ISSN: 2229-6948(ONLINE)
DOI: 10.21917/ijct.2019.0308

SECURE DATA AGGREGATION IN WSN USING SYNOPSIS DIFFUSION

S. Vimalnath[1] and G. Ravi[2]

[1]Department of Electronics and Communication Engineering, Paavai Engineering College, India
[2]Department of Electronics and Communication Engineering, Sona College of Technology, India

Abstract

In a wireless sensor network, a sensor node is severely constrained in terms of communication bandwidth, computation capability and energy reserves. To reduce energy consumption and the amount of communication, many systems also perform in-network data aggregation. A resilient aggregation framework called synopsis diffusion approach which combines multipath routing schemes with duplicate-insensitive algorithms to accurately compute aggregates, such as Count and Sum in spite of message losses resulting from node and transmission failures. Consequently, these systems are vulnerable to wide variety of attacks. In particular we consider two major attacks called false sub aggregate and wormhole attack. In this paper, the synopsis diffusion approach secure against attacks was provide by designing a novel light weight verification algorithm which can be used by the base station to determine the attacks and also the falsified value. The performance was evaluated via both analysis and extensive simulation study which shows that our verification algorithm outperforms other existing approaches.

Keywords:

Light Weight Verification, WSN, Synopsis Diffusion

1. INTRODUCTION

A Wireless Sensor Network (WSN) typically consists of a sink node sometimes referred to as a Base Station and a number of small wireless sensor nodes. The base station is assumed to be secure with unlimited available energy while the sensor nodes are assumed to be unsecured with limited available energy [9]. The sensor nodes monitor a geographical area and collect sensory information [4]. Sensory information is communicated to the Base Station through Wireless hop by hop transmissions. To conserve energy this information is aggregated at intermediate sensor nodes by applying a suitable aggregation function on the received data. Aggregation reduces the amount of network traffic which helps to reduce energy consumption on sensor nodes [2]. It however complicates the already existing security challenges for wireless sensor networks and requires new security techniques tailored specifically for this scenario. Providing security to aggregate data in Wireless Sensor Networks is known as Secure Data Aggregation in WSN [1].

WSN often consist of a large number of low-cost sensor nodes that have strictly limited sensing, computation, and communication capabilities. The data aggregation in WSN. Due to resource restricted sensor nodes, it is important to minimize the amount of data transmission. So that the average sensor lifetime and the overall bandwidth utilization are improved [3].

Data aggregation is the process of summarizing and combining sensor data in order to reduce the amount of data transmission in the network. As wireless sensor networks are usually deployed in remote and hostile environments to transmit sensitive information, sensor nodes are prone to node compromise attacks and security issues such as data confidentiality and integrity are extremely important [6]-[8].

The nature of wireless sensor network was very attractive to attackers because the sensors have their limited battery, power, memory and processing capabilities [5]. In large WSNs, computing aggregates in-network (i.e., combining results at intermediate nodes during message routing) significantly reduces the amount of communication and hence the energy consumed [10]. To combine multipath routing and accurately compute aggregates the synopsis diffusion approach is proposed. The important aggregates considered here is Count and Sum. The synopsis diffusion do not include any provision for security and in particular a compromised node may inject false data that leads to incorrect aggregates being computed at the base station [11].

One of the potent form of denial of Service attacks called wormhole attack which makes the sensor to drop their packets that may disable whole sensor network. This is an important problem since sensor networks are highly vulnerable to node compromises due to the unattended nature of sensor nodes and the lack of tamper-resistant hardware. To address this problem we present a novel light weight verification algorithm by which the base station can determine if the computed aggregates (predicate Count or Sum) includes any false contribution and also it detect and locate the worm hole attack. Through theoretical analysis and extensive simulation study show that our algorithm outperforms other existing approaches.

2. PROPOSED METHOD

In the synopsis diffusion approach, nodes are classified into multiple rings determined by their hop counts from the base station, illustrated in Fig.1. During the query distribution phase, the base station’s aggregation request is broadcast hop-by-hop, and each node *X* on any hop keeps track of the nodes on the previous hop from which *X* has received the aggregation request, and node *X* considers all of them as its parents. Once the aggregation request reaches all of the nodes in the network, ring construction is completed. In the subsequent query aggregation phase, starting in the outermost ring, each node generates a local synopsis relevant to the query, and broadcasts it to its neighbors. A node in ring T_i will receive broadcasts from all of the nodes in its range in ring T_i+1. It will then aggregate its own local synopsis with the synopses received from its children, and then broadcast the updated synopsis. Thus, the fused synopses propagate level-by-level until they reach the base station, which combines the received synopses to derive the final aggregate. This approach is robust against communication loss because each node contribution to the aggregate reaches the base station via multiple paths.

