Clustering Strategies in Wireless Sensor Network & Security Concern in Data Aggregation

Sonal Dudhat[1] Jayesh Rathod[2]
[1,2]M.E. in Computer Engineering
[1,2]Atmiya Institute of Technology & Science

Abstract — A sensor network is a network of low-powered, energy-constrained nodes which are equipped with sensors, processors, memory and wireless communication devices. The sensor nodes are operated by lightweight batteries. The capability of an individual sensor node is limited, and that are densely deployed within the sensing field. Clustering is method of grouping sensor nodes. It is a basic approach for designing energy-efficient, robust and highly scalable distributed sensor networks. By using clusters we can reduce the communication overhead, thereby decreasing the energy consumption and interference among the sensor nodes. In clusters we have a cluster head node all other nodes of the cluster communicate through this cluster head so communication overhead is reduce and energy is saved. Most existing proposals for data aggregation are subject to attack. Once a single node is compromised, it is easy for an adversary to inject false data into the network and mislead the aggregator to accept false readings. Because of this, the need for secure data aggregation is raised. A general definition for secure data aggregation is the efficient delivery of the summary of sensor readings that are reported to an onsite user in such a way that ensures these reported readings have not been altered.

I. INTRODUCTION

A wireless sensor network is a wireless network consisting of small devices which can monitor physical or environmental conditions like temperature, pressure, motion or pollutants etc. at different regions.

The small device, known as sensor node, consists of a radio transceiver, microcontroller, power supply, and the actual sensor. Initially sensor network were used for military applications but now they are widely used for civilian application area including environment and habitat monitoring, healthcare application and so on.

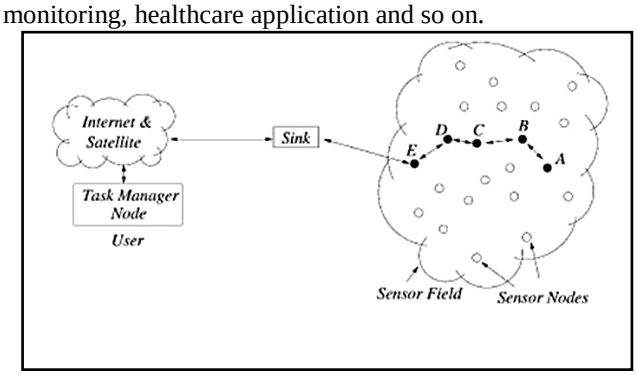

Fig. 1

Normally sensor nodes are distributed throughout the region which has to be monitored; they self-organize in to a network through wireless communication, and collaborate with each other to complete the common task. Now days, sensor nodes are becoming smaller, cheaper, and more powerful so we can easily deploy a large-scale sensor network.

Basic features of sensor networks are self-organizing capabilities, dynamic network topology, limited power, node failures & mobility of nodes, short-range broadcast communication and multi-hop routing, and large scale of deployment. The strength of wireless sensor network lies in their flexibility and scalability.

There are several key limitations in WSNs, that clustering schemes must consider:

A. Limited Energy:

Unlike wired designs, wireless sensor nodes are ”off-grid”, meaning that they have limited energy storage and the efficient use of this energy will be important in determining the range of suitable applications for these networks. The limited energy in sensor nodes must be considered as proper clustering can reduce the overall energy usage in a network.

B. Network Lifetime:

The energy limitation on nodes results in a limited network lifetime for nodes in a network. Proper clustering should attempt to reduce the energy
usage, and also increase network lifetime.

C. Limited Abilities:

The small physical size and small amount of stored energy in a sensor node limits many of the abilities of nodes in way of processing and communication abilities. A good clustering algorithm should make use of shared resources within an organizational structure so ability can be maximized.

D. Application Dependency:

Normally a given application will heavily rely on cluster organization. When designing a clustering algorithm, application robustness must be considered as a good clustering algorithm should be able to adapt to a variety of application requirements.

II. CLUSTERING IN WSN

It is widely accepted that the energy used in one bit of data transfer can be used to perform a large number of arithmetic operations in the sensor processor. In a densely deployed sensor network the physical environment would produce very similar data in close-by sensor nodes and transmitting such data is more or less redundant. Therefore, using some kind of grouping of nodes such that data from sensor nodes of a group can be combined or compressed together in an

intelligent way and transmit only compact data.

There are some issues involved with the process of clustering in a wireless sensor network. First issue is, how many clusters should be formed that could optimize some performance parameter. Second could be how many nodes should be taken in to a single cluster. Third important issue is the selection procedure of cluster-head in a cluster.

