SECURE DATA TRANSMISSION IN WIRELESS SENSOR NETWORKS USING AGGREGATION TECHNIQUE

[1]R.Rajani, Professor, Dept of MCA, Narayana Engineering College, Nellore

[2]K .Chandana, Student, Dept of MCA, Narayana Engineering College, Nellore

[3]N.Kavya, Student, Dept of MCA, Narayana Engineering College, Nellore

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 desire for lustiness of observance and low price of the nodes, wireless sensing element networks (WSNs) area unit sometimes redundant. Information from multiple sensors is aggregate at aggregator node that then forwards to the base station solely the combination values. At present, owing to limitations of the computing power and energy resource of sensing element nodes, information is aggregate by very simple algorithms like averaging. However, such aggregation is thought to be terribly prone to faults, and additional significantly, malicious attacks. This can not be remedied by cryptographical ways, as a result of the attackers typically gains complete access to information holds on within the compromised nodes. For that reason aggregation at the aggregator node must be amid associate assessment of trustworthiness of knowledge from individual sensing element nodes. Thus, better, additional refined algorithms area unit required for data aggregation within the future WSN. Such associate formula ought to have 2 options.

1. Within the presence of random errors such formula ought to turn out estimates that area unit near the optimal ones in information theoretic sense. Thus, for instance, if the noise gift in every sensing element may be a Gaussian severally distributed noise with zero mean, then the estimate made by such associate formula ought to have a variance near the CramerRao bound (CRLB) , i.e, it ought to be near the variance of the most chance figurer (MLE). However, such estimation ought to be achieved while not provision to the formula the variances of the sensors, unobtainable in observe.

2. The formula ought to even be strong within the presence of non-stochastic errors, like faults and malicious attacks, and, besides aggregating information; such formula ought to conjointly offer associate assessment of the reliableness and trustworthiness of the information received from every sensing element node.

Trust and reputation are recently recommended as an efficient security mechanism for Wireless sensing element Networks. Though sensing element networks square measure being progressively deployed in several application domains, assessing trait of reportable knowledge from distributed sensors has remained a difficult issue. Sensors deployed in hostile environments could also be subject to node compromising attacks by adversaries shall inject false knowledge into the system. During this context, assessing the trustworthiness of the collected knowledge becomes a difficult task.

Iterative Filtering (IF) algorithms are a gorgeous possibility for WSNs as a result of they solve each problems—data aggregation and data trustworthiness assessment—using single iterative procedure .Such trustworthiness estimate of every sensing element is predicated on the gap of the readings of such a sensing element from the estimate of the right values, obtained within the previous spherical of iteration by some style of aggregation of the readings of all sensors. Such aggregation is sometimes a weighted average; sensors whose readings significantly disagree from such estimate are allotted less trustworthiness and consequently within the aggregation method within the gift spherical of iteration their readings are given a lower weight.

This paper presents a replacement subtle collusion attack situation against variety of existing IF algorithms supported the false knowledge injection. In such AN attack situation, colluders arrange to skew the combination price by forcing such IF algorithms to converge to skew values provided by one in every of the attackers.

As we are going to see, vulnerability to classy collusion attacks comes from the actual fact that these IF algorithms begin the iteration method by giving AN equal trust price to any or all sensing element nodes. During this paper, we have a tendency to propose {a solution| an Answer} for such vulnerability by providing an initial trust estimate that relies on a strong estimation of errors of individual sensors. Once the character of errors is random, such errors basically represent AN approximation of the error parameters of sensing element nodes in WSN like bias and variance. However, such estimates additionally sway be strong in cases once the error isn't random however attributable to coordinated malicious activities. Such initial estimation makes IF algorithms strong against delineated subtle collusion attack, and, we believe, additionally strong beneath significantly additional general circumstances; for instance, it's additionally effective within the presence of a whole failure of a number of the sensing element nodes.

We validate the performance of our algorithm by simulation on synthetically generated data sets. Our simulation results illustrate that our strong aggregation technique is effective in terms of lustiness against our novel subtle attack situation additionally as efficient in terms of the computational cost.

