International Journal of Computer Science Engineering & Technology
APPROVED BY
NATIONAL SCIENCE LIBRARY (NSL)
NATIONAL INSTITUTE OF SCIENCE COMMUNICATION AND INFORMATION RESOURCES (NISCAIR)
COUNCIL OF SCIENTIFIC AND INDUSTRIAL RESEARCH (CSIR)– NEW DELHI INDIA .
ISSN :2455-9091

SHELTERED DATA AGGREGATION IN WIRELESS SENSOR NETWORKS: STRAINING OUT THE ATTACKER'S COLLISION

[1] S. Saravanan, [2] S. Hamsappriya
[1] Assistant Professor, [2] Research Scholar,
[1, 2] Department of Computer Science,
[1, 2] Sengunthar College of Arts and Science,
[1, 2] Tiruchengode,Tamilnadu.

ABSTRACT: Serious security threat is originated by node capture attacks in hierarchical data aggregation where a hacker achieves full control over a sensor node through direct physical access in wireless sensor networks. It makes a high risk of data confidentiality. In this study, we propose a securing node capture attacks for hierarchical data aggregation in wireless sensor networks. Initially network is separated into number of clusters, each cluster is headed by an aggregator and the aggregators are directly connected to sink. The aggregator upon identifying the detecting nodes selects a set of nodes randomly and broadcast a unique value which contains their authentication keys, to the selected set of nodes in first round of data aggregation. When any node within the group needs to transfer the data, it transfers slices of data to other nodes in that group, encrypted by individual authentication keys. Each receiving node decrypts, sums up the slices and transfers the encrypted data to the aggregator. The aggregator aggregates and encrypts the data with the shared secret key of the sink and forwards it to the sink. The set of nodes is reselected with new set of authentication keys in the second round of aggregation.

Key Words: [Wireless Sensor Network (WSN),Cluster-Head (CH), LEACH(Low Energy Adaptive Clustering Hierarchy),TAG(Tiny Aggregation), order- and duplicate-insensitive (ODI), Privacy Preserving Secure In-Network Data Aggregation (PPSDA).]

1. INTRODUCTION

The wireless sensor network is an ad-hoc network. It consists of small light weighted, low powered wireless nodes called sensor nodes, which are shown in with limited memory, computational, and communication resources, it measures physical parameters such as sound, force per unit area, temperature, and humidity. These sensor nodes are envisioned to play an important part in a broad diversity of fields ranging from critical military surveillance applications to forest fire monitoring and building security monitoring in the near future. In these networks, a big bit of sensor nodes are deployed to monitor a huge domain. Withal, the nodes in WSNs have severe resource constraints due to their lack of processing power, limited memory and vitality. Since these networks are commonly deployed in distant offices and left unattended, they should be fitted with security mechanisms to guard against attacks such as node capture, physical tampering, eaves dropping,denial of service, etc. Unfortunately, traditional protection mechanisms with high budget items are not feasible in resource constrained sensor nodes. The researchers in WSN security have proposed various security schemes which are optimized for these networks with resource constraints.Aim of data aggregation protocols

is to combine and summarize data packets of several sensor nodes so that the amount of data transmission is reduced. An example data aggregation WSN is presented in where a group of sensor nodes collects the information from a target region. When the base station queries the network, instead of sending each sensor node‘s data to base station, one of the sensor nodes, called data aggregator, collects the information from its neighboring nodes, aggregates them and sends the aggregated data to the base station over a multihop path.

ISSUES IN DATA AGGREGATION

A sensor network is a special type of ad hoc network. So it shares some common property of traditional networks. The security requirements of a wireless sensor network can be classified as follows:

1. Data Confidentiality: Data confidentiality is the most important issue in network security. Every network with any security focus will typically address this problem first. In sensor networks, the confidentiality relates to the following:

- A sensor network should not leak sensor readings to its neighbors. Especially in a military application, the data stored in the sensor node may be highly sensitive.
- In many applications nodes communicate highly sensitive data, e.g., key distribution; therefore it is extremely important to build a secure channel in a wireless sensor network.
- Public sensor information, such as sensor identities and public keys, should also be encrypted to some extent to protect against traffic analysis attacks.

