

A Dynamic Level-based Secure Data Aggregation in Wireless Sensor Network

Tamer AbuHmed and DaeHun Nyang

Information Security Research Laboratory
Graduate School of IT & Telecommunication
InHa University
tamer@seclab.inha.ac.kr, nyang@inha.ac.kr
<http://seclab.inha.ac.kr>

Abstract. The resource constraints and security are the main challenges for data aggregation in WSN. In this paper, we present a dynamic and secure scheme for data aggregation in WSN. The proposal scheme includes level-based key derivation, data aggregation, and a new node join phases. Furthermore, we do a security analysis for a related Level-based Key Management (LBKM) scheme proposed by Kim *et al.* The analysis shows that LBKM is insecure for one node compromising and neighbor nodes misbehavior. To this end, we propose different level-based key management scheme for secure data aggregation. Our scheme is secure and more efficient than LBKM scheme in term of communication overhead and security.

Key words: WSN, Secure data aggregation, In-network processing

1 Introduction

Wireless sensor networks (WSNs) have gained much attention in recent years because of their capability of providing an interaction between humans and environment, which consequently can help to build an intelligent environment [1]. Therefore, WSN becomes commonly used in various fields and applications. From the military applications to the civilian applications, WSNs attracts researchers and industrials due to their low cost solutions for many real life challenges. WSN is involved in various applications such as traffic monitoring, forecasting, health monitoring, and building infrastructure safety. Sensor nodes in previous applications communally work to gather data from sensing area and feed them to central point for further processing. However, sensor nodes which are used in such applications have introduced several requirements because of the large number of nodes, the dynamic topology, the traffic randomness, and resources limitation (e.g., memory, computation, communication, and battery power). Therefore, designing and developing an efficient data gathering technique would have a strong impact on WSN life and its work efficiency. Also, the high density nature of sensor nodes in WSN application incurs a high redundancy of data reads gather from sensing area. As a consequence, sending back all of the sensing data will

cause a high communication overhead. Instead, raw data is aggregated by sensor nodes in the middle between source and final destination according to application functionality (e.g., return average, summation, min., and max.). However, deploying WSN in unobserved or hostile environment makes sensor nodes subjects to either passive adversary or active adversary who compromises deployed nodes as a result of the physical tampering. For this reason, secure aggregation is considered as a one of the direct solutions for the case of an attacker existence in WSN. However, providing security requirements considered as challenging in which the sensing nodes or the aggregation nodes may have been compromised then attacker gets the secret information which is combined with the data aggregation and finally subverts the aggregated data and affect the WSN's application functionality. Therefore, the simplest way is hardening the attacker from gathering secret information which is used to secure data gathering. This goal can be archived by distributing secret using the secret sharing scheme. The secret sharing scheme is one of powerful tools that can be used in which the aggregator puts constraint on number of participant nodes m from n nodes that need to recover the secret. This property is very useful in WSN, where (i) not all sensors are participating (e.g., asleep), or (ii) a part of nodes are vulnerable to either passive or active adversary.

1.1 Contribution of the paper

The main purpose of the paper is to design a secure data aggregation scheme which satisfies the main security requirements such as: integrity, confidentiality, reliability, and freshness. Moreover, we take advantage of the routing structure of the WSN in constructing a level based key management of secure data aggregation such that nodes in the upper level can decrypt lower level node's data by one key. That means to decrease the key materials each node holds. Also, because of the routing tree alteration due to the sudden node failure, sleeping, or compromising, designing a dynamic data aggregation scheme is one of the main goals in for the WSN environment.

- We enhance the security of the key management scheme of Kim *et al.* [2] which generates a hierarchical level-based key for secure data aggregation on WSN.
- We propose a secure data aggregation scheme to aggregate data over the hierarchical routing tree with low communication overhead comparing to Kim *et al.* scheme. Our scheme ensures that the adversary cannot misbehave the aggregation process if the compromised nodes are less than t nodes. A comparison between our work and the related work of [2] is presented in term of security and communication overhead.
- We propose a mechanism for the new nodes joins the WSN in the case of replacing the dead or failing nodes.

