Secure Data Aggregation Technique for Wireless Sensor Networks in the Presence of Collusion Attacks

Mr. S.Rajadurai[1], Mr.D.Vinod[2]
[1]PG Student, [2]Professor
[1 2] Department of CSE, [1 2] Sri Ramanujar Engineering College, Tamil Nadu, India.

Abstract-- Due to limited computational power and energy resources, aggregation of data from multiple sensor nodes done at the aggregating node is usually accomplished by simple methods such as averaging. However such aggregation is known to be highly vulnerable to node compromising attacks. Since WSN are usually unattended and without tamper resistant hardware, they are highly susceptible to such attacks. Thus, ascertaining trustworthiness of data and reputation of sensor nodes is crucial for WSN. As the performance of very low power processors dramatically improves, future aggregator nodes will be capable of performing more sophisticated data aggregation algorithms, thus making WSN less vulnerable. Iterative filtering algorithms hold great promise for such a purpose. Such algorithms simultaneously aggregate data from multiple sources and provide trust assessment of these sources, usually in a form of corresponding weight factors assigned to data provided by each source. In this paper we demonstrate that several existing iterative filtering algorithms, while significantly more robust against collusion attacks than the simple averaging methods, are nevertheless susceptive to a novel sophisticated collusion attack we introduce. To address this security issue, we propose an improvement for iterative filtering techniques by providing an initial approximation for such algorithms which makes them not only collusion robust, but also more accurate and faster converging.

Keywords: *Distributed detection, distributed hash table, node clone attack, randomly directed exploration, wireless sensor networks (WSNs)..*

I. INTRODUCTION

Due to a need for robustness of monitoring and low costof 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. This cannot be remedied by cryptographic meth-ods, because the attackers generally gain complete access to information stored in the compromised nodes. For that rea-son data aggregation at the aggregator node has to be accompanied by an assessment of trustworthiness of data from individual sensor nodes. Thus, better, more sophisti-cated algorithms are needed for data aggregation in the future WSN. Such an algorithm should have two features.

In the presence of stochastic errors such algorithm should produce estimates which are close to theoptimal 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 algo-rithm should have a variance close to the Cramer-Rao lower bound (CRLB), i.e, it should be close to the variance of the Maximum Likelihood Estimator (MLE). However, such estimation should be achieved without supplying to the algorithm the var-iances of the sensors, unavailable in practice.

The algorithm should also be robust in the presence of non-stochastic errors, such as faults and malicious attacks, and, besides aggregating data, such algo-rithm should also provide an assessment of the reli-ability and trustworthiness of the data received from each sensor node.

Trust and reputation systems have a significant role in supporting operation of a wide range of distributed sys-tems, from wireless sensor networks and e-commerce infrastructure to social networks, by providing an assess-ment of trustworthiness of participants in such distributed systems. A trustworthiness assessment at any given moment represents an aggregate of the behaviour of the participants up to that moment and has to be robust in the presence of various types of faults and malicious behav-iour. There are a number of incentives for attackers to manipulate the trust and reputation scores of participants in a distributed system, and such manipulation can severely impair the performance of such a system. The main target of malicious attackers are aggregation algo-rithms of trust and reputation systems.

Trust and reputation have been recently suggested as an effective security mechanism for Wireless Sensor Net-works. Although sensor networks are being increas-ingly deployed in many application domains, assessing trustworthiness of reported data from distributed sensors has remained a challenging issue. Sensors deployed in hostile environments may be subject to node compromis-ing attacks by adversaries who intend to inject false data into the system. In this context, assessing the trust-worthiness of the collected data becomes a challenging task.

As the computational power of very low power processors dramatically increases, mostly driven by demands of mobile computing, and as the cost of such technology drops, WSNs will be able to afford hardware which can implement more sophisticated data aggregation and trust assessment algorithms; an example is the recent emer-gence of multicore and multi-processor systems in sensor nodes.

Iterative Filtering (IF) algorithms are an attractive option for WSNs because they solve both problems—data aggregation and data trustworthiness assessment—using a single iterative procedure. Such trustworthiness esti-mate of each sensor is based on the distance of the read-ings of such a sensor from the estimate of the correct values, obtained in the previous round of iteration by some form of aggregation of the readings of all sensors. Such aggregation is usually a weighted average; sensors whose readings significantly differ from such estimate are assigned less trustworthiness and consequently in the aggregation process in the present round of iteration their readings are given a lower weight.