The standard approach for keeping sensitive data secret is to encrypt the data with a secret key that only intended receivers possess, thus achieving confidentiality.

2. Data Integrity and Freshness: With the implementation of confidentiality, an adversary may be unable to steal information. However, this doesn‘t mean the data is safe. The adversary can change the data, so as to send the sensor network into disarray. For example, a malicious node may add some fragments or manipulate the data within a packet. This new packet can then be sent to the original receiver. Data loss or damage can even occur without the presence of a malicious node due to the harsh communication environment. Thus, Data integrity guarantees that data being transferred is never been corrupted in transit.

3. Source Authentication: Even if confidentiality and data integrity are assured, we also need to ensure the freshness of each message. Informally, data freshness suggests that the data is recent, and it ensures that no old messages have been replayed. This requirement is especially important when there are shared-key strategies employed in the design. Typically shared keys need to be changed over time. However, it takes time for new shared keys to be propagated to the entire network. In this case, it is easy for the adversary to use a replay attack. Also, it is easy to disrupt the normal work of the sensor, if the sensor is unaware of the new key change time. To solve this problem a nonce, or another time-related counter, can be added into the packet to ensure data freshness.

4. Data Availability: Adjusting the traditional encryption algorithms to fit within the wireless sensor network is not free, and will introduce some extra costs. Some approaches choose to modify the code to reuse as much code as possible. Some approaches try to make use of additional communication to achieve the same goal.

What‘s more, some approaches force strict limitations on the data access, or propose an unsuitable scheme (such as a central point scheme) in order to simplify the algorithm. But all these approaches weaken the availability of a sensor and sensor network for the following reasons:

- Additional computation consumes additional energy. If no more energy exists, the data will no longer be available.
- Additional communication also consumes more energy. What‘s more, as

communication increases so too does chance of incurring a communication conflict.

- A single point failure will be introduced if using the central point scheme. This greatly threatens the availability of the network.

The requirement of security not only affects the operation of the network, but also is highly important in maintaining the availability of the whole network.

5. Self-Organization: A wireless sensor network is a typically an ad hoc network, which requires every sensor node be independent and flexible enough to be self-organizing and self-healing according to different situations. There is no fixed infrastructure available for the purpose of network management in a sensor network. This inherent feature brings a great challenge to wireless sensor network security as well. For example, the dynamics of the whole network inhibits the idea of pre-installation of a shared key between the base station and all sensors. Several random key pre-distribution schemes have been proposed in the context of symmetric encryption techniques. In the context of applying public-key cryptography techniques in sensor networks, an efficient mechanism for public-key distribution is necessary as well. In the same way that distributed sensor networks must self-organize to support multi-hop routing, they must also self-organize to conduct key management and building trust relation among sensors. If self-organization is lacking in a sensor network, the damage resulting from an attack or even the hazardous environment may be devastating.

2. RELATED WORK

Ko et al. described three applications that exemplify these problems and the solutions they developed. First, they show how temporal over-sampling can simplify the analysis of a slow process such as the avian nesting cycle. Then, they show how to overcome temporal under-sampling in order to detect birds at a feeder station. Finally, they show how to exploit temporal consistency to reliably detect pollinators as they visit flowers in the field.

Corke et al. concerned with the application of wireless sensor network (WSN) technology to long-duration and large-scale environmental monitoring. The Holy Grail is a system that can be deployed and operated by domain specialists not engineers, but this remains some distance into the future. They present their views as to why this field has progressed less quickly than many envisaged it would over a decade ago. They use real examples taken from their own work in this field to illustrate the technological difficulties and challenges that are entailed in meeting end-user requirements for information gathering systems. Reliability and productivity are key concerns and influence the design choices for system hardware and software.

Madden et al. discussed various generic properties of aggregates, and show how those properties affect the performance of their in network approach. They include a performance study demonstrating the advantages of their approach over traditional centralized, out-of-network methods, and discuss a variety of optimizations for improving the performance and fault tolerance of the basic solution.