A. Essential part of the Clustering:

1) Sensor Node:

A sensor node is the component of a WSN. Sensor nodes can take on multiple roles in a network, like simple sensing; data storage; routing; and data processing.

2) Clusters:

Clusters are the main unit for WSNs. The dense nature of these networks require the need for them to be broken down into clusters to simplify tasks such a communication

3) Cluster heads:

Cluster heads are the leader of a cluster. They often are required to organize activities in the cluster. These tasks include but are not limited to data-aggregation and organizing the communication schedule of a cluster.

4) Base Station:

The base station is at the upper level of the hierarchical WSN. It provides the communication link between the sensor network and the end-user.

5) End User:

The data in a sensor network can be used for a wide-range of applications. Therefore, a particular application may make use of the network data over the internet, using a PDA, or even a desktop computer. In a queried sensor network (where the required data is gathered from a query sent through the network). This query is generated by the end user.

There are several key attributes that must carefully consider, which are of particular importance in wireless sensor networks:

1) Cost of Clustering:

Clustering plays an important role in organizing sensor network topology; there are normally many resources like communication and processing tasks needed in the creation and maintenance of the clustering topology. Such costs as the required resources are not being used for data transmission or sensing tasks.

2) Selection of Cluster heads and Clusters:

The clustering strategy gives good benefits for wireless sensor networks. When designing for a particular application, designers must carefully examine the formation of clusters in the network. Depending on the application, some requirements for the number of nodes in a cluster or its physical size may play an important role in its operation. This pre work may have an impact on how cluster heads are selected in this application.

3) Real-Time Operation:

WSN is generally used for sensing the data. So it should give the response in real time. In applications such as habitat monitoring, simply receiving data is sufficient for analysis, meaning delay is not an important issue. When we look at a military tracking, the issue of real-time data getting becomes much more important. When looking at clustering algorithms, important attention must be paid to the delay created by the clustering scheme itself. In addition, the time required for cluster recovery mechanisms is also very important factor.

4) Synchronization:

One of the primary limitations in Wireless Sensor Networks is the limited energy capacity of nodes. Slotted transmission schemes (such as TDMA); allow nodes to regularly schedule sleep intervals to minimize energy used. Such schemes require synchronization mechanisms to setup and maintain the transmission schedule. When considering a clustering scheme, synchronization and scheduling will have a considerable effect on network lifetime and the overall network performance.

5) Data Aggregation:

One major advantage of wireless sensor networks is the ability for data aggregation to occur in the network. In a densely populated network there are normally multiple nodes sensing similar information. Data aggregation allows the differentiation between sensed data and useful data. As such, the amount of data transferred in network should be minimized. Many clustering schemes provide data aggregation capabilities, and as such, the requirement for data aggregation should be carefully considered when selecting a clustering approach.

6) Repair Mechanisms:

Due to the nature of Wireless Sensor Networks, they are having node mobility, node death and interference. All of these situations can result in link failure. When looking at clustering schemes, it is important to look at the mechanisms in place for link recovery and reliable data communication.

7) Quality of Service (QoS):

From an overall network point, we can look at QoS requirements in Wireless Sensor Networks. Many of these requirements are application dependant and as such, it is important to look at these services when choosing a clustering scheme. Implementations can vary widely in terms of these metrics, and as a result, the design process should consider these things.

B. Advantages of Clustering:

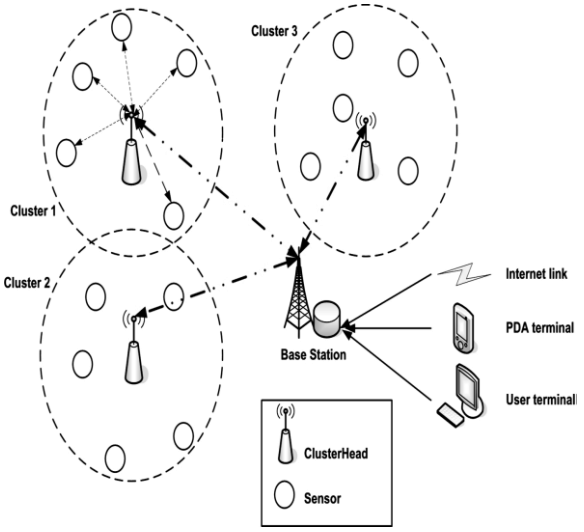

Fig. 2

- Transmit aggregated data to the data sink.
- Reducing number of nodes taking part in transmission.
- Useful energy consumption.
- Scalability for large number of nodes.
- Reduces communication overhead for both single and multi-hop.

