A Group-based Security Mechanism for Distributed Sensor Networks

Md. Abdul Hamid and Choong Seon Hong

Department of Computer Engineering, Kyung Hee University, 449-701, Korea

hamid@networking.khu.ac.kr and cshong@khu.ac.kr

Abstract

This paper presents a group-based Secure Data Aggregation (SDA) scheme for distributed Wireless Sensor Networks. The scheme is devised for dealing with sensory data aggregated by groups of collocated sensors. We model the network for secure routing with 3-tier sensor network composing three types of nodes: Base Station (BS), Group Dominator (GD) and ordinary Sensor Nodes (SN). More than 85% network connectivity can be achieved with the proposed model. Key pre-distribution is performed with the concept of star key graph where one superior sensor node (GD) dominates other ordinary nodes. We devise a light-weight scheme that combines the star graph to form a secure weakly connected dominating set where GDs form the backbone of the entire network. Analysis on network robustness is presented using the properties of Poisson Process. Simulation results are demonstrated to validate the analysis. The scheme is simple and scalable and it offers a stronger resilience against node capture attacks.

1. Introduction

Much works are going on in designing storage and computationally inexpensive mechanism [1,11] for WSN security. There are some important issues in engineering a particular security protocol. Firstly, key storage for individual sensor node needs to be reasonably small. For example, if there are *N* nodes in the network, then we cannot expect that a node can store $N - 1$ keys to share a secrete key with each of the other nodes. Secondly, in case where quite a good amount of sensor nodes are compromised by an adversary, the communications among other nodes should still be secure. Forward secrecy and backward secrecy must be maintained to devise a stalwart network. Thirdly, it should be ensured that both local and global connectivity is maintained. A sensor node should be able to securely communicate with its local neighbors. Connectivity among local zones should provide global network connectivity [10]. Finally, asymmetric cryptography to WSN is too expensive, because they require expensive computations and long messages that might easily exhaust the sensor's resources [9]. That's why we take symmetric cryptographic operations. The two types of nodes (ordinary sensor nodes and group dominator nodes) perform different task in our secure data aggregation scheme. An ordinary sensor senses the events and provides its group dominator a proof for any sensory report it has agreed. A group dominator collects raw sensing data from ordinary sensors, makes a report and sends the report to the base station.

Rest of the paper is organised as follows. Section 2 outlines the network model and preliminaries; Section 3 presents our scheme in details. We analyze our scheme in Section 4. Related works and comparisons are noted in Section 5 and Section 6 concludes this article.

This paper was supported by ITRC and MIC

2. Assumptions: Network Model and Preliminaries

Clustering is a good approach to alleviate the scalability problem and to simplify the overall network structure [10]. We adopt a graph theoretic approach of dominating sets [10,12]. The edges covered by one vertex in a vertex cover are the edges incident to it and they form a star. We assume our network model as a heterogeneous network, where three types of entities are present; Sink/Base Station (BS), Group Dominators (GD), and ordinary Sensor Nodes (SN).

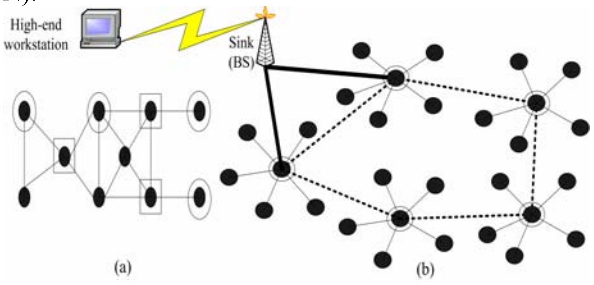


Fig 1. (a) Minimal dominating set (circles) and minimum dominating set (squares). (b) Proposed weakly connected dominating set induced by circles as stars for our scheme.

