

# Data Fusion Techniques in Wireless Sensor Networks: Structured Vs. Structure-Free Approaches

Mahnaz Koupaee  
UC Santa Barbara  
Iran  
[koupaee@cs.ucsb.edu](mailto:koupaee@cs.ucsb.edu)

Mohammad Reza  
Kangavari  
Iran University of Science and Technology  
Iran  
[kangavari@comp.iust.ac.ir](mailto:kangavari@comp.iust.ac.ir)



**ABSTRACT:** Recent advancements in sensor technology, wireless networks and consequently wireless sensor networks and the increase of their applications in different fields have led to their great importance. One of the most important challenges is the management of huge amount of data produced by sensors in network in order to reduce data traffic in network and minimize the energy consumption. This paper gives a survey on different data fusion techniques dividing them into two main groups: structured and structure-free. The main properties of structured approaches are discussed and the motivations to introduce structure-free approaches are emphasized via describing the shortcomings of structured approaches.

**Keywords:** Wireless Sensor Network, Data Fusion, Structured and Structure-free fusion, Event-based Applications

**Received:** 10 December 2017, Revised 19 January 2018, Accepted 28 January 2018

© 2018 DLINE. All Rights Reserved

## 1. Introduction

Wireless sensor networks cause a lot of challenges according to the inherent limitations of sensors and the environments they are deployed in [1-2]. One of the main challenges of these networks is the limitation of available energy due to the low capacity of batteries embedded in sensors. Since the sensors are usually deployed in environments in which they cannot be accessed, the batteries cannot be replaced or recharged [3-5]. Therefore, the lifetime of a wireless sensor network is highly dependent on the lifetime of sensors, so saving energy of sensors can expand the lifetime of a network [6-7]. The main responsibilities of sensors are divided in three parts: sensing, processing and communication. The communication part, including both sending and receiving the packets, consumes the largest portion of available energy in a sensor [3]. There have been a lot of attempts to reduce the energy consumption and prolong network lifetime [8]. Meanwhile, the number of sensors deployed in an environment is increased because of possible errors in sensors and the necessity to cover an area both spatially and temporally [9-10]; this causes the scalability issue in wireless sensor networks. So a mechanism to reduce this amount of data generated by numerous

sensors is needed [11-12]. In order to manage the huge amount of produced data and reduce the energy consumption, a lot of methods were intended to take advantage of the sensor computation capacity and perform the desired fusion algorithm while data is being routed towards the sink [13-19].

The rest of this paper is organized as follows. Section 2 reviews the structured approaches by dividing them into two main groups. Section 3 explains structure-free methods which perform data fusion dynamically. The comparison of these two groups is described in section 4 and finally section 5 concludes the paper.

## **2. Structured Approaches**

The main challenges in wireless sensor networks include fault tolerance, scalability, adaptability to topology changes and energy consumption [12]. In-network data fusion can be used to overcome the scalability issue and reduce network traffic in a distributed manner. Performing data fusion in this way can reduce the communication cost which is the most expensive task of a sensor by aggregating same-type data and performing application dependent fusion before reaching the sink, therefore reducing the energy consumption [17]. The most popular distributed paradigm in wireless sensor networks is in-network aggregation. The main idea of this method is to use the computational capacity of sensor nodes to perform data fusion while routing data to the sink. This paradigm is also known as data-centric routing [5]. These methods are mainly divided into two groups: tree-based approaches and cluster-based approaches [16]. In the first group, a routing tree is constructed rooted at the sink. The data is forwarded from source nodes to the sink. The child nodes send their data to their parents. Each parent performs fusion after receiving all data from its children and then sends the result to its parent [16-18]. In the second group, the nodes are divided into clusters. A node acts as a cluster head; it receives data from cluster members, performs fusion and sends data to the sink. Some of the approaches in each category are briefly described as follows:

### **2.1 Tree-based Approaches**

One of the early methods proposed is TAG framework [20]. In this method users send the queries to the nodes. The nodes which can answer this query send their data through the routing tree rooted at the sink to the user. Fusion occurs at intermediate nodes while data is moving towards the sink. There is another method called EADAT in which sensors with higher residual energy are chosen as non-leaf nodes in order to prolong the network lifetime [21]. In another method called cascading timeout [22] the nodes in the routing tree choose their waiting time according to their distance to the sink; so the nodes further from the sink, consider lower waiting time. A tree-based data aggregation method using centralized method to manage data is introduced in [23].

