Secure Data Aggregation in Wireless Sensor Network: a survey

Hani Alzaid **Ernest Foo** **Juan Gonzalez Nieto**

Information Security Institute
Qeensland University of Technology,
PO Box 2434, Brisbane, Queensland 4001,
Email: halzaid@isi.qut.edu.au , {e.foo,j.gonzaleznieto}@qut.edu.au

Abstract

Recent advances in wireless sensor networks (WSNs) have led to many new promising applications including habitat monitoring and target tracking. However, data communication between nodes consumes a large portion of the total energy consumption of the WSNs. Consequently, data aggregation techniques can greatly help to reduce the energy consumption by eliminating redundant data traveling back to the base station. The security issues such as data integrity, confidentiality, and freshness in data aggregation become crucial when the WSN is deployed in a remote or hostile environment where sensors are prone to node failures and compromises. There is currently research potential in securing data aggregation in the WSN. With this in mind, the security issues in data aggregation for the WSN will be discussed in this paper. Then, the adversarial model that can be used in any aggregation scheme will be explained. After that, the "state-of-the-art" proposed secure data aggregation schemes will be surveyed and then classified into two categories based on the number of aggregator nodes and the existence of the verification phase. Finally, a conceptual framework will be proposed to provide new designs with the minimum security requirements against certain type of adversary. This framework gives a better understanding of those schemes and facilitates the evaluation process.

Keywords: security, aggregation, wireless sensor networks, survey.

1 Introduction

The WSN is defined as highly distributed networks of small, lightweight wireless nodes, deployed in large numbers to mentor the environment or system by the measurement of physical parameters such as temperature, pressure, or relative humidity (Murthy & Manoj 2004, p 647). Sensor nodes are deployed in large numbers and they collaborate to form an ad-hoc network capable of reporting to a data collection sink. Recently, WSN networks have been used in many promising applications including habitat monitoring (Mainwaring et al. 2002) and target tracking (He et al. 2006). However, WSNs are resource constrained with limited energy lifetime, slow computation, small memory, and limited communication capabilities. The current version of sensors such as mica2 (Corporation 2006) uses a 16 bit, 8MHz Texas Instruments MSP430 microcontroller with only 10 KB RAM, 48KB Program space, 1024 KB External flash, and is powered by two AA batteries. Therefore, the energy impact of the added security feature should be considered when implementing a cryptographic technique for securing data aggregation in the WSN. For example, data authentication in TinyOS increases the consumed energy by almost 3% while data authentication and encryption puts 14% (Guimarães et al. 2005). Furthermore, the embedded processors in sensor nodes are generally not as powerful as those in the nodes of a wired network. As such, complex cryptographic algorithms are impractical for WSNs.

Not only the resources limitations affect the WSN performance but the deployment nature does also. Most of the WSNs are deployed in remote or hostile environments and then nodes cannot be protected from physical attacks since anyone can access the deployment area. Moreover, the only way to manage and control the network is via wireless communication which makes any physical operation such as battery replacement difficult. Another factor that affects the WSN performance is communication instability. For example, if two sensors that have same aggregator node start sending packets at the same time, conflicts will occur near the aggregator node and the transfer process will fail. In addition, packets might get dropped at highly congested nodes, since the packet based routing of the WSN is connectionless, which is inherently unreliable. As a result, any proposed protocol might also lose critical security packets such as keys, if it does not maintain a reasonable channel error rate.

Due to these limitations, devising security protocols for the WSN is complicated and may not be successfully accomplished by the simple adaptation of security solutions designed for wired networks. Studies such as Wagner (2004) and Krishnamachari et al. (2002) show that data transmission consumes much more energy than computation. Data transmission accounts for 70% of the energy cost of computation and communication for the SNEP protocol (Perrig et al. 2002). Data aggregation can greatly help to reduce this consumption by eliminating redundant data. However, the aggregators are vulnerable to attack especially they are not equipped with tamper-resistant hardware. When an aggregator node is compromised, it is easy for the adversary to change the aggregation result and inject false data into the WSNs. Unfortunately, the security mechanisms that are used in a similar network environment are not appropriate for the WSN since they are based on public key crytpography which is too expensive for sensor nodes.

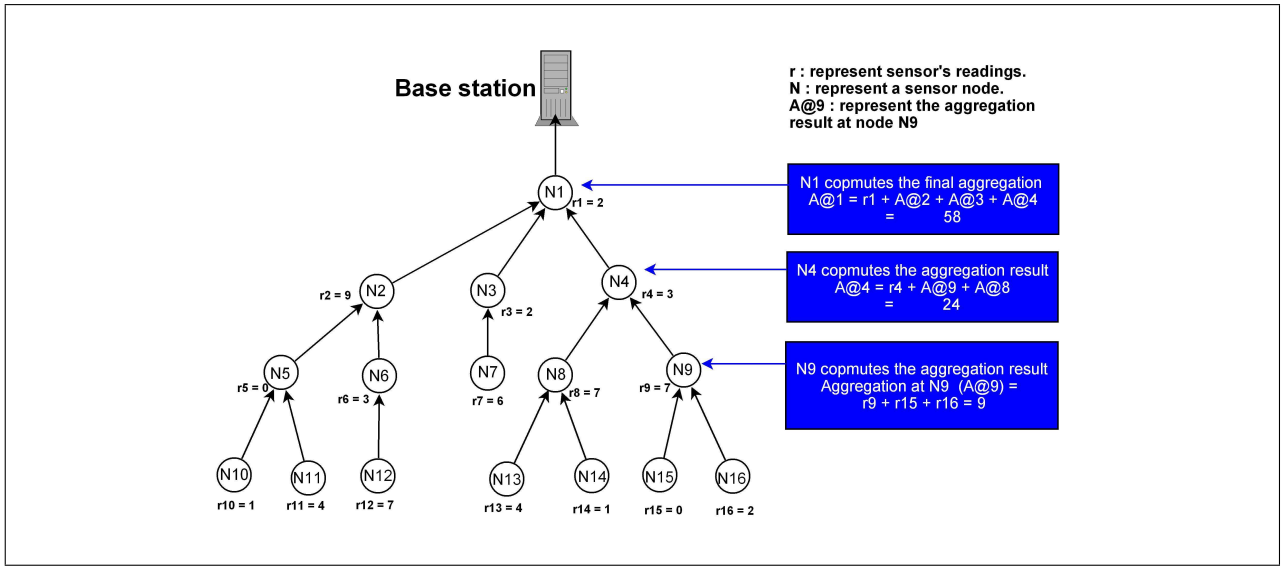

Figure 1: An aggregation scenario using sum function.

The only existing survey on secure data aggregation schemes, which is done by Sang et al. (2006), classifies them into: hop-by-hop and end-to-end encrypted data aggregation. However, this classification does not detail the security analysis of these schemes nor their performance. They are classified here based on how many times the data is aggregated when it travels to the base station.

Our contributions in this paper include the following:

- The security issues in data aggregation for the WSN are discussed and then secure data aggregation is defined informally.
- An adversarial model is proposed that can be expected in any secure data aggregation scheme. This model covers different types of adversaries where the computational strength, the network access, and node's secrets access may vary.
- A survey of the "state-of-the-art" existing secure data aggregation schemes is presented and then classified into two groups according to the number of aggregator nodes and whether the integrity of the aggregated result is considered or not.
- Finally a conceptual secure data aggregation framework is proposed to help in comparing secure data aggregation schemes.

