Efficient Group Key Distribution Schemes with Self Healing Property and Distributed Hash Table-Based Routing and Data Management in Wireless Sensor Networks

**Abstract:** The main property of the self-healing key distribution scheme is that users are capable of recovering lost group keys on their own, without requesting additional transmission from the group manager. Self-healing key distribution schemes allow group managers to broadcast session keys to large and dynamic groups of users over unreliable channels. Such schemes are quite suitable in supporting secure communication in wireless networks and mobile wireless ad-hoc networks. Recent papers have focused on self-healing key distribution, and have provided definitions, stated in terms of the entropy function, and some constructions. Secure key distribution schemes for group communications allow to establish a secure multicast communication between a group manager and group members through an unreliable broadcast channel. The article classifies, analyzes and compares the most significant key distribution schemes, by looking at the selective key distribution algorithms, at the predistributed secret data management, and at the self-healing mechanisms. In this paper, we propose a new selfhealing key distribution scheme, which is optimal in terms of user memory storage and more efficient in terms of communication complexity than the previous results. The last few years have witnessed the emergence of several approaches that build Distributed Hash Tables (DHTs) over WSN. DHTs are initially conceived for efficient data lookup in large-scale wired networks. The main objective of this combination is to manage location-independent data and nodes identification. DHT mapping over WSN brings however new challenges. This paper presents an analytical survey on applying DHT techniques in WSNs. It describes existing DHT-based routing and data management protocols and includes a detailed classification of them.

1. Introduction:

A major problem area in secure group communications is group key management which is concerned with the secure distribution and refreshment of keying material. For large groups with frequent membership changes, the group key has to be updated on every membership change and securely redistributed to the existing members of the group. This is referred to as group rekeying. Self-healing key distribution, recently introduced in [22], enable a dynamic group of users to establish a group key over an unreliable network. In such a scheme, a group manager, at the beginning of each session, in order to provide a key to each member of the group, sends packets over a broadcast channel. Every user, belonging to the group, computes the group key by using the packets and some private information. The group manager can start multiple sessions during a certain time-interval, by adding/removing users to/from the initial group. Self-healing group key distribution schemes has recently received a lot of attention from the researchers, as a method enabling large and dynamic groups of users to establish group keys over unreliable network for secure multicast communication. In such schemes, time is divided into epochs called sessions. At the beginning of each session, a Group Manager transmits some broadcast message, in order to provide a common key to each member of the group.

Every user, belonging to the group, computes the group key using the message and some private information. The main property of the scheme is that, if some broadcast message gets lost, then users are still capable of recovering the group key for that session by using the message they received at the beginning of a previous session and the message they will receive at the beginning of a subsequent one, without requesting additional transmission from the Group Manager. This approach decreases the workload on the Group Manager and reduces network traffic as well as the risk of user exposure through traffic analysis. The article reviews the most significant self-healing group key distribution schemes, and it gives insight into open research problems in this area. Self-healing key distribution schemes are stateless and noninteractive, i.e., users do not need to update the secret information they receive in the setup phase, and they do not need to send any key-request message to the group manager. Some benefits of such an approach are reduction of network traffic, reduction of the work load on the group manager, and a lower risk of user exposure through traffic analysis.
Applications. The relevance of self-healing key distribution has been well motivated in [22] and, later on, in [18]. Self-healing key distribution schemes can be used to efficiently achieve secure communication in wireless networks and mobile wireless ad-hoc networks. International peace operations and rescue missions, where there is no network infrastructure support and the adversary may intercept, modify, and/or partially interrupt the communication, are important examples of cases in which reliability, confidentiality and authenticity of the communication is a major concern. In the above settings, all techniques developed for secure group communication in traditional networks might be used. However, some unique features of mobile and ad-hoc networks identify a new scenario: nodes/devices in mobile networks may move in and out of range frequently. Moreover, devices are powered by batteries.

