of the IEEE International Conference on Communications *(ICC'10)*, pages 1–6, May 2010.

[14] J. Kulik, W.R. Heinzelman, H. Balakrishnan, "Negotiation–Based Protocols for Disseminating Information in Wireless Sensor Networks," In: Wireless Networks, Vol 8, pp. 169-185, 2002.

[15] Do-Seong Kim and Yeong-Jee Chung, "Self-Organization Routing Protocol Supporting Mobile Nodes for Wireless Sensor Network," In Proc. Intl. Multi-Symposiums on Computer and Computational Sciences, IMSCCS 2006.

Table 1.
Summary of the basic characteristics of the Routing Protocols.

Algo. Char.	**BED[3]**	**LEACH-MF[4]**	**OEERP[5]**	**SRTSD-DA[1]**	**MSPIN[2]**	**RDAG[7]**	**CREEC [6]**
Aggregation Method	Cluster-based	Cluster-based	Cluster-based	Tree-based	Tree-based	Structure Free	Chain-based
Energy Saving Methods	Prioritize the network into higher and lower energy nodes	Multi-layer Clustering	Cluster head is changed in every time slot.	Smaller hop count and shorter distance from sink	Selective Transmission	Judiciously Waiting Policy	Base Station can maintain the same energy level at all nodes
Timing Strategy	Periodic per hop	Periodic per hop	Periodic per hop	Periodic per hop	Periodic per hop	No Timing Strategy	Periodic per hop
Throughput	Better than LEACH[9]	-	It reaches higher levels as the new cluster formation requires additional broadcast and reply packets.	-	-	-	-
Life Time	Better than LEACH[9]	Better than LEACH[9]	It works 30% more compared to LEACH[9]	It works 4.5 times more compared to SRTSD and 1.6 times more compared to LEACH-MSN[15]	Better than SPIN-BC[14]	Better than DASDR[13]	It is more compared to PEGSIS[10], LEACH[9], PEDAP[11]
Energy Consumption	Less compared to LEACH[9]	Less compared to LEACH[9]	Less compared to LEACH[9]	Less compared to LEACH[9]	Less compared to SPIN-BC[14]	Less compared to DASDR [13]	Less compared to LEACH[9], PEGSIS[10], PEDAP[11]
Reliability	Reliable with respect to network lifetime, throughput and energy resources.	-	-	-	-	Dynamic routing in intermediate nodes, so it is very reliable	-
Scalability	No Knowledge of the precise position location of the nodes ,so it is high	-	-	-	-	-	-
Future Work	-	Application of this protocol in water area monitoring	Effective multi path data delivery to the access point.	Optimal value of time allocated to routing tree construction phase remains to be investigated	Few sensor nodes may be used several times remains to be investigated.	Apply in Mobile WSNs, Latency is evaluated	-