The rest of the paper is organized as follows: In section 2, data aggregation is explained In section 3, the expected attacks that threaten secure data aggregation schemes in the WSN is discussed. Then, different types of attacks are listed. In section 4, the expected adversary in any secure aggregation scheme is classified. In section 5, the current research is surveyed and classified into two models. Then in section 6, the discussed secure schemes are compared and then a conceptual framework is given.Finally, section 7 concludes the paper.

2 Data Aggregation in the WSN

Typically, there are three types of nodes in WSN: normal sensor nodes, aggregators, and a querier. The aggregators collect data from a subset of the network, aggregate the data using a suitable aggregation function and then transmit the aggregated result to an upper aggregator or to the querier who generates the query. The querier is entrusted with the task of processing the received sensor data and derives meaningful information reflecting the events in the target field. It can be the base station or sometimes an external user who has permission to interact with the network depending of the network architecture. Data communication between sensors, aggregators and the querier consumes a large portion of the total energy consumption of the WSN. The WSN in figure 1 contains 16 sensor nodes and uses SUM function to minimize the energy consumption by reducing the number of bits reported to the base station. Node 7, 10-16 are normal nodes that are collecting data and reporting them back to the upper nodes whereas nodes 1-6, 8, 9 are aggregators that perform sensing and aggregating at the same time. In this example 16 packets traveled within the network and only one packet is transmitted to the base station. However, the number of traveling packets would increase to 50 packets if no data aggregation exists. This number of packets has been computed for one query.

Most existing proposals for data aggregation are subject to attack (Wagner 2004). Once a single node is compromised, it is easy for an adversary to inject false data into the network and mislead the aggregator to accept false readings. Because of this, the need for secure data aggregation is raised and its importance needs to be highlighted. However, it was found that the design principles of secure data aggregation schemes have no standard and sometimes are poorly understood. Moreover, there is no clear definition of what secure data aggregation should mean and what requirements they should have. These proposals might have one or more of the security requirements that are discussed in section 2.1 depending on how the secure aggregation looks like to the authors. Unfortunately, following this method to address the security in data aggregation is not practical and has several problems. For example, secure data aggregation has been addressed in (Przydatek et al. 2003) from the point of view of detecting forged data aggregation values. This does not cover security issues such as how to elect aggregators or how to set up trust between aggregators and sensor nodes. Some proposed protocols provide more security requirements than others, or send more bits than others as seen in section 6. How to compare between them? Are they all called secure data aggregation protocols?

A general definition for secure data aggregation is the efficient delivery of the summary of sensor readings that are reported to an off-site user in such a way that ensures these reported readings have not been altered (Przydatek et al. 2003). They consider an aggregation application where the querier is located outside the WSN and the base station acts as an aggregator. Moreover, a detailed definition of secure data aggregation is proposed as the process of obtaining a relative estimate of the sensor readings with the ability to detect and reject reported data that is significantly distorted by corrupted nodes or injected by malicious nodes (Shi & Perrig 2004). However, rejecting reported data that is injected by malicious nodes consumes the network resources, specifically the nodes' batteries, since each time the suspicious packet will be processed at the aggregator point. The damage caused by malicious nodes or compromised nodes should be reduced by adding a self-healing property to the network. This property helps the network in learning how to handle new threats through extensive monitoring of network events, machine learning and network behavior modeling. Consequently, it is believed that a secure data aggregation scheme for the WSN should have the following properties:

- Fair approximation of the sensor readings although a limited number of nodes are compromised.
- Ability to reduce the size of the data transmitted through the network.
- Data freshness and integrity are important and should be included in the scheme. However, the application type of the WSN affects the scheme designer's decision regarding whether to add the data confidentiality and availability or not.
- Dynamic response to attack activities by executing of a self-healing mechanism.
- Dynamic aggregator election/rotation mechanism to balance the workload at aggregators.

These properties should work together to provide accurate aggregation results securely without exhausting the network.

2.1 Requirements for Data Aggregation Security

Since WSNs share some properties with the traditional wireless networks, the data security requirements in the WSNs are similar to those in traditional networks (Perrig et al. 2002, Shi & Perrig 2004). However, there are some unique specifications that can only be found in WSNs, as discussed in Section 1, that require more attention during design process. In this section the required security properties to strengthen the security in aggregation schemes will be defined.

- **Data Confidentiality:** ensures that information content is never revealed to anyone who is not authorized to receive it. It can be divided (in secure data aggregation schemes) into a hop-by-hop basis and an end-to-end basis. In the hop-by-hop basis, any aggregator point needs to decrypt the received encrypted data, apply some sort of aggregation function, encrypt the aggregated data, and send it to the upper aggregator point. This kind of confidentiality implementation is not practical for the WSN since it requires extra computation. On the other basis, the aggregator does not need to decrypt and encrypt data and instead of this, it needs to apply the aggregation functions directly on the encrypted data by using homomorphic encryption (Westhoff et al. 2006). The interested reader in homomorphic encryption is referred to Appendix A.
- **Data Integrity:** ensures that the content of a message has not been altered, either maliciously or by accident, during transmission process. Confidentiality itself is not enough since an adversary is still able to change the data although it knows nothing about it. Suppose a secure data aggregation scheme focuses only on data confidentiality. An adversary near the aggregator point will be able to change the aggregated result sent to the base station by adding some fragments or manipulating the packet's content without detection. Moreover, even without the existence of an adversary, data might be damaged or lost due to the wireless environment.
- **Data Freshness** ensures that the data are recent and that no old messages have been replayed to protect data aggregation schemes against replay attacks. In this kind of attack, it is not enough that these schemes only focus on data confidentiality and integrity because a passive adversary is able to listen to even encrypted messages transmitted between sensor nodes can replay them later on and disrupt the data aggregation results. More importantly when the adversary can replay the distributed shared key and mislead the sensor about the current key.
- **Data Availability** ensures that the network is alive and that data are accessible. It is highly recommended in the presence of compromised nodes to achieve network degradation by eliminating these bad nodes. Once an attacker gets into the WSN by compromising a node, the attack will affect the network services and data availability especially in those parts of the network where the attack has been launched. Moreover, the data aggregation security requirements should be carefully implemented to avoid extra energy consumption. If no more energy is left, the data will no longer be available. When the adversary is getting stronger, it is necessary that a secure data aggregation scheme contains some of the following mechanisms to ensure reasonable level of data availability in the network:
  - **Self-healing** that can diagnose, and react to the attacker's activities especially when he gets into the network and then start corrective actions based on defined policies to recover the network or a node.
  - **Aggregator rotation** that rotates the aggregation duties between honest nodes to balance the energy consumption in WSN.
- **Authentication:** There are two types of authentication; entity authentication, and data authentication. Entity authentication allows the receiver to verify if the message is sent by the claimed sender or not. Therefore, by applying authentication in the WSNs, an adversary will not be able to participate and inject data into the network unless it has valid authentication keys. On the other hand, data authentication guarantees that the reported data is the same as the original one. In a secure data aggregation, both entity and data authentication are important since entity authentication ensures that

some exchanged data between sensors. For instance, electing an aggregator point or reporting invalid aggregated results are authenticated using their identity while data authentication ensures that raw data are received at the aggregators at the same time as they are being sensed.

- **Non-repudiation:** ensures that a transfered packet has been sent and received by the person claiming to have sent and received the packet. In secure aggregation schemes, once the aggregator sends the aggregation results, it should not be able to deny sending them. This gives the base station the opportunity to determine what causes the changes in the aggregation results.
- **Data Accuracy:** One major outcome of any aggregation scheme is to provide an aggregated data as accurately as possible since it is worth nothing to reduce the number of bits in the aggregated data but with very low data accuracy. A trade-off between data accuracy and aggregated data size should be considered at the design stage because higher accuracy requires sending more bits and thus needs more power.

