Trustworthy and Protected Data Collection for Event Detection Using Networked Sensing Systems

Md Zakirul Alam Bhuiyan[⋆,†] and Jie Wu[⋆]
[⋆]Department of Computer and Information Sciences, Temple University, Philadelphia, PA 19122, USA
[†]Department of Computer and Information Sciences, Fordham University, New York, NY 10458, USA
Email: zakirulalam@gmail.com and jiewu@temple.edu

***Abstract*—Data collection in wireless networked sensing systems (WNSS) is usually not reliable due to sensor faults and/or security attacks. This makes detection of an event (e.g., structural damage) through data aggregation unreliable. In this paper, we propose a trustworthy and protected data collection (TPDC) framework for event detection in WNSS. This framework facilitates reliable data for aggregation at clusters of WNSS. The key idea of TPDC is to allow a cluster head to check whether or not the transmitted data is trustworthy (i.e., unaltered estimated at the sensor node level) and protected (i.e., received without alteration after the transmission) before aggregating the data at a cluster. For the trustworthy data, we propose an algorithm to make sure that transmitted data is unaltered. For the protected data, we present a truth discovery approach, whose goal is to infer truthful facts from unreliable sensor data. Through simulations , we demonstrate that the collected data in TPDC is trustworthy and protected that can make aggregation for event detection reliable.**

***Index Terms*—Networked sensing systems, data collection, fault tolerance, trustworthy, security, correlation**

I. Introduction

With the capabilities of pervasive surveillance, wireless networked sensing systems (WNSS) have strong practical applications in many domains, e.g., crowd sensing, structural health monitoring (SHM) or damage event detection for industrial machines or infrastructures, chemical explosions, and military surveillance intrusion tracking [1]–[4]. In most sensing applications, the quality of the data or the quality of the monitoring and timely detection of an even are the utmost important issues. This is particularly true for events like structural damage or fire where an employed system should be able to detect the acquired data faults online and take recovery actions immediately to avoid meaningless monitoring operations [5]. In fact, reliability is highly desired in structural damage event detection, as an "alert" about a structural event conveys a serious concern with public safety and economic losses.

However, transmitted data provided by sensors is usually unreliable due to various reasons such as sensor fault, lack of sensor calibration, background noise, incomplete views of observations, and alterations (by security attacks) [6]. Without identifying any alternation of the acquired data considering these reasons, when the data is transmitted towards the upstream sensors (such as clusters) for aggregation, the data cannot be "trustworthy." Such data highly impacts the overall monitoring quality.

In addition, the power of networked sensing can be unleashed only by properly aggregating unreliable information from different sensors whose submitted data may be altered before transmission at the sensor. Regardless of if the transmitted data is trustworthy or not, they can be further altered during transmission from sensors to clusters by the third party. Some sensors constantly provide truthful data while others may generate biased, compromised, or even fake data, due to security attacks such as the collusion attack and the malicious attack [7], [8]. For that reason, data aggregation at a cluster head node has to be accompanied by an assessment of trustworthiness of data from individual sensor nodes. Therefore, the received data should be "protected" data before aggregation. Thus, it is indispensable to distinguish whether received data is protected or not before aggregation at a cluster head (CH).

Taking these two aspects into account, decision-making in existing event detection frameworks through traditional aggregation methods (e.g., voting, avg, sum) that regard all the users equally would not be able to derive accurate aggregated results for the event detection [9], [10]. Also, existing event detection based on results of weighted average, sum, or voting may not provide true facts in terms of trustworthy or untrustworthy acquired data [2], [11], [12].

In this paper, we propose a trustworthy and protected data collection (TPDC) framework for event detection in WNSS. This framework facilitates data collection for reliable aggregation at clusters of WNSS. The key idea of TPDC is to allow a cluster head to check whether or not the transmitted data is trustworthy (unaltered estimated at the sensor node level) and protected (received without alteration after the transmission) before aggregating the data at a cluster. For the trustworthy data, we use a general measurement model, mutual information independence (MII), between two signals from two different sensors or the same sensor for evaluating results in the absence of the ground truth. We think that mutual statistical information could be used as an indicator to check whether or not the acquired data is trustworthy in conjunction with structural damage event detection. Once the data passes this check, the data is trustworthy and can be transmitted towards the CH.

