ITERATIVE FILTERING ALGORITHM BASED ON ROBUST DATA AGGREGATION METHOD FOR WIRELESS SENSOR NETWORK IN THE PRESENCE OF ADVERSARY ENVIRONMENT

1UMA ANGADI, 2G.F ALI AHAMMED

1Dept. of Computer Science &Engg, VTU PG Centre, Mysuru, Karnataka, INDIA
2Associate Professor, Dept. of Computer Science &Engg, VTU PG Centre, Mysuru, Karnataka, INDIA
E-mail: umaangadiu@gmail.com

Abstract— As we have limited computational power and energy resources, aggregation of data from the multiple sensor nodes is done at the aggregator node is usually accomplished by simple method is averaging. WSNs are usually unattended and without tamper resistant hardware, they are highly vulnerable to such node compromising attacks. Thus, making it necessary to ascertain data trustworthiness and reputation of sensor nodes is crucial for WSN. Iterative filtering algorithms were found out to be very helpful in this purposes. Such algorithms provide the aggregate the data from multiple data from multiple sources and also provide the trust assessment of these sources, usually in a form corresponding weight factors assigned to data provided by each source. In this paper demonstrates the several existing iterative filtering algorithms, while significantly more robust against the collusion attacks than the simple averaging method. These algorithms are susceptible to the most sophisticated collusion attack scenario presented in this paper, we propose an improvement for iterative filtering techniques by providing initial approximation for these algorithms which make them not only collusion robust, but also get more accurate and faster converging.


I. INTRODUCTION

Wireless sensor networks (WSNs) are composed of sensor nodes with the capabilities of sensing, communication and computation. For that we have a tendency to apply data aggregation technique on the detected data by the sensor hubs. Due to robustness of monitoring and low cost of the nodes, wireless sensor networks (WSNs) are usually redundant. Data from multiple sensors is aggregated at an aggregator node which then forwards to the base station only the aggregate values. At present, due to limitations of the computing power and energy resource of sensor nodes, data is aggregated by extremely simple algorithms such as averaging. However, such aggregation is known to be very vulnerable to faults, and more importantly, malicious attacks [1]. This cannot be remedied by cryptographic methods, because the attackers generally find complete access to information stored in the compromised nodes. For that reason data aggregation at the aggregator node has to be accompanied by an assessment of trustworthiness of data from individual sensor nodes. Thus, better, more sophisticated algorithms are needed for data aggregation in the future WSN. Such an algorithm should have two features.

1. In the presence of stochastic errors such algorithm should produce estimates which are close to the optimal ones in information theoretic sense. Thus, for example, if the noise present in each sensor is a Gaussian independently distributed noise with zero mean, then the estimate produced by such an algorithm should have a variance close to the Cramer-Rao lower bound (CRLB) [2], i.e., it should be close to the variance of the Maximum Likelihood Estimator (MLE). However, such estimation should be achieved without suppling to the algorithm the variances of the sensors unavailable in practice.

2. The algorithm should also be robust in the presence of non-stochastic errors, such as faults and malicious attacks, and, besides aggregating data, such algorithm should also provide an assessment of the reliability and trustworthiness of the data received from each sensor node. The main goal of data aggregation method to gather and aggregate data in any energy efficient manner so that network lifetime is enhanced.

Trust and reputation system have a significant role in supporting operation of a range of distributed systems from wireless sensor networks and e-commerce infrastructure to social networks, by providing an assessment of trustworthiness of participants in such distributed systems. A trustworthy assessment at any given moment represents an aggregate of the behavior of the participants up to that moment and has to robust in the presence of various attacks and malicious users. There are number of incentives for attackers to manipulate the trust and reputation scores of participants in a distributed system. The main target of the malicious attackers are aggregation algorithms of trust and reputation systems. Iterative Filtering algorithms are an attractive for WSNs because they solve the data aggregation and data trustworthiness assessment using single iterative procedure. Such trustworthiness estimate of each sensor is biased on the distance of the readings of such a sensor from the...
estimate of the correct values obtained in the previous round of iteration by some from of aggregation of the readings significantly differ from such estimate are assigned less trustworthiness and consequently in the aggregation process in the present round iteration their readings are given a lower weight. If the attackers have a high level of knowledge about the aggregation algorithm and its parameters, they can conduct the sophisticated attacks on WSNs by exploiting false data injection through a number of compromised nodes. This paper presents a new sophisticated collusion attack scenario against a number of existing IF algorithms based on the false data injection. In such an attack scenario, colluders attempt to skew the aggregate value by forcing such IF algorithms to converge to skewed values provided by one the attackers.