3 Types of Attacks on WSN Aggregation

WSNs are vulnerable to different types of attacks (Roosta et al. 2006) due to the nature of the transmission medium (broadcast), remote and hostile deployment location, and the lack of physical security in each node. However, the damage caused by these attacks varies from scheme to scheme according to the assumed adversarial model (to be discussed in section 4) which is assumed by the scheme's designers. In this section, these attacks that might affect the aggregation in the WSN are discussed.

- **Denial of Service Attack(DoS):** is a standard attack on the WSN by transmitting radio signals that interfere with the radio frequencies used by the WSN and is sometimes called jamming. As the adversary capability increases, it can affect larger portions of the network. In the aggregation context, an example of the DoS can be an aggregator that refuses to aggregate and prevents data from traveling into the higher levels.
- **Node Compromise:** is where the adversary is able to reach any deployed sensor and extract the information stored on it which is some times called supervision attack. Considering the data aggregation scenario, once a node has been taken over, all the secret information stored on it can be extracted.
- **Sybil Attack:** is where the attacker is able to present more than one identity within the network. It affects aggregation schemes in different ways. Firstly, an adversary may create multiple identities to generate additional votes in the aggregator election phase and select a malicious node to be the aggregator. Secondly, the aggregated result may be affected if the adversary is able to generate multiple entries with different readings. Thirdly, some schemes use witnesses to validate the aggregated data and the data is only valid if n out of m witnesses agreed on the aggregation results. However, an adversary can launch a Sybil attack and generate n or more witness identities to make the base station accept the aggregation results.
- **Selective Forwarding Attack:** With no consideration about security, it is assumed in the WSN that each node will accurately forward received messages. However, a compromised node may refuse to do so. It is up to the adversary that is controlling the compromised node to either forward the received messages or not. In the aggregation context, any compromised intermediate nodes have the ability to launch the selective forwarding attack and this subsequently affects the aggregation results.
- **Replay Attack:** In this case an attacker records some traffic from the network without even understanding its content and replays them later on to mislead the aggregator and consequently the aggregation results will be affected.
- **Stealthy Attack:** The adversary aims to inject false data into the network without revealing its existence. In a data aggregation scenario, the injected false data value leads to a false aggregation result. A compromised node can report significantly biased or fictitious values, and perform a Sybil attack to affect the aggregation result.

4 Adversarial Model

In this section, we describe the different capabilities that an adversary may have against the WSN.

4.1 Adversary Type

Secure data aggregation schemes are threaten by two types of adversaries: passive and active. Passive adversary affects the data confidentiality property while active adversary affects data integrity property.

- **Passive Adversary:** is the adversary that takes advantage from the communication nature of the wireless (broadcasting) and eavesdrop on the traffic to obtain any important information about the sensed data. For example, if the adversary is able to hear the traffic near the aggregator point, it can gain some knowledge about the aggregated result especially if the secure data aggregation scheme does not ensure data confidentiality.
- **Active Adversary:** is the adversary that interacts with the WSN by injecting packets, destroying nodes, stopping/delaying packets from being delivered to the querier, compromising nodes and extracting sensitive data, etc.

4.2 Network Access

Each WSN has three different types of components: sensor, aggregator, and base station with different functionalities and capabilities. In this section, the adversary ability to compromise these three elements is discussed.

- **Total Access:** The adversary that has total access to the network is powerful and has access to the whole WSN. If the adversary is passive, this means that he can listen to all communications between nodes. On the other hand, if the adversary is active, this means that he can interact with all types of components in the WSN.
- **Partial Access:** The adversary in this type has less power compared to the previous one. Its goal is to listen to communications between a subset of nodes in the network, if the adversary is passive. On the other hand, if the adversary is active, this means that he can only interact with a subset of nodes in the WSN.

Table 1: Adversarial Models in the Aggregation schemes

	Adv. Type		Network Access		Secrets Access		Adv. Classification		
	Active	Passive	Total	Partial	Total	No	Strong	Med	Light
CDA		x		x		x			x
SDA	x			x	x			x	
SIA	x		x		x		x		
SHDA	x		x		x		x		
RA	x			x	x			x	
WDA	x			x	x			x	
SecureDAV	x			x	x			x	
SRDA		x		x		x			x
SDAP	x			x	x			x	
ESA	x			x	x			x	
EDA		x	x			x			x

4.3 Access to Secret Data

This refers to the sensitive information kept in the nodes. Based on the assumptions made by the designers, each scheme considers an adversary with different access levels to the secrets that are kept on the sensors.

- **Total Access:** The adversary that has total access to the node's secrets is able to extract *all* the sensitive information that is stored in the sensor's memory and then harm the aggregation results.
- **Partial Access:** The adversary with this type of access is able to only extract some of the secret data that is stored in the sensor's memory.

4.4 Adversary Classification

Table 1 summaries the adversary types that are exist in different secure aggregation proposals based on the designers' assumptions. It is concluded that the existing adversaries can be divided into three types:

- **Strong Adversary:** refers to an active adversary that has the ability to compromise any component in the WSN. In other words, it can be described as an active adversary with no limitation on its network access and its ability to extract secret data on sensor's memory since the aggregated data are highly important to the adversary.
- **Medium Adversary:** refers to an active adversary that has low computational strength to launch an attack against the secure system because the aggregated data are not that important as in the previous type and also because the network access is limited.
- **Light Adversary:** refers to a passive adversary with limited access to the network and it is interested to reveal the encrypted aggregated data.

It is believed that this adversary classification should help in the better evaluation of the proposed schemes and facilitate making decisions on which scheme is more suitable for specific conditions (as described in section 6).

5 Classification of Existing Secure Data Aggregation Schemes

This section classifies the proposed secure data aggregation schemes into two models: the one aggregator model and the multiple aggregator model. Under each model, each scheme is examined to see whether it has a verification phase or not.

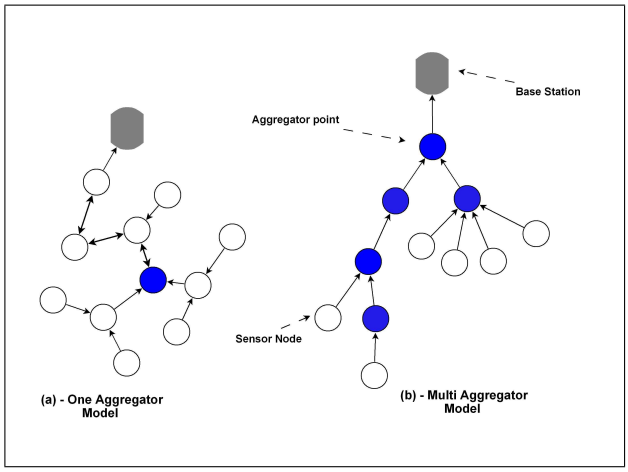

Figure 2: Sketch of single and multi aggregator model.

5.1 Single Aggregator Model

In this model, the aggregation process takes place once between the sensing nodes and the base station or the external user. In other words, all individual collected data in the WSN travels to only one aggregator point in the network before reaching the querier. This aggregator node should be powerful enough to perform the expected high computation and communication. The main role of the data aggregation might not be satisfied fully since redundant data will still travel in the network for a while until they reach the aggregator as in Figure 2-a. This model is useful when the network is small or when the querier is not in the same network. However, large networks are not suitable places to implement this model especially when data redundancy at the lower levels is high. The data aggregation schemes that fit in this model can be divided into two categories: whether they have a verification phase or not.

