International Journal of Innovative Research in Science and Engineering
Vol. No.2, Issue 12, December 2016
www.ijirse.com
ISSN (O) 2454-9665
ISSN (P) 2454-0663

SECURE DATA AGGREGATION IN WIRELESS SENSOR NETWORKS: FILTERING OUT THE ATTACKER'S IMPACT

Kalyan D. Bamane[1], Dr. Ravindranath C. Cherukuri[2]

[1]*Asst.Prof, IT Dept., DYPCOE Akurdi, Pune (India)*

[2]*VTU_RRC, Belagavi (India)*

ABSTRACT

Research community is paying attention to Wireless sensor networks as it contains latest technology. WSN are self-organizing ad hoc systems which consists of several small devices with low cost. Sensor network senses the surrounding environment, gather information and transmit the same information to other sink nodes in the network. There could be transmission failure or node failure in WSN due to which communication failure occurs. Tree-based aggregation approaches are highly affected due to this. In order to deal with this problem, multi-path routing techniques can be used for forwarding sub-aggregates. Multi-path routing technique is good for duplicate insensitive aggregates such as Min and Max. It provides a fault tolerant solution. But for duplicate sensitive aggregates, such as Count and Sum, multi-path routing is not good as it counts sensor readings twice. Our algorithm securely compute aggregates, such as Count and Sum despite the falsified sub aggregate attack.

Keywords: WSN, Aggregation, Multi-Path Routing

I INTRODUCTION

Wireless sensor networks consist of the latest technology that has attained notable consideration from the research community. Sensor networks consist of numerous low cost, little devices and are in nature self-organizing ad hoc systems. The job of the sensor network is to monitor the physical environment, gather and transmit the information to other sink nodes. Generally, radio transmission ranges for the sensor networks are in the orders of the magnitude that is lesser that of the geographical scope of the unbroken network. Hence, the transmission of data is done from hop-by-hop to the sink in a multi-hop manner.

Reducing the amount of data to be relayed thereby reduces the consumption of energy within the network. Wireless sensing element network consists of an enormous range of little electromechanical sensing devices that are capable of sensing, computing and communication. These mechanical sensing element devices are created for gathering sensory data, like measuring of temperature from an in depth geographical region . Several options of the wireless sensing element networks have given rise to difficult issues . The foremost vital 3 characteristics are:

•Sensor nodes are exposed to most failures.

ISSN (O) 2454-9665
ISSN (P) 2454-0663

•Sensor nodes that build use of the broadcast communication pattern and have severe information measure restraint.

•Sensor nodes have inadequate quantity of resources

Communication losses resulting from node and transmission failures, which are common in WSNs, can adversely affect tree-based aggregation approaches. To address this problem, we can make use of multi-path routing techniques for forwarding sub-aggregates. For duplicate insensitive aggregates such as Min and Max, this approach provides a fault-tolerant solution. Unfortunately, for duplicate sensitive aggregates, such as Count and Sum, multi-path routing leads to double-counting of sensor readings.

Recently, several researchers have presented clever algorithms to solve this double-counting problem. A robust and scalable aggregation framework called synopsis diffusion has been proposed for computing duplicate-sensitive aggregates. This approach uses a ring topology where a node may have multiple parents in the aggregation hierarchy. Furthermore, each sensed value or sub-aggregate is represented by a duplicate-insensitive bitmap called synopsis.

The possibility of node compromise introduces more challenges because most of the existing in-network aggregation algorithms have no provisions for security. A compromised node might attempt to thwart the aggregation process by launching several attacks, such as eavesdropping, jamming, message dropping, message fabrication, and so on. This project focuses on a subclass of these attacks in which the adversary aims to cause the BS to derive an incorrect aggregate.

By relaying a false sub-aggregate to the parent node, a compromised node may contribute a large amount of error to the aggregate. As an example, during the Sum computation algorithm a compromised node X can in-ject an arbitrary amount of error in the final estimate of Sum by falsifying X's own sub-aggregate.

The rest of the paper is organized as : Section 2 is about literature review; Section 3 describes proposed approach; Section 4 depicts experimental results along with comparison; Section 5 represents conclusion of the paper.

II LITERATURE REVIEW

The literature review covers the background, latest development of and related techniques for secure data aggregation in Wireless Sensor Network (WSN). Partial result at intermediate node during message routing in WSN is combined in data aggregation technique.

