Secure End to End Data Aggregation using Public Key Encryption in Wireless Sensor Network

Kaushal J. Patel
M.E. Student
Department of Information Technology
GCET, V V Nagar, Affiliated to G.T.U.
Gujarat, India

Nirav M. Raja
Assistant Professor
Department of Information Technology
GCET, V V Nagar, Affiliated to G.T.U.
Gujarat, India

ABSTRACT
Wireless Sensor Networks is made of a large number of low-cost and low power sensor nodes that contains strictly limited sensing, computation, as well as communication abilities. Because of resource restricted sensor nodes, it is very important to reduce the amount of data transmission so the average sensor lifespan and also the total bandwidth utilization are generally increased. Data aggregation is the process of summarizing and merging sensor data to be able to minimize the number of data transmission in the network. Cluster based Data Aggregation approach, the data sensed by the sensor nodes are transferred to the cluster head and in the cluster head performs data aggregation and forwards results to the sink. Data aggregation could conserve energy (power) as well as bandwidth of the networks. As wireless sensor networks are generally used in remote and hostile environments in order to transfer very sensitive information, sensor nodes are susceptible to node compromise attacks and security problems such as data confidentiality and integrity are extremely crucial. Therefore, Secure Data aggregation protocol, must be designed with security in mind and investigates the relationship between security and data aggregation process within wireless sensor networks. In this paper, the secure data aggregation schemes are categorized into hop by hop aggregation and end to end aggregation. Here we explore a Homomorphic aggregation system based on a public key encryption (PKE) scheme to protect sensor data secure is proposed. Security analysis shows that our proposed protocol can guarantee end-to-end confidentiality and privacy.

General Terms
Secure Data Aggregation, WSN, Privacy Homomorphism(PH).

Keywords
Wireless Sensor Network, Data Aggregation, Data Confidentiality, Data Integrity, Public Key Encryption, Homomorphic Encryption.

1. INTRODUCTION
Wireless sensor network is really a well-known area for research now days, because of huge probable use of sensor networks in several areas. A sensor network is a consist of sensing, processing, communication capability which really helps to observe, instrument, respond to events and phenomena in a specified environment. This type of network allows connecting the physical world to environment. By networking small sensor nodes, it becomes easy to get the data about physical phenomena which was very much difficult with conventional ways. Wireless sensor network typically consist of tens to thousands of nodes. These nodes obtain process and cooperatively pass this collected info to a main location.

Wireless Sensor Network (WSN) is a collection of large number of sensor nodes that are deployed in a particular region. A wireless sensor network is a distributed system interacting with physical environment to measure the surrounding environment (e.g. temperature, pressure) and sensitive information (e.g. movement, target tracking). WSN have wide range of application like Environmental monitoring, Military application, Home and Industry monitoring etc[1].

Wireless Sensors are equipped with limited range of sensing, computational, storage and communication resources [2].Extensive utilization of computational, communication resources can potentially reduce the battery life of a Wireless Sensor[2].Life time of a WSN depends on the Life time of Sensor nodes[2].

Data Aggregation is defined as the process of aggregating the data from multiple sensors to eliminate redundant transmission and provide combined information to the base station[2].Data aggregation can significantly reduce the amount of data transmitted to the base station, therefore improve the energy efficiency and prolong the wireless network lifetime[6]. The main concept of data aggregation is to aggregate multiple sensing data by performing Effective type of data aggregation functions like max, min, average, duplicate suppression. Such functions can be performed in each sensor node[2][5].

Data aggregation protocols are divided into tree based data aggregation protocols and cluster based data aggregation protocols based mainly on the topology used for data aggregation[12]. Cluster based data aggregation protocols reduces the latency in the tree-based data aggregation by grouping the nodes in WSNs into clusters[12].This process is called clustering. In cluster based protocols, cluster head performs data aggregation and parent nodes in the path to the base station perform data aggregation in tree based data aggregation protocols[2].

2. LEACH PROTOCOL
One well-known data-aggregation algorithm for WSN is Low Energy Adaptive Clustering Hierarchy Aggregation (LEACH) algorithm[3]. It is a distributed cluster-based aggregation algorithm. LEACH has recorded good results in increasing the network lifetime. The scenario of data aggregation in the cluster based network of LEACH protocol is shown in Figure.1. The LEACH protocol is distributed and sensor nodes organize themselves into clusters for data fusion.

