International Journal of Ethics in Engineering & Management Education
Website: www.ijeee.in (ISSN: 2348-4748, Volume 1, Issue 4, April 2014)

Trust Based Data Aggregation in Wireless Sensor Networks

B Niket, Prof. Basavaraj S.Prabha
Department of Computer Science and Engineering
Bheemanna Khandre Institute of Technology, Bhalki-585328, India,
Visvesvaraiya Technological University Belgaum-590014, India.
niketb17@gmail.com, bsprabha@yahoo.com

Abstract- Sensor networks are collection of sensor nodes which co-operatively send sensed data to base station. As sensor nodes are battery driven, an efficient utilization of power is essential in order to use networks for long duration hence it is needed to reduce data traffic inside sensor networks, reduce amount of data that need to send to base station. The main goal of data aggregation algorithms is to gather and aggregate data in an energy efficient manner so that network lifetime is enhanced. Wireless sensor networks (WSN) offer an increasingly Sensor nodes need less power for processing as compared to transmitting data. In addition, sensor devices are limited in resources and vulnerable to node capture, so even a few malicious adversaries can easily compromise sensor devices and inject forged data into the networks to make the networks be in confusion. Therefore, a novel trust management scheme is necessary to distinguish illegal nodes from legal ones, and filter out malicious nodes' deceitful data in the networks. In this paper, to make resilient wireless sensor networks, we propose a secure data aggregation scheme based on trust evaluation model which can identify trustworthiness of sensor nodes. This model suggests a defensible approach against insider attacks incipiently beyond standard authentication mechanisms and conventional key management schemes.

Keywords—resilient sensor networks, trust evaluation, security, secure Aggregation.

1. INTRODUTION

Wireless sensor networks suggest potentially beneficial solutions for various applications including climate and temperature monitoring, freeway traffic analyzing, people's heart rates sensing, and many other military applications [1]. A major feature of these systems is that sensor nodes in networks assist each other by passing data, in network process and control packets from one node to another. It is often termed an infrastructure-less, self-organized, or spontaneous network [2].However, sensor networks tend to be organized in open environment, thus some false data broadcasted from irrespective nodes might be injected into the networks regardless of their intention. In addition, sensor networks are susceptible to a variety of attacks, for example node capture, eavesdropping, denial of services, wormhole, and sybil attack [4]. So, a certain amount of sensor nodes can be compromised by adversaries, and the compromised nodes can successfully authenticate bogus reports to their neighbors, which have no way to distinguish false data from legitimate ones [3]. In either case, to differentiate false data from legal ones is an essential process for a normal and effective function of sensor networks, because false reports can drain out the finite amount of energy resources in a battery powered sensor networks, and even a small amount of compromised nodes can influence the whole sensor networks critically. A few recent research efforts have proposed mechanisms to provide authentication for sensor networks to prevent false data injection by an outsider attacker. Their basic approaches for security are to use Message Authentication Codes (MACs) and probabilistic key predistribution schemes. The approaches prevent naïve impersonation of a sensor node, however they cannot prevent the injection of forged or false data from malicious or compromised insider nodes which have already been authenticated as legal ones in the networks. Once authenticated as a legitimate node, broadcasting data from that node cannot help being accepted as trustful data in the networks without any question. Therefore, a smart trust management scheme is needed to identify trustworthiness of sensor nodes in order to distinguish between malicious nodes and innocuous nodes, and to strengthen reliable nodes and weaken suspicious nodes. However, there have not been many of researches for trust evaluation models which are applicable to wireless sensor networks properly. Here, we propose a trust-based aggregation scheme for resilient wireless sensor networks, which helps the networks to operate normally with high probability although some nodes or data would be compromised. General direction for resilience is to gather multiple and redundant sensing data and crosscheck them for consistency. For reasonable crosschecking, we compare them with the expected sensing results within the possible and legitimate sensing range. Based on the result of that crosschecking, each node estimates its neighbour nodes' trust values. As the sensor nodes operate trustworthily, they will get higher trust values from its neighbor nodes. On the other hand, of course, as the sensor nodes operate maliciously or inconsistently, they will get lower trust values.