II. RELATED WORK

In this paper, robust data aggregation is a serious concern in WSNs and there are a number of papers investigating malicious data injection by taking into account the various adversary models. There are three bodies of work related to our research: IF algorithms, trust and reputation systems for WSNs, and secure data aggregation with compromised node detection in WSNs.

There are a number of published studies introducing IF algorithms for solving data aggregation problem [1], [2], [3], [4]. Ayday et al. proposed a slight different iterative algorithm in [3]. Their main differences from the other algorithms are: 1) the ratings have a time-discount factor, so in time, their importance will fade out; and 2) the algorithm maintains a blacklist of users who are especially bad raters. Liao et al. in [4] proposed an iterative algorithm which beyond simply using the rating matrix, also uses the social network of users. The main objective of Chen et al. in [5] is to introduce a “Bias smoothed tensor model”, which is a Bayesian model of rather high complexity. Although the existing IF algorithms consider simple cheating behaviour by adversaries, none of them take into account sophisticated malicious scenarios such as collusion attacks.

Xiao et al. in [6] proposed a trust-based framework which employs correlation to detect faulty readings. Moreover, they introduced a ranking framework to associate a level of trustworthiness with each sensor node based on the number of neighboring sensor nodes are supporting the sensor. Li et al. in [7] proposed PRESTO, a model-driven predictive data management architecture for hierarchical sensor networks. PRESTO is a two tier framework for sensor data management in sensor networks. The main idea of this framework is to consider a number of proxy nodes for managing sensed data from sensor nodes. Lim et al. in [6] proposed an interdependency relationship between network nodes and data items for assessing their trust scores based on a cyclical framework. Reputation and trust concepts can be used to overcome the compromised node detection and secure data aggregation problems in WSNs.

Ho et al. in [8] proposed a framework to detect compromised nodes in WSN and then apply a software attestation for the detected nodes. They reported that the revocation of detected compromised nodes can not be performed due to a high risk of false positive in the proposed scheme. The main idea of false aggregator detection in the scheme proposed in [9] is to employ a number of monitoring nodes which are running aggregation operations and providing a MAC value of their aggregation results as a part of MAC in the value computed by the cluster aggregator. High computation and transmission cost required for MAC-based integrity checking in this scheme makes it unsuitable for deployment in WSN.

III. OBJECTIVE

- The main objective of this project is introduced a novel collusion attack scenario against a number of existing IF algorithms.
- To improve the IF algorithms by providing an initial approximation of the trustworthiness of sensor nodes which makes the algorithms not only collusion robust, but also more accurate and faster converging.

IV. METHODOLOGY

A. Network Model

For sensor network topology, we consider the abstract model proposed by Wagner in [4], is considered Fig. 1 shows our network model assumption. The sensor nodes are divided into disjoint clusters, and each cluster has cluster head which act as an aggregator. In this paper we assume that the aggregator node itself is not compromised which makes data aggregation more secure when individual sensor nodes might be compromised and might be sending false data to the aggregator.

B. Iterative Filtering in reputation systems

Kerchove and Van Dooren proposed in [8] an IF algorithm for computing the reputation of objects and raters in a rating system. We briefly describe the

---

Iterative Filtering Algorithm Based On Robust Data Aggregation Method For Wireless Sensor Network In The Presence Of Adversary Environment
algorithm in the context of data aggregation in WSN and explain the vulnerability of the algorithm for a possible collusion attack. We note that our improvement is applicable to other IF algorithms as well.

**Algorithm**: Iterative filtering algorithm.

**Output**: The reputation vector $r$

$\begin{align*}
&l \leftarrow 0; \\
&w(0) \leftarrow 1; \\
&\text{Repeat} \\
&\text{Compute } r(l+1) \\
&\text{Compute } d; \\
&\text{Compute } w(l+1) \\
&l \leftarrow l + 1 \\
&\text{until reputation has converged;}
\end{align*}$

- Where, we consider WSN with $n$ sensors and we assume that aggregator work on one block of reading at a time each block consists of reading from $m$ consecutive instants. Therefore the block of reading is represented by matrix $X$.

- $r$ denotes the aggregated values it is called as a reputation vector computed with the sequence of weight $w$.

- The iterative procedure starts with giving equal credibility to all the sensors with initial value $w(0)$.

- The value of the reputation vector $r(l+1)$ in round of iteration $l + 1$ is obtained from the weights of the sensors obtained in the round of iteration $l$.