In recent years, there has been an increasing amount of literature on IF algorithms for trust and reputation sys-tems. The perfor-mance of IF algorithms in the presence of different types of faults and simple false data injection attacks has been studied, for example in where it was applied to com-pressive sensing data in WSNs. In the past literature it was found that these algorithms exhibit better robustness compared to the simple averaging techniques; however, the past research did not take into account more sophisti-cated collusion attack scenarios. If the attackers have a high level of knowledge about the aggregation algorithm and its parameters, they can conduct 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 of the attackers.

Although such proposed attack is applicable to a broad range of distributed systems, it is particularly dangerous once launched against WSNs for two reasons. First, trust and reputation systems play critical role in WSNs as a method of resolving a number of important problems, such as secure routing, fault tolerance, false data detec-tion, compromised node detection, secure data aggrega-tion, cluster head election, outlier detection, etc.,.Second, sensors which are deployed in hostile and unat-tended environments are highly susceptible to node compromising attacks. While offering better protec-tion than the simple averaging, our simulation results demonstrate that indeed current IF algorithms are vulner-able to such new attack strategy.

As we will see, such vulnerability to sophisticated collusion attacks comes from the fact that these IF algorithms start the iteration process by giving an equal trust value to all sensor nodes. In this paper, we propose a solution for such vulnerability by providing an initial trust estimate which is based on a robust estimation of errors of individual sensors. When the nature of errors is stochastic, such errors essentially represent an approximation of the error parameters of sensor nodes in WSN such as bias and variance. However, such estimates also prove to be robust in cases when the error is not stochastic but due to coordinated malicious activities. Such initial estimation makes IF algorithms robust against described sophisticated collusion attack, and, we believe, also more robust under significantly more general circumstances; for example, it is also effective in the presence of a complete failure of some of the sensor nodes. This is in contrast with the traditional non iterative statisti-cal sample estimation methods which are not robust against false data injection by a number of compromised nodes [18] and which can be severely skewed in the presence of a com-plete sensor failure.

Since readings keep streaming into aggregator nodes in WSNs, and since attacks can be very dynamic (such as orchestrated attacks), in order to obtain trustworthiness of nodes as well as to identify compromised nodes we apply our framework on consecutive batches of consecu-tive readings. Sensors are deemed compromised only rel-ative to a particular batch; this allows our framework to handle on-off type of attacks.

We validate the performance of our algorithm by simulation on synthetically generated data sets. Our simulation results illustrate that our robust aggregation technique is effective in terms of robustness against our novel sophisticated attack scenario as well as efficient in terms of the computational cost.

Our contributions can be summarized as follows:[1]

1. Identification of a new sophisticated collusion attack against IF based reputation systems which reveals a severe vulnerability of IF algorithms.
2. A novel method for estimation of sensors' errors which is effective in a wide range of sensor faults and not susceptible to the described attack.
3. Design of an efficient and robust aggregation method inspired by the MLE, which utilises an esti-mate of the noise parameters obtained using contri-bution 2 above.
4. Enhanced IF schemes able to protect against sophisticated collusion attacks by providing an initial estimate of trustworthiness of sensors using inputs from contributions 2 and 3 above.

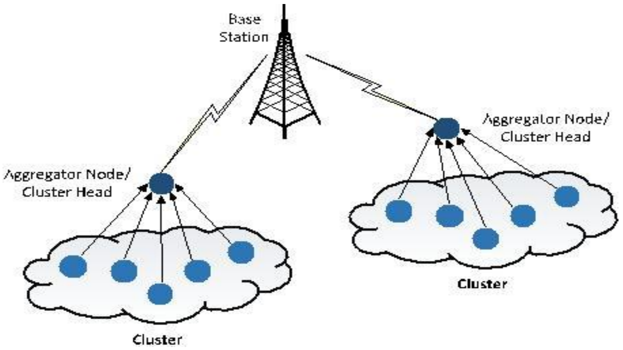

Fig. 1.Network model for WSN.

