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

Mohsen Rezvani, *Student Member, IEEE,* Aleksandar Ignjatovic, Elisa Bertino, *Fellow, IEEE* and Sanjay Jha, *Senior Member, IEEE,*

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.

Index Terms—Wireless sensor networks, robust data aggregation, collusion attacks.

1 Introduction

Due to a need for 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 gain 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 a 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* supplying 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 the sensor nodes.

Trust and reputation systems have a significant role in supporting operation of a wide 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 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 behaviour. 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

- *M. Rezvani, A. Ignjatovic and S. Jha are with School of Computer Science and Engineering, University of New South Wales, Sydney, Australia. E-mail: {mrezvani,ignjat,sanjay}@cse.unsw.edu.au*
- *E. Bertino is with Department of Computer Science, Purdue University. E-mail: bertino@cs.purdue.edu*

severely impair the performance of such a system [3]. The main target of malicious attackers are aggregation algorithms of trust and reputation systems [4].

Trust and reputation have been recently suggested as an effective security mechanism for Wireless Sensor Networks (WSNs) [5]. Although sensor networks are being increasingly 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 compromising attacks by adversaries who intend to inject false data into the system. In this context, assessing the trustworthiness of the collected data and announcing decision makers for the data trustworthiness becomes a challenging task [6].

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 emergence of multi-core and multi-processor systems in sensor nodes [7].

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 [8]. Such trustworthiness estimate of each sensor is based 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 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 systems [9], [8], [10], [11], [12], [13], [14], [15]. The performance of IF algorithms in the presence of different types of faults and simple false data injection attacks has been studied, for example in [16] where it was applied to compressive 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 sophisticated 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 detection, compromised node detection, secure data aggregation, cluster head election, outlier detection, etc [17]. Second, sensors which are deployed in hostile and unattended environments are highly susceptible to node compromising attacks [18]. While offering better protection than the simple averaging, our simulation results demonstrate that indeed current IF algorithms are vulnerable 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 statistical 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 complete sensor failure.

Since readings keep streaming into aggregator nodes in WSNs, and since attacks can be very dynamic (such as *orchestrated* attacks [4]), in order to obtain trustworthiness of nodes as well as to identify compromised nodes we apply our framework on consecutive batches of consecutive readings. Sensors are deemed compromised only relative to a particular batch; this allows our framework to handle on-off type of attacks (called orchestrated attacks in [4]).

We validate the performance of our algorithm by simulation on synthetically generated datasets. 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. An extended version of this paper has been published as a technical report in [19]

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TDSC.2014.2316816, IEEE Transactions on Dependable and Secure Computing

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 estimate of the noise parameters obtained using contribution 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;

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

The rest of this paper is organized as follows. Section 2 describes the problem statement and the assumptions. Section 3 presents our novel robust data aggregation framework. Section 4 describes our experimental results. Section 5 presents the related work. Finally, the paper is concluded in Section 6.

2 Background, Assumptions, Threat Model and Problem Statement

In this section, we present our assumptions, discuss IF algorithms, describe a collusion attack scenario against IF algorithms, and state the problems that we address in this paper.

2.1 Network Model

For the sensor network topology, we consider the abstract model proposed by Wagner in [20]. Fig. 1 shows our assumption for network model in WSN. The sensor nodes are divided into disjoint clusters, and each cluster has a cluster 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 compromised and might be sending false data to the aggregator. We assume that each data aggregator has enough computational power to run an IF algorithm for data aggregation.

2.2 Iterative Filtering in Reputation Systems

Kerchove and Dooren proposed in [8] an IF algorithm 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 vulnerability of the algorithm for a possible collusion attack. We note that our improvement is applicable to other IF algorithms as well.

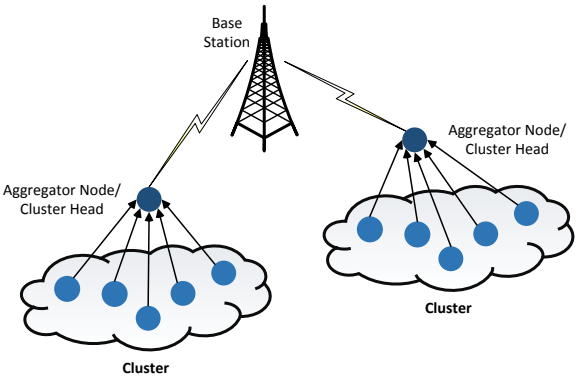


Fig. 1: Network model for WSN.

We consider a WSN with n sensors S_i, $i = 1, \ldots, n$. We assume that the aggregator works on one block of readings at a time, each block comprising of readings at m consecutive instants. Therefore, a block of readings is represented by a matrix $X = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ where $\mathbf{x}_i = [x_i^1 \; x_i^2 \; \ldots \; x_i^m]^T$, $(1 \leq i \leq n)$ represents the i^{th} m-dimensional readings reported by sensor node S_i. Let $\mathbf{r} = [r_1 \; r_2 \; \ldots \; r_m]^T$ denote the aggregate values for instants $t = 1, \ldots, m$, which authors of [8] call a *reputation vector*[2], computed iteratively and simultaneously with a sequence of weights $\mathbf{w} = [w_1 \; w_2 \; \ldots \; w_n]^T$ reflecting the trustworthiness of sensors. We denote by $\mathbf{r}^{(l)}$, $\mathbf{w}^{(l)}$ the approximations of $\mathbf{r}$, $\mathbf{w}$ obtained at l^{th} round of iteration ($l \geq 0$).

The iterative procedure starts with giving equal credibility to all sensors, i.e., with an initial value $\mathbf{w}^{(0)} = \mathbf{1}$. The value of the reputation vector $\mathbf{r}^{(l+1)}$ in round of iteration $l+1$ is obtained from the weights of the sensors obtained in the round of iteration l as

$$\mathbf{r}^{(l+1)} = \frac{X \cdot \mathbf{w}^{(l)}}{\sum_{i=1}^{n} w_i^{(l)}}$$

Consequently, the initial reputation vector is $\mathbf{r}^{(1)} = \frac{1}{n} X \cdot \mathbf{1}$, i.e., $\mathbf{r}^{(1)}$ is just the sequence of simple averages of the readings of all sensors at each particular instant. The new weight vector $\mathbf{w}^{(l+1)}$ to be used in round of iteration $l+1$ is then computed as a function $g(\mathbf{d})$ of the normalized *belief divergence* $\mathbf{d}$ which is the distance between the sensor readings and the reputation vector $\mathbf{r}^{(l)}$. Thus, $\mathbf{d} = [d_1 \; d_2 \; \ldots \; d_n]^T$, $d_i = \frac{1}{m} \left\| \mathbf{x}_i - \mathbf{r}^{(l+1)} \right\|_2^2$ and $w_i^{(l+1)} = \mathrm{g}(d_i)$, $(1 \leq i \leq n)$.