Hence, expensive computations like the ones required by public key cryptography are not suitable. Previous Work. Self-healing key distribution was introduced in [22]. Definitions and lower bounds on the resources required for implementing such schemes, stated in terms of the entropy function, and some constructions were provided. In [18] this definition was generalized and more efficient constructions were presented. Other constructions were given in [16]. Finally, in [3], some efficient constructions were presented, and it was pointed that some of the constructions given in [22] are flawed. The above papers have mainly considered unconditionally secure schemes. Related Work. Broadcast Encryption is a closely related research area. Loosely speaking, in broadcast encryption a broadcaster delivers in a secure way to a privileged subset of recipients of a given universe a session key. The recipients then use this key for decrypting broadcast transmissions. Broadcast Encryption is static, i.e., the family of possible privileged subsets is specified during the setup phase of the scheme.

The rest of the article is organized as follows: in section 2 describe most important characteristics of self healing approach. section 3 describes self organizing networks(SON)functionanilities with self-healing reference model.section 4 specifies distributed hash table(DHT) with different pros and cons of mapping DHT over WSN.section 5 highlights existing data management and routing protocols that combine DHT to WSN. Finally, section 6 conclude the paper with future directions.
2. Characteristics Of Self Healing Approach:

An active research area in recent years. In this section we introduce the basic idea and the most important characteristics of the self healing approach to the group key distribution.

1. Overview of the network model

Self-healing group key distribution schemes can be used in various network scenarios, thus, to make their analysis and comparison easier, we introduce an abstract model of the network in which these schemes are applicable. The network consists of a single Group Manager (GM) and a finite universe of User Nodes (U). Group Manager is a resource rich node with high computational power, large memory space, and unlimited energy resources. User nodes, on the other hand, have limited computational power, limited memory, and limited energy resources. GM communicates with nodes in U through an unreliable broadcast channel. She transmits broadcast messages which are received by all users. Because of nodes mobility and channel communications errors, some messages can be lost. Message retransmission should be avoided, if possible, since it is costly and requires feedback connection from receiver nodes to GM, which may not always be available.

The main goal is to establish secure multicast communication between GM and members of a group of nodes \( G \subseteq U \), which is a subset of U. Group G is dynamic, user nodes can join and leave. Communications security is achieved by message encryption and authentication using shared symmetric secret group key \( K \). A shared key is convenient, but it can be disclosed by nodes leaving the group, or by group members intercepted by an adversary. To achieve high security level the key shall be changed frequently throughout the group lifetime. To do so, secure group key distribution mechanism, with GM acting as a Trust Anchor, is needed for key replacement. A prospective group key distribution scheme should satisfy the following requirements:

i. Authorization: The scheme should prevent adversaries or unauthorized user nodes, which are not in \( G \), from learning the group key.

ii. Key freshness: Key distribution scheme has to provide fresh keys.

iii. Efficiency: Group key distribution scheme shall be efficient with respect to communication, computational, memory, and energy cost. It shall take into account user nodes limitations.
iv. *Scalability:* Network size usually ranges from dozens to hundreds of thousands nodes, so the scheme has to be scalable to be practically applied.

v. *Communications Model:* The scheme should be applicable to the network model described in this section.

2. **Group lifetime divided into sessions**

   Common group key is frequently updated to ensure secure multicast communication. Group lifetime is divided into epochs called sessions, single key instance is valid only throughout one session. Group membership can change between consecutive sessions (denoted by Gj in session j). At the beginning of session j, GM distributes a new session key Kj to nodes in Gj. Session duration is determined by the GM. It can vary over time, depending on security policy, group membership changes and nodes’ behavior. Session key changes have to be performed, with some predefined minimum frequency to protect the system from cryptanalysis attacks. Moreover, to effectively remove a node Ui from multicast group Gj, who is willing to leave, or is forced to leave, a new session must begin and nodes from Gj+1 shall start protecting group communication using a new Kj+1, which is not accessible to Ui. Thus, the choice of session length is a tradeoff between key distribution cost in terms of communication and computational overhead, and the required security level.

