IJCTA, 10(03), 2017, pp. 119-125

A Secure Data Aggregation Technique for Wireless Sensor Networks

A. Anitha* K. Arthi Krishna* and K.R. Eswaran Aasan**

Abstract : At present, due to limited computational power and energy resources of sensor nodes, aggregation of data from multiple sensor nodes done at the aggregating node is usually accomplished by simple averaging methods. However, such aggregation has been known to be highly vulnerable to node compromising attacks. Since Wireless Sensor Networks are usually unattended and untamper resistant hardware, they are highly susceptible to collusion attacks. Thus, ascertaining trust-worthiness of data and reputation of sensor nodes has become crucially important for WSN. As the performance of very low power processors dramatically improves and their cost is drastically reduced, future aggregator nodes will be capable of performing more sophisticated data aggregation algorithms, which will make WSN less vulnerable to severe impact of compromised nodes. Iterative filtering algorithms hold great promise for such 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. This work demonstrates a number of existing iterative algorithms, while significantly more robust against collusion attacks than the simple averaging methods, are nevertheless susceptive to a novel sophisticated collusion attack. To address this security issue, improvements for iterative filtering techniques are proposed by providing an initial approximation for such algorithms which makes them not only collusion robust, but also more accurate and faster converging.

Keywords : WSN, Collusion attack, Iterative filtering.

1. INTRODUCTION

A wireless network consists of nodes capable of grouping information from the environment and communication with each other via wireless transceivers. The collected information are delivered to one or extra sinks, generally via multi-hop communication. The nodes unit of measurement typically expected to figure with batteries and unit of measurement typically deployed to not-easily-accessible or hostile surroundings, generally in large quantities. It will be difficult or inconceivable to change the batteries of the nodes. On the alternative hand, the sink is commonly created in energy. Since the energy is that the foremost precious resource among the, economical utilization of the energy to prolong the network fundamental quantity has been the most target of plenteous of the analysis on the. The communications among this has the several-to-one property in this information from AN outsized sort of nodes tend to be centered into many sinks. Since multi-hop routing is generally needed for distant nodes from the sinks to avoid wasting energy, the nodes near a sink are going to be burdened with relaying AN outsized amount of traffic from completely different nodes. Network management is that the tactic of managing, monitoring, and dominant the behavior of a network.

* Noorul Islam Centre for Higher Education, Department of Computer Science and Engineering, Kumaracoil, Tamilnadu, India. *E-mail: anidathi@yahoo.co.in*

** Noorul Islam Centre for Higher Education, Department of Electrical and Electronics Engineering, Kumaracoil, Tamilnadu, India. *E-mail : arthikrishna7@gmail.com*

Wireless Sensor Networks (WSNs) produce distinctive challenges for network management that make ancient network management techniques impractical. In ancient networks the primary goals square measure minimizing latent amount and providing comprehensive information, but in detector networks the primary goal is minimizing energy use and conjointly the most suggests that for doing this will be by reducing the quantity of communication between nodes. Optimizing the operational and purposeful properties of WSNs may wish a singular declare each application draw back. Network failures square measure common events rather than exceptional ones. Thus, in WSNs, a bent to face live primarily committed observance and dominant node communication thus on optimize the efficiency of the network, make certain the network operates thoroughly, maintain the performance of the network, and management large numbers of nodes whereas not human intervention.

Monitoring individual nodes really terribly giant detector network is additionally impractical. it's spare to regulate the network 2 by guaranteeing specific network coverage. What is extra, sensor nodes unit generally deployed in remote or harsh conditions so the configuration of nodes in WSNs changes dynamically. Thus, a detector network management system got to alter the network to self-forming, self-organize, and ideally to self-configure at intervals the event of failures whereas not previous data of the topology. Despite the importance of detector network management, there's not any existing generalized resolution for WSN management. However, most detector network applications unit of activity designed with network management in mind then no additional network management layer is needed.

Cluster Head could be a node collect all the knowledge and aggregates the info. That information are forward to base station. Avoiding misconduct activities here cluster head act as associate degree critic of all the cluster members Nodes. Here nodes generated the profiles. Information from multiple sensors is collective at associate degree someone node that then forwards to the bottom station solely the mixture values. At present, as a result of limitations of the computing power and energy resource of device nodes, information is collective by very simple algorithms like averaging. There is no assurance for CH won't act as Malicious Node. Difficult generate Keys.

