Weakly connected dominating set-based secure clustering and operation in distributed sensor networks

Al-Sakib Khan Pathan and
Choong Seon Hong*

Department of Computer Engineering,
Kyung Hee University,
1 Seocheon, Giheung, Yongin,
Gyeonggi, 446-701 Korea
Fax: +82 31 204-9082
E-mail: spathan@networking.khu.ac.kr
E-mail: sakib.pathan@gmail.com
E-mail: cshong@khu.ac.kr
*Corresponding author

Abstract: This paper presents an efficient approach to secure network formation in distributed sensor networks (DSNs), which could be used for secure communications among the nodes after the start of the network's operation. The structure of the network is formed on the basis of offline rank assignments by pre-distribution of secret keys to the participating sensors. Our approach uses the notion of weakly connected dominating set (WCDS) to reduce the number of cluster-heads in the network for greater resource efficiency. The clusters in the network are formed in a secure and efficient way so that no hostile entity could be included in any cluster during network structuring process. We propose an efficient algorithm to form a network-wide secure WCDS, which includes an optional re-keying mechanism to increase the level of security. Our objective is to ensure security in distributed sensor network from the bootstrapping stage up to the beginning of its operation. Along with the description of our approach, we present detailed analysis, simulation results and comparisons of our approach with other related schemes.

Keywords: cluster; graph; operation; security; weakly connected dominating set; WCDS.

Reference to this paper should be made as follows: Pathan, A-S.K. and Hong, C.S. (2009) 'Weakly connected dominating set-based secure clustering and operation in distributed sensor networks', *Int. J. Communication Networks and Distributed Systems*, Vol. 3, No. 2, pp.175–195.

Biographical notes: Al-Sakib Khan Pathan received his PhD from Networking Lab, Department of Computer Engineering in Kyung Hee University, South Korea in 2009 and his BSc in Computer Science and Information Technology from Islamic University of Technology (IUT), Bangladesh in 2003. Before joining Networking Lab, he was with the CSE Laboratories, North South University, Bangladesh. He is actively involved in various research activities and received some awards and recognitions for his works. His research interests include wireless networking, wireless sensor networks, network security and e-services technologies.

Choong Seon Hong received his BS and MS in Electronic Engineering from Kyung Hee University, Seoul, Korea, in 1983 and 1985, respectively and his PhD from Keio University, Japan in 1997. He worked for the Telecommunications Network Lab, KT as a Senior Member of technical staff and as a Director of the Networking Research Team until August 1999. Since September 1999, he has been working as a Professor at the School of Electronics and Information, KHU. His research interests include ad hoc networks, network security and network management. He is a member of IEEE, IPSJ, KIPS, KICS and KIISE.

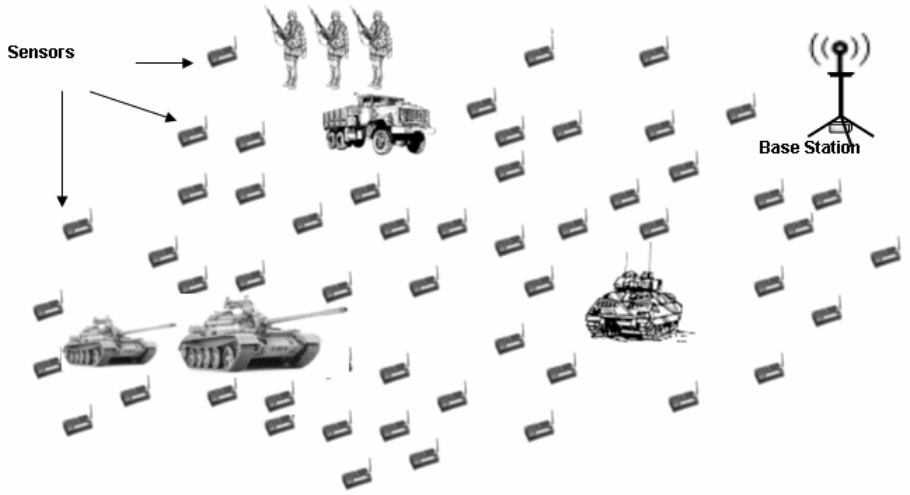
1 Introduction

The nature of services expected from wireless sensor networks often demands the utilisation of efficient security mechanisms. While the method of routing of data in the network should have some associated security mechanisms, the initial formation of the network should also be done securely so that any harmful attempt by potential adversaries can be restrained. Security should be given priority from the very early state of the network's formation and operation. This issue is of one of the prime concerns for distributed sensor networks as these types of networks are envisaged to operate in the presence of adverse or hostile entities.

Distributed sensor network (DSN) is basically a wireless sensor network (WSN) with a large number of sensors and a large coverage area. It differs from the traditional wireless sensor network in the sense that it contains considerably huge number of sensors which are intended to be deployed over hostile and hazardous areas. In such areas, the communications among the sensors could be monitored by outside entities, the sensors are under constant threat of being captured by the enemy or can be manipulated by the adversaries. A DSN is dynamic in nature in the sense that new sensors can be added or deleted whenever necessary (Carman et al., 2000). Basically, these types of networks are suitable for covering large areas for monitoring, target tracking, surveillance and moving object detection which are very crucial tasks in many military and public-oriented operations.

Let us consider a military reconnaissance scenario. Surveillance and detection of enemy vehicle movements are very crucial tasks in such a case. In order to detect the positions and movements of the enemy units, wireless sensors could effectively be used. The acoustic, magnetic, pressure or other types of signals from the sensors could be accumulated in the sink or group gateways. For the fidelity of the reported data, it is needed to check the sensing outputs of a certain number of sensors (which depends on the application at hand).

Figure 1 A sample application scenario (battlefield) of DSN



In Figure 1 we show an example scenario where hundreds of sensors are dispersed over the area of interest (AOI). Here, accuracy, timeliness and fidelity of sensed data are very crucial. Inclusion of any rogue entity in such a network during its formation might be extremely harmful for the network. If an adversary can thwart the work of the network by perturbing the generated information, stopping production or pilfering information, then the usefulness of the network is drastically curtailed. In fact, if the data are wrong or somehow altered by any active saboteur, that can instigate the command centre to take disastrous military decisions. Therefore, it is necessary to ensure that no hostile entity can be included in the network during its bootstrapping phase. Considering such types of critical scenarios, our approach tries to solve the issue of secure formation of DSN for its effective use. However, it should be noted that dealing with physical security issues is beyond the scope of this paper (it can be ensured by camouflaging or other tactics of the sensors in the battlefields).

In this paper, we present an efficient approach of key pre-distribution among the sensors which helps for offline rank assignments of the sensors and eventually plays the crucial role to form a network-wide weakly connected dominating set. Our target is to ensure security in the network and to minimise the number of cluster heads while forming the clusters, so that a relatively small number of cluster heads can securely cover the whole network. We assume that the base station is fully secure and the adversaries can not compromise the base station in any way. The scope of this paper is restricted to the method of secure bootstrapping/clustering and starting of the operation of DSN. Our analysis and performance evaluation show that our approach performs well to form secure clusters in a distributed sensor network with minimum number of cluster heads and the formed network can be used for secure data transmissions within the network.