3. **Broadcast distribution of session keys**

   At the beginning of each session j, GM transmits a broadcast message Bj to distribute session key Kj to all nodes in Gj. Since Bj is broadcast, it can be received by any node in U, including the ones which are not authorized to obtain Kj. Thus, to prevent from unauthorized access, Kj has to usually achieved in three phases:

   i. **Set Up:** GM generates personal key Si for each node belonging to a multicast group, and distributes it to Ui over secure channel. Personal key Si is secret information used only by Ui to recover any future session keys from broadcast messages. Nodes can be equipped with personal keys before network deployment, if set of potential group members can be predetermined, or they can obtain their personal keys from GM when they join multicast group.

   ii. **Broadcast:** GM creates message Bj from Kj in a way fulfilling the following requirements:
✓ There is an efficient algorithm $\eta$, which for all $i : U_i \in G$, can be used to recover $K_j$ knowing $S_i$, that is: $K_j = \eta(B_j, S_i)$.

✓ There is no computationally viable algorithm $\varsigma$, which for any set of nodes $R \subset U \setminus G_j$, can be used to recover $K_j$ knowing personal keys of all nodes in $R$, that is: $K_j = \varsigma(B_j, \{S_l\} : U_l \in R)$ is infeasible.

✓ Session Key Recovery: Every member $U_j \in G_j$ recovers key $K_j$ from received message $B_j$ using her personal key $S_i$, by calculating $K_j = \eta(B_j, S_i)$.

4. User join and revocation Group $G$ is dynamic, which means user nodes can spontaneously join and leave it. A new node $U_i$ joining the group receives from GM her personal key $S_i$ valid in consecutive sessions $(r, \ldots, s)$, where $r$ is a session in which the user joins the group, and $s+1$ is a session in which she is supposed to leave the group. $U_i$ must not be able to obtain session keys, which are used in sessions before $r$ and after $s$. $U_i$ can also be forcibly removed from multicast group before $S_i$ expires. To revoke node $U_i$ in session $l : r < l \leq s$, GM has to construct all messages $(B_l, \ldots, B_s)$ in a way rendering $S_i$ useless.

5. Stateless key distribution: Self-healing key distribution schemes are stateless, that is they always permit group members $U_i \in G_j$ to obtain valid session key $K_j$ from the last broadcast message $B_j$, also when they miss some previous key distribution messages. This property is usually achieved by storing all the necessary and up-to-date group state information at GM and transmitting it in every broadcast $B_j$. User nodes do not update their personal keys upon receiving broadcast messages, so the keys stay valid even if they lack some messages. Nodes, which recover from temporary communication loss, can obtain current session key immediately after receiving the next broadcast message, and become fully functional group members.

6. Key recovery through self-healing: The purpose of self-healing is to add some additional information to message $B$, which would allow user nodes to recover keys from previous sessions lost due to communication errors. User nodes shall be able to recover lost session keys on their own, without any additional interaction with GM. $B_l$ preceding the lost packet $B_j$ with information from any message $B_r$ following it. Formally, given any $l$, $r$, such that $l < r$, there is an efficient algorithm $\zeta$ which for all $j : l < j < r$ can be used by node $U_i \in G_l \cap G_j \cap G_r$ to recover $K_j$ from known $B_l$ and $B_r$. 
that is \( K_j = \zeta(B_l, B_r, S_i)l \). It should be noted that self-healing mechanism allows to recover only past session keys. Users, who have lost the most recent message \( B_j \), are not able to obtain current session key \( K_j \) until the next session \( j + 1 \), in which they correctly receive message \( B_{j+1} \). Thus, self-healing mechanism does not guarantee availability of the most recent keys. It is worth recalling here, that user nodes can obtain the most recent session key \( K_j \) despite previous communication problems, upon receiving correct \( B_j \), even without self-healing mechanism, because this is guaranteed by the stateless property of key distribution scheme. In systems, where only the most recent session key is used to protect multicast group communication, self-healing property does not yield significant profits. It can be only used by user nodes to validate and decrypt messages collected in previous sessions, but not processed yet due to lack of appropriate session keys. However, with slight modification of key distribution scheme described so far, self-healing property can become really valuable.