For protecting the data, we present a truth discovery approach whose goal is to infer truthful facts from unreliable sensors. To achieve this, TPDC performs sensor status value calculation and data encryption. An intended recipient, like

the CH, for example, should check if the received data is protected. We conduct a performance evaluation of TPDC through extensive simulations. We use real-world data set in the simulations, and demonstrate that the collected data in TPDC is trustworthy and protected that can makes aggregation for event detection reliable.

The remainder of this paper is organized as follows. Section II briefly discusses framework. Section III provides the trustworthy data collection approach. Section IV truth status discovery approach. Section V evaluates our TPDC framework. Finally, we conclude the paper in Section VI.

II. TPDC Framework

In this section, we describe our trustworthy and protected data collection (TPDC) framework. Let us consider a hierarchical WNSS with a set of sensors to be deployed for a particular monitoring application, e.g., monitoring the health of civil structures, e.g., building, bridge, aircraft. A reference 2D building model is shown in Fig. 1a, where sensors (white circle) are deployed according to a engineering-driven deployment method [1] and a remote monitoring center or a base station (BS) station location (colored circle) which is at a remote place. The deployed sensors are self-organized into clusters using some clustering algorithm [13], [14]. Every cluster head (CH) forwards a final decision of an event or aggregated data to the BS. We assume sensors can have different types of application tasks (e.g., sensing the vibration, strain, and damping, pressure, temperature, etc., in the context of SHM) and sending its measurements to neighboring nodes or a CH. For simplicity, we consider vibration signals in this paper.

Sensors acquire data using a state-space model [15], [16], analyze it locally, identify whether the acquired data is trustworthy or not, and finally transmit the trustworthy data. To identify trustworthy data, we focus on the following set of sensor faults that occur in a real wireless SHM system, sensor debonding fault, faulty signals by precision degradation, breakage, etc., especially in vibration signal capturing, faults in offset, bias, and the amplification gain factor of signals. Sensors may also produce abnormal signals from security attacks.

We use 'mutual information independence (MII)' as an indirect signal measurement, assuming that a prior correlation model denoted by C presents [17]. Model C can be given by a reference data set. This data set is immediately-stored data in the sensor local memory after the sensor network system initialization. A MII function between two signals of sensors i and j at time t in a cluster is applied to check abnormal signals.

In practical WNSS systems, the security threats mainly come from the parties themselves (i.e., from any sensors). The CH may try to deduce the observation of each sensor. On the other hand, each sensor may also try to infer the information of other parties. Therefore, it is of paramount importance to preserve sensor observation values (without alteration). Data received at the CH with some alteration are not considered

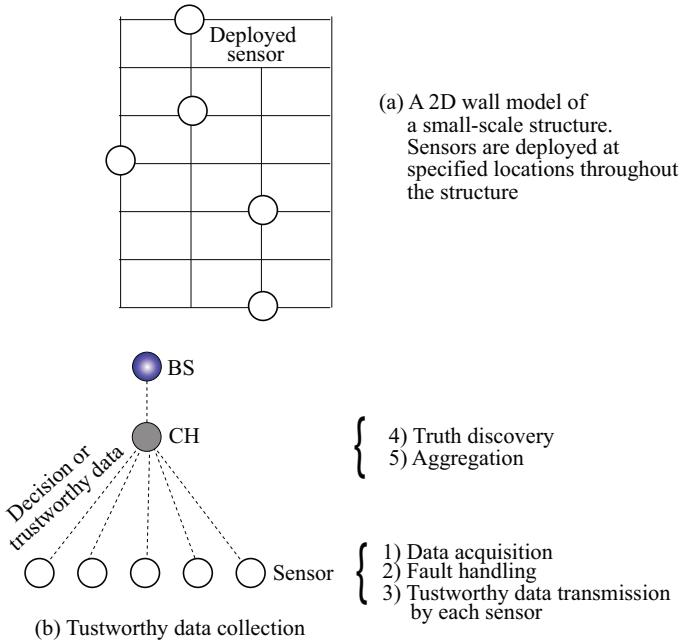

Fig. 1. WNSS-based SHM frameworks.

for aggregation for event detection. We use a sensor status value truth discovery for this purpose. In the truth discovery, the sensors that provide true information will be considered having truthful facts more often and the information that is supported by reliable sensors will be regarded as truths.