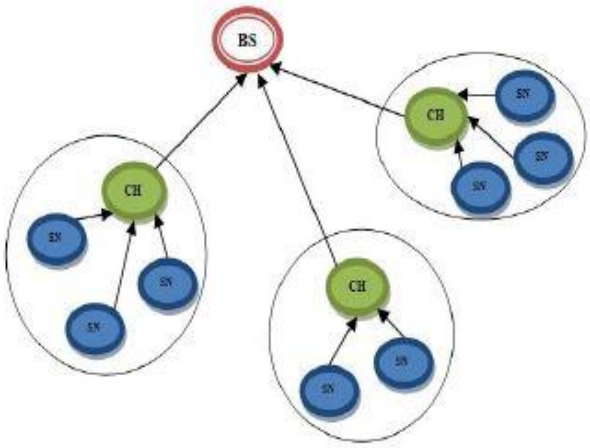

Figure 1 LEACH protocol

A node is designated as the cluster head (determined by the probability of becoming cluster head) in each cluster, transmitting the data received from the sensors in its cluster to the sink.
The probability is given by

$$P_{CH=C*\frac{E_{residual}}{E_{max}}}$$

Where *C* denotes the initial percentage of cluster heads (specified by the user). Residual is the estimated current residual energy of the node. E_{max} is its initial energy corresponding to a fully charged battery.

The idea is to form cluster of sensor nodes based on signal strength and use the cluster-head as a router to forward data of other nodes in cluster to the base station. The data processing is performed at cluster-heads. In this protocol, nodes are classified into two categories: CHs and SNs. The nodes are organized into local clusters and the communication process is divided into rounds. A dedicated node selected as CH is responsible for creating and manipulating a TDMA slots and aggregating the data coming from different nodes and sending it to the BS.
LEACH protocol works in rounds. Each round is divided into two phases:

- Setup phase
- Steady phase

In Setup phase, at the beginning of the round, each node decides independently of other nodes whether or not to become a cluster head for current round. Each sensor node generates a random number such that 0< random < 1 and compares it to a pre-defined threshold T (n). If random < T (n), the sensor node becomes cluster-head in that round, otherwise it is cluster member. The threshold is given T (n) below:

$$\mathbf{T(n) = \frac{P}{1 - P\left(r \bmod \left(\frac{1}{P}\right)\right)} \quad if\ n \in G}$$

Where,
P is the probability of the node being selected as a cluster-head node. r is the number of rounds of selection. G is the set of nodes that haven't been cluster-heads in the last 1/P rounds mod denotes modulo operator[1][3].

Nodes that are cluster heads in round r shall not be selected in the next 1/P rounds. After CH selection, the CH will broadcast an advertisement message using CSMA MAC protocol to its neighbors that it is the new cluster-head. The nodes will send the join-request message containing their IDs by using CSMA (carrier sensing multiple access) to join a cluster from which they receive strongest strength signal. After that, each CH knows its own cluster members information. The CH node sets up a TDMA schedule for data transmission coordination within cluster and broadcast it to its cluster members.[3] The TDMA schedule prevents collision among data messages and conserves energy among non-cluster head nodes. So all the member nodes know their TDMA slots, and then the steady-state phase begins.[3]

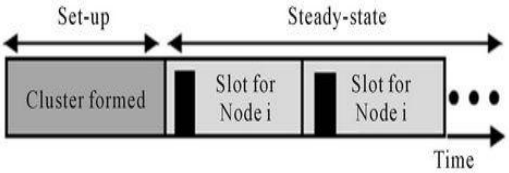

Figure 2 Period of LEACH

In the steady-state phase, cluster members sense the surroundings and transmit the sensed data to their CH depending on the TDMA schedule received at the setup phase. SNs go into sleep mode to save energy for other slots. When the CH receives all the data sent by its cluster members, it will aggregate them and then send the aggregated data to BS. After a certain time, the network goes back into the setup phase again and enters another round of selecting new CH.