### **2.2 Cluster-based Approaches**

One of the most famous protocols in this group is LEACH [17]. Cluster heads are randomly elected in order to distribute the energy consumption evenly among all nodes in the network. The cluster head is responsible to receive data from all cluster nodes and fuse it. The result is directly sent to the sink. LEACH-C is an improvement made on original LEACH to improve its efficiency [24]. In this protocol the procedure of choosing cluster heads is done with the help of the sink. Another group of these methods are chain-based algorithms. PPEGASIS [25] and Hierarchical PEGASIS [26] are the most famous ones. PEGASIS is based on following assumptions: the sink is located further away from the sensors and each node is able to perform perfect aggregation on its own data and the received data from its neighbor. This algorithm uses a greedy method to construct a chain starting from the furthest point in the network. A node is randomly chosen as the head. The head node is responsible to fuse the whole data received from nodes and send it to the sink. In order to balance energy consumption, each node will perform as the head. Hierarchical PEGASIS is an improvement on PEGASIS. In this algorithm there is the possibility of parallel transmission of data. The last node which receives data becomes the head node. There were also some other improvements on PEGASIS to reduce transmission delay and improve the chances for parallel transmission of data [27].

The algorithms in this group, assume that all nodes in the network are possible to reach the sink in one hop. This assumption limits the size of network. Moreover, if there is no possibility to fuse data in the head node, it must send a lot of packets towards sink. This may cause the head die soon because of the high amount of energy used for transmission [18].

Tree-based and cluster-based approaches perform well in stable environments in which the nodes work all the time. But in practical environments where the nodes may stop working or some nodes may be added because of application requirements, and in event-based networks the cost of construction and maintenance of a structure is high. Therefore, constructing a structure cannot be beneficial.

These methods perform well in data gathering applications where the traffic patterns do not change, the error rate of sensors is low and the topology of the network do not change due to sensor failures or adding some sensors to network. However, in event based networks where source nodes change with the event and in dynamic scenarios where the nodes may be added or removed from the network, the construction and maintenance of a structure may be too expensive both in terms of energy and traffic in network [19]. Figure (1) shows the poor performance of tree-based and cluster-based approaches in event-based networks. Although the nodes sensing the event are spatially close to each other, they are not able to be fused before reaching the sink. There are also some applications in which there are different types of data. A static fusion algorithm cannot be optimal while a dynamic approach can fuse the same-type data as early as possible on its way to the sink. So the fusion algorithm is performed sooner and the number of transmissions is reduced.

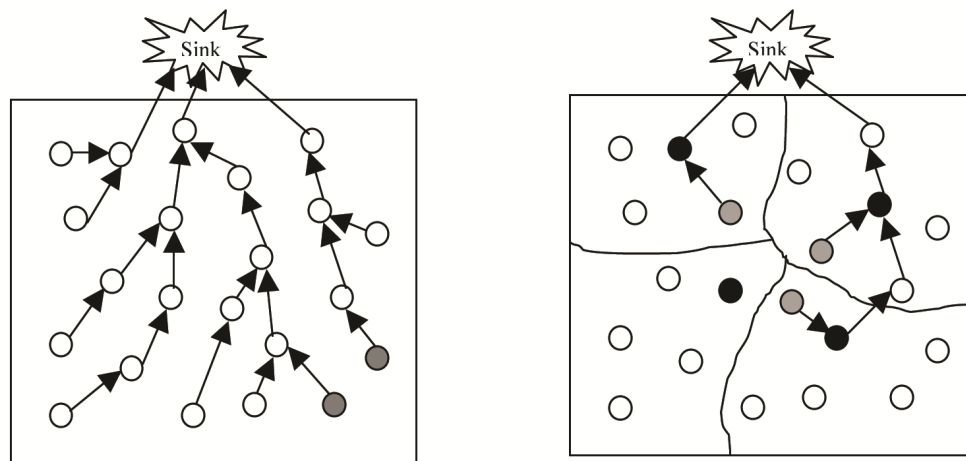


Figure 1. Poor performance of structured approaches: (a) Tree-based (b) Cluster-based

### 3. Structure-free Approaches

Based on the challenges issued for structured approaches, structure-free methods are intended to fuse data using local information of the nodes. The fusion can take place efficiently by realizing two important conditions known as temporal and spatial convergence. These two conditions which are necessary for in-network data fusion can determine the performance of different algorithms.

#### 3.1 Temporal Convergence

Power efficient data fusion algorithms require data to be met at intermediate nodes before reaching the sink; to do so, packets must be present in the same node at the same time. In tree-based approaches parent nodes wait until they receive data from all their children and then perform fusion and forward data. Cluster-based approaches achieve this by making the cluster heads wait for data from all cluster members. On the other hand, in structure-free approaches the destinations of nodes are not known in advance; so other timing mechanisms are needed to achieve temporal convergence and increase the chances of data fusion at intermediate nodes. Different approaches apply different timing mechanisms.