III. Trustworthy Data Collection

We consider data collection for SHM applications [1]. SHM techniques rely on measuring structural responses to ambient vibrations, strain signals, or forced excitation. A variety of sensors, such as accelerometers, strain gauges, or displacement can be used to measure structural responses. The civil or structural engineering communities use various data collection techniques. We consider the *state space model*, which is widely accepted for data collection that can accurately capture the structural dynamics [15].

We propose an algorithm that simply presents the data collection method in a cluster denoted by D. While theoretically this procedure involves multi-hop communication, consider the fact that for SHM application, the radio communication range of current sensor nodes exceeds the area in which the sensors gather signals. We limit sensors to communicate within the one-hop neighboring nodes. This algorithm has three steps. In step 1 of the algorithm, every sensor acquires signals captured from the vibration responses of the structure (bridge, building, etc.), and buffers them temporarily. Then, it transmits and receives the measured signals. In step 2, the sensors check if there are any faulty signals.

Step 3 executes another algorithm called "decision-making on the faulty signals" that identifies if collected sensor signals are faulty or not. When a remarkable change appears in a sensor's signals, there is a possibility that a sensor is faulty or

compromised. The MII is used to detect sensor signals. Let us consider the statistical dependency between the two sensors' signals quantified by MII. ω measures the information about one sensor that is shared by another sensor in the set of signals in D. It is seen that ω changes as soon as a sensor signal fault occurs because the faulty signal is not present in the reference or in other sensor signals.

Decision-making on the faulty signals. In order to make a decision about the sensor acquired signals, we use a joint Gaussian distribution based correlation model. Multivariate Gaussian distribution has been used to accurately model the correlation of many types of signals in literature [10]. Each signal is broadcasted to sensors in cluster D, where ith sensor signal $y_i^t \in y_D^t$ and jth sensor signal $y_j^t \in y_D^t$, $i, j \in D$. For simplicity, y_i^t as u and y_j^t as v are denoted hereafter.

Hence, it would be worth considering how to find joint probability density between two signals u and v. The statistical dependency/independency between the two Gaussian distributed time signals u and v can be expressed in the form of the joint probability density $p(u, v)$ of signals, which is given as follows:

$$p(u,v) = \frac{1}{2\pi\tau_u\tau_v\sqrt{1-\rho_{uv}}} e - \frac{1}{2(1-\rho_{uv}^2)} \left[\left(\frac{u-\mu_u}{\tau_u^2}\right)^2 - 2\rho_{uv}\frac{(u-\mu_u)(v-\mu_v)}{\tau_u\tau_v} + \left(\frac{v-\mu_v}{\tau_v^2}\right)^2\right] \quad (1)$$

where μ_u, μ_v, τ_u, and τ_v are the means and the standard deviations of the signals u and v, respectively. ρ_{uv} is the correlation coefficient between the two signals. The coefficient is given by:

$$\rho_{uv} = \frac{E\{(u-\mu_x)(v-\mu_y)\}}{\tau_u\tau_v} \quad (2)$$

The correlation coefficient can also sometimes be used to determine if two signals are statistically independent. On one hand, if $|\rho_{uv}| = 1$, there is a strong correlation between the two signals. On the other hand, if $|\rho_{uv}| = 0$, the two signals are not correlated. The correlation can be interpreted as a weak form of statistical dependency. In [12], it is shown that two random variables, which are not correlated, can be statistically dependent. This is why we take the statistical dependency or independency. The product of the marginal densities ρ_u and ρ_v of the signals u and v, respectively, is given by:

$$p(u,v) = p(u)p(v) \quad (3)$$

If the expression in (1) is equal to the product of the marginal densities in (3), the signals are completely independent. One possibility to quantify the statistical dependency between two signals is to calculate the MII of them, as follows:

$$\omega(u,v,C) = \int\int p(u,v)\log\frac{p(u,v)}{p(u)p(v)}du\,dv \quad (4)$$

The base of the logarithm determines the units in which information is measured. (4) shows that if u and v are independent, ω becomes zero. A forward approach is to divide the range of u and v into finite bins and count the number of sampled pairs of $h_o = (u_o, v_o), o = 1, 2, \cdots, n$, falling into these finite bins. This count allow us to approximately determine the probabilities, replacing (15) by the finite sum:

$$\omega_{bin}(u,v,C) = \sum_{a,b} p_{uv}(a,b)\log\frac{p_{u,v}(a,b)}{p_u(a)p_v(b)} \quad (5)$$

where $p_u(a) \approx n_u(a)/n$ and $p_u(b) \approx n_u(b)/n$ are the probabilities based on the number of points $n_u(a)$ and $n_v(b)$ falling into the ath bin of u and the bth bin of v, respectively. The joint probability is $p_{uv}(a,b) \approx n(a,b)/n$ based on the number $n(a,b)$ of points falling into box nos. a, b. MII is non-negative and symmetric:

$$\omega(u,v,C) = \omega(v,u,C) \geq 0 \quad (6)$$

The MII for all possible combinations of sensor outputs y_r and y_s (except $r = s$, $i = 1, 2, \cdots, r, j = 1, 2, \cdots, s$) is computed, which leads to an ω-matrix for all combinations of r and s. The basic idea is that the MII changes when a signal fault f_r is present. Suppose that it is in the rth channel or index:

$$\tilde{y}_r = y_r + f_r \quad (7)$$

This fault appears only in the rth channel. Thus, we should expect that all combinations with index r should show a reduction of ω. This allows us to localize the faulty signals. One or more sensors' faulty signals can be simultaneously detected in the same way. One possibility to visualize the faulty signals is to use the relative change as a signal fault indicator $\lambda_{y_r}^{\omega}$:

$$\lambda_{y_r}^{\omega} = \frac{|\omega_{y_r} - \omega_{ref}|}{\omega_{y_r}} \quad (8)$$

where y_r is an actual data set and the lower index *ref* is one reference data set. The method based on MII is able to detect sensor faults in different combinations of them.

The faulty signal detection can execute in a distributed manner when each sensor makes a decision on the collected signals locally. In this algorithm, if the local decision on a sensor's signals, $\lambda_{y_r}^{\omega} > 0.5$, the signals are faulty. This means that MII is high on the sensor's faulty signals. The algorithm based on MII is able to detect different kinds of faults (as discussed in Section II). Using the algorithm, a sensor is able to know whether its collected data is trustworthy or not and forward the data towards the CH.

IV. Protected Data Collection for Aggregation

Once a sensor has trustworthy data which has been identified earlier, it may be altered at the sensor or intermediate sensor before/after transmission, i.e., a CH may receive unprotected (or altered) data for aggregation. We need to ensure the data protection. To discover a unreliable sensor or unprotected data at the CH, we use the truth discovery approach in TPDC. Traditionally, truth discovery is used in many domains in order to resolve conflicts with multiple noisy data sensors. The insight is a truth discovery algorithm that begins with a random guess of ground truths, and iteratively conducts status value updates and truth updates until convergence [9].

In our approach, we compute sensor status value to check whether or not the data is altered at the transmission. The basic idea is that a sensor's status value can be given a high value if the sensor transmitted trustworthy data is close to the estimated ground truths. Typically, the sensor status values are computed as follows:

$$S_k = \log(\frac{\sum_{k'=1}^{K}\sum_{m=1}^{M} d(x_m^{k'}, x_m^*)}{\sum_{m=1}^{M} d(x_m^k, x_m^*)}) \quad (9)$$

$d(.)$ is the distance function which measures the difference between sensors observation values $x_m^{k'}$ and the estimated ground truths x_m^* [18]. $d(.)$ relies on particular sensing application scenarios. The proposed framework TPDC is intended to deal with SHM applications,such as structural damage event detection, for example. For SHM applications, where the sensory data is continuous, like with acceleration or strain, we adopt the following normalized squared distance function:

$$d(x_m^k, x_m^*) = \frac{(x_m^k - x_m^*)^2}{std_m} \quad (10)$$

where std_m is the standard deviation of all observation values.