3. SECURE DATA AGGREGATION IN WSN

In WSN security issues, data confidentiality and integrity becomes an important in data aggregation in Wireless Sensor Networks. In this paper, the secure data aggregation schemes are categorized into hop by hop aggregation and end to end aggregation.

3.1 Hop by Hop Secure Data Aggregation

In hop by hop data aggregation sensor nodes sensed data and encrypted this data and send to aggregator node. Aggregator node decrypt the data and use aggregation function and aggregate data from various node then encrypt it again and send to Base station. Every time CH should decrypt data then perform aggregation and transmit result to base station in Encrypted form, it will very time and power consuming and attacker or adversary get a chance to forge our data.

3.2 End to End Secure Data Aggregation

In comparison with hop by hop data aggregation, end to end data aggregation comparatively flexible structure and routing protocol [6]. In end to end data aggregation sensor node sensed data and encrypted this data then send to the aggregation node. The aggregator directly aggregated the encrypted sensor data without the encryption keys of each sensor nodes. so, the sensor data provide end to end confidentiality and integrity. This two security requirements provide end to end data privacy. Secure end-to-end data aggregation has higher computation cost on the sensor nodes[5], but achieves stronger security, in comparison with the framework for Secure hop by hop data aggregation[5].

Table 1: Comparison of Secure Data Aggregation Methods

Parameters	Hop by Hop Data Aggregation	End to End Data Aggregation
End to End Privacy	No	Yes
Aggregation performed on	Plain Sensor data	Encrypted Sensor Data
Computational Cost	Low	High
Memory requirement	High	Low
Vulnerable to attack	More to passive attack	More to active attack
Energy Consumption	High	Low
Data Confidentiality	Lesser security	High security

4. SECURITY THREATS AND ISSUES IN WIRELESS SENSOR NETWORK

Previous work in data aggregation assumes that every mote is honest and only transmits their correct readings. Shih-I Huang, Shiuhpyng Shieh [14] proposed a data-centric diffusion method to aggregate data. Their method enables diffusion to achieve energy savings by selecting empirically good paths and by caching and processing data in-network. Though their method can achieve significant energy savings, security is not put into consideration in their design.

Sanjeev Setia, Sankardas Roy and Sushil Jajodia. [15] further examined the problem that a single compromised sensor mote can render the networks useless, or worse, misleads the operator into trusting a false reading. They proposed an aggregation protocol that is resilient to both intruder devices and single device key compromises, but their scheme suffers a problem that the aggregated data will be expanded every time when it was aggregated and forwarded by any intermediate sensor mote using some heuristic rules.

Suat Ozdemir[10] proposed a secure information aggregation protocol to answer queries over the data acquired by the sensors. In particular, their proposed protocols are designed especially for secure computation of the median and the average of the measurements, for the estimation of the network size and for finding the minimum and maximum sensor reading. Even though their scheme provided data authentication to provide secrecy, the data is still delivered in plaintext format which provides no privacy during transmission.

Wenbo He, Hoang Nguyen, Xue Liu [16] presented a paper studying related attacks on data aggregation in sensor networks. He thoroughly examined current aggregation functions and proved that these aggregation functions are vulnerable and insecure under several attacks. He also proposed a theoretical framework for evaluating data aggregation resiliently in sensor networks and in its security against these attacks. Still privacy is not guaranteed in his scheme.

Without encryption, adversaries can monitor and inject false data into the network. Encryption can solve this problem, but it is require some approach to transmit data in encrypted form, received data encrypted form and only end points can decrypt data .Intermediate nodes can perform operation only on encrypted data.

Adversaries can use the following attacks:

- Adversaries can deploy sensors near existing sensors to determine their likely value.
- Adversaries can use common key encryption systems (which always encrypt common sensor data in the same way) to see when two readings are identical. By using nearby sensors under the adversaries' control, adversaries can conduct a known-plaintext attack.
- Adversaries can tamper with sensors to force them to predetermined values (such as heating a temperature sensor) and thus conduct a chosen-plaintext attack.
- Adversaries can inject false readings or resend logged readings from legitimate sensor motes to manipulate the data aggregation process, conducting a man-in-the middle attack.
- Eavesdropping and Stealthy attack also used by attacker.

