Estimation based Efficient and Resilient Hierarchical In-Network Data Aggregation Scheme for Wireless Sensor Network

N. Chitradevi
Assistant Professor, IT Department
Kumaraguru College of Technology
Coimbatore - 641 006

V.Palanisamy
Principal
INFO Institute of Engineering
Coimbatore - 641 007

K. Baskaran
Assistant Professor, CSE Department
Government College of Technology
Coimbatore - 641 013

K. Swathithya,
Kumaraguru College of
Technology, Coimbatore,
India

ABSTRACT

In a large sensor network, in-network data aggregation is inherently used as a communication paradigm which reduces the number of packets transmitted and hence the energy consumed. However the unattended and hostile operation of sensor network makes the system vulnerable to node compromise attack. The compromised nodes can inject false data in to the network which deteriorates the accuracy of the aggregate data. So the research on resilient data aggregation with a focus on data integrity and accuracy becomes a major issue. In this paper, we propose a statistical based robust estimate to design a resilient in-network aggregation scheme which detects and isolates the outliers from computed aggregate value. Simulation results demonstrate that our approach provides a powerful mechanism for detecting outliers even in the presence of multiple compromised nodes.

Categories and Subject Descriptors

C.2.0 [Computer Communication Networks]: General - Security and protection; C2.1 [Computer - Communication Networks] Network Architecture & Design - Wireless Communication; D4.6 [Operating Systems]; Security and Protection; K.6.5 [Management of Computing and Information systems]: Communication Networks-Security and Protection.

General Terms

Algorithms, Design, Security

Keywords

Sensor Networks, Resilient aggregation, Outlier detection, data integrity.

1. INTRODUCTION

A wireless sensor network (WSN) consist of spatially distributed autonomous sensor devices that cooperatively monitor physical or environmental conditions, such as temperature, sound, vibration, pressure, motion, pollutants etc, at different locations. The monitored parameters which may vary with time and space are collaboratively sent to a collection center for further processing through a multihop network architecture due to their energy-efficiency and scalability features [17][6]. In reality, as the data measured by the sensors are highly correlated either spatially or temporally [26][23] data aggregation seems to be the commonly used communication technique which exploits data redundancy to dramatically decrease the amount of information to be transmitted [13][16][4][14] thereby reducing communication traffic. Nowadays, sensor networks are increasingly being used in many real time monitoring applications ranging from military(e.g., surveillance, intelligent gathering) to environmental monitoring (e.g., habitat monitoring, wild-life monitoring).

Due to the deployment nature of WSN, sensor nodes are highly vulnerable to many physical attacks like DoS, Sybil etc. However, in this paper we concentrate on false data insertion attack. In a false data insertion attack, an adversary can compromise single or a set of node(s) and makes it to send arbitrary false data to its parent. This attack becomes more damaging when multiple compromised nodes collude in injecting false data. When a few sensor nodes are compromised by an attacker then a part of the sample received by aggregator may be corrupted causing the aggregator to perform an inaccurate aggregation which makes the final aggregation result to far deviate from true measurement [19][9]. The main objective of attacker in false data attack is to make the user accept false aggregation results, which are significantly different from the true results determined by the measured values. Thus designing attack resistant data aggregation has been identified as a major challenging research topic for sensor network.

Most of the applications designed for WSN have high demands for efficient performance of the network. The integrity of data has tremendous effect on the performance for any such mission critical data gathering system. Recently many data aggregation protocols have been proposed [7][8] [13] [16] [4] [14] with a focus on improving energy efficiency by eliminating the data redundancy in sensor data of the network. However, there are only a few studies addressing data integrity and accuracy issues in sensor network and a lot of related issues still remain opens.

In this paper, we take a step towards providing resilient data aggregation mechanism for WSN. The goal of this study is to design an energy efficient in-network algorithm for determining faulty readings even in the presence of multiple outlier nodes. The rest of the paper is organized as follows: in section 2 we describe some of the related work in the literatures. In section 3 we provide attack model and basic assumptions. In section 4 we introduce our system model and problem formulation. In section 5 we explain how the aggregator detects the outlier in distributed fashion. We discuss the simulation results and its interpretation in section 6. In section 7 we conclude our work.

2. RELATED WORK

The research area in sensor networks is relatively broad and interdisciplinary predominantly dealing with computation and communication. Most of the challenges and bottlenecks in sensor network research deal with energy efficient design and development of software and /or hardware components [10][25][10]. The focus of this research paper is to present novel ideas that have practical implementations in developing outlier aware aggregation protocols for sensor network.