III. DATA AGGREGATION

In sensor network, the energy is mainly consumed for three purposes: data transmission, signal processing, and hardware operation. It is said in that 70% of energy consumption is due to data transmission. So for maximizing the network lifetime, the process of data transmission should be optimized. The data transmission can be optimized by using efficient routing protocols and effective ways of data aggregation.

Routing protocols have their ways to save energy of nodes in the network by providing or creating an optimal route from sensor nodes to base station or sink. Data aggregation plays an important role in energy conservation of sensor network. Data aggregation methods are used not only for finding an optimal path from source to destination but also to eliminate the redundancy of data, since transmitting large volume of raw data is an energy intensive operation, and minimizing the number of data transmission. This also reduces the amount of transmission power expended by nodes.

Data aggregation techniques show how the data is to be routed in the network as well as the processing method that are applied on the packets received by a node. They have a great impact on the energy consumption of nodes and thus on network efficiency by reducing number of transmission or length of packet.

A. Aggregation functions:

- Average (AVG)
- Maximum (MAX)
- Sum (SUM)
- Count (COUNT)
- Minimum (MIN)

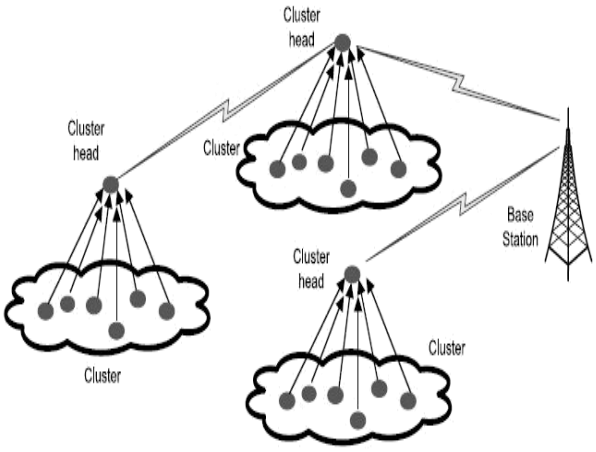

Fig. 3

Data aggregation reduces the communication overhead in the network, thus saving the sensor energy resources. In addition, aggregation reduces channel contention and packet collisions.

IV. RELAY NODE

Sensor networks have uses in disaster prediction, security, environmental monitoring, and traffic control. A wide variety of network sizes are used in these applications. In environmental monitoring, for example, hundreds or thousands of sensor nodes are deployed in a large monitoring area. In such large scale sensor networks, scalability and robustness are very important. In addition, sensor nodes are highly power constrained, and they must work at very low power to make long the lifetime of the sensor network.

The idea is to use relay nodes to take some load away from the sensor nodes. The deployment of relay nodes in sensor networks has also been proposed for energy- efficient data gathering, load-balancing as well as to make the network fault tolerant. These relay nodes can be making with higher energy, as compared to the sensor nodes, and can be used as cluster-heads in hierarchical sensor networks. However, similar to the sensor nodes, the relay nodes are also battery operated and hence, power constrained.

In hierarchical sensor network architecture, if a sensor node depletes its energy, it may result in a localized degradation of performance, but will not affect the whole network. But the total depletion of the power of a relay node, especially in a hierarchical architecture, can impact the functionality of the network more badly than the depletion of the battery of a single sensor node. This is because, when the battery of a relay node is totally depleted, the sensor nodes which are transmitting to this relay node will no longer be able to send their data to the base station and an entire area within the network becomes non operative. It is very important to properly distribute the total load for data communication, over the different relay nodes.

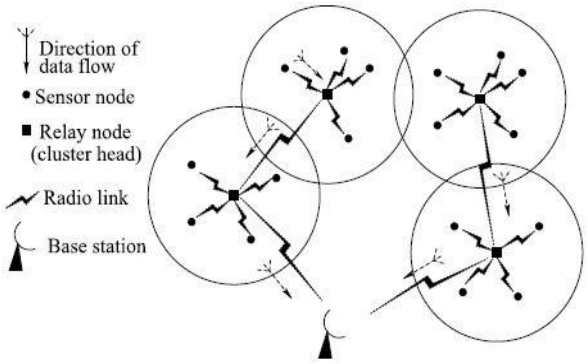

Fig. 4

According to literature analysis the max load of network is decided by the cluster headers neighbour to the sink. This is because is where traffic gathers. In addition, when cluster size increases to maximum size, the transmission range also increases to guarantee communication connectivity among cluster headers. If this causes more interference with traffic gathering in the sink, then packet delivery decreases.

The decrease of packet delivery can be improved most effectively by positioning relay nodes between the sink and cluster headers that are one-hop close to the sink. Because these relay nodes decrease transmission range of cluster

headers that are one-hop close to the sink, the interference of cluster headers that are one-hop close to the sink decreases, resulting improvement of packet delivery.