2. LITERATURE SURVEY

Reputation systems provide mechanisms to produce a metric encapsulating reputation for a given domain for each identity within the system. These systems seek to generate an accurate assessment in the face of various factors including but not limited to unprecedented community size and potentially adversarial environments [1].

The purpose of trust and reputation systems is to strengthen the quality of markets and communities by providing an incentive for good behavior and quality services, and by sanctioning bad behavior and low quality services. However, trust and reputation systems will only be able to produce this effect when they are sufficiently robust against strategic manipulation or direct attacks. Currently, robustness analysis of TRSs is mostly done through simple simulated scenarios implemented by the TRS designers themselves, and this cannot be considered as reliable evidence for how these systems would perform in a realistic environment. In order to set robustness requirements it is important to know how important robustness really is in a particular community or market. This paper discusses research challenges for trust and reputation systems, and proposes a research agenda for developing sound and reliable robustness principles and mechanisms for trust and reputation systems [2].

The concept of trust has become very relevant in the late years as a consequence of the growth of fields such as internet transactions or electronic commerce. In general, trust has become of paramount importance for any kind of distributed networks, such as wireless sensor networks (WSN in the following). By considering trust as a factor to take into account on the relationship between two peers, it is possible to deal with the inherent uncertainty of the cooperation process. Differing on the underlying model trust management systems are classified into credential-bases trust management systems (*i.e.* based on the identity of a node) or behavior based trust (*i.e.* based on the actions of a node) [3].

As sensor networks are being increasingly deployed in decision making infrastructures such as battlefield monitoring systems and SCADA (Supervisory Control and Data Acquisition) systems, making decision makers aware of the trustworthiness of the collected data is a crucial. The trust scores of data items are computed from their value similarity and provenance similarity. The value similarity comes from the principle that "the more similar values for the same event, the higher the trust scores". The provenance similarity is based on the principle that "the more different data provenances with similar values, the higher the trust scores". Experimental results show that our approach provides a practical solution for trustworthiness assessment in sensor networks [4].

Wireless Sensor Network 'WSN' is an active research field which explores many technological challenges, while the WSN node design is one of the most challenging areas. Nowadays, many wireless sensor nodes are implemented such as BT nodes, l'ESB/2 nodes, SmartTags, EYES node, Tiny Node, Mote, Mica2, Tmote Sky, Atlas and I mote. Note that, all these wireless sensor nodes are quite similar in term of functionality. They are in general based on 8 or 16 bit RISC (Reduced Instruction-Set Computer) microcontroller (ATMEGA128 or MSP430) equipped with a unique Bluetooth or ZigBee wireless access medium having 200 meter LOS (Line Of Sight) range, and enable to implement a simple wireless sensor node. The robustness and the reliability of wireless sensor node are important for many applications such as smart care, smart home and outdoor applications [5].

Ranking problem of web-based rating system has attracted many attentions. A good ranking algorithm should be robust against spammer attack. Here we proposed a correlation based reputation algorithm to solve the ranking problem of such rating systems where user votes some objects with ratings. In this algorithm, reputation of user is iteratively determined by the correlation coefficient between his/her rating vector and the corresponding objects' weighted average rating vector. Comparing with iterative refinement (IR) and mean score algorithm, results for both artificial and real data indicate that, the present algorithm shows a higher robustness against spammer attack [6].

With the explosive growth of accessible information, especially on the Internet, evaluation-based filtering has become a crucial task. Various systems have been devised aiming to sort through large volumes of information and select what is likely to be more relevant. This system analyze a new ranking method, where the reputation of information providers is determined self-consistently [7].

Advances in information technology reduce barriers to information propagation, but at the same time they also induce the information overload problem. For the making of various decisions, mere digestion of the relevant information has become a daunting task due to the massive amount of information available. This information, such as that generated by evaluation systems developed by various web sites, is in general useful but may be noisy and may also contain biased entries [8].

