An Efficient and Secured Data gathering using Data Aggregation Technique in Wireless Sensor Network

Siddhartha Choubey[1], Abha Choubey[2] ,Anil Magendra[3], Ashay Rajimwale[4]

siddhartha00@rediffmail.com , niceabha1@rediffmail.com , anil.magene@gmail.com , ashay.rajimwale@gmail.com,

[1] *Reader, CSE Dept, SSCET, Bhilai*

[2] *Sr. Lecturer, CSE Dept, SSCET, Bhilai*

[3]*ISTE member , CSE, SSCET, Bhilai*

[4]*ISTE member,CSE , SSCET , Bhilai*

Abstract- **In wireless sensor network (WSN) security issue, Data confidentiality, integrity, and elimination of data redundancy becomes vital, when the sensor network is deployed in a hostile environment. In this hostile environment there is requirement of efficient and secured data gathering of sensed data and forward that data to the required users.**

In this paper our main focus is to achieve data confidentiality and integrity and to eliminate data redundancy using Data Aggregation technique .We adopt a Data Aggregation technology where an Aggregator nodes can compute the sum, average, minimum or maximum of the data from its children sensors, and send the aggregation results to a higher-level aggregator .In this way this method gives larger latency to the transmission of network's data and ultimately affect the accuracy of network 's efficiency and thus prolong the lifetime of network. This paper presents a Data Aggregation technology configuring the aggregator node's timeout by timing control scheme which is to achieve a good trade-off between energy efficiency and Data accuracy.

Key Words**: Aggregator nodes, Data confidentiality and integrity, Data Redundancy.**

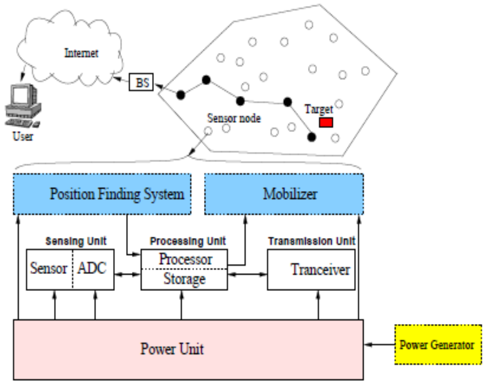
I. INTRODUCTION

What is a Wireless Sensor Network (WSN): A wireless sensor network is a network of multiple sensing nodes that can perform sensing, computation and communication among different sensor nodes and their respective base stations. The network can consist of any number of sensing nodes, and each sensor node has the ability to store and send information across the network.

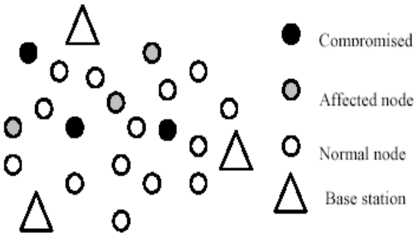
How and why WSN's are used: Wireless Sensor Networks (WSNs) are comprised of many small and resource constrained sensor nodes [8] that are deployed in an environment to gather sensed data and forward that data to interested legal users. Advances in micro-electro-mechanical systems (MEMS) technology allow sensors to be reprogrammable, self localizing, and to support low-energy [2], wireless, multi-hop networking, while requiring only minimal pre-configuration.

To support the reliability of coordinated control, management, and reporting functions, the sensor networks are self organizing with both decentralized control and autonomous sensor behavior, resulting in a sophisticated processing capability.

- The sensing nodes have the ability to communicate with each other and collect information about the area of interest.
- The information can be stored in a special node called the sink node, or it can be sent to a neighbor node (a node with short distance).



(Fig. 1 Components of sensor nodes & wireless sensor network)



(Fig 2. In a sensor network, compromised nodes spoof, inject, modify, or represents false identity to affect normal sensor node to collect sensed data.)

Significance of Data Aggregation technique in WSN:

In wireless sensor network (WSN) security issue, Data confidentiality [1], integrity, and elimination of data redundancy are major requirements especially when the sensor network is deployed in a hostile environment. In that situation Data Aggregation technique helps to cope up with these issues. Data aggregation is a widely used technique in wireless sensor networks. Data aggregation can reduce the number of data packets transmitted and the data conflict, thus raise the data accuracy and data collection efficiency through dealing with the redundant data in-network.