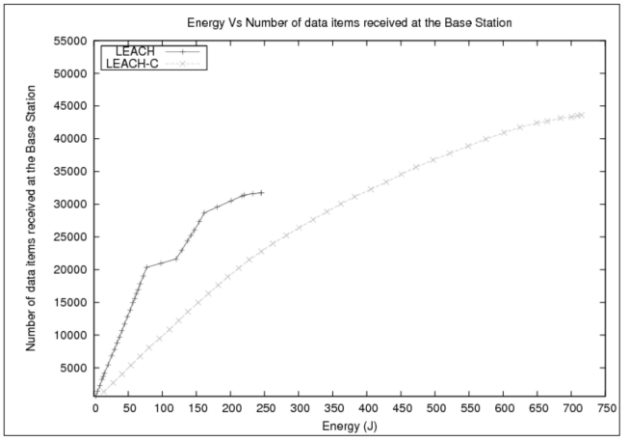

Fig. 4
Plot of Energy vs. Number of data signals received at BS

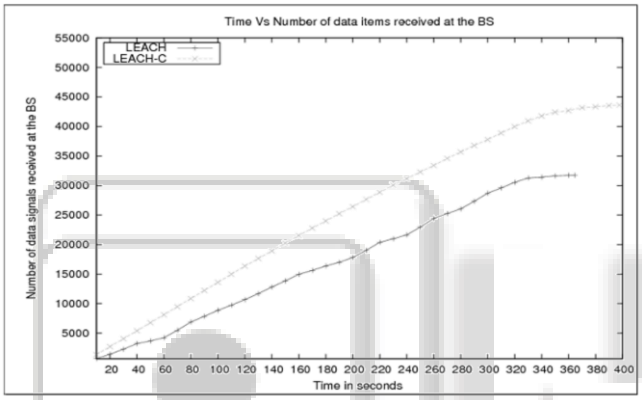

Fig. 5
Plot of Time vs. Number of data signals received at BS

V. SECURITY CONCERN

Data Confidentiality: ensures that information content is never exposed to anyone who is not authorized to receive it. It can be divided (in secure data Aggregation schemes) into a hop-by- hop basis and an end-to-end basis. In the hop-by-hop basis, any aggregator point needs to decrypt the received encrypted data, apply some sort of aggregation function, encrypt the aggregated data, and send it to the upper aggregator point. This kind of confidentiality implementation is not practical for the WSN since it requires extra computation. On the other hand in end-to-end basis, the aggregator does not need to decrypt and encrypt data and instead of this, it needs to apply the aggregation functions directly on the encrypted data by using encryption.

Data Integrity: ensures that the content of a message has not been altered, either maliciously or by accident, during transmission process. Confidentiality It is not enough since an adversary is still able to change the data although it knows nothing about it. Suppose a secure data aggregation scheme focuses only on data confidentiality. An adversary near the aggregator point will be able to change the aggregated result sent to the base station by adding some fragments or manipulating the packet's content without detection. Moreover, even without the existence of an adversary, data might be damaged or lost due to the wireless environment.

Data Freshness: ensures that the data are recent and that no old messages have been replayed to protect data aggregation schemes against replay attacks. In this kind of attack, it is not enough that these schemes only focus on data confidentiality and integrity because a passive adversary is able to listen to even encrypted messages transmitted between sensor nodes can replay them later on and disrupt the data aggregation results. More importantly when the adversary can replay the distributed shared key and mislead the sensor about the current key.

Data Availability: ensures that the network is alive and that data are accessible. It is highly recommended in the presence of compromised nodes to achieve network degradation by eliminating these bad nodes. Once an attacker gets into the WSN by compromising a node, the attack will an affect the network services and data availability especially in those parts of the network where the attack has been launched. Moreover, the data aggregation security requirements should be carefully implemented to avoid extra energy consumption. If no more energy is left, the data will no longer be available. When the adversary is getting stronger, it is necessary that a secure data aggregation scheme contains some of the following mechanisms to ensure reasonable level of data availability in the network.

- Self-healing that can diagnose, and react to the attacker's activities especially when he gets into the network and then start corrective actions based on defined policies to recover the network or a node.
- Aggregator rotation that rotates the aggregation duties between honest nodes to balance the energy consumption in WSN.

