A Secure Composite Data Algorithmic Scheme for Wireless Sensor Network in Presence of Illegitimate Attacks

[1] M. Shamini, [2] D. Kalaiyarasi
[1]PG Scholar ,[2]Associate Professor,
Department of Electronics and Communication Engineering
Panimalar Engineering College,Chennai, India

Abstract— **Composite of data from multiple sensor nodes is usually done by simple methods such as averaging or, more sophisticated, iterative filtering methods. However, such aggregation methods are highly vulnerable to malicious attacks where the attacker has knowledge of all sensed values and has ability to alter some of the readings. In this work, we develop and evaluate algorithms that eliminate or minimize the influence of altered readings. The basic idea is to consider altered data as outliers and find algorithms that effectively identify altered data as outliers and remove them. Once the outliers have been removed, use some standard technique to estimate a true value. Thus, the proposed composite data algorithm operates in two phases: removal of outliers and computation of an estimated true value from the remaining sensor data. Extensive evaluation of the proposed algorithms shows that they significantly outperform all existing methods.**

I. INTRODUCTION

The present level of processes automation requires extensive use of various sensors. Due to unreliability of sensors they are deployed redundantly. Data from multiple sensor nodes is accumulated and combined by an aggregator node. An aggregator node not only collects readings from sensors, but also minimizes or eliminates the influence of readings from faulty or compromised sensors. Secure data aggregation algorithms for sensor networks aim to provide mechanisms for eliminating or resisting data distortion. These algorithms are usually run on an aggregator node or a base station.

Data Aggregation Methods: Probably the earliest and easiest method of data aggregation is simple averaging of readings from all sensors. However, simple averaging method has some major drawbacks, because it does not consider the existence of bias errors or faulty sensors and not to mention malicious attacks. Only one faulty sensor may reduce the accuracy of aggregated result significantly. In addition, the method does not verify any sensor's reading. This makes the method highly vulnerable even to a simple attack, where the attacker skews reading of one or more sensors to a certain degree to alter estimated reading. Iterative Filtering (IF) algorithms offer refined approaches They initially assign one weight to the reading of each sensor and then weights are recalculated at each iteration based on the distance of the readings from the estimated value obtained in the previous iteration. They reduce the effect of a simple attack. But the weakest point of these iterative filtering algorithms is the use of a predetermined procedure for assigning an initial weight to each sensor's reading. An iterative algorithm is vulnerable to a malicious attacker who has knowledge of all readings and has power to alter two or more readings. A malicious attacker can force the iterative filtering algorithm to converge to a desired value by altering readings of the compromised sensors. To overcome weakness of the iterative filtering algorithm.

A Robust Data Aggregation Method (RDAM) was proposed in. The main idea of the algorithm is to estimate a set of non- equal initial weights for the readings. The objective of the method is to calculate smaller initial weights for the readings of the compromised sensors. Results of extensive empirical evaluation of the RDAM method against other methods demonstrated highest accuracy for both simple and illegitimate attacks. However, any estimation of mean and standard deviation is extremely sensitive to presence of outliers. Thus, any aggregation algorithm that estimates true values from sensor readings before removing outliers is susceptible to errors. Our extensive evaluations discovered that in certain conditions the RDAM is vulnerable to a malicious attacker. In this work, we propose and evaluate a set of two phase data aggregation algorithms. The first phase of the proposed method employs a variant of Local Outlier Factor (LOF) calculation method to estimate the degree that an object is an outlier because of collusion attack, sensor fault, noise, or a combination of them. This gives a flexible instrument to exclude suspected sensor-readings before estimation of a true value. Note that this method, unlike methods described previously, removes sensor-readings from compromised or bad sensors, which improve estimated values, while decrease the number of calculations in the second phase of the proposed method. We use a set of different methods, including IF and RDAM, after LOF method. Finally, we introduce a LOF-based non-iterative algorithm that we found most accurate almost in all considered attacks. Moreover, the method works without IF portion, that excludes disadvantages and vulnerabilities related to it. The rest of paper is organized as follows. Section II describes sensor network model used here and reviews related work, including, outliers detection methods. Section III presents our novel algorithms. We describe different combinations of using LOF method that we apply in our algorithms. Section IV shows our experimental results. Finally, the conclusion is provided in Section V.