Truth Update. Suppose that the status value of each sensor is fixed. Then, we can estimate the ground truth for the m-th events (m is used as there can be multiple events or an event includes multiple type of signals' information, such as with accelerometer, stain, displacement, etc.

$$x_m^* \leftarrow \frac{\sum_{k=1}^{K} S_k x_m^k}{\sum_{k=1}^{K} S_k} \quad (11)$$

Since acceleration and strain data are continuous, x_m^* denotes the estimated ground truth value. The truth discovery process begins with randomly guessing the ground truth for an event, then iteratively updates sensors status values and estimated ground truths until some convergence criterion is satisfied. Normally, the convergence criterion is set regarding the requirements of specific applications. It should be a threshold of the change in the estimated ground truths in two consecutive iterations.

Truth Status Value Discovery. Here, we discuss the details of our secured truth discovery approach. We assume a semantically secure (p, t)-threshold Paillier cryptography, adopted from [19]. Here p is the number of sensors including both clusters and sensors, and t is the minimum number of (clusters and sensors) needed to complete the decryption. Thus, each sensor TPDC has known the public encryption key $p_k = (g, n)$, while the private decryption key has been divided and distributed to all nodes in D (i.e., node i gets its private key share sk_i). At the data transmission, sensors iteratively conduct the following two procedures:

- *Status Value Update.* Each sensor computes the distances between its trustworthy data (observation values) and the estimated ground truths given by the cluster regarding the distance functions, then encrypts the distance information and yields the ciphertexts to the CH. Once the ciphertexts from all sensors are received, the CH securely updates the status value in encrypted form for each sensor. Then, the ciphertext of the updated status value is sent to each corresponding sensor.
- *Secure Truth Estimation.* Based on the encrypted status value received from the CH, each sensor computes the ciphertexts of status value observation without decrypting the status value, and then sends them to the CH. Once the CH gets all the ciphertexts of status value from sensors, it can compute the ground truth for the trustworthy data.

The above two procedures begin with a random initialization of the ground truth for the trustworthy data, and are then iteratively conducted until convergence. Throughout these procedure, the processes are conducted on encrypted data. Therefore, it is ensured that the status value observations of each sensor are known only to itself and the sensor status values are not disclosed to any other sensors.

For status value update and secure truth estimation, we use a secure sum protocol designed to calculate the summation of the data collected from sensor without disclosing them to any unintended party of the system. According to Eqn. (9) and Eqn. (11), a CH calculates the summation of the data collected from sensors in order to update status values and estimate ground truths. However, the plaintext of each sensors data should not be accessible to other due to privacy concerns. We deal with this problem with a the secure sum protocol which can be achieved through threshold Paillier cryptosystem [20].

V. Performance Evaluation

A. Simulation Methods

We conduct simulations using MATLAB to evaluate TPDC that includes the trustworthy and protected data collection methods. We use real data sets collected by the SHM system employed on the high-rise Guangzhou National TV Tower (GNTVT) [5], [21] and a SHM toolsuite [22]. The dataset includes data collected from 800 sensors. We use the data sets for the 100-sensor case in our simulations. We perform the WNSS deployment via our WNSS-based deployment scheme suggested in [23] The simulation environment is a 450×50 sensing field regarding structural environment, e.g., bridge, building, aircraft.

The background data is simulated as vibration influenced by the 100 sensor locations in the field. A random Gaussian noise is added to all the data. The mean of the noises is zero, and the standard deviation is 10% of the real signals. From the data sets, a set of data is used as *reference data* to train the joint distribution, and another set of similar data is used for testing. The noise is present in both the data sets. Thus, the trained correlation model reflects the noises. After a sensor receives a decision, it recomputes its MII.

As a baseline approach for status value truth discovery, we use the state-of-the-art truth discovery scheme in our simulations, i.e., CRH (conflict resolution on heterogeneous data) [18], which does not take any actions

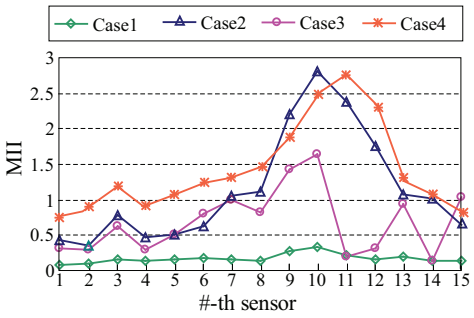

Fig. 2. Performance of TPDC in trustworthy data collection: achieved MII under sensor signal faults.