- The new weight vector $w(l+1)$ to be used in round of iteration $l + 1$ is then computed as a function $g(d)$ of the normalized belief divergence $d$ is the distance between the sensor reading and reputation vector $r(l)$.

This algorithm is the iterative filtering algorithm which is vulnerable to collusion attacks improvement to this algorithm is also applicable to other IF algorithms as well. We show that the algorithm converges in little iteration.

### C. Adversary Model

Sensors are deployed in a hostile unattended environment. In this situation some sensor nodes can be easily compromised by the adversary. Through the compromised node the adversary can send false data to the aggregator with a purpose of distorting the aggregate values. We assume that all compromised nodes can be under control of a colluding group of adversaries. We also assume that the adversary has enough knowledge about the aggregation algorithm and its parameters. The following collusion attack scenario describes how the adversary attempt to skew the aggregate values. An adversary compromise three nodes in the networks in order to launch a collusion attack. It instructs the two compromised nodes to insert aggregate value as far from the true value. Then it computes the skewed value of the simple average of all sensor readings. The third compromised node reports value that is very close to the true value. Thus the IF algorithms quickly converges to the value provided by the third compromised node, because it assigns an equal initial weight to all sensors in the first iteration. To overcome this type of shortcoming we estimate an initial trust vector based on error parameters of sensor nodes. Thus the new trust vector is used to consolidate the algorithms against collusion attack.

### D. Collusion attack scenario

Most of the IF algorithms occupy simple assumptions about the initial values of weights for sensors. In case of our opponent model, an attacker is able to misinform the aggregation system from side to side cautious range of report data standards. Assume that ten sensors report the values of temperature which are aggregated using the IF algorithm planned in with the reciprocal discriminated function.

In scenario 1, all sensors are reliable and the result of the IF algorithm is close to the actual value. In scenario 2, an adversary compromises two sensor nodes, and alters the readings of these values such that the simple average of all sensor readings is skewed towards a lower value. As these two sensor nodes report a lower value, IF algorithm penalizes them and assigns them lower weights, because their values are far from the values of the sensors. The algorithm assigns very low weights to these two sensor nodes and consequently their contributions decrease.

In scenario 3, an adversary employs three compromised nodes in order to launch a collusion attack. It listens to the reports of sensors in the network and instructs the two compromised sensor nodes to report values far from the true value of the measured quantity. It then computes the skewed value of the simple average of all sensor readings and commands the third compromised sensor to report such skewed average as its readings.

![Fig.2 Attack scenario against IF algorithm.](Image)

### V. PROBLEM STATEMENT
In the prior work there is techniques are proposed to attain the reachable data security while the aggregation but there is not satisfying the data security and integrity or the accuracy of the process while they process with the sensor data after storing into the aggregator. Thus, we cannot rely on cryptographic methods for preventing the attacks, since the adversary may extract cryptographic keys from the compromised nodes.

VI. PROPOSED SYSTEM

The proposed system is mainly to avoid the attacks availability on the each sensor nodes reading. An improvement is made on iterative filtering technique by providing an initial approximation which not only makes the algorithm collusion robust, but also faster converging. Iterative Filtering algorithms are an efficient and reliable option for wireless sensor networks because they solve both problems of data aggregation and data trustworthiness estimation using a single iterative procedure. This algorithm is for robust aggregation along with which different collusion attacks are identified and avoided in the proposed system. These attacks are described by estimating sensor’s error and uses MLE for robust aggregation. The trustworthiness of nodes is estimated from their data aggregated from them. The computational cost is also reduced by the proposed method.

A. Proposed system architecture

In the wireless sensor network consists of sensor nodes these sensor nodes are scattered then deployed environment in the network and then to form the cluster ,each cluster has a cluster head and then data send to the aggregator node before sending base station to verify the data if any error in the data then to estimate the value using parameters such as bias and variance and also estimate MLE using iterative filtering algorithm. The proposed system architecture view can be shown in Fig 3.

![Proposed system architecture](image)

**Fig.3. Proposed system architecture.**

B. Enhanced Iterative Filtering Algorithm

IF algorithm is robust against the simple outlier injection by the compromised nodes. An adversary employs three compromised nodes in order to launch a collusion attack. It listens to the reports of sensors in the network and instructs the two compromised sensor nodes to report values far from the true value of the measured quantity. It then computes the skewed value of the simple average of all sensor readings and commands the third compromised sensor to report such skewed average as its readings. In other words, two compromised nodes distort the simple average of readings, while the third compromised node reports a value very close to such distorted average thus making such reading appear to the IF algorithm as highly reliable reading. As a result, IF algorithms will meet to the values provide by the third compromised node, because in the first iteration of the algorithm the third compromised node will achieve the highest influence, radically dominate the weights of all other sensors. Initial test vector based on the IF method provide a robust nature of the security system.