Literature survey is that the most vital step in software system development method. Before developing the tool it's a necessity to ascertain the time issue, economy and company strength. Once this stuff is done, then next steps is to see that which OS and language will be used for developing the tool. A lot of external support is needed by programmer once they begin building the tool. This support will be obtained from senior programmers, from book or from websites. Before building the system the above factors are taken into consideration for developing the projected system. We have to analyze the Parallel and distributed systems Survey:

International Journal of Innovative Research in Science and Engineering
Vol. No.2, Issue 12, December 2016
www.ijirse.com
ISSN (O) 2454-9665
ISSN (P) 2454-0663

Parallel and distributed computing

-“A distributed system is a collection of independent computers that appear to the users of the system as a single computer.”-“A distributed system consists of a collection of autonomous computers linked to a computer network and equipped with distributed system software.” -“A distributed system is a collection of processors that do not share memory or a clock.” -“Distributed systems are a term used to define a wide range of computer systems from a weakly-coupled system such as wide area networks, to very strongly coupled systems such as multiprocessor systems.” Distributed systems are groups of networked computers, which have the same goal for their work. The terms "concurrent computing", "parallel computing", and "distributed computing" have a lot of overlap, and no clear distinction exists between them. The same system may be characterized both as "parallel" and "distributed"; the processors in a typical distributed system run concurrently in parallel. Parallel computing may be seen as a particular tightly-coupled form of distributed computing, and distributed computing may be seen as a loosely-coupled form of parallel computing. Nevertheless, it is possible to roughly classify concurrent systems as "parallel" or "distributed" using the following criteria: -In parallel computing, all processors have access to a shared memory. Shared memory can be used to exchange information between processors. -In distributed computing, each processor has its own private memory (distributed memory). Information is exchanged by passing messages between the processors.

Fig 1. Distributed System

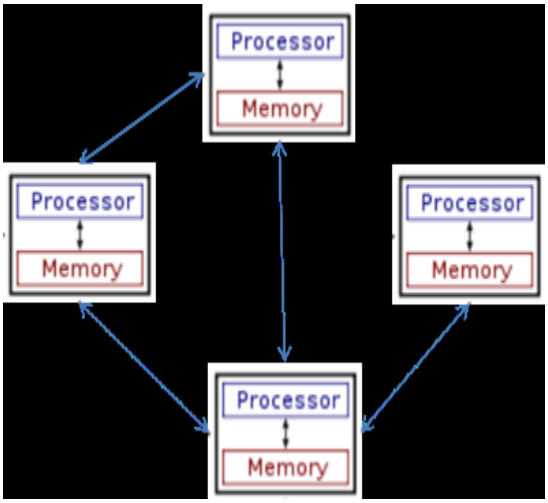

Fig 2. Distributed System

International Journal of Innovative Research in Science and Engineering
Vol. No.2, Issue 12, December 2016
www.ijirse.com
ISSN (O) 2454-9665
ISSN (P) 2454-0663

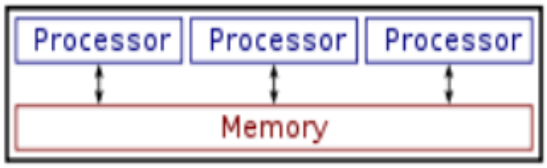

Fig 3. Parallel System

The figure on the right illustrates the difference between distributed and parallel systems. Figure 3(a) is a schematic view of a typical distributed system; as usual, the system is represented as a network topology in which each node is a computer and each line connecting the nodes is a communication link. Figure 3(b) shows the same distributed system in more detail: each computer has its own local memory, and information can be exchanged only by passing messages from one node to another by using the available communication links.

III PROPOSED WORK

This topic discusses the security issues of in-network aggregation algorithms to compute Count and Sum when a fraction of nodes in the network have been compromised. Prior research has addressed some of these security issues, but most of the existing schemes are limited to tree-based aggregation. We analyze the vulnerabilities of the ring-based aggregation within the synopsis diffusion framework. We also propose solutions for securely computing Count and Sum within this framework.

A verification algorithm and an attack-resilient computation algorithm are proposed with goal to detect whether an attack has been launched and to compute the aggregate despite the presence of the attack,repectively.

Synopsis Diffusion

The synopsis diffusion framework uses a ring topology. 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 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() 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 functions SG(), SF (), and SE() depend upon the target aggregation function, e.g. Count, Sum, etc. We now describe the duplicate-insensitive synopsis diffusion algorithms for Count and Sum. These algorithms are based on Flajolet and Martin's probabilistic algorithm for counting the number of distinct elements in a multi-set.

In this algorithm, each node X generates a local synopsis Q X which is a bit vector of length $\eta > \log N$, where N is the upper bound on Count. To generate Q X , node X executes the function CT (X, η) given below

ISSN (O) 2454-9665
ISSN (P) 2454-0663

(Algorithm 1), where X is the node's identifier. 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, say i, until the first head occurs or η + 1 if η tosses have occurred with no heads occurring. 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 (η−i) with probability 2 −i .