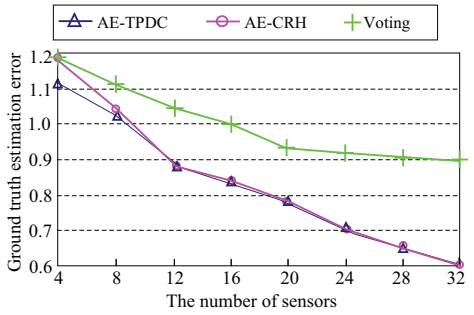

Fig. 3. Performance of data alternation and detection observing ground truth estimation.

to break sensor security during the whole procedure. A $(p, \lfloor \frac{p}{2} \rfloor)$-threshold Paillier cryptosystem is used in simulations (http://cs.utdallas.edu/dspl/cgi-bin/paillierttoolbox/). The status value truth discovery is implemented by following the Paillier Threshold Encryption Toolbox.

We also consider Voting framework for comparison. Voting is used to eliminate conflicts for decision-making based on collected data, which is used to conduct majority voting so that information with the highest number of occurrences, mean, or median is regarded as the correct answer. In Voting, it is assumed that all the sensors are equally reliable, and thus the votes from different sensors are uniformly weighted. We adopted a state-of-the-art network based voting algorithm from [24].

B. Performance measures

We consider the following measures for evaluating the estimation accuracy of sensor status value truth in TPDC: mean of absolute error (AE) and error rate. The data used in TPDC is the continuous data and we use AE to measure the mean of absolute distance between the estimated results and ground truths. Error rate is calculated as the percentage of the approachs output that are different from the trustworthy data and ground truths. The simulation is repeated for 50 times for confidence.

C. Results

In the first set of simulations, we implemented trustworthy data collection under the sensor signal fault injection, which was achieved through modifying a number of sensors' signals randomly in the data sets. A fraction of the sensors is randomly selected and the modified faulty signals are fed into their acquisition modules. We vary the number of faulty signals from 15% to 25%. Each sensor broadcasts its readings towards the neighboring sensors. Each of the faulty readings is replaced by a random number independently drawn from a uniform distribution in the deployment field (0, 450). Such a fault model is selected since it yields uncorrelated data in the same magnitude as the collected signals in practice.

In Fig. 2, MII achieved in the first four successful simulation cases, with varying sensor fault injection. Case 1 has no any signal fault injection. This means that the acquired data is not altered almost in all the sensors by any signal faults or security attack. Case 2 shows the high MII value at some sensors including sensor 9 and sensor 10. Their signals are faulty or partly altered, which is apparently detected. When the rate of signal fault injection increases, we can see that the MII values at sensors in the neighborhood becomes highest. Data from these sensors cannot be trustworthy. This justifies the correctness of the untrustworthy data detection. Whenever a CH receive such data, it may drop the data from aggregation or a data reconstruction method may be used for the portion of untrustworthy data.

We compare the accuracy of ground truths between TPDC and CRH. The estimation errors of TPDC are introduced by randomly guessing the ground truths, since we randomly initialize the estimated ground truths, and use a threshold of the change in estimated ground truths in two consecutive iterations as the convergence criterion. Fig. 3 shows the ground truth estimation errors of TPDC, CRH, ad Voting under different random values. The error is measured in terms of AE. The figure shows that TPDC almost has the same estimation errors as CRH while the number of sensors is varying. In some cases, even TPDC is better performing than the CRH; particularly, when the number of sensors is small. Also, we can see that, the estimation errors decrease with the increase of the number of sensors. When compared to both CRH and TPDC, the Voting framework shows the worst performance. One possible reason is that reliability estimation based on the maximum number of packets or votes cannot reflect the true facts in the networked systems. Though the commonly used approach to eliminate conflicts for a decision on an event or faulty sensor is to conduct majority voting so that information with the highest number of occurrences or median is regarded as the correct answer. The issue of such Voting/Averaging frameworks is that they assume all the data packets from sensors or the sensors are equally reliable, and thus the votes from different sources are uniformly weighted. This does not include information when the packets or voted are altered.