With the growth of the Internet and E-commerce, bipartite rating networks are ubiquitous. In such bipartite rating networks, there exist two types of entities: the users and the objects, where users give ratings to objects. A fundamental problem in such networks is how to rank the objects by user's ratings. Although it has been extensively studied in the past decade, the existing algorithms either cannot guarantee convergence, or are not robust to the spammers. In this proposed system, six new reputation-based algorithms, where the user's reputation is determined by the aggregated difference between the user's ratings and the corresponding object's rankings.

Trust and reputation play critical roles in most environments wherein entities participate in various transactions and protocols among each other. The recipient of the service has no choice but to rely on the reputation of the service provider based on the latter's prior performance. This proposed system introduces an iterative method for trust and reputation management referred as ITRM. The proposed algorithm can be applied to centralized schemes, in which a central authority collects the reports and forms the reputations of the service providers as well as report/rating trustworthiness of the (service) consumers. The proposed iterative algorithm is inspired by the iterative decoding of low-density parity-check codes over bipartite graphs. The scheme is robust in filtering out the peers who provide unreliable ratings. We provide a

detailed evaluation of ITRM via analysis and computer simulations. Further, comparison of ITRM with some well-known reputation management techniques (*e.g.*, Averaging Scheme, Bayesian Approach and Cluster Filtering) indicates the superiority of our scheme both in terms of robustness against attacks (*e.g.*, ballot-stuffing, bad-mouthing) and efficiency.

3. PROPOSED SYSTEM

The base station consists of some data aggregators. In the proposed system shown in Fig. 1, it considers two data aggregators (*i.e*, sender and receiver) is connected with the base station. The data aggregator includes number of clusters. The clusters are the collection of data or information. If the data is send from the cluster through the wireless sensor networks means, then there will be a chance of data loss or hacking (*i.e*, insecurity of data). So the proposed system includes the improved iterative filtering algorithm. The repeated process will leads in securing the data. The data which is send from the cluster will first reaches the base station in the format of coded information. The original data which is multiplied with the keyword in matrix form to get the coded information. Thus the coded information is then transferred to the data aggregator2. If anyone tries to hack the information means, then a notification message will be send to each cluster connected with the base station. The receiver will decode the coded information to get the original data. Thus the data is securely send in wireless sensor network by the proposed system [10].

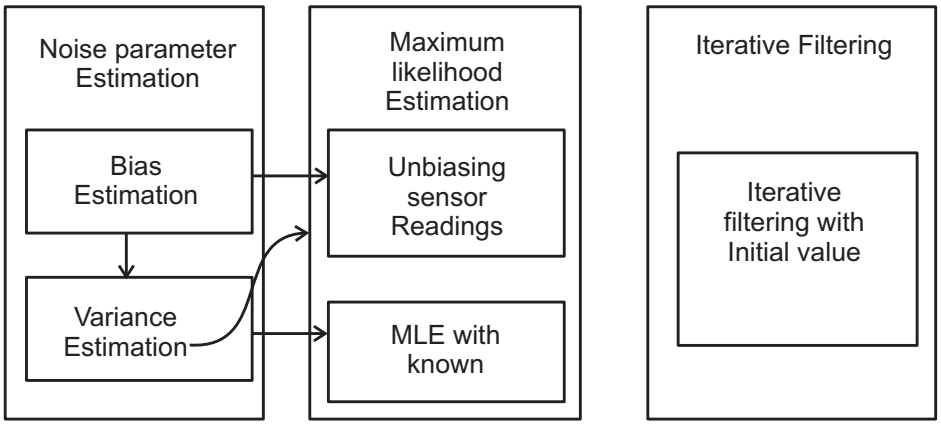

Figure 1: Architecture of the Proposed System

Advantages : When CH acts as malicious node it can remove the CH also. This method is highly secured than the existing systems. The CH of the proposed system has low work load compared to the existing systems.

A. Cluster Formation

The device nodes square measure divided into disjoint clusters, and every cluster includes a cluster head that acts as associate degree soul. Knowledge square measure sporadically collected and aggregative by the soul. Soul itself isn't compromised and concentrates on algorithms that build aggregation secure once the individual device nodes may well be compromised and may well be causation false knowledge to the soul. we have a tendency to assume that every knowledge soul has enough process power to run associate degree IF algorithmic program for knowledge aggregation.