Data Aggregation is a process of merging sensed data from multiple sensor nodes at any of the aggregator node in the network. Every sensor node must send its data to its upstream neighbor.Intermediate nodes along the path to the aggregator fuse the data received from the downstream nodes with their own data and forward the local aggregated value towards aggregator. The Aggregator must perform final aggregation on the data received from its neighbors and then forward the result to the BS through the sensor nodes. The purpose of Data Aggregation is to eliminate redundant data transmission and provide fused information to the BS, to conserve energy and bandwidth of resource constrained sensor nodes and hence reducing the communication cost between sensor nodes in the network. Since the compromised nodes have access to cryptographic keys and also compromised nodes that pose as an authorized node in the network cannot be detected easily using cryptographic primitives in the data aggregation process. To overcome these drawbacks, Reputation and Trust management can be employed in the process of data aggregation In our proposed scheme, to perform trust and reputation mechanism we have identified the different roles of sensor nodes in the context of trust evaluation in data aggregation process of sensor networks in order to prolong the network life time, have a reliable aggregation as well as reliable data delivery to the destination.

Each sensor node must maintain Reputation Table and Trust Table to record the reputations and trust values of all its neighbors based on their roles respectively. The values in the reputation table and trust table maintained by each node is periodically updated based on their trustworthiness

2. PROBLEM DEFINITION

Data aggregation protocols aims at eliminating redundant data transmission and thus improve the lifetime of energy constrained wireless sensor network. In wireless sensor network, data transmission took place in multi-hop fashion where each node forwards its data to the neighbor node which is nearer to sink. Since closely placed nodes may sense same data, above approach cannot be considered as energy efficient. An improvement over the above approach would be clustering where each node sends data to cluster-head (CH) and then cluster-head perform aggregation on the received raw data and then send it to sink.

We consider the security of Data Aggregation in which merging of sensed data from multiple sensor nodes at any of the aggregator node in the network. Every sensor node must send its data to its upstream neighbor. The Aggregator must perform final aggregation on the data received from its neighbors and then forward the result to the BS through the sensor nodes and for this the intermediate nodes should be trustworthy. The process continues until the data reaches the destination. This is an attractive approach for aggregating the data in WSNs due to its low overhead and localized communication. Here, nodes only need to interact with their upstream one-hop neighbors to exchange the information and forward the data. It is widely used in ad hoc and wireless sensor networks due to its scalability.

3. RELATED WORK

3.1 Trust Evaluation Model

Current security researches for trust management schemes mainly focus on more powerful ad hoc networks than sensor networks. Z.Yan, P. Zhang, and T. Virtanen proposed a trust evaluation model [5].In this model, each node should evaluate trust values of all the other nodes globally in the networks. Such a global computation for all the other nodes cannot be accomplished in practical resource-constraint sensor networks. In addition, trust evaluation factors used in that model cannot reflect malice of the illegal nodes, rather just check experience statistics such as communication success, reference count, and personal preference. So, the previous trust evaluation model can hardly exclude maliciously forged data in the networks.

3.2 Inconsistency Check

Efficient inconsistency check mechanisms are mainly researched in intrusion detection research areas. Generally, intrusion detection systems consider unexpected results or events which are out of their learned pattern as intrusions. In order to train the intrusion detection system, some machine-learning models, for example hidden Markov model, are adopted to the system and the system is trained by a large number of training data . Such an anomaly detection scheme is necessary in wireless sensor networks to find out malicious or compromised sensor nodes which act inconsistently. However, how to define such an anomaly model based on which training data is still a main challenge.

3.3 Key Management

Because of the infrastructure-less and resource-constrained features of sensor networks, traditional asymmetric key mechanisms, such as digital signature and public key encryption, are seldom applied to sensor networks. So, key management schemes using a small number of symmetric keys, while security level of the system is still remained high, are studied and proposed for wireless sensor networks However, such a key management scheme alone cannot help legal nodes in the networks to identify legitimacy of neighbor nodes which they communicate with. Moreover, it is a reasonable assumption that some nodes are likely to be deprived of secret keys by physical attacks[3]. So, a novel trust management scheme is necessary for secure and resilient wireless sensor networks.

3.4 Data Aggregation

In recent, secure data aggregation schemes are proposed by many researchers. Some researches are for resiliency of primitive aggregation function itself , and the other researches are for secure in-network aggregation methods which can be resilient even if some compromised nodes inject false data into the networks. They have proposed several schemes which can filter out injected false data in sensor networks. However, the proposed methods so far can only differentiate the deceitful data of outsiders or unauthenticated illegal nodes from authenticated legal ones. In sensor networks, detecting and filtering out forged data injected by already authenticated as legitimate insider node of the networks is a great research challenge. According to our surveys, there have not been proposed such a data aggregation schemes which could purge false data from compromised nodes previously authenticated as legitimate participants of the networks.