Table 2 presents encryption policies, possible attacks, and vulnerabilities in data aggregation schemes.
First, we provide a lightweight data aggregation mechanism which protects data when data are processed in aggregators. Aggregators can help to eliminate redundant data without decrypting data. Thus, aggregators do not need to spend extra power in data decryption, and more network lifetime can be guaranteed. Second, our proposed scheme is resilient to known-plaintext attacks, chose plaintext attacks, cipher text-only attacks, and man-in the middle attacks.

In secure data aggregation require encryption technology in which directly performs data aggregation on encrypted data. CH directly perform data aggregation on received data and transform result to the sink so there is no need to decrypt data

Table 2: Encryption policies, attacks and vulnerabilities in data aggregation schemes [21][22]

Encryption policy	Possible attacks	Confidentiality	Privacy	Data aggregation
Sensors transmit readings without encryption[19][21]	Man-in-the-middle Eavesdropping	No	No	Generating wrong aggregated results[21]
Sensors transmit encrypted readings with permanent keys[19][21]	Known-plaintext attack Chosen-plaintext attack Man-in-the-middle	Yes	No	Data aggregation cannot be achieved when data are encrypted unless the aggregator has encryption keys[21][22]
Sensors transmit encrypted readings with dynamic keys[19][21]	None of above	Yes	Yes	End to End Secure Data Aggregation[21]

to perform aggregation again and encrypt and forward result to the sink.

5. PROPOSED SCHEME

It is an encryption transformation that allows direct computation on encrypted data's. Privacy homomorphic encryption achieves both end to end confidentiality and data aggregation. Additive PH and multiplicative PH are the two variations of privacy homomorphism[12].

If an encryption algorithm E() is said to be additive homomorphic, then it support additive operations on encrypted data without the decryption of individual data's. ie. E(x+y) =E(x) +E(y).It is more suitable in wireless sensor network due to their less expensive operation than multiplicative PH[12].

If an encryption algorithm E() is said to be multiplicative homomorphic, then it support multiplicative operations on encrypted data without the decryption of individual data's. ie. E(x*y) =E(x) *E(y)[12].

Cryptographic algorithms that support privacy homomorphism are divided into two. These are Symmetric PH [10], [28] and Asymmetric PH/ public key encryption.

Table 3: Difference between Symmetric PH and Asymmetric PH [12]

Parameters	Symmetric PH	Asymmetric PH
Encryption and Decryption	using same key	using different key
Number of key	Single key	Multiple key
Aggregation Speed	Fast	Slow
Type of key stored by sensor node	Sensor node needs to store secret key	Sensor node needs to store non sensitive public key
Overall system security	Less	Better
Chosen plain text attacks	Insecure	Secure
Node compromise attack	Largely affected	Less affected
Example	Domingo-Ferrer [17],CMT[18]	ECC,RSA

A homomorphic encryption scheme allows arithmetic operations on cipher texts. One example is a multiplicatively homomorphic scheme, where the decryption of the efficient manipulation of two cipher texts yields the multiplication of the two corresponding plaintexts. Homomorphic encryption schemes are especially useful whenever some party not having the decryption key(s) needs to perform arithmetic operations on a set of cipher texts. RSA is homomorhic encryption technique in which there is several steps to perform arithmetic operation. Homomorphic encryption also applies on data aggregation results.

RSA algorithm is applied on sensor reading. During LEACH protocol second round, there are require several modification on transmission of data.

Key Generation Steps of RSA

Step1: Choose two distinct large prime numbers p and q.
Step2: Calculate the value of n.
n=p*q, n will be used as the modulus for both public and private keys
Step3: Find the totient of n, fi(n)
fi(n)=(p-1)*(q-1).
Step4: Choose an e such that 1<e<fi(n).and such that e and fi(n) no divisors other than 1.
gcd (e, fi(n))=1.
Step5: Calculate the value of d based on relation,
de≡1 (mod fi(n))
Step6: keep d is private,
Public key is (e, n) : public key is available to cluster members and CH.
Private key is (d, n) : private key is only available to the sink or base.