HEF: Without a prior data (such as network period, residue energy state, and therefore the energy consumption for clusters), it's not possible for any cluster head choice rule to get sensible results for prolonging the network period. The core plan of the HEF cluster rule is to decide on the highest-ranking energy residue sensing element as a cluster head. The HEF cluster rule is outlined as follows. Some researchers have claimed that HEF is associate degree economical cluster choice rule that prolongs network period supported simulations. However, their measurements and simulation results area unit random processes. A theoretical roof to demonstrate the optimality of HEF underneath sure conditions is provided during this paper.

B. Adversary Model

In this model, we have a tendency to produce some cluster of sensors inject any false knowledge through the compromised nodes. All the data that is within the node becomes accessible by the soul. soul model, the assaulter node tried to vary minimum 2 sensors report. It listen the important report and it\'ll turn out to be skew report.

C. Robust Data Aggregation

Robust information aggregation, aggregation node aggregates all the info that area unit comes from varied sensors within the cluster. Here we want secure information aggregation, for this purpose we have a tendency to invariably analyze trait of the detector nodes. If error distribution of sensors is either glorious or calculable, our algorithms are often custom-made to alternative distributions to attain associate degree optimum performance. Our aggregation technique operates on batches of consecutive readings of sensors, continuing in many stages. Finally, it estimate the trait of every detector supported distance of the readings.

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

The collusion attack scenario in the IF algorithms are maintained with the accuracy. It can be seen that the IF algorithms with reciprocal discriminant function are highly vulnerable to such attack scenario, while the affine discriminant function generates more robust results in this case. However, the accuracy of the affine discriminant function is still much worse than the previous experiment without the collusion attack. This simulation results shows that the collusion attack scenario can circumvent all the IF algorithms. Moreover, the accuracy of the algorithms dramatically decreases by increasing the number of compromised nodes participated in the attack scenario. As explained before, the algorithms converge to the readings of one of the compromised nodes, namely, to the readings of the node which reports values very close to the skewed mean. This demonstrates that an attacker with enough knowledge about the aggregation algorithm employed can launch a sophisticated collusion attack scenario which defeats IF aggregation systems.

The accuracy of this approach is made by taking into account the IF algorithm with reciprocal and affine discriminant functions, respectively. This proposed approach is superior to all other algorithms in terms of the accuracy for reciprocal discriminant functions, while the approach has a very small improvement on affine function. Moreover, comparing the accuracy of our approach in this experiment with the results from no attack and simple attack methods that this approach with reciprocal discriminant function is robust against the collusion attack scenario. The reason that this approach not only provides the highest accuracy for this discriminant functions, it actually approximately reaches the accuracy of No Attack scenarios. The IF algorithms in the proposed attack scenario is that they quickly converge to the sample mean in the presence of the attack scenario. In order to investigate the shortcoming, we conducted an experiment by increasing the sensor variances as well as the number of colluders. In this experiment, we quantified the number of iterations for the IF algorithm with reciprocal discriminant function (dKVD-Reciprocal and Robust Aggregate-Reciprocal algorithms). The results obtained from this experiment show that the original version of the IF algorithm quickly converges (after around five iterations) to the skewed values provided by one of the attackers, while starting with an initial reputation provided by our approach, the algorithms require around 29 iterations, and, instead of converging to the skewed values provided by one of the attackers, it provides a reasonable accuracy. The results of this experiment validate

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

4. RESULTS AND DISCUSSION

In the simulation all the nodes will select a base station through Cluster Head by LEACH protocol. CH will collects the data and it will send to destination. CH will always validate the data originality. This process will be done CH, it will filtering the data and compared with neighbour. If any false node detected in the network, CH will eliminate the node.