II. SENSOR NETWORK MODEL AND RELATED WORK

Any sensor reading further from true value are categorized into several classes. 1) Noise: data with greater variance, 2) Spike: data with one or more out- of-bound readings, 3) Stuck-at: data with quasi zero variance, and 4) Corrupted: data altered by malicious attackers. Because

sensor networks often operate in unattended environments and are deployed distributive, they are highly susceptible to failure and physical attacks

A. Sensor Network Model

The sensor network topology used for our work is an abstract model proposed in . A sensor network is built of a base station and a set of sensor clusters. Each cluster has a Composite node (also known as cluster head) that gathers data from all sensor-nodes connected to it. The main functions of an aggregator node are collecting data from its sensor nodes, aggregating the raw data to produce an estimated reading, and communicating the processed data to the base station. Each sensor node has a micro-controller with one or more sensors. The micro-controller is equipped with relatively small memory and computing power, while an aggregator has bigger memory and higher computing power. A base station has larger memory and computing power, in addition to communication Capabilities. It is assumed that the aggregator nodes and the base station are not compromised. After receiving readings from multiple sensors, an aggregator applies a filtering technique, e.g. Iterative Filtering, for reducing the contributions of reading from unreliable sensors.

Problem Statement: Although in general, security can be defined as the combination of availability, confidentiality, and integrity, in our work we focus only on integrity. Any sensor reading further from true value is considered as an outlier. Filtering algorithms attempt to reduce influence of outliers by reducing weights on them. However, filtering techniques cannot overcome a malicious attack, if and when the attacker has knowledge of all readings, details about filtering technique, and capability to alter some of the readings.

Enhanced Iterative Filtering Algorithm : To reduce influence of collusion attack, Rezvani et al. proposed a Robust Data Aggregation Method (RDAM), which is an improvement to IF method. In this method, unequal initial weights are calculated using the readings available to the aggregator. This enhancement not only makes an IF algorithm collusion attack resistant, but it also reduces the number of required iterations to converge to an estimated value. The method is based on the assumption that the distribution of stochastic components of sensor errors is known or can be estimated. Next we describe Algorithms useful for outlier detection

B. Outlier Detection Algorithms

In general, outlier detection methods form five main classes: statistical, nearest neighbor, clustering, classification, and spectral decomposition . In statistical methods outliers are defined as objects that do not fit in assumed distribution. Nearest neighbor methods use a distance as a mean to distinguish outliers Clustering algorithms use similarity metrics. Most popular and probably old clustering algorithm is k-means. The newest and popular for detecting outliers in big data is local outlier method.

Local Outlier Factor: The Local Outlier Factor (LOF) algorithm proposed by Breunig M. et al. for finding anomalous objects . The idea of LOF is based on the concept of local density. The algorithm requires the user to provide only one parameter k, the minimum number of local neighbors of an object to be considered for computing LOF.

III. PROPOSED TWO PHASE COMPOSITE DATA ALGORITHMS

Logically and functionally our algorithms consist of two phases: 1) Detection and removal of outliers, and 2) True value estimation with remaining data.

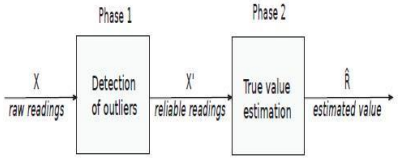

Figure 1. Two phases of our algorithms

A. Detection of Outliers

Following the logic of the construction of our algorithms, in this section we describe method that we apply in the first phase of our algorithms. This phase is dedicated to the detection of outliers.

Detection of Outliers Iteratively: A variation of the previous method, where one outlier with the maximum LOF value is removed at each iteration. At the beginning of each iteration, LOF values of the remaining data set is calculated and the one data point with highest LOF is removed. For space Limitation we omit the details. Interested reader can find the details from. Detection of Outliers using Row-Wise Votes: Two outlier detection algorithms presented in previous sections, compute an average of all readings from a sensor. Then outlier detection algorithms identify outliers from these averages. The algorithm described in this section computes LOFk(x(t)s) for every Observation x(t) s of sensor s. The average LOF(x (t)) value is calculated for every vector of observation x (t). For those x (t) s whose LOF values indicate them as reliable corresponding elements of votes [m, n] array are assigned 1, otherwise it is assigned 0. This gives a table of votes with average values of all votes for each sensor, which is referred to reliability. Sensors that have reliability greater than a certain threshold are considered as reliable. The advantage of this method is that it allows to set a threshold for excluding outliers; a higher threshold will eliminate readings from a larger number of sensors, but the readings from remaining sensors is expected to be highly reliable. The data shown in the Table III identify s1, s3, s4, s5 as reliable, if threshold is set to 0.5. But for same data only sensors s1, s3, s4 are found to be reliable data, if the threshold is set to 0.8. We omit pseudo code of the algorithm for conserving space.

TABLE I. VOTES TABLE

	Sensors				
Instant	S_1	S_2	S_3	S_4	S_5
t = 1	1	0	1	1	1
t = 2	1	0	1	1	0
t = 3	1	0	1	0	0
t = 4	1	1	1	1	1
t = 5	1	0	1	1	1
Reliability	1	0.2	1	0.8	0.6

IV. EXPERIMENTAL RESULTS

The experimental evaluation of the proposed algorithms= aims to show efficacy in the presence of a collusion attack. Since the RDAM performs best in the presence of collusion attack, we use this algorithm's performance as a benchmark for measuring that of the proposed algorithms. It is assumed that an attacker compromises less than half of the n sensor nodes. This is

done with help of ns2 software. Figure 2 shows the initialization of the 12 sensor nodes, 4 cluster heads and one base station with respect to x and y axis.

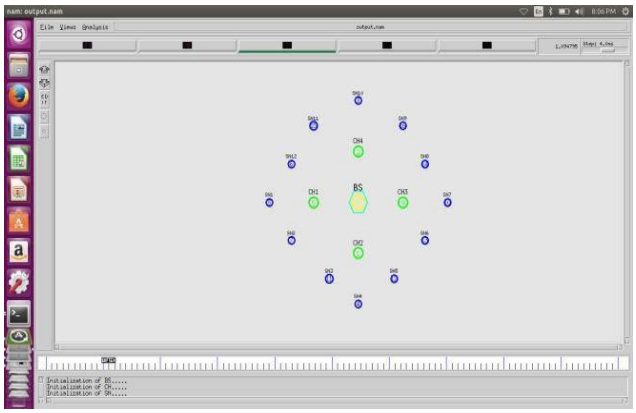

Figure 2. Initialization of sensor nodes with a base station

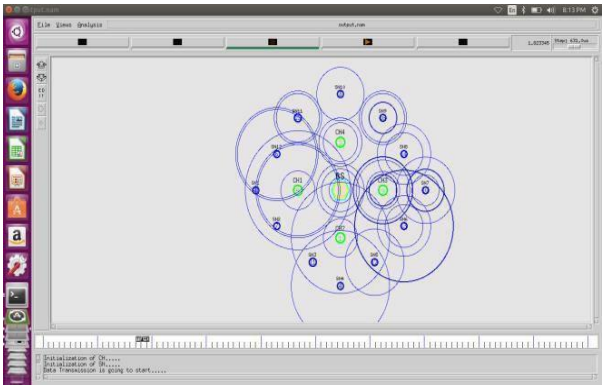

Figure 3. Shows the transmitting of data from sensor nodes to cluster heads and from cluster heads to base station.

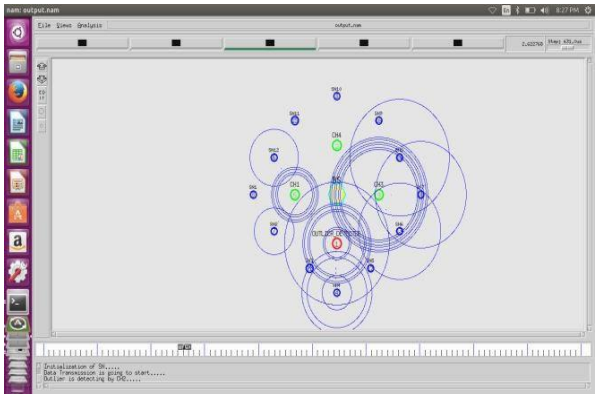

Figure 4. Show the detection of outlier with the help of sensor node behavior.

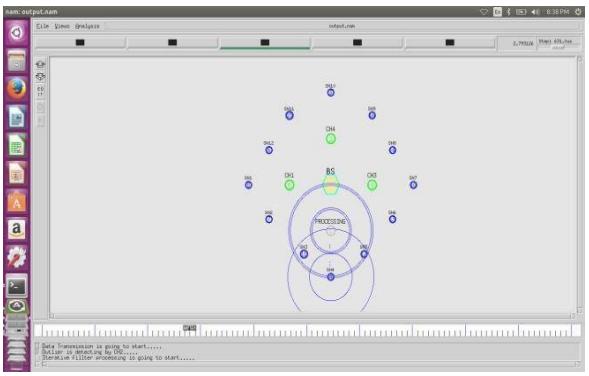

Figure 5. Shows the detected outlier and iterative filtering method using RDMA algorithm.

V. CONCLUSION

In this work, first we developed outlier detection methods using local outlier factor. Then we used these methods for several two phase data aggregation algorithms. The main feature of our algorithms is detection and removal of outliers before estimation true value. First, this increases the accuracy by removing the influence of outliers in aggregated result. Second, having only reliable data true values can be estimated.

Future research will be directed towards experimental verification of the obtained Life time results and enhancing the SSC evaluation methods to be more consistent with practical sensor model.

REFERENCES

[1] M. Rezvani, A. Ignatovich, E. Bertino, and S. Jha, Secure Data Aggregation Technique for Wireless Sensor Networks in the Presence of Collusion Attacks. IEEE Transactions on Dependable and Secure Computing, January/February 2015, vol. 12.

[2] C. de Kerchove and P. Van Dooren, "Iterative filtering in reputation systems," SIAM. J. Matrix Anal. Appl., vol. 31, no. 4, pp. 1812– 1834, 2010.

[3] M.Abou-Nasr, Real world data mining applications. Springer Publishing Company, Incorporated, 2014.

[4] V. Barnett and T. Lewis, Outliers in statistical data. John Wiley, 1994.

[5] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J.Sander, LOF: Identifying density-based local outliers. Dalles, Texas, USA: ACM, 2000.

[6] K. Ni, N. Ramanathan, M. N. H. Chehade, L. Balzano, S. Nair, S. Zahedi, E. Kohler, G. Pottie, M. Hansen, and M.Srivastava, "Sensor network data fault types," ACM Trans. Sen. Netw., vol. 5, no. 3, pp. 25:1–25:29, Jun. 2009.

[7] D.Wagner,"Resilient aggregation in sensor networks," in Proceedings of the 2nd ACM Workshop on Security of Ad Hoc and Sensor Networks, ser. SASN '04. New York, NY, USA: ACM, 2004, pp. 78–87.

[8] E.Ayday and F. Fekri, "Iterative trust and reputation management using belief propagation," IEEE Transactions on Dependable and Secure Computing, vol. 9, no. 3, pp. 375–386, 2012.