As it was suggested in [4], introduction of a delay of \( d \) sessions between distribution of a session key, and usage of this key for the protection of multicast group communication, allows to improve availability of the currently used keys at the cost of key freshness. At the beginning of session \( j \) every node \( U_i \in G_j \) recovers \( K_j \) from the received \( B_j \) and stores it in a key vector \((K_{j-d+1}, \ldots, K_j)\). During session \( j \), key \( K_{j-d} \) is used to protect group communication. Due to the buffering of recent \( d \) keys, user nodes can seamlessly handle loss of up to \((d-1)\) consecutive broadcast messages, since knowledge of messages \( B_{j-d} \) and \( B_j \) allows to recover all session keys \((K_{j-d+1}, \ldots, K_{j-1})\) using self-healing property.

Unfortunately, delay between session key distribution and session key usage has an impact on node revocation. Node revocation performed by \( GM \) is effective only after \( d \) sessions, i.e. when group communication is protected by the session key, which is no longer accessible for the revoked node.

G. Formal definition of self-healing group key distribution To clarify scheme analysis, we introduce a general model of a self-healing group key distribution scheme. It is a generalized version of the model proposed by Blundo et al. in [5], so that it can cover both unconditionally secure and computationally secure schemes. Let us assume that \( m \) is the number of sessions supported by the scheme, and \( t \) is the maximum number of users, that can be revoked during the scheme lifetime. Self-healing group key is the original definition of the self-healing property, introduced by Staddon et al. in [3], however in some techniques presented in Next Section recovery of lost key \( K_j \) can be performed based
on the single message Br and also requirement that \( U_i \in G_l \cap G_j \cap G_r \) may be significantly relaxed, so that \( U_i \) must only be a member of the group \( G_j \).

Note, that there exist solutions with unlimited number of sessions, or unlimited number of users that can be revoked, but they should be considered as a limit case, with \( m = \infty \), or \( t = |U| \) respectively.

3. **Self Organizing Networks (Son) Functionanlities With Self-Healing Reference Model:**

3.1. **The Main SON Functionalities Are:**

3.1.1. **Self-Configuration:** includes functions for network deployment and configuration of its parameters. Network elements, thanks to autoconfiguration, can start in an autonomous manner, run setup routines, configure initial parameters, etc.

3.1.2. **Self-Optimization:** responsible for auto-tuning of parameters, which should be dynamically recalculated when traffic and network conditions change. Self-optimization includes tuning parameters related to: list of neighboring cells, traffic balance, handover, coverage, etc.

3.1.3. **Self-Healing:** includes functions to cope with service degradation or outage, including fault detection and diagnosis and mechanisms for outage compensation.

3.1.4. **Self-Healing Reference Model:** The unified framework for self-healing in wireless networks is proposed paradigm gathers and integrates the different functionalities and terminology related to selfhealing that have been separately identified by network operators and vendors, used in standards, and partially studied in existing references. All these functions should be performed autonomously with minimal human intervention.

3.1.5. **Information Collection:** This function is responsible for collecting input information used by the self-healing process. The more complete the information is, the faster the failures are identified and solved. The sources of information normally at the disposal of network operators are the following:
3.1.6. **Configuration Parameters:** This is the information on the actual configuration of NEs and network resources.

3.1.7. **Alarms:** Alarms are messages generated by NEs when there are faults.

3.1.8. **Network Counters:** Measurements from the network elements (including data from measurement reports), which are periodically transferred to the Operations & Management system (O&M). They can be related to traffic load, resource availability, etc.

3.1.9. **Mobile Traces:** This is information from specific users or mobiles, which can be collected from most NEs.

3.1.10. **Real Time Monitoring:** these are online measurements of specific items, such as traffic load and handovers performed in certain NE (e.g. eNB).