#### 3.2 Spatial Convergence

This condition is achieved by sending packets to nodes which still have packets for aggregation. In structured approaches, a structure is constructed in advance and packets are routed through this structure; so some packets may not be aggregated while being in nodes spatially close to each other. Structure-free approaches dynamically choose each node's destination based on local information thus increasing the spatial convergence of the method. In these methods, destinations are chosen dynamically according to each event.

The first attempt to propose a structure-free protocol is introduced in [18]. There are two mechanisms used in this protocol to improve chances of aggregation: randomized waiting time to improve temporal convergence and data-aware any casting to make data more spatially convergent. The early aggregation only occurs when the nodes closer to sink choose longer delays. A dynamic mechanism to fuse data is introduced in [28]. To converge packets spatially, the potential field theory is used to improve

chances of same-type packets visit. This waiting time policy may cause long delays for sensors in large-scale networks. RAG is a real-time structure-free protocol introduced in [18]. To improve chances of early fusion, waiting time is assigned to packets so that packets are delayed at intermediate nodes according to their distance to sink. When the waiting time is over, the same type packets are aggregated and forwarded to next hop. Data aware any casting is also used to route packets forward. An attribute aware aggregation method is introduced in [29] it also applies the potential field theory to spatially fuse packets. An adaptive timing algorithm is used to delay packets at intermediate nodes to make fusion process more effective by reducing the number of transmissions. Data from different types and different applications are considered in this method.

[30] Assigns waiting time to the node rather than packets. This makes all nodes, even those which are not in the event area, have the possibility to fuse data rather than just forwarding separate packets. The calculation of the waiting time depends on two factors; distance to the sink and distance to the furthest point of the network. These two factors will make the desired difference between the waiting times of different nodes which is especially suitable for dense wireless sensor networks. Data types can be different so we can have a network of heterogeneous sensors. For each event and after the expiration of the waiting time, nodes are chosen to be fused according to the value of these attributes. So that the volume of the data is reduced and the redundant data is not forwarded to the sink. The spatial and temporal mechanisms which are used in this method, makes the whole fusion convergent; thus data of different nodes are met and fused before reaching the sink.

Table (1) compares some of the famous structure-free methods. All methods in this group try to realize two conditions in order to perform fusion. Those two conditions are temporal convergence and spatial convergence. In order to perform fusion, packets containing data need to be met at the same type at the same place.

	<b>Convergent</b>	<b>Data type</b>	<b>Spatial convergence</b>	<b>Assignment of waiting time</b>	<b>Temporal convergence</b>
DAA+RW [18]	-	Homogenous sensors with the same data type	Data aware any cast at MAC layer	Packet	Randomized waiting time
DASDR [28]	✓	Homogenous sensors with the same data type	Potential field theory	-	-
RAG [19]	-	Heterogeneous sensors with different data type	Data aware any cast at MAC layer	Packet	Judiciously waiting time Policy
ADA [29]	✓	Heterogeneous sensors with the same data type	Potential field theory	Packet	Packet-Driven Adaptive Timing Scheme
BN+WT [30]	✓	Heterogeneous sensors with different data type	Graph summarization using sensors attributes	Node	Waiting time according to the distance to the sink and furthest point of the network

Table 1. Comparison of structure-free methods

#### 4. Structured Approaches Vs. Structure-free Approaches

Table 2 compares these two approaches by naming their advantages and disadvantages. The first positive point of structured approaches is their speed to deliver fused data to the sink. Since the destination of each node is known beforehand, at transmission time the data is sent immediately to the destination which is known in advance. Structured approaches intend to construct a structure such as a tree or cluster before network starts its mission. Therefore, if the change rate of network is high due to sensors failures or changes in their positions, the structure needs to be reconstructed which may cause the increase of energy consumption. In event-based sensor networks, for each event happening, there are only some nodes sensing the event; so having a static structure is not suitable as previously mentioned in section 1. Static structures prevent same-type data which are spatially close to each other to be fused. So those data will take partially or completely different routes to the sink. For monitoring applications, having a static structure makes the fusion process faster. In these applications, sensors contain data

all the time. So having pre-known routes to forward data will reduce the response time and traffic load by performing in-network fusion.