Zhao et al. illustrated architecture for sensor network monitoring, then focus on one aspect of this architecture: continuously computing aggregates (sum, average, count) of network properties (loss rates, energy levels etc., packet counts). Their contributions are two-fold. First, they propose a novel tree construction algorithm that enables energy-efficient computation of some classes of aggregates. Second, they show through actual implementation and experiments that wireless communication artifacts in even relatively benign environments can significantly impact the computation of these aggregate properties. In some cases, without careful attention to detail, the relative error in the computed aggregates can be as much as 50%. However,

by carefully discarding links with heavy packet loss and asymmetry, they can improve accuracy by an order of magnitude.

Considine et al. presented new methods for approximately computing duplicate-sensitive aggregates across distributed datasets. An elegant building block which enables their techniques are the duplicate-insensitive sketches of Flajolet and Martin, which give us considerable freedom in their choices of how best to route data and where to compute partial aggregates. In particular, use of this duplicate-insensitive data structure allowed us to make use of dispersity routing methods to provide fault tolerance that would be inappropriate otherwise.

Nath et al. proposed synopsis diffusion, a general framework for achieving significantly more accurate and reliable answers by combining energy-efficient multi-path routing schemes with techniques that avoid double-counting. Synopsis diffusion avoids double-counting through the use of order- and duplicate-insensitive (ODI) synopses that compactly summarize intermediate results during in-network aggregation. They provide a surprisingly simple test that makes it easy to check the correctness of an ODI synopsis. They show that the properties of ODI synopses and synopsis diffusion create implicit acknowledgments of packet delivery. They show that this property can, in turn, enable the system to adapt message routing to dynamic message loss conditions, even in the presence of asymmetric links. Finally, they illustrate, using extensive simulations, the significant robustness, accuracy, and energy-efficiency improvements of synopsis diffusion over previous approaches.

Yang et al demonstrated SDAP, a Secure Hop-by-hop Data Aggregation Protocol for sensor networks. The design of SDAP is based on the principles of divide-and-conquer and commit and attest. First, SDAP uses a novel probabilistic grouping technique to dynamically partition the nodes in a tree topology into multiple logical groups (sub trees) of similar sizes. A commitment based hop-by-hop aggregation is performed in each group to generate a group aggregate. The base station then identifies the suspicious groups based on the set of group aggregates. Finally, each group under suspect participates in an attestation process to prove the correctness of its group aggregate. Their analysis and simulations show that SDAP can achieve the level of efficiency close to an ordinary hop-by-hop aggregation protocol while providing certain assurance on the trustworthiness of the aggregation result. Moreover, SDAP is a general-purpose secure aggregation protocol applicable to multiple aggregation functions.

Yu aimed to enable aggregation queries to tolerate instead of just detecting the adversary. To this end, they propose a novel tree sampling algorithm that directly uses sampling to answer aggregatqueries. It leverages a novel set sampling technique to overcome a key and well-known obstacle in sampling traditional sampling technique is only effective when the predicate count or sum is large. Set sampling can efficiently sample a set of sensors together, and determine whether any sensor in the set satisfies the predicate (but not how many). With set sampling as a building block, tree sampling can provably generate a correct answer despite adversarial interference, while without the drawbacks of traditional sampling techniques.

Roy et al. showed that even if a few compromised nodes contribute false sub-aggregate values, this results in large errors in the aggregate computed at the root of the hierarchy. They present modifications to the aggregation algorithms that guard against such attacks, i.e., they present algorithms for resilient hierarchical data aggregation despite the presence of compromised nodes in the aggregation hierarchy. They evaluate the performance and costs of their approach via both analysis and simulation. Their results show that their approach is scalable and efficient.

PROBLEM DESCRIPTION

EXISTING SYSTEM:

NMS monitors the neighborhood nodes using neighbor list checking approach for detecting packet-dropping attacks. Although most of the existing algorithms use multipath routing approach to mitigate selective forwarding attacks, NMS uses a single path routing mechanism. If the packet encounters a malicious node on its way, it attempts to circle around in an efficient manner and return to the single shortest path. The new route is selected using sending broadcasts in the neighborhood of the malicious node. At neighbor discovery phase, in addition to its one hop neighbors the protocol keeps the information about its two hop neighbors i.e. neighbors of neighbors. In addition, a key is shared between a node and its one-hop neighbors. When a node sends a packet to its neighbor, it keeps a copy of the packet, encrypts it with the shared key and forwards it to the next-hop. Since the key is shared between the node and all its neighbors, all one-hop neighbors can overhear and monitor whether the encrypted packet is forwarded any further or whether it is forwarded intact. This method also relies on node additional overhearing capabilities.

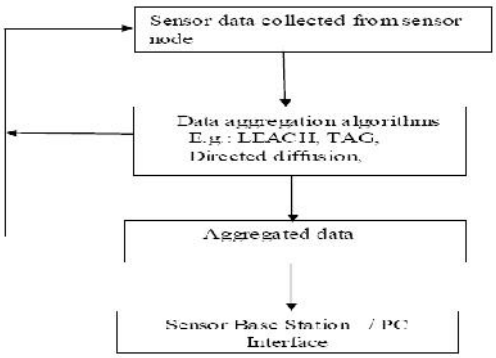
PROPOSED SYSTEM:

As previously described, the goal of a secure routing protocol is to ensure confidentiality, integrity and availability of the network traffic. Generally, attacks on different layers try to break one or all of the CIA properties and respectively a security policy for a system may demand fulfillment of one or all of these properties. Since they are nearly stateless and do not keep any history or information about the entire network topology. However, attacks on path selection such as routing misdirection or data forwarding phase attacks such as sinkhole attack or selective forwarding attack are very powerful and can cripple the functionality of Geographic routing protocols. finally we propose a mitigation technique for each attack.

MODULES SYSTEM MODEL AND SECURITY MODEL OVERVIEW

Here present a comprehensive overview of secure data aggregation concept in wireless sensor networks and survey on data aggregation protocols. Although the presented research addresses the many problems of data aggregation, there are still many research areas that need to be associated with the data aggregation process, especially from the security point of view.

The general data aggregation algorithm works as shown in the below figure. The algorithm uses the sensor data from the sensor node and then aggregates the data by using some aggregation algorithms such as centralized approach, LEACH(low energy adaptive clustering hierarchy),TAG(Tiny Aggregation) etc. This aggregated data is transfer to the sink node by selecting the efficient path.



General Architecture Of The Data Aggregation Algorithm

Centralized Approach: This is an address centric approach where each node sends data to a central node via the shortest possible route using a multihop wireless protocol. The sensor nodes simply send the data packets to a leader, which is the powerful node. The leader aggregates the data which can be queried. Each intermediate node has to send the data packets addressed to leader from the child nodes. So a large number of

messages have to be transmitted for a query in the best case equal to the sum of external path lengths for each node.

QUERY PROCESSING

A query layer to support aggregate queries. With the interface provided, the clients can issue queries without knowing how the results are generated, processed and returned by the sensor network to them. The query layer processes declarative queries and generate a cost effective query plan. They follow a database approach to design a query interface for sensor networks. The view of cost is different for sensor networks. The major factor under consideration is the communication cost, involving the cost of routing the queries and aggregating data over the sensor networks.

Proposes a Query Agent that provides application independent query interface and an API support to map the user specified queries to lower level semantics corresponding to underlying routing and aggregating protocols. It supports different communication models - anycast, unicast, multicast and broadcast. Query agent will support a wide variety of routing and aggregation protocols selecting the best combination based on the type of the query.

DATA AGGREGATION

Data aggregation is considered as one of the basic dispersed data processing measures to save the energy and minimize the medium access layer contention in wireless sensor networks. It is used as an important pattern for directing in the wireless sensor networks. The fundamental idea is to combine the data from different sources, redirect it with the removal of the redundancy and thereby reducing the number of transmissions and also saves energy. The inbuilt redundancy in the raw data gathered from various sensors can be banned by the in-network data aggregation. In addition, these operations utilize raw materials to obtain application specific information. To conserve the energy in the system thereby maintaining longer lifetime in the network, it is important for the network to preserve high incidence of the in-network data aggregation

Data Confidentiality: In particular, the fundamental security issue is the data privacy that protects the transmitted data which is sensitive from passive attacks like eavesdropping. The significance of the data confidentiality is in the hostile environment, where the wireless channel is more prone to eavesdropping. Though cryptography provides plenty of methods, such as the process related to complicated encryption and decryption, like modular multiplication of large numbers in public key based on cryptosystems, utilizes the sensor's power speedily. Data Integrity: It avoids the modification of the last aggregation value by the negotiating source nodes or aggregator nodes. Sensor nodes can be without difficulty compromised because of the lack of the expensive tampering-resistant hardware. The otherwise hardware that has been used may not be reliable at times. A compromised message is able to modify, forge and discard the messages. Generally, in wireless sensor networks for secure data aggregation, two methods can be used. They are hop by hop encrypted data aggregation and end to end encrypted data aggregation.