- **Verification Phase:** informs on the secure data aggregation schemes that aggregate data once in its way to the querier. This phase enhances the querier's ability to distinguish between the valid and invalid aggregated readings.
- **No Verification Phase:** informs on the secure data aggregation scheme that does not contain a verification phase because data integrity has not been considered by the scheme's designers. In other words, the type of expecting adversary is honest but has some interest in knowing about

sensitive information while the one in the previous phase is not honest and can inject false readings.

5.2 Multiple Aggregator Model

In this model, collected data in the WSN are aggregated more than one time before reaching the last destination (querier). This model achieves greater reduction in the number of bits transmitted within the network especially in the large WSNs, as illustrated in Figure 1. A sketch of the multi-aggregator model can be found in Figure 2-b. The importance of this model appears as the network size is getting bigger especially when data redundancy at the lower levels is high. The data aggregation schemes that fit in this model can be divided into two categories: whether they have a verification phase or not.

- **Verification Phase:** Secure data aggregation scheme that contains a verification phase to enhance the querier ability in distinguishing between the valid and invalid aggregated readings. This phase is more complicated than the same phase in the single aggregator model since the data is aggregated many times at different aggregation points. The querier is interested to know whether the final aggregated result is altered or not by one of these points.
- **No Verification Phase:** informs on the secure data aggregation scheme that does not contain a verification phase because data integrity has not been considered by the scheme's designers.

6 Comparison of the secure aggregation schemes

This section, attempts to compare the secure data aggregation schemes that were reviewed in section 5. Comparisons of security schemes can be difficult since the designers solve secure aggregation from different angles. Therefore, these schemes are compared in a number of different ways: security services provided, cryptographic primitives used, resilience against attacks described in section 3, and the number of bits transmitted in the aggregation phase. For each method an attempt is made to show the differences between these secure aggregation schemes.

6.1 Description of Existing Schemes

The first secure data aggregation (SDA) was proposed by Hu & Evans (2003) who studied the problem of data aggregation once one node is compromised. This protocol achieves resilience against a node compromise by delaying the aggregation and authentication at the upper levels. Therefore, sensors measurements are forwarded unchanged and then aggregated at the second hop instead of aggregating them at the immediate next hop. Thus, the sensor needs to buffer the data to authenticate it once the shared key is revealed by the base station. Moreover, the proposed scheme only offers data integrity, freshness and authentication. Even though it increases the confidence in the sensor readings integrity the data can be altered once a parent and child in the hierarchy are compromised. Once a compromised node is detected, no practical action is taken to reduce the damage caused by this compromise which affects the data availability in the network. Much worse, once a grandfather node detects a node compromise, it could not decide whether the cheating node is the child or the grandchild.

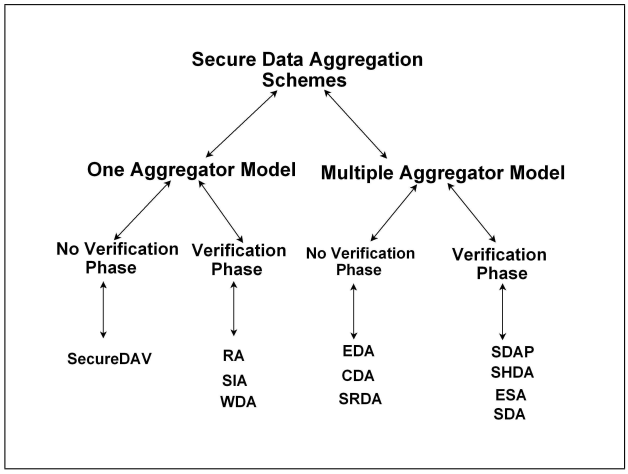

Figure 3: Classification of Existing Secure Data Aggregation Schemes.

In addition, SDA scheme is improved in ESA by Jadia & Mathuria (2004). Instead of using μTESLA to authenticate the base station's broadcast in the validation process to reveal the shared key with sensors, the authors used one-hop pairwise keys (to encrypt data between a node and its parent) and two-hop pairwise keys (to encrypt data between a node and its grandparent). This will improve the secure aggregation scheme by adding data confidentiality and reducing the memory overhead since data does not need to be stored until the key is revealed. However, the system will still break as soon as two consecutive nodes in the hierarchy are compromised.

Przydatek et al. (2003) proposed a secure information aggregation (SIA) framework for WSNs called aggregate-commit-prove. This framework provides resistance against a special type of attack called stealthy attacks aggregate manipulation where the attacker's goal is to make the user accept false aggregation results without revealing its presence to the user. It consists of three node categories: a home server, a base station, and sensor nodes. SIA assumes that each sensor has a unique identifier and shares a separate secret cryptographic key with both the home server and the aggregator. The keys enable message authentication and encryption if data confidentiality is required. Moreover, it assumes that the home server and base station can use a mechanism, such as μTESLA (Perrig et al. 2002), to broadcast authentic messages. SIA consists of three parts: collecting data from sensors and locally computing the aggregation result, committing to the collected data, and reporting the aggregation result while proving the correctness of the result. SIA offers data integrity, authentication, data freshness, and confidentiality (if required).

Wagner (2004) proposed a mathematical framework (RA) for evaluating the security of several resilient aggregation techniques. The paper measures how much damage an adversary can cause by compromising a number of nodes and then using them to inject erroneous data. Wagner described a number of better methods for securing the data aggregation such as how the median function is a good way to summaries statistics. Furthermore, Wagner claimed that trimming and truncation can be used to strengthen the security of many aggregation primitives by eliminating possible outliers. However, this work only focused on examining the received aggregated data (at the base station) without

Table 2: Attacks Against Existing Aggregation Schemes.

Scheme	Denial of services	Node compromise	Sybil	Selective forwarding	Replay	Stealthy	Adversary Classification
CDA	x	x					light
SDA	x	x		x		x	med
SIA	x	x		x			high
SHDA	x	x		x			high
WDA	x	x	x	x	x	x	med
SecureDAV	x	x		x	x	x	med
SRDA	x						light
SDAP	x	x		x		x	med
ESA	x	x		x		x	med
EDA	x						light

studying how these data are aggregated. Thus, when the network size increases, the communication cost will be very high for the transmission of all the sensor readings to the base station. Moreover, eliminating abnormal data with no further reasoning is impractical especially for applications such as monitoring bush-fire.

A witness based data aggregation (WDA) scheme for the WSN is being proposed by Du et al. (2003) to assure the validation of the data sent from an aggregator node to the base station. In order to prove the validity of the aggregated result, the aggregator node has to provide proofs from several witnesses. A witness is one who also performs data aggregation like the aggregator node, but does not forward its result to the base station. Instead, each witness computes the message authentication code (MAC) of the result and then sends it to the aggregator node which must forward the proofs to the base station. WDA offers only integrity property to the data aggregation security and this is required to send multiple copies similar to the original aggregated result, to the aggregator point. Thus, the aggregator point must forward these reports as well as the aggregated result to the base station. Since the aggregator point is fixed and responsible to handle so much traffic, the aggregator resources will not last long.