	Disadvantages	Advantages
<b>Structure-free approaches</b>	<ul style="list-style-type: none"> <li>• The overhead of construction and maintenance of a structure</li> <li>• Being static so expensive to adapt them to network changes</li> <li>• Not suitable for event-based networks</li> <li>• Not able to fuse spatially related data</li> <li>• High energy consumption for transmitting non fused data</li> </ul>	<ul style="list-style-type: none"> <li>• High speed in transmitting data to the sink</li> <li>• Suitable for monitoring applications and static networks or networks with low change rate</li> </ul>
<b>Structure approaches</b>	<ul style="list-style-type: none"> <li>• Lower speed of transmission since the routes are chosen dynamically</li> <li>• Highly delay time and response time</li> </ul>	<ul style="list-style-type: none"> <li>• The early fusion of data while being forwarded to the sink</li> <li>• Being dynamic and thus low overhead and its adaptability to the changes</li> <li>• Fault-tolerant</li> <li>• Suitable for event-based networks and dynamic networks with high rate of changes</li> <li>• Routing data using only local information without a centralized mechanism</li> </ul>

Table 2. Structured approaches vs. structure-free approaches

Considering the mentioned problems for structured approaches and the need for a fusion method regarding the special characteristics of event-based sensor networks, there are some efforts designing structure-free approaches. Paying attention to the point sensors including data may vary by time, the structure needs to be dynamic. Thus in these approaches, according to the local information of the sensors, its destination is determined without any centralized management. Being dynamic, enables spatially convergent data to be fused close to the event. This causes reduction in network traffic. Since structure free approaches are dynamic, they are flexible to changes; so failures of sensors or changes in network topology do not reduce the performance of network and the maintenance can take place easier without high cost. On the other hand, the lack of a structure and therefore not knowing the destinations in advance may cause longer delays because each time each sensor needs to choose its destination for fusion and transmission.

## 5. Conclusion

In this paper a brief discussion of available data fusion techniques is introduced. Considering the shortcomings of the tree-based and cluster-based approaches known as structured approaches, dynamic methods to fuse and route packets are introduced. Comparing both groups and knowing the advantages of each category can lead to a better choice of fusion method for different applications.

## References

- [1] Hammoudeh, M., Newman, R. (2013). "Adaptive Routing in Wireless Sensor Networks: QoS Optimization for Enhanced Application Performance. *Journal of Information Fusion* 22, p. 3-15.
- [2] Yick, J., Mukherjee, B., Ghosal, D. (2008). "Wireless Sensor Network Survey, *Journal of Computer Networks* 52 (12) 2292-2330.
- [3] Akyildiz, I. F., Su, W., Sankarasubramaniam, Y., Cayirci, E. (2002). Wireless sensor networks: a survey, *Journal of Computer*