The rest of this paper is organised as follows: Section 2 notes down some related works that motivated us to propose our approach, Section 3 presents our network model and preliminaries, Section 4 contains the detailed description of our approach, Section 5 presents our performance analysis and simulation results and Section 6 concludes the paper delineating the achievements from our work with future research directions.

2 Motivation and related works

We define a cluster as a subset of the total set of sensors in a network that might contain at least one cluster head which is capable of manipulating sensed data locally and of sending the gist of that to base station. Grouping nodes into clusters is a good idea as it divides the network into several distinct but interrelated and manageable regions. It can also be helpful for efficient routing in the network.

Some of the previous works have addressed the issue of clustering or group formation in sensor networks. But, most of the previous works on clustering in wireless sensor networks have not addressed the security issues from the beginning stage of network's formation or considered a secure environment for bootstrapping of the network. We argue that in many cases such an ideal environment might not exist. For example, if sensor networks are profoundly used in the military reconnaissance scenarios, both sides might try to take benefit from the technology and while forming the friendly network, there could be a hidden and active enemy sensor network present in prior in the target area. If the friendly network is to be formed later than the enemy network in that area, the hostile sensors might actively try to participate in the clustering process or can try to hinder the formation of any other network within that particular region. This type of problem can also arise for other applications of DSN which need commensurate security protections.

We have investigated a number of prior works which consider the network structuring with clustering mechanisms. But unfortunately, most of them overlook the issue of secure formation of the network. Here we note down some of those works that have motivated us to devise our approach. Nowak (2003) presents a distributed expectation-maximisation (EM) algorithm suitable for clustering and density estimation in sensor networks. Works like Halgamuge et al. (2003), Lee et al. (2004), Younis and Fahmy (2004), Ye et al. (2005) and Liu and Lin (2003) address only energy-aware clustering. Gupta and Younis (2003) propose a load-balanced clustering scheme which increases the lifetime of the network. Other works on clustering in sensor networks are Tzevelekas et al. (2005), Wokoma et al. (2003), Banerjee and Khuller (2001), etc. In fact, very few works address the bootstrapping (with clustering) issues in wireless sensor networks considering prior presence of hostile entities in the target area.

Mathew et al. (2005) propose a new approach termed slotted sensor bootstrapping (SSB) protocol, which focuses on avoiding collisions in the bootstrapping phase but it ignores the security threats present in the beginning state of sensor networks. Prasad and Alam (2006) analyse the security issues, threats, attacks and requirements for wireless sensor networks. This paper also proposes security framework and security architecture to integrate existing technologies with WSN technology. The authors mention the bootstrapping issues; however, they do not provide any specific solution how to bootstrap the network with secure formation of the network, rather address general security architecture and discuss the considerable aspects for security in WSN.

One of the works addressing secure clustering is Bohge and Trappe (2003), in which the authors consider a hierarchical sensor network with three-tier topology and propose an authentication framework to check the validity of newly joining nodes in the network. They focus on achieving two goals; firstly, to ensure that the data received by the application is sent by an approved sensor node; secondly, to verify that the information has not been modified on its way to the application. Their authentication service authenticates incoming nodes, establishes shared secrets among them and with the

application, keeps track of changes in the network topology and provides data origin authentication for sensor node data.

Ferreira et al. (2005) consider the network model used in LEACH (Heinzelman et al., 2000) and add a security scheme taking the ideas from SPINS as its building blocks. In this work, though the authors have tried to add security to the cluster-based network model, their basic network model LEACH is basically inefficient because of its unrealistic assumptions. LEACH was proposed in the early days in this research field considering that the sensor nodes in the network can directly communicate with the base station using long transmission ranges (if necessary). Also, it considers a novel type of routing that randomly rotates routing nodes among all nodes in the network. This is in fact not a desirable case for wireless sensor networks especially when the resources of the tiny devices are scarce.

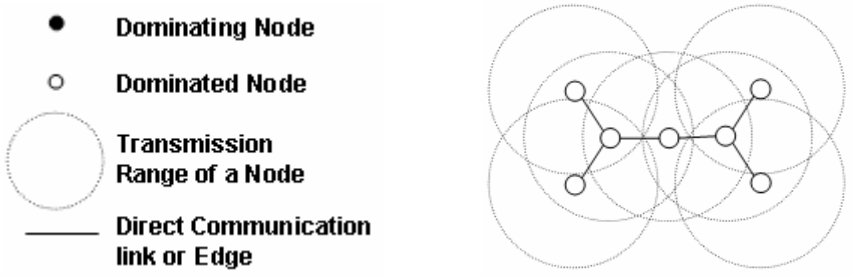
Recently another similar work like Ferreira et al. (2005) is proposed in Oliveira et al. (2006). This work presents SecLEACH which is a protocol for securing none-to-node communication in LEACH-based networks. The authors show how random key pre-distribution, which basically is studied in the context of flat networks, could be used for secure communication in cluster-based network models like that is used in LEACH.

Our work is different than all of the mentioned clustering approaches as we use the notion of WCDS considering the whole network as a graph. Our key pre-distribution helps for offline rank assignments of the sensors and eventually plays the main role to create a network-wide secure WCDS. It should be mentioned that the details of key generation and selection or prevention of denial-of-service (DoS) attacks caused by attacks like jamming (Wood et al., 2003) or other physical layer attacks are beyond the scope of this paper.

3 Our network model, assumptions and preliminaries

We consider the topology of the whole distributed sensor network as a unit-disk graph (UDG) (Clark et al., 1990), G = (*V*,*E*) (see Figure 2), where *V* is the set of sensors (vertices) in the network and *E* is the set of direct communication links (edges) between any two sensors.

Figure 2 (a) legend used in the paper (b) UDG

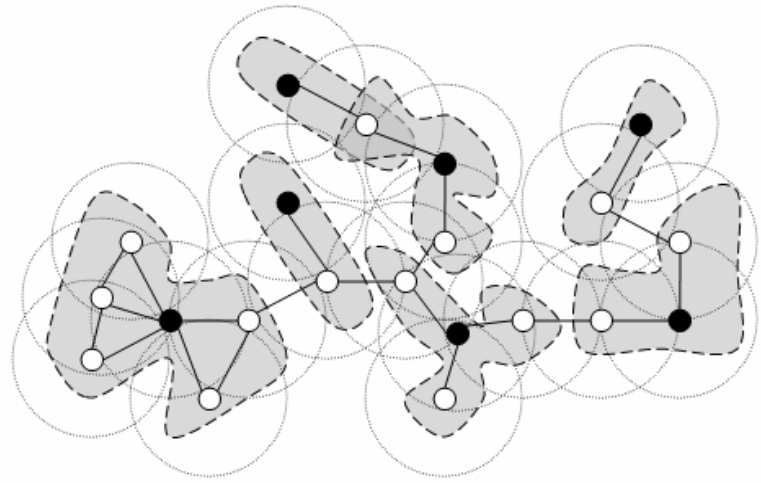


(a) (b)

Definition 1. A dominating set S is a subset of the vertex set V of a graph G = (V,E) (i.e., $S \subseteq V$), so that all other vertices in the graph are adjacent to the vertices of S. For a dominating set S, $N_G[S]$ = V, where $N_G[S]$ is the set of vertices including the vertices in S and the vertices adjacent to a vertex of S (Figure 3). In this case, each of the nodes in the set S is called a 'dominator' (or 'dominating node') and all other sub-ordinate nodes under it are called 'dominated nodes'. However, finding a minimum-size-dominating-set in a general graph is NP-complete (Garey and Johnson, 1979).

Definition 2. A connected dominating set (CDS), S_C is a dominating set of a given graph G = (V,E) where the induced subgraph of S_C is connected. Figure 4(a) shows the connected dominating set for our graph model (i.e., all the black vertices). The connected dominating set for any type of ad hoc network can be used for efficient routing or message transmission throughout the network. However, for CDS, a large number of dominating nodes are needed for maintaining the connectivity requirements of the network.

Figure 3 Dominating set consisting of black vertices



Definition 3. A WCDS, SW is a dominating set where the graph induced by the stars of the vertices in SW is connected. A star of a vertex is comprised of the vertex itself and all the vertices adjacent to it [all the black nodes in Figure 4(b)]. For any given graph,

$$|WCDS| \leq |CDS| \tag{1}$$

where, |.| denotes the size of the set. So, in case of WCDS, comparatively less number of dominating nodes is needed for establishing network-wide connectivity than that is required for CDS. For example, in Figure 4, $|WCDS|$ = 8 whereas, $|CDS|$ = 13. This difference increases as the size (number of nodes/vertices in the network) of the network increases.

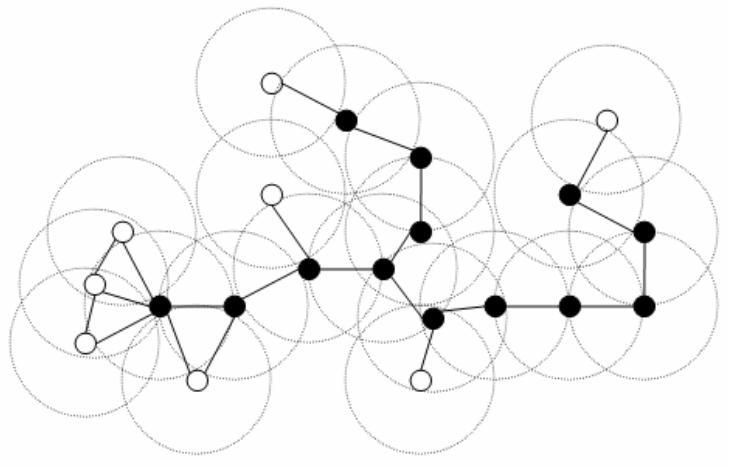
The weakly connected dominating set underpins our scheme (Pathan and Hong, 2006). In fact, it could be noticed from Figure 4(b) that each dominating node (or vertex) in the weakly connected dominating set is at the centre of a star structure (dominated nodes in each star are shown in white colour). Thus for each dominating node in a WCDS

of the overall network, we have one star where all the other nodes in the star are just one hop apart [Figure 5(a)]. Also it can be observed that; between two stars there is at least one common dominated node which can be used for the communication purpose between two adjacent stars. We term such a common dominated node between two distinct stars as a 'mediator' [Figure 5(b)].

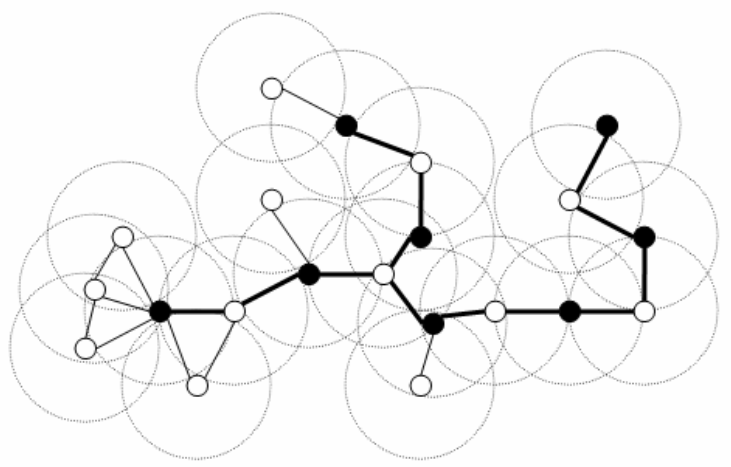
Assumption 1. Once the sensors are deployed, they remain relatively static in their respective positions.

Assumption 2. In a unit disk or transmission range of a sensor, all the neighbouring sensors do not necessarily have direct communication links among themselves. If two nodes i and j have a direct communication link, it is bidirectional $\forall_{i,j},(i,j) \in E \Rightarrow (j,i) \in E$ and it exists if and only if i and j have a common secret key.

Figure 4 (a) CDS (b) WCDS

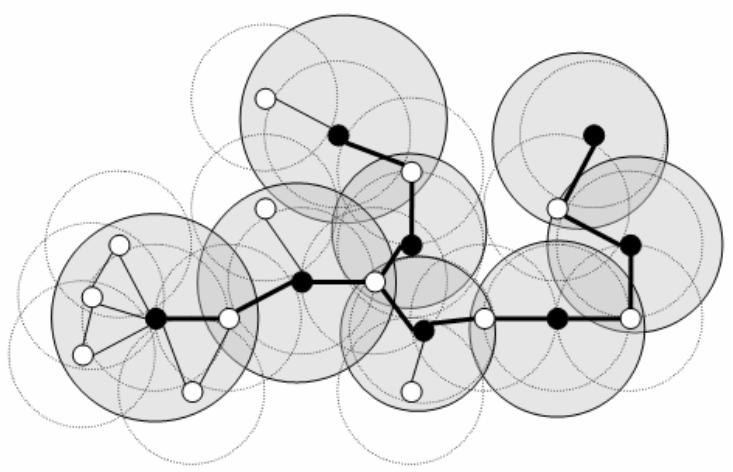


(a)

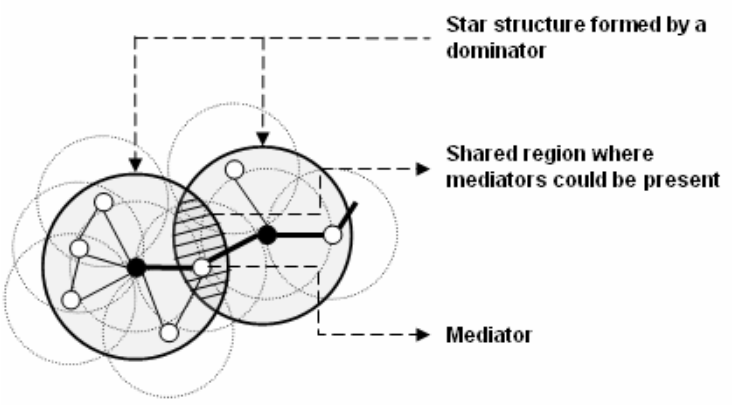


(b)

Figure 5 (a) stars (disks) formed by each dominator in WCDS (b) a mediator between two stars



(a)



(b)

4 Our proposed approach

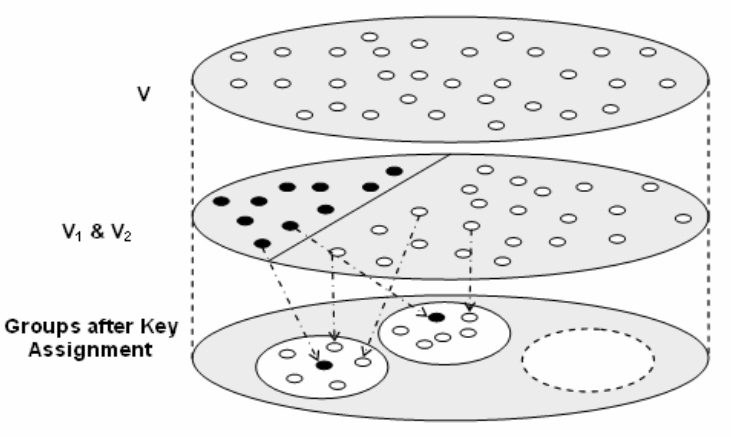
In this section we describe the details of our approach. We first apply two stage operations for secure formation of clusters in the network.

4.1 Offline rank assignments with key pre-distribution

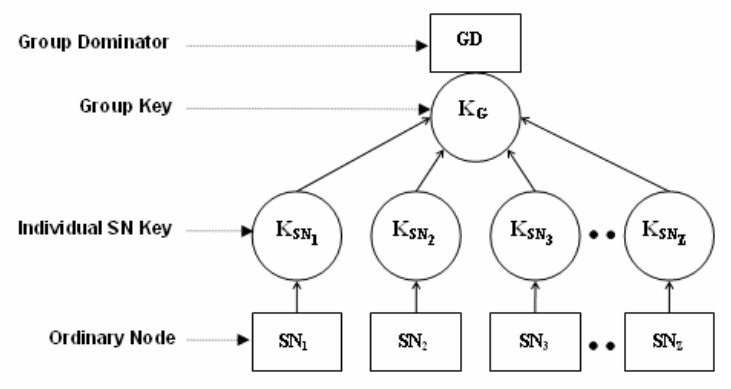
There are mainly three types of entities in the underlying network that hold the secret keys for the communications among the nodes: base station (BS), group dominator (GD), and ordinary sensor node (SN). Some of the SNs play the roles of mediators. Before describing the offline rank assignment, we note the following terms that will be used throughout the rest of the paper:

- *Group key (K_G)*: this key is shared by all the sensor nodes in a particular group (including the dominator).
- *Individual Key (K_{SN_i}):* each SN and corresponding GD pair holds a unique common individual key.
- *Group dominator (GD):* contains the K_G and all the individual keys for its own group. It acts like a data manipulator and manager of a group for key manipulation (if needed).
- *Ordinary Sensor Node (SN_i):* SN_i is an ordinary sensor node. Each SN_i holds a K_G and an K_{SN_i} shared with the corresponding GD.

Figure 6 (a) ranking of the sensors based on key pre-distribution (b) logical pre-distribution of secret keys using star key graph



(a)



(b)

The sensors in the network are assigned their ranks based on offline key pre-distribution. For this, we divide the whole set of sensors V into two subsets, V_1 and V_2, where V_1 contains the probable GD or cluster heads and V_2 consists of ordinary sensors (SNs). The set V_2 is further divided into several subsets $w_i \subset V_2$, i=1, 2, 3,….,N and N is the maximum number of possible proper subsets of V_2. Each w_i is assigned to one element in the set of V_1. The sensors in the subset w_i (SN_1, SN_2,….SNs_η) and corresponding one sensor from V_1 (let, GD_i, i = 1) are taken for group-wise key-predistribution [Figure 6(a)]. Each of the sensors in the set w_i is assigned two keys: K_G and its K_{SN_i}. The K_G is common for all the SNs in a w_i but the K_{SN_i} is shared by the particular sensor and the GD_i. The GD_i contains all the individual keys of the sensors in its w_i and its own K_G. The star key graph in Figure 6(b) shows the hierarchical ordering of the keys for the sensors in the network.

Assumption 3. All the transmission ranges of the sensors are same. Transmission of each node is isotropic (i.e., in all directions) within its transmission range so that each message sent is a local broadcast.

Assumption 4. The BS contains all the individual keys and K_Gs used in the network.

Assumption 5. The number of SNs (value of η) in each group is decided on demand. It can be group specific or can be set to a common value for all of the groups in the network. η is actually the maximum degree ($\Delta(GD_i)$) of a GD in a group.

4.2 Post deployment cluster formation

The groups of sensors are deployed over the target region with one group at a time. After deployment, each SN tries to find out its own GD within its transmission range by sending a join request packet encrypted with its K_{SN_i}. The corresponding GD in turn sends the join approval message encrypted with the K_G. In both cases, both the GD and the SN can decrypt the messages and form the group without disclosing any secret. In some cases, the corresponding GD of an SN might not be within its one-hop transmission range (disk). Such a SN is called an ‘orphan’.

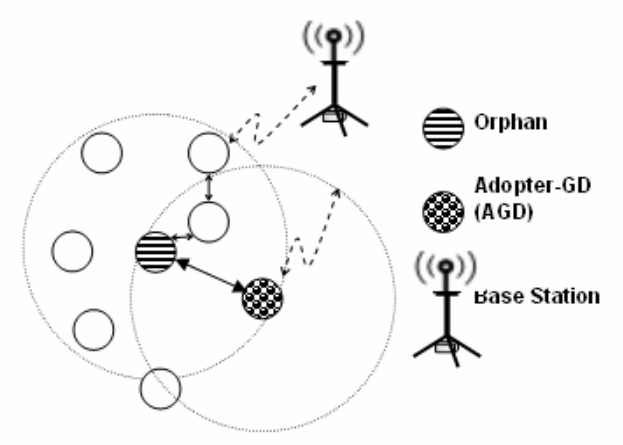
Two cases might occur in such a situation:

Case I *The orphan has other dominator(s) of other group(s) within its transmission range*. In this case, the SN detects the presence of other GDs of other groups in its surroundings, collects their ids and sends an error message to the BS with this information. The GDs within its one-hop transmission range can also detect such erroneous SN and report to the BS. The BS in turn assigns one of the neighbouring GDs as the adopter (AGD or adopter GD) of the orphan SN.

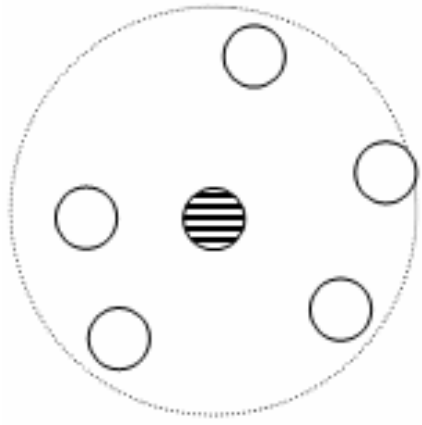
Case II *No other dominator is present within its transmission range*. This is the worst case. In this case, The BS assigns the rank of a GD (let us call it GDSN) to that particular SN, though it does not contain any other sub-ordinate (i.e., dominated) sensors.

For both the cases, corresponding pre-distributed keys (SNs use individual keys and GDs use K_Gs) are used for the communications. Figure 7 illustrates the two cases.

Figure 7 (a) Case I: an orphan with a neighbouring dominating node (GD) of another group (b) Case II: an orphan with no dominating neighbour



(a)



(b)

Note: In this case, the rank of a GD is assigned to that stranded node

Creation of mediators: an SN which gets its own GD and another GD of another group in its one-hop transmission range sets itself as a mediator. The neighbouring GDs also can detect the presence of sensors of another group within its own disk (i.e., transmission range) and note down the ids of these sensors as potential mediators. These mediators act as communication gateways between two separate groups or clusters. As stated earlier, all the stars thus shaped can use mediators for the inter-group (inter-star or inter-cluster) communications [see Figure 5(b)].

In this way, eventually the resultant logical model of the whole network contains a secure weakly connected dominating set where the GDs of the logical groups (stars) are dominating nodes and all other nodes in the network are dominated by dominators. This logical model now can be used for secure message delivery within the network (using the secret keys). The pseudo code for secure cluster formation algorithm is presented in Figure 8.

Figure 8 Pseudo code for secure clustering algorithm

```
Let, V = (VSN ⋃ V_GD)
each sensor s in the set V_SN
    broadcast JOIN_REQ locally
    if group-key-encrypted JOIN-APRV message from any node in V_GD and within hop = 1
        establish a link with g
        mark g as the dominator
    else
        flood encrypted GD_ERR towards BS
    end if
    if unknown-group-key-encrypted JOIN_APRV from any node in V_GD and hop=1
        mark g as the neighbour_dominator
    end if
each g in V_GD
    if ind-key-encrypted JOIN_REQ from any sensor s in V_SN
        send group-key encrypted JOIN_APRV
        establish a link with s
        mark s as sub-ordinate
    end if
if unknown-ind-key-encrypted JOIN_REQ from any s in V_SN
        mark s as mediator
    end if
    if ind-key-encrypted GD_ERR from any s in V_SN and hop = 1
        report group-key-encrypted ORP_ERR message to BS
    end if
#For the BS:
if ind-key-encrypted GD_ERR from any s in V_SN and group-key-encrypted ORP_ERR from any
g in V_GD
        if same s is reported in both messages: set the reporting g as the adopter_GD of s
```

4.3 Authenticated message delivery

Once the network is logically covered with a secure WCDS, the sensory data from the sensors could be transmitted securely to the base station. GDs are responsible to aggregate data collected from different sensors. As the sensors in the same group (or cluster) cover more or less the same part of the target area, for fidelity and correctness of data, we consider that; if there are η number of ordinary sensors (SN) in a group, the corresponding GD waits for the same (or almost same) sensing reports from at least τ ($\tau \leq \eta$) number of SNs, where τ is the threshold value set for that particular group or the network. The value of τ and η depend on the settings and the situation at hand. The timeliness of the sensing reports is also crucial; hence within a threshold time h, if the

GD does not get at least τ same reports, it simply discards the report as it could be some falsely injected data from the adversary(s).

Each SN in a cluster prepares a sensed report with a message authentication code (MAC) of the message using its K_{SN_i} (that is shared with the corresponding GD) as $Msg_i = MAC(K_{SN_i}, nonce \parallel E_v)$, where *nonce* is the time of detection, E_v is the event sensed by SN. The SN then sends key index *i* and Msg_i to GD. GD checks the MACs of at least τ number of messages and then sends the final report to BS, encrypted with the K_G of that particular group. The mediators help for delivering the authenticated data to the base station. In fact, when the mediators get messages (encrypted with particular K_G) they do not decrypt rather only forward those to another GD. Note that, for each of these communications all the sensors use their local transmission ranges. In this way, the aggregated authenticated data securely reaches to the BS. RC5 can be used to calculate the MACs. RC5 is a symmetric block cipher designed to be suitable for both software and hardware implementation. It is a parameterised algorithm, with a variable block size, a variable number of rounds and a variable length of key. This provides the opportunity for greater flexibility in both performance characteristics and the level of security (Rhee, 2003).

4.4 Optional re-keying mechanism

In this subsection, we describe the optional re-keying mechanism which can be employed if the level of security needs to be increased for the network. All the groups of sensors can be used at a time or more groups can be used later based on demand. If it is needed, some sensors in a group could be deployed later. During the offline key pre-distribution, all the nodes are assigned the keys but all the nodes might not be deployed. When any of those remaining nodes is newly deployed, it follows the procedure of joining a group. If it is authorised by the access list of GD, it joins the group. Otherwise, GD forwards the id of this sensor to the BS. BS informs GD about the K_{SN_i} of that particular SN if it is a legitimate node. If the SN is authenticated by BS, GD generates a new K_G and encrypts the new K_G with the newly added node's K_{SN_i} and sends it to that particular SN. All the other nodes in the group know about the change of K_G from a local broadcast by the GD of that group. In this case, the previous K_G is used for encrypting the new K_G. For leaving a group or cluster, a node simply leaves a message to inform the GD which in turn generates a new K_G and multicasts it within the group members. Figure 9 shows the joining and leaving of a node.

As an example, let's suppose SN_4 wants to join the existing group (left) in Figure 9. GD changes the K_G to a new key K_G', and sends the following re-keying messages:

$GD \rightarrow [SN_1, SN_2, SN_3]: E_{KG}(K_G')$, new group key encrypted with the old K_G

$GD \rightarrow SN_4: E_{K_{SN4}}(K_G')$, new group key encrypted with the joining SN's K_{SN_i} known either from GD's previous knowledge or from the BS.

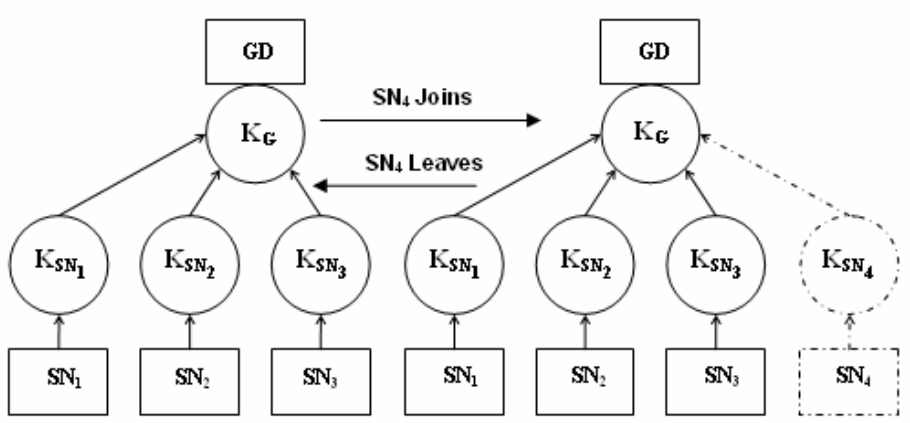
Similarly, when any SN wants to leave any group, it just sends a *leave* message. GD deletes the leaving SN, updates K_G to new K_G' and unicasts a message. So,

$If, SN_4 \rightarrow GD : E_{K_{SN4}}(leave)$, SN_4 wants to leave the group.

$GD \rightarrow SNi : E_{K_{SNi}}(K_G')$, I = 1,2,.. $\eta - 1$, GD unicasts the new K_G encrypted with remaining SNs' individual keys.

As stated earlier, this kind of regeneration and updating of keys is optional and based on the requirements of the level of security for the network. If not needed, the nodes join or leave the groups using the general procedure without any regeneration of keys.

Figure 9 Optional re-keying feature for join/leave scenario for a local group



5 Performance evaluation and comparison

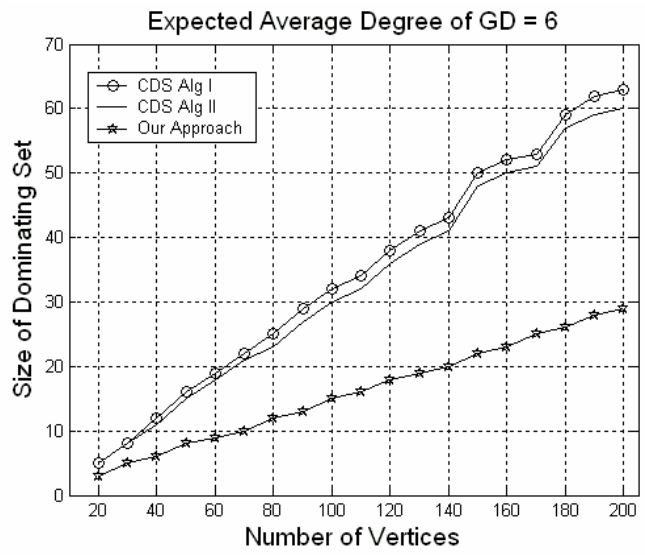
We form a secure WCDS to cover almost all of the nodes in the network with minimum effort. The offline rank assignment reduces the burden of executing resource-hungry operations to form clusters. As shown in equation (1), WCDS requires less number (or equal to) of dominating nodes to cover the whole network than that of a CDS requires. Depending on the requirements we can increase or decrease the value of η (the expected maximum degree, Δ of a GD in a group). In ideal case, the size of the dominating set created in our approach could be obtained by,

$$\text{Size of dominating set} = \frac{\text{Number of vertices in the graph}}{\eta + 1} = \frac{\text{Number of vertices in the graph}}{\Delta(GD) + 1} \quad (2)$$

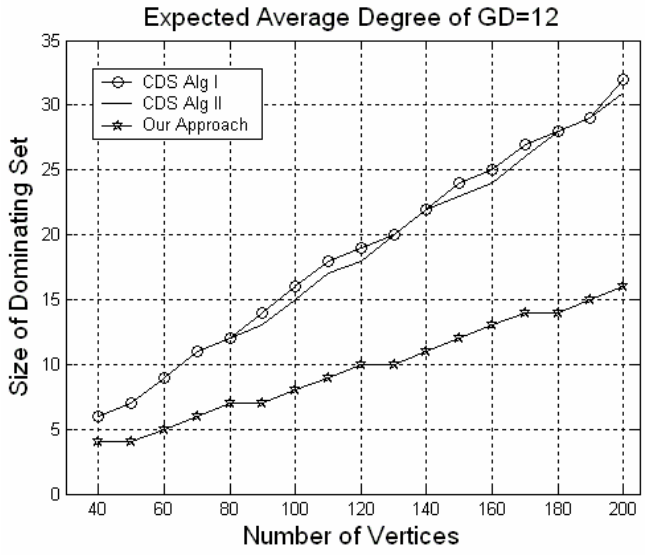
In our simulation, we generated random graphs of 20–200 and 40–200 nodes with expected average degree 6 and 12 respectively. To simulate the structure of the sensor network, we placed the vertices randomly over a 2-D rectangular plane. The network size and density were set by changing the number of vertices and transmission ranges of the nodes. Applying our approach and two algorithms (I and II) of Das and Bharghavan (1997), we have found that our approach generates much smaller number of group

dominators or cluster heads. For a large number of sensors in the network, it works effectively. Figure 10 shows the size of dominating set in comparison with that of Algorithm I and Algorithm II of Das and Bharghavan (1997). One of the major advantages of our approach is the flexibility of setting the value of η according to the requirements of deploying the network.

Figure 10 Number of vertices versus size of the dominating set (a) when expected average degree of GD is 6 (b) when expected average degree is 12



(a)



(b)

We use distinct K_Gs for each of the GDs and distinct individual keys for each SN. So, in general case, the number of distinct keys required for our network depends directly on the number of sensors in the whole network [Figure 11(a)]. Each GD in the network has to remember one K_G and all the individual keys of the SNs of that particular group. So, the storage requirement for each GD in number of bits is,

$$\gamma_{GD} = (\sigma + 1) \times k \tag{3}$$

and for each SN,

$$\gamma_{SN} = 2 \times k \tag{4}$$

where, σ is the average number of SNs in the groups/clusters and k is the minimum number of bits required to represent a key. If total number of sensors in the network is T, $k = ceil\left(\log_2 (T-1)\right)$, as this number of bits is enough to represent the keys for T number of sensors.

Therefore, equations (3) and (4) could be written as,

$$\gamma_{GD} = (\sigma + 1) \times ceil(\log_2 (T-1)) \tag{5}$$

and

$$\gamma_{SN} = 2 \times ceil(\log_2 (T-1)) \tag{6}$$

As the value of σ increases, the storage loads for GDs increase [Figure 11(b)]. Hence, this value is set according to the requirements or a particular situation at hand. If initially we have α number of GDs and β number of SNs, the network-wide storage usage (in bits) for storing the keys is,

$$\begin{aligned} \Gamma_{network-wide} &= \alpha \times \left((\sigma + 1) \times k\right) + \beta \times (2 \times k) \\ &= k \times \left(\alpha \times (\sigma + 1) + 2 \times \beta\right) \\ &= ceil(\log_2 (T-1)) \times \left(\alpha \times (\sigma + 1) + 2 \times \beta\right) \end{aligned} \tag{7}$$

Here, $T = \alpha + \beta$.

After formation of clusters within the network, the mediators are used for communications among clusters. From the higher level view, we can consider the clusters (or groups) as nodes in a random graph, G = (n, p), where n is the number of nodes (i.e., clusters in our case) for which the probability that an edge (i.e., communication link via mediator) exists between two nodes is p. p = 0 when there is no edge and p = 1 when the graph is fully connected. According to Erdös and Rényi (1959), for monotone properties, there exists a value of p such that the property moves from 'nonexistent' to 'certainly true' in a very large random graph. The function defining p is called the threshold function of a property. Given a desired probability P_c for graph connectivity, the threshold function p is defined by,

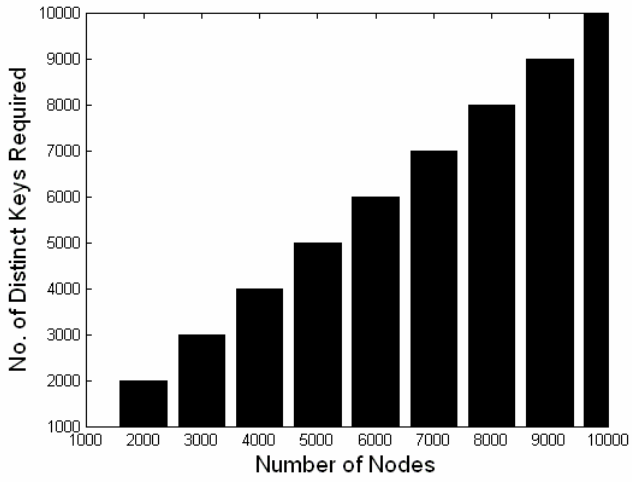
$$P_c = \lim_{n \to \infty} P_r\left[G(n, p) \text{ is connected}\right] = e^{e^{-c}}$$

where, $p = \dfrac{\ln(n) - \ln(-\ln(P_c))}{n}$

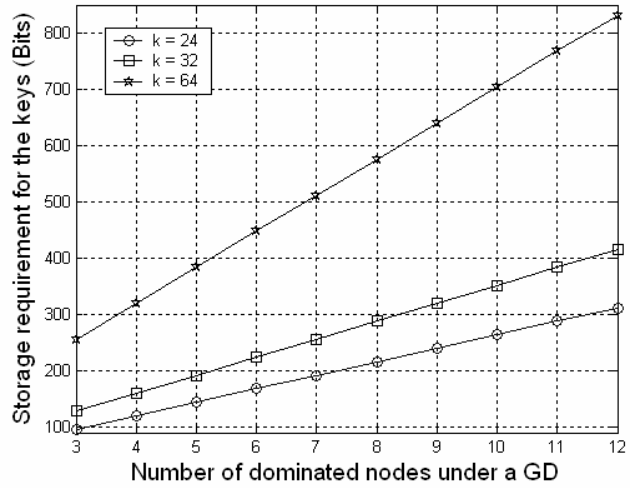
Let, p be the probability that an edge (communication link via mediator) exists between two GDs of two clusters, n be the number of nodes (i.e., clusters/groups in the entire network in this case) and d be the expected degree of each GD, then,

$$\begin{aligned} \mathrm{d} &= \mathrm{p}(\mathrm{n}-1) \\ &= \frac{(n-1)(\ln(n)-\ln(-\ln(P_c)))}{n} \end{aligned} \tag{8}$$

Figure 11 (a) number of distinct keys required to support the size of the network (b) storage requirement for a GD for storing the keys for various values of σ



(a)



(b)

Figure 12 illustrates the plot of the expected degree of a node d, as a function of the network size n (i.e., here the number of clusters or groups), for various values of P_c. The figure shows that the expected degree of a GD needs to be increased by two to increase

the probability that a random graph is connected by one order. Moreover, the curves of the plot are almost flat when n is large, indicating that the size of the network has insignificant impact on the expected degree of a node (here, clusters) required to have a connected graph.

In our approach, the sensors can be added later on, rather than deploying all of them at a time. Sometimes the entire terrain information and deployment diagram might be available (consider a battlefield scenario where the sensors are deployed prior to the enemy forces' invasion). In this case, the extra sensors could be deployed within the range of their appropriate group or cluster. If the sensors are deployed randomly, in the worst case, all the newly added sensors might not be within the ranges of their intended group dominators and even no other GD might be available in their surroundings. Hence, in the worst case, all the newly added sensors would be included in the dominating set which would increase the size of the dominating set. Still it would be less than the number of dominators needed in case of a CDS in a DSN.

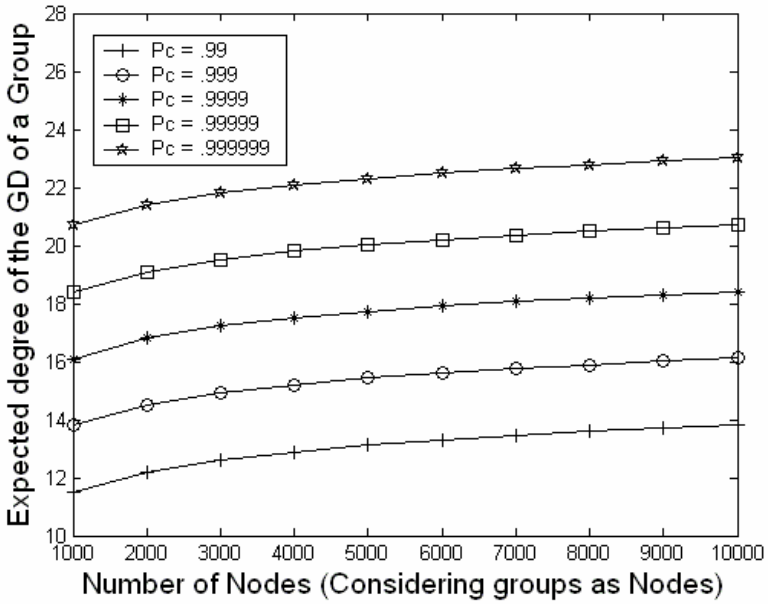
Our scheme ensures that each of the GDs and the corresponding SNs can directly form the groups (i.e., clusters) maintaining the security of the network from the bootstrapping state. As encryption is used for message-transmission within the network from the very beginning of the network's formation, our scheme reduces the chance of hello flood attack or most of the other sorts of attacks against sensor networks (readers can go through Karlof and Wagner (2003) and Pathan et al. (2006) for types of attacks in wireless sensor networks). Again, as each node carries distinct individual and K_Gs, compromising one node affects only one link in the network while other links remain safe from the attacks by the adversaries. If the K_G of a particular group is compromised, still the adversary needs valid individual keys of the SNs for decrypting the information sent from an SN. In case of the compromise of a GD, the base station gets involved for revoking the keys and even in this case, only one group is affected while others can still operate securely.

As the group dominators rule over all other sensors in the group for data transmission, the dominators might require more energy, processing and storage power. For this, a set of sensors with greater resources could be considered as dominators. In our approach we kept the number of cluster heads small; hence, it can reduce the overall cost in comparison with a network which requires a large number of cluster heads. To avoid traffic concentration on a few cluster heads, the cluster size should be evenly distributed among the cluster heads. Our approach performs well in this case. To reduce inter-cluster-head traffic, the number of clusters should be controlled and in our scheme, as the number of cluster heads is relatively less, the number of clusters is relatively less. The aspect of keeping the size of dominating set (i.e., number of cluster heads) to a minimum also helps for better security in the network in the sense that there is comparatively smaller number of entry-paths (to the base station) for injecting false reports by the adversaries and each dominator in the set can check the validity of the reports before forwarding those to the base station. For this, as stated earlier, the GD can set a particular value τ, which is the number of sensors in a particular group/cluster that should send the same report to the GD for convincing it that the report is true. The GDSN that could be formed in the network does not have any subordinate sensors; hence, it can easily carry an extra K_G in its memory. Basically in that case, only the rank of the SN changes but it does not incur any significant load on it.

Our approach also has some limitations. If a sensor network is deployed via random scattering (e.g., from an airplane), the sensors could be well-scattered even if one group is

released at a time (the worst case as mentioned earlier) and the nodes of the same group could be out of the communication ranges of one other after deployment. Even if the nodes are deployed by hand, the large number of nodes involved in distributed sensor networks makes it costly to predetermine the location of every individual node. The re-keying feature ensures robust security as with each addition of new sensor, the K_G is renewed but, it could also be resource-exhaustive. In such a case, the key renewal mechanism can be omitted. However, in some cases like say for example in military networks, as ensuring security is the major issue, we can consider the schemes with a slight increase of overall costs.

Figure 12 Given a connectivity probability, expected degree of a GD from the high level view



6 Conclusions and future works

In this paper, we have proposed an efficient approach of secure bootstrapping and beginning the operation of distributed sensor networks based on key pre-distribution and prior rank assignments. We have addressed secure clustering, starting of operation and secure message delivery within a DSN. We believe that our work opens the opportunity of many other future works. Based on the secure formation of the network, a secure and efficient routing protocol can be devised. As our future works, we intend to perform more detailed analysis on the potential applications of star based key assignment for misbehaviour detection, secure distributed storage and secure routing. In addition, we will develop an efficient intrusion detection mechanism that can work side-by-side our approach to detect the presence of any rogue node in the network.

Acknowledgements

This research was supported by the MKE, Korea, under the ITRC support program supervised by the IITA (IITA-2009-(C1090-0902-0016)).

References

Banerjee, S. and Khuller, S. (2001) 'A clustering scheme for hierarchical control in multi-hop wireless networks', *Proceedings of IEEE INFOCOM*, Vol. 2, pp.1028–1037.

Bohge, M. and Trappe, W. (2003) 'An authentication framework for hierarchical ad hoc sensor networks', *Proceedings of ACM WiSE'03*, San Diego, CA, USA, pp.79–87.

Carman, D.W., Kruss, P.S. and Matt, B.J. (2000) 'Constraints and approaches for distributed sensor network security', *NAI Labs Technical Report # 00–010*, dated 1 September.

Clark, B.N., Colbourn, C.J. and Johnson, D.S. (1990) 'Unit disk graphs', *Discrete Mathematics*, Vol. 86, pp.165–177.

Das, B. and Bharghavan, V. (1997) 'Routing in ad-hoc networks using minimum connected dominating sets', *Proceedings of the IEEE International Conference on Communications (ICC'97)*, pp.376–380.

Erdös, P. and Rényi, A. (1959) 'On random graphs', *Publicationes Mathematicae*, Vol. 6, pp.290–297.

Ferreira, A.C., Vilaça, M.A., Oliveira, L.B., Habib, E., Wong, H.C. and Loureiro, A.A. (2005) *On the Security of Cluster-Based Communication Protocols for Wireless Sensor Networks*, *ICN 2005, LNCS 3420*, pp.449–458, Springer-Verlag.

Garey, M.L. and Johnson, D.S. (1979) *Computers and Intractability: A Guide to the Theory of NP-Completeness*, W.H. Freeman, San Francisco.

Gupta, G. and Younis, M. (2003) 'Load-balanced clustering of wireless sensor networks', *Proceedings of IEEE International Conference on Communications (ICC'03)*, Vol. 3, pp.1848–1852.

Halgamuge, M.N., Guru, S.M. and Jennings, A. (2003) 'Energy efficient cluster formation in wireless sensor networks', *Proceedings of the 10th International Conference on Telecommunications*, Vol. 2, pp.1571–1576.

Heinzelman, W.R., Chandrakasan, A. and Blakrishnan, H. (2000) 'Energy-efficient communication protocol for wireless microsensor networks', *Proceedings of IEEE Hawaii International Conference on System Science*, pp.4–7.

Karlof, C. and Wagner, D. (2003) 'Secure routing in wireless sensor networks: attacks and countermeasures', *Elsevier's Ad Hoc Network Journal, Special Issue on Sensor Network Applications and Protocols*, pp.293–315.

Lee, S., Yoo, J. and Chung, T. (2004) 'Distance-based energy efficient clustering for wireless sensor networks', *Proceedings of the 29th Annual IEEE International Conference on Local Computer Networks*, pp.567–568.

Liu, J-S. and Lin, C-H.R. (2003) 'Power-efficiency clustering method with power-limit constraint for sensor networks', *Proceedings of the 2003 IEEE International Performance, Computing, and Communications Conference*, pp.129–136.

Mathew, R., Younis, M. and Elsharkawy, S.M. (2005) 'Energy-efficient bootstrapping for wireless sensor networks', *Innovations Syst. Softw. Eng.*, Vol. 1, No. 2, pp.205–220, Springer, London.

Nowak, R.D. (2003) 'Distributed EM algorithms for density estimation and clustering in sensor networks', *IEEE Transactions on Signal Processing*, Vol. 51, No. 8, pp.2245–2253.

Oliveira, L.B., Wong, H.C., Bern, M., Dahab, R. and Loureiro, A.A.F. (2006) 'SecLEACH – a random key distribution solution for securing clustered sensor networks', *Proceedings of the Fifth IEEE International Symposium on Network Computing and Applications*, pp.145–154.

Pathan, A-S.K. and Hong, C.S. (2006) 'A key-predistribution-based weakly connected dominating set for secure clustering in DSN', *HPCC 2006, Lecture Notes in Computer Science*, Vol. 4208, pp.270–279, Springer-Verlag.

Pathan, A-S.K., Lee, H-W. and Hong, C.S. (2006)'Security in wireless sensor networks: issues and challenges', *Proceedings of the 8th IEEE ICACT 2006*, Vol. 2, Phoenix Park, Korea, pp.1043–1048.

Prasad, N.R. and Alam, M. (2006) 'Security framework for wireless sensor networks', *Wireless Personal Communications*, Vol. 37, Nos. 3–4, pp.455–469, Springer, Netherlands.

Rhee, M.Y. (2003) *Internet Security: Cryptographic Principles, Algorithms and Protocols*, pp.84–95, John Wiley & Sons, ISBN: 0-470-85285-2.

Tzevelekas, L., Ziviani, A., Amorim, M.D.D., Todorova, P. and Stavrakakis, I. (2005) 'Towards potential-based clustering for wireless sensor networks', *Proceedings of The 2005 ACM Conference on Emerging Network Experiment and Technology*, Toulouse, France, pp.292–293.

Wokoma, I., Sacks, L. and Marshall, I. (2003) 'Clustering in sensor networks using quorum sensing', *London Communications Symposium*, University College London.

Wood, A.D., Stankovic, J.A. and Son, S.H. (2003) 'JAM: a jammed-area mapping service for sensor networks', *24th IEEE Real-Time Systems Symposium (RTSS 2003)*, pp.286–297.

Ye, M., Li, C., Chen, G. and Wu, J. (2005) 'EECS: an energy efficient clustering scheme in wireless sensor networks', *Proceedings of the 24th IEEE International Performance, Computing, and Communications Conference*, pp.535–540.

Younis, O. and Fahmy, S. (2004) 'Distributed clustering in ad-hoc sensor networks: a hybrid, energy-efficient approach', *IEEE Transactions on Mobile Computing*, Vol. 3, No. 4, pp.366–379.