Applications of WSN:

WSN have broad applications in either controlled environments (such as home, office, warehouse, etc) or uncontrolled environments (such as hostile or disaster areas, toxic regions, etc). Some important applications are:

- Area monitoring: gathering information from a region where it is located. Generally data like heat, pressure, sound, light, vibration, electromagnetic field etc.
- Environmental monitoring: measurement of temperature, rainfall etc.

II Network model and assumptions:

We assume a wireless area network which consists of numbers of sensor nodes deployed in a region where security is main issue. Each sensor node can communicate with each other and has a communication range such that if the distance between two sensors is more than this range, they can not communicate. We also assume that the communications channels are bidirectional, i.e. if a node x can receive a message from y, and then it can also send a message to *y*.

We consider a network model in which the nodes in the WSN can be divided into four sets:

S: set of sensing nodes which sense their environment

A: set of aggregator nodes which combine the sensing values from **S** by aggregation functions [6]
F: set of forwarders which transfer the aggregation results from **A** towards **R** hop by hop.

R: is the set of readers of the WSN, which may be base stations (BS), or merely the sinks which provide an access to the outside for the WSN.

It is assumed that values of **S, A, F** and **R** may change over time and their intersections may not be null. This network can represent 2 types of network as defined in [9]

1. Hierarchical network (HWSN): where nodes are deployed hierarchically according to their Capabilities. The whole network is composed of base stations (**R**), cluster heads (**A** *and* **F***)* and sensor nodes (**S**).
2. Distributed network (DWSN): In which nodes are deployed randomly in the environment. After nodes are deployed, a transmission structure should be constructed to aggregate data

III Achieving Data Aggregation technology in WSN

Data Aggregation technique – an overview: In this section we describe an overview of data aggregation technique and its methods. Aggregator node plays important role in establishing secure and accurate exchange of data. An aggregator node can compute the sum, average, minimum or maximum of the data from its children sensors, and send the aggregation results to a higher-level aggregator. WSN can choose its aggregators dynamically according to their power remnant to optimize the total power consumption of the aggregation.

Generally, two methods can be used for secure data aggregation in WSN:

1. Hop by hop encrypted data aggregation
2. End to end encrypted data aggregation

Hop by hop encrypted data aggregation [4]**:** data is encrypted by the sensing nodes and decrypted by aggregator nodes. The aggregator nodes then aggregate the data and encrypt the aggregation result again. At last the sink node gets the final encrypted aggregation result and decrypts it.

End to end encrypted data aggregation**:** In this type of data aggregation intermediate nodes haven't decryption keys and they can only do aggregations on the encrypted data

Data aggregation using hop by hop technique:

In this section we present an idea to secure data aggregation by using hop by hop data encryption. This technique is implemented by using these steps:

- There is bootstrapping defined in [5] which secures links among the nodes.
- Aggregating the data should be done inside the network.
- After aggregation of data there should be authentication of data to be done to achieve integrity of aggregation results.

(i) Bootstrapping: It helps to establish a secure communication infrastructure from a collection of sensor nodes which may have been initialized with some secret information but have had no prior direct contact with each other. The bootstrapping of hop-by-hop encryption can be realized by two methods:

1) Pair-wise key distribution among each pair of sensor nodes; (its used in distributed network) as mentioned in [4] 2) Group-wise key distribution among a cluster of sensor nodes.(its used in hierarchical network)In pair wise key distribution: keys are stored in sensors before sensors are deployed. After the deployment, each sensor establishes a secret link with its neighbour using a common pair-wise key which has been stored in it. Key connectivity, the probability of one sensor

node finds a common key with its neighbour, is an important factor to be considered in the pair-wise key distribution schemes.

In random key distribution key each sensor node receives a random subset of **k** keys from a large key pool of **K** keys. The probability of key share among two sensor nodes is: $[((K\text{-}k)!)^2] \div [((K\text{-}2k)!K!)]$ mentioned in [3]

With this calculation we come to know the sharing of keys among a pair of nodes which will help to ensure enhanced infrastructure for communication among different nodes.