VII. ROBUST DATA AGGREGATION

A. Robust Data Aggregation Framework

Robust Data Aggregation model is operates on batches of consecutive readings of sensors, proceeding in several stages. In the first stage provide an initial estimate of two noise parameters for sensor nodes, bias and variance details of the computations for estimating bias and variance of sensors. A novel approach for estimating the bias and variance of noise for sensors based on their readings. The variance and the bias of a sensor noise can be interpreted as the distance measures of the sensor readings to the true value of the signal. In fact, the distance measures obtained as our estimates of the bias and variances of sensors also make sense for non-stochastic errors. Based on such an estimation of the bias and variance of each sensor, the bias estimate is subtracted from sensors readings and in the next phase of the proposed framework, we provide an initial estimate of the reputation vector calculated using the MLE as shown in Fig 4.

B. Bias Estimation

All sensors may have some errors in their readings. Such error is denoted as \( e^s_t \) of sensor \( s \) and it is modelled by the Gaussian distribution random variable with bias \( b^s \) and variance \( \sigma^s \). Let \( r_s \) denotes the true value of the sensor at time \( t \). Sensor readings \( x^s_t \) can be written as

\[
x^s_t = r_s + e^s_t
\]

(1)

Since there is no true value, the error value of sensor \( s \) is not to be found. But the difference values of such
sensors are calculated with the equation given below. Let \( \delta = \delta(i,j) \) be an estimator for mutual difference of sensor bias.

\[
\delta(i,j) = \frac{1}{m} \sum_{l=1}^{m} e_l^i - e_l^j = \frac{1}{m} \sum_{l=1}^{m} e_l^i - \frac{1}{m} \sum_{l=1}^{m} e_l^j
\]  

(2)

Let \( a_j = \frac{1}{m} \sum_{l=1}^{m} e_l^j \) be the sample mean of the random variable and \( m \) be the number of readings for each sensor. Then the expected value is calculated by minimizing the obtained value with respect to the mean value and the equation is given below

\[
\delta(i,j) = a_i - a_j \approx b_i - b_j
\]  

(3)

C. Variance Estimation

With the known values of bias estimated from the equation 3 the variance of sensor errors are calculated. Each sensor bias value is substracted from the sensor readings. By using the error difference value from the equation 2 we can get the variance value as a squared difference of each sensor error and the bias value. This varies up to the last sensor reading and is defined as

\[
\beta(i,j) = \frac{1}{m-1} \sum_{l=0}^{m-1} (e_l^i - b_i)^2 + \frac{1}{m-1} \sum_{l=0}^{m-1} (e_l^j - b_j)^2
\]  

(4)

Where \( \sigma_i^2 = v_i \) is the variance of sensor from the matrix \( \beta = \{ \beta(i,j) \} \).

D. Maximum Likelihood Estimation

The unbiassing sensor readings are extracted and take place with help of the bias estimated result which is calculated from the above section. After that the variance estimated result from equation 4 is considered and the extracted unbiassing sensor is used to make the maximum likelihood estimation with variance value. By differentiating the likelihood function the true values are obtained and are measured in the form of weighted average. It is defined as

\[
r = \sum_{i=1}^{n} w_i x_i
\]  

(5)

Thus it estimates the reputation vector without any iteration. Hence the computational complexity of the estimation is less than the existing IF algorithms.

E. Iterative Filtering

For the proposed collusion attack the results from the above is considered as an initial reputation for this filtering. It estimates the trustworthiness of each sensor based on the distance of sensors readings. From this process the estimation is made with an initial level itself. Using this initial reputation the efficiency of the IF algorithm is improved and reduces the required number of iterations.

**CONCLUSION AND FUTURE WORK**

In this paper, introduced the novel collusion attack scenario is considered against the number of prior IF algorithms. An initial approximation of the trustworthiness of sensor nodes is proposed for an improvement of the IF algorithm. Furthermore, this have a novel data collection technique from the sensor reading data in the presence of the collusion attack and it prevents the sensor faults. The whole performance will be evaluated in terms of time consumption. It makes the IF algorithms not only collusion robust but also gives more accurate and faster. In the future we will extend the proposed robust aggregation for protecting the aggregator node from the collusion attack and also plan to improve the data security while transmitting the data from the sensor node to the aggregator node. We also plan to implement our approach in a deployed sensor network.

**REFERENCES**