AN ENERGY-EFFICIENT AND SCALABLE SLOT-BASED PRIVACY HOMOMORPHIC ENCRYPTION SCHEME FOR WSN-INTEGRATED NETWORKS

Suraj Verma*, Prashant Pillai and Yim Fun Hu

Faculty of Engineering and Informatics, University of Bradford, Bradford, United Kingdom

ABSTRACT

With the advent of Wireless Sensor Networks (WSN) and its immense popularity in a wide range of applications, security has been a major concern for these resource-constraint systems. Alongside security, WSNs are currently being integrated with existing technologies such as the Internet, satellite, Wi-Max, Wi-Fi, etc. in order to transmit data over long distances and hand-over network load to more powerful devices. With the focus currently being on the integration of WSNs with existing technologies, security becomes a major concern. The main security requirement for WSN-integrated networks is providing end-to-end security along with the implementation of in-processing techniques of data aggregation. This can be achieved with the implementation of Homomorphic encryption schemes which prove to be computationally inexpensive since they have considerable overheads. This paper addresses the ID-issue of the commonly used Castelluccia Mykletun Tsudik (CMT) [12] homomorphic scheme by proposing an ID slotting mechanism which carries information pertaining to the security keys responsible for the encryption of individual sensor data. The proposed scheme proves to be 93.5% lighter in terms of induced overheads and 11.86% more energy efficient along with providing efficient WSN scalability compared to the existing scheme. The paper provides analytical results comparing the proposed scheme with the existing scheme thus justifying that the modification to the existing scheme can prove highly efficient for resource-constrained WSNs.

KEYWORDS

Homomorphic encryption, wireless sensor networks, end-to-end data confidentiality, integrated networks, performance evaluation

1.INTRODUCTION

Wireless Sensor Networks (WSNs) have gained immense popularity in the fields of remote monitoring due to their low-cost and low-power consumption nature [1]. Their primary function is to obtain sensor readings which can then be processed and transmitted across the network. In most cases, WSNs are working in cohesion with other communication technologies such as the Internet [2], IoT [3], Wi-Fi [4], RFID [5], satellites [6], etc. This is mainly due to the resource-

DOI:10.5121/ijassn.2015.5201

constraint nature of WSNs and their short communication range (10m-100m) whereby WSNs depend on backbone networks to either share the resources or transport data over long distances. Most commonly WSNs are integrated with the Internet in order to transform sensor data packets into IP packets and transport them over the Internet to the end-user. Apart from the integration challenges such as packet transformation and protocol adaptation, there are several security challenges that arise when resource-constrained devices such as WSNs are integrated with resource-rich networks such as the Internet. One of the main security challenges that arise with the implementation of in-processing techniques of data aggregation is the failure of providing end-to-end data confidentiality and data aggregation [7] of multi-sensor data [8] which is a vital security requirement in WSN-integrated networks. Traditionally when data aggregation is implemented in WSNs the sensor nodes transmit encrypted data to their cluster head nodes or intermediate nodes which then decrypts the data, performs data aggregation over the plaintext sensor data and then re-encrypts the aggregated plaintext data prior to transmission towards the next node. This gives rise to hop-by-hop security which consumes more energy for every decryption/encryption and also induces a delay for the additional processing, all of which prove expensive for WSNs in terms of resources. Apart from the added security feature of end-to-end data confidentiality, end-to-end security mechanisms remove the need for additional utilization of resources thereby proving to be more efficient compared to hop-by-hop security [9]. However, implementing data aggregation along with traditional end-to-end data confidentiality mechanisms is impossible since the data aggregation algorithm on the cluster head node or intermediate nodes would require access to the plaintext sensor data in order to aggregate the sensor data **[10]**. Research in this aspect has led to the design and development of a special category of encryption schemes called Privacy Homomorphic (PH) encryption schemes which ensure that end-to-end data confidentiality along with the implementation of data aggregation can be achieved [11].