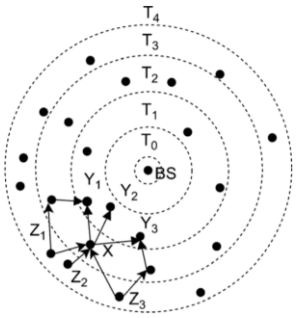

Fig.1. Synopsis diffusion over a ring topology - A node may have multiple parents e.g., X has three parents, Y_1, Y_2, Y_3

A robust and scalable aggregation framework called synopsis diffusion which uses a ring topology which uses a ring topology as illustrated. 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. By T_i we denote the ring consisting of the nodes which are 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() is the synopsis generation function and v is the sensor value relevant to the query.

A node in ring T_i will receive broadcasts from all of the nodes in its communication range in ring T_i+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 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. The duplicate-insensitive synopsis diffusion algorithm was described for Count and Sum.

2.1 SYNOPSIS DIFFUSION

Synopsis Diffusion addresses the problem of a reliable computer system in a wireless sensor network (for example, the average temperature recorded by the sensors). In-network aggregation via a spanning tree rooted at the base station is a traditional approach for computing aggregates. A branch, though, which is vulnerable to node and connectivity errors, does in fact have incorrect responses. For examples, the inaccuracy can be as high as 75% under a message loss rate (20%-30%), which is common for true sensor applications. Another way of making routing reliable is to use redundancy, for instance by multi-path routing. Nevertheless, using this compatibility of conventional in-network aggregation methods, double-counting would be added as sensor readings and incomplete tests would be sent over many routes.

Given its inaccuracies, researchers were held to the topology of the tree in this double-counting issue. In other terms, routing is defined by the criteria of in-network aggregation strategies in conventional network aggregation approaches. Overview Diffusion decouples aggregation and routing to allow for individual optimization. Overview Diffusion obtained topology freedom by using order and ODI synopses. Double counting and overview intermediate tests during the in-network aggregation are omitted from ODI synopses, a special class of synopses used for conventional data stream analysis. It makes it possible to use a stable topology of aggregation. Synopsis Diffusion will render the aggregation cycle considerably more efficient than traditional methods, without excessive overhead resources, against standard node and contact failures.

Synopsis Diffusion can improve aggregation exactness by about 85%, for example, at a typical message loss rate. We have also presented novel (roughly) computational synopsis diffusion algorithms for a number of useful aggregates as well as surprisingly simple methods for checking the consistency and approximation errors of any Synopsis Diffusion algorithm as well as strategies for taking advantage of the unique feature of Synopsis Diffusion as implied message delivery.

2.2 SYNOPSIS DIFFUSION ON A RINGS OVERLAY

During the query distribution phase, nodes form a set of rings around the querying node q as follows: q is in ring R_0, and a node is in ring R_i if it receives the query first from a node in ring R_{i1}, thus a node is in ring R_i if it is i hops away from q. The subsequent query aggregation period is divided into epochs and one aggregate answer is provided at each epoch. We presume that nodes in various circles are linked loosely and that particular hours are assigned when they are open to synopsis from other nodes. The length of the allocated time is calculated a priori depending on the distribution distance, so that all sensors have room enough to transfer messages once, even when the sensors do carrier sensing.

As the underlying wireless communication is broadcast, each node transmits exactly the same number of messages as tree-based approaches. Because therefore, synopses scatter through multiple paths from the sensor nodes to the querying node, rings are much more resilient. It quantifies this additional robustness.

2.3 DUPLICATE SENSITIVE AGGREGATES

The aggregation can be carried out via routing topologies via arbitrary post. In a synopsis distribution algorithm, the main challenge is to properly support duplicate-sensitive aggregates for all potential multi-path propagation systems. To do this, we need a series of ODI synopses and fusion functions to model the goal sum (e.g. counting).