Function $\mathrm{g}(x)$ is called the *discriminant function* and it provides an inverse relationship of weights to distances $\mathbf{d}$. Our experiments show that selecting a discriminant function has a significant role in stability and robustness of IF algorithms. A number of alternatives for this function are studied in [8]:

2. We find such terminology confusing, because reputation should pertain to the level of trustworthiness rather than the aggregate value, but have decided to keep the terminology which is already in use.

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*;

- reciprocal: $\mathrm{g}(\mathbf{d}) = \mathbf{d}^{-k}$;
- exponential: $\mathrm{g}(\mathbf{d}) = e^{-\mathbf{d}}$;
- affine: $\mathrm{g}(\mathbf{d}) = 1 - k_l\mathbf{d}$, where $k_l > 0$ is chosen so that $g(\max_i\{d_i^{(l)}\}) = 0$.

Algorithm 1 illustrates the iterative computation of the reputation vector based on the above formulas. TABLE 1 shows a trace example of this algorithm. The sensor readings in the first three rows of this table are from sensed temperatures in Intel Lab dataset [21] at three different time instants. We executed the IF algorithm on the readings; the discriminant function in the algorithm was a reciprocal of the distance between sensor readings and the current computed reputation. The lower part of the table illustrates the weight vector in each iteration as well as the obtained reputation values for the three different time instants ($t1$, $t2$, $t3$) in the last three columns. As can be seen, the algorithm converges after six iterations.

2.3 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 [22]. 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 knowledge 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.

2.4 Collusion Attack Scenario

Most of the IF algorithms employ simple assumptions about the initial values of weights for sensors. In case of our adversary model, an attacker is able to mislead the aggregation system through careful selection of reported data values. We use visualisation techniques from [18] to present our attack scenario.

Assume that ten sensors report the values of temperature which are aggregated using the IF algorithm proposed in [8] with the reciprocal discriminant function. We consider three possible scenarios; see Fig. 2.

- 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 penalises them and assigns to them lower weights, because their values are far from the values of other sensors. In other words, the algorithm is robust against false data injection in this scenario because the compromised nodes individually falsify the readings without any knowledge about the aggregation algorithm. TABLE 2 illustrates a trace example of the attack scenario on Intel dataset; sensors 9 and 10 are compromised by an adversary. As one can see, the algorithm assigns very low weights to these two sensor nodes and consequently their contributions decrease. Thus, the IF algorithm is robust against the simple outlier injection by the compromised nodes.
- 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. 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 a highly reliable reading. As a result, IF algorithms will converge to the values provided by the third compromised 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,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TDSC.2014.2316816, IEEE Transactions on Dependable and Secure Computing

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

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 dataset; sensors 8, 9 and 10 are compromised 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 control 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 experiments, we investigate how the behaviour of the IF algorithm depends on the number of compromised nodes; see Section 4.4.

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 vector 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 estimate 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 scenario of the type described.

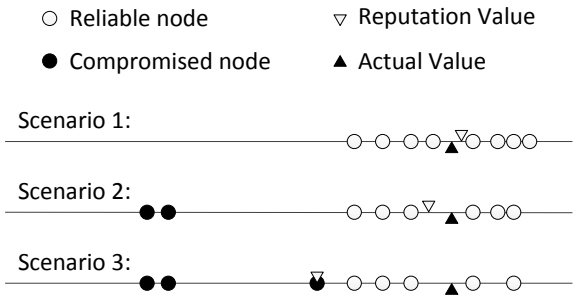


Fig. 2: Attack scenario against IF algorithm.

3 Robust Data Aggregation

In this section, we present our robust data aggregation method. TABLE 4 contains a summary of notations used in this paper.

3.1 Framework Overview

In order to improve the performance of IF algorithms against the aforementioned attack scenario, we provide a robust initial estimation of the trustworthiness of sensor nodes to be used in the first iteration of the IF algorithm. Most of the traditional statistical estimation methods for variance involve use of the sample mean. For this reason, proposing a robust variance estimation method in the case of skewed sample mean is an essential part of our methodology.

In the rest of this paper, we assume that the stochastic components of sensor errors are independent random variables with a Gaussian distribution; however, our experiments show that our method works quite well for other types of errors without any modification. Moreover, if error distribution of sensors is either known or estimated, our algorithms can be adapted to other distributions to achieve an optimal performance.

Fig. 3 illustrates the stages of our robust aggregation framework and their interconnections. As we have mentioned, our aggregation method operates on

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

TABLE 3: A trace example of the proposed collusion 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	**13.2001**	**13.5609**	**18.4546**	
round#	**sensor weights**										**t=1**
1	1	1	1	1	1	1	1	1	1	1	18.4546
2	0.6113	0.7414	0.5755	0.3268	0.2590	1.0106	0.3540	0.0362	0.0418	**6.25E+08**	18.4546
3	0.6113	0.7414	0.5755	0.3268	0.2590	1.0105	0.3540	0.0362	0.0418	**1.78E+16**	18.4546

TABLE 4: Notation used in this paper.

n	number of sensors
m	number of readings for each sensor
r^t	true value of the signal at time t
x_s^t	data from sensor s at time t
e_s^t	noise (error) of sensor s at time t
b_s	bias of sensor s
σ_s	standard deviation of noise of sensor s
v_s	variance of sensor s

batches of consecutive readings of sensors, proceeding in several stages. In the first stage we 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 are presented in Section 3.2 and 3.3, respectively.

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. The detailed computation operations of such estimation are described in Section 3.4.

In the third stage of the proposed framework, the initial reputation vector provided in the second stage is used to estimate the trustworthiness of each sensor based on the distance of sensor readings to such initial reputation vector. This idea will be described in Section 3.5.

3.2 Estimating Bias

We assume that all sensors in WSN can have some error; such error e_s^t of a sensor s is modelled by the

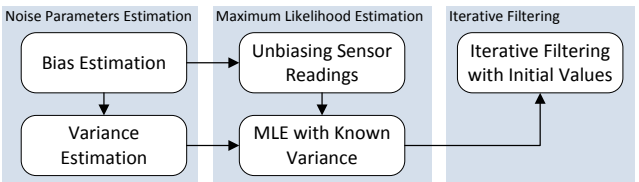


Fig. 3: Our robust data aggregation framework.