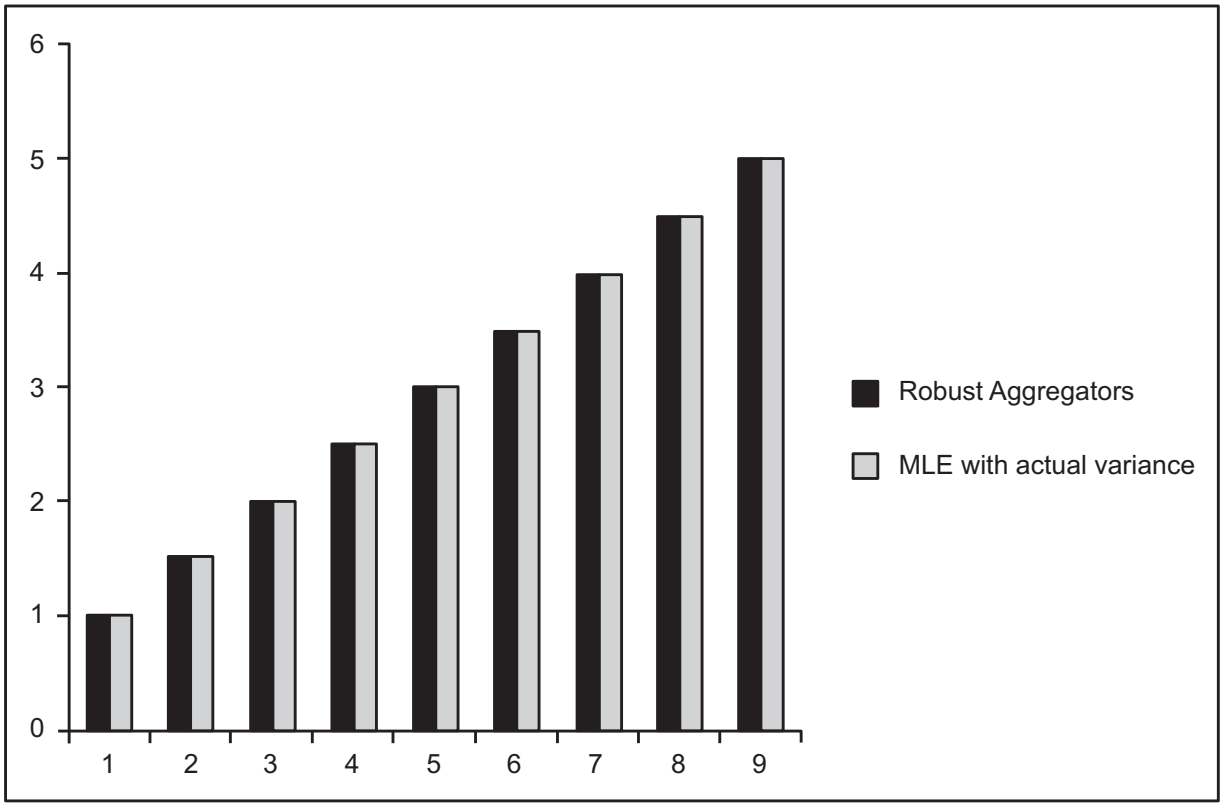

Figure 2: Unbiased Error

Unbiased error

Here various distributions of the variance across the set of sensors are considered and obtained similar results. We have chosen to present the case with the error of a sensor s at time *t* is given, considering different values for the baseline sensor variance shows the results of the MLE with our noise parameter estimation 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 this algorithm. The proposed approach nearly exactly achieves the minimal possible variance coming from the information theoretic lower bound. The performance of this approach for the initial trustworthiness assessment of sensors with different discriminant functions as well as other IF algorithms. It shows that in this experiment, the performance of this approach with both discriminant functions is very similar to the original IF algorithm.

5. CONCLUSION

Clustering mechanism is one of the finest algorithms for improving QOS in WSN. Our system is specially designed for both security and QOS. In secure data aggregation mechanism data validation by sensors Iteration Filtering algorithm. If any false data is detected, which node send that data that node will be eliminated by a network also compression mechanism will improves the QOS of network. In future the

network will be specially designed for reduce no of over heads in the network. Clustering algorithm may be changed into dual clustering algorithm.