In the final set of simulations in this paper, we observe the error rate in different frameworks. In Fig. 4, we illustrate

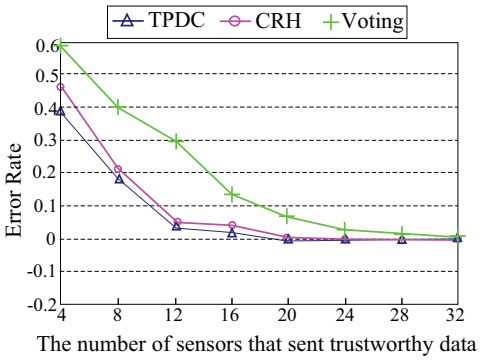

Fig. 4. Performance on the error rate under both trustworthy data and altered data situations in all the frameworks.

the performance of them in terms of error rate on the SHM data set. We can see that the proposed TPDC obtains better performance on the SHM data set compared with the CRH and Voting. From the figure, we can observe that the CRH framework consider all the data collected where some data are compromised, i.e., faulty signals were injected. Trustworthy data is not identified in both the CRH and Voting. In Voting, sensor data are considered equally reliable and/or the sensor is reliable which has maximum votes. The sensors in TPDC include the trustworthy data. When a portion of the trustworthy data is further altered at the transmission, this data is dropped before aggregation. As a result, the error rate in TPDC is lower than that of the CRH.

VI. Conclusion

In this paper, we have presented a trustworthy and protected data collection (TPDC) framework for event detection in WNSS. It facilitates reliable collection for aggregation at a cluster of WNSS. For the trustworthy data, we propose an algorithm to make sure that transmitted data is unaltered. For the protected data, we present a truth discovery approach, whose goal is to infer truthful facts from unreliable sensor data. Through extensive simulations, we demonstrate that the collected data in TPDC is trustworthy and protected that may provide a reliable decision-making in event detection. Our future work includes performance evaluation of event detection using TPDC framework, detailed performance analysis of the proposed approaches, and comparative studies. Furthermore, the data transmission security and authenticity with the sensor status value discovery approach is not described in this paper, which will be focused in the future.

Acknowledgment

This research was supported in part by NSF grants CNS 1449860, CNS 1461932, CNS 1460971, CNS 1439672, CNS 1301774, and ECCS 1231461.

References

[1] M. Z. A. Bhuiyan, G. Wang, J. Cao, and J. Wu, "Sensor placement with multiple objectives for structural health monitoring," *ACM Transactions on Sensor Networks*, vol. 10, no. 4, pp. 1–45, 2014.

[2] Y. Cheng, X. Li, Z. Li, S. Jiang, Y. Li, J. Jia, and X. Jiang, "Aircloud: a cloud-based air-quality monitoring system for everyone," in *Proc. of ACM Sensys14*, 2014, pp. 1–14.

[3] G. Xing, M. Li, T. Wang, W. Jia, , and J. Huang, "Efficient rendezvous algorithms for mobility-enabled wireless sensor networks," *IEEE Transactions on Mobile Computing*, vol. 11, no. 1, pp. 47–60, 2012.

[4] M. Z. A. Bhuiyan, J. Wu, G. Wang, , and J. Cao, "Sensing and decision-making in cyber-physical systems: The case of structural health monitoring," *IEEE Transactions on Industrial Informatics*, pp. 1–11, 2016, http://dx.doi.org/10.1109/TII.2016.2518642.

[5] B. Li, D. Wang, F. Wang, and Y. Q. Ni, "High quality sensor placement for SHM systems: Refocusing on application demands," in *Proc. of IEEE INFOCOM*, 2010, pp. 1–9.

[6] M. Z. A. Bhuiyan and J. Wu, "Collusion attack detection in networked systems," in *Proc. of the 14th IEEE International Conference on Dependable, Autonomic and Secure Computing (IEEE DASC 2016)*, 2016, pp. 1–8.