In a graph *G*, a set $S \subseteq V(G)$ is a dominating set if every vertex in *S* has a neighbor in *S*. The dominating number $\gamma(G)$ is the minimum size of a dominating set in *G*. For example, in Fig. 1a, the graph *G* has a minimal dominating set of size *4* (circles) and a minimum dominating set of size *3* (squares). Also, we note that, covering the vertex set with stars may not require as many stars as covering the edge set. When a graph *G* has no isolated vertices, every vertex cover is a dominating set [12]. A weakly connected dominating set (WCDS), S_W, is a dominating set where the graph induced by the stars of the vertices in S_W is connected. A star of a vertex comprised of the vertex itself and all the vertices adjacent to it (All the circles in Fig. 1b).

We assume that each SN is simple, inexpensive and stringent in resources (power, memory and computation), while each GD is rich is resources and more compromise-tolerant and having transmission range more than $2R_{SN}$, where R_{SN} is the transmission range of an ordinary sensor node. We also assume that one GD can communicate with its neighbor GD to forward aggregated messages towards the base station. There is no communication link among SNs within one group and between SNs in different group (Fig. 1(b)). Though neighbor nodes can overhear each other, but we assume that no message is exchanged among ordinary sensor nodes either in one group or in different groups. We consider the sensors in the whole network as a graph $G = (V,E)$, where V is the set of sensors in the network and E is the set of direct communication links. A direct communication link is present between SNs and its corresponding GDs if and only if they are within the transmission range of one another and share the same key. The underpinning of our proposed scheme is the star based weakly connected dominating set. In fact, it is easy to see that each dominating node (or vertex) in the star based WCDS is at the center of a star (Fig. 1b).

3. Proposed Secure Data Aggregation Scheme

We assign heavy works to the sink (BS) that is responsible for smooth functioning of the overall network. In our approach, key pre-distribution is performed using the concept of star key graph [2]. This is the special class of a secure group where each sensor node has only *3* keys to maintain: its individual key (shared between SN and GD), and a local group key that is shared by every user in the star graph with GD and a pairwise key between SN and BS. BS stores all the keys of SNs and GDs. We use the notations in table 1 to describe our scheme.

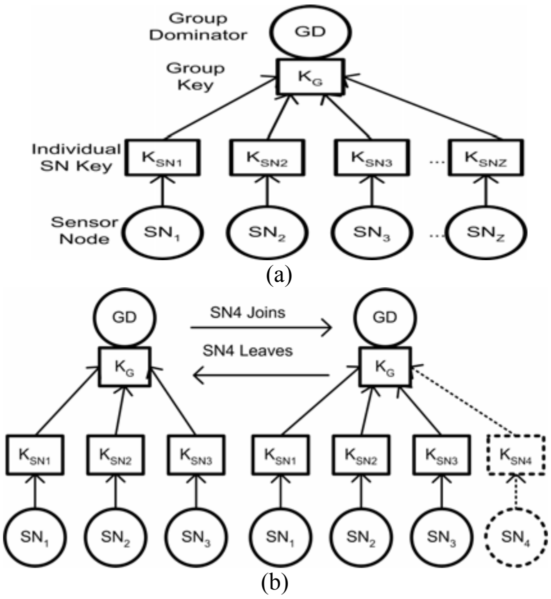
Table 1. Notations

Notation	Definition
i (0 ≤ i ≤ N)	Ordinary sensor node i (SN_i)
j (0 ≤ j ≤ Y)	Group dominator j of ordinary sensor i (GD_j)
ID_i	ID of the ordinary sensor node i
ID_j	ID of the group dominator j
K_{Gj}	Group key shared by all sensors in a group j and GD_j
$K_{(SNi,GDj)}$	Pairwise key between a sensor i and GD_j
$K_{(SNi,B)}$	Pairwise key between sensor i and Sink/BS
$K_{(GDj,B)}$	Pairwise key between GD_j and Sink/BS
M_i	Event sensed by SN_i
M_{GDj}	Message aggregated by GD_j
MAC(K,M)	Computation of Message Authentication Code of message M using key K
E(K,M)	Encryption of message M using key K
X|Y	Concatenation of X and Y

3.1 Pre-deployment Key Pre-distribution and Rekeying

Key Pre-distribution: In the offline key pre-distribution phase, we assign the group keys and individual keys to a group of nodes. For this, the key assignment is accomplished according to Fig. 2a. Each GD holds group key and all individual keys of SNs and a shared key with BS. Each SN holds group key and its individual key shared with GD and a shared key with BS. In this phase, all the SNs are also assigned unique ID_i ($1 \leq i \leq N$) which are also stored by the respective GDs. Each GD is also assigned ID_j ($1 \leq j \leq Y$), where Y is the total number of group dominators in the network. We set the ID of ordinary SN and GD to be *2* bytes long.

Rekeying: During the offline key pre-distribution, all the nodes are assigned the keys but not all the nodes are deployed. When any of those remaining nodes is deployed, it sends the JOIN_REQ_NEW message using its own individual key. If authorized by the access list of GD, it joins the group. Otherwise, GD forwards this to BS. BS informs GD about the individual key of that SN. If authenticated by BS, GD generates a new group key and encrypts the new group key with the newly added node's individual key and sends it to the SN. All other nodes in the group know about the change by a multicasting by the GD of that group. For leaving a star graph, the node simply leaves a message to inform the GD, which in turn generates a new group key and unicasts it within the group members (Fig. 2b).



(a)

(b)

Fig 2. (a) Pre-distribution of secret keys using star key graph. (b) Re-keying: Scenario when an ordinary sensor node joins/leaves a group.

For example, let's say, SN_4 wants to join the existing group in Fig. 2b shown above. GD changes the group key K_G to a new key K_{Gnew}, and sends the following rekeying messages:

$GD_j \rightarrow all\ SN_i$**:** $E_{K_G}(K_{Gnew})$, Encrypted new group key with the old group key.

$GD \rightarrow SN_4$**:** $E_{K_{SN4}}(K_{Gnew})$, Encrypted new group key with the joining SN's individual key. Similarly, when any SN wants to leave the group, it just sends a leave message. GD deletes the leaving SN and updates the K_G to new K_{Gnew}

and unicasts the following message:
$SN_4 \rightarrow GD$: E_{KSN4}(leave), SN_4 wants to leave the group.
$GD \rightarrow SN_{i-1}$: $E_{KSNi-1}(K_{Gnew})$, GD unicasts the new group key encrypted with remaining SN's individual key.

3.2 Secure Data Aggregation

GDs are responsible to aggregate data sensed from different sensors. If there are Z number of ordinary sensors (SN) in a group, for fidelity and correctness of data, the GD waits for the same sensing event from at least q ($q \leq Z$) number of the SNs, where q is the threshold value set for a particular group. For example let us consider any one group with ordinary SNs and its corresponding GD. Once an event occurs, q out of Z ($0 \leq q \leq Z$) ordinary sensors (ID_1, ID_2, . . ., ID_q) within the sensing area detect the event and send information to the GD as

$$SN_i \rightarrow GD_j : ID_i|E(M_i|MAC(M_i, K_{(SNi,GDj)}), K_{(SNi,GDj)}))$$

Upon receiving the message sensed by SN_i (ordinary sensor), Dominator GD_j verifies every single MAC and generates an aggregated report (and discards the false packet if any). GD broadcasts the aggregated results M_{GDj} and MAC to all sensors as

$$GD_j \rightarrow all\ SN_i : ID_{GDj}|\ M_{GDj}|MAC(M_{GDj}, K_{Gj})$$

All SNs in a particular group j verifies the aggregated report whenever they receive it for the consistency with its raw sensing result. It creates a MAC only to be verified by the BS but to be relayed by its GD. The message format is

$$SN_i \rightarrow GD_j : ID_i|MAC(K_{(SNi,B)}, ID_i|M_{GDj})$$

Now, GD collects all the MACs from ordinary sensor nodes and sends q MACs, q IDs, ID_{GDj}, and M_{GDj} to the sink directly or via its neighboring GD (multi-hop path through consecutive GDs towards the Sink) as

$$GD_j \rightarrow Sink\text{: } ID_{GDj}, E(K_{(GDj,B)}, ID_{GDj}|M_{GDj}|ID_1|MAC_{(SN1,B)}|\ldots..ID_q|MAC_{(SNq,B)})$$

The q MACs and aggregation report M_{GDj} are sent securely to the base station. Upon receiving an aggregation report, the sink first decrypts the message using the corresponding key $K_{(GDj,B)}$ and checks the report. If no less than q MACs is correct, the event is accepted to be legitimate; otherwise it is discarded. Table 2 shows the overhead of our scheme. We assume the length of ID, key and MAC is 2, 16 and 4 bytes respectively.

Table 2. Overhead of our scheme (Excluding encryption)

Overhead Type	SN	GD
Storage	48 bytes (3 keys)	352 bytes (1+1+Z keys)
Communication	12 bytes (2 ID + 2 MAC)	1220 bytes ((1 ID_{GD} + (1 ID_{sn} + 1 MAC) Z) H hops)
Computation	3 MACs	21 MACs (1 MAC_{GD} + Z MAC_{SN})

4. Security Analysis

We present security analysis of our scheme from various perspectives as described below.

Ordinary sensors (SNs) are captured: A compromised ordinary sensor in a particular group may produce an invalid MAC by providing wrong guarantee for an aggregated report. Our scheme is robust against this kind of attack as long as no more than q sensors within a local group are compromised. Since we devise our scheme where each aggregated report carries q number of MACs from ordinary sensors. However, a compromised ordinary sensor may forge false event's information with valid MAC to its GD.

A GD is captured: When a GD is captured, it may fabricate a report. But to do that, at least q MACs need to be forged. The probability that at least q out of Z MACs is correct is given by

$$p_{GD} = \sum_{j=q}^{Z} \binom{Z}{q} p^{q} (1 - p)^{Z-q}$$

where, p = $1/2^L$ and L is the MAC size in bits. We claim that this probability is negligible for a MAC size of 4-bytes; moreover, only one group out of entire network is in fact affected while other groups are not hampered.

Both GD and ordinary SNs are captured: We consider the situation where an adversary has compromised GD and some number x ($0 \leq x \leq q$) ordinary sensor nodes. To inject a false report, GD needs at least q valid MACs. Since GD has to forge ($q-x$) more MACs, the probability that ($q-x$) out of ($Z-x$) is valid is given by

$$p_{GD}^{x} = \sum_{j=q-x}^{Z-x} \binom{Z-x}{j} p^{j} (1 - p)^{Z-x-j}$$

Again, this probability is almost negligible. If an individual key is compromised, the attacker at best could send false report to the GD but when any message from the GD comes encrypted with the group key, it cannot decrypt it. So, for successful attack, it needs both the individual and group key at the same time. Hence, it definitely ensures resilience of the network, as compromising one key doesn't affect the rest of the keys used among other nodes and links in the network.

Network-wide Compromise of SNs and GDs: Let us suppose that an adversary has captured Q SNs from entire network (i.e. Q out of N). We use the properties of Poisson Process [13] to analyze the robustness of our scheme. We consider the case where number of compromised SNs Q is much less than total number of SNs N (i. e. Q « N). We denote the status of any sensor in a particular group as C_i, $i = 1, 2, \ldots, Z$ and consider it as Bernoulli random variable. Let $C_i = 1$, if the i^{th} sensor is captured and $C_i = 0$, if not captured. Considering the Q captured SNs are uniformly distributed across the network, probability of any compromised SN can be given by $p[C_i = 1] = Q / N, i = 1, 2, \ldots, Z$. If the condition $Q \ll N$ holds, then, $C_1, C_2, \ldots, C_Z$ are said to be independent according to the properties of Poisson Process [13] and the number of captured sensors in one particular group follows the Poisson distribution with approximated mean value Q/Y. So, if q sensors are captured in a group, we get the probability as follows

$$p[q] \approx e^{-Q/Y} \frac{(Q/Y)^q}{q!}.$$

Now, if the total number of groups is Y, we calculate the expected value of number of groups Y having q sensors captured as

$$E[Y_q] \approx Y.e^{-Q/Y} \frac{(Q/Y)^q}{q!}.$$

Next, considering the case where an adversary has captured some groups having the x ($x \geq q$) SNs captured and we call this situation as group capture. Let X be the number of completely captured groups from entire networks. We can compute expected value $E[X]$ by the following equation:

$$E[X] \approx \sum_{q=x}^{Z} Y_q \approx \sum_{q=x}^{Z} Y.e^{-Q/Y} \frac{(Q/Y)^q}{q!}.$$

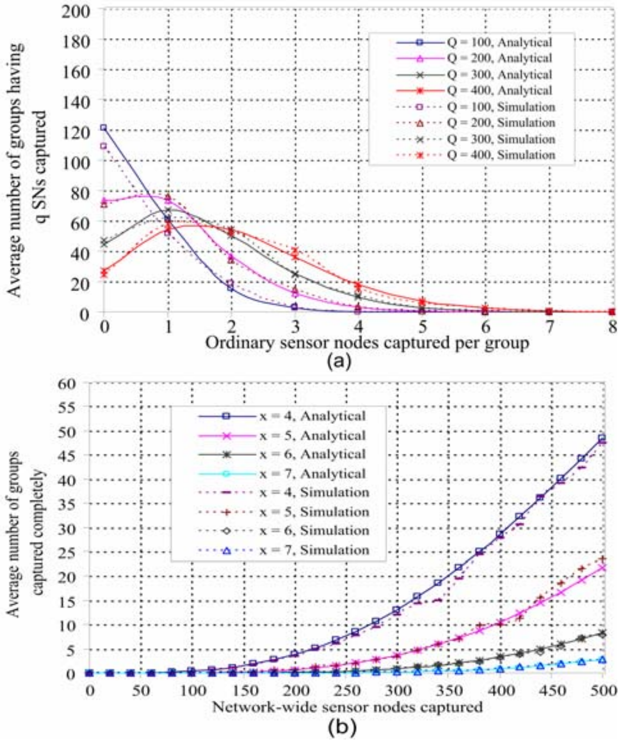


Fig 3. Network performance (N = 2000, Y = 100, Z = 20). (a) Av erage number of groups in which q SNs are captured. (b) Average number of groups that are completely captured when Q SNs are c ompromised in entire network.

Robustness of our scheme against node compromise has been plotted in Fig. 3 using the analytical result calculated from Poisson Process as well as the results obtained by simulation. Analytical and simulated results are almost similar as can seen in Fig. 3. We consider 100 group dominators (GD) and under each GD there are 20 ordinary sensor nodes (SN). So, total number of ordinary sensor nodes are N = 2000 for the entire network. Fig. 3a shows the average number of groups where q ($0 \leq q \leq Z$) ordinary SNs are compromised by the attacker in each individual group. *37* groups having *0* and *18* groups having *2* SNs compromised when $Q = 100$ SNs are captured from the entire network. Worst-case scenario in our approach, when *400* SNs are capture network-wide, only *18* groups out of *100* having *4* SNs captured. Fig. 3b shows the average number of groups that are fully captured when SNs are captured crosses the threshold value q ($q \leq x \leq Z$). For the values x equals *4*, *5*, *6* and *7*, we realize that for $x = 4$, *73%* groups are fully captured when *25%* SNs are captured in entire network. But, as the value of x increases, its robustness gets stronger. For example, $x = 7$, *23%* (*23* out of *100*) groups are captured for *25%* (*500* out of *2000*) network-wide captured SNs. So, significant improvement in network robustness can be achieved by increasing the value of x.

5. Related Works and Comparison

Extensive effort has been put so far to develop security protocol for wireless sensor networks. In [4] a defense against Sybil attack is presented using the random key predistribution and position verification and code attestation mechanism is used to detect Sybil entity. Using master key to generate other keys and HMAC authentication are used to defend Wormhole attack in TIK [5] protocol. Random key predistribution techniques are presented in [6] to defend data and information spoofing, attacks in information transit. REWARD [7] is a good scheme that protects Black hole attacks using (no cryptographic keys) broadcast messages to identify malicious node and suspicious area. SPINS [8] protocol is devised to provide semantic security, data authentication, replay protection and weak freshness while using low communication overhead. In SPINS [8], each node is given a master key and all other keys are derived from this master key. Global key pool is used in [3] and each node is randomly assigned m number of keys from one of the partitions of the global pool. All these protocols described here use traditional wireless network.

In our scheme, intuitively, WCDS will, in general, be smaller than connected dominating sets and the resulting induced graph will have smaller edges. This corresponds to fewer clusters and a sparser abstracted network. For comparison, Statistical En-Route Filtering [3] scheme, each intermediate forwarding node verifies one MAC and five hash computations (for Bloom filters) probabilistically if it has one of the keys in common. In our scheme, only the GDs forward the aggregated report, but they don't perform this intermediate checking. SEF [3] is constrained by sensor's storage since to increase one hop detection probability; the number of keys a sensor stores should be large. But in our scheme only *3* keys are required. SEF performs better when the number of hops a packet travels is very high, but our scheme has much smaller hops and the overhead on forwarding aggregated reports gets to the powerful GDs only. Each report is about *15* bytes long in

SEF scheme and communication overhead is (*15.h*) bytes, where the number of hops *h*, a report travels is very high (necessary for better performance).

6. Conclusions

A distributed architecture consisting of relatively powerful node each servicing a number of less powerful sensor nodes in a star-like configuration is proposed. The powerful sensor nodes assumed to for a connected communications graph. We evaluate our scheme through analysis and simulation to show that our mechanism is resilient to an increasing number of compromised nodes. For further investigation, some important research issues are worth mentioning: a) to investigate and design the model for group based deployment. b) how to optimize the value *Y* (number of GDs) and *Z* (number of SNs under one GD) to achieve better performance.

References

[1] Karlof, C. and Wagner, D., *Secure routing in wireless sensor networks: Attacks and countermeasures*, Elsevier's Ad Hoc Network Journal, Special Issue on Sensor Network Applications and Protocols, September 2003, pp. 293-315.

2. C.K. Wong, M.Gouda, and S.S.Lam., *Secure Group Communications using Key Graphs,* IEEE/ACM Transactions on Networking, vol. 8, no. 1, February 2000.
3. Ye, F., Luo, H., Lu, S, and Zhang, L, *Statistical en-route filtering of injected false data in sensor networks*, IEEE Journal on Selected Areas in Communications, Volume 23, Issue 4, April 2005, pp. 839 – 850.
4. Newsome, J., Shi, E., Song, D, and Perrig, A, *The sybil attack in sensor networks: analysis & defenses*, Proc. of the third international symposium on Information processing in sensor networks, ACM, 2004, pp. 259 – 268.
5. Hu, Y.-C., Perrig, A., and Johnson, D.B., *Packet leashes: a defense against wormhole attacks in wireless networks*, Twenty-Second Annual Joint Conference of the IEEE Computer and Communications Societies. IEEE INFOCOM 2003, Vol. 3, 30 March-3 April 2003, pp. 1976 – 1986.
6. Du, W., Deng, J., Han, Y. S., and Varshney, P. K., *A pairwise key pre-distribution scheme for wireless sensor networks*, Proc. of the 10th ACM conference on Computer and communications security, 2003, pp. 42-51.
7. Karakehayov, Z., *Using REWARD to detect team black-hole attacks in wireless sensor networks*, in Workshop on Real-World Wireless Sensor Networks (REALWSN'05), 20-21 June, 2005, Stockholm, Sweden.
8. Perrig, A., Szewczyk, R., Wen, V., Culler, D., and Tygar, J. D., *SPINS: Security Protocols for Sensor Networks*, Wireless Networks, vol. 8, no. 5, 2002, pp. 521-534.
9. A. Menezes, P. Oorschot, S. Vanstone, *Handbook of Applied Cryptography*, CRC Press, Boca Raton, 1996.
10. Y.P. Chen and A. L. Liestman, *A Zonal Algorithm for Clustering Ad Hoc Networks*, International Journal of Foundations of Computer Science. 14(2):305-322, April 2003.
11. J. P. Walters, Z. Liang, W. Shi, and V. Chaudhary, *Wireless Sensor Network Security: A Survey*, 2006 Auerbach Publications, CRC Press.
12. Douglas B. West, *Introduction to Graph Theory*, August 23, 2000, 2nd Edition.
13. S. M. Ross, *Introduction to Probability Models*, Academic Press, 2003, 8th Edition.