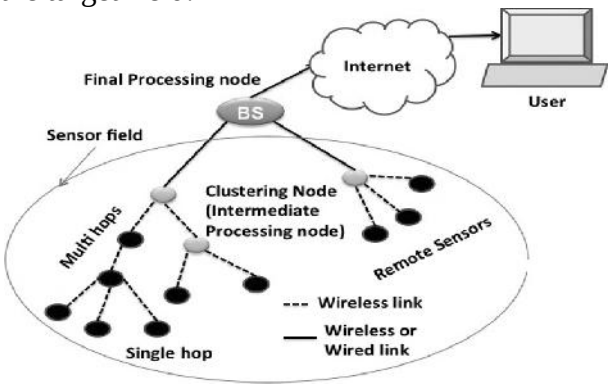
PRIVACY-PROTECTING DATA AGGREGATION

The data aggregation is a widely used mechanism in Wireless Sensor Networks (WSNs) to increase lifetime of a sensor node, send robust information by avoiding redundant data transmission to the base station. The privacy preserving data aggregation is a challenge in wireless communication medium as it could be eavesdropped; however it enhances the security without compromising energy efficiency. Thus the privacy protecting data aggregation protocols aims to prevent the disclosure of individual data though an adversary intercept a link or compromise a node's data. We present a study of different privacy preserving data aggregation

techniques used in WSNs to enhance energy and security based on the types of nodes in the network, topology and encryptions used for data aggregation.

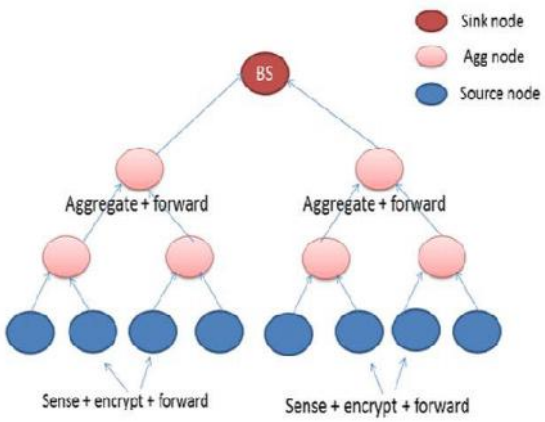
The Wireless Sensor Networks is an ad hoc network consisting of a large number of distributed autonomous wireless devices called sensors, which is densely deployed in remote areas to detect the environmental conditions such as temperature, pressure, humidity, sound, vibration, motion, pollutants etc. The sensor nodes are resource constrained in terms of energy, memory and computation capabilities. There are three types of nodes: normal sensor nodes, intermediate nodes, base station. A sensor node has the capability of sensing, processing and communicating the data collected from the environment in which it is deployed and report it to the base station located at remote places.

An aggregator aggregates sensed data with other data received from multiple sensor nodes based on some preferred aggregation functions and forward aggregation results to another aggregator or BS. Finally, the BS processes the received data and derives the significant information reflecting the events in the target field.



TYPICAL WIRELESS SENSOR NETWORK ARRANGEMENT

The dense deployment of resource constraint sensor nodes in terms of energy, memory, bandwidth, communication and computational capabilities in close proximity sense the same data, which in turn increases the redundant data in the network. The transmission of these redundant data incurs the sensor node energy. A data aggregation mechanism avoids the redundant data transmission employed in WSNs at aggregator level, which reduces the energy consumption of a node and thereby increases the network lifetime. The impact of data aggregation in WSNs.