A PH encryption scheme is an encryption transformation which allows direct computation over ciphertext data. For instance, let *Q* denote a modulus ring and + denotes the addition operations on the ring; if *K* is the key space and $a, b \in Q$ and $k, k_1, k_2 \in K$ then $a + b = Dec_k\,(Enc_k(a) + Enc_k(b))$ is termed as additively homomorphic with a single secret key *k* and $a + b = Dec_{k1,2}(Enc_{k1}(a) + Enc_{k2}(b))$ as additively homomorphic with *k1* and *k2* as the multiple secret keys [10]. Apart from the primary benefit of data aggregation and end-to-end data confidentiality, PH schemes have other benefits when implemented in WSNs, such as: the knowledge of encryption keys at intermediate nodes are not required since there are no intermediate decryptions of the ciphertext; the overall energy consumption of the network components and the induced latency can be reduced since there is no need for intermediate decryptions/encryptions. While most of the PH schemes presented in literature prove to be expensive for resource-constraint WSNs, there are certain low-power algorithms presented in [10] that may be used in WSNs. Out of all the schemes proposed in [10], the Castelluccia-Mykletun-Tsudik (CMT) [12] scheme is chosen due to its low computational costs and simplicity which proves efficient for resource-constraint WSNs. However, despite the low computational costs and simplicity, the CMT scheme has the ID-issue which increases the overhead when the number of responding sensor nodes increase. This is mainly because the IDs of individual sensor nodes that participate in the data aggregation process are appended to the aggregated ciphertext during each transmission towards the end-user thereby increasing the bandwidth and energy consumption. This paper proposes a modified version of the original CMT scheme in order to counter the ID-issue faced by the scheme and ensure that the proposed scheme is more energy and bandwidth efficient for large scale WSNs. The proposed scheme is a slot-based CMT scheme (SCMT) which transmits information pertaining to the security keys used during encryption in order to use the corresponding security keys for decryption such that the correct aggregated plaintext data is retrieved from the aggregated ciphertext data. This paper explains the working principle of the proposed SCMT scheme and through a simulation based performance evaluation proves that the SCMT scheme is more efficient than the original CMT scheme in terms of packet overhead induced and energy consumed which proves to be suitable for scalable resource-constrained WSNs.

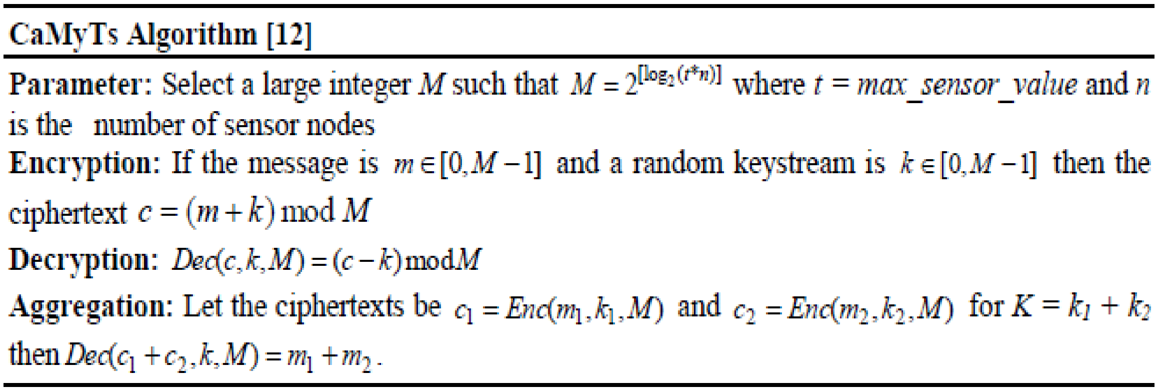
The remainder of this paper is as follows: Section 2 introduces the CMT algorithm and methods proposed by other authors to reduce the bandwidth consumption. Section 3 briefly highlights the security requirements along with the network architecture of the WSN-satellite integrated network. Section 4 presents the working of the proposed SCMT scheme and the initial setup phases prior to data transmission. Section 5 presents the analysis of the SCMT scheme in comparison to the original CMT scheme in order to prove the efficiency in terms of energy and induced overhead. Section 6 presents the future work and concludes the paper.

2. RELATED WORK

There are several Privacy Homomorphic (PH) encryption schemes that have been proposed for WSNs wherein their performances have been measured with respect to different metrics [10] [13]. PH schemes can be broadly classified into asymmetric [14] and symmetric PH schemes. This paper mainly focuses on the symmetric PH schemes due to their low computational expense which proves to be suitable for resource-constrained networks and their computational delay which is several orders of magnitude lower as compared to asymmetric PH schemes [5]. The following subsections provide a brief overview of the existing symmetric PH schemes.

2.1. Castelluccia Mykletun Tsudik Privacy Homomorphic Scheme (CMT)

The authors of **[12]** present an additively homomorphic stream cipher scheme known here as the CMT scheme, that allows data aggregation and end-to-end data confidentiality. The main advantage of this scheme is that it uses modular arithmetic with very small moduli making this highly suitable for resource-constrained WSNs. The scheme can be used to compute several calculations such as mean, variance and standard deviation of sensor data along with achieving a significant bandwidth gain. Each node is assumed to be an aggregator node which forwards aggregated data towards the sink node. The main concept behind the scheme is that the *xor* operation of the traditional stream cipher is replaced by the modular $\oplus$ addition operation. The keystream *k* can be generated by using a stream cipher, such as RC4, keyed with the node's secret key *si* and a unique message ID. This secret key is pre-computed and shared between the node and the sink. The message ID can either be included in the query from the sink or it can be derived from the time period in which the node is sending its value (assuming some form of synchronization). The compromise of an aggregator node by an adversary will not reveal any secret information since the plaintext data is not available at the aggregator node and hence the system proves to be more secure compared to the hop-by-hop encryption. However, a drawback of the proposed scheme is that the identities of non-responding nodes or responding nodes need to be sent along with the aggregated data value to the sink node which significantly increases the overhead costs. Another important drawback in terms of a potential attack is that the encryption transformation is malleable, i.e. an attacker can modify the content of the ciphertext into another known form of plaintext without knowing the plaintext or the secret key.



CaMyTs Algorithm [12]

Parameter: Select a large integer M such that $M = 2^{[\log_2(t*n)]}$ where t = *max_sensor_value* and n is the number of sensor nodes

Encryption: If the message is $m \in [0, M-1]$ and a random keystream is $k \in [0, M-1]$ then the ciphertext $c = (m+k) \bmod M$

Decryption: $Dec(c,k,M) = (c-k) \bmod M$

Aggregation: Let the ciphertexts be $c_1 = Enc(m_1,k_1,M)$ and $c_2 = Enc(m_2,k_2,M)$ for $K = k_1 + k_2$ then $Dec(c_1 + c_2, k, M) = m_1 + m_2$.

2.1.1. ID-Issue of the CMT Scheme

In order for the receiver to compute the aggregated plaintext data from the aggregated ciphertext data, the receiver should be able to determine the keysum value *K* which is the summation of all the security keys used during the secure data aggregation process at the cluster head node or intermediate nodes. The original CMT scheme transmits the nodes IDs (16-bit or 64-bit source addresses) along with the aggregated ciphertext value such that the receiver node can determine which security keys it needs to consider in order to calculate the keysum *K*. This proves to be expensive in terms of bandwidth consumption and induced packet overhead since as the number of responding sensor nodes increases the payload which contains the IDs also increases in size. Further increase in the number of responding sensor nodes leads to an increase in the number of data transmissions since the MAC frame of the IEEE 802.15.4 is limited to only 127 Bytes leaving a payload of around 100 Bytes in length. This is termed as the ID-issue which is being addressed in this paper with respect to reducing the induced packet overhead and number of transmissions leading to conservation of valuable WSN energy and bandwidth.

2.2. Domingo-Ferrer Privacy Homomorphic Scheme (DF)

The author of [15] introduced a symmetric probabilistic PH scheme wherein addition, subtraction, multiplication and division can be performed on ciphertext. The proposed homomorphism scheme was the first one to allow full arithmetic operations while being secure against the known-ciphertext attacks which requires that the ciphertext splitting is always used when encrypting with a splitting factor and the ciphertext space is much larger than the plaintext space. The DF algorithm uses a single symmetric key in every sensor node to encrypt the sensor data wherein the aggregator performs aggregation on the ciphertext and the end node decrypts the result using the same secret key. Thus, there is no need of appending the node IDs along with the ciphertext. However, this reduces the level of security despite the implementation of probabilistic encryption (randomness introduced during encryption) since the physical compromise of a single sensor node can disclose the secret key in use and induce several security attacks on the whole network. However, in the CMT scheme if a single sensor is physically compromised only a single security key is compromised and the decryption of the aggregated ciphertext is not possible with only a single security key since multiple security keys are used. The authors of [16] showed that with reasonable parameters the DF algorithm can be implemented in resource constrained devices. It is seen in [16] that value of *d* ranging from 2 to 4 proves to be beneficial for WSNs wherein the power consumption of the transmitting sensor nodes increases linearly with packet size.

2.3. Hybrid PH Scheme

The authors of [17] evaluate three homomorphic algorithms suitable for WSNs and propose two approaches to mitigate the weaknesses of the algorithms in terms of the potential attacks and low security level. The first approach is the successful combination of two algorithms (DF and CMT) which increases the security and the minimization of additional efforts by the careful selection of parameters such as the splitting factor. The second approach tackles specific weaknesses by considering homomorphic message authentication codes and also describes in detail the ID-issue of the CMT scheme. The authors of [17] integrate the CMT algorithm with the Concealed Data Aggregation (CDA) algorithm which is the DF scheme in order to create a cascading encryption such that $E_2(E_1(a)) \oplus E_2(E_1(b)) = E_2(E_1(a+b))$. The inner encryption function is CMT which suffers from the ID-issue and its malleable nature. However, the malleability can be countered by the implementation of an outer encryption transformation such as the DF, where the knowledge of the secret key is needed to modify the content of a single data packet. The advantage of this combination is that the aggregation requires exactly the same computational effort as the standalone DF. In order to counter the ID-issue of CMT scheme the authors of [17] applied the expression in Eq. 1, to determine the minimum number of bits required for reporting *e* randomly distributed nodes out of *n* sensor nodes. It was seen that this hybrid scheme of DF and CMT proves to be 5 times more efficient in terms of bandwidth gain. However, this hybrid scheme is not scalable to large number of sensor nodes as the further increase of responding nodes *e* increases the bandwidth utilization. The proposed scheme proves to be more efficient compared