Outlier detection has been extensively studied in the database research community [12][15] and there are several different definitions of outliers in the literature. Hu et al. [27] consider large sensor networks where the sensor nodes organize themselves into a tree rooted at the base station. The packets are transmitted through the tree constructed. The authors adopt a delayed aggregation and delayed authentication approach to handle single node compromise attack.

Przydatek et al. present cryptography based countermeasures against the attacker who wants to distort the aggregate in [19]. The author uses a cryptographic commitment and interactive proof technique for enabling secure aggregation where the aggregator sends the aggregate statistics to the home server together with a Merkle hash tree based commitment. During the verification phase, the home server sends a query to the aggregators for a randomly selected subsample and verifies the received subsample against the commitment by interacting with the aggregator. If this check is successful, i.e., the home server concludes that the subsample really comes from the sample used for the calculation of the commitment and sent by the corresponding sensor.

In [18][3][22] the authors utilize the naturally existing correlation between the measurement of the sensors. The authors [11][24] propose an alternate reputation based approach which relies on reports from other nodes to detect anomalies. This method also reduces the chances of accepting and forwarding false data. The main drawback with reputation based system is the high rated nodes if compromised can still inject malicious data silently. By contrast, our work directly deals with data to look for outliers.

3. ATTACK MODEL AND ASSUMPTIONS

An attacker who is able to alter the parameters of the sensor nodes within its proximity, represents a serious threat which cannot be circum vented by crypto graphical methods, since the nodes that measure the distributed phenomena generate cryptographically sound messages. Handling the problem of such an attack is a must in order to realize security in sensor networks.

3.1. ATTACK MODEL

We consider a setting where an attacker is able to produce some kind of "noise" that is added to the measurements of the sensors. In particular, we assume the Byzantine fault model [15] where a compromised node is under the full control of the attacker. Thus the noise introduced is completely under the control of the adversary, but still considered to be independent and identically distributed. Due to the inherent noisy nature of the physical measurement, we assume that an attacker in order to have significant gain aims to inject false values that deviate from a true measurement in a noticeable scale. We note that we do not restrict the number of samples an adversary can compromise but we assume that the adversary does not have any knowledge regarding the distribution and size of sample. Finally, we do not consider any particular distribution for the attacker's noise.

4. SYSTEM MODEL AND PROBLEM STATEMENT

This section describes our system model, design goals and problem statement.

4.1. SYSTEM MODEL

We consider a static WSN which automatically gets organized into non-overlapped tree-based aggregation groups where each node joins in only one group as shown in Figure 1. Data collected within the network are reported to a powerful base station located nearby the sensor network through some special nodes called aggregators. The aggregators have more memory space, computational power and battery life compared to the sensor nodes. The issue regarding aggregator selection is out of the scope of this paper and we simply assume that there exist some nodes acting as aggregators at a given time. However, the ratio of number of sensor nodes to an aggregator is limited and user specified.

We assume that each sensor has a unique identifier and shares a separate secret cryptographic key with the base station and with the aggregator. We assume that there is a reliable transmission mechanism. We also assume that every sensor node has an individual secret key shared with base station and the aggregator as in [1], which enables message authentication and data confidentiality. Moreover, we also assume that the base station and aggregators employ a secure mechanism [1][2] to enable an authenticated broadcast to all the nodes in the network.

4.2 Design Goals

Our design goal is to defend against fault data injection attack making the base station accept false aggregation result. Specifically, our design goal includes:

- Low communication overhead
- Efficient detection
- Low false alarm

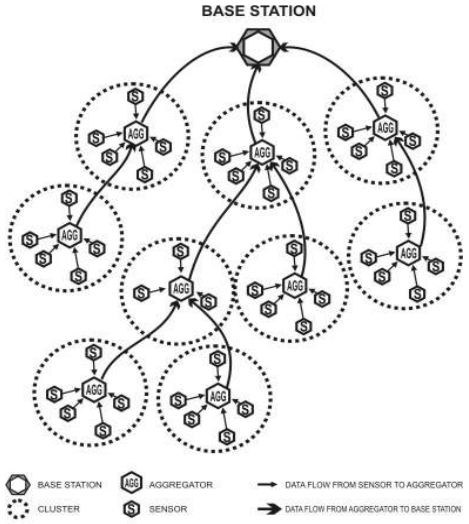

Figure 1. A clustered sensor network topology

4.3 Problem statement

Sensor nodes are vulnerable to physical capture (node compromise) as they are usually deployed in hostile areas. Once a sensor node is compromised, the adversary gains the control of the node and can arbitrarily make changes in the sensed data. The aggregation performed on such false data makes the ultimate aggregate result inaccurate. Since in many applications the information received by the base station provides a basis for critical decisions, false information may cause catastrophic consequences. This raises the necessity to have an outlier aware aggregation in order to keep the aggregate data received at base station intact with the aggregate computed based on the original measurement, even in spite of occurrence of multiple compromised nodes.