Intuitively, such a set of functions guarantee, irrespective of the interference on those paths and overlap with propagation routes, that a partial outcome of node u is calculated through the readings of sensor nodes with distributed pathways to u. Regardless of the configuration of the fusion features, the effect is the same. Therefore, if the propagation path from the sensor node to the querying node occurs, a sensor read is registered (exactly once) on an aggregate, and never more than once.

2.4 COUNT ALGORITHM

In this algorithm, each node X generates a local synopsis Q_x which is a bit vector of length n>logN_0, where N_0 is the upper bound on Count. To generate Q_x, node X executes the function $CT(X,n)$ given below (Algorithm 1), where X is the node's identifier. Algorithm 1 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, say i, until the first head occurs or n+1 if n tosses have occurred with no heads occurring.

Algorithm 1:

```
$CT(X,n)$
begin
$i$=1;
while $i < n+1$ AND $h(X,i) = 0$ do
    $i = i + 1$;
end return $i$;
end
```

In the synopsis generation function SG_{count}, the i^{th} bit of Q_x is set to 1 while all other bits are 0. Thus, Q_x is a bit vector of the form 0(*i*-1)10(*n*-i) with probability 2-*i*.

The synopsis diffusion function *SF*() is the bitwise Boolean OR of the synopses being combined. Each node *X* fuses its local synopsis Q_x with the synopses it receives from its children. Let *B* denote the final synopsis computed by BS by combining all of the synopses received from its child nodes. We observe that *B* will be a bit vector of length *n* of the form $|z|_0[0,1]_{n-z}$, where *z* is the lowest-order bit in *B* that is 0. BS can estimate Count from *B* via the synopsis evaluation function *SE*(). The count of nodes in the network is 2*z*-1/0:7735. Intuitively, the number of sensor nodes is proportional to 2*z*-1 since no node has set the z^{th} bit while computing CT(*X*,*n*).

The fused synopsis of a node *X*, B_x, is recursively defined as follows. If *X* is a leaf node (i.e., *X* is in the outermost ring), B_x is its local synopsis Q_x. If *X* is a non-leaf node, B_x is the logical OR of *X*'s local synopsis Q_x with *X*'s child fused synopses. If node *X* receives synopses $B_{x1}, B_{x2},\ldots, B_{xd}$ from *d* child nodes $X_1, X_2,\ldots,X_d$ respectively, then *X* computes *BX* as follows

$$BX = Q_x||B_{x1}||B_{x2}||...||B_{xd}$$

where || denotes the bitwise OR operator. Note that B_x represents the subaggregate of node *X*, including its descendant nodes. We note that *B* BS is same as the final synopsis *B*.

2.5 SUM ALGORITHM

The synopsis generation function *SG*() for Sum is a modification of that for count, while the fusion function *SF*() and the evaluation function *SE*() for Sum are identical to those for Count.

To generate the local synopsis Q_x to represent its sensed value v_x, node *X* invokes the function *CT*(), used for Count synopsis generation, v_x times. In the i^{th}, $1 \leq i \leq v_x$ in vacation, node *X* executes the function $CT(X_i,n)$ where Xi is constructed by concatenating its *ID* and integer *i* (i.e. $X_i = <X,i>$), and n is the synopsis length.

The value of *n* is taken as $\log_2 S+4$, where *S* is an upper bound on the value of Sum aggregate. Note that the local synopsis of a node for Count, more than one bit in the local synopsis of anode for sum may be equal to one. The pseudo code of the synopsis generation function, SGsum(*X*,v_x,*n*), is presented in algorithm 2.

Algorithm 2:

```
SGsum($X,v_x,n$),
begin
$Q_x[j] = V_j$ for $1 \leq j \leq n$; $i$=1;
while $i \leq v_x$ do
    $X_i = <X,i>$;
    $j$ = CT($X_i,n$);
    $Q_x[j] = 1$;
    $i = i + 1$;
end return $Q_x$;
end
```

Note that count can be considered as a special case of Sum where each node sensor reading is equal to one unit.

3. RESULTS AND DISCUSSIONS

In wireless sensor network an aggregate computation and verification algorithm was designed to establish reliable and secure communication. A robust and scalable aggregation framework called synopsis diffusion has been proposed for computing aggregates, such as Count and Sum. This approach uses a ring topology where a node may have multiple parents in the aggregation hierarchy, and each sensed value or subaggregate is represented by a duplicate in-sensitive bitmap called synopsis. Here the node creation and topology are formed by using NS-2.3 Tool command language coding. In sensor network a compromised nodes might attempt to thwart or drop the data aggregation process by launching several attacks such as falsified local value attack, falsified subaggregate attack, and wormhole attack and so on. Our project focusing on most two vexing attacks as falsified subaggregate attack, in which a compromised node relays a false subaggregate to the parent node with the aim of injecting error to the final value of the aggregate computed at the base station and another attack called wormhole attack which can destabilize or disable wireless sensor networks. To address this problem the simple verification algorithm is designed by using C++ language. The verification algorithm enable the base station to detect the various attack and also verify if the computed aggregate is valid. The performance of our algorithm is evaluated via both theoretical analysis and simulation.

Table.1. Simulation Parameters

Parameter	Value
MAC Protocol	IEEE 802.11
Propagation Model	Two-ray Ground
Initial Energy in Joules	15 Joules
Number of Nodes	101 Nodes
Ring Distance	80m
Antenna	Omni Directional Antenna

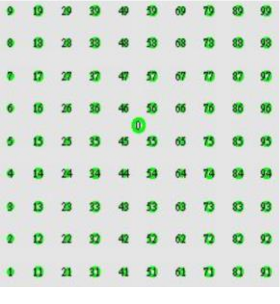

Fig.2. Node Creation Output - Initial node position output

The simulation result of node creation is shown in Fig.2 has one base station and hundred nodes which are located around the base station. The nodes are created by using tool command language.

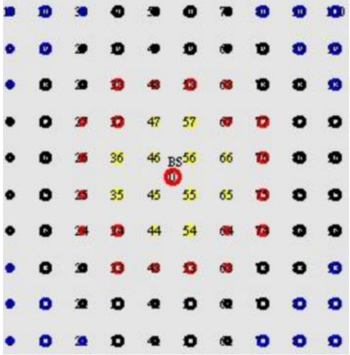

Fig.3. Ring formation output

The simulation result of ring formation of nodes is shown in Fig.3. The result shows the position of nodes after broadcasting query from base station. The nodes form set of ring around the base station based on their distance.

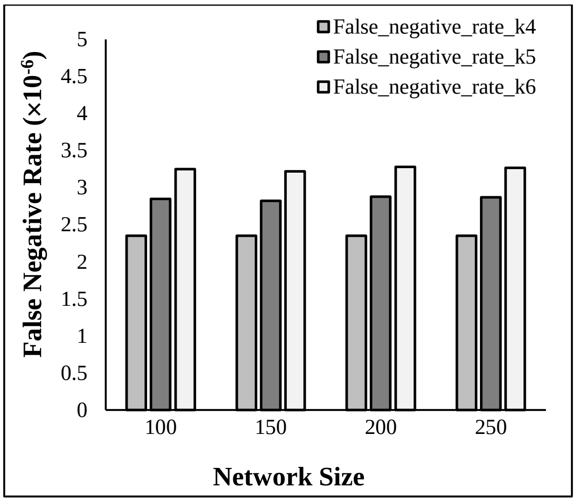

Fig.4. Network size vs. False Negative rate

The Fig.4 shows the simulation result of Network size vs False Negative rate. This result shows the simulation of our verification algorithm for different values of network size (100, 150, 200 and 250 grid sizes) and value of the parameter *k* (4, 5 and 6). The false negative rate of our verification algorithm was developed in C++ language and simulates it by using NS2.34 network simulator.

The simulation result of network size vs. average byte sent per node is shown in Fig.5. This result plots the number of bytes a node transmits on average during the verification protocol considering different network sizes. This result illustrates the per-node byte overhead of the original synopsis diffusion approach.