In group wise key distribution scheme mainly used in hierarchical network in which 2 type of distribution can be done:

1) Symmetric group-wise key distribution: a symmetric key can be generated among *t* nodes by evaluating a symmetric multivariate polynomial p(x1 …xt) at each node.

2) Asymmetric group-wise key distribution: the memory of each sensor node is preloaded with the ECC (elliptic curve cryptography) domain parameters. After deployment, each sensor will compute its EC-public/private key pair and broadcast its public key to all nodes within the cluster.

(ii) Data Integrity: we introduce a concept which will ensure integrity of data in network as mentioned in [7]. According to this concept each node (e.g., node A) is initialized before deployment with a symmetric pair-wise key, e.g., *K (AS)*, shared with the base station S. In the network in *i*-th data *transmission* phase, a leaf node *A* computes a temporary key $K^{i}_{AS}(=E(K_{AS}, i))$ based on K_{AS}

- it sends its data reading R_A, node id ID_A and message authentication code $MAC(K^{i}_{AS}, R_A)$ on *RA* to its parent.
- The parent node *B* calculates the aggregation of its children nodes readings, sends the result *Aggr*, node id ID_B and message authentication code $MAC\ (K^{i}_{BS}, Aggr)$ on *Aggr* to its parent.
- The final aggregation and its *MAC* is sent to the base station.

In the *data validation* phase, the base station verifies the final aggregation, and broadcasts the temporary keys (K^{i}_{AS}, K^{i}_{BS}, ...). *i AS,RA*) Using these pair-wise keys, the intermediate aggregation results can be verified by the intermediate aggregators.

Data Aggregation using End to end data aggregation technique:

Hop-by-hop encrypted data aggregation leaves aggregator nodes vulnerable to attacks because the sensor readings will be decrypted on those aggregators. But End-to-end encrypted data aggregation is an alternative to address this vulnerability issue.

It provides end-to-end privacy between sensor nodes and the sink. The aggregators aggregate the encrypted sensor readings without decrypting them, so the end-to-end privacy should be realized by homomorphic cryptosystems.

A. Network wise key distribution: End-to-end privacy needs to establish a network-wise key between the sink and all the sensor nodes. According to [10] and [11] It includes:

B. Master key based solution: In this technique sensor nodes S_i ($1 \leq i \leq n$) sends their encrypted readings $E_k(R_i)$ to the aggregator node and then it calculates the encrypted aggregation $E_K(f(R_1,$

...,R_n)) based on $E_K(Ri)$ and sends it to the sink. The sink decrypts $E_K(f(R_1,...,R_n))$ and gets the aggregation result.

C. public key based solution: in this technique each sensor node uses the public key of the base station to encrypt its reading employing some homomorphic public key encryption scheme.

D. Data integrity: there isn't any aggregator node inside the WSN. Each sensor node sends its reading to the sink using end-to-end encryption. The sink employs truncation and trimming on the readings to achieve robust aggregation result against spoofed sensor readings. But when the network size is very large, the communication cost will be very high for the transmission of all sensor readings to the sink.

IV. Framework design for data aggregation technique: We present two general frameworks for two cases respectively:

- Framework for end to end encrypted data aggregation(it has higher computation cost than hop by hop and more secured)
- Framework for hop by hop encrypted data aggregation

A. *For hop by hop encrypted data aggregation:*

Bootstrapping phase: for controlled environment HWSN is to be constructed and group wise key would be generated, whereas in uncontrolled environment DWSN and pair wise key would be used.

Aggregator selection phase: The sink or base station can select aggregators to construct a transmission structure with minimum energy cost.

Data aggregation phase: n is the number of the children of an aggregator A, the children nodes Si ($1 \leq i \leq n$) encrypt their readings xi as $E_{K, Si, A}(xi)$, and sends it to A. $K (Si, A)$ is the pair-wise key between A and $S_{i.}$

Data transmission phase: Each aggregator encrypts its aggregation result and sends it to the upper level aggregator. The upper level aggregator decrypts the aggregation results and aggregates them as a new aggregation results.

B. Framework for End to end encrypted data aggregation: In this bootstrapping phase as well as aggregator selection phase are same except end of the bootstrapping phase the sink broadcasts a network-wise public key *K*.