[7] M. Rezvani, A. Ignjatovic, E. Bertino, and S. Jha, "Secure data aggregation technique for wireless sensor networks in the presence of collusion attacks," *IEEE Transactions on Dependable and Secure Computing*, vol. 12, no. 1, pp. 98–110, 2015.

[8] D. Jiao, M. Li, Y. Yu, and J. Ou, "Self-healing key-distribution scheme with collusion attack resistance based on one-way key chains and secret sharing in wireless sensor networks," *International Journal of Distributed Sensor Networks*, vol. 2015, pp. 1–7, 2015.

[9] Y. Li, J. Gao, C. Meng, Q. Li, L. Su, B. Zhao, W. Fan, , and J. Han, "A Survey on Truth Discovery," *ACM SIGKDD Explorations Newsletter*, vol. 17, no. 2, pp. 1–16, 2015.

[10] P. Schaffer and I. Vajda, "CORA: Correlation-based resilient aggregation in sensor networks," *Ad Hoc Network*, vol. 7, no. 6, pp. 1035–1050, 2009.

[11] M. Rezvani, A. Ignjatovic, E. Bertino, and S. Jha, "Secure data aggregation technique for wireless sensor networks in the presence of collusion attacks," *IEEE Transactions on Dependable and Secure Computing*, vol. 12, no. 1, pp. 98–110, 2015.

[12] T. Clouqueur, K. K. Saluja, and P. Ramanathan, "Fault tolerance in collaborative sensor networks for target detection," *ACM Transactions on Computers*, vol. 53, no. 3, pp. 320–333, 2004.

[13] X. Liu, J. Cao, S. Lai, C. Yang, H. Wu, and Y. Xu, "Energy efficient clustering for WSN-based structural health monitoring," in *Proc. of IEEE INFOCOM*, 2011, pp. 2768–2776.

[14] M. Z. A. Bhuiyan, G. Wang, J. Cao, , and J. Wu, "Deploying wireless sensor networks with fault-tolerance for structural health monitoring," *IEEE Transaction on Computers*, vol. 64, no. 2, pp. 382–395, 2015.

[15] D. N. Mitiku and M. E. Getachew, "Data processing algorithms in wireless sensor networks for structural health monitoring," Master Thesis, Royal Institute of Technology, January, 2012.

[16] M. Z. A. Bhuiyan, G. Wang, J. Wu, J. Cao, X. Liu, and T. Wang, "Dependable structural health monitoring using wireless sensor networks," *IEEE Transactions on Dependable and Secure Computing*, 2016, DOI: http://dx.doi.org/10.1109/TDSC.2015.2469655.

[17] P. Zhuang, D. Wang, and Y. Shang, "Distributed faulty sensor detection," in *Proc. of IEEE GLOBECOM* , 2009, pp. 1–6.

[18] Q. Li, Y. Li, J. Gao, B. Zhao, W. Fan, and J. Han, "Resolving conicts in heterogeneous data by truth discovery and source reliability estimation," in *Proc. of ACM SIGMOD*, 2014.

[19] I. Damgard and M. Jurik, "Generalisation, a simplification and some applications of paillier's probabilistic public-key system," in *Proc. of PKC*, 2001, pp. 119–136.

[20] R. Cramer, I. Damgard, and J. B. Nielsen, "Multiparty computation from threshold homomorphic encryption," in *Proc. of EUROCRYPT*, 2001, pp. 280–300.

[21] Y. Ni, Y. Xia, W. Liao, and J. Ko, "Technology innovation in developing the structural health monitoring system for Guangzhou New TV Tower," *Struct. Cont. and Health Monit.*, vol. 16, no. 1, pp. 73–98, 2009.

[22] ISHMP Toolsuite. [Online]. Available: http://shm.cs.uiuc.edu/

[23] S. Wei, Y. Meng, and C. Jean-Pierre, "Resilient secure localization and detection of colluding attackers in WSNs," in *Proc. of Ad-hoc, Mobile, and Wireless Networks*, 2012, pp. 181–192.

[24] H.-T. Pai, , and Y. S. Han, "Power-Efficient Direct-Voting Assurance for Data Fusion in Wireless Sensor Networks," *IEEE Transactions on Computers*, vol. 57, no. 2, pp. 261–273, 2008.