Gaussian distribution random variable with a sensor bias b_s and sensor variance σ_s, $e_s^t \sim \mathcal{N}(b_s, \sigma_s^2)$. Let r_t denotes the true value of the signal at time t. Each sensor reading x_s^t can be written as:

$$x_s^t = r_t + e_s^t \tag{1}$$

The main idea is that, since we have no access to the true value r_t we cannot obtain the value of the error e_s^t; however, we can obtain the values of the differences of such errors. Thus, if we define $\delta(i,j) = \frac{1}{m}\sum_{t=1}^{m}\left(x_i^t - x_j^t\right)$, we get:

$$\delta(i,j) = \frac{1}{m}\sum_{t=1}^{m}\left(x_i^t - x_j^t\right) = \frac{1}{m}\sum_{t=1}^{m} e_i^t - \frac{1}{m}\sum_{t=1}^{m} e_j^t$$

where e_i^t is a random variable with Gaussian distribution $e_i^t \sim \mathcal{N}(b_i, \sigma_i^2)$. Let $\bar{e}_i = \frac{1}{m}\sum_{t=1}^{m} e_i^t$ be the sample mean of this random variable. As the sample mean is an unbiased estimator of the expected value of a random variable, we have

$$\delta(i,j) = \bar{e}_i - \bar{e}_j \approx b_i - b_j$$

Let $\boldsymbol{\delta} = \{\delta(i,j) : 1 \le i,j \le n\}$; this matrix is an estimator for mutual difference of sensor bias. In order to obtain the sensor bias from this matrix, we solve

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TDSC.2014.2316816, IEEE Transactions on Dependable and Secure Computing

the following minimization problem.

$$\begin{aligned} \underset{\mathbf{b}}{\text{minimize}} \quad & \sum_{i=1}^{n}\sum_{j=1}^{i-1}\left(\frac{b_i-b_j}{\delta(i,j)}-1\right)^2 \\ \text{subject to} \quad & \sum_{i=1}^{n} b_i = 0. \end{aligned} \tag{2}$$

To justify our constraint, it is clear that if the mean of the bias of all sensors is not zero, then there would be no way to account for it on the basis of sensor readings. On the other hand, bias of sensors, under normal circumstances, comes from imperfections in manufacture and calibration of sensors as well as from the fact that they might be deployed in places with different environmental circumstances where the sensed scalar might in fact have a slightly different value. Since by the very nature we are interested in obtaining a most reliable estimate of an average value of the variable sensed, it is reasonable to assume that the mean bias of all sensors is zero (without faults or malicious attacks). We chose the above objective rather than $\sum_{i=1}^{n}\sum_{j=1}^{i-1}(b_i-b_j-\delta(i,j))^2$ to improve the performance in case when biases can be of very different magnitudes.

We introduce a Lagrangian multiplier λ and look at extremal values of the following function:

$$F(\vec{b}) = \sum_{i=1}^{n}\sum_{j=1}^{i-1}\left(\frac{b_i-b_j}{\delta(i,j)}-1\right)^2 + \lambda\sum_{i=1}^{n} b_i$$

By setting the gradient of $F(\vec{b})$ to zero we obtain a system of linear equations whose solution is our approximation of the the bias values. If we let

$$\mathrm{d}(i,j) = \begin{cases} -\delta(j,i) & i<j \\ \delta(j,i) & i\geq j \end{cases}$$

then these equations can be written in the following compact form:

$$\begin{cases} \sum\limits_{\substack{i=1\\ i\neq k}}^{n} \frac{2}{\mathrm{d}(i,k)^2} b_i - \sum\limits_{\substack{i=1\\ i\neq k}}^{n} \frac{2}{\mathrm{d}(i,k)^2} b_k - \lambda = 2\sum\limits_{\substack{i=1\\ i\neq k}}^{n} \frac{1}{\mathrm{d}(i,k)}, \\ \qquad\qquad\qquad\qquad\qquad\qquad \text{for all } k=1,\cdots,n \\ \sum\limits_{i=1}^{n} b_i = 0. \end{cases} \tag{3}$$

Note that the obtained value of b_i is actually an approximation of the sample mean of the error of sensor i, which, in turn is an unbiased estimator of the bias of such a sensor.

3.3 Estimating Variance

In this section, we propose a similar method to estimate variance of the sensor noise using the estimated bias from previous section. Given the bias vector $\mathbf{b} = [b_1, b_2, \cdots, b_n]$ and sensor readings $\{x_s^t\}$, we can define matrices $\{\hat{x}_s^t\}$ and $\boldsymbol{\beta} = \{\beta(i,j)\}$ as follows:

$$\hat{x}_s^t = x_s^t - b_s \tag{4}$$

$$\begin{aligned} \beta(i,j) &= \frac{1}{m-1}\sum_{t=1}^{m}\left(\hat{x}_i^t - \hat{x}_j^t\right)^2 \\ &= \frac{1}{m-1}\sum_{t=1}^{m}\left(\left(x_i^t - x_j^t\right) - \left(b_i - b_j\right)\right)^2 \end{aligned}$$

By (1) we have $x_i^t - x_j^t = (r_t + e_i^t) - (r_t + e_j^t) = e_i^t - e_j^t$; thus, we obtain

$$\begin{aligned} \beta(i,j) = \frac{1}{m-1}\sum_{t=1}^{m}\left(e_i^t - b_i\right)^2 + \frac{1}{m-1}\sum_{t=1}^{m}\left(e_j^t - b_j\right)^2 \\ - \frac{2}{m-1}\sum_{t=1}^{m}\left(e_i^t - b_i\right)\left(e_j^t - b_j\right) \end{aligned}$$

We assume that the sensors noise is generated by independent random variables[3]; as we have mentioned, our approximations of the bias b_i are actually approximations of the sample mean; thus

$$\frac{1}{m-1}\sum_{t=1}^{m}\left(e_i^t - b_i\right)\left(e_j^t - b_j\right) \approx Cov(e_i, e_j) = 0$$

and similarly

$$\begin{aligned} \beta(i,j) &= \frac{1}{m-1}\sum_{t=1}^{m}\left(e_i^t - b_i\right)^2 + \frac{1}{m-1}\sum_{t=1}^{m}\left(e_j^t - b_j\right)^2 \\ &\approx \sigma_i^2 + \sigma_j^2 \end{aligned}$$

The above formula shows that we can estimate the variance of sensors noise by computing the matrix $\boldsymbol{\beta}$. We also compute the sum of variances of all sensors using the following Lemma.

Lemma 3.1 (Total Variance)**.** *Let $\bar{x}^t$ be the mean of readings in time t, then, using* (4) *and our assumption that $\sum_{i=1}^{n} b_i = 0$, we have*

$$\bar{x}^t = \frac{1}{n}\sum_{j=1}^{n} x_j^t = \frac{1}{n}\sum_{j=1}^{n}\hat{x}_j^t,$$

and the statistic

$$S(t) = \frac{n}{m(n-1)}\sum_{i=1}^{n}\sum_{t=1}^{m}\left(\hat{x}_i^t - \bar{x}^t\right)^2$$

is an unbiased estimator of the sum of the variances of all sensors, $\sum_{i=1}^{n} v_i$.