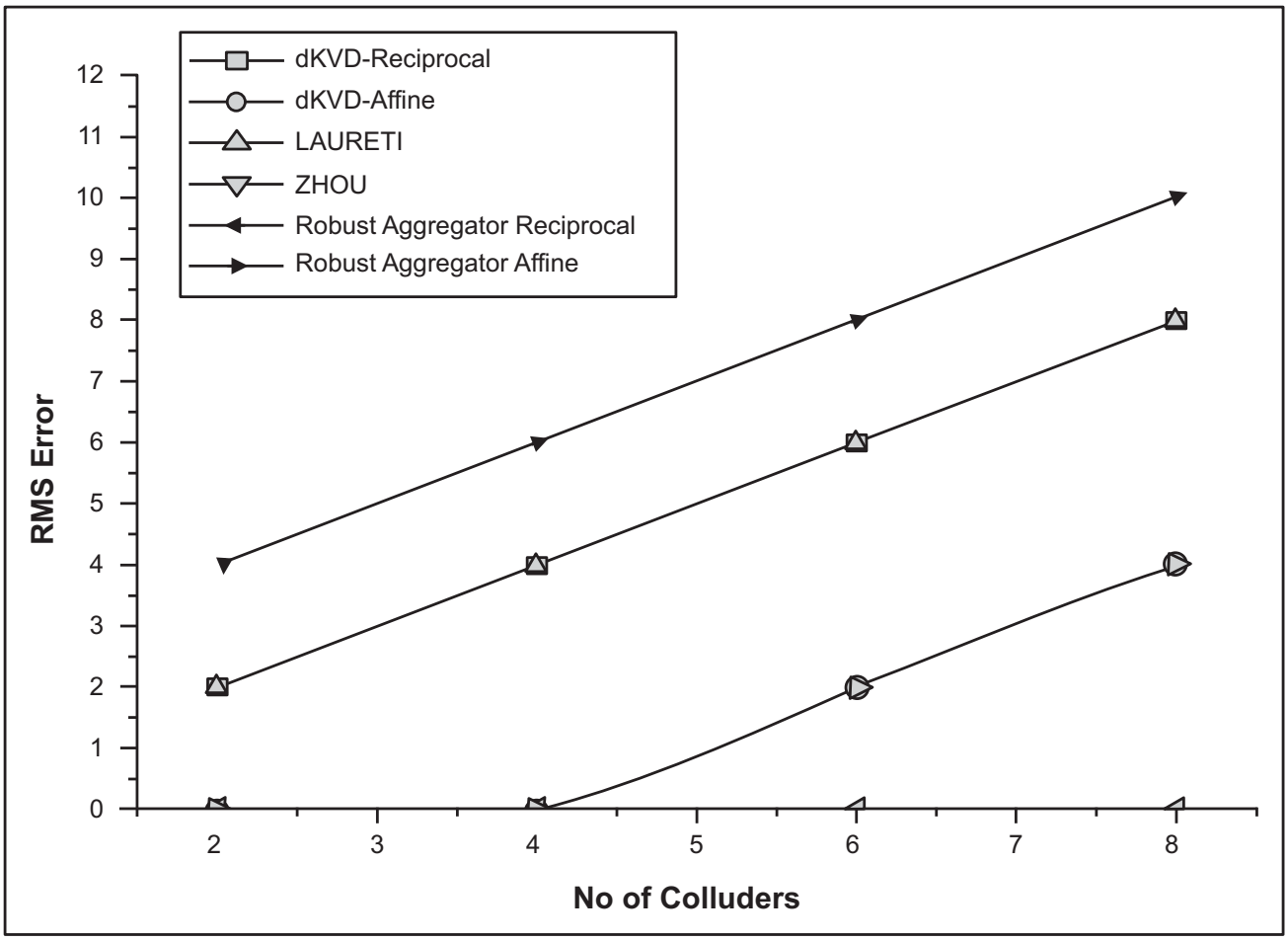

Figure 3: Accuracy of no attack scenarios

6. REFERENCES

1. S. Ozdemir and Y. Xiao, "Secure data aggregation in wireless sensor networks: A comprehensive overview," Comput. Netw., vol. 53, no. 12, pp. 2022–2037, Aug. 2009.
2. L. Wasserman, All of Statistics : A Concise Course in Statistical Inference. New York, 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.