We provide a thorough empirical evaluation of effectiveness and efficiency of our proposed aggregation method. The results show that our method provides both higher accu-racy and better collusion resistance than the existing methods.

II. RELATED WORK

A. *Network Model*

For the sensor network topology, we consider the abstract model proposed by Wagner in. Fig. 1 shows our assumption for network model in WSN. The sensor nodes are divided into disjoint clusters, and each cluster has a clus-ter head which acts as an aggregator. Data are periodically collected and aggregated by the aggregator. In this paper we assume that the aggregator itself is not compromised and concentrate on algorithms which make aggregation secure when the individual sensor nodes might be compro-mised and might be sending false data to the aggregator. We assume that each data aggregator has enough computa-tional power to run an IF algorithm for data aggregation.

B. *Iterative Filtering in Reputation Systems*

Kerchove and Van Dooren proposed in an IF algo-rithm for computing reputation of objects and raters in a rating system. We briefly describe the algorithm in the context of data aggregation in WSN and explain the vul-nerability of the algorithm for a possible collusion attack. We note that our improvement is applicable to other IF algorithms as well.We consider a WSN with n sensors Si, i¼1;. . .; n. We assume that the aggregator works on one block of read-ings at a time, each block comprising of readings at m.

C. *Adversary Model*

In this paper, we use a Byzantine attack model, where the adversary can compromise a set of sensor nodes and inject any false data through the compromised nodes. We assume that sensors are deployed in a hostile unattended environment. Consequently, some nodes can be physically compromised. We assume that when a sensor node is compromised, all the information which is inside the node becomes accessible by the adversary. Thus, we cannot rely on cryptographic methods for preventing the attacks, since the adversary may extract cryptographic keys from the compromised nodes. We assume that through the compromised sensor nodes the adversary can send false data to the aggregator with a purpose of distorting the aggregate values. We also assume that all compromised nodes can be under control of a single adversary or a colluding group of adversaries, enabling them to launch a sophisticated attack. We also consider that the adversary has enough knowl-edge about the aggregation algorithm and its parameters. Finally, we assume that the base station and aggregator nodes cannot be compromised in this adversary model; there is an extensive literature proposing how to deal with the problem of compromised aggregators; in this paper we limit our attention to the lower layer problem of false data being sent to the aggregator by compromised individual sensor nodes, which has received much less attention in the existing literature.

TABLE 1
A Trace Example of Iterative Filtering Algorithm

	sensor readings								aggregate values		
instant	s1	s2	s3	s4	s5	s6	s7	s8			
t=1	19.3612	19.42	19.0084	18.5674	17.95	22.153	18.0088	20.4			
t=2	19.3612	19.4102	19.0084	18.5478	21.282	21.347	18.0088	20.4098			
t=3	19.3612	19.42	19.0084	17.117	21.3408	20.813	21.625	19.7924			
round#	sensor weights								t=1	t=2	t=3
1	1	1	1	1	1	1	1	1	19.3586	19.6719	19.8097
2	1.01E+01	1.34E+01	2.4896	0.3282	0.4335	0.2581	0.3806	1.8413	19.4008	19.439	19.4318
3	2.38E+02	2.24E+03	5.7843	0.4381	0.328	0.2286	0.3412	1.4486	19.4137	19.4052	19.4139
4	4.01E+02	2.96E+04	6.1705	0.446	0.3199	0.2267	0.3404	1.4116	19.4192	19.4095	19.4192
5	3.31E+02	1.59E+06	6.02	0.4433	0.3206	0.2278	0.3403	1.4273	19.42	19.4102	19.42
6	3.22E+02	6.47E+09	5.9971	0.4428	0.3207	0.2279	0.3402	1.4297	19.42	19.4102	19.42

```
Algorithm 1: Iterative filtering algorithm.
Input: $X, n, m$.
Output: The reputation vector $\mathbf{r}$
$l \leftarrow 0$;
$\mathbf{w}^{(0)} \leftarrow \mathbf{1}$;
repeat
    Compute $\mathbf{r}^{(l+1)}$;
    Compute $\mathbf{d}$;
    Compute $\mathbf{w}^{(l+1)}$;
    $l \leftarrow l + 1$;
until reputation has converged;
```

TABLE 2
A Trace Example of a Simple Attack Scenario

	sensor readings										
instant	s1	s2	s3	s4	s5	s6	s7	s8	s9	s10	
t=1	19.7336	19.6160	19.7728	20.2040	20.4196	19.4494	20.1354	19.0084	**13.2001**	**13.5609**	
round#	sensor weights										t=1
1	1	1	1	1	1	1	1	1	1	1	18.5097
2	6.68	8.17	6.27	3.48	2.74	11.41	3.78	40.21	0.35	0.41	19.3390
3	64.21	130.29	53.13	13.36	8.56	872.81	15.77	91.52	0.27	0.30	19.4811
4	156.81	549.28	117.50	19.13	11.35	8.1E+3	23.36	44.76	0.25	0.29	19.4676
5	141.35	454.18	107.37	18.44	11.03	2.1E+4	22.42	47.42	0.25	0.29	19.4536
6	127.57	379.24	98.16	17.76	10.72	1.7E+5	21.51	50.45	0.26	0.29	19.4468
7	121.61	349.49	94.12	17.44	10.57	1.4E+7	21.09	52.02	0.26	0.29	19.4460
8	120.91	346.06	93.64	17.40	10.55	1.0E+11	21.04	52.22	**0.26**	**0.29**	19.4460

In other words, two compro-mised 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 a highly reliable reading. As a result, IF algorithms will con-verge to the values provided by the third compro-mised node, because in the first iteration of the algorithm the third compromised node will achieve the highest weight, significantly dominating the weights of all other sensors. This is reinforced in every subsequent iteration; therefore, the algorithm quickly converges to a reputation which is very close to the initial skewed simple average, as shown in Fig. 2. Table 3 shows the same attack scenario on Intel Lab data set; sensors 8, 9 and 10 are compro-mised by an adversary. As one can see, the algorithm converges quickly to the readings of sensor 10 which is essentially equal to the simple average value of the sensors.

In the third scenario, how much the aggregate value is skewed directly depends on the number of compromised nodes which distort the sample average of readings. Moreover, in this scenario, the attacker needs to gain con-trol over at least two sensor nodes; one which will reports readings which distort the sample average and another one which reports such distorted average. In our experi-ments, we investigate how the behaviour of the IF algo-rithm depends on the number of compromised nodes.

Clearly, the main source of the above vulnerability comes from the fact that the algorithm assigns an equal initial weight to all sensor nodes in the first iteration. Moreover, the reciprocal discriminant function has a pole at zero which makes the algorithm unstable in the presence of sensors exhibiting a very small belief divergence at any given round of iteration. Therefore, under an attack of the kind described, the reputation value of the first iteration is equal to the simple average of readings, and the second vec-tor of weights is computed based on the distance of each sensor to the simple average provided by the first iteration. As most of the IF algorithms in the literature make the same assumption about the initial trustworthiness of sensors, we argue that an adversary with sufficient knowledge of such algorithms can launch an attack as we have described and deceive the aggregator node.In the case in which the nodes use cryptography to ensure the confidentiality of readings they send to the aggregator, the adversary can still estimate these readings by sensing the measured quantity using the malicious nodes.

To address the shortcoming of existing IF methods, we focus on estimating an initial trust vector based on an esti-mate of error parameters of sensor nodes. After that, we use the new trust vector as the initial sensor trustworthiness in order to consolidate the algorithms against an attack sce-nario of the type described.

III. PROPOSED METHOD

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 mod-els. 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 reviewed three of them in our comparative experiments. Li et al. in proposed six different algorithms, which are all iterative and are similar. The only difference among the algorithms is their choice of norm and aggregation function. Ayday et al. proposed a slight different iterative. Their main differences from the other algorithms are: 1) the ratings have a time-discount factor, so in time, their impor-tance will fade out; and 2) the algorithm maintains a black-list of users who are especially bad raters. Liao et al. in 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 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.

Our work is also closely related to the trust and reputa-tion systems in WSNs. Ganeriwal et al. in proposed a general reputation framework for sensor networks in which each node develops a reputation estimation for other nodes by observing its neighbors which make a trust communication for sensor nodes in the network. Xiao et al. in 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 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 sen-sor nodes.