- [4] Guiyi, W., Yun, L., Binfeng, G., Bin, X., Athanasios, V. (2011). Prediction-based data aggregation in wireless sensor networks: Combining grey model and Kalman Filter, *Computer Communications* 34, p. 793-802
- [5] Mikami, Sh., Aonishi, T., Yoshino, H., Ohta, Ch., Kawaguchi, H., Yoshimoto, M. (2006). Aggregation Efficiency-Aware Greedy Incremental Tree Routing for Wireless Sensor Networks, *IEICE Transactions on Communications* 89 (10) 2741-2751.
- [6] Zhang, K., Li, C., Zhang, W., (2013). Wireless Sensor Data Fusion Algorithm Based on the Sensor Scheduling and Batch Estimate, *International Journal of Future Computer and Communication* 2 (4) 333.
- [7] Qingqing, L., Chunlin, L., Jun, L. (2010). Data Aggregation Algorithm Based on Grid and Adaptive Genetic Algorithm for Wireless Sensor Networks with a Mobile Sink, *In: 2nd International Workshop on Intelligent Systems and Applications*.
- [8] Maraiya, K., Kant, K., Gupta, N. (2011). Study of Data fusion in Wireless Sensor Network, *In: Proc. of the International Conference on Advanced Computing and Communication Technologies (ACCT 2011)*.
- [9] Nakamura, E. F., Loureiro, A. A. F., Frery, A. C. (2007). Information Fusion for Wireless Sensor Networks: Methods, Models, and Classifications, *ACM Computing Surveys* 39 (3) 9.
- [10] Zhu, Y., Vedantham, R., Park, S-J., Sivakumar, R. (2008). A scalable correlation aware aggregation strategy for wireless sensor networks, *Information Fusion* 9, p. 354-369
- [11] Younis, O., Fahmy, S. (2004). Distributed Clustering in Ad-hoc Sensor Networks: A Hybrid, Energy-Efficient Approach, *Twenty-third Annual Joint Conference of the IEEE Computer and Communications Societies*.
- [12] Wenz, M., Wom, H. (2006). Event-based Production Rules for Data Aggregation in Wireless Sensor Networks, *IEEE International Conference on Multi-sensor Fusion and Integration for Intelligent Systems*, Germany, p. 59- 64
- [13] Du, W., Deng, J., Han, Y. S., Varshney, P. K. (2003). A Witness-Based Approach for Data Fusion Assurance in Wireless Sensor Networks, *In: Global Telecommunications Conference, GLOBECOM'03. IEEE*, 3, p. 1435- 1439
- [14] Dai, X., Xia, F., Wang, Z., Sun, Y. (2005). A Survey of Intelligent Information Processing in Wireless Sensor Network, *Mobile Ad-hoc and sensor networks*, p. 123-132.
- [15] Yu, B., Scerri, P., Sycara, K., Xu, Y., Lewis, M. (2006). Scalable and Reliable Data Delivery in Mobile Ad Hoc Sensor Networks, *In: Proceedings of the Fifth International Joint Conference on Autonomous agents and multi agent systems*, p. 1071-1078
- [16] Chao, Ch-M., Hsiao, T-Y. (2009). Design of Structure-Free and Energy-Balanced Data Aggregation in Wireless Sensor Networks, *11th IEEE International Conference on High Performance Computing and Communications, HPCC'09*, p. 222-229.
- [17] Heinzelman, W., Chandrakasan, A., Balakrishnan, H. (2000). Energy-Efficient Communication Protocol for Wireless Microsensor Networks, *In: Proceedings of the 33rd Hawaii International Conference on System Sciences*, p. 10.
- [18] Fan, K-W., Liu, Sh., Sinha, P. (2007). Structure-free Data Aggregation in Sensor Networks, *IEEE Transactions on Mobile Computing* 6, p. 929-942.
- [19] Yousefi, Y., Yeganeh, M. H., Alinaghipour, N., Movaghar, A. (2012). Structure-Free Real-Time Data Aggregation in Wireless Sensor Networks, *Computer Communications* 35, p. 1132-1140.
- [20] Madden, S., Franklin, M. J., Hellerstein, J., Hong, W. (2002). TAG: A Tiny Aggregation Service for Ad-Hoc Sensor Networks, *In: Proceedings of the Fifth Symposium on Operating Systems Design and implementation* 36(SI) p. 131-146.
- [21] Ding, M., Cheng, X., Xue, G. (2003). Aggregation tree construction in sensor networks, *In: Proceedings of the 58th IEEE Vehicular Technology Conference*, p. 2168-2172
- [22] Solis, I., Obraczka, K. (2004). The impact of timing in data aggregation for sensor networks, *In: Proceedings of the IEEE International Conference on Communications (ICC' 04)* p. 3640-3645.
- [23] Du, H., Hu, X., Jia, X. (2006). Energy Efficient Routing and Scheduling for Real-Time Data Aggregation in WSNS, *Computer Communications* 29, p. 3527-3535.
- [24] Heinzelman, W. R., Chandrakasan, A., Balakrishnan, H. (2002). An Application-Specific Protocol Architecture for Wireless Microsensor Networks, *IEEE Transactions on Wireless Communications*, 1 (4) 660-670.
- [25] Lindsey S., Raghavendra, C. (2002). PEGASIS: Power-efficient gathering in sensor information systems, *In: Proceedings of*

IEEE Aerospace Conference, 3.

- [26] Lindsey, S., Raghavendra, C., Sivalingam, K. M. (2001). Data Gathering in Sensor Networks using the Energy\*Delay Metric, *In: Proceedings 15th International Parallel and Distributed Processing Symposium*, p. 2001–2008.
- [27] Lindsey, S., Raghavendra, C., Sivalingam, K. M. (2002). Data Gathering Algorithms in Sensor Networks Using Energy Metrics, *In: IEEE Transactions on Parallel and Distributed Systems* 13, p. 924–935
- [28] Zhang, J., Wu, Q., Ren, F., He, T., Lin, C. (2010). Effective Data Aggregation Supported by Dynamic Routing in Wireless Sensor Networks, *In: Proceedings of the IEEE International Conference on Communications (ICC'10)*, p. 1–6.
- [29] Ren, F., Zhang, J., Wu, Y., He, T., Chen, C., Lin, Ch. (2013). Attribute-Aware Data Aggregation Using Potential- Based Dynamic Routing in Wireless Sensor Networks, *IEEE Transactions On Parallel and Distributed Systems*, 24(5) 881-892.
- [30] Koupaee, M., Kangavari, M. R., Amiri, M. J. (2017). Scalable structure-free data fusion on wireless sensor networks, *Journal of Supercomputing*, Accepted.