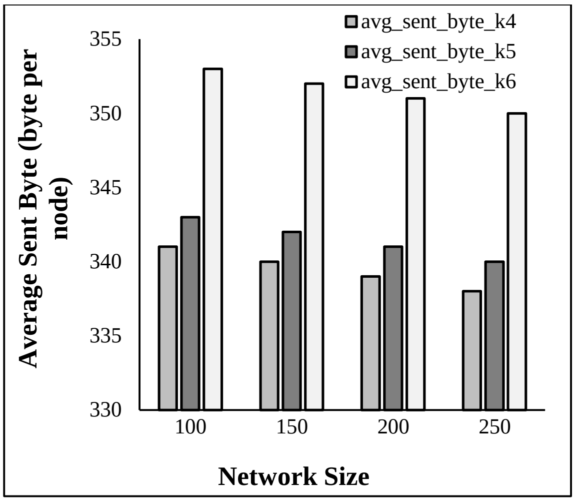

Fig.5. Network Size vs. Average Byte Sent per Node

4. CONCLUSION

In this paper, we presented security in-network aggregation algorithms for wireless sensor networks considering the possibility that a fraction of nodes might become compromised. In particular, we designed verification algorithms and aggregation computation to compute basic aggregates, such as Sum and Count. Using a verification algorithm, the base station can verify the correctness of the computed aggregate and the presence of wormhole attack and denied the attack. This algorithm reduces the amount of communication and increases the packet delivery ratio in a large sensor network. Therefore reliable communication was achieved in the sensor networks.

REFERENCES

[1] S. Roy, M. Conti, S. Setia and S. Jajodia, “Secure Data Aggregation in Wireless Sensor Networks: Filtering out the Attackers Impact”, *IEEE Transactions on Information Forensics and Security*, Vol. 9, No. 4, pp. 681-694, 2014.

[2] H. Yu, “Secure and Highly-Available Aggregation Queries in Large-Scale Sensor Networks Via Set Sampling”, *Distributed Computing*, Vol. 23, pp. 373-394, 2011.

[3] He Ronghui, Ma Guoqing, Wang Chunlei, and Fang Lan, “Detecting and Locating Wormhole Attacks in Wireless Sensor Networks using Beacon Nodes”, *World Academy of Science, Engineering and Technology*, Vol. 2, No. 7, pp. 1741-1746, 2009.

[4] G. Kiruthiga and M. Mohanapriya, “An Adaptive Signal Strength Based Localization Approach for Wireless Sensor Networks”, *Cluster Computing*, Vol. 22, No. 5, pp. 10439-10448, 2019.

[5] G. Kiruthiga, K. Kalaiselvi, R.S. Shudapreyaa and V. Dineshbabu, “Detection of Faults in Flying Wireless Sensor Networks using Adaptive Reinforcement Learning”, *International Journal of Recent Technology and Engineering*, Vol. 8, No. 4, pp. 761-763, 2019.

[6] Ritesh Maheshwari, Jie Gao and Samir R. Das, “Detecting Wormhole Attacks in Wireless Networks using Connectivity Information”, *Proceedings of IEEE International Conference on Computer Communications*, pp. 6-12, 2007.

[7] M. Garofalakis, J.M. Hellerstein and P. Maniatis, "Proof Sketches: Verifiable in Network Aggregation", *Proceedings of 23rd IEEE International Conference on Data Engineering*, pp. 132-136, 2007.

[8] H. Chan, A. Perrig and D. Song, "Secure Hierarchical in-Network Aggregation in Sensor Networks", *Proceedings of 13th ACM Conference on Computer and Communications Security*, pp. 278-287, 2006.

[9] L. Buttyan, P. Schaffer and I. Vajda, "Resilient Aggregation with Attack Detection in Sensor Networks", *Proceedings of 2nd ACM Conference on Sensor Networks and Systems for Pervasive Computing*, pp. 331-336, 2006.

[10] Y. Yang, X. Wang, S. Zhu and G. Cao, "SDAP: A Secure Hop-by-Hop Data Aggregation Protocol for Sensor Networks", *Proceedings of 7th ACM Symposium on Mobile Ad Hoc Networking and Computing*, pp. 889-893, 2006.

[11] D. Wagner, "Resilient Aggregation in Sensor Networks", *Proceedings of 2nd ACM Workshop on Security of Ad Hoc and Sensor Networks*, pp. 78-87, 2004.

[12] J. Considine, F. Li, G. Kollios and J. Byers, "Approximate Aggregation Techniques for Sensor Databases", *Proceedings of IEEE International Conference on Data Engineering*, pp. 345-354, 2004.