- In Data aggregation phase S_i sends the encrypted reading $E_K(xi)$ to its aggregator A. A calculates the aggregation result $E_K(f(x1, ..., xn))$ based on $E_K(xi)$ for $1 \leq i \leq n$.
- In Data integrity verification phase The sink checks whether the commitment is the hash of all $E'_{K,Si,R}(K)$ If it's right, the sink decrypts all $E'_{K,Si,R}(x_i)$ to check whether the final aggregation result is right.

V. Discussion

We have achieved data confidentiality and integrity with above methods which is independent of network whether its hierarchical type (HWSN) or distributed one (DWSN).We have also proposed two general frameworks for

1. Hop by hop data aggregation
2. End to end data aggregation separately which will ensure security of data in the network.

In both frameworks, data confidentiality can be protected by the keys established in the bootstrapping phase. For pair-wise keys and group-wise keys, the compromising of a small number of nodes won't lead to the compromising of the whole network because they are only partially used. The network-wise key in end-to-end encryption is safe because it's a public key. In Framework 2, public key encryption scheme is used, so it's less efficient than Framework 1. However, in Framework 1 sensor readings are decrypted before they are aggregated, so the compromising of an aggregator will make the adversary easy to read the sensor readings and aggregation result.

VI .Conclusion

Our work described security issues and data integrity which is major requirement in wireless sensor network. We studied Data aggregation technique and classify it into 2 major criteria: Hop by hop and End to end encrypted data aggregation. For every case we summarize various techniques for protecting data confidentiality and data integrity.

We also present frameworks for two cases i.e. framework for Hop by hop and for End to end data encrypted data aggregation which supports the security issue. Framework designed for Hop by hop data encrypted model is more efficient than End to end data encrypted model. There are some issues which is to be improved like in Hop by hop data model the sensor readings may be leaked to adversary if the aggregator is compromised. With this work we tried to achieve robust and efficient data gathering using Data aggregation technology and also ensure security and integrity of data in wireless sensor network.

VII References

[1] W. Lou, W. Liu, Y. Fang, "SPREAD: Enhancing Data Confidentiality in Mobile Ad Hoc Networks", *IEEE INFOCOM 2004*, 2004.

[2] A. Perrig, R. Szewczyk, V. Wen, D. Cullar, and J. D. Tygar, "Spins: Security protocols for sensor networks", in *Proceedings of the 7th Annual ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom)*, Rome, Italy, July 2001, pp.189C199

[3] L. Eschenauer and V. Gligor, "A Key Management Scheme for Distributed Sensor Networks", In *ACM CCS 2002*, Washington DC, 2002.

[4] S. A. Camtepe and B. Yener, "Key Distribution Mechanisms for Wireless Sensor Networks: a Survey", Rensselaer Polytechnic Institute, technical report TR-05-07, March 2005.

[5] *H. Chan, A. Perrig, and D. Song, "Random Key Predistribution Schemes for Sensor networks", in Proc.of the IEEE Security and Privacy Symposim 2003, 2003.*

[6] B. Krishnamachari, D. Estrin, S. Wicker, "The Impact of Data Aggregation in Wireless Sensor Networks", in *Proceedings of the 22nd International Conference on Distributed Computing Systems (ICDCS)*, Pages 575 – 578.

[7] D. Wagner, "Resilient aggregation in sensor networks", in *SASN '04: Proceedings of the 2nd ACM workshop on Security of ad hoc and sensor networks*, pages 78-87. ACM Press, 2004.

[8] L. Hu and D. Evans, "Secure aggregation for wireless networks", In *Workshop on Security and Assurance in Ad hoc Networks*, Jan 2003.

[9] K. Yuen, B. Li, B. Liang, "Distributed Minimum Energy Data Gathering and Aggregation in Sensor Networks", in *IEEE International Conference on Communications (ICC 2006).*

[10] C. Castelluccia, E. Mykletun and G. Tsudik, "Efficient Aggregation of Encrypted Data in Wireless Sensor Networks", *ACM/IEEE Mobiquitous Conference*,July 2005, San Diego,USA

[11] J. Girao, D. Westhoff, and M. Schneider, "CDA: Concealed Data Aggregation for Reverse Multicast Traffic in Wireless Sensor Networks", *40th International Conference on Communications, IEEE ICC2005*, May 2005, Korea