We presented the proof of the lemma in [19]. To obtain an estimation of variances of sensors from

3. We analyze our estimation method with synthetic correlated data and the experimental results show that the our method produces excellent results even for correlated noise.

the matrix $\boldsymbol{\beta} = \{\beta(i,j)\}$ we solve the following minimization problem:

$$\begin{aligned} \underset{\mathbf{v}}{\text{minimize}} \quad & \sum_{i=1}^{n}\sum_{j=1}^{i-1}\left(\frac{v_i+v_j}{\beta(i,j)}-1\right)^2 \\ \text{subject to} \quad & \sum_{i=1}^{n} v_i = \frac{n}{m(n-1)}\sum_{i=1}^{n}\sum_{t=1}^{m}\left(\hat{x}_i^t-\bar{x}^t\right)^2 \end{aligned} \tag{5}$$

Note that the constrain of the minimisation problem comes from Lemma 3.1. We again introduce a Lagrangian multiplier λ and by solving the minimization problem, we obtain linear Equations (6):

$$\begin{cases} \sum\limits_{\substack{i=1\\ i\neq k}}^{n} \frac{1}{\beta(i,k)^2} v_i + \sum\limits_{\substack{i=1\\ i\neq k}}^{n} \frac{1}{\beta(i,k)^2} v_k + \frac{\lambda}{2} = \sum\limits_{i=1}^{n} \frac{1}{\beta(i,k)}, \\ \qquad\qquad\qquad\qquad \text{for all } k = 1,\cdots,n \\ \sum\limits_{i=1}^{n} v_i = \frac{n}{m(n-1)} \sum\limits_{i=1}^{n}\sum\limits_{t=1}^{m} \left(\hat{x}_i^t - \bar{x}^t\right)^2. \end{cases} \tag{6}$$

3.4 MLE with Known Variance

In the previous sections, we proposed 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.

Given matrix $\{x_s^t\}$ where $x_s^t \sim r_t + \mathcal{N}(b_s, \sigma_s^2)$ and estimated bias and variance vectors $\mathbf{b}$ and $\boldsymbol{\sigma}$, we propose to recover r_t using (an approximate form of) the MLE applied to the values obtained by subtracting the bias estimates from sensors readings. As it is well known, in this case the MLE has the smallest possible variance as it attains the CRLB.

From a heuristic point of view, we removed the "systematic component" of the error by subtracting a quantity which in the case of a stochastic error corresponds to an estimate of bias; this allows us to estimate the variability around such a systematic component of the error, which, in case of stochastic errors, corresponds to variance. We can now obtain an estimation which corresponds to MLE formula for the case of zero mean normally distributed errors, but with estimated rather than true variances. Therefore, we assume that the expected value r_t of the measurements is the true value of the quantity measured, and is the only parameter in the likelihood function. Thus, in the expression for the likelihood function for normally distributed unbiased case,

$$\mathcal{L}_n(r_t) = \prod_{i=1}^{n} \frac{1}{\sigma_i\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x_i^t - r_t)^2}{\sigma_i^2}}$$

we replace σ_i^2 by the obtained variance v_i from Equation (6). Moreover, by differentiating the above formula with respect to r_t and setting the derivative equal to zero we get

$$r_t = \sum_{i=1}^{n} \frac{\frac{1}{v_i}}{\sum_{j=1}^{n}\frac{1}{v_j}} x_i^t \quad \text{for all} \quad t = 1,\cdots,m. \tag{7}$$

Equation (7) provide an estimate of the true value of the quantity measured in a form of a weighted average of sensor readings, with the sensor readings given a weight inversely proportional to the estimation of their error variance provided by our method:

$$\mathbf{r} = \sum_{s=1}^{n} w_s \mathbf{x}_s \tag{8}$$

Note that this method estimates the reputation vector without any iteration. Thus, the computational complexity of the estimation is considerably less than the existing IF algorithms.

3.5 Enhanced Iterative Filtering

According to the proposed attack scenario, the attacker exploits the vulnerability of the IF algorithms which originates from a wrong assumption about the initial trustworthiness of sensors. Our contribution to address this shortcomings is to employ the results of the proposed robust data aggregation technique as the initial reputation for these algorithms. Moreover, the initial weights for all sensor nodes can be computed based on the distance of sensors readings to such an initial reputation. Our experimental results illustrate that this idea not only consolidates the IF algorithms against the proposed attack scenario, but using this initial reputation improves the efficiency of the IF algorithms by reducing the number of iterations needed to approach a stationary point within the prescribed tolerance; see Section 4.2.

4 Simulation Results

In this section, we report on a detailed numerical simulation study that examines robustness and efficiency of our data aggregation method. [4] The objective of our experiments is to evaluate the robustness and efficiency of our approach for estimating the true values of signal based on the sensor readings in the presence of faults and collusion attacks. For each experiment, we evaluate the accuracy based on Root Mean Squared error (RMS error) metric and efficiency based on the number of iterations needed for convergence of IF algorithms.

4. Unfortunately, we are unable to prove mathematically that our system is secure; however we demonstrated higher robustness compared to the state of the art by thorough empirical evaluation. We are working on obtaining a rigorous proof but this appears to be a challenging problem.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TDSC.2014.2316816, IEEE Transactions on Dependable and Secure Computing

4.1 Experimental Settings

All the experiments have been conducted on an HP PC with 3.30GHz Intel Core i5-2500 processor with 8Gb RAM running a 64-bit Windows 7 Enterprise. The program code has been written in MATLAB R2012b. Although there are a number of real world datasets for evaluating reputation systems and data aggregation in sensor networks such as Intel dataset [21], none of them provides a clear ground truth. Thus, we conduct our experiments by generating synthetic datasets. The experiments are based on simulations performed with both correlated and uncorrelated sensor errors. If not mentioned otherwise, we generate synthetic datasets according to the following parameters:

- Each simulation experiment was repeated 200 times and then results were averaged;
- Number of sensor nodes is $n = 20$;
- Number of readings for each sensor is $m = 400$;
- For statistical parameters of the errors (noise) used to corrupt the true readings, we consider several ranges of values for bias, variance and covariance of noise for each experiment;
- The level of significance in K-S test is $\alpha = 0.05$.

In all experiments, we compare our robust aggregation method against three other IF techniques proposed for reputation systems. For all parameters of other algorithms used in the experiments, we set the same values as used in the original papers where they were introduced.

The first IF method considered computes the trustworthiness of sensor nodes based on the distance of their readings to the current state of the estimated reputation [8]. We described the details of this approach in Section 2.2. We investigate two discriminant functions in our experiments $\mathrm{g}(\mathbf{d}) = \mathbf{d}^{-1}$ and $\mathrm{g}(\mathbf{d}) = 1 - k_l \mathbf{d}$, and call these methods *dKVD-Reciprocal* and *dKVD-Affine*, respectively.

The second IF method we consider is a correlation based ranking algorithm proposed by Zhou et al. in [9]. In this algorithm, trustworthiness of each sensor is obtained based on the correlation coefficient between the sensors readings and the current estimate of the true value of the signal. In other words, this method gives credit to sensor nodes whose readings correlate well with the estimated true value of the signal. Based on this idea, the authors proposed an iterative algorithm for estimating the true value of the signal by applying a weighted averaging technique. They argued that correlation coefficient is a good way to quantify the similarity between two vectors. Thus, they employed Pearson correlation coefficient between sensor readings and the current state of estimate signal in order to compute the sensor weight. We call this method *Zhou*.

The third algorithm considered has been proposed by Laureti et al. in [10] and is an IF algorithm based on a weighted averaging technique similar to the algorithm described in Section 2.2. The only difference between these two algorithms is in the discriminant function. The authors in [10] exploited discriminant function $\mathrm{g}(\mathbf{d}) = \mathbf{d}^{-0.5}$. We call this method *Laureti*.

TABLE 5: Summary of different IF algorithms.