Moreover, SecureDAV (Mahimkar & Rappaport 2004) improved the data integrity vulnerability in SDA and ESA by signing the aggregated data. In SecureDAV, each sensor within a cluster will have its share of its secret cluster key and then it will be able to generate a partial signature on the aggregated data. Once an aggregator receives sensor readings in the same cluster, it aggregates them and broadcasts the average value of the readings. Each sensor in the cluster compares its reading with the average value received from the aggregator. Then, it partially signs the average value only and only if the difference between the received average value and its reading is less than a certain value (threshold). Then, the aggregator (cluster-head) combines partial signatures to form a full signature of the aggregated results and sends it to the base station. SecureDAV provides data confidentiality, data integrity, and authentication. The drawbacks of this scheme are: it requires high communication costs on data validation, and supports only the AVG aggregation function.

Yang et al. (2006) proposed a secure hop-by-hop data aggregation protocol (SDAP) that can tolerate more than one compromised node. SDAP is based on two principles: divide-and-conquer and commit-and-attest. In order to reduce the damage caused by compromising an aggregator at a high level in the per-hop aggregation scheme, SDAP uses the divide-and-conquer principle to divide the network tree into multiple logical subtrees which increases the number of aggregators and reduces the number of nodes in each subtree. Consequently, the damage caused by compromising an aggregator of a subtree is reduced. The other principle, that is commit-and-attest, enhances the ordinary hop-by-hop aggregation scheme by adding a commitment property, and helps the base station to prove the correctness of the aggregated data. Once an aggregator of a logical subtree commits its aggregation result, it can not deny it later on. This scheme needs to send much data to ensure reasonable level of security (as explained in Appendix B).

Furthermore, Chan et al. (2006) extended the work in SIA by applying the aggregate-commit-prove framework in fully a distributed network instead of single aggregator model. In general, this scheme (SHDA) offers exactly what the SIA does data integrity, authentication, and confidentiality. Each parent sensor performs an aggregation function whenever it has heard from its child nodes. In addition, it has to create a commitment to the set of the input used to compute the aggregated result by using a merkle hash tree. Then, it forwards the aggregated data and the commitment to its parent until it reaches the base station. Once the base station received the final commitment values, it rebroadcasts them into the rest of the network in an authenticated broadcast. Each node is responsible for checking whether its contribution was added to the aggregated data or not. Once its readings are added, it sends an authentication code to the base station where the authentication code for node R is $MAC_{KR}(\mathrm{N}\|\mathrm{OK})$. For communication efficiency, the authentication codes are aggregated along the way to the base station. However, missing one authentication code for any reason leads the base station to reject the aggregated result. Furthermore, noticeable delay, too much transmission and computation will be added as consequences of adding security to the scheme.

Sanli et al. (2004) developed a new data aggre-

Table 3: Comparison between different secure data aggregation schemes

Scheme	Confidentiality	Integrity	Freshness	Availability	Authentication
CDA	x				
SDA		x	x		x
SIA	x	x	x		x
SHDA		x	x		x
WDA		x			x
SecureDAV	x	x			x
SRDA	x		x		
SDAP	x	x	x		x
ESA	x	x	x		x
EDA	x				

gation technique called the Secure Reference-Based Data Aggregation scheme (SRDA) that sends only the difference between sensed data and the reference value (called differential value) instead of raw data. Deference value is taken as the average value of previous sensor readings. In SRDA scheme, each sensor computes the differential data (sensed data - reference value), encrypts it, and then sends it to the cluster-head. The authors claim that the security level of the network should be gradually increased as the data is traveled to higher level cluster-heads. Therefore, they suggest using a cryptographic algorithm (RC6) with adjustable parameters such as the number of rounds, to achieve different level of security in the WSN. Increasing or decreasing the number of rounds changes the security strength of the RC6 that can be measured by the security margin. The security margin is the deviation of the actual number of rounds from the minimum number of rounds for which the algorithm is considered to be secured. The SRDA uses a higher security margin at higher level cluster-heads compared to low level cluster-heads.

Moreover, the problem of aggregating encrypted data in the WSN is being addressed in (Westhoff et al. 2006). The proposed protocol, called Concealed Data Aggregation (CDA), uses an additive and multiplicative homomorphic encryption scheme that allows the aggregator to aggregate encrypted data. In this paper, the authors argued that the security level is still reasonable and the privacy homomorphism (PH) (Domingo-Ferrer 2002) helps to implement encryption in the WSN, although Wagner (2003) proved that PH is unsecure against chosen plain text attacks. However, they admitted that the encryption in CDA is very expensive and adds between 0%-22% additional data overhead compared to RC5 which increases the power consumption of the sending node. Genrally speaking, CDA ensures only data confidentiality.

Furthermore, a new secure data aggregation scheme based on homomorphic encryption (EDA) is proposed by (Castelluccia et al. 2005) This allows an aggregator to execute the aggregation function and aggregate the encrypted data that are received from its children with no need for decryption and to recover the original messages. It uses a modular addition instead of the xor (Exclusive-OR) operation that is found in the stream ciphers. Thus, even if an aggregator is being compromised, original messages can not be revealed by an attacker. The authors claimed that the provided privacy protection by this scheme is comparable to the privacy protection that is provided by a scheme that performs end-to-end encryption with no aggregation. However, they admit that their proposed scheme generates significant overhead if the network is unreliable since sensors' identities of non-responding nodes must be sent together with the aggregated result to the base station. More importantly, this scheme concerns only one security property which is data confidentiality.

6.2 Security Services Provided

Since the considered type of adversary varies from one scheme to another, each proposed scheme has different requirements. Data authentication is a must in each secure scheme that defeats against any type of active adversary. Table 3 shows that data confidentiality is the minimum security requirement that should be provided when a light type of adversary is considered. Once the adversary capability increases and reaches the medium type, some of the proposed schemes protect against it by providing data integrity and authentication. However, the medium adversary type is able to at least launch the replay attack. It is believed that the minimum security requirements that provide reasonable security levels against this type of adversary should include data freshness too. As the adversary is getting stronger, a combination of data confidentiality, data integrity, authentication, and data freshness should exist in the proposed scheme. If the network lifetime is a concern for the designers, then the data availability should be provided as well. However, none of the existing proposals consider the data availability even when a strong adversary exists.

As a conclusion, the CDA, EDA, and SRDA have met the minimum security requirements when a light type of adversary is around. When the medium type is considered, the SDA, ESA, and SDAP have met the minimum requirements. However, the WDA and SecureDAV have not met the minimum requirements because they did not offer data freshness. Finally, the minimum requirements have been met in the SIA and SHDA when an adversary with strong capability is considered.

6.3 Cryptographic Primitives Used

The cryptographic primitives used in each of the proposed schemes are varied according to how the authors employ different primitives to achieve a certain

Table 4: Cryptographic primitives used

Scheme	Message authen-tication	Digital signa-ture	Symmetric key	Public key	Readings commit-ment	Privacy homom-orphic	Broadcast authen-. tication	Inter-active protocol	Voting sche-em
CDA			x			x			
SDA	x		x				x		
SIA	x		x		x		x	x	
SHDA	x		x		x		x	x	
RA									
WDA	x		x						x
SecureDAV		x		x	x				
SRDA			x						
SDAP	x		x		x		x	x	
ESA	x		x						
EDA			x			x			

service. Table 4 shows that most of these schemes use a message authentication code (MAC) to exclude unauthorized parties from sending forged aggregated data. However, SecureDAV uses a digital signature instead. Furthermore, the MAC is used to protect the original message from being altered. Symmetric and public keys have been used to achieve hop-by-hop or end-to-end encryptions which prevent the passive adversary from eavesdropping on the traffic. Moreover, a verification process in these schemes has been set up in different ways, such as: using interactive protocols, broadcast authentication by the base station, or voting system.