Encryption

Step1: Nodes have public key (e, n) [e is public]
Step2: **C= M^e mod(n).**
Step3: Message is encrypted

Decryption

Step1: Base Station has private key (d, n) [d is private]
Step2: **M= C^d mod(n).**
Step3: Message is decrypted

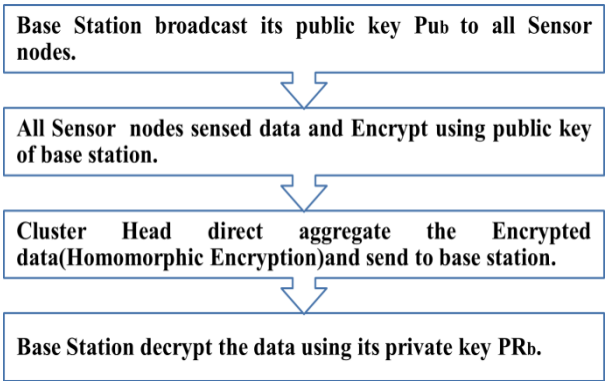

Figure 3 Public Key Encryption in WSN

6. SIMULATION PARAMETERS

Simulations are conducted using Castalia simulator to run topologies & get precise plots. Castalia is Omnet++ platform based simulator.

There are several parameters consider during simulation

Table 4: Simulation parameters

No.	Parameters	Value
1	Routing Protocol	LEACH
2	Number of nodes	100,200…600
3	Node as sink or base station	Node 0
4	Packet Size	2000 bits
5	Election Probability value of CHs (p)	0.05
6	Initial energy per node (E0)	1.0 J

.

In simulation, there is implementation of LEACH protocol then in LEACH protocol apply secure RSA encryption-decryption technique to Modified LEACH protocol. This Modified LEACH protocol provides end to end confidentiality and end to end privacy from sensor nodes to base station.

6.1 Performance Evaluation

Analysis (Evaluation) can be done on the following parameter:

1. **Energy:** A scheme is energy efficient if it maximizes the functionality of the sensor network. This idea is captured by network lifetime which quantifies energy efficiency. As my proposed scheme is prolonging the Lifetime of sensor network.

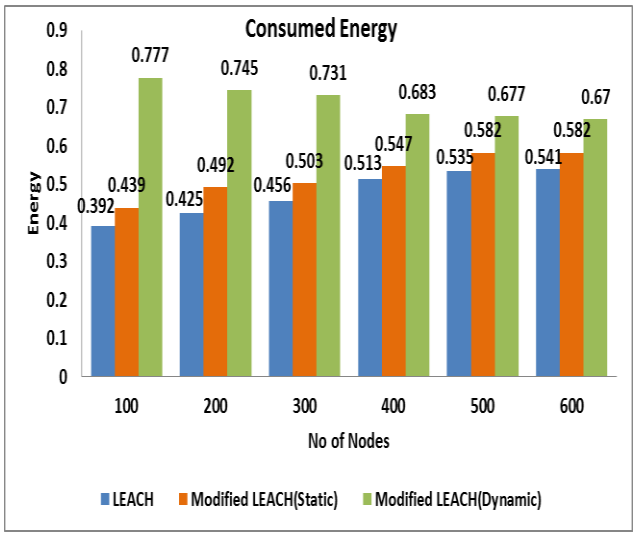

Figure 4 Consumed Energy vs. No. of Nodes

- This graph represents the average consumed energy with different no. of nodes.
- In modified LEACH(Static Environment), Consumed energy increases because in Modified LEACH encryption process takes place at every Sensor Node which consumes more energy than in normal scenario in which sensor nodes just have to send the sensed data as it as.
- In Dynamic Environment nodes are mobile and sensor node have to establish the connection before sending data. So, it consumes more energy than simple LEACH and Modified LEACH (Static).
- In Dynamic Environment, when no of nodes are increased the energy consumption per node decreases as there may be more than one paths are available for sending data to base station.

2. **Latency** can be measured as the time taken from sending the data generated at the source nodes and the data packets received at the BS.