Name	Discriminant Function
dKVD-Reciprocal	$w_i^{l+1} = \left(\frac{1}{m} \left\| \mathbf{x}_i - \mathbf{r}^{l+1} \right\|_2^2 \right)^{-1}$
dKVD-Affine	$w_i^{l+1} = 1 - k \frac{1}{m} \left\| \mathbf{x}_i - \mathbf{r}^{l+1} \right\|_2^2$
Zhou	$w_i^{l+1} = \frac{1}{m} \sum_{i=1}^{m} \left(\frac{x_i^t - \bar{\mathbf{x}}^t}{\sigma_{\mathbf{x}_i}} \right) \left(\frac{r^t - \bar{\mathbf{r}}}{\sigma_r} \right)$
Laureti	$w_i^{l+1} = \left(\frac{1}{m} \left\| \mathbf{x}_i - \mathbf{r}^{l+1} \right\|_2^2 \right)^{-0.5}$

We apply *dKVD-Reciprocal*, *dKVD-Affine*, *Zhou*, *Laureti* and our robust aggregation approach to synthetically generated data. Although we can simply apply our robust framework to all existing IF approaches, in this paper we investigate the improvement which addition of our initial trustworthiness assessment method produces on the robustness of *dKVD-Reciprocal* and *dKVD-Affine* methods (We call them *RobustAggregate-Reciprocal* and *RobustAggregate-Affine*, respectively.).

TABLE 5 shows a summary of discriminant functions for all of the above four different IF methods.

We first conduct experiments by injecting only Gaussian noise into sensor readings. In the second part of the experiments, we investigate the behaviour of these approaches by emulating a simple, non-colluding attack scenario presented in the second case of Fig. 2. We then evaluate these approaches in the case of our sophisticated attack scenario.

4.2 Accuracy and Efficiency without an Attack

In the first batch of experiments we assume that there are no sensors with malicious behaviour. Thus, the errors are fully stochastic; we consider Gaussian sensors errors. In order to evaluate the performance of our algorithm in comparison with the existing algorithms, we produce the following four different synthetic datasets.

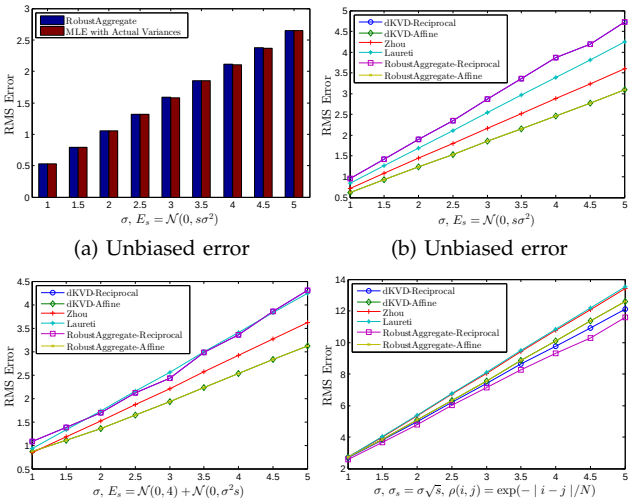
1) **Unbiased error**: We considered various distributions of the variance across the set of sensors and obtained similar results. We have chosen to present the case with the error of a sensor s at time t is given by $e_s^t \sim \mathcal{N}(0, s \times \sigma^2)$, considering different values for the baseline sensor variance σ^2. Fig. 4(a) shows the results of the MLE with our noise parameter estimation (steps ① and ② in Fig. 3) and the information theoretic limit for the minimal variance provided by the CRLB, achieved, for example, using the MLE with the *actual, exact variances* of sensors, which are NOT available to our algorithm. As one can see in this figure, our proposed approach

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TDSC.2014.2316816, IEEE Transactions on Dependable and Secure Computing

nearly exactly achieves the minimal possible variance coming from the information theoretic lower bound. Furthermore, Fig. 4(b) illustrates the performance of our approach for the initial trustworthiness assessment of sensors with different discriminant functions as well as other IF algorithms. It shows that in this experiment, the performance of our approach with both discriminant functions is very similar to the original IF algorithm.

2) **Bias error**: In this scenario, we inject bias error to sensor readings, generated by Gaussian distribution with different variances. Therefore, the error of sensor s in time t is generated by $e_s^t \sim \mathcal{N}(\mathcal{N}(0, \sigma_b^2), s \times \sigma^2)$ with the variance of the bias $\sigma_b^2 = 4$ and increasing values for variances, where the variance of sensor s is equal to $s \times \sigma^2$. Thus, the sensors bias is produced by a zero mean Gaussian distribution random variable. Fig. 4(c) shows the RMS error for all algorithms in this scenario. As can be seen in this figure, since all of the IF algorithms, along with our approach, generate an error close to their errors in the unbiased scenario, we can conclude that the methods are stable against bias but fully stochastic noise.
3) **Correlated noise**: The heuristics behind our initial variance estimation assumed that the errors of sensors are uncorrelated. Thus, we tested how the performance of our method degrades if the noise becomes correlated and how it compares to the existing methods under the same circumstances. So in this scenario, we assume that the errors of sensors are no longer uncorrelated. Possible covariance functions can be of different types, such as *Spherical*, *Power Exponential*, *Rational Quadratic*, and *Matern*; see [23]. Although our proposed method can be applied to all covariance functions, we present here the results for the case of the *Power Exponential* function $\rho(i,j) = e^{\frac{-|i-j|}{n}}$. Moreover, the variance of a sensor s is again set to $\sigma_s^2 = s \times \sigma^2$. From the corresponding covariance matrix $\mathbf{\Sigma} = \{\Sigma_{ij} = \rho(i,j)\sigma_i\sigma_j \ : \ i,j = 1 \cdots n\}$, the noise values of sensors are generated from multivariate Normal distribution $\mathbf{Noise} \sim \mathcal{N}(\mathbf{Bias}, \mathbf{\Sigma})$. In this scenario, we take into account different values of σ for generating the noise values of sensors in order to analyse the accuracy of the data aggregation under various levels of noise. Fig. 4(d) shows the RMS error of the algorithms for this scenario. As can be seen in this figure, our approach with reciprocal discriminant function improves *dKVD-Reciprocal* algorithm for all different values of variance, although our method with affine function generates very similar RMS error to the original *dKVD-Affine* algorithm. Moreover, the scale of RMS error is in general larger than in scenarios with uncorrelated noise, as one would expect. This can be explained by our assumption that the sensors noise is generated by independent random variables; see Section 3.3.

The results of our simulations also show that the use of our initial variance estimation in the second phase of our proposed framework as the initial reputation of IF algorithms decreases the number of iterations for the algorithms. We evaluate the number of iterations for the IF algorithm proposed in [8] by providing the initial reputation from the results of the our approach for both unbiased and biased sensors errors. The results of this experiment show that the proposed initial reputation for the IF algorithm improves the efficiency of the algorithm in terms of the number of iterations until the procedure has converged. In other words, by providing this initial reputation, the number of iterations for IF algorithm decreases approximately 9% for reciprocal and around 8% for affine discriminant functions in both biased and unbiased circumstances. This can be explained by the fact that the new initial reputation is close to the true value of signal and the IF algorithm needs fewer iterations to reach its stationary point. In the next part of our experiments, we employ this idea for consolidating the IF algorithm against the proposed attack scenario.