PRIVACY PRESERVING SECURE IN-NETWORK DATA AGGREGATION (PPSDA)

The steps of the proposed PPSDA are as follows:

1. Let the i S is the sensed value by node i.
2. The sensed value i.e. i S , is appended by multiplying the sensed value with a predefined threshold, k.
3. Leaf node encrypts the appended value, i.e. ( )i e kS
4. Suppose m is a message to be encrypted such that m€Zn .
5. Select random s where n s€Z
6. Compute cipher text as: C € rm.snmod n2
7. The parent node performs the aggregation operation on encrypted values €Agg €e €Si€ and forwards the encrypted data to the next node in the communication chain until the data reaches the sink.
8. At the sink, the data is decrypted to obtain the final aggregated value.
9. For calculating the SUM at sink, the aggregated value i.e. X is divided by the predefined threshold of k. The

floor of the decimal will give the SUM of all the sensed data.

10. For finding the COUNT, modulo of X is calculated.
11. Finally, the MEAN is obtained by dividing the SUM by the COUNT.
12. The intermediate nodes are used only for the aggregation. The homomorphic property of paillier cryptosystem allows addition in the data without decrypting the original data.
13. The proposed PPSDA approach makes use of the additive homomorphic property of paillier algorithm.
14. All nodes perform SUM on the aggregated values without decrypting them, thus maintaining the confidentiality and integrity of the original data.

SECURE DATA AGGREGATION MECHANISMS

Several data agagregation techniques have been proposed to enhance data availability. The aggregation functionalities with the advantages provided by a reputation system in order to enhance the network life time and the accuracy of the aggregated data. By monitoring neighbourhood's activities, each sensor node evaluates the behaviour of its cell members in order to filter out the inconsistent data in the presence of multiple compromised nodes.

Accomplish data trustworthiness by extending Josang's trust model. Based on the multilayer aggregation architecture of network, they design a trust-based framework for data aggregation with fault tolerance with a goal to reduce the impact of erroneous data and provide measurable trust-worthiness for aggregated results.

The important and challenging problem of assuring trustworthiness of sensor data in the presence of malicious adversaries. They developed a game theoretic defense strategy to protect sensor nodes from attacks and to guarantee a high level of trustworthiness for sensed data. The objective of the defense strategy is to ensure that sufficient sensor nodes are protected in each attack/defense round.

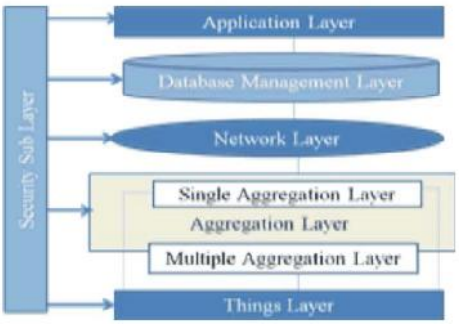


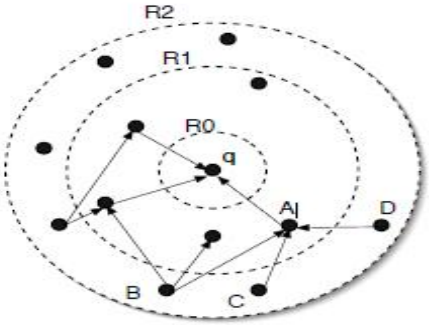
Figure. 3.2: data aggregation architecture

ATTACK MODEL AND SECURITY GOALS

We consider a setting with a polynomially-bounded attacker, which can compromise some of the sensors as well as the aggregator. Actions of a corrupted device are totally determined by the adversary (i.e., a compromised node or aggregator is Byzantine in its behavior). In particular, the adversary can arbitrarily change the measured values reported by a corrupted sensor. We assume that the adversary can only compromise a (small) fraction of the sensor nodes.

ROBUST AGGREGATION

A brief overview of the synopsis diffusion approach for robust aggregation. Figure 3.3 illustrates how the synopsis diffusion approach uses a rings topology for aggregation.