6.4 Attacks Existence

This section analyses the secure aggregation schemes described in Section 5 and investigates whether they are vulnerable to types of attacks described in section 3 or not. Table 2 shows that all these schemes are vulnerable to DoS and Physical attacks as long as the existence of at least the light type of adversary. As the capability of the adversary varies from light to strong, the damage caused by these attacks varies, too. The stronger adversary can jam a wider range of the network, while the light adversary can jam only a limited area. Moreover, since the light type of adversary is not interested in affecting the data integrity of the system, the adversary will not be interested in launching Sybil, Replay, Selective forwarding, stealthy, and Spoofed-altered attacks such as in CDA. Replay attack may be launched in these schemes by the medium type of adversary or above, unless these schemes provide data freshness (Table 3). For example, the WDA and SecureDAV schemes are vulnerable to replay attack. This type of adversary, can also launch Sybil attack in secure aggregation schemes unless the identity is checked upon receiving any message.

6.5 Framework for Evaluation New Schemes

Based on the analyses in the previous sections, a conceptual framework, helps the new schemes designers to strengthen their proposed scheme against the considered adversarial model. As far as is known this work is the first that tries to build such a framework that can suggest the minimum security requirements that should exist in any scheme according to its specifications. It is believed that the security level can be determined by considering one of the adversarial models (discussed in section 4). Then, the network size helps the designers to choose the proper aggregation model. Figure 4 shows that the minimum requirements for a proposed secure scheme to resist against the light adversary are data confidentiality and freshness. It is suggested that these requirements be offered in the single aggregator model if the network size is small, otherwise, in the multiple aggregator model. The minimum security services that are required to resist against the medium adversary are data integrity, authentication, and freshness, while the required services to resist against the strong adversary needs data confidentiality as well as.

7 Conclusion and Future Work

By reviewing the existing data aggregation security in the WSN, an adversarial model that can threaten any secure aggregation scheme has been proposed. Consequently, these schemes were classified into two groups: the one aggregator model, and the multiple aggregator model. Based on this classification and the adversarial model, a conceptual framework that leads to better evaluation of secure aggregation schemes was also proposed. In the future, it is planned to evaluate more secure schemes and extend the framework if necessary.

References

Castelluccia, C., Mykletun, E. & Tsudik, G. (2005), Efficient Aggregation of Encrypted Data in Wireless Sensor Networks., *in* 'MobiQuitous', IEEE Computer Society, pp. 109–117.

Chan, H., Perrig, A. & Song, D. (2006), Secure hierarchical in-network aggregation in sensor networks., *in* A. Juels, R. N. Wright & S. D. C. di Vimercati, eds, 'ACM Conference on Computer and Communications Security', ACM, pp. 278–287.

Corporation, C. (2006), 'Mica2 datasheet'. Reviewed 10th of October 2007, `http://www.xbow.com/Products/Product_pdf_files/Wireless_pdf/MICA2_Datasheet.pdf`.

Domingo-Ferrer, J. (2002), A provably secure additive and multiplicative privacy homomorphism., *in* A. H. Chan & V. D. Gligor, eds, 'ISC', Vol. 2433 of *Lecture Notes in Computer Science*, Springer, pp. 471–483.

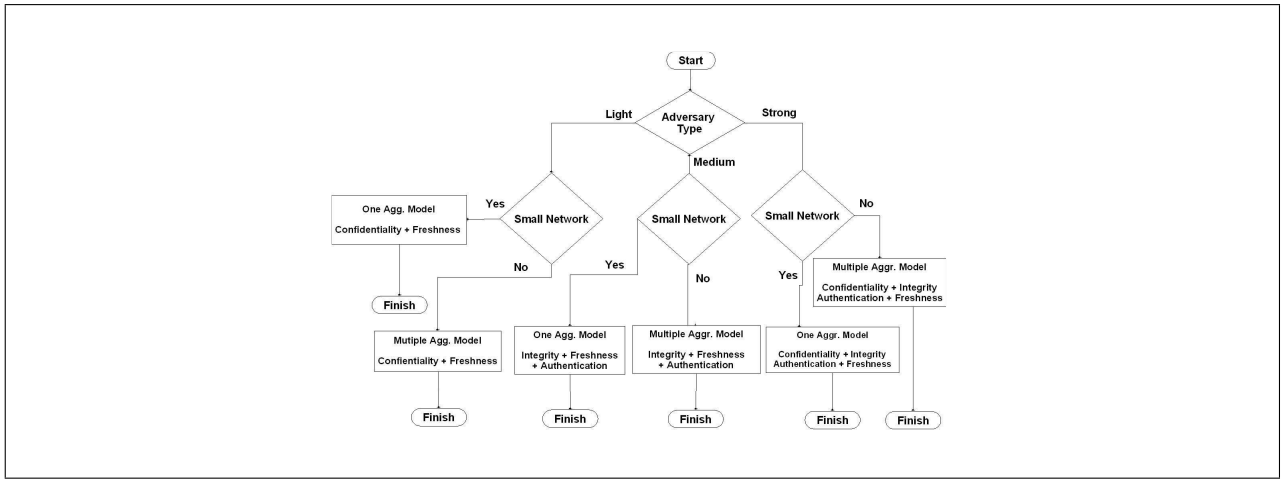

Figure 4: Framework for Evaluation New Schemes.

Du, W., Deng, J., Han, Y. S. & Varshney, P. (2003), A witness-based approach for data fusion assurance in wireless sensor networks, *in* 'IEEE Global Communications Conference (GLOBECOM)', Vol. 3, pp. 1435– 1439.

Grigoriev, D. & Ponomarenko, I. V. (2003), 'Homomorphic public-key cryptosystems over groups and rings', *CoRR* **cs.CR/0309010**.

Guimarães, G., Souto, E., Sadok, D. F. H. & Kelner, J. (2005), Evaluation of security mechanisms in wireless sensor networks., *in* 'ICW/ICHSN/ICMCS/SENET', IEEE Computer Society, pp. 428–433.

He, T., Vicaire, P., Yan, T., Luo, L., Gu, L., Zhou, G., Stoleru, R., Cao, Q., Stankovic, J. A. & Abdelzaher, T. F. (2006), Achieving real-time target tracking usingwireless sensor networks., *in* 'IEEE Real Time Technology and Applications Symposium', IEEE Computer Society, pp. 37–48.

Hu, L. & Evans, D. (2003), Secure aggregation for wireless network., *in* 'SAINT Workshops', IEEE Computer Society, pp. 384–394.

Jadia, P. & Mathuria, A. (2004), Efficient secure aggregation in sensor networks., *in* L. Bougé & V. K. Prasanna, eds, 'HiPC', Vol. 3296 of *Lecture Notes in Computer Science*, Springer, pp. 40–49.

Krishnamachari, B., Estrin, D. & Wicker, S. B. (2002), The impact of data aggregation in wireless sensor networks., *in* 'ICDCS Workshops', IEEE Computer Society, pp. 575–578.

Mahimkar, A. & Rappaport, T. S. (2004), SecureDAV: A secure data aggregation and verification protocol for sensor networks., *in* 'Global Telecommunications Conference', Vol. 4, pp. 2175– 2179.

Mainwaring, A. M., Culler, D. E., Polastre, J., Szewczyk, R. & Anderson, J. (2002), Wireless sensor networks for habitat monitoring., *in* C. S. Raghavendra & K. M. Sivalingam, eds, 'WSNA', ACM, pp. 88–97.