(a) Unbiased error (b) Unbiased error

(c) Bias error (d) Correlated noise

Fig. 4: Accuracy for *No Attack* scenarios.

4.3 Accuracy with Simple Attack Scenario

Lim et al. in [18] introduced an attack scenario against traditional statistical aggregation approaches. We described the scenario in Section 2.4 and the second round of Fig. 2 as a simple attack scenario using a number of compromised node for skewing the simple average of sensors readings. In this section, we investigate the behavior of IF algorithms against the simple attack scenario. Note that the objective of this attack scenario is to skew the sample mean of

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TDSC.2014.2316816, IEEE Transactions on Dependable and Secure Computing

sensors readings through reporting outlier readings by the compromised nodes.

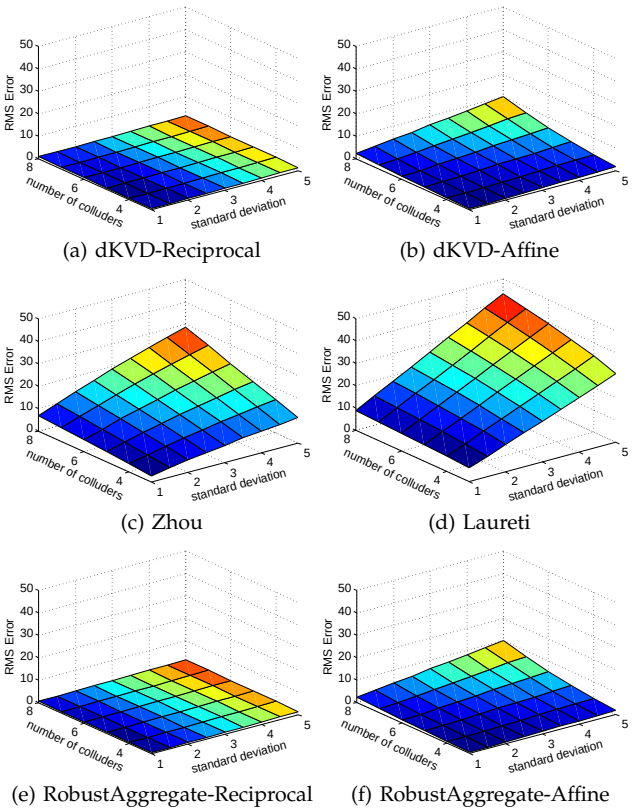
In order to evaluate the accuracy of the IF algorithms against the simple attack scenario, we assume that the attacker compromises c ($c < n$) sensor nodes and reports outlier readings by these nodes. We generate synthetically datasets for this attack scenario by taking into account different values of variance for sensors errors as well as employing various number of compromised nodes. Moreover, we generate biased readings for all sensor nodes with bias provided by a random variable with a distribution $\mathcal{N}(0, \sigma_b^2)$ with the variance of bias chosen to be $\sigma_b^2 = 4$.

Fig. 5 shows the accuracy of the IF algorithms and our approach in the presence of such simple attack scenario. It can be seen that the estimates provided by the three approaches, *dKVD-Affine*, *Zhou* and *Laureti* are significantly skewed by this attack scenario and their accuracy significantly decreases by increasing the number of compromised nodes. On the other hand, *dKVD-Reciprocal* provides a reasonable accuracy for all parameter values of this simple attack scenario (see Fig. 5(a)). The robustness of this discriminant function can be explained by the fact that the function sharply diminishes the contributions of outlier readings through assigning very low values of weights to them. In our sophisticated collusion attack scenario, we exploit this property in order to compromise systems employing such discriminant function.

The results of this experiment clearly show that our initial trustworthiness has no negative effects on the performance of the IF algorithm with both discriminant functions in the case of the simple attack scenario. In next section, we show that how this initial values improve the IF algorithm in the case of proposed collusion attack scenario, while both *dKVD-Affine* and *dKVD-Reciprocal* algorithms are compromised against such an attack scenario.

4.4 Accuracy with a Collusion Attack

In order to illustrate the robustness of the proposed data aggregation method in the presence of sophisticated attacks, we synthetically generate several datasets by injecting the proposed collusion attacks. Therefore, we assume that the adversary employs c ($c < n$) compromised sensor nodes to launch the sophisticated attack scenario proposed in Section 2.4. The attacker uses the first $c-1$ compromised nodes to generate outlier readings in order to skew the simple average of all sensor readings. The adversary then falsifies the last sensor readings by injecting the values very close to such skewed average. This collusion attack scenario makes the IF algorithm to converge to a wrong stationary point. In order to investigate the accuracy of the IF algorithms with this collusion attack scenario, we synthetically generate several datasets with different values for sensors variances as well as various number of compromised nodes (c).



(a) dKVD-Reciprocal (b) dKVD-Affine

(c) Zhou (d) Laureti

(e) RobustAggregate-Reciprocal (f) RobustAggregate-Affine

Fig. 5: Accuracy with a simple attack scenario.

Fig. 6 shows the accuracy of the IF algorithms and our approach in the presence of the collusion attack scenario. It can be seen that the IF algorithms with reciprocal discriminant function are highly vulnerable to such attack scenario (see Fig. 6(a) and Fig. 6(d)), while the affine discriminant function generates more robust results in this case (see Fig. 6(b)). However, the accuracy of the affine discriminant function is still much worse than the previous experiment without the collusion attack.

This experiment shows that the collusion attack scenario can circumvent all the IF algorithms we tried. Moreover, the accuracy of the algorithms dramatically decreases by increasing the number of compromised nodes participated in the attack scenario. As explained before, the algorithms converge to the readings of one of the compromised nodes, namely, to the readings of the node which reports values very close to the skewed mean. This demonstrates that an attacker with enough knowledge about the aggregation algorithm employed can launch a sophisticated collusion attack scenario which defeats IF aggregation systems.