3.1.11. **Drive Tests:** These are field measurements, e.g. related to coverage and interference, performed in a certain area by specialized equipment such as measurement terminals and GPS.

3.1.12. **Key Performance Indicators (KPIs):** These are combinations of other measurements, e.g. a formula of counters that provide a meaningful performance measure. Typically, KPIs describe the success/failure rates of the most important events such as handovers or dropped calls.

3.1.13. **Context Information:** This is information on the environment, such as type of area (e.g. urban area) or typical weather in the cell (e.g. rainy). A self-healing system could consider all or only part of these sources of information, and it should be able to provide acceptable results even with partial information. Normally, the main data considered for manual troubleshooting are KPIs calculated from counters and alarms, due to their easier accessibility.

Currently, the main sources of information used by operators to identify faults are the statistics from the O&M. The drawback of using those sources of information is that it normally implies a centralized off-line approach, which is not always possible, e.g. in enterprise femtocell environments.
4. Distributed Hash Table (DHT) With Different Pros And Cons Of Mapping DHT Over WSN:

DHT is a popular decentralized distributed system. The main advantage of DHT is its efficient lookup service. Data in DHT are organized into (key, value) pairs. The keys are the products of a hash function that should balance the keys distribution in the whole network. Each node in DHT has a unique identifier. These identifiers belong to the same output space of the used hash function. In order to retrieve data of a certain key, it is useful to know the node that stores this key. So, first of all, the searching node hashes the key, then it routes the query to the node having the closest identifier to the key hashing. If a new node (having the identifier A) wants to join the DHT system, it sends a join message to an existing node. This message will be forwarded from a node to another until reaching the node master of the identifier A. At the reception of this message, this latter node updates its neighbors list and shares its own space as well as its own keys with the new neighbor.

4.1. Benefits:

Applying DHTs over WSNs can bring a lot of advantages to these networks. In WSN, there are thousands of sensors that send a huge amount of messages simultaneously which should be handled and treated efficiently. When a base station wants to get information, it should have an efficient and a realtime response. DHT is able to perform an efficient lookup in a scalable and reliable manner. Most of WSN are heavily deployed in emergency cases such as natural disasters and battlefields. In such cases, sensors are scattered from a plane and consequently, they have no information about their locations. This means that the network in these situations is auto-deployed. Consequently, proposed routing and data management
protocols should ensure their functionalities without worrying. Meanwhile, DHT protocols are based on nodes identifiers, which have no location semantic. Furthermore, there are different types of applications that are built successfully thanks to DHT systems (file and resource sharing, P2P web search engine). These applications are also useful in WSN. Hence applying DHT to ensure these applications would be beneficial to WSN.

4.2. Drawbacks:

There are some negative points that make the mapping of DHT over WSN a real challenge. Conventional DHT are ill-suited for WSNs. DHT function at application level for wired Internet. This means that DHT do not consider the physical structure of the underlying network. Indeed, two overlay neighbors are not necessarily two physical neighbors. An overlay hop can hence cost a lot of unnecessary physical hops. DHT routing tables should be always updated to avoid message misrouting and to ensure low path stretch. Thus, a lot of messages are sent in order to know which nodes are still available and which ones resigned. However, a big traffic of signaling messages degrades significantly the energy of sensors which leads to a dramatic reduce of the network lifetime. In addition, this maintenance traffic causes further network instability and reduces drastically the packet delivery ratio.

5. Highlights Existing Data Management And Routing Protocols That Combine DHT To WSN:

Because of the diversity of DHT based routing protocols in WSN, the need of a general taxonomy becomes a necessity. We present a classification of these routing protocols according to several criteria reflecting fundamental design and implementation choices:

i. Structure: Is there any distinction between the different nodes?
ii. Design options of DHT combination to WSN: How DHT is combined to WSN? Is DHT applied on the application layer and lying atop another routing protocol? Is it integrated or cross layered with the routing protocol?
iii. Identifier assignment: How the node obtains its identifier? Randomly? Depending on its physical neighbors’ identifiers? Do hierarchical routing protocols assign nodes identifiers hierarchically?
iv. Routing: How the message is routed? Is source route employed in routing? Are there any constraints that should be taken into account before forwarding a message to a
given node? At the case of hierarchical routing protocols, what is the relationship between master nodes and slave nodes? Is a slave node totally dependent to its master node? Can a slave node route a message to the destination or only master nodes have this ability?