Authentication: There are two types of authentication; entity authentication, and data authentication. Entity authentication allows the receiver to verify if the message is sent by the claimed sender or not. Therefore, by applying authentication in the WSNs, an adversary will not be able to participate and inject data into the network unless it has valid authentication keys. On the other hand, data authentication guarantees that the reported data is the same as the original one. In a secure data aggregation, both entity and data authentication are important since entity authentication ensures that some exchanged data between sensors. For example, electing an aggregator point or reporting invalid aggregated results are authenticated using their identity while data authentication ensures that raw data are received at the aggregators at the same time as they are being sensed.

Non-repudiation: ensures that a transferred packet has been sent and received by the person claiming to have sent and received the packet. In secure aggregation schemes, once the aggregator sends the aggregation results, it should not be able to deny sending them. This gives the base station the opportunity to determine what causes the changes in the aggregation results.

Data Accuracy: One major outcome of any aggregation scheme is to provide an aggregated data as accurately as possible since it is worth nothing to reduce the number of bits in the aggregated data but with very low data accuracy. A trade-o between data accuracy and aggregated data size should be considered at the design stage because higher

accuracy requires sending more bits and thus needs more power.

Finally conclude that indifferent data aggregation approaches in wireless sensor networks which are focusing on optimizing important performance measures such as network lifetime, data latency, data accuracy and energy consumption LEACH protocol has been selected. It achieves greater network lifetime and also consumes less energy.

After data aggregation in network we can achieve security in terms of data integrity, confidentiality and freshness. This can be done in future by using one of the key management techniques of WSN.

REFERENCES

[1] Inbo Sim and Jaiyong Lee, "Performance Analysis According to the change of Cluster Size in Large Scale Wireless Sensor Networks", IJCSNS International Journal of Computer Science and Network Security, VOL.9 No.4, April 2009.

[2] Ataul Bari, Arunita Jaekel, Subir Bandyopadhyay, "Clustering strategies for improving the lifetime of two-tiered sensor networks", Computer Communications 31 (2008) 3451–3459.

[3] Stanislava Soro, Wendi B. Heinzelman, "Cluster head election techniques for coverage preservation in wireless sensor networks", Ad Hoc Networks 7 (2009) 955–972.

[4] Pranay Tiwari, "Data Aggregation in Cluster-based Wireless Sensor Networks", M.Tech. - Thesis Indian Institute of Information Technology, Allahabad, July 2008.

[5]Suat Ozdemir, Yang Xiao b, "Secure data aggregation in wireless sensor networks: A comprehensive overview", Elsevier B.V, computer netwrks,2009.

[6] D. J. Dechene, A. El Jardali, M. Luccini, and A. Sauer," A Survey of Clustering Algorithms for Wireless Sensor Networks", the University Of Western Ontario London, Ontario, Canada.

[7] Ossama Younis, Sonia Fahmy, "An Experimental Study of Routing and Data Aggregation in Sensor Networks", University of Arizona, Tucson, AZ 85721, USA.

[8] Vinay Kumar1, Sanjeev Jain and Sudarshan Tiwari, Energy E_cient Clustering Algorithms in Wireless Sensor Networks: A Survey, IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 5, No 2, September 2011

[9] Arati Manjeshwar and Dharma P. Agrawal. TEEN: A Routing Protocol for Enhanced E_ciency in Wireless Sensor Networks In Proc, IEEE 2001

[10]Harneet Kour and Ajay K. Sharma. Hybrid Energy E_cient Distributed Protocol for Heterogeneous Wireless Sensor Network. International Journal of Computer Applications (0975 8887) Volume 4 No.6, July 2010

[11]Stephanie Lindsey and Cauligi S. Raghavendra. PEGASIS: Power-E_cient Gathering in Sensor Information Systems, Computer Systems Research Department The Aerospace Corporation, April 2003.

[12]Nalin VimalKumar Subramanian, Survey on Energy-Aware Routing and Routing Protocols for Sensor, Department of Computer Science University of North Carolina Charlotte, North Carolina, USA

[13]Dr. Kemal Akkaya, Routing Protocols for Sensor Networks Hierarchical and Location-based and QoS Protocols, Department of Computer Science Southern Illinois University Carbondale

[14]Wendi Rabiner Heinzelman, Anantha Chandrakasan, and Hari Balakrishnan Energy-E_cient Communication Protocol forWireless Microsensor Networks, 33rd Hawaii International Conference on System Sciences – 2000

[15]Suraj Sharma and Sanjay Kumar Jena "A Survey on Secure Hierarchical Routing Protocols in Wireless Sensor Networks" ICCCS'11 February 12-14, 2011, Rourkela, Odisha, India.

[16]Kiran Maraiya, Kamal Kant, Nitin Gupta "Wireless Sensor Network: A Review on Data Aggregation" International Journal of Scientific & Engineering Research Volume 2, Issue 4, April -2011 ISSN 2229-5518.