The synopsis fusion function SF () is the bitwise Boolean OR of the synopses being

Algorithm 1 CT (X, η)
i=1;
while i < η + 1 AND h(X, i) = 0 do
i = i + 1;
end while
return i;

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 η of the form 1 z−1 0[0, 1] η−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 [9]. The synopsis evaluation function SE() is based on Property 2 below.

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, η). 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 children's fused synopses.

If node X receives synopses B X 1 , B X 1 , . . . , B X d from d child nodes X 1 , X 2 , . . . , X d , respectively, then X computes B X as follows:

B X = Q X ||B X 1 ||B X 1 || . . . ||B X d ,

where || denotes the bitwise OR operator.

Below we present a few important properties of the final synopsis B computed at BS. The first three properties have been derived in [7,10], while Property 4 is documented from our observation. Let B[i], $1 \leq i \leq \eta$ denote the i-th bit of B, where bits are numbered starting from the left. Also, N is the number of nodes present in the network.

Property 1. For i < log 2 N −2 log 2 log 2 N , B[i] = 1 with probability ≈ 1. For i ≥

3/2log 2 N , B[i] = 0 with probability ≈ 1.

ISSN (O) 2454-9665
ISSN (P) 2454-0663

This result implies that for a network of N nodes, we expect that B has an initial prefix of all ones and a suffix of all zeros, while only the bits around B[log 2 N] exhibit much variation. This provides an estimate of the number of bits, η, required for a node's local synopsis. In practice, η = log 2 N + 4 bits are sufficient to represent B with high probability [10], where N is the upper bound of Count. This result also indicates that the length of the prefix of all ones in B can be used to estimate N . Let z = min {i|B[i] = 0}, i.e., z is the location of the leftmost zero in B. Then R = z − 1 is a random variable representing the length of the prefix of all ones in the synopsis. The following results hold for R.

Property 2. The expected value of R, E(R) ≈ log 2 (φN), where the constant φ is approximately 0.7735.

This result implies that R can be used as an unbiased estimator of log 2 (φN), and it is the basis for the synopsis evaluation function SE(), which estimates N as 2 R /φ.

Property 3. The standard deviation of R, σ R ≈ 1.1213.

This property implies that estimates of N derived from R will often be off by a factor of two or more in either direction. To reduce the standard deviation of R, Flajolet et al. proposed an algorithm named PCSA, where m synopses are computed in parallel.

The single synopsis computation algorithm is extended to the PCSA algorithm as follows:

In synopsis generation function SG count , one synopsis out of m synopses is randomly chosen before CT () is invoked and then, only the chosen synopsis is updated. The synopsis fusion function SF() for each synopsis is bitwise Boolean OR as in the original algorithm. In synopsis evaluation function SE(), the new estimator is the average of all individual R's of these synopses.

Property 4. If N nodes participate in Count algorithm, the expected number of nodes that will contribute a '1' to the i-th bit of the final synopsis B is N/2 i . We call these nodes contributing nodes for bit i of B.

This property is derived from the observation that each node X sets the i-th bit of its local synopsis Q X with probability 2 −i . As an example, for bit r = E(R) = log 2 (φN), the expected number of contributing nodes is 1/φ ≈ 1.29. This result also implies that the expected number of nodes that contribute a '1' to the bits right to the i-th bit (i.e.,bits j, where i < j ≤ η) is approximately N/2 i . As an example, the expected number of contributing nodes for bits j ≥ r + 1 is approximately 1/φ.

Algorithm 2 SG sum (X, v X , η)
Q X [j] = 0 ∀j 1 ≤ j ≤ η;
i=1;
while i ≤ v X do
X i =< X, i >;
j = CT (X i , η);
Q X [j] = 1;

ISSN (O) 2454-9665
ISSN (P) 2454-0663

i = i + 1;

end while

return Q X ;

Let B[i], $1 \leq i \leq \eta$ denote the i-th bit of the final synopsis B, where bits are numbered starting from the left. Furthermore, S is the Sum of the sensed values of the nodes present in the network.

Property 1. For i < log 2 S − 2 log 2 log 2 S, B[i] = 1 with probability ≈ 1. For i ≥

3/2log 2 S,

B[i] = 0 with probability ≈ 1.

Property 2. Let R represent the length of the prefix of all ones in B, i.e., R = z − 1 where z = min {i|B[i] = 0}. The expected value of R, E(R) ≈ log 2 (φS), where the constant φ is approximately 0.7735.

Property 3. The standard deviation of R, σ R ≈ 1.1213.