Lim et al. in proposed an interdependency relationship between network nodes and data items for assessing their trust scores based on a cyclical framework. The main contribution of Sun et al. in is to propose a combination of trust mechanism, data aggregation, and fault tolerance to enhance data trustworthiness in Wireless Multimedia Sensor Networks (WMSNs) which considers both discrete and continuous data streams. Tang et al. in proposed a trust framework for sensor networks in cyber physical systems such as a battle-network in which the sensor nodes are employed to detect approaching ene-mies and send alarms to a command center. Although fault detection problems have been addressed by applying trust and reputation systems in the above research, none of them take into account sophisticated collusion attacks scenarios in adversarial environments.

Reputation and trust concepts can be used to overcome the compromised node detection and secure data aggrega-tion problems in WSNs. Ho et al. in proposed a frame-work 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 pro-posed scheme. The main idea of false aggregator detection in the scheme proposed in is to employ a number of monitoring nodes which are running aggregation opera-tions and providing a MAC value of their aggregation results as a part of MAC in the value computed by the clus-ter aggregator. High computation and transmission cost required for MAC-based integrity checking in this scheme makes it unsuitable for deployment in WSN. Lim et al. in proposed a game-theoretical defense strategy to protect sensor nodes and to guarantee a high level of trustworthi-ness for sensed data. Moreover, there is a large volume of published studies in the area of secure tiny aggregation in WSNs. These studies focus on detecting false aggregation operations by an adversary, that is, on data aggregator nodes obtaining data from source nodes and producing wrong aggregated values. Consequently, they address neither the problem of false data being provided by the data sources nor the problem of collusion. However, when an adversary injects false data by a collusion attack scenario, it can affects the results of the honest aggregators and thus the base station will receive skewed aggregate value. In this case, the compromised nodes will attest their false data and consequently the base station assumes that all reports are from honest sensor nodes. Although the aforementioned research take into account false data injec-tion for a number of simple attack scenarios, to the best of our knowledge, no existing work addresses this issue in thecase of a collusion attack by compromised nodes in a man-ner which employs high level knowledge about data aggre-gation algorithm used.

IV. CONCLUSIONS

In this paper, we introduced a novel collusion attack sce-nario against a number of existing IF algorithms. More-over, we proposed an improvement for the IF algorithms by providing an initial approximation of the trustworthi-ness of sensor nodes which makes the algorithms not only collusion robust, but also more accurate and faster converging. In future work, We will investigate whether our approach can protect against compromised aggrega-tors. we also plan to implement our approach in a dep-loyed sensor network.

REFERENCES

[1]. S. Ozdemir and Y. Xiao, “Secure data aggregation in wireless sen-sor networks: A comprehensive overview,” Comput. Netw., vol. 53, no. 12, pp. 2022–2037, Aug. 2009.

[2]. L. Wasserman, All of Statistics : A Concise Course in Statistical Inference. New York, NY, USA: Springer.

[3]. A. Jøsang and J. Golbeck, “Challenges for robust trust and reputa-tion systems,” in Proc. 5th Int. Workshop Security Trust Manage., Saint Malo, France, 2009, pp. 253–262.

[4]. K. Hoffman, D. Zage, and C. Nita-Rotaru, “A survey of attack and defense techniques for reputation systems,” ACM Comput. Sur-veys, vol. 42, no. 1, pp. 1:1–1:31, Dec. 2009.

[5]. R. Roman, C. Fernandez-Gago, J. Lopez, and H. H. Chen, “Trust and reputation systems for wireless sensor networks,” in Security and Privacy in Mobile and Wireless Networking, S. Gritzalis, T. Karygiannis, and C. Skianis, eds.,Leicester, U.K.: Troubador Publishing Ltd, 2009 pp. 105–128,.

[6]. H.-S. Lim, Y.-S. Moon, and E. Bertino, “Provenance-based trust-worthiness assessment in sensor networks,” in Proc. 7th Int. Work-shop Data Manage. Sensor Netw., 2010, pp. 2–7.

[7]. H.-L. Shi, K. M. Hou, H. ying Zhou, and X. Liu, “Energy efficient and fault tolerant multicore wireless sensor network: E2MWSN,” in Proc. 7th Int. Conf. Wireless Commun., Netw. Mobile Comput., 2011