5. Estimation based Resilient Aggregation

This section presents our resilient data aggregation protocol named EBRA. This protocol makes use of a robust estimation of mean and standard deviation to localize all extreme values and preclude them from the aggregation process.

In EBRA, after the network deployment, we assume that all the sensor nodes establishes pair wise keys with its aggregator and base station using [1][2] to enable data confidentiality and secure transmission. The protocol is designed for a continuous data gathering application where each sensor measures the feature F (like temperature, humidity, pressure etc.) of application interest and sends them periodically to its aggregator for further processing. The protocol relies on outlier definition in [21] to efficiently label the anomalous data. The observation that is sufficiently far from most other observations in the data set is concluded as outlier data. The aggregator uses the robust estimates determined using EBRA protocol to filter the non-consistent data from the sample and compute the arithmetic mean of the remaining values. The protocol makes use of winsorisation process to converge the estimation to an acceptable degree of accuracy and uses the z-score method to identify abnormal data. Table 1 shows the definition of z-score test adopted

Table 1. z-score definition

H0: There are no outlier in the data set
Ha: There is at least one outlier in the data set
Test Statistic $Z = \frac{(x - \hat{\mu}_h)}{\hat{\sigma}_h}$

6. Performance Evaluation

In this section, through simulations, we show that the protocol EBRA increases the resilient data aggregation ability in the presence of comprised nodes.

6.1 Simulation Environment

We consider a simulation setup, where N sensors are deployed over a region of 50 * 50m to monitor a specified parameter which may be corrupted by additive gaussian white noise. We assume the sensors communicate in multi-hop fashion. The sample was generated by the randomGr function. We have kept Pc as the probability for determining the number of compromised nodes at a time and Pr as the corruption rate that defines the rate at which an adversary makes the alteration of value. The noise was simulated by a function which alters the measurements in sample according to the corruption rate Pr with a probability of Pc. To obtain the maximum distortion reachable we have made 100 simulation runs for different values of Pr for a given probability Pc. We evaluated the performance using two metrics namely detection rate and misdetection rate. The detection rate is defined as the rates between the number of correctly detected attacks to the total number of attacks. While misdetection rate is the ratio between the number of normal condition that are incorrectly misclassified as attacks and the total number of normal conditions.

6.2 Simulation Result

The performance of EBRA protocol is compared with the Grubb's approach in [28] for different scenarios involving single and multiple compromised nodes with different degree of alterations. Figure 2 and Figure 3 compares the detection rate and misdetection rate of the EBRA with Grubb's when there is one malicious node launching the false data attack. In Figure 2 and Figure 3, the x-axis represents the faulty sensor rate which is the ratio of the number of faulty or compromised sensors and the total number of sensors deployed.

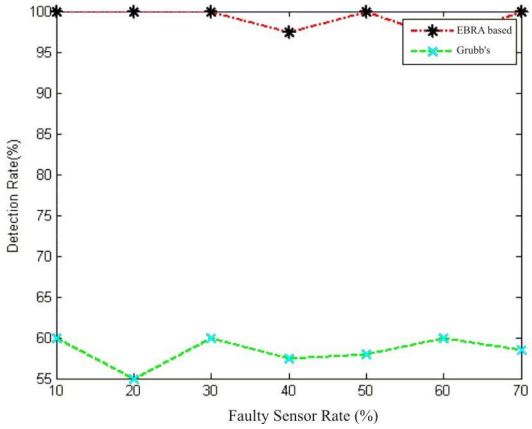

Figure 2. Detection rate for single attack

Each compromised sensor will report noise or faulty readings according to the parameter Pr. Thus all the compromised nodes will report faulty data with a common corruption ratio.. In addition to above, the scenario of noise insertion in variable corruption rate is also considered. Figure 2 and Figure 3 shows the performance for the scenario where the compromised nodes insert malicious data with different noise ratio. From Figure 2, we can see that EBRA based approach is effective in detecting the attacks as the detection probability rapidly increases to a high value (³ 95%). While Grubb's test offer only an average of 60% detection rate. Moreover, the misdetection rate in EBRA remains zero until the faulty sensor rate reaches 50% but in case of Grubb's test, the misdetection rate increases with faulty rate.