Unlike the above properties, Property 4 is not a straightforward extension of its counterpart for Count synopsis. From the construction of the synopsis generation function, SG sum () (Algorithm 2), we observe that if the Sum is S, then the function CT () is invoked S times in total considering synopsis generation of all nodes. Each node X gets a chance to set the i-th bit of Q X , its local synopsis, v X times—each time with probability 2 −i . So, the expected number of contributing nodes for the i-th bit of B not only depends on the total number of nodes N and the value of i but also on the distribution of sensor readings.

Property 4. The expected number of invocations of CT () that will contribute a '1' to the i-th bit of the final synopsis B is S/2 i , where S is the value of Sum.

As an example, with r = E(R) = log 2 (φS), the expected number of invocations of

CT () which set the r-th to '1' is 1/φ ≈ 1.29. This result also implies that the expected number of contributing nodes for bit r is less than 1/φ. Furthermore, the expected number of invocations of CT () that contribute a '1' to the bits right to the i-th bit (i.e., bits j, where $i < j \leq \eta$) is approximately S/2 i . As an example, the expected number of invocations of CT () that contribute a '1' to the bits right to the r-th bit is approximately 1/φ, which implies that the expected number of contributing nodes for the bits to the right of the r-th bit is less than 1/φ.

In synopsis generation function SG sum , one synopsis out of m synopses is randomly chosen before each invocation of CT () and only the chosen synopsis is updated after that particular invocation of CT ().

IV CONCLUSION AND FUTURE SCOPE

Securing Node Capture Attacks for Hierarchical Data Aggregation in Wireless Sensor Networks is proposed in this paper. Initially in data aggregation, the aggregator first identify the detecting nodes then selects random set of nodes and broadcast a unique value containing authentication keys, to the selected set of nodes. When node

ISSN (O) 2454-9665
ISSN (P) 2454-0663

from set wants to send the data, it sends fragments of data to other nodes in the same set, encrypted with their respective authentication keys. After receiving data, node decrypts data, sums up the fragments and sends 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. In the next round of aggregation, reselection of set of nodes with new set of authentication keys is done. By simulation results, we have shown that the proposed approach rectifies the security threat of node capture attacks in hierarchical data aggregation.

REFERENCES

[1] M. Liu, N. Patwari, and A. Terzis, "Scanning the is-sue," Proc. IEEE, vol. 98, no. 11, pp. 1804–1807, Apr. 2010.

[2] T. Ko, J. Hyman, E. Graham, M. Hansen, S. Soatto, and D. Estrin, "Embedded imagers: Detecting, localiz-ing, and recognizing objects and events in natural habi-tats," Proc. IEEE, vol. 98, no. 11, pp. 1934–1946, Nov. 2010.

[3] P. Corke, T. Wark, R. Jurdak, W. Hu, P. Valencia, and D. Moore, "Environmental wireless sensor networks," Proc. IEEE, vol. 98, no. 11, pp. 1903–1917, Nov. 2010.

[4] (2006). James Reserve Microclimate and Video Remote Sensing [Online]. Available: http://research. cens.ucla.edu/projects/2006/terrestrial/microclimate/ defau%lt.htm.

[5] S. Madden, M. J. Franklin, J. Hellerstein, and W. Hong, "TAG: A tiny aggregation service for ad hoc sen-sor networks," in Proc. 5th USENIX Symp. Operating Syst. Des.Implement., 2002, pp. 1–3.

[6] J. Zhao, R. Govindan, and D. Estrin, "Computing aggregates for monitoring sensor networks," in Proc. 2nd Int. Workshop Sensor Netw. Protocols Appl., 2003, pp. 139–158..

[7] J. Considine, F. Li, G. Kollios, and J. Byers, "Approxi-mate aggregation techniques for sensor databases," in Proc. IEEE 20th Int. Conf. Data Eng. (ICDE), 2004, pp. 449–460.

[8] S. Nath, P. B. Gibbons, S. Seshan, and Z. Anderson, "Synopsis diffusion for robust aggregation in sensor networks," in Proc. 2nd Int. Conf. Embedded Netw. Sensor Syst. (SenSys), 2004, pp. 250–262.

[9] M. Garofalakis, J. M. Hellerstein, and PManiatis, "Proof sketches: Verifiable in-network aggregation," in Proc. 23rd Int. Conf. Data Eng. (ICDE), 2007, pp. 996– 1005. [10] Y. Yang, X. Wang, S. Zhu, and G. Cao, "SDAP: A se-cure hop-byhop data aggregation protocol for sensor networks," in Proc. ACM MOBIHOC, 2006, pp. 356– 367.