DIFFUSION OVER A RINGS TOPOLOGY

In the query distribution phase, nodes form a set of rings around the querying node q based on their distance in hops from q. During the subsequent query aggregation period, starting in the outermost ring each node generates a local synopsis s = SG(v) where v is the sensor reading relevant to the query, and broadcasts it. (SG() is the synopsis generation function.) A node in ring Ri will receive broadcasts from all the nodes in its range in ring Ri+1. It will then combine its own local synopsis with the synopses received from its children using a synopsis fusion function SF(), and then broadcast the updated synopsis. Thus, the fused synopses propagate level by-level until they reach the querying node, who first combines the received synopses with its local synopsis using SF() and then uses the synopsis evaluation function SE() to translate the final synopsis to the answer to the query.

In this algorithm, each node generates a local synopsis which is a bit vector ls of length k >logn, where n is an upper bound on the nodes in the network. To generate its local synopsis, each node executes the function CT(X, k) given below, where X is the node's identifier and k is the length of ls in bits. CT() can be interpreted as a coin-tossing experiment (with a cryptographic hash function h(), modeled as a random oracle whose output is 0 or 1, simulating a fair coin-toss), which returns the number of coin tosses until the first heads occurs or k+1 if k tosses have occurred with no heads occurring. In the local synopsis ls of node X, a single bit i is set to 1, where i is the output of CT(X, k). Thus ls is a bitmap of the form 0i−11⋯ with probability 2−i.

Algorithm CT(X, k)

```
i=1;
while i < k+1 AND h(X, i) = 0 do
i = i+1;
end while
return i;
```

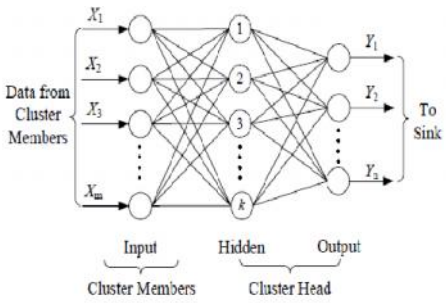
The synopsis fusion function SF() is simply the bitwise Boolean OR of the synopses being combined. Each node fuses its local synopsis ls with the synopses it receives from its children by computing the bit-wise OR of all the synopses. Let S denote the final synopsis computed by the querying node by combining all the synopses received from its children and its local synopsis. We observe that S will be a bitmap of length k of the form 1r−10⋯. The querying node can estimate Count from S via the synopsis evaluation function SE(): if r is the lowest-order bit in S that is 0, the count of nodes in the network is 2r−1/0.7735.

BACK PROPAGATION NEURAL NETWORK FOR DATA AGGREGATION

Back propagation neural network (BPNN) training is used for the data aggregation. Back propagation neural network for data aggregation is as shown in Figure.

BPNN is a multi layer feed forward neural network. The input layer is located in the leaf nodes (cluster members) and hidden and output layer is in the cluster head (CH).This neural network involves supervised learning. Back propagation network have: 1. Input layer consists of the back propagation network inputs. 2. Hidden layer consists of neurons; these neurons are responsible for adjusting the weights to determine the correct

weights. 3. Output layer consists of back propagation network outputs and represents the final decision of training operation.



DATA AGGREGATION USING BACK PROPOGATION NEURAL NETWORK

In a feed forward neural network the information flows from leaf nodes to cluster

head. Every node processes the data in following manner. The flowchart for data aggregation using BPNN is as shown in Fig.7. First initialize the weights for all nodes. Leaf node data are the input to the neural network. Then transfer the data to the cluster head for training or processing.

The aggregated data Xj(n) of node Cj is given by

$$X_j(n) = y_j(n)w_j(n) + \sum_{K=k} Y_k(n)w_k(n)$$

Where,

Cj = Cluster head

wj(n) =Associated weight of cluster head

yj(n) = Observation made by Cj

K = set of indices of all leaf nodes of node Cj.

wk(n) = Weights of leaf nodes

Yj(n)= output after applying sigmoid function to Xj(n)

The sigmoid function is defined as Y=1/(1+e-X) is used at intermediate nodes to get the binary decision. The decision

Yj(n) of node Cj is given by,

Yj(n)=1/(1+e-Xj)

The binary decision Yj(n) of each node is propagated to cluster head. The binary decision made by CH estimates the event hypothesis Hest(n)=YCH(n).