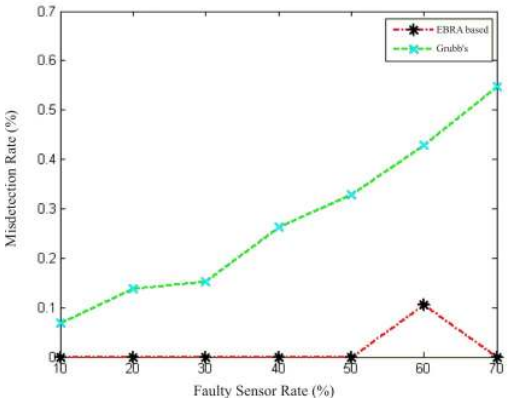

Figure 3. Misdetection rate for single attack

In addition to performing single node compromised attack, we also checked for multiple node compromise. We randomly choose multiple nodes and simulated the corruption according to Pc and Pr. Figure 4 shows the detection rate EBRA and Grubb's test for multiple attacks. As can be seen from Figure 4, the detection rate of EBRA remains 100% until faulty rate becomes 60% but later on decreases to zero percent as the outlier dominate normal.

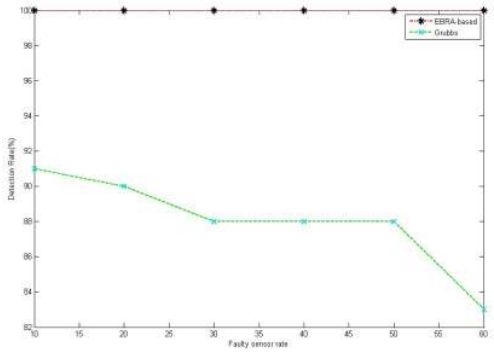

Figure 4. Detection rate for multiple attack

Figure 5 shows the detection rate for multiple attacks where the corruption rate Pr is 20%. From the Figure 5, it can be found that Grubb's test detection rate rapidly increases to a high value (³80%) after the data is altered by a certain degree but as the corruption rate increases, the detection rate also increases until it finally reaches 100% while EBRA detection rate is found to remain 100% in all cases.

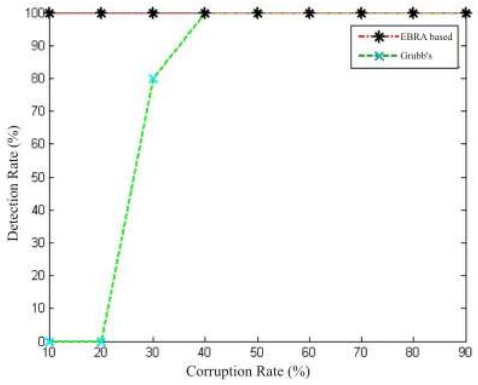

Figure 5. Detection rate for multiple attack

7. CONCLUSION

In this paper, we presented a serious threat against the sensor network where the adversaries modify the measured parameters of the environment with in the proximity of sensor nodes. We propose EBRA protocol to design "resilient aggregators" that uses a robust estimator to detect fault data. Our protocol suits for any data distribution. Simulations are conducted to characterize the effectiveness of our algorithm in terms of misdetection and detection rate probability.

REFERENCES

[1] Adrian Perrig, Robert Szewczyk, Victor Wen, David Culler and J.D. Tygar. SPINS: Security Protocols for Sensor Networks. In Wireless Networks Jounal (WINET), 8(5):521-279,1981.

[2] Adrian Perrig, Ran Cnetti, J.D. Tygar, and Dawn Song. The TESLA Broadcast Authentication Protocol. RSA CryptoBytes, 2002.

[3] Antonios Deligiannakis et al. Outlier-Aware Data Aggregation in Sensor Networks, In ICDE, 2008.

[4] A. Boulis, S. Ganeriwal, and M.B. Srivastava. Aggregation in Sensor Networks: An Energy-Accuracy Trade-Off. In Proc. 2003 IEEE Int'l Workshop Sensor Network Protocols and Applications, pp. 128-138,2003.

[5] V. Barnett and T. Lewis Outliers in Statistical Data. John Wiley & Sons, 1994.

[6]. J. Chou, D. Petrovic, and K. Ramchandran. A Distributed and Adaptive Signal Processing Approach to Reducing Energy Consumption in Sensor Networks. In Proc. IEEE INFOCOM 2003, vol.2, pp.1054-1062, 2003.