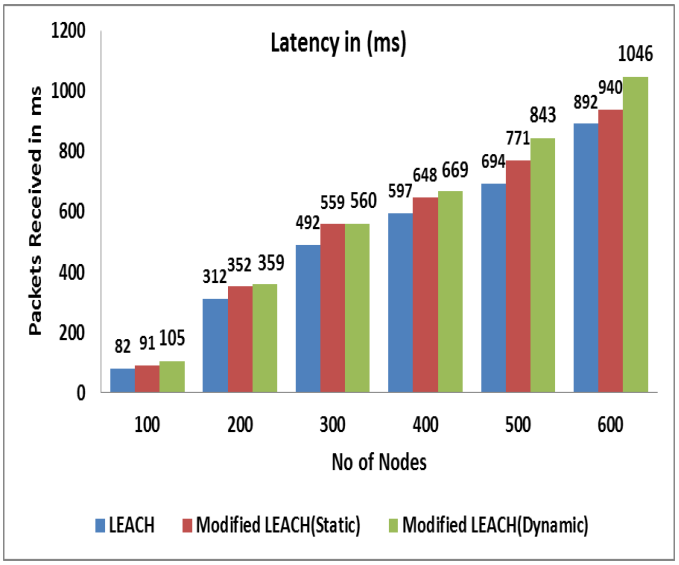

Figure 5 Latency vs. No. of Nodes

- This graph represents the average Latency of network when different no. of nodes taken.
- In modified LEACH(Static topology), Delay increases as every sensor node takes more time to first encrypt and then send data and decryption at Base station requires much time than general LEACH protocol.
- In Dynamic environment, Latency is even much higher than Static environment because in dynamic environment it takes time to find the current location of node and then sending message.

3. **Packet Delivery Ratio:** (∑ Number of packet receive / ∑ Number of packet send): the ratio between the numbers of data packets delivered at receiver and no. of packets sent by the sender node. This illustrates the level of delivered data to the destination.
   - This graph represents the packet delivery ratio when different no. of nodes are employed in network.
   - In modified LEACH(Static Environment), PDR slightly increases in modified LEACH as attacker can't forged the encrypted packet, so packet are almost delivered in its original form.
   - In Dynamic environment, the packet delivery ratio slightly decreases than Static environment because mobility of nodes. So, both environments provide better PDR than general LEACH.

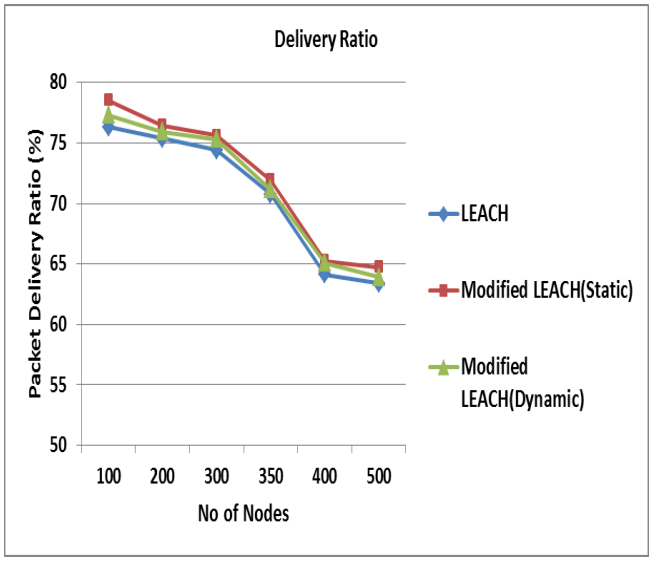

Figure 6 Packet Delivery Ratio vs. No. of Nodes

7. CONCLUSION

In WSN, It is required to find routing protocol that perform data aggregation in efficient way. There are several approaches to perform data aggregation. But Cluster based approach is best in all of existing approaches. In this approach, Network life time is increase and less number of transmissions performs. LEACH is mostly used as cluster based routing protocol in WSN. There are several possible attacks in WSN which can disturb aggregation process. So, there is much requirement of secure channel for transmission of data. For that, all the transmission performs in encrypted form and decrypt at base station. Modified LEACH routing protocol based on security parameter. Data transmissions are perform in secure form and prevent from attacks. End to End Secure Data Aggregation provides Confidentiality with energy efficiency of network. We have analyzed the performance of our proposed algorithm under static and dynamic environment. In dynamic environment Energy consumption, latency goes high because of mobility of sensor nodes.

8. ACKNOWLEDGMENTS

We are extremely thankful to entire Information Technology Department, faculty and staff, for helping us in every conceivable way during my course of study for this work.