Murthy, C. S. R. & Manoj, B. (2004), *Ad Hoc Wireless Sensor Networks Architectures and Protocols*, Prentice Hall PTR, Upper Saddle River, NJ, USA.

Perrig, A., Szewczyk, R., Tygar, J. D., Wen, V. & Culler, D. E. (2002), 'SPINS: Security Protocols for Sensor Networks.', *Wireless Network* **8**(5), 521–534.

Przydatek, B., Song, D. X. & Perrig, A. (2003), SIA: Secure Information Aggregation in Sensor Networks., *in* I. F. Akyildiz, D. Estrin, D. E. Culler & M. B. Srivastava, eds, 'SenSys', ACM, pp. 255–265.

Rivest, R. L., Shamir, A. & Adleman, L. M. (1978), 'A Method for Obtaining Digital Signatures and Public-Key Cryptosystems.', *Communication. ACM* **21**(2), 120–126.

Roosta, T., Shieh, S. & Sastry, S. (2006), Taxonomy of security attacks in sensor networks, *in* 'The First IEEE International Conference on System Integration and Reliability Improvements', IEEE International, Washington, DC, USA.

Sang, Y., Shen, H., Inoguchi, Y., Tan, Y. & Xiong, N. (2006), Secure data aggregation in wireless sensor networks: A survey, *in* 'Proceedings of the Seventh International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT '06)', IEEE Computer Society, Washington, DC, USA, pp. 315–320.

Sanli, H. O., Ozdemir, S. & Cam, H. (2004), SRDA: Secure reference-based data aggregation protocol for wireless sensor networks, *in* 'Vehicular Technology Conference', pp. 4650– 4654.

Shi, E. & Perrig, A. (2004), 'Designing secure sensor networks.', *IEEE Personal Communications* **11**(6), 38–43.

Wagner, D. (2003), Cryptanalysis of an algebraic privacy homomorphism., *in* C. Boyd & W. Mao, eds, 'ISC', Vol. 2851 of *Lecture Notes in Computer Science*, Springer, pp. 234–239.

Wagner, D. (2004), Resilient aggregation in sensor networks., *in* S. Setia & V. Swarup, eds, 'SASN', ACM, pp. 78–87.

Westhoff, D., Girao, J. & Acharya, M. (2006), 'Concealed data aggregation for reverse multicast traffic in sensor networks: Encryption, key distribution, and routing adaptation', *IEEE Transactions on Mobile Computing* **05**(10), 1417–1431.

Yang, Y., Wang, X., Zhu, S. & Cao, G. (2006), SDAP: a secure hop-by-hop data aggregation protocol for sensor networks., *in* 'Proceedings of the 7th ACM Interational Symposium on Mobile Ad Hoc Networking and Computing, MobiHoc 2006, Florence, Italy, May 22-25, 2006', ACM, pp. 356–367.

Appendix A: Homomorphic Cryptosystems

During the last few years homomorphic encryption schemes have been studied extensively since they are proved to be useful in many cryptographic protocols such as electronic elections (Grigoriev & Ponomarenko 2003), sensor networks (Westhoff et al. 2006, Castelluccia et al. 2005) and so on. Homomorphic cryptosystem is a cryptosystem that allows direct computation on encrypted data by using an efficient scheme. It is an important tool that can be used in a secure aggregation scheme to provide end to end privacy if needed. The scheme is called additively homomorphic if the message space ($\mathcal{M}$) is an additive while it is called multiplicative homomorphic if $\mathcal{M}$ is a multiplicative. Moreover, the encryption algorithm ($\mathcal{E}$) is called probabilistic if it gets a uniform random number as an input otherwise it is called deterministic.

The classical RSA scheme (Rivest et al. 1978) is a good example of a deterministic, multiplicative homomorphic cryptosystem on $\mathcal{M} = /N$ where $\mathcal{N}$ is the product of two large primes. Thus, $\mathcal{C} = /N$ is the ciphertext space while the key space is

$$\mathcal{K} = \{(k_e, k_d) = ((N, e), d) \mid N = pq, \\ ed \equiv 1 \ mod \ \varphi(N)\}$$

The encryption of any message m $\in \mathcal{M}$ is defined as:

$$\mathcal{E}_k e(m) = m^e \ mod \ N$$

while the decryption of any ciphertext c $\in \mathcal{C}$ is defined as:

$$\mathcal{D}_{ke,kd}(c) = c^d \ mod \ N = m \ mod \ N$$

Obviously, the encryption of the product of two messages $m_1, m_2 \in \mathcal{M}$ can be computed by multiplying the corresponding ciphertexts:

$$\begin{aligned} \mathcal{E}_{ke}(m_1 \cdot m_2) &= (m_1 \, m_2)^e \ mod \ N \\ &= (m_1^e \ mod \ N) \, (m_2^e \ mod \ N) \\ &= \mathcal{E}_{ke}(m_1) \cdot \mathcal{E}_{ke}(m_2) \end{aligned}$$

Appendix B: Performance Analysis

A Notations

For simplicity to perform our calculations in the following sections, we only consider the case which the leaf nodes transmit their readings and no readings are expected from aggregator nodes. We assume a general tree hierarchy in which every node has b children and the depth of the tree is d as in figure 5. This means the distance between the base station and the leaf nodes are d. Therefore, this kind of tree has b^d leaf nodes. In single aggregator model we consider the root of the tree to be the aggregator. Let us denote the length in bits of reported message from the leaf nodes toward upper level as x where x can be raw data or encrypted data. The sensor node ID in bits will be denoted as y. Also, we denote the MAC's length in bits as z. Also, let us denote query nonce as qn. Since TinyOS packet is pre-configured with a maximum size of 36 byte, 29 byte payload and 6 byte header, we denote header as oh to compute the overhead bits transmitted within the network. Finally, the total number of nodes N in this type of tree is n bits long and can be computed as

$$\left(\frac{b^{d+1}-1}{b-1}\right) \tag{1}$$

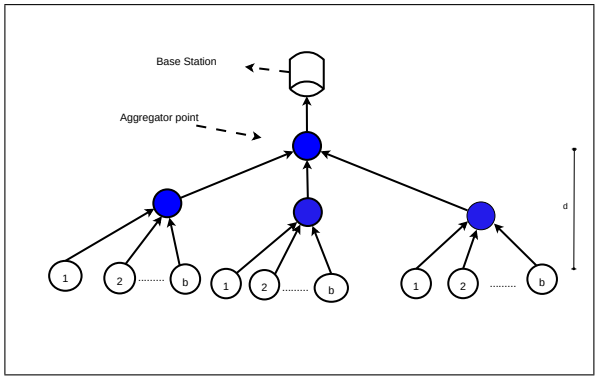

Figure 5: Tree model used to analyze the performance of schemes.