3. CONCLUSION

Data aggregation mechanisms along with data averaging techniques are analysed. Network model proposed by Wagner is described for sensor network. Adversary models with their assumptions are reviewed. New sophisticated collu-sion attack scenarios along with its impact on wireless sensor networks is explained. As soon as computational power of very low power processors significantly improves, future aggregator nodes will be capable of performing more difficult data aggregation algorithms, thus making wireless sensor networks less vulnerable. In future an enhanced strategy against collusion attack is introduced which makes is not only collusion robust, but also more accurate and faster converging.

We have studied the two most important parts of data communication in sensor networks- query processing, data aggregation and realized how communication in sensor networks is different from other wireless networks. Wireless sensor networks are energy constrained network. Since most of the energy consumed for transmitting and receiving data, the process of data aggregation becomes an important issue and optimization is needed. Efficient data aggregations not only provide energy conservation but also remove redundancy data and hence provide useful data only. The simulation result shows that when the data from source node is send to sink through neighbors nodes in a multihop fashion by reducing transmission and receiving power, the energy consumption is low as compared to that of sending data directly to sink that is aggregation reduces the data transmission then the without aggregation. We have showed how aggregate queries are efficiently executed in wireless sensor networks.

4. FUTURE WORK

In addition to all the discussed problems, some open issues are remained that are worthy of a complementary research. We handled the trust value adjustment in a simple and elementary way. Typically, trust value adjustment is a critical and delicate matter. The system efficiency could be improved even more in case of further investigation and empirical tests on trust value adjustment. All scenarios were tested for grid networks. Some mitigation techniques such as neighbor analysis approach might generate false alarms in sparse and random graph networks. In addition, GPSR always works better than GEAR in sparse networks. Thus, results could be different for both GPSR and GEAR if the network topology was a random graph. Finally, we did not study the security of the Face Routing component. With the assumption that the algorithm of the component is complex, we then assumed that the adversary has less incentive to attack it. However, the door is open and the adversary

can attack this component with more efforts. To the best of our knowledge, no research has been done in this context up to this moment.

5. BIBLOGRAPHY

[1] B. Settles, "Active Learning Literature Survey," technical report, 2010.

[2] R. Huang and W. Lam, "Semi-Supervised Document Clustering via Active Learning with Pairwise Constraints," Proc. Int'l Conf. Date Mining, pp. 517- 522, 2007.

[3] P. Mallapragada, R. Jin, and A. Jain, "Active Query Selection for Semi-Supervised Clustering," Proc. Int'l Conf. Pattern Recognition, pp. 1-4, 2008.

[4] Q. Xu, M. Desjardins, and K. Wagstaff, "Active Constrained Clustering by Examining Spectral Eigenvectors," Proc. Eighth Int'l Conf. Discovery Science, pp. 294-307, 2005.

[5] L. Breiman, "Random Forests," Machine learning, vol. 45, no. 1, pp. 5-32, 2001.

[6] M. Al-Razgan and C. Domeniconi, "Clustering Ensembles with Active Constraints," Applications of Supervised and Unsupervised Ensemble Methods, pp. 175-189, Springer, 2009.

[7] O. Shamir and N. Tishby, "Spectral Clustering on a Budget," J. Machine Learning Research - Proc. Track, vol. 15, pp. 661-669, 2011.

[8] K. Voevodski, M. Balcan, H. Ro¨ glin, S. Teng, and Y. Xia, "Active Clustering of Biological Sequences," J. Machine Learning Research, vol. 13, pp. 203-225, 2012.

[9] L. Breiman, "RF/Tools: A Class of Two-Eyed Algorithms," Proc. SIAM Workshop, Statistics Dept., 2003.

[10] T. Shi and S. Horvath, "Unsupervised Learning with Random Forest Predictors," J. Computational and Graphical Statistics, vol. 15, pp. 118-138, 2006.

[11] A. Frank and A. Asuncion, "UCI Machine Learning Repository," http://archive.ics.uci.edu/ml, 2010.

[12] O. Mangasarian, W. Street, and W. Wolberg, "Breast Cancer Diagnosis and Prognosis via Linear Programming," Operations Research, vol. 43, no. 4, pp. 570-577, 1995.