Our contributions are summarized as follows:

1.Identification of a replacement subtle collusion attack against IF based mostly reputation systems that reveals a severe vulnerability of IF algorithms.

2. A completely unique methodology for estimation of sensing elements errors that is effective during a wide selection of sensor faults and not vulnerable to the represented attack.

3. Style of an efficient and strong aggregation methodology galvanized by the MLE that utilizes an estimate of the noise parameters obtained using contribution 2 above.

4. Increased IF schemes ready to shield against subtle collusion attacks by providing an initial estimate of trustworthiness of sensors using inputs from contribution 2 and 3 above.

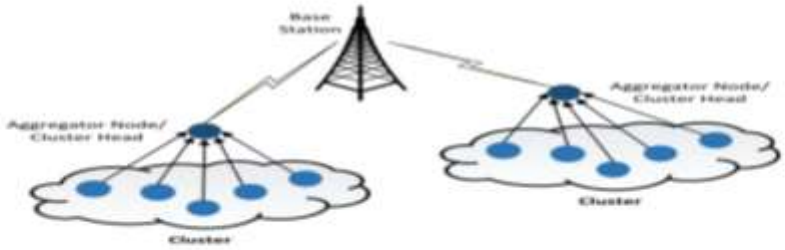


Fig. 1. Network model for WSN.

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

Fig. 1 shows our assumption for network model in WSN. The sensor nodes area unit divided into disjoint clusters, and every cluster incorporates a cluster head that acts as aggregator node. Data are periodically collected and aggregated by the aggregator. In this paper we tend to assume that the aggregator itself isn't compromised and consider algorithms that build aggregation secure once the individual sensor nodes may be compromised and may be causing false information to the aggregator. We tend to assume that every data aggregator has enough procedure power to run associate degree IF formula for data aggregation.

2.2 Iterative Filtering in Reputation Systems

We consider a WSN with n sensors Si, i =1,...,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 $=\{x_i^1, x_i^2 .. x_i^m\}$ where xi $=\{x_i^1, x^2_i \ldots x^m_i\}^T$, $(1 \leq i \leq n)$ represents the ith m -dimensional reading reported by sensor node S_i. Let r $[r_1, r_2, \ldots. r_m]^T$ denote the aggregate values for instants t =1,...,m, which authors of call a reputation vector,2 computed iteratively and simultaneously with a sequence of weights $w=[w_1\ w_2 \ldots w_n]^T$ reflecting the trustworthiness of sensors. We denote by $r^{(l),}$ $w^{(l)}$ the approximations of r; w obtained at lth round of iteration $(l \geq 0)$. The iterative procedure starts with giving equal credibility to all sensors, i.e., with an initial value $w^{(0)}=1$. The value of the reputation vector $r^{(l+1)}$ in round of iteration l+1 is obtained from the weights of the sensors obtained in the round of iteration l as

$r(l+1)=x.w(l)/\sum^n{}_{i=1} w_i^{(l)}$

Consequently, the initial reputation vector is $r^{(1)}=1/nX\text{-}1$, i.e., $r^{(l)}$ is just the sequence of simple averages of the readings of all sensors at each particular instant. The new weight vector $w^{(l+1)}$ to be used in round of iteration l+1 is then computed as a function g(d) of the normalized belief divergence d which is the distance between the sensor readings and the reputation vector $r^{(l).}$ Thus, $d=[\ d1\ d2 \ldots dn]^T$,d i = $l/m\|x_i - r^{(l+1)}\|_2^2$ and $w_i^{(l+1)}=g(d_i),(1 \leq i \leq n)$.

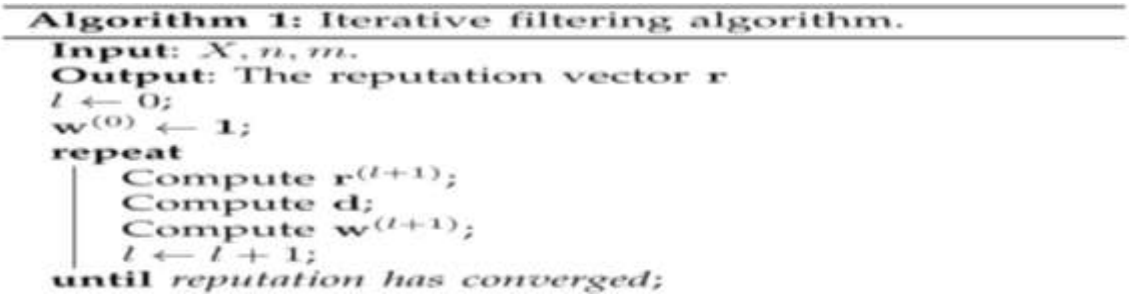
Function g(x) is called the discriminant function and it provides an inverse relationship of weights to distances d. Our experiments show that selecting a discriminant function has a significant role in stability and robustness of IF algorithms. A number of alternatives for this function are studied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

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

2.3 Adversary Model

In this paper, we tend to use a Byzantine attack model, wherever the adversary will compromise a collection of sensor nodes and inject any false data through the compromised nodes.

We tend to assume that sensors are deployed in a hostile unattended environment. Consequently, some nodes is physically compromised. We tend to assume that once a sensor node is compromised, all the data that is within the node becomes accessible by the adversary. Thus, we cannot have faith in crypto logic strategies for preventing the attacks, since the adversary might extract crypto logic keys from the compromised nodes. We tend to assume that through the compromised sensor nodes the adversary will send false information to the aggregator with a purpose of distorting the aggregate values. we tend to assume that every compromised node is in check of one adversary or a colluding cluster of adversaries, enabling them to launch a sophisticated attack. We tend to conjointly think that the adversary has enough data about the aggregation formula and its parameters. Finally, we tend to assume that the base station and aggregator nodes can't be compromised during this adversary model; there's an in depth literature proposing the way to agitate drawback of compromised aggregators; during this paper we tend to limit our attention to the lower layer problem of false data being sent to the aggregator by compromised individual sensor nodes, that has received a lot of less attention within the existing literature.



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

2.4 Collusion Attack Scenario

Most of the IF algorithms use easy assumptions concerning the initial values of weights for sensors. Just in case of our adversary model, an attacker is able to mislead the aggregation system through careful choice of reported data values. We have a tendency to use visualization techniques from to present our attack scenario.

Assume that ten sensors report the values of temperature that square measure aggregate using the IF algorithmic program planned in with the reciprocal discriminant function. We have a tendency to take into account 3 attainable scenarios; see Fig. 2.

- In scenario one, all sensors are reliable and also the results of the IF algorithmic program is near to the actual value.
- In scenario two, associate adversary compromises 2 sensor nodes, and alters the readings of those values specified the easy average of all sensor readings is skew towards a lower value. As these 2 sensor nodes report a lower value, IF algorithmic program penalises them and assigns to them lower weights, as a result of their values square measure removed from the values of different sensors. In different words, the algorithmic program is powerful against false data injection during this scenario as a result of the compromised nodes severally falsify the readings with none data concerning the aggregation algorithmic program.
- In scenario three, adversary employs 3 compromised nodes so as to launch a collusion attack. It listens to the reports of sensors within the network and instructs the 2 compromised sensor nodes to report values removed from true value of the measured amount. It then computes the skew value of the easy average of all sensor readings and commands the third compromised sensor to report such skew average as its readings. In different words, 2 compromised nodes distort the easy average of readings, whereas the third compromised node reports a worth terribly near such distorted average so creating such reading seem to the IF algorithmic program as a extremely reliable reading. As a result, IF algorithms can converge to the values provided by the third compromised node, as a result of within the first iteration of the algorithmic program the third compromised node can succeed the best weight, significantly dominating the weights of all different sensors. This can be reinforced in each future iteration; thus, the algorithmic program quickly converges to a reputation that is extremely near the initial skew easy average, as shown in Fig. 2.

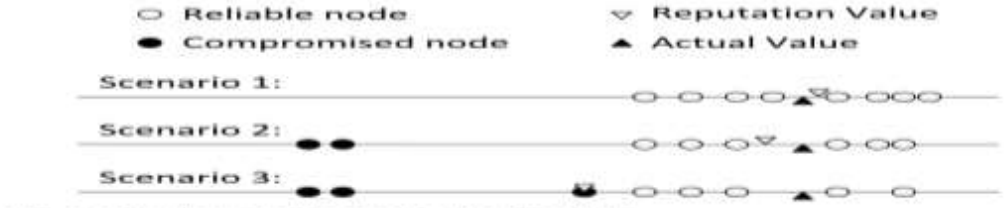


Fig. 2. Attack scenario against IF algorithm.

3 ROBUST DATA AGGREGATION

In this section, we present our robust data aggregation method.

3.1 Framework Overview

In order to enhance the performance of IF algorithms against the said attack situation, we offer a strong initial estimation of the trustiness of sensing element nodes to be employed in the first iteration of the IF algorithmic program. Most of the standard applied mathematics estimation strategies for variance involve use of the sample mean. For this reason, proposing a strong variance estimation technique within the case of inclined sample mean is an important a part of our methodology.

Fig. 3 illustrates the stages of our strong aggregation framework and their interconnections. As we've got mentioned, our aggregation methodology operates on batches of consecutive readings of sensors, continuing in many stages. within the first stage we provide an initial estimate of 2 noise parameters for sensor nodes, bias and variance; details of the computations for estimating bias and variance of sensors square measure bestowed in Sections 3.2 and 3.3, severally.

Based on such associate estimation of the bias and variance of every sensing element, the bias estimate is subtracted from sensors readings and within the next part of the planned framework, we offer an initial estimate of the reputation vector calculated exploitation the MLE.

In the third stage of the projected framework, the initial name vector provided within the second stage is employed to estimate the trustiness of every detector supported the space of detector readings to such initial name vector.

3.2 Estimating Bias

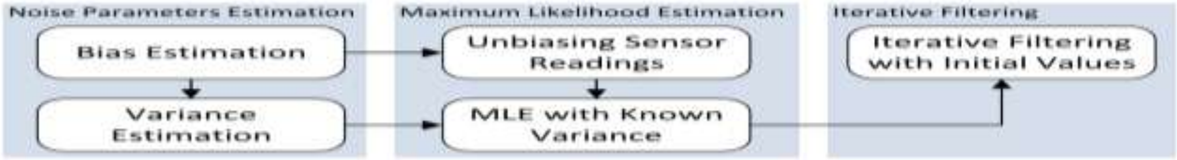
We assume that all sensors in WSN can have some error; such error e^t_s of a sensor s is modeled by the Gaussian distribution random variable with a sensor bias b_s and sensor variance σ_{st} $e^t_s \sim N(b_s,\sigma^2_s)$. Let r_t denotes the true value of the signal at time t. Each sensor reading x^t_s can be written as

$$x^t_s = r_t + e^t_s. \qquad (1)$$

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^t_s; however, we can obtain the values of the differences of such errors. Thus, if we define

$\delta(i,j) = \frac{1}{m}\sum_{t=1}^{m}(x_i^t - x_j^t)$, we get

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


Noise Parameters Estimation
Bias Estimation
Variance Estimation
Maximum Likelihood Estimation
Unbiasing Sensor Readings
MLE with Known Variance
Iterative Filtering
Iterative Filtering with Initial Values

ig. 3. Our robust data aggregation framework.

where e^t_i is a random variable with Gaussian distribution $e^t_i \sim \mathcal{N}(b_i, \sigma_i^2)$. Let $\bar{e}_i = \frac{1}{m}\sum_{t=1}^{m} e^t_i$ 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 $\delta = \{\delta(i,j) \,:\, 1 \le i,j \le n\}$; this matrix is an estimator for mutual difference of sensor bias. In order to obtain the sensor bias from this matrix, we solve the following minimization problem.

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

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 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 and look at external 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

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

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

$$\begin{cases} \sum_{i=1, i\neq k}^{n} \frac{2}{d(i,k)^2} b_i - \sum_{i=1, i\neq k}^{n} \frac{2}{d(i,k)^2} b_k - \lambda = 2\sum_{i=1, i\neq k}^{n} \frac{1}{d(i,k)}, \\ \qquad\qquad \text{for all } k = 1,\ldots,n \\ \sum_{i=1}^{n} b_i = 0. \end{cases} \qquad (3)$$

Note that the obtained value of bi 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 $b=[b_1,b_2...b_n]$and sensor readings $\{x^t_s\}$, we can define matrices $\{x^t_s\}$and β={β(i, j)} as follows:

$$\tilde{x}^t_s = x^t_s - b_s. \qquad (4)$$

$$\begin{aligned} \beta(i,j) &= \frac{1}{m-1}\sum_{t=1}^{m}(\tilde{x}^t_i - \tilde{x}^t_j)^2 \\ &= \frac{1}{m-1}\sum_{t=1}^{m}((x^t_i - x^t_j) - (b_i - b_j))^2. \end{aligned}$$

By (1) we have $x^t_i - x^t_j = (r_t + e^t_i) - (r_t + e^t_j) = e^t_i - e^t_j$; thus, we obtain

$$\begin{aligned} \beta(i,j) = & \frac{1}{m-1}\sum_{t=1}^{m}(e^t_i - b_i)^2 + \frac{1}{m-1}\sum_{t=1}^{m}(e^t_j - b_j)^2 \\ & - \frac{2}{m-1}\sum_{t=1}^{m}(e^t_i - b_i)(e^t_j - b_j). \end{aligned}$$

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

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

and similarly

$$\begin{aligned} \beta(i,j) &= \frac{1}{m-1}\sum_{t=1}^{m}(e^t_i - b_i)^2 + \frac{1}{m-1}\sum_{t=1}^{m}(e^t_j - b_j)^2 \\ &\approx \sigma_i^2 + \sigma_j^2. \end{aligned}$$

The above formula shows that we can estimate the variance of sensors noise by computing the matrix β.

3.4 MLE with Known Variance

In the previous sections, we tend to project a completely unique approach for estimating the bias and variance of noise for sensors supported their readings. The variance and therefore the bias of a sensing element noise may be taken because the distance measures of the sensing element readings to truth worth of the signal. In fact, the space measures obtained as our estimates of the bias and variances of sensors additionally add up for non-stochastic errors.

Given matrix $\{x^t_s\}$ where $x^t_s \sim r_t + \mathcal{N}(b_s, \sigma_s^2)$ and estimated bias and variance vectors b and σ, 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^2_i by the obtained variance vi from Equation .Moreover, by differentiating the above formula with respect to r_t and setting the derivative equal to zero we get

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

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:

$$\mathbf{r} - \sum_{s=1}^{n} w_s \mathbf{x}_s. \qquad (8)$$

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 projected attack situation, the assailant exploits the vulnerability of the IF algorithms that originates from a wrong assumption concerning the initial trustiness of sensors. Our contribution to deal with these short comings is to use the results of the projected sturdy information aggregation technique because the initial name for these algorithms. Moreover, the initial weights for all sensor nodes will be computed supported the space of sensors readings to such associate initial name. Our experimentalresultsillustratethat thisconcept not solely consolidatestheIFalgorithmsagainstthe projected attack situation, however victimizati on this first name improves the efficiency of the IF algorithms by reducing the amount of iterations required to approach a stationary purpose at intervals the prescribed tolerance; see Section four.2.

4 RESULTS

The target of our experiments is to judge the hardiness and efficiency of our approach for estimating verity values of signal supported the detector readings within the presence of faults and collusion attacks.

4.1 Experimental Settings

All the experiments have been conducted on an HP PC with 3.30 GHz Intel Core i5-2500 processor with 8 GB RAM running a 64-bit Windows 7 Enterprise. The program code has been written in MATLAB R2012b. we generate synthetic data sets according to the following parameters:

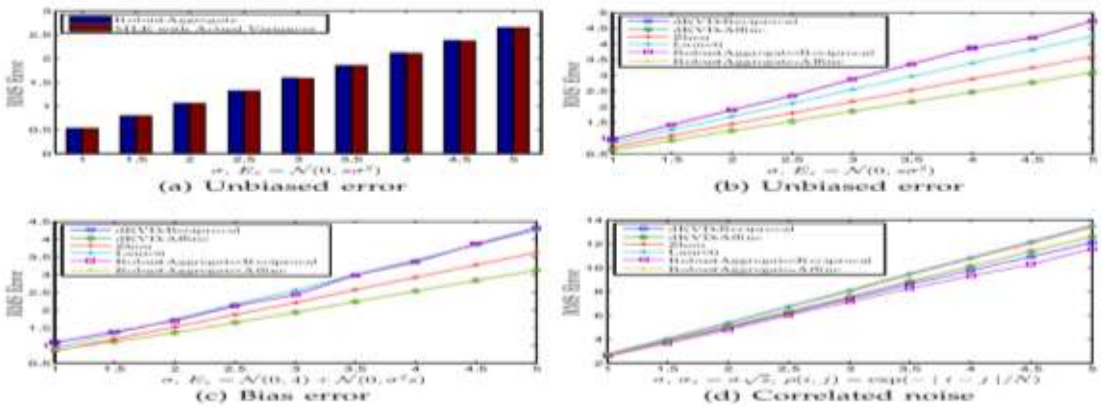
- Each simulation experiment was repeated 200 times and then results were averaged.
- Number of sensor nodes is n = 20.
- Number of readings for each sensor is m = 400.
- For statistical parameters of the errors (noise) used to corrupt the true readings, we consider several ranges of values for bias, variance and covariance of noise for each experiment.
- The level of significance in K-S test is α= 0:05.

4.2 Accuracy and Efficiency without an Attack

In the first batch of experiments we assume that there are no sensors with malicious behavior. 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 data sets.

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 $e^{t}\sim_{N(0,s\times\sigma^{2})}$, considering different values for the baseline sensor variance s2. Fig. 4a shows the results of the MLE with our noise parameter estimation(step1and 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 nearly exactly achieves the minimal possible variance coming from the information theoretic lower bound

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 $e^{t}_{s\,\sim N(N(0,\sigma^{2}_{b});s\times\sigma^{2})}$with the variance of the bias $\sigma^{2}{}_{b}$ = 4 and increasing values for variances, where the variance of sensor s is equal to s×σ^2. Thus, the sensors bias is produced by a zero mean Gaussian distribution random variable. Fig. 4c 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.



(a) Unbiased error (b) Unbiased error
(c) Bias error (d) Correlated noise

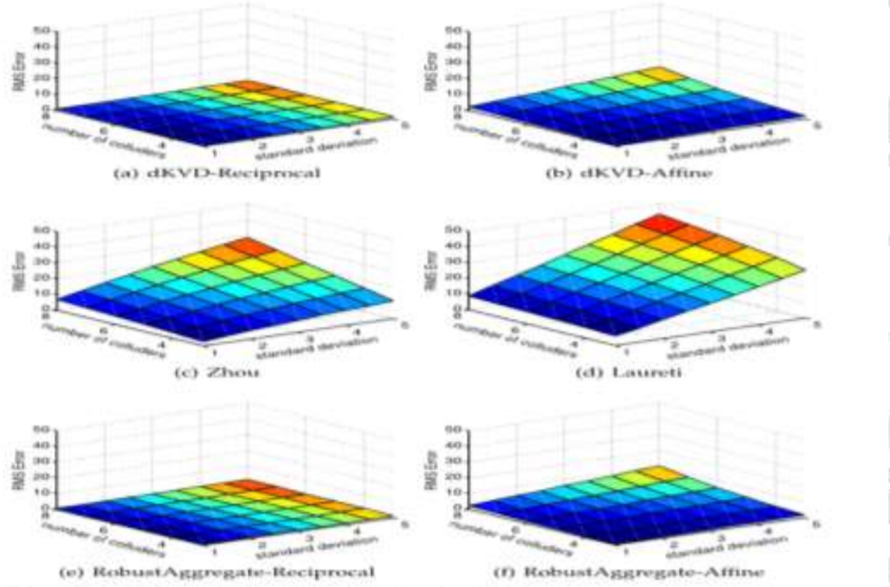
Fig. 4. Accuracy for No Attack scenarios

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. Moreover, the scale of RMS error is in general larger than in scenarios with uncorrelated noise, as one would expect.

4.3 Accuracy with Simple Attack Scenario

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 nodes 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 sensors readings through reporting outlier readings by the compromised nodes.

Fig. 5 shows the accuracy of the IF algorithms and our approach in the presence of such simple attack scenario



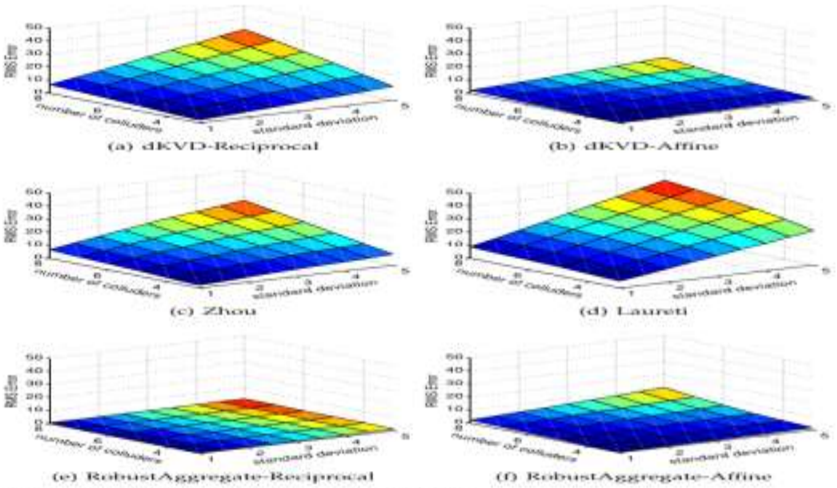
(a) dKVD-Reciprocal (b) dKVD-Affine
(c) Zhou (d) Laureti
(e) RobustAggregate-Reciprocal (f) RobustAggregate-Affine

Fig. 5. Accuracy with a simple attack scenario.

4.4 Accuracy with a Collusion Attack

Fig. 6 shows the accuracy of the IF algorithms and our approach within the presence of the collusion attack state of affairs. It will be seen that the IF algorithms with reciprocal discriminant operate are extremely prone to such attack state of affairs (see Figs. 6a and 6d), whereas the affine discriminant operate generates a lot of strong ends up in this case (see Fig. 6b). However, the accuracy of the affine discriminant operate remains abundant worse than the previous experiment while not the collusion attack.

Figs. 6e and 6f show the accuracy of our approach by taking under consideration the IF algorithmic program in with reciprocal and affine discriminant functions, severally. Joined will see, our projected approach is superior to any or all alternative algorithms in terms of the accuracy for reciprocal discriminant functions, whereas the approach encompasses a terribly little improvement on affine operate.



(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 knowledge aggregation may be a serious concerning WSNs and there square measure variety of papers investigation malicious knowledge injection by taking into consideration the assorted oppose models. There square measure 3 bodies of labor associated with our research: IF algorithms trust and name systems for WSNs, and secure knowledge aggregation with compromised node detection in WSNs.

6 CONCLUSIONS

In this paper, we have a tendency to introduced a completely unique collusion attack situation against variety of existing IF algorithms. Moreover, we have a tendency to planned associate degree improvement for the IF algorithms by providing associate degree initial approximation of the trait of detector nodes that makes the algorithms not solely collusion sturdy, however conjointly additional correct and quicker convergence.In future work, we are going to investigate whether or not our approach will defend against compromised aggregators. We have a tendency to conjointly arrange to implement our approach during a deployed detector network.

REFERENCES

[1] S. Ozdemir and Y. Xiao, “Secure data aggregation in wireless sensor networks: A comprehensive overview,” Compute. 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 reputation 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. Surveys, 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 trustworthiness assessment in sensor networks,” in Proc. 7th Int. Workshop 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, 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,” Europhys. Lett., vol. 94, p. 48002, 2011.

[10] P. Laureti, L. Moret, Y.-C. Zhang, and Y.-K. Yu, “Information filtering via iterative refinement,” Europhys. Lett., 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: Statist. Mech. Appl., vol. 371, pp. 732–744, Nov. 2006.

[12] R.-H. Li, J. X. Yu, X. Huang, and H. Cheng, “Robust reputationbased ranking on bipartite rating networks,” in Proc. SIAM Int. Conf. Data Mining, 2012, pp. 612–623.