4. GOALS AND ASSUMPTIONS

4.1 Goals

We focus on making resilient wireless sensor networks which work normally even though some sensor nodes might be compromised. Without any trust evaluation mechanisms, we cannot guarantee the sensor networks to work appropriately even if the networks adopt cryptographic key management approaches because of the sensor networks' vulnerability to physical attacks. For the purpose of the resilience of sensor networks, we direct our approaches to evaluate trustworthiness of sensor nodes in consideration of some trust evaluation factors. As a result of the crosschecked and evaluated trust values from accumulated histograms, we can filter out inconsistent and deceitful data from the malicious or compromised nodes.

4.2 Assumption

We have some assumptions in our trust evaluation model as follows: (1) The network consists of a set of sensor nodes of unknown location and a set of specially equipped nodes, anchor nodes, with known location and orientation. Anchor nodes are trusted during localization step. (2) Sensor nodes are deployed densely enough to be able to sense some identical events redundantly with their neighbour nodes. (3) Malicious nodes do not collude with each other. That is,they do not manipulate or intentionally increase trust values for each other, but just try to inject spurious data into the networks.

5. TRUST-BASED AGGREGATION MODEL

We describe our trust-based secure aggregation model which consists of four steps. First, we divide sensing areas into some logical grids and assign a unique identification to each grid (Section4.1). Second, sensor nodes estimates their deployed location and a corresponding grid (Section 4.2). Third, each node evaluates trustworthiness of its neighbor nodes by crosschecking the neighbor nodes' redundant sensing data with its own result. Inconsistent data from malicious or compromised nodes can be detected. Fourth, special nodes, aggregators, aggregate sensing data from their grids and transmit the computed results to the destination node, or sink. Inconsistent data from malicious nodes can be excluded in this step (Section 4.3).

4.1 Step 1: Grid Definition

In this step, consider first sensing areas in which sensor nodes will be deployed and ready for some events. We can easily know the location information of the sensing areas before we deploy sensor nodes. Then, we divide the sensing areas into some logical grids in proportion to the sensing range of a sensor device. We define *r* to be the sensing range of a sensor device. The main focus on dividing in this step is to set the size of a logical grid to the extent that one sensor device's sensing range can cover the grid entirely which it belongs to.

Consider an ideal case that a sensor node is deployed at the center of a grid. There is a tradeoff between accuracy and cost between the two cases. Although there can be so many choices between the two extreme cases, in our model, we intend to use as many redundant sensing data from multiple sensor nodes as possible to identify inconsistent data among them. After dividing sensing areas into some logical grids, we assign a unique identification to each grid.

4.2 Step 2: Trust Evaluation

In this step, we propose a trust evaluation process. At the beginning each node is assigned a value of '1'meaning they all are trustworthy nodes and as the time passes when we detect the compromised node by the TWO ACK Mechanism we make value of that node to '0'meaning it is not trustworthy and hence we won't send the data through that node. In this model we record the behaviour of each and every node for secure data aggregation. The trust defined in our model is the confidence of a node on another node. The trust value means the level of trustworthiness of a node, which is computed based upon several trust evaluation factors, such as battery lifetime, sensing communication ratio, sensing result, and consistency level. In our scheme, sensor nodes do not compute all the other nodes' trust values in the networks [5], but compute only their neighbour nodes' trust values accumulatively. This local trust evaluation mechanism is suitable for resource-limited sensor networks. An accumulated evaluation mechanism using histogram makes the evaluation more accurate than just contemporary evaluation mechanism does. The scheme used in evaluation is described below:

A. The TWOACK Scheme

The TWOACK scheme can be implemented on top of any source routing protocol such as DSR. This follows from the fact that a TWOACK packet derives its route from the source route established for the corresponding data packet. The TWOACK scheme uses a special type of acknowledgment packets called TWOACK packets, which are assigned a fixed route of two hops (or three nodes) in the direction opposite to that of data packets. Figure 1 illustrates the operational details of the TWOACK scheme.

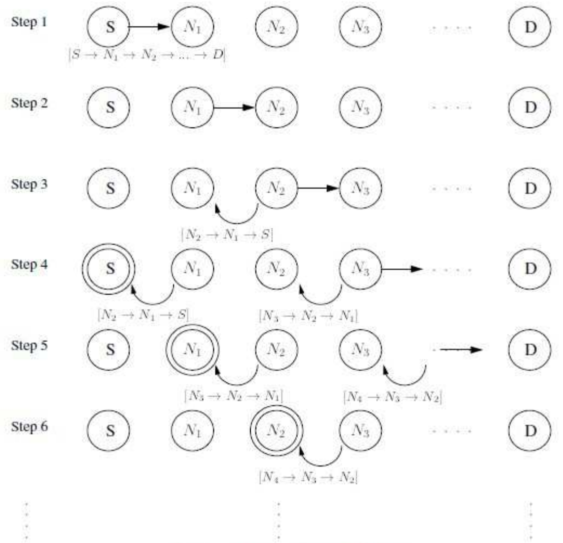
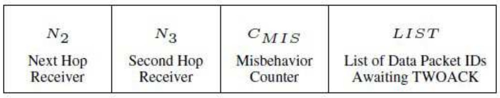


Fig. 1. The TWOACK Scheme

Suppose that the process of Route Discovery has already yielded a source route $[S \to N1 \to N2 \to N3 \to \cdot \quad \cdot \to D]$ from a source node S to destination node D. For instance, when $N1$ forwards a data packet to $N2$, to be forwarded on to $N3$, $N1$ has no way of knowing if the packet reached $N3$ successfully or not. Listening on the medium, as suggested in [9], would only tell $N1$ whether $N2$ is sending out the packet or not. However, the reception status at $N3$ is unclear to node $N1$. The possibility of collisions at both $N1$ and $N3$ makes the overhearing technique vulnerable to medium access problems and false detections [9]. The TWOACK scheme is designed to solve these problems: when $N3$ receives a data packet, it sends out a TWOACK packet over two hops back to $N1$, carrying the packet ID of the corresponding received data packet. The route $[N3 \to N2 \to N1]$ for the TWOACK packet is extracted from the source route of the original data packet. The aim of the TWOACK packet is to notify $N1$ that the data packet has successfully reached a node that is two-hop away, namely $N3$. Such a procedure will be carried out by every set of three consecutive nodes, termed *triplet*, along the source route. Note that the ACK packets at the TCP layer have a similar effect as our TWOACK packets do. The main differences are the following: First, ACK packets in TCP are used for the purpose of flow-control and reliable end-to-end communication, while selfishness is more a problem that should be solved by the underlying IP layer. In the absence of a lower layer acknowledgment scheme, the source and other intermediate nodes have no way of finding out which of the downstream nodes is misbehaving. It will be inefficient to conclude that the entire route is misbehaving when indeed there is only one misbehaving node. To correctly detect and isolate such a misbehaving node, additional techniques such as the TWOACK scheme need to be employed. Second, ACK packets in TCP have to travel all the way from the final destination back to the source. Therefore, depending on the length of the path used for data packets, it is likely that ACK packets will arrive after significant delays. In contrast, TWOACK packets travel exactly two hops, making the timeout period shorter and more predictable. To detect misbehaviour, the sender or router of a data packet maintains a list of data packet IDs that have yet to receive a TWOACK acknowledgment packet from a node two hops away. Each node maintains a unique list for each forwarding link that it is using. Each item on the list has the following data members (cf. Fig. 2):



N_2	N_3	C_{MIS}	$LIST$
Next Hop Receiver	Second Hop Receiver	Misbehavior Counter	List of Data Packet IDs Awaiting TWOACK

Fig. 2. Data Structure maintained for misbehavior detection

• $N2$ and $N3$: the receivers of the next two hops after this node, along the source route being used;
• *CMIS*: counter for number of instances of misbehaviour by forwarding link $N2 \to N3$;
• *LIST*: list of data packet IDs that are awaiting the TWOACK packets. When a node, say, $N1$, sends or forwards a data packet
along a particular route, say, $N1 \to N2 \to N3$, it adds the ID of the packet to *LIST* on its list corresponding to $N2 \to N3$. When it receives a TWOACK packet, it checks for the $N2 \to N3$ combination, and then removes the packet ID from the corresponding *LIST*. If a data packet ID stays on *LIST* longer than a certain period of time, termed *timeout*, misbehavior of link $N2 \to N3$ is suspected. Every time misbehavior is suspected, a non-negative misbehavior counter *CMIS* is increased by one. When *CMIS* exceeds a certain level, termed *thresh*, a node declares the corresponding link, $N2 \to N3$, misbehaving and sends out an RERR packet informing the source about the same.2 Every node receiving or overhearing

such an RERR packet should identify link *N*2 → *N*3 as misbehaving. Every node maintains a list of misbehaving links that it has learned. Such links will not be chosen when it selects routes for data transmission later on.3 It might be unclear how the TWOACK scheme distinguishes *genuine route failures* from misbehaving nodes (links). Indeed, genuine route failures may take place due to mobility or excessive traffic in the vicinity of a forwarding node, e.g., *N*2. When such failures appear, *N*2 will voluntarily send an RERR packet to notify the source, as described in the routing protocol. Such an RERR packet is different from the RERR packet sent out by *N*1 reporting a misbehaving link *N*2 → *N*3. The values assigned to *thresh* and *timeout* play an important role in determining the effectiveness of the TWOACK scheme. These parameters should be large enough so that intermittent failures or excessive transmission delays (due to collisions) of TWOACK packets are not interpreted as misbehavior. On the other hand, they should not be so large that a significant number of data packets are lost before a misbehaving node (link) is detected

4.3 Step 3: Data Aggregation

In this step, we propose a data aggregation scheme based on trust value of each node evaluated in step 3. Data aggregation is an essential process in wireless sensor networks to eliminate redundancy of sensing data, to minimize communication overhead, and to save energy. First, sensing data of multiple nodes are aggregated per grid. To aggregate data, we elect one node as an aggregator per each grid. Then, the aggregator obtains sensing data from the other member nodes in its grid and aggregates them to a representative value in consideration of the trust values of the member nodes. After that, an aggregator node aggregates the data and sends it to the base station.

6. CONCLUSION

We proposed a trust-based data aggregation scheme for wireless sensor networks. Though it takes long path for some cases, the proposed idea provides highly secured data aggregation and achieves better performance in terms of network lifetime as it has considered the effect of certain nodes out of energy to the whole sensing function of the networks. As we are considering the data aggregation, it reduces the communication overhead and hence increases network lifetime .As we referred, the security for wireless sensor networks is still in its infancy and there are not clear trust evaluation models which can be applied to sensor networks properly. Our aggregation scheme does not employ cryptographic approaches or certification mechanisms, so it is light enough to fit well with wireless sensor networks without great overheads. To the best of our knowledge, our approach is one of the incipient researches on a secure aggregation scheme based on trust evaluation model for wireless sensor networks, which can detect malicious and compromised sensor nodes, and filter out the inconsistent sensing data of them.

REFERENCES

[1]. H. Chan and A. Perrig, *Security and Privacy in Sensor Networks*, IEEE Computer 2003.

[2]. A. Pirzada, C. McDonald, *Establishing Trust In Pure Ad-hoc Networks*, Proceedings of the 27th conference on Australasian computer science, 2004.

[3]. A. Perrig, J. Stankovic, D.Wagner, *Security in Wireless Sensor Networks*, Communication of the ACM, June 2004.

[4]. C. Karlof, D. Wagner, *Secure Routing in Wireless Sensor Networks: Attacks and Countermeasures*, NEST 2003.

[5]. Z. Yan, P. Zhang, T. Virtanen, *Trust Evaluation Based Security Solution in Ad Hoc Networks*, NordSec 2003, Proceedings of the Seventh Nordic Workshop on Secure IT Systems, 15th-17th October 2003.

[6]. N. Sastry, U. Shankar, D. Wagner, *Secure Verification of Location Claims*, Proceedings of the 2003 ACM workshop on Wireless security.

[7]. B. Przydatek, D. Song, A. Perrig, *SIA: Secure Information Aggregation in Sensor Networks*, SenSys 2003.

[8]. K. Balakrishnan, “Prevention of Node Selfishness in Mobile Ad Hoc Networks”, *M.S. Thesis*, Department of EECS, Syracuse University, Syracuse, NY, USA, August 2004.

[9]. S. Buchegger and J-Y. Le Boudec, “Performance Analysis of the CONFIDANT Protocol: Cooperation Of Nodes, Fairness In Dynamic Ad-hoc Networks”, *Proc. of the IEEE/ACM Symposium on Mobile Ad Hoc Networking and Computing (MobiHOC)*, June 2002.