v. Auto-deployment: Can DHT-based data management and/or routing protocols be applied in the case of network auto-deployment? If yes how do they bootstrap?

vi. Asymmetric link detection and utilization: Do routing protocols detect link asymmetry that can happen in WSN? If so, is such asymmetry taken into account in routing?

vii. Master nodes selection: In hierarchical routing protocols, how master nodes are selected? Does this selection depend on nodes’ features?

viii. Master nodes rotation: Do master nodes change over the time? If yes, what is the approach used to choose master nodes in each round?

Design options of DHT combined to WSN can be classified into three classes: layered, cross-layered and integrated1. The first contributions that combined DHT to WSN, applied DHT at the top of the application layer and used a MANET routing protocol at the network layer. This combination is called layered approach. Other protocols use a cross-layered approach. In this mapping design, the DHT system acts also at the application layer. However, contrarily to the layered approach, DHT interacts with the routing protocol. This means that there is a permanent communication between the application and the network layers. For example, the nodes identifiers needed in DHT data management are assigned depending on the nodes positions in order to facilitate the routing process: two virtual neighbors are also physical neighbors. Such combination reduces the implementation complexity of DHT combination to WSN.

In the integrated approach, DHT is directly and fully integrated with the routing process on the network layer or the Medium Access Control process on the link layer. In such a case, the nodes identifiers are assigned independently of any information related to other layers. These identifiers can be the hash of the nodes’ IP addresses or the nodes’ MAC addresses. In DHT-based protocols, each node should have a unique identifier. In several protocols, a node identifier can be randomly assigned, or be the hash of the node MAC address. In other protocols, physical neighbors are also virtual neighbors. This virtual relative position based technique avoids overhead resource usage caused by protocols that rely on under layer routing techniques.
Some hierarchical protocols use hierarchical identifier assignment so that each node has at a given level, a level-related identifier. The advantage of this method is that the node identifier reflects the level to which it belongs. However, hierarchical identifiers assignment process is more complex than that of flat identifiers.

6. Conclusions And Future Direction:

This article presented a survey on self-healing group key distribution schemes. A survey and analysis by the authors have shown that all existing schemes can be decomposed into three elements: selective key distribution mechanism, predistributed secret data management and self-healing mechanism. Techniques applied in each element were grouped into several categories. Each category was discussed by introducing example solutions. This way of categorizing the available schemes gives one the opportunity to establish a deeper insight into available self-healing key distribution approaches and to discuss the strengths and weaknesses of each one. SON has arisen as one of the key technologies for minimizing OPEX in future wireless networks. Within the SON aspects, self-healing is the one that, so far, has received the least attention. Information is fragmented over existing references and there is a lack of uniformity in the concepts and methods used. This article has presented a framework that unifies concepts and it has proposed a reference model for designing self-healing systems. Due to the fact that the field has been roughly explored, the open lines of research are numerous. In this article, the most important technical challenges have been outlined. There is no consensus on the best protocol since each protocol has its advantages and its limits. Each category of protocols is suitable for a specific applicability domain. Flat DHT-based protocols are mainly used in simple systems made of homogeneous sensors. These protocols can be applied in indoor surveillance for domotic applications. Hierarchical DHT-based protocols are more accommodated to complex systems composed of a great number of heterogeneous sensors. The applicability domains of these protocols are dependent to the relationship between master and slave nodes. Existing DHT-based routing and data management protocols present good performance in terms of data lookup and storage in WSN. However, sensors dynamism as well as asymmetric link detection and bootstrapping are rarely considered. Further researches are needed to address these issues.
References:


