International Journal of Engineering & Technology, 7 (4.20) (2018) 598-605

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

SPEEDA: A Secure Protocol and Energy Efficient for Data Aggregation in Wireless Sensor Networks

Sarah Wayzani [1], Chérif Diallo [2”]

[1,2] Laboratoire ACCA, UFR SAT, Université Gaston Berger, BP 234 Saint-Louis, Sénégal
”Corresponding author E-mail: cherif.diallo@ugb.edu.sn

Abstract

Wireless Sensor Networks (WSN) consists of a set of small and resources-constrained devices. They have received much attention over the last few years for the study and development of a plethora of potential applications. However, the common denominator of all applications of these sensor networks is the vulnerability of micro-sensors because of their limited material resources the most constraining of which is energy. In fact, wireless sensors are limited in terms of calculation, storage, battery, etc. Therefore, every possible solution that aims to conserve these resources is extensively sought. Thus, a great deal of researches has been conducted leading to an effective technique answering the established problem. This solution concerns the aggregation of data which is one of the techniques that is actually considered as an essential paradigm for sensor networks since it tends to save computation and communication resources. Data aggregation allows in-network processing which leads to lesser packet transmissions and reduces redundancy, and therefore, helps in increasing network's overall lifetime. However, sensor networks are usually deployed in unattended and hostile environments.
Thus, the designer should not only consider the limited resources of the sensor nodes but also the security threats that can occur in an easily accessible network to the attacker. In this paper, we will give an overview study of the existing data aggregation solutions in wireless sensor networks then, we will propose a new efficient and secure approach to aggregation which can be evaluated with specific criteria.

Keywords: *Wireless Sensor Network (WSN); Security; Aggregation; Routing; SPEEDA.*

1. Introduction

Since their inception, wireless telecommunications networks have been increasingly successful in industrial and scientific communities because of its many practical advantages: ease of deployment, ubiquity of information, and low installation costs. Thus, several derived architectures were born from the evolution of the wireless paradigm, among which we can mention: cellular networks (GSM), wireless local area networks and ad hoc networks (MANET i.e. Mobile ad hoc NETworks). Technological advances in the fields of microelectronics, wireless communications, coupled with miniaturization efforts and reduced production costs of electronic components, have enabled the development of new generation wireless networks. These offer a lot of advantages and one of them promises to revolutionize our life, our work and our way of interacting with the physical environment around us: it would be wireless sensor networks (WSN). These are ad hoc networks usually composed of a large number of communicating nodes and distributed over a given area in order to convey information to treatment points (e.g. Figure 1). This information comes from the collection of events occurring in the coverage area where the sensors are deployed. They often relate to environmental parameters such as temperature, atmospheric pressure, humidity, light intensity, seismic movements, noise, etc. Natural successors of ad hoc networks both in history and in the inheritance of techniques, wireless sensor networks share with MANETs (Mobile Ad hoc NETworks) several common properties such as lack of infrastructure, wireless communications, and the characteristics of self-configuration and spontaneity. However, they differ in their fields of application Unlike MANET networks, which have so far been modest in their success, wireless sensor networks have been successful in attracting large numbers of users and large companies such as IBM, Sun, Intel and Philips given their realism, their concrete contribution as well as the flexibility offered by their method of use.

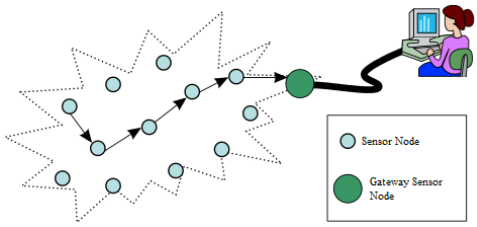


Fig. 1: Illustration of a Wireless Sensor Network (WSN)

However, it is not without saying that the WSN have generated a lot of interest in scientific research because of the problems they contain. Indeed, despite the miniaturization and the reduction of the cost of manufacture, the sensors are generally endowed with limited resources in terms of transmission power, processing capacity and storage of data and energy (e.g. Figure 2).
These constraints have influenced many of the domain's research issues. The durability constraint of the network is therefore a major concern given the often inaccessible nature of surveillance zones

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

(hostile areas). It is also difficult to imagine finding another source of energy than batteries. Itis for this reason that sensors are considered as autonomous devices and their lifetime is therefore equal to the service life of the network. As a result, energy conservation has become a preponderant performance criterion and arises first. The design of such networks must therefore be based on a modular architecture and its operation on a strategy ensuring network performance and extending its lifetime.

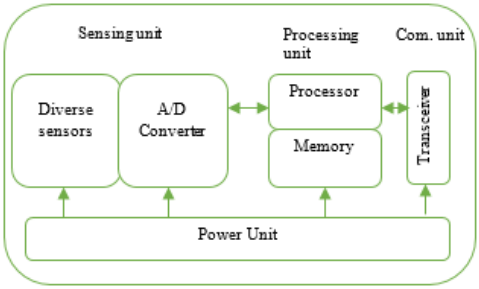


Fig. 2: Anatomy of a sensor node

From all these observations, research topics were created to find solutions for optimizing the energy resources of sensor nodes for a better longevity of the network. One of the proposed solutions then is the aggregation of data which is one of the most powerful techniques used for these purposes. Aggregation is a concept that allows sensor nodes to send their collected data to an aggregator node that will summarize them to reduce the number of transmissions and send them to the base station (e.g. Figure 3 & 4).
However, despite its status as a solution for the optimization of sensor resources, this technique can be a source of attacks and vulnerabilities because the many proposals made on this subject do not generally take into account the security problems of these networks. Among the many attacks [2] that exist we can mention the best known to destroy the efficiency, reliability, quality of service and security of applications: denial-of-Service attack, Sybil attack, selective forwarding attack, replay attack, stealthy attack, and node compromise attack.
Data aggregation and security do not go very well together because it should be noted that they have opposite objectives. Thus, the first

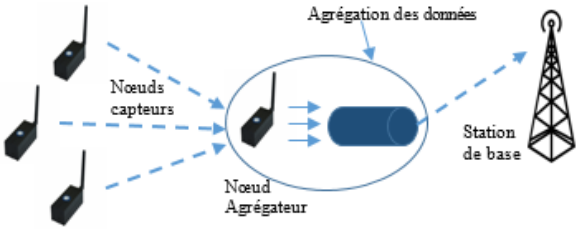


Fig. 3: Data aggregation process

tries to minimize the amount of data transmitted and the second adds a computation and communication load not insignificant to ensure the verification of some properties of security. On the other hand, since the sensors are endowed with limited computing resources, the implementation of the security primitives is a challenge and these limitations must be taken into consideration. The security requirements of data aggregation [3] are therefore summarized in these points:

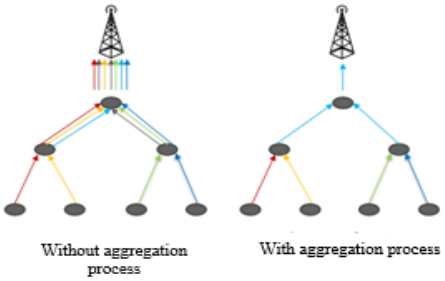


Fig. 4: Importance of data aggregation process in WSN

Data Confidentiality: ensures that the content of the information is not disclosed to an unauthorized entity.
Data Integrity: It ensures that the content of messages is not intentionally or accidentally altered during transmission.
Data Freshness: This property ensures that data flowing through the network is fresh or new.
Data Availability: Ensures that the network is ready to satisfy requests and that data is available.
Data Authentication: Allows a recipient to verify that received messages come from a trusted source.
In summary, data aggregation protocols must be designed in conjunction with security protocols, in order to provide a good compromise between the complexity of the overall protocol and the level of security provided, while maintaining an acceptable consumption of energy. Ensuring the security of aggregation is therefore a big challenge. This observation motivates us to study in Section 2, the security of data aggregation in WSN in existing solutions in order to bring out their limits and propose in Section 3, a new secure algorithm that can guarantee an optimal level of security and a better energy consumption for a better longevity of the network.

2. Evaluation of existing solutions of data aggregation

Without data aggregation in WSN, sensor nodes route all collected data directly to the base station. This data often redundant, can lead to several disadvantages: Redundant data makes no sense for the application, the chances of network congestion increase dramatically, network capacity is wasted, energy consumption is increasing. Thus, considered as a preventive approach, the operation of aggregation performs additional processing on the data captured from the environment. An aggregator node combines data from multiple nodes into meaningful information. This significantly reduces the amount of data transmitted by consuming less power and thus extending network lifetime. Data aggregation protocols must necessarily satisfy security needs. There is a multitude of proposals for secure solutions for data aggregation and each model proposed, has its own challenges that must be considered carefully.

2.1. Przydatek and al. scheme

One of the first solutions proposed by the scientific community is Przydatek et al's solution [4] which ensures the security of data aggregation by using the calculation of MAC and the Merkle hash tree with TESLA between simple sensors and their aggregators without data encryption. This schema is able to verify the validity of the result of the aggregation, but without any other action to remove or isolate the node that resulted in inconsistency in the results. Also, the use of two safety mechanisms namely the calculation of the MAC and TESLA can generate an overconsumption of energy and slowdown the lifespan of the sensors and the network in general.
Thus, it would be preferable to act in this way and introduce a mechanism for isolating a compromised node but also reduce the computational load at the sensor nodes while ensuring good security.

2.2. Mahimkar and Rappaport's scheme

The scheme of Mahimkar and Rappaport [5] is similar to [4] and provides an additional security service: confidentiality of data. It uses elliptic curve cryptography to encrypt the data sent to the clusterhead, Merkle's digital signature concept to sign aggregation results, and Merkle hash tree to verify the integrity of the aggregation results declared once verification of the signature failed. The clusterhead is based on the calculation of the average of the physical phenomena detected. It then broadcasts the average for all members of the cluster to allow them to compare their physical phenomena with the average. If the difference is less than a certain threshold, the node creates an average partial signature using its share of the cluster's private key, and then sends it to the clusterhead. The clusterhead combines these signatures into a full signature and sends it with the average value to the base station. Despite enhanced security by using in addition MAC and the Merkle tree, the cryptography of the elliptic curve, a dense exchange of verification information can saturate the storage memories and impact the energy of the sensors. Also, with this scheme, the clusterhead transmits 2 times more data because of the signatures received from the simple nodes that accompany the aggregated data to the base station.
We will therefore avoid in our scheme overloading the simple sensor nodes of calculations and transmissions by dispensing them from the verification phase.

2.3. Sanli and al. Scheme

Sanli et al. As for them, they propose a schema [6] based solely on references, and encrypt the results of the aggregation by applying a variable security force at different levels of the hierarchy of aggregators. So this scheme depends only on encryption (RC6 encryption block) to provide a precise and secure aggregation of data. Also, the security primitives, used here, are not practical for use in multi-constraint devices such as sensor nodes. In sum, the security provided by this scheme is not very robust because the number of laps for RC6 encryption can be as low as 10 rounds once the aggregator node is 10 hops away from the base station.
We can deduce that RC6 encryption is not an adequate solution for good data security in these networks.

2.4. Hu and Evans Scheme

The scheme proposed by Hu and Evans [7] meanwhile, achieves resilience against the compromise of nodes by delaying aggregation and authentication at the higher nodes. Thus, the parent nodes must keep the physical phenomena's data in memory to authenticate them once the shared key is revealed by the base station. This schema is able to detect a single node compromise, but with no other action to remove or isolate this compromised node. Even worse, once a 'grandfather' node detects a node compromise, it cannot decide whether the compromised node is its child or grandchild. The system also suffers from memory overhead due to delayed authentication and the need to keep in memory the data received by the parents to be authenticated later. Finally, parents lose energy by listening to some of the revealed keys that are not intended for them. Like schema [4], this schema also encounters difficulties in isolating a compromised node and storing the received data for an authentication requirement does not prove to be a good approach. In addition, the periodic transmissions performed by the base station to the sensor nodes deplete them and affect their energy consumption. Thus, we will take into account these limits by proposing in our diagram, a mechanism of detection and isolation of a compromised node but also we will avoid a storage overload in the sensor nodes by facilitating the authentication of the data by mechanisms like the calculation of the MAC.

2.5. Jadia and Mathuria Scheme

Jadia and Mathuria [8] extended the Hu and Evans program [7] to improve security services by adding data privacy. The aggregation phase is thus performed in the same way as the Hu and Evans scheme, with the exception of two differences: The leaf nodes calculate two MACs on the encrypted data, the leaf nodes encrypt their physical phenomena before sending them.
Jadia and Mathuria have therefore added data confidentiality to the security services provided by the Hu and Evans system, but their schemas have the same weaknesses. However, memory overhead weakness is not visible in this scheme because it uses paired keys and does not need to keep copies of MAC information until the base station reveals temporary keys.
The computation of two MAC per node, certainly reinforces the security of the schema, but increases the computing load and the volume of data transmitted. We will try in our diagram, using the computation of a MAC, to guarantee a similar level of security while offering a better advantage vis-à-vis the optimization of the nodes.

2.6. Westhoff and al. Scheme

Westhoff et al. [9], for their part, introduce secure aggregation on the encrypted data by relying on a homomorphous additive and multiplicative encryption scheme (Privacy Homomorph PH) by promoting end-to-end encryption. The security primitive used in this scheme to defeat a Type III adversary is PH. This primitive is impractical for use in constraint devices, such as sensor nodes, because of its high computing cost [9]. Westhoff and al. also aimed to defeat passive adversaries who spy on communication between sensor nodes, aggregators and the base station. Since in this scheme an adversary is able to compromise the aggregation nodes, it can start the replay attack by replacing the old valid encrypted messages until the encryption keys of the leaf nodes have been updated / renewed. Also, once an aggregator is compromised, the opponent is easily able to launch the selective retransmission attack.
The compromise of a node in this diagram leads to a possibility of a replay attack that could be taken into account by introducing for example a time phenomenon that can attest the non-freshness of the data. Also, given the high cost of the calculation of the PH primitive used, we will be able to spread some end-to-end encryption by promoting a partial encryption (between aggregators and base station) to optimize the resources of the sensors.

2.7. Yang and al. Scheme

The scheme [10] proposed by Yang et al. uses a probabilistic clustering technique that partitions the nodes of a tree topology into multiple logical groups and then an aggregation based on hop-by-hop commitment is performed in each group to generate a group aggregate. The base station then identifies the suspect groups based on a set of group aggregates. Each suspected group participates in a certification process to prove the validity of its group aggregation result. To do this, Yang et al. use an adaptive Grubbs test [11] to check the anomaly in the aggregation results before accepting or rejecting them.
This adaptive test can be attacked when certain nodes are compromised. This will affect the decision of the base station and force it to start the process of certifying honest groups instead of malicious groups. In addition, invalid aggregation results are documented (or verified) by a centralized audit that results in high communication costs.

2.8. Chan and al. Scheme

In [12], Chan et al. propose a schema that aggregates data according to a hierarchy of several aggregators. The aggregation process is initiated by the base station that broadcasts across the network, triggering the construction of an aggregation tree resembling a hash

tree. Each node sends its data to its father who performs the aggregation until all the aggregation results arrive at the base station. This makes a final aggregation and distributes the result to all the nodes for a verification phase. The base station finally receives and accepts the aggregation result by verifying that all sensors have sent an authentication code. In this diagram, we will be able to judge the verification phase of the aggregated data as a robust process but which, however, makes the sensor nodes work 2 times more. In fact, the aggregator node first sends its data in addition to the data that it receives and aggregates, and then receives again the final aggregate broadcast by the base station in order to verify the taking into account of its data. In the final result and finally retransmits the whole to the base station accompanied by an authentication code.
This diagram, although robust, greatly affects the energy of the nodes because the number of transmission carried out is important. So it could not be suitable for large networks. Also, there are several aggregation operations before arriving at the base station that require a large storage memory and that could affect the life of the batteries. Thus, we can promote our scheme with fewer aggregator nodes and with a verification phase lighter but offering the same level of security.

2.9. Cam and al. Scheme

Cam et al. proposed in [13] a secure aggregation scheme (ESPDA) applicable to hierarchical networks organized in clusters. The protocol is based on code patterns to ensure data aggregation. The codes patterns are representative data extracted in the measurements of the nodes and make it possible to characterize these measurements. Nodes generate and send pattern codes to clusterheads. These clusterheads examine the code patterns and allow only one node to send them the data representing the code patterns they end up transmitting to the base station.
We will be inspired by a similar scheme to put in place a mechanism that does not collect all the data from all sensors in order to reduce the number of transmissions in the network while having accurate and reliable results.

2.10. Ozdemir Scheme

The protocol proposed by Ozdemir in [14] (CDAP) is also based on homomorphic encryption like [9] to ensure the security of aggregated data. The author claims that symmetric homomorphic encryption used in some protocols like [15] has security problems because of the unique key shared between the nodes. For this, it uses an asymmetric homomorphic encryption and because of the additional costs of computation, the scheme uses particular nodes called AGGNODE which hold sufficient resources to carry out the aggregation.
This diagram proves that it cannot be deployed in any network due to the need for powerful sensor nodes to perform asymmetric homomorphic encryption. So we can rule out the possibility of using such a process. Also, we will avoid using symmetric homomorphic encryption because of the security problems it contains because of the unique key shared between the nodes.

2.11. Onen and al. Scheme

Onen et al. [16] propose a secure level aggregation scheme based on additive homomorphic encryption. A pseudo-random key distribution mechanism is set up allowing nodes to share symmetric keys. The authors combine homomorphic encryption with a multiple encryption process. In the schema, the network is structured as a tree in which each leaf node sends its measurements to its father who extracts the shared key with his son, adds his measurements and the key he shares with his father in the tree before sending the result to this one. This process is repeated until all data arrives at the base station that performs the result authentication. In this scheme, the intermediate nodes cannot read the data encrypted by the simple nodes, which is all its strength. However, shared symmetric key sizes occupy a large part of the sensor memory and greatly increase the size of the transmitted data. Also, the authentication of the result is performed only from the base station which can be a security breach.
Given the limitations of this scheme, we will try to work with small encryption keys while always ensuring a good level of security but also we will implement an authentication process closer to the sensors, so that final result is safer and more reliable.

3. SPEEDA: secure protocol and energy efficient for data aggregation

After the analysis done on the methods and schemes dealing with the security of data aggregation in WSN, we will now in this part, propose a diagram that covers the few limits noted during this analysis. Thus our scheme, SPEEDA, will aim to bring a new touch to improve the security of aggregation in wireless sensor networks.

3.1. Description of the SPEEDA proposed scheme

To define our protocol, we have taken into account the limited resources of the sensors. Thus, we will use symmetric cryptography with which we can optimize the energy and storage memory of the nodes that will encrypt the data with a single key whose size is acceptable for the WSN. Our protocol uses a two-level architecture, in which aggregators aggregate data from their member nodes and forward them to the base station. In other words, the aggregated data will go through a single aggregator node, this means that the transmissions between each aggregator (clusterhead) and the base station will be done in one hop and will only be decrypted once they have arrived at the base station. Our protocol can be summarized as follows. Nodes X_i collect the environmental data and send it to their clusterhead accompanied by the transmission date T_i, their identification number ID_i and a MAC_i (Message Authentication Code) calculated from a common key shared with their clusterheads. The computation of the MAC performed at each node will ensure the authentication and integrity of the data transmitted to each clusterhead. Upon receipt of the data, each clusterhead CH_i, proceeds to a phase of verification of the data received and applies an aggregation function f which can be for example the average or the sum of the measurements PP_i collected.

3.2. Details of the proposed scheme

Our SPEEDA protocol is based on the following assumptions:
The sensor network is static (non-mobile nodes).
The sensor nodes are homogeneous: similar in their processing, communication, energy and storage capacity.
The deployment is random: the neighbours of any node are not known before deployment
The base station has no constraints on computing, storage, and cannot be compromised
The communication channels are bidirectional: in other words, if a node *u* can receive a message from the node *v* then *u* can send a message to *v*.
The network structure is clustered
The routing protocol used is APTEEN
The pre-distribution of keys is probabilistic: for networks with a large number of nodes, the probabilistic approach is very efficient. With this approach, the existence of one or more common keys between any nodes is guaranteed with a certain probability. The basic idea is to distribute randomly before deployment to each sensor a subset of keys *k* from a set of keys *m*. The number of keys of the subset *k* is chosen such that two random subsets of *m* will have a certain probability *p* of having at least one common key. This mechanism supports scaling [17], [18].
The cryptography used is based on symmetric Encryption: The same key is used between the aggregator nodes and the base station

to encrypt and decrypt messages exchanged between them using a symmetric encryption algorithm.
The cryptography is also based on a MAC: It is a cryptographic system that verifies the authenticity of the origin and the integrity of the information received. It is produced by symmetric key hash functions and sent with the data. The receiver recalculates the MAC code with the same key and compares it to the received code. If both codes are identical, the source is authenticated and the data is saved, otherwise the source is suspected and the data is considered corrupted.
The main idea of SPEEDA is to distribute a range of keys *m* to each node of the network before deployment. In other words, the algorithm makes it possible to create a key management scheme based on the probability of distributing for each node a key range, resulting from a finite set *U*. Also, any two nodes will be able to exchange secure messages if they have a common key. This key will also be used to encrypt the data during a communication between a clusterhead and the base station. Another key, common to all nodes, is also shared by all members of the network. This key will allow each elected clusterhead to communicate and warn other nodes in the cluster as well as the base station of its election. Thus, SPEEDA takes place in 3 main phases: initial phase, data transmission and data aggregation.

3.2.1. The SPEEDA initial phase

The SPEEDA initial phase consists of the following operations:
Pre-distribution of keys: A large set *U* of keys is generated. For each node, *m* keys are randomly selected from the set and an identical key k_c is also assigned to all nodes. These *m* keys are stored in the node's memory and form the keyring of the node. The keychain is then made up of small keys belonging to a specific range, so as to follow a certain logical sequence in the same cluster *C*. The number of keys |U| of the set is chosen in such a way that two random subsets of *U* of size *m* will have a certain probability *p* of having at least one key k_i in common.
The discovery of the shared key: The nodes discover their neighbors and especially those with whom they are able to communicate in a secure way because they have an identical key in their respective keyring (e.g. Figure 5). The shared key becomes the session key of the link between the two nodes.
The key way establishment: After the discovery phase of shared keys, the network becomes a connected graph formed of some secure links. Also, the nodes not sharing a common key with their neighbours can use the key kc identical to all the nodes to establish a secure link of communication.

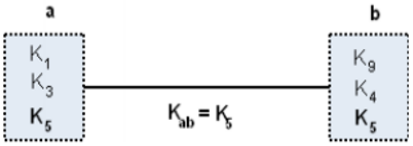


Fig. 5: Discovering the shared key for establishing secure communications

Key revocation: The revocation of a compromised node is done by the elimination of its keyring. For this, a controller node (which has a high connectivity usually the clusterhead) announces a simple revocation message containing a signed list of the keys of the compromised node so that they are removed from the keyrings of the other nodes. Some links will disappear because of the deletion of keys of the compromised node which requires a reconfiguration of these links (by the discovery of shared keys or the key way establishment).

3.2.2. SPEEDA data transmission phase

Each node X_i collects and sends its measurements PP_i to the aggregator node CH_i. For transmission, the node X_i associates with the data PP_i collected, a MAC calculated from one of the keys of the keychain also accompanied by the date of transmission T_i (date + Time) and its identifier ID_i (e.g. Figure 7). Each CH_i locally maintains a table containing all the identifiers of each sensor in its cluster.

The sent message then becomes:

$$C_{k_{X_I}}^{CH_i} = (PP_{X_i} \| T_i) \| MAC_{X_i}\left(K_{X_i}^{CH_i}\right) \| ID_{X_i} \quad (1)$$

Where PP_{Xi} are the collected data (Physical Phenomena), T_i the date of data transmission PP_i, MAC_{Xi} the MAC of the node X_i calculated using the key $K_{X_i}^{CH_i}$ shared with the clusterhead CH_i and ID_{Xi} the identifier of the node X_i.

3.2.3. SPEEDA data aggregation process

To perform the aggregation operation, the clusterhead must first receive the data transmitted by its member nodes to compute the useful information using an aggregation function such as: average, maximum, minimum etc. This reception is followed by determining the total number of messages received to verify that there are no intruders (e.g. Figure 8). Thus, on receiving the data $C_{k\ X_i}^{CH_i}$ of each initial node X_i the aggregator node first verifies the date the message was sent.
If the difference between the current time and the send time exceeds a certain value δ it ignores the message and informs the other nodes of the cluster that the node ID_i must be isolated by deleting the key K_i shared with this same node. Otherwise, the aggregator continues to check the identifier of the node X_i. If it is incorrect, it ignores the message and also informs the other nodes of the isolation of the node in question. Otherwise, it checks the MAC with the common key shared with the X_i node. If it is correct, it applies the aggrega tion function f and then encrypts the result with the common key shared with the base station. Finally, it sends the encrypted result to the base station.

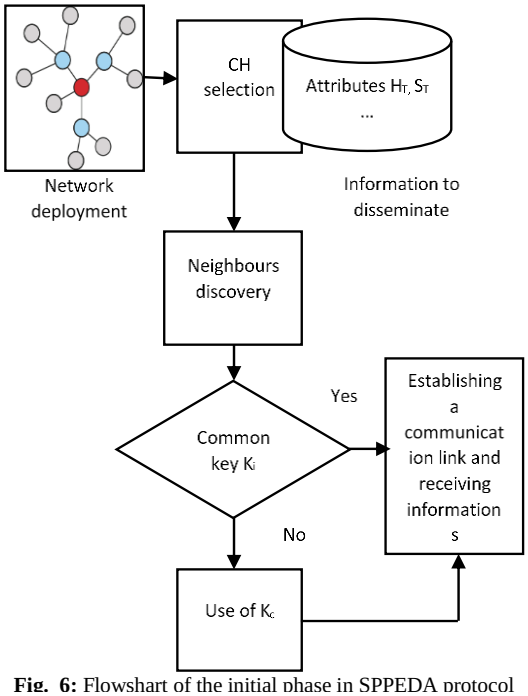


Fig. 6: Flowshart of the initial phase in SPPEDA protocol

4. Evaluation of the proposed scheme

We will analyze the security of our SPEEDA protocol by focusing on security services that any secure aggregation scheme must satisfy and on its resistance to attacks on data aggregation.

4.1. SPEEDA Insured security services

Data aggregation, while a very efficient technique in wireless sensor networks, often faces security issues. Thus, the main purpose of securing data aggregation is to ensure that the resulting data from the aggregation process is reliable, which allows for good interpretation when it is transmitted to the human operator. In the proposed scheme, we have tried to take into account all the security requirements of the aggregation process, but we have relied more on integrity and authentication because they are considered very important when it comes to collecting environmental phenomena at specific times. Thus, our schema guarantees the integrity of data. Indeed, only two kinds of data transmission are possible: nodes X_i to the aggregators CH_i and aggregators CH_i to the base station S.

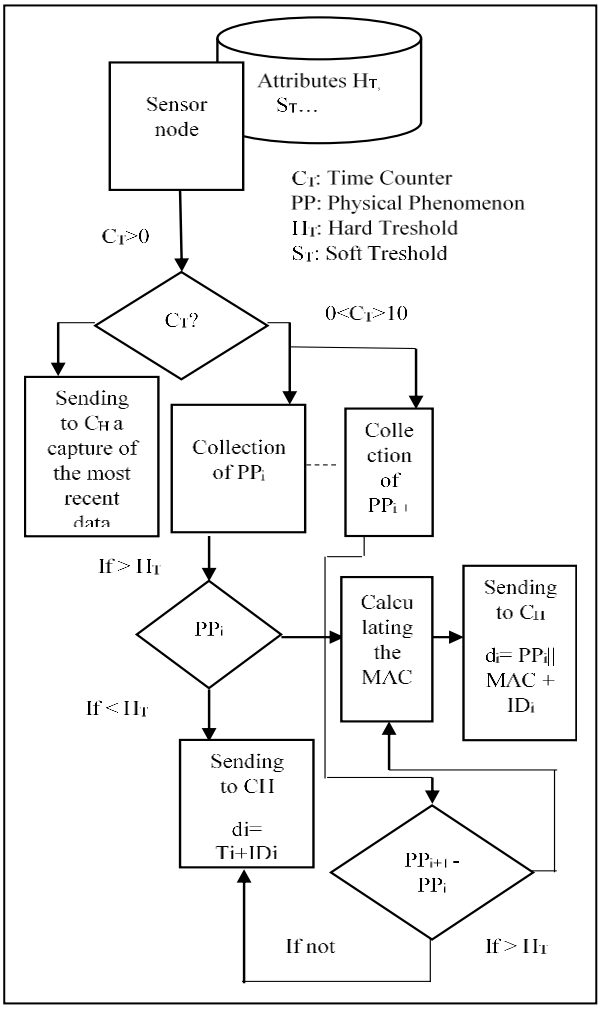


Fig. 7: Flowchart of periodic data transmission (from a single node to its clusterhead) in SPEEDA protocol

First, we assume that an opponent is in the vicinity of cluster C and listens for traffic between node X_i and aggregator CH_i.
It is also assumed that the data sent by the collecting nodes is reliable. After collecting the data PP_i, the node X_i sends to CH_i the message $C_k{}^{CH_i}_{X_i}$. We assume that the attacker intercepts the message D_i and whose purpose is to make the aggregator node CH_i accept erroneous data. To succeed, the attacker must modify the PP_i. With the computation of the MAC carried out at the level of the collector node, without knowledge of the key $K_{X_i}^{CH_i}$ shared between the node X_i and the aggregator CH_i, there is no other way for the attacker to reach the end of his action.
In a second case, we assume this time that the attacker is between the aggregator node CH_i and the base station S. For this type of link, the two parties communicate using the key k_i which encrypts the data to the base station, and the link is secure enough to ensure data integrity.

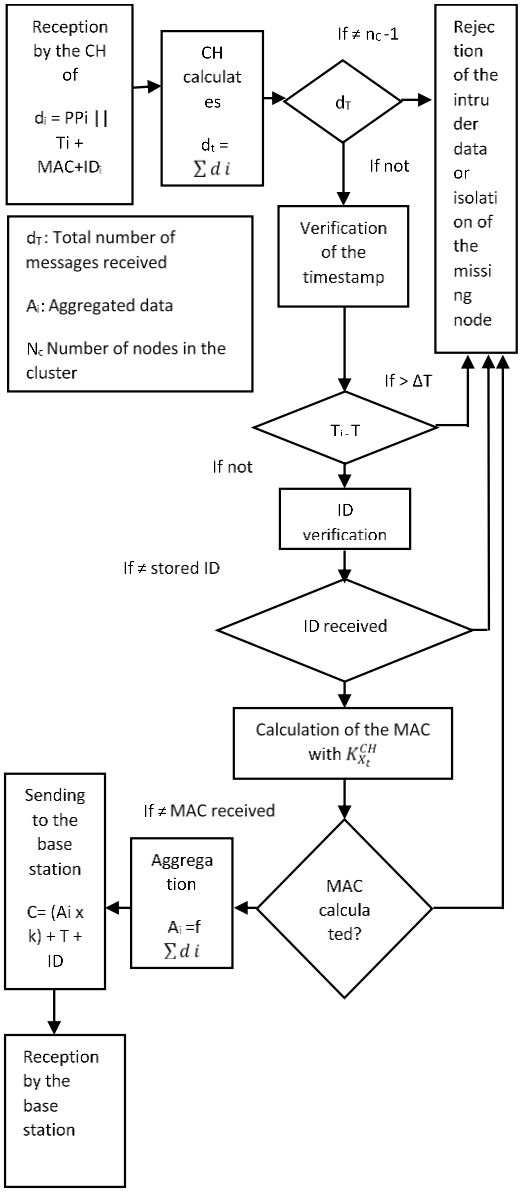


Fig. 8: Flowchart of data aggregation process in SPEEDA protocol

From this integrity analysis, it follows that in our scheme, integrity can only be broken by an attacker who will have the ability to physically capture a node and extract the cryptographic keys. On the other hand, we can say that our schema perfectly guarantees the authentication of data and nodes. We adopted a technique that combines these two types of authentication. Each sensor has an identifier ID_i and each aggregator (clusterhead) CH_i knows the identifiers of all the nodes of its cluster. At the reception of a message, each aggregator will be able to ensure that the data received comes from reliable sources by verifying that the identifier contained in the message is correct. Thus, the proposed schema allows the aggregator to work only on correct data. An attacker cannot add malicious nodes until he has physically captured an aggregator node and updated his information. Data confidentiality is also emphasized when transmitting aggregated data but only between clusterheads and base station. Indeed, the data is encrypted after aggregation by the clusterheads before being sent to the base station, so these data streams do not contain any information in clear. Therefore, our attacker will not be able to read the information he is intercepting. With our schema, the data will only be revealed to attackers who first know

of a secret key, which means that it must first succeed other attacks. Our schema also ensures data freshness by comparing at the clusterhead level, the difference between the date a message was sent by a collector node and the current date at the set stamp. In this way, an attacker will not be able to replay an old message, which ensures the freshness of the data.

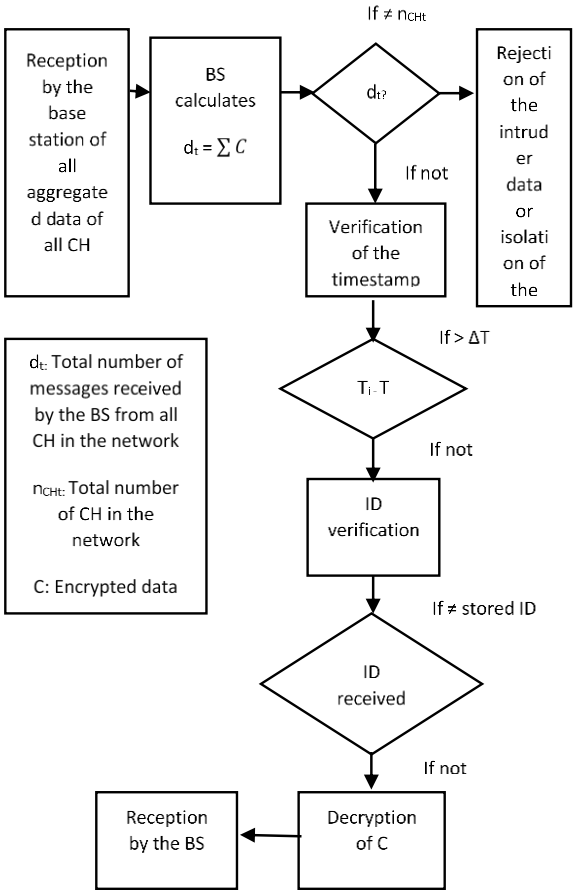


Fig. 9: Flowchart of the reception of aggregated data by the base station and their verification in SPEEDA protocol

4.2. SPEEDA resistance against attacks

We will now analyse the security of our schema by considering here, the most known attacks of sensor networks. Among these attacks, we can mention the attack by replay that is impractical in our scheme. Indeed, to fight against this type of attack, each aggregator node CH_i including the base station S compares the difference between the date of sending the message of a node T_i and the current date to the stamp defined δT. If the difference is greater than δT, then the replayed message is rejected. Another known attack of these networks is the Sybil attack, in which an attacker can alter the system by "endorsing" several identities, which makes it possible to create several routes through the malicious node, which are actually only one path. This attack can be circumvented given the need to hold in addition to an identity, a common key k_i between two nodes for the establishment of a communication link.

Also, the base station as well as the aggregators hold the identifiers of all nodes of the network. It is therefore impossible for the attacker to present several identities and to have his data accepted at the base station or an aggregator, which reinforces the security against this type of attack. Thanks to the adopted routing protocol (APTEEN [19]), our schema integrates a dynamic cluster formation in which the clusterheads are chosen randomly and rotatably. This allows our scheme to be resilient to node compromise attacks. At some point, if an attacker compromises an aggregator node, it is not certain that the same node is always aggregator at another date. Our protocol is based on an initial key k_c loaded in each of the nodes of the network. We suppose that to compromise a node, the adversary may require a minimal time T_{min}: this is the time to connect a serial cable and copy the contents of the memory of the compromised node. We use this (trusted) time to allow two neighbouring nodes that do not have a common key k_i to establish, in a secure manner, a session symmetric key from the initial key k_c. After T_{min}:, the key k_c is transferred to the volatile memory of the node. However, the difficulty is that there is always communication with some nodes sharing with the compromised node a common key k_i. The rotation and selection of a clusterhead are tracked by the clusterhead information of the number of nodes present in the cluster. Thus, to combat the selective retransmission attack, the clusterhead expects to receive *n* transmissions (where n is the number of nodes in the cluster). If it does not receive *n*, it automatically detects the attack and isolates the affected node, by checking the unused key for the computation of the MAC of the node to be isolated. There is also in WSN a sinkhole attack in which the malicious node tries to pull as many paths to it as possible allowing control over most data circulating in the network and this by presenting optimal routes.

Our schema is optimal for this type of attack because the routes are established thanks to the common keys shared between the nodes, which are chosen by a probabilistic key distribution law before deployment. In addition, the chosen network structure (in cluster) allows a regular communication that goes from the nodes to the clusterhead and does not necessarily favour the change of trajectory of the communications. Finally, the well-known denial of service attack can be circumvented by adopting the clustered structure and information held by the clusterhead that can easily detect a malfunction in the network. However it is important to note that this attack takes several forms depending on the layers of the network (flooding, jamming, spying etc.) and therefore, it would be wise to affirm the non-vulnerability against all types of attacks.

5. Conclusion

Wireless sensor networks are undoubtedly a technology that plans to assist us in the coming years as part of our daily lives. However, there are still many issues that need to be addressed for efficient operation [20] of these networks in real-world applications. Among these fundamental issues we mention the problem of energy saving which generally impacts on the lifetime of the network and which is an absolute necessity to which adequate solutions must be proposed. Thus, one of the most effective techniques that respond to this problem is aggregation, which promotes a reduction in the data circulating in the network by specific functions and thus contributes particularly to the energy optimization of the sensors and that of the network in general. Also, because of their deployment in hostile environments, these networks often face several threats and vulnerabilities that can paralyse their activities. Therefore, it would be necessary to take into account the security aspect in the aggregation techniques, aspect often put out in the numerous solutions proposed. Thus, we have set for mission in this paper, the proposal of a new diagram of data aggregation taking into account the security needs of sensor networks but also contributing to a strong energy optimization in these networks. The various existing solutions answering in a certain way the security problems of the aggregation were noted and classified for a better study in order to better guide us and to fix the objectives to be reached. Thus, the main objectives set are the security of the aggregation by ensuring the authentication and the integrity of the aggregated data arriving at the base station but also the energy optimization of the sensors which will allow a better lifetime of the network. Finally, it would be important now to implement our solution in simulators dedicated to wireless sensor networks and compare it to other protocols that support the energy and safety constraint according to specific performance criteria as in [21].

Acknowledgement

The authors would like to thank:
Direction des Technologies de l'Information et de la Communication (DTIC) au Ministère de la Communication, des Télécommunications, des Postes et de l'Economie Numérique (MCTPEN) du Sénégal.
Centre d'Excellence en Mathématiques, Informatique et TIC (CEA-MITIC); UFR Sciences appliquées et de Technologies (UFR SAT), Université Gaston Berger, Saint-Louis, Sénégal. (http://www.ceamitic.sn/).

References

[1] Diallo, C. (2017), Deployment Strategies and Clustering Protocols Efficiency. *In Sensors & Transducers*, Vol. 213, Issue 6, June 2017, pp. 9-23, [International Journal ISSN: 2306-8515, e-ISSN: 1726-5479].

[2] Diallo, C., Sawaré, A., Sow, M., T. (2017), Security Issues and Solutions in Wireless Sensor Networks. In *International Journal of Computer Science and Information Security, IJCSIS*, ISSN 1947-5500, VOL.15 No.3, March 2017.

[3] Diallo, C. (2017), Security Issues and Solutions related to Data Aggregation Process in WSN. *In International Journal of Computer Science and Network Security, IJCSNS*, ISSN 1738-7905, VOL.17 No.4, pp.59-71, April 2017.

[4] B. Przydatek, D. X. Song, and A. Perrig. SIA (2003), Secure Information Aggregation in sensor networks. In Proceedings of the *1st International Conference on Embedded Networked Sensor Systems, SenSys'03*, Los Angeles, California, USA, pp. 255-265, Nov. 2003

[5] A. Mahimkar and T. S. Rappaport. SecureDAV (2004), A secure data aggregation and verification protocol for sensor networks. In Proceedings of the *Global Telecommunications Conference (Globecom'04)*, Dallas, USA, vol. 4, pp 2175-2179, Nov. 2004.

[6] H. O. Sanli, S. Ozdemir, and H. Cam. SRDA (2004), Secure reference-based data aggregation protocol for wireless sensor networks. In Proceeding of the *60th IEEE Vehicular Technology Conference, VTC'04*, Los Angeles, USA, vol. 7, pp. 4650-4654, Sept. 2004.

[7] L. Hu and D. Evans (2003), Secure aggregation for wireless network. In Proceedings of the *2003 Symposium on Applications and the Internet Workshops, SAINT'03*, Orlando, FL, USA, pp. 384-394, Jan. 2003

[8] P. Jadia and A. Mathuria (2004), Efficient secure aggregation in sensor networks. In Proceedings of the *11th conference on High Performance Computing, HiPC'04*, Banga-lore, India, vol. 3296 of Lecture Notes in Computer Science, pp. 40-49. Springer, Dec. 2004

[9] D. Westhoff, J. Girao, and M. Acharya (2006), Concealed data aggregation for reverse multicast traffic in sensor networks: Encryption, key distribution, and routing adaptation. *IEEE Transactions on Mobile Computing*, 5(10):1417-1431, 2006.

[10] Y. Yang, x. Wang, S. Zhu, and G. Cao (2008), SDAP: a secure hop-by-hop data aggregation protocol for sensor networks. In *ACM Transactions on Information and Systems Security*, Vol. 11, No. 4, Article 18, July 2008.

[11] F. E. Grubbs (1969), Procedures for detecting outlying observations in samples. Techno metrics, 11(1):1-21, 1969.

[12] H. Chan, A. Perrig, and D. Song (2006), Secure hierarchical in-network aggregation in sensor networks. In Proceedings of the *13th ACM Conference on Computer and Communications Security, CCS'06*, pp. 278-287, New York, NY, USA, 2006.

[13] H. Cam, S. Ozdemir, P. Nair, D. Muthuavinashiappan, and H. O. Sanli (2006), Energy efficient secure pattern based data aggregation for wireless sensor networks. *Comput. Commun*. 29(4):446-455, Feb. 2006.

[14] S. Ozdemir (2007), Concealed data aggregation in heterogeneous sensor networks using privacy homomorphism. *IEEE, 8 2007.*

[15] J. Girao, D. Westhoff, and M. Schneider (2005), Cda: Concealed data aggregation for reverse multicast trafic in wireless sensor networks. In *IEEE International Conference on Communications (ICC2005)*, Seoul, Korea, May 2005.

[16] M. Onen and R. Molva (2007), Secure data aggregation with multiple encryption. In *Koen Langendoen and Thiemo Voigt, editors, EWSN, vol. 4373 of Lecture Notes in Computer Science*, pp. 117-132. Springer, 2007.

[17] L. Eschenauer and V. D. Gligor (2002), A key-management scheme for distributed sensor networks. In Proceedings of the *9th ACM conference on Computer and communications security*, pp. 41-47, 2002

[18] H. Chan, A. Perrig, and D. Song (2003), Random key predistribution schemes for sensor networks. In *Proceedings of the Symposium on Security and Privacy*, pp. 197–213, IEEE, 2003.

[19] Manjeswar, A. Agrawal, D.P (2007), APTEEN: A hybrid protocol for efficient routing and comprehensive information retrieval in wireless sensor networks. In Proceedings of *2nd International Workshop on Parallel and Distributed Computing Issues in Wireless Networks and Mobile Computing*, Fort Lauderdale, FL, USA, Apr. 2002; pp. 195-202

[20] Sow, M., T., Diallo, C. (2017), Energy over-consumption induced by securing network operations. Proceedings of *2nd IEEE International Conference on Frontiers of Sensors Technologies (IEEE-ICFST 2017)*, Shenzhen, China, Apr. 2017, ISBN: 978-1-5090-4858-8/17/ c 2017 IEEE, pp. 154- 160 (2017)

[21] Diallo, C., Sow, M., T. (2017), LQI-DCPSec: Secure Distributed d-Cluster Formation in Wireless Sensor Networks. In Proceedings of the *10th EAI International Wireless Internet Conference, WiCON 2017*, Tianjin, China, December 2017. Published by Springer.