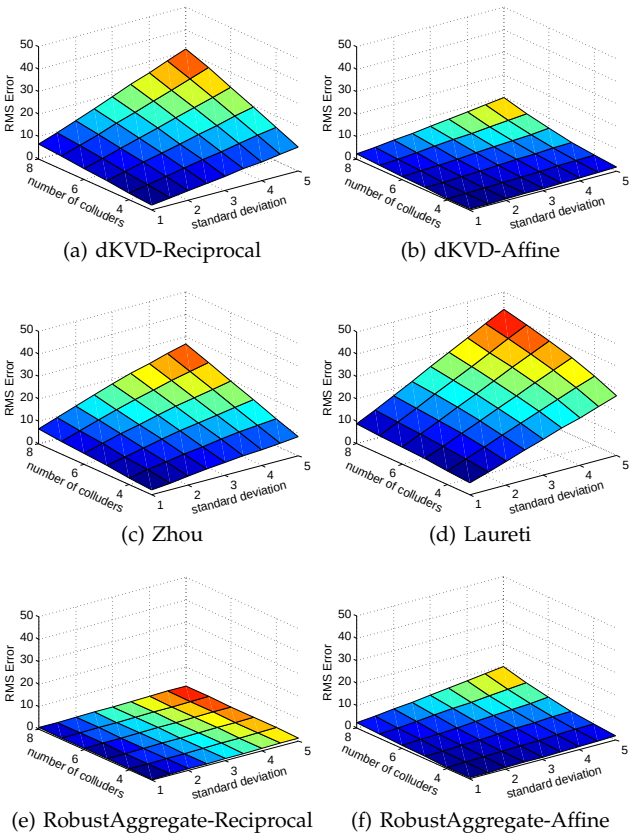
Fig. 6(e) and Fig. 6(f) show the accuracy of our approach by taking into account the IF algorithm in [8] with reciprocal and affine discriminant functions, respectively. As one can see, our proposed approach is superior to all other algorithms in terms of the accuracy for reciprocal discriminant functions, while the approach has a very small improvement on affine function. Moreover, comparing the accuracy of our approach in this experiment with the results from *no*

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TDSC.2014.2316816, IEEE Transactions on Dependable and Secure Computing

attack and simple attack experiments in Fig. 4 and Fig. 5, we can argue that our approach with reciprocal discriminant function is robust against the collusion attack scenario. The reason is that our approach not only provides the highest accuracy for this discriminant functions, it actually approximately reaches the accuracy of *No Attack* scenarios.

As we described, the main shortcoming of the IF algorithms in the proposed attack scenario is that they quickly converge to the sample mean in the presence of the attack scenario. In order to investigate the shortcoming, we conducted an experiment by increasing the sensor variances as well as the number of colluders. In this experiment, we quantified the number of iterations for the IF algorithm with reciprocal discriminant function (*dKVD-Reciprocal* and *RobustAggregate-Reciprocal* algorithms). The results obtained from this experiment show that the original version of the IF algorithm quickly converges (after around five iterations) to the skewed values provided by one of the attackers, while starting with an initial reputation provided by our approach, the algorithms require around 29 iterations, and, instead of converging to the skewed values provided by one of the attackers, it provides a reasonable accuracy.

The results of this experiment validate that our sophisticated attack scenario is caused by the discovered vulnerability in the IF algorithms which sharply diminishes the contributions of benign sensor nodes when one of the sensor nodes reports a value very close to the simple average.



(a) dKVD-Reciprocal (b) dKVD-Affine

(c) Zhou (d) Laureti

(e) RobustAggregate-Reciprocal (f) RobustAggregate-Affine

Fig. 6: Accuracy with our collusion attack.

5 Related Work

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 [8], [9], [10], [11], [12], [13], [14], [15]. We reviewed three of them in our comparative experiments in Section 4. Li et al. in [12] 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 algorithm in [13]. 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 black-list of users who are especially bad raters. Liao et al. in [14] proposed an iterative algorithm which beyond simply using the rating matrix, also uses the social network of users. The main objective of author in [15] 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 reputation systems in WSNs. Authors in [24] 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 community for sensor nodes in the network. Xiao et al. in [25] 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 [26] 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. Authors in [6] proposed an interdependency relationship between network nodes and data items for assessing their trust scores based on a cyclical framework. The main contribution of authors in [27] 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 [28] 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 enemies 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 aggregation problems in WSNs. Ho et al. in [29] 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 [30] 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. Lim et al. in [18] proposed a game-theoretical defense strategy to protect sensor nodes and to guarantee a high level of trustworthiness for sensed data. Moreover, there is a large volume of published studies in the area of secure tiny aggregation in WSNs [31], [32], [33]. 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 injection for a number of simple attack scenarios, to the best of our knowledge, no existing work addresses this issue in the case of a collusion attack by compromised nodes in a manner which employs high level knowledge about data aggregation algorithm used.

6 Conclusions

In this paper, we introduced a novel collusion attack scenario against a number of existing IF algorithms. Moreover, we proposed an improvement for 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. In future work, We will investigate whether our approach can protect against compromised aggregators. we also plan to implement our approach in a deployed sensor network.

References

[1] S. Ozdemir and Y. Xiao, "Secure data aggregation in wireless sensor 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: Springer.

[3] A. Jøsang and J. Golbeck, "Challenges for robust trust and reputation systems," in *Proceedings of the 5 th International Workshop on Security and Trust Management*, Saint Malo, France, 2009.

[4] K. Hoffman, D. Zage, and C. Nita-Rotaru, "A survey of attack and defense techniques for reputation systems," *ACM Comput. Surv.*, 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. Troubador Publishing Ltd, 2009, pp. 105–128.

[6] H.-S. Lim, Y.-S. Moon, and E. Bertino, "Provenance-based trustworthiness assessment in sensor networks," in *Proceedings of the Seventh International Workshop on Data Management for Sensor Networks*, ser. DMSN '10, 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: E^2MWSN," in *Wireless Communications, Networking and Mobile Computing (WiCOM), 2011 7th International Conference on*, 2011, pp. 1–4.

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

[9] Y. Zhou, T. Lei, and T. Zhou, "A robust ranking algorithm to spamming," *CoRR*, vol. abs/1012.3793, 2010.

[10] P. Laureti, L. Moret, Y.-C. Zhang, and Y.-K. Yu, "Information filtering via Iterative Refinement," *EPL (Europhysics Letters)*, vol. 75, pp. 1006–1012, Sep. 2006.

[11] Y.-K. Yu, Y.-C. Zhang, P. Laureti, and L. Moret, "Decoding information from noisy, redundant, and intentionally distorted sources," *Physica A Statistical Mechanics and its Applications*, vol. 371, pp. 732–744, Nov. 2006.

[12] R.-H. Li, J. X. Yu, X. Huang, and H. Cheng, "Robust reputation-based ranking on bipartite rating networks," in *SDM'12*, 2012, pp. 612–623.

[13] E. Ayday, H. Lee, and F. Fekri, "An iterative algorithm for trust and reputation management," in *Proceedings of the 2009 IEEE international conference on Symposium on Information Theory - Volume 3*, ser. ISIT'09, 2009, pp. 2051–2055.

[14] H. Liao, G. Cimini, and M. Medo, "Measuring quality, reputation and trust in online communities," *ArXiv e-prints*, Aug. 2012.

[15] B.-C. Chen, J. Guo, B. Tseng, and J. Yang, "User reputation in a comment rating environment," in *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, ser. KDD '11, 2011, pp. 159–167.

[16] C. T. Chou, A. Ignatovic, and W. Hu, "Efficient computation of robust average of compressive sensing data in wireless sensor networks in the presence of sensor faults," *Parallel and Distributed Systems, IEEE Transactions on*, vol. 24, no. 8, pp. 1525–1534, Aug 2013.

[17] Y. Yu, K. Li, W. Zhou, and P. Li, "Trust mechanisms in wireless sensor networks: Attack analysis and countermeasures," *Journal of Network and Computer Applications*, vol. 35, no. 3, pp. 867 – 880, 2012, ¡ce:title¿Special Issue on Trusted Computing and Communications¡/ce:title¿.

[18] H.-S. Lim, G. Ghinita, E. Bertino, and M. Kantarcioglu, "A game-theoretic approach for high-assurance of data trustworthiness in sensor networks," in *Data Engineering (ICDE), 2012 IEEE 28th International Conference on*, april 2012, pp. 1192 –1203.

[19] M. Rezvani, A. Ignjatovic, E. Bertino, and S. Jha, "Secure data aggregation technique for wireless sensor networks in the presence of collusion attacks," School of Computer Science and Engineering, UNSW, Tech. Rep. UNSW-CSE-TR-201319, July 2013.

[20] 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, 2004, pp. 78–87.

[21] "The Intel lab data," *Data set available at: http://berkeley.intel-research.net/labdata/*, 2004.

[22] B. Awerbuch, R. Curtmola, D. Holmer, C. Nita-rotaru, and H. Rubens, "Mitigating byzantine attacks in ad hoc wireless networks," Department of Computer Science, Johns Hopkins University, Tech, Tech. Rep., 2004.

[23] M. C. Vuran and I. F. Akyildiz, "Spatial correlation-based collaborative medium access control in wireless sensor networks," *IEEE/ACM Trans. Netw.*, vol. 14, no. 2, pp. 316–329, Apr. 2006.

[24] S. Ganeriwal, L. K. Balzano, and M. B. Srivastava, "Reputation-based framework for high integrity sensor networks," *ACM Trans. Sen. Netw.*, vol. 4, no. 3, pp. 15:1–15:37, Jun. 2008.

[25] X.-Y. Xiao, W.-C. Peng, C.-C. Hung, and W.-C. Lee, "Using SensorRanks for in-network detection of faulty readings in wireless sensor networks," in *Proceedings of the 6th ACM international workshop on Data engineering for wireless and mobile access*, ser. MobiDE '07, 2007, pp. 1–8.

[26] M. Li, D. Ganesan, and P. Shenoy, "PRESTO: feedback-driven data management in sensor networks," in *Proceedings of the 3rd conference on Networked Systems Design & Implementation - Volume 3*, ser. NSDI'06, 2006, pp. 23–23.

[27] Y. Sun, H. Luo, and S. K. Das, "A trust-based framework for fault-tolerant data aggregation in wireless multimedia sensor networks," *IEEE Trans. Dependable Secur. Comput.*, vol. 9, no. 6, pp. 785–797, Nov. 2012.

[28] L.-A. Tang, X. Yu, S. Kim, J. Han, C.-C. Hung, and W.-C. Peng, "Tru-Alarm: Trustworthiness analysis of sensor networks in cyber-physical systems," in *Proceedings of the 2010 IEEE International Conference on Data Mining*, ser. ICDM '10, 2010, pp. 1079–1084.

[29] J.-W. Ho, M. Wright, and S. Das, "ZoneTrust: Fast zone-based node compromise detection and revocation in wireless sensor networks using sequential hypothesis testing," *Dependable and Secure Computing, IEEE Transactions on*, vol. 9, no. 4, pp. 494 –511, july-aug. 2012.

[30] S. Ozdemir and H. Çam, "Integration of false data detection with data aggregation and confidential transmission in wireless sensor networks," *IEEE/ACM Trans. Netw.*, vol. 18, no. 3, pp. 736–749, Jun. 2010.

[31] H. Chan, A. Perrig, and D. Song, "Secure hierarchical in-network aggregation in sensor networks," in *Proceedings of the 13th ACM Conference on Computer and Communications Security*, ser. CCS '06. New York, NY, USA: ACM, 2006, pp. 278–287.

[32] Y. Yang, X. Wang, S. Zhu, and G. Cao, "SDAP: a secure hop-by-hop data aggregation protocol for sensor networks," in *MobiHoc*, 2006, pp. 356–367.

[33] S. Roy, M. Conti, S. Setia, , and S. Jajodia, "Secure data aggregation in wireless sensor networks," *Information Forensics and Security, IEEE Transactions on*, vol. 7, no. 3, pp. 1040–1052, 2012.

Mohsen Rezvani got his Bachelors and Masters degrees in Computer Engineering at the Amirkabir University of Technology and Sharif University of Technology, respectively. He is a PhD candidate in the School of Computer Science and Engineering at the University of New South Wales, Sydney, Australia. His research focuses on trust and reputation systems in WSNs. He is also a student member of the IEEE and the IEEE Communications Society.

Aleksandar Ignjatovic got his Bachelor and Master degrees in Mathematics at the University of Belgrade, former Yugoslavia, and Ph.D. in Mathematical Logic at the University of California at Berkeley. After graduation he was an Assistant Professor at the Carnegie Mellon University, where he taught for 5 years at the Department of Philosophy and subsequently had a startup in the Silicon Valley. Aleks joined in 2002 the School of Computer Science and Engineering at UNSW, where he is teaching algorithms. His research interests include sampling theory and signal processing, applications of mathematical logic to computational complexity theory, algorithms for embedded systems design and most recently trust based data aggregation algorithms.

Elisa Bertino is Professor of Computer Science at Purdue University, and serves as research director of the Center for Education and Research in Information Assurance and Security (CERIAS) and Interim Director of Cyber Center (Discovery Park). Previously, she was a faculty member and department head at the Department of Computer Science and Communication of the University of Milan. Her main research interests include security, privacy, digital identity management systems, database systems, distributed systems, and multimedia systems. She is currently serving as chair of the ACM SIGSAC and as a member of the editorial board of the following international journals: IEEE Security and Privacy, IEEE TRANSACTIONS ON SERVICE COMPUTING, ACM Transactions on Web. She also served as editor-in-chief of the VLDB Journal and editorial board member of ACM TISSEC and IEEE TDSC. She coauthored the book Identity ManagementConcepts, Technologies, and Systems. Dr. Bertino is a fellow of the ACM. She received the 2002 IEEE Computer Society Technical Achievement Award for outstanding contributions to database systems and database security and advanced data management systems and the 2005 IEEE Computer Society Tsutomu Kanai Award for pioneering and innovative research contributions to secure distributed systems.

Sanjay Jha received the PhD degree from the University of Technology, Sydney, Australia. He is a professor and head of the Network Group at the School of Computer Science and Engineering at the University of New South Wales. He is an associate editor of the IEEE Transactions on Mobile Computing. He was a Member-at-Large, Technical Committee on Computer Communications (TCCC), IEEE Computer Society for a number of years. He has served on program committees of several conferences. He was cochair and general chair of the Emnets-1 and Emnets-II workshops, respectively. He was also the general chair of the ACM Sensys 2007 symposium. His research activities include a wide range of topics in networking including wireless sensor networks, adhoc/community wireless networks, resilience/quality of service (QoS) in IP networks, and active/programmable networks. He has published more than 100 articles in high quality journals and conferences. He is the principal author of the book Engineering Internet QoS and a coeditor of the book Wireless Sensor Networks: A Systems Perspective. He is a senior member of the IEEE and the IEEE Computer Society.