9. REFERENCES

[1] I. Akyildiz, W. Su, M. Vuran, and E. Cayirci, "A Survey On Sensor Networks", *IEEE Communications Magazine, Volume 40, Number, 2002.*

[2] P.N.Renjith, E. Baburaj, "An Analysis on Data Aggregation in wireless sensor networks", ICRCC © 2012.

[3] Chanjuan Wei, Yanjie Gao, J. Yang, "Cluster-Based Routing protocols in Wireless Sensor Network: A Survey", IEEE © 2011.

[4] Yan-Xiao Li, Lian-Qin, Qian-Liang, "Research On Wireless Sensor Network Security", IEEE © 2010.

[5] Yingpeng Sang, Hong Shen, "Secure Data Aggregation in Wireless Sensor Networks: A Survey", IEEE © 2006.

[6] Jia Guo, Jian'an Fang, Xuemin Chen, "Survey on Secure Data Aggregation for Wireless Sensor Networks", IEEE © 2011.

[7] A.S.Poornima, B.B.Amberker, "SEEDA : Secure End-to-End Data Aggregation in Wireless Sensor Networks", IEEE © 2010.

[8] Hung-Min Sun, Chiung-Hsun Chen, and Po-Chi Li, "A Lightweight Secure Data Aggregation Protocol for Wireless Sensor Networks", IEEE © 2011.

[9] Juan Wei, Shanqing Guo, Qiuliang Xu, "Secure Homomorphic Aggregation Algorithm of Mixed Operations in Wireless Sensor Networks", IEEE © 2011.

[10] Suat Ozdemir, Yang Xiao, "Secure data aggregation in wireless sensor networks: A Comprehensive overview", ELSEVIER © 2009.

[11] Soufiene Ben Othman, Abdullah Ali Bahattab, Abdelbasset Trad & Habib Youssef, "Confidentiality and Integrity for Data Aggregation in WSN Using Homomorphic Encryption", Springer 2014.

[12] Josna Jose, Joyce Jose, "Asymmetric Concealed Data Aggregation Techniques in Wireless Sensor Networks: A Survey", Modern Education and Computer Science Press 2014.

[13] Soufiene Ben Othman, Abdelbasset Trad, Habib Youssef, Hani Alzaid, "Secure Data Aggregation in Wireless Sensor Networks", IEEE © 2013.

[14] Shih-I Huang, Shiuhpyng Shieh, J. D. Tygar." Secure encrypted-data aggregation for wireless sensor networks", Springer Science 2009.

[15] Sanjeev Setia, Sankardas roy and Sushil jajodia "Secure Data Aggregation in Wireless Sensor Networks" IEEE.

[16] Wenbo He, Hoang Nguyen, Xue Liu, Klara Nahrstedt, Tarek Abdelzaher. "SPDA: Secure and Privacy-preserving Data Aggregation in Wireless Sensor Networks".

[17] T. Okamoto and S. Uchiyama. A New Public-Key Cryptosystem as Secure as Factoring. Proc. Ann. Int'l Conf. Theory and Applications of Cryptographic Techniques (EUROCRYPT '98), 1998, pp. 308-318.

[18] C. Castelluccia, E. Mykletun, and G. Tsudik. Efficient Aggregation of Encrypted Data in Wireless Sensor Networks. Proc. Second Ann. Int'l Conf. Mobile and Ubiquitous Systems, July 2005, pp. 109-117.

[19] Jacques Bahi, Christophe Guyeux, Abdallah Makhoul. Secure Data Aggregation in Wireless Sensor Networks.

[20] Homomorphism versus Watermarking Approach. ADHOCNETS 2010, 2nd Int. Conf. on Ad Hoc Networks, Dec 2009, Canada.

[21] Kaushal J. Patel, Nirav M. Raja," An Overview of Secure Data Aggregation in Wireless Sensor Network" International Journal of Advanced Research in Computer Science and Software Engineering, Volume 5, Issue 1, January 2015.(ISSN 2277 128X).

[22] Shih-I Huang, Shiuhpyng Shieh, J.D.Tygar "Secure Encrypted-data aggregation for wireless sensor network" Springer © 2010.