B Number of transmitted bits

First, we start analyzing the number of transmitted bits by considering the situation where no aggregation and no security are used. In this case each sensor forwards data as soon as it receives it. Suppose each node at level d sends its ID and sensed data toward its parents which are $x+y+oh$ bits long. So, the total number of bits generated at level d will be $b^d(x+y+oh)$. Since each intermediate node needs to forward $b(x+y)$ bits which adds oh bits as header into each transmission , the total number of bits traveled within the network is approximately

$$(d+1)b^d(x+y+oh) \tag{2}$$

Next, SDA as being discussed in section 5.2 achieves resilient against a node compromise by delaying the aggregation and authentication at the upper levels. So, each leaf node (at depth d) needs to send its ID, data, and one message authentication code toward its parent. The length of this message in bits will be $x+y+z+oh$. Therefore, the total number of bits sent by all the leaf nodes is $b^d(x+y+z+oh)$. Also, each node at depth *d-1* needs to forward the received data unchanged and add one MAC. Thus, the length of this message in bits will be $b(x+y+z)+z+oh$ and the total number of bits sent by all the nodes at level *d-1* should be $b^{d-1}[b(x+y+z)+z+oh]$. We compute the length of transmitted message at level *d-2*, *d-3*, and so on till we reach the base station. The approximate number of transmitted bits in this scheme is

$$b^d(x+y+z+oh)+\left(\frac{b^d-b}{b-1}+1\right)[b(x+y+z)+z+oh] \tag{3}$$

However, the improvement in ESA that adds confidentiality requires each node to add one more message authentication into the message. So, instead of sending $x+y+z+oh$ bits in SDA at each node at depth d, $x+y+2z+oh$ bits are needed to be sent by ESA. The total number of bits sent by all leaf nodes is $b^d(x+y+2z+oh)$ and consequently the total number of bits sent by ESA will be approximately

$$b^d(x+y+2z+oh)+\left(\frac{b^d-b}{b-1}+1\right)[b(x+y+z)+z+oh] \tag{4}$$

Table 5: Number of bytes transmitted within WSN

	b=2			b=3			b=4		
	d=2	d=3	d=4	d=2	d=3	d=4	d=2	d=3	d=4
No Aggregation	180	480	1200	405	1620	6075	720	3840	19200
SDA	210	462	966	399	1155	3423	546	2226	8946
ESA	234	510	1062	471	1470	4467	792	2520	13032
SDAP	289	623	1334	636	1722	6374	1119	4704	19709
CDA	91	248	512	215	660	1997	347	1403	5627
SIA	261	709	1829	541	2305	9109	933	5413	28709
WDA	825	2265	5865	1725	7395	29265	2985	17385	92265

In SDAP, it uses divide-and-conquer principle to divide the network tree into multiple logical subtrees which increases the number of aggregators and reduces the number of nodes in each subtree. For simplicity, we assume each subtree has an average size of s and therefore the number of subtrees is $(N/s)+1$ considering the base station as a subtree. Also, the hight of a subtree can be approximated by $d/2$ and the distance from each subtree's leader and the base station is $d/2$. Each leaf node needs to send its ID, aggregation flag (one bit), an encrypted sensed data that is concatenated with a MAC. This transmission is about $x+y+z+1+oh$ bits long. Thus, the total number of bits transmitted by all nodes at level d is $b^d(x+y+z+1+oh)$ and consequently the number of bits reaches the subtree's leader is around $(s-1)b^d(x+y+z+1+oh)$. Next, each subtree's leader will forward the aggregation result toward the base station and this increases the number of traveled bits within the network by $(N/s)(d/2)(x+y+z+1+oh)$ bits. Therefore, the total number of bits is approximated by

$$b^d(s-1)(x+y+z+1+oh)+\left(\frac{N}{s}\frac{d}{2}\right)(x+y+z+1+oh)$$

$$(x+y+z+1+oh)[b^d(s-1)+\left(\frac{Nd}{2s}\right)] \quad (5)$$

Moreover, the authors of CDA admitted that the encryption in CDA is very expensive and adds between 0%-22% additional data overhead compared to RC5 which increase the power consumption of the sending node. For example, the size of encrypted one Byte sensed data will increase to 9 Bytes by using RC5 while the size will range between 9 - 11 Bytes by using PH encryption (Westhoff et al. 2006). In other words, we can assume that the ecypted data has cx bits where c is constant that represent the overhead caused by encrypting 1 bit using PH. This means, the total encrypted messages in bits is approximatetly

$$N(cx+oh) \quad (6)$$

It obvious that the number of bits is increased linearly with the plain-text size (raw sensed data at the leaf nodes) and the number of leaf nodes. Consequently, the power consumption will increase linearly.

Subsequently, SIA proposed a secure information aggregation framework for WSNs called aggregate-commit-prove. In the aggregate phase, each leaf sensor needs to send its ID, data, query nonce, and two message authentication codes with two shared keys between the sensor and the aggregator and the base station. The length of this message in bits is $x+y+qn+2z+oh$. This message travels all the way toward the aggregator that is d hops away in our example. Therefore, the total number of bits traveled within the network till the sensed data reaches the aggregator is $db^d(x+y+qn+2z+oh)$ in each event. Then in the commit phase, the aggregator constructs a merkle hash tree of the received messages and sends the root of this tree as a commitment value, the number of leaves of the hash tree (the number of leaf nodes), and aggregated result. Let us assume for simplicity the length of the commitment value is $(x+y+qn+2z)$ bits long and the length of the aggregated result as long as the reported data x. Thus, the total number of bits sent to the remote user by the aggregator is $n+2x+y+qn+2z+oh$. Thus, the total number of traveled bits within the network will be

$$db^d(x+y+qn+2z+oh)+n+2x+y+qn+2z+oh \quad (7)$$

In WDA, authors assume that leave nodes are honest and the sensed data reaches the aggregator and wetnesses correctly. Let us assume that each sensor needs to send at least its ID and sensed data. The length of this message in bits is $x+y+oh$. Therefore, $db^d(x+y+oh)$ bits needs to be traveled withing the network to reach the aggregator for each event. Also, the same number of bits goes to each witness (w) and consequently the total number of traveled bits is $wdb^d(x+y+oh)$. Each witness computes the aggregated data and send message authentication code (MAC) contains its ID, aggregation result, and its shared key with the base station and sends its MAC to the aggregator. Finally, the aggregator forwards its ID, aggregation result computed by him, and all MACs received from the witnesses. Therefore, the total number of traveled bits is

$$db^d(x+y+oh)+wdb^d(x+y+oh)+w(z+oh)+(x+y+wz+oh)$$

$$db^d(x+y+oh)(1+w)+2wz+x+y+oh(w+1) \quad (8)$$

C Example

In this section, we give an example with numbers to give a better understanding about the previous equations in section B. Let us select x, y, z, w to be 7 bytes, 2 bytes, 6 bytes, 5 witnesses respectively. We compute the number of bytes that each secure aggregation scheme transmits to achieve the aggregation result with no consideration about the extra

overhead comes from the verification process (if exist) since not all of the schemes cover this process.

In Tabel 5, we investigate some of proposed secure data aggregation schemes in the context of the number of transmitted bits during the aggregation phase. We started with a system that has no aggregation to be as a reference in our comparison. CDA needs to send less number of bits compared to the system with no aggregation because it assumes light adversary type and performs multiple aggregation before sending the readings into the base station. Assuming light adversary means less security services provided which leads to less number of bits must be sent. Also, performing multiple aggregation helps the scheme to minimize the transmitted bits by removing redundant data and also removing the extra bits such as headers. As the adversary capability increases, the number of bits increases as well to provide reasonable level of security. For example, when the designers consider medium type of adversary, the security services provided should be different to those in the light type. From Tabel 5, SDA, ESA, SDAP, and WDA consider medium adversary type and send more bits than schemes that assumed to defeat against light adversary such as CDA. The reason that WDA sends more bits than other schemes, that defeat against the same type of adversary, is because it belongs to the one aggregator model. Finally, SIA considers high adversary type and can achieve its security goals by transmitting number of bits within the results of other secure proposals that defeat against medium type of adversary.