Securing Data Aggregation with Synopsis Diffusion in Wireless Sensor Networks

Aleena Sebastian
Master of Technology in Computer Science & Engineering
Department of Computer Science and Engineering
Jawaharlal College of Engineering and Technology
Palakkad, India

Manoj M
Assistant Professor
Department of Computer Science and Engineering
Jawaharlal College of Engineering and Technology
Palakkad, India

Abstract – Sensor networks are collection of sensor nodes which send sensed data to base station. Sensors are battery powered and have limited energy. To minimize the energy consumption on sensors it is necessary to reduce data traffic inside the sensor networks. Data aggregation is the mechanism by which the energy consumption in wireless sensor networks is reduced. Here data from various sensor nodes are collected and merged in a single data point. The aggregation process generates new security issues. The basic approaches of network security are confidentiality and integrity. Securing data aggregation with synopsis diffusion in wireless sensor networks is a method which combines multipath routing scheme with synopsis diffusion and provide confidentiality and integrity through encryption and authentication.

Keywords- Aggregation, data authentication, encryption

I. INTRODUCTION

A wireless sensor network consists of a number of sensors distributed spatially to monitor physical or environmental conditions. These sensor devices are limited in their energy, computation and communication capabilities. Data aggregation minimizes energy consumption. Data transfer in sensors is an important application. Nowadays most of the data transfer occurs through internet with advance in technology. The main issue with data transfer is to ensure security. The basic approaches of network security include confidentiality and integrity. In case of wireless sensor networks security is an important issue. It is challenging to design a security mechanism for wireless sensor networks. Synopsis diffusion combines multipath routing scheme with duplicate-insensitive algorithm to compute aggregate (sum). Aggregations collect information from various sensors and merge them in a single data point. The most important aggregates considered by the research community include Count, Sum, Uniform Sample, and Median. Previously tree based approaches are used which are not resilient to communication losses resulting from node and transmission failures, which are relatively common in WSNs. count and sum are duplicate-sensitive and the multipath routing scheme may lead to double counting problem. Synopsis diffusion eliminates this double counting problem and reduces energy consumption. This approach uses a ring topology where a node may have multiple parents in the aggregation hierarchy, and each sensed value or sub aggregate is represented by a duplicate-insensitive bitmap called synopsis. .

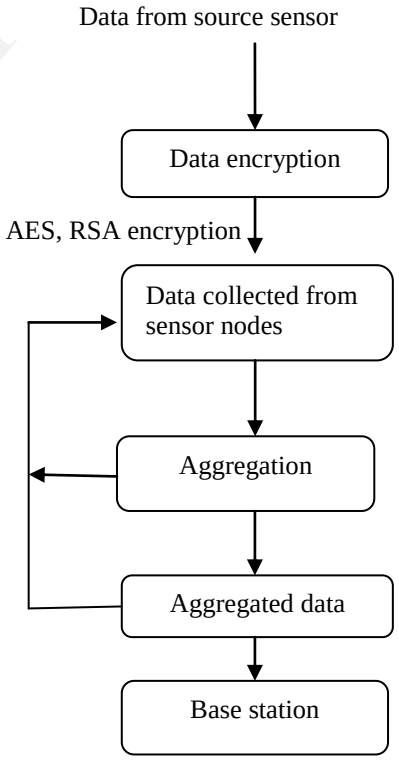

Fig.1 System architecture

To ensure confidentiality for the data which is transmitted through the sensor nodes to the base station encryption can be

applied. Message authentication codes are generated for integrity checking. A verification algorithm is used to check whether the aggregate computed at the base station is correct or not. Data confidentiality ensures that the information which is transferred is never revealed to anyone who is not an authorized one to receive it. In a hop-by-hop encryption scheme the aggregator node needs to decrypt the received data to perform some aggregation process and then encrypt it before forwarding it to the parent node. In an end-to-end encryption scheme the aggregator does not need decrypt the data instead of this; it can perform aggregation directly on the encrypted data with homomorphic encryption. Data integrity ensures that the data has not been altered. Confidentiality alone is not sufficient since the adversary can still alter the data. Tree based aggregation approaches are proposed. The drawback of tree based approach is the limited robustness of the system. To overcome this drawback, a new approach was proposed by many researchers .in which instead of sending partially aggregated data to single parent node in aggregation tree, a node could send data over multiple paths. Here each and every node can send data packets to its possibly multiple neighbor's. Hence data packet flow from source node to the sink node along multiple path ,there are a number of intermediate node between source node to sink node so aggregation done in every intermediate node. Using this approach make the system robust but there is some extra overhead. The example of this approach like ring topology, where network is divided in to concentric circle with defining level according to hop distance from sink.

II. DATA AGGREGATTION USING SUM

An aggregation framework called synopsis diffusion which uses a ring topology is proposed. During the query distribution phase, nodes form a set of rings around the base station (BS) based on their distance in terms of hops from BS. Ti denote the ring consisting of the nodes which are i hops away from BS. In the subsequent aggregation period, starting in the outermost ring, each node generates and broadcasts a local synopsis SG(v) where SG(v) is the synopsis generation function and v is the sensor value relevant to the query. Node in ring Ti will receive broadcasts from all of the nodes in its communication range in ring Ti+1. It will then combine its own local synopsis with the synopses received from its children using a synopsis fusion SF() function and then broadcast the updated synopsis. Thus, the fused synopses propagate level-by-level until they reach BS, which first combines the received synopses using SF () and then uses the synopsis evaluation function SE()to translate the final synopsis to the answer to the query.

A. Synopsis generation

The sensor nodes sense the value and generate synopsis corresponding to the value. The sum function is used for synopsis generation. For a node X, it will generate the local synopsis Q^X corresponding to the sensed value $V_{X.}$

B. Synopsis fusion

For a node X the fused synopsis $B^{X,}$ is its local synopsis if X is a leaf node otherwise B^X is the logical OR of X's local synopsis Q^X with X's children's fused synopses.

C. Synopsis evaluation

The synopsis evaluation function is given by the equation $2^{z-1}/0.7735$, where z is the lowest order bit in B that is 0.

III. ENCRYPTION USING AES

The data is encrypted with Advanced Encryption Standard (AES) before it is transmitted to the BS. AES operates on a block size of 128 bit. It is a symmetric key encryption algorithm; the key size can be 128,192 or 256 bit. The key size specifies the number of rounds required for the transformation of the plaintext into cipher text. For 128 bit keys there will be 10 rounds. Each round requires a round key which is generated from the 128 bit cipher key. The 128 bit data block is divided into 16 bytes. These bytes are mapped into a 4*4 arrays called state array. All the rounds are same except the last round. Each round has substitute bytes, shift rows, mix columns, add round key. In the final round there are no mix columns. AES algorithm is illustrated in Fig.2. Different rounds involved in the AES encryption algorithm are the following.

i. Key Expansion— round keys are derived from the cipher key using Rijndael's key schedule.
ii. Initial Round

- Add Round Key—each byte of the state is combined with the round key using bitwise XOR

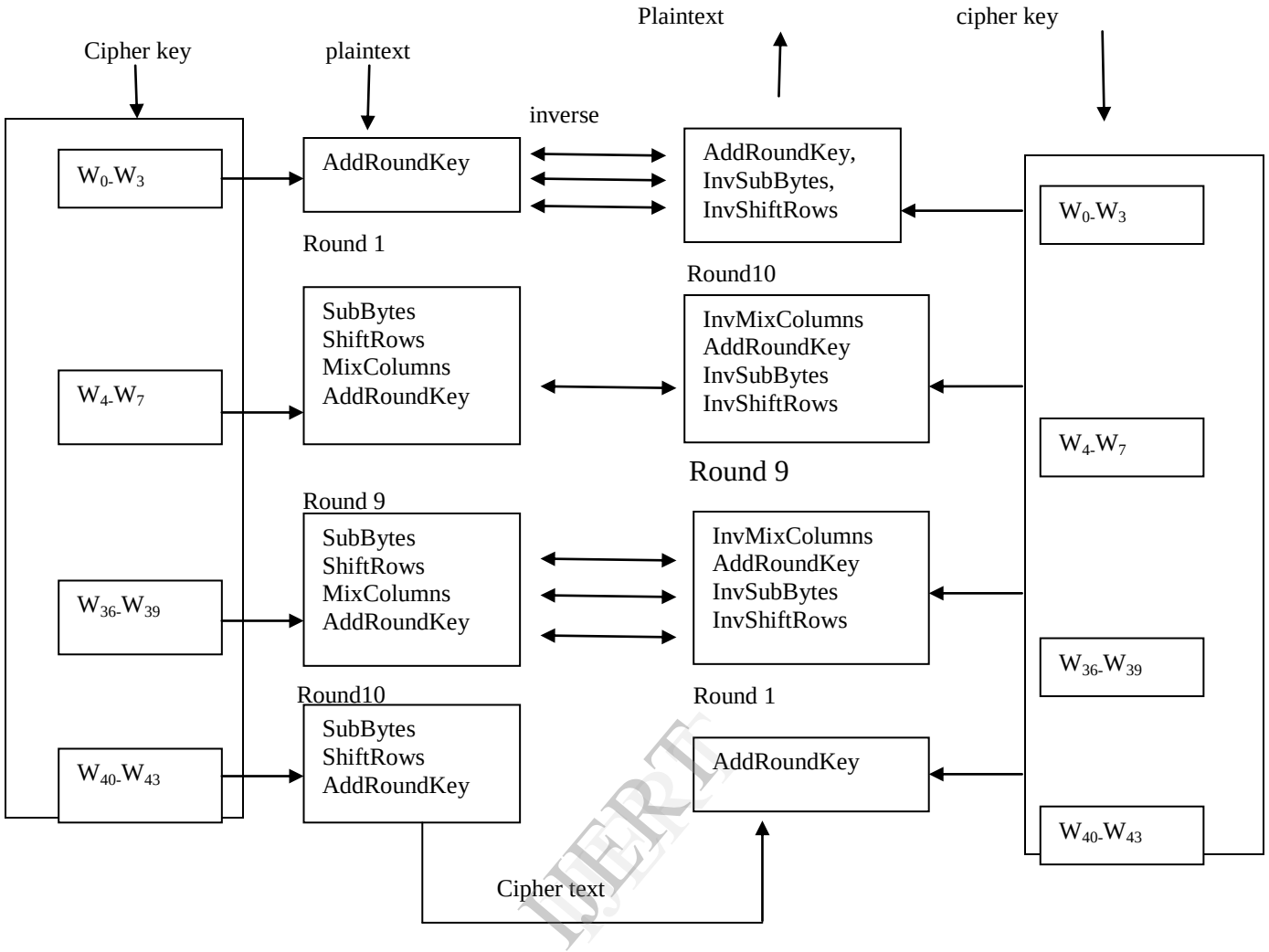

Fig. 2 Block diagram of AES algorithm

iii. ROUNDS

- Substitute Bytes—a non-linear substitution step where each byte is replaced with another according to a lookup table.
- Shift Rows—a transposition step where each row of the state is shifted to left cyclically a certain number of steps.
- Mix Columns—a mixing operation which operates on the columns of the state, combining the four bytes in each column.
- Add Round Key

iv. FINAL ROUND

Involves substitute bytes, shift rows.

IV. ENCRYPTION USING RSA

RSA is a public key encryption algorithm. Here the data is encrypted with public key and is decrypted with private key. To generate the key two prime numbers p and q are selected and N = p*q is calculated. Then Φ (n) = (p-1) (q-1) is obtained. Now select an integer e less than Φ (n) and relatively prime to Φ (n). From e the value of d is calculated as, de mod Φ (n) =1. Then the public key is KU= {e, n} and the private key is KR= {d, n}. During encryption the value of C is calculated with public key as $C = M^e \mod n$, and in decryption M is calculated as $M = C^d \mod n$. RSA is a homomorphic encryption algorithm. The main advantage of using homomorphic encryption is that we can carry out

operations on cipher text without needing the decryption key. This helps in protecting the decryption key by not exposing it at hostile places. It requires less computation as no decryption is needed.

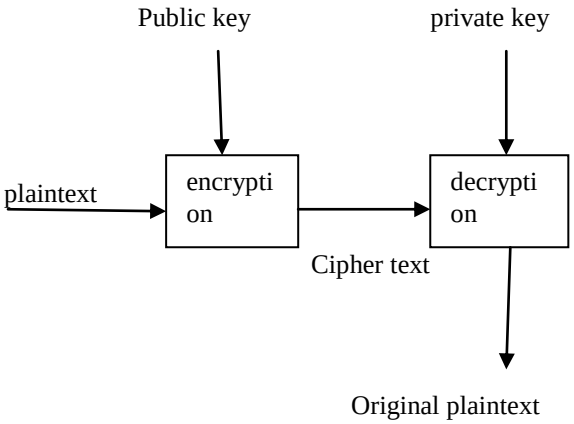

Fig. 3 Asymmetric cryptography

V. DATA VERIFICATION

BS broadcasts an aggregation query which includes a random value, Seed associated to the current query. In the aggregation phase, along with the fused synopsis B^X, each node X also sends a MAC towards BS authenticating its sensed value V_X . Node uses Seed and its own ID to compute its MAC. BS verifies the final synopsis if it receives one valid MAC for each 1 bit in the final synopsis. To verify a particular "1" bit BS need not to receive authentication messages from all the nodes which contribute to that bit. It requires only a single authentication message from one of these nodes. Thus there is only need to forward one MAC corresponding to each " 1" bit. The verification algorithm presented here reduces the communication overhead per node Here each node forwards one MAC each for at most k bits in the synopsis, where k is a small constant (e.g., 3) and BS will authenticate the rightmost k "1" bits in the final synopsis. Then, BS can securely compute R with very high probability, where R is the length of the prefix of consecutive "1"s in the final synopsis. The higher the value of k, the greater is the probability that a false "1" bit in the final synopsis is detected. The verification process involves the following steps.

Step 1: Receive the synopsis values along with MAC from d child nodes

Step 2: Aggregate the received synopsis with local synopsis generated.

Step 3: Generate one MAC for the k rightmost 1 bits

Step 4: Construct the union of the received MACs and the self generated one and randomly select MACs from it.

Step 5:Broadcast the final synopsis and the MACs.

The BS when receives the messages from its child nodes, compute the final synopsis, perform the decryption and verifies the received MAC. The verification process succeeds if it receives one valid MAC for the rightmost k "1"s present in final synopsis.

VI.EXPERIMENTAL RESULTS

We implemented the proposed synopsis diffusion scheme by using Java 2 Standard Edition (Jdk 1.7) in a personal computer running Microsoft Windows 7 with the specifications such as processor above 2 GHz, hard disk of 80 GB and RAM of 1 GB. The technology used is Java Swing. The platform framework for running the application is NetBeans IDE. We experimented the synopsis diffusion approach. If no attack is launched, BS will receive atleast one MAC for each of the k rightmost "1"s in the final synopsis. The encryption strategies ensure confidentiality

VII. CONCLUSION AND FUTURE WORK

We discussed the security issues of in-network aggregation algorithms to compute aggregates such as Sum. We discussed how a compromised node can corrupt the aggregate estimate of the base station, keeping our focus on the ring-based hierarchical aggregation algorithms. To address this problem, we presented a lightweight verification algorithm which would enable the base station (BS) to verify whether the computed aggregate was valid. AES and RSA encryption provides confidentiality and the authentication process provides integrity. For future work an attack resilient computation algorithm can be generated which would guarantee the successful computation of the aggregate even in the presence of an attack.

REFERENCES

1. Sankardas Roy, Mauro Conti, Sanjeev Setia , and Sushil Jajodia: *"Secure Data Aggregation in Wireless Sensor Networks".2012.*
2. Y. Yang, X. Wang, S. Zhu, and G. Cao, "SDAP: A secure hop-by-hop data aggregation protocol for sensor networks," in *Proc. Seventh ACM . Symp. Mobile Ad Hoc Networking and Computing (MobiHoc)*, 2006.
3. H. Chan, A. Perrig, and D. Song, "Secure hierarchical in-network aggregation in sensor networks," in *Proc. ACM Conf. Computer and Communications Security (CCS)*, 2006.
4. S. Nath, H. Yu, and H. Chan, "Secure outsourced aggregation via one-way chains," in *Proc. 35th SIGMOD Int. Conf. Management of Data*, 2009.
5. S. Nath, P. B. Gibbons, S. Seshan, and Z. Anderson, "Synopsis diffusion
for robust aggregation in sensor networks," in *Proc. 2nd Int. Conf. Embedded Networked Sensor Systems (SenSys)*, 2004.
6. M. Garofalakis, J. M. Hellerstein, and P. Maniatis, "Proof sketches: Verifiable in-network aggregation," in *Proc. 23rd Int. Conf. Data Engineering(ICDE)*, 2007.