[7] C. Castelluccia, E. Mykletun, and G. Tsudik. Efficient Aggregation of Encrypted Data in Wireless Sensor Networks. In Mobile and Ubiquitous Systems: Networking and Services (MobiQuitous'05),2005.

[8] J.Y. Chen, G. Pandurangan, and D. Xu. Robust Computation of Aggregates in Wireless Sensor Networks: Distributed Randomized Algorithms and Analysis. In Proceedings of the International Symposium on Information Processing in Sensor Networks (IPSN'05). 348-355,2005.

[9] David Wagner. Resilient Aggregation in Sensor Networks. In Workshop on Security of Ad Hoc and Sensor Networks, 2004.

[10] A. Deligiannakis, Y. Kotidis, and N. Roussopoulos. Bandwidth Constrained Queries in Sensor Networks. In VLDB Journal, 2007.

[11] S. Ganeriwal, M.B. Srivastava. Reputation-based Framework for High Integrity Sensor Networks. In ACM Security for Ad-hoc and Sensor Networks (SASN 2004).

[12] Hodge V, and Austin J. A Survey of Outlier Detection Methodologies. In Artificial Intelligence Review, 22:85-126. 2004.

[13] C. Intanagonwiwat, R. Govindan, D. Estrin, J. Heidemann and F. Silva. Directed Diffusion for Wireless Sensor Networking. In IEEE/ACM Trans. Networking, vol.11, no.1, pp 2-16, 2003.

[14] B. Krishnamachari, D. Estrin, and S. Wicker. The impact of data aggregation in Wireless Networks. In Internationl Workshop on Distributed Event-Based Systems, (DEBS '02), Vienna, Austria, 2002.

[15] L. Lamport, R. Shostak, and M. Pease. The Byzantine Generals Problem. In ACM Transactions on Programming Languages and Systems, Vol. 4, No.3, 1982.

[16] S. Olariu, A. Wada, L. Wilson, and M. Eltoweissy. Wireless Sensor Networks: Leveraging the Virtual Infrastucture. In IEEE Network, vol. 18, no. 4, pp.51-56, . 2004.

[17] S. Pattem, B. Krishnamachar, and R. Govindan. The Impact of Spatial Correlation on Routing with compression in Wireless Sensor Networks. In Proc. Third International Symposium. Information Processing in Sensor Networks (IPSN), pp.28-35, 2004.

[18] Péter Schaffer et al. Correlation-based Resilient Aggregation in Sensor Networks. In MSWiM'07, October 22-26, 2007.

[19] B. Przydatek, D. Song, and A. Perrig. SIA: Secure Information Aggregation in Sensor Networks. In SenSys '03: Proceedings of the 1st International Conference on Embedded Networked Sensor System, pp.225-265,2003.

[20] A. Sharaf, J. Beaver, A. Labrinidis, and P. Chrysanthis. Balancing Energy Efficiency and Quality of Aggregate Data in Sensor Networks. In VLDB Jounal, 2004.

[21] S. Subramaniam, T. Palpanas, D. Papadopoulos, V. Kalogeraki, and D. Gunopulos. Online Outlier Detection in Sensor Data Using Non-Parametric Models. In VLDB, pages 187-198, 2006.

[22] Sapon Tanachaiwiwat and Ahmed Helmy. Correlation Analysis for Alleviating Effects of Inserted Data in Wireless Sensor Networks. In IEEE International Conference on MobiQuitous, 2005.

[23] C.C. Shen, C. Srisathapornphat, and C. Jaikaeo. Sensor Information Networking Architecture and Applications. In IEEE Personal Comm., col. 8, no. 4, pp.52-59, 2001.

[24] Yannis Kotidis et al. Robust Management of Outliers in Sensor Network Aggregate Queries. In MobiDE, 2007.

[25] S. Yoon and C. Shahabi. Exploiting Spatial Correlation Towards Energy Efficient Clustered Aggregation in Sensor Networks. In Proc. of ICC, 2005.

[26] J. Zhu and S. Papavassiliou. On the Connectivity Modeling and the Tradeoffs between Reliability and Energy Efficiency in Large Scale Wireless Sensor Networks. In Proc. IEEE Wireless Comm. and Networking Conf., Vol. 2, pp. 1260-1265, 2003.

[27] L.Hu. Evans. Secure Aggregation for Wireless Networks. In Proc. of SACNT, 2003.

[28] Yi Yang et al. A Secure Hop-by-Hop Data Aggregation Protocol for Sensor Networks. In MobiHoc, 2006.