1.2 Notation

We will use the notation shown in table 1 to describe our scheme.

Table 1. Notations

Term	Declaration
N	The overall number of nodes in the network
FA_i	Node i status factors (e.g., its power, tree level, etc).
S	Secret key which is used in decryption of aggregated data. it's difference in each level.
W_i	Secret shared value of node i .
$E_k(M)$	Symmetric encryption for message M by secret key k .
K_{PA}	Pairwise key between node P and A .
\hat{K}_{SA}	Symmetric level-based encryption key which is used in encrypt the aggregated data of node A .
α_i	ID of sensor node which will participate in secret sharing

1.3 Threshold Secret Sharing Scheme (T-SSS)

Secret sharing concept was firstly introduced by Shamer [3] to protect from secret key exposure. Secret sharing is one of the tools which is widely used to share a secret key S among predetermine group of participants in such a way that a predetermined subgroup of participants can recover the secret. The secret shared scheme can be generated in such a way that a threshold number of participants t cooperatively can recover the secret S . This scheme called threshold secret sharing scheme (n,t) , where n is the number of participants and t is the threshold number from n needed to recover the secret. Threshold (n,t) secret sharing scheme has two phases i) Secret shares set W generation as illustrated in equation 1, and ii) secret key S recovery as illustrated in equation 2

$$\begin{pmatrix} W_1 \\ W_1 \\ W_2 \\ \vdots \\ W_N \end{pmatrix} = \begin{pmatrix} 1 & \alpha_1 & \dots & \alpha_1^{t-1} \\ 1 & \alpha_2 & \dots & \alpha_2^{t-1} \\ 1 & \alpha_3 & \dots & \alpha_3^{t-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \alpha_N & \dots & \alpha_N^{t-1} \end{pmatrix} \times \begin{pmatrix} S \\ r_1 \\ r_2 \\ \vdots \\ r_{t-1} \end{pmatrix} \quad (1)$$

$$\begin{pmatrix} S \\ \bar{r}_1 \\ \bar{r}_2 \\ \vdots \\ \bar{r}_t \end{pmatrix} = \begin{pmatrix} 1 & \alpha_{j1} & \dots & \alpha_{j1}^{t-1} \\ 1 & \alpha_{j2} & \dots & \alpha_{j2}^{t-1} \\ 1 & \alpha_{j3} & \dots & \alpha_{j3}^{t-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \alpha_{jt} & \dots & \alpha_{jt}^{t-1} \end{pmatrix} \times \begin{pmatrix} W_{j1} \\ W_{j1} \\ W_{j2} \\ \vdots \\ W_{jt} \end{pmatrix} \quad (2)$$

2 Related Works

Recently, data aggregation attracted many researchers in WSN. Thus, different aggregation protocols had been proposed to achieve in-network processing capability in WSN. These aggregation protocols can be divided into two main

categories. First, the aggregation protocols which assumed a trusted environment such as, directed diffusion [4], LEACH [5], greedy aggregation [6], and TAG aggregation service for TinyOS motes which supports SQL-like language for expressing aggregation queries over sensor nodes data [7]. Otherwise, in the other category, the aggregation protocols assume the existence of the passive and active adversary whose try to eavesdrop or to subvert the functionality of WSN. The latter category called the secure data aggregation protocols. Hu and Evans [8] were the first who study secure data aggregation. In [8], Hu and Evans proposed a data aggregation scheme which is secure as long as no two consecutive nodes are compromised in WSN routing tree. However, their protocol is vulnerable if a parent and a child node in their hierarchy are compromised. Recently, in [2], Kim and Ramakrishna proposed level-based aggregation protocol which works on tree based WSN structure. Their work is much related to our proposal. In the scheme, Kim and Ramakrishna proposed a key generation technique that depends on hash chain where ancestor's hash function use by descendant nodes to derive it's key for data encryption. This feature of key derivation can help ancestor node to decrypt any encrypted data which is sent by descendants. There are also many other schemes such as Zhu *et al.* [9], Kim *et al.* [2], for more details see [10].

3 Our Scheme

3.1 Assumptions & Network Model

We assume a hierarchical multi hops structured tree constructed from N sensors as shown in Fig. 1(a) and clearly shown in Fig. 3. We assume a bounded attacker, which can compromise some of the sensors as well as the aggregators. Hence, the behavior of a compromised sensor is totally determined by the adversary (i.e., a compromised nodes or aggregators have a Byzantine behavior). We assume that BS is secure and each sensor has ID and key material for pairwise key establishment with its neighbors. In this WSN hierarchical construction a node in level L_i aggregates predecessor nodes data in level L_j , where $j < i$ as shown in Fig. 3. In WSN, a node failure is common due to resource limitation, node compromising, programming failure and other reasons. Thus, providing a dynamic mechanism for secure in-network processing capability is our goal in this proposal.

3.2 Key Generation

First step for secure data aggregation is the key generation and management which will be an important step for the secure data aggregation phase. The key generation phase is executed after the tree construction of the deployed WSN. Hence, the routing protocol like TinyOS beaconing [11] initially constructs the routing tree and each node knows its parent and children. Afterword, The key generation phase is taking over. The key generation depends on the hash function

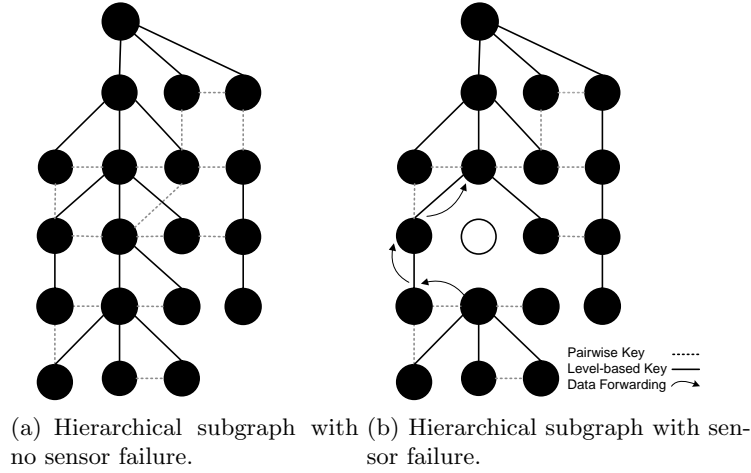


Fig. 1. An instance subgraph from WSN shows the effect of the sensor failure on data delivery and processing.

and the threshold secret sharing scheme as mentioned in subsection 1.3. Broadly speaking, these cryptographic tools are considered as light tools which can be applied to WSN and not gain any burden in terms of computation. Also, the key generation phase gives a decryption capability to aggregation nodes bottom-up way in the routing tree. That is the higher level nodes in WSN tree hierarchy can decrypt lower level data of the child nodes. This capability gives flexibility to real WSN application scenarios in the case of the dedicated aggregation nodes in the higher level of WSN tree are asleep, dead, or compromised. Therefore, Kim *et al.* scheme [2] which uses key derivation technique that depends on hash chain of the parent hash with node ID inspired us for our key derivation scheme. However, our key generation uses a level-based hash chain for key derivation which means that we will not use one hash to all WSN tree. Aggregation protocol use different key derivation hash chain for each sub tree of nodes in tree as illustrated in Fig. 2. This level-based key derivation technique gets fewer burden in term of key update. Hence, in the case of re-keying, key derivation does not have to be executed for the whole tree when part of nodes are compromised. The key generation and shares of secret key for data aggregation is executed after tree relation as follows:

- 1-a Child nodes of BS receive S_{BS} via secure channel (e.g., pairwise key), then each node in level i ($\forall i \leq$ key derivation chain limit) of tree generates $S_i = H(ID_i|ID_{i-1}|S_{i-1})$ and pass it to its children (e.g., child node ID_1 will $S_1 = H(ID_1|ID_{BS}|S_{BS})$). This technique of key derivation mechanism relies on the node identity (ID_i) and the edge between node and its parent which is provably secure against key recovery [12].

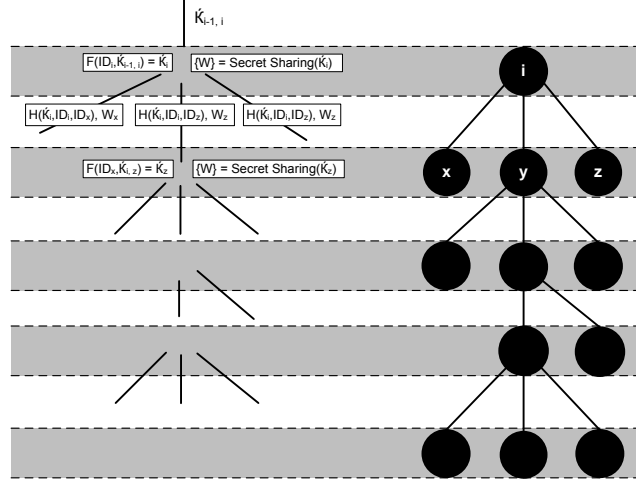


Fig. 2. Key generation in random routing tree.

- 1-b Then, each node i , at level j of key chain, execute secret sharing scheme for its key as illustrated in equation 1 where $S = H(ID_i | ID_{j-1} | S_{j-1})$. However, setting the parameters (t, n) of the secret sharing scheme in equation 1 will depend on the density of the node children and the node's power.
- 2 Each node i store hash value of W set with node ID_i s of each node received secret share, as $H(W_i | ID_i)$. This step is crucial for verifying shares later when data is aggregated.
- 3 Finally, each node delete $\{W\}$ set of secret shares and secret key S .

The basic construction of our level-based key generation is slightly different from the [2] construction in order to prevent the security weakness of the last scheme in the case of the equal level nodes misbehaves. More details would be in section 4.

3.3 Data aggregation

Data aggregation is one of the most preferable techniques to significantly reduce energy consumption in WSN, where data is combined according to aggregation function (e.g., sum, max, min, \dots etc) to reduce data redundancy of the transmitted data. In our proposed protocol, the aggregation process fully depends on co-operation between the sensor nodes in sensing area and their parents. Thus, aggregation process will be activated in the aggregation node according to the number of the children nodes participated to recover the decryption key. We exploit the routing tree construction of children nodes which are connected to one parent node. Moreover, another advantage of our protocol is the dynamic aggregation capability, where any parent node works as aggregator if their children who send data are greater than a threshold number t . Else, the parent node

passes its children data packet into the higher level node. This flexibility of the dynamic aggregation technique is helpful in WSN where sensors sleep or die. Thus, aggregation can be delayed to the upper levels of the WSN routing tree instead of being aggregated directly. Also, using the threshold secret sharing in our data aggregation scheme gives the scheme more advantages such as:

- Availability: greater than or equal to t parties can recover S .
- Confidentiality: less than t parties have no information about S .
- Flexibility: the owner of S can increase n and add new shares for the new joined nodes without affecting other shares. Also, secret shares can be updated if subgroup of nodes less than t is being compromised. In this case, attacker will not get any benefit from shared he already had.

Our aggregation scheme assumes tree-based structure for WSN whatever tree structure. The general procedure for secure data aggregation process would be as follows:

1. Sensor nodes set $\{ID_1, ID_2, \dots, ID_t\}$ which are belong to parent P encrypt their data R_i by their level key \hat{K}_{P_i} .
2. Also, each node encrypts th secret shared W_i received later in key generation phase.
3. At parent node, if the number of children nodes which participate is $\geq t$, then secret shares are verified. Else, the node passes the children packets to its parent.
4. Finally, the node recovers its key as shown in section 1.3, aggregates the children data, and repeats step (1).

According to the tree structure shown in Fig. 3, a secure data aggregation (with $t = 3$) is executed as follows:

- A \rightarrow P:
 $ID_A, N_i, E_{\hat{K}_{SA}}(R_A|N_i), E_{K_{PA}}(W_A|N_i)$
- B \rightarrow P:
 $ID_B, N_i, E_{\hat{K}_{SB}}(R_B|N_i), E_{K_{PB}}(W_B|N_i)$
- C \rightarrow P:
 $ID_C, N_i, E_{\hat{K}_{SC}}(R_C|N_i), E_{K_{PC}}(W_C|N_i)$
- * Data should be received from at least $t \geq 3$ nodes.
- Parent P then decrypts and collects the secret shares from at least t (e.g., $S_{W_A}, S_{W_B}, \dots, S_{W_t}$) and verifies its correctness before constructing the secret \hat{K}_{SP} . Verification occurs by comparing each node secret share with hash value parent node stored (i.e., $H(\hat{W}_i) =? H(W_i)$). This verification of secret shares makes an implicit authentication of children.
- P \rightarrow A_{gg} :

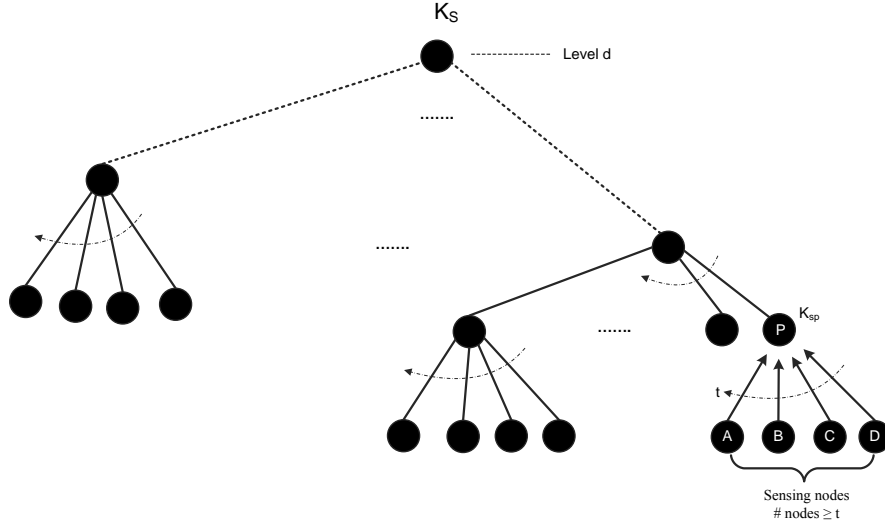


Fig. 3. Key reconstruction in the level-based data aggregation protocol, example: construct two level aggregation keys K_{s1} , K_{s2}

3.4 Join Node

Scalability is one of the requirements of WSN where a new node deployment is occurred securely and without any adversary misused. Usually, node joining or deployment is a common scenario occurs when many nodes died or compromised in the already deployed WSN. Therefore, secure node joined is critical to security of the aggregation protocol. Join node protocol started when a new deployed nodes broadcast join request for their neighbors. Afterwards, neighbor nodes send back their ID's and other parameters for connectivity (e.g., its power, tree level, etc). Consequently, according to the replay messages from neighbor nodes, the new joined node calculates a pairwise key with the desired neighbor node. After node authentication, the parent of the joined node verifies the correctness and trustiness of the joined node request. The parent node tries to recover generate secret S by other nodes co-operation, and sends a secret share to the new joined node without changing the other children shares. However, if the parent node of a new joined node did not have the ability to reconstruct the secret as in equation 2, due to some children nodes failure or being asleep. In this case, the node asks its parent to the secret S_{i-1} construction and to pass it by secret channel. Thus, the parent node have two choices which makes a node join algorithm flexible and immune against any failure and sleeping mode situation. Further details have been introduced in algorithm 1, where node J join node P as in Fig. 3.

<p>Input : Sub tree of nodes T, Joint node J</p> <p>Output: Aggregation key K_{PJ} and shared W_{PJ} installed into node J.</p> <ol style="list-style-type: none"> 1 J broadcast join REQ. to T in deployed area; all nodes $\in T$ overhear.; 2 Each node $\in T$ replies to J with ID_i, FA_i and other information for pairwise key establishment (PKE).; 3 J evaluate the proper node to join according to received FA_i and PKE; 4 J replies to node P, which if join success will be the parent of J, with ID_J, ID_P, N_i, $MAC(K_{JP}, ID_J ID_P N_i)$; 5 P will authenticate received MAC by pairwise key K_{PJ}; 6 if <i>Authenticated</i> then 7 if <i>has $\geq t$ active children</i> then 8 P performs secret share S_P recovery.; 9 Generate secret share W_P for J; send W_P and K_{PJ} via secret channel.; 10 else 11 P send aggregation key request to its parent.; 12 Parent of P repeat steps 7 and send its aggregation key to P.; 13 P performs secret share S_P recovery.; 14 Generate secret share W_P for J; send W_P and K_{PJ} via secret channel.; 15 else 16 Discard Join REQ.
--

Algorithm 1: Node join protocol for tree based WSN.

4 Analysis and Evaluation

In this section, we discuss the security and the communication overhead of our scheme and the related scheme of Kim *et al.* [2]. First, we show the security requirements of our scheme such as: the confidentiality, integrity, refreshness, and authenticity. Our scheme satisfies the confidentiality property of the raw and aggregated data by encryption. The integrity property is implicitly satisfied if the parent node decrypts child data and gets the correct structure of data, e.g., $(R_B|N_i)$ as in 3.3. Also, the refreshness fulfilled by adding the nonce N_i in each transaction which can be a counter value stored between nodes. Thus, replay attack in this scenario is not applicable. Regarding to the physical node compromising, attacker in our scheme cannot recover the level-based key S_i except if the attacker compromises $\geq t$ of nodes in one hit. Otherwise, if the intrusion detection system detects the attacker before that, the secret shares of other nodes will be updated. In this way, the attacker will not take any benefit from the compromised nodes. Finally, the authenticity of nodes interaction is derived implicitly from the encrypted secret shared which each node sends to the parent node. In [2], Kim *et al.* proposal of level-based aggregation scheme, which depends on hash chain key derivation only, is vulnerable to one node compromising. Thus, in [2] scheme, it's enough that attacker compromises one node in WSN routing tree and gets the aggregation key. By doing so, the attacker will control all the

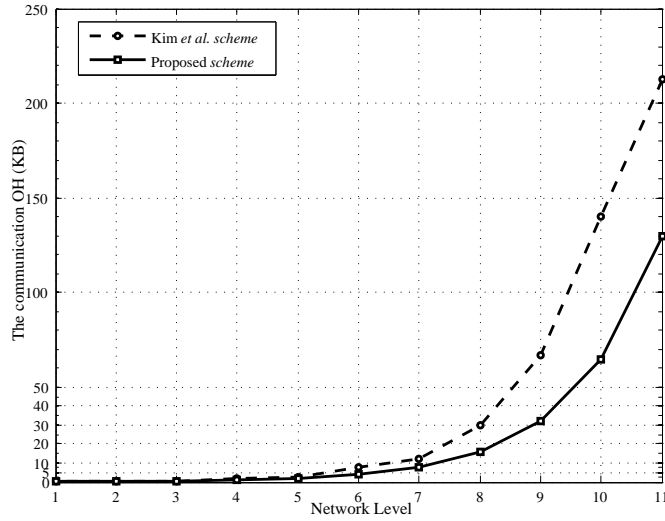


Fig. 4. Comparison of the communication overhead between Kim *et al.* and the proposed scheme.

lower sub tree nodes and subverts WSN aggregation protocol by inject faked aggregated value which will deviate data aggregation result. The attack in this type of protocols is critical as the attacker compromise node near the root of the tree. Since the attacker compromises a node near the root of routing tree, the attacker can influence the aggregation result much more because of the big branch of lower level nodes that attacker controls. Also, in [2], their aggregation protocol suspect to another security breach which is the replay attack. Replay attack can be held by attacker, who has control over communication channel and eavesdrops the encrypted packets and buffers these packets, and resends early packets again over the communication channel. Moreover, even if aggregation protocol executes a re-keying protocol periodically, it will not be sufficient to prevent the replay attack due to the period in between the re-keying. Finally, Kim *et al.* key derivation gives the nodes at the same level (i.e., not aggregation nodes) the capability to decrypt their neighbors' data. This will severely affects the confidentiality of the aggregation scheme. Regarding to the communication overhead (COH), we analyzed the COH of our scheme and comparing it with the related work of Kim *et al.* We computes the COH by considering an ideal routing tree with root as BS and each node has two children. The observation is that the COH of our scheme is less than the related work of Kim *et al.* about 40% as tree level increased or number of the children increased. A comparison between the proposed aggregation scheme and Kim *et al.* scheme in the COH is shown in Fig. 4.

5 Conclusion

Data aggregation and in-network processing are important techniques for WSN. With the existence of the attacker threat, secure data aggregation scheme with immunity against malicious attacker is crucial for WSN correctness and security. In this paper, we proposed a secure and flexible data aggregation scheme with hierarchical key generation, aggregation, and node join protocols. The key are protected using secret sharing primitive such that the level key will be secure as long as the number of compromised nodes is less than t node. To show the superiority of our scheme, performance and security analyses were performed for both proposed and Kim *et al.* schemes.

References

1. Akyildiz, I.F., Su, W., Sankarasubramaniam, Y., Cayirci, E.: Wireless sensor networks: a survey. *Computer Networks* **38** (2002) 393–422
2. Kim, K.T., Ramakrishna, R.S.: A level-based key management for both in-network processing and mobility in wsns. In: *Mobile Adhoc and Sensor Systems (MASS)*, IEEE International Conference. (2007)
3. Shamir, A.: How to share a secret. *Communications. ACM* **22**(11) (1979) 612–613
4. Intanagonwiwat, C., Govindan, R., Estrin, D.: Directed diffusion: a scalable and robust communication paradigm for sensor networks. In: *MOBICOM*. (2000) 56–67
5. Heinzelman, W.R., Chandrakasan, A., Balakrishnan, H.: Energy-efficient communication protocol for wireless microsensor networks. In: *HICSS*. (2000)
6. Intanagonwiwat, C., Estrin, D., Govindan, R., Heidemann, J.S.: Impact of network density on data aggregation in wireless sensor networks. In: *ICDCS*. (2002) 457–458
7. Madden, S., Franklin, M.J., Hellerstein, J.M., Hong, W.: Tag: A tiny aggregation service for ad-hoc sensor networks. In: *OSDI*. (2002)
8. Hu, L., Evans, D.: Secure aggregation for wireless network. In: *SAINT Workshops*. (2003) 384–394
9. Zhu, S., Setia, S., Jajodia, S., Ning, P.: An interleaved hop-by-hop authentication scheme for filtering of injected false data in sensor networks. In: *IEEE Symposium on Security and Privacy*. (2004) 259–271
10. Sang, Y., Shen, H., Inoguchi, Y., Tan, Y., Xiong, N.: Secure data aggregation in wireless sensor networks: A survey. *Parallel and Distributed Computing, Applications and Technologies*, 2006. *PDCAT '06. Seventh International Conference on* (Dec. 2006) 315–320
11. Hill, J., Szewczyk, R., Woo, A., Hollar, S., Culler, D., Pister, K.: System architecture directions for networked sensors. *SIGPLAN Not.* **35**(11) (2000) 93–104
12. Atallah, M.J., Friksen, K.B., Blanton, M.: Dynamic and efficient key management for access hierarchies. In: *ACM Conference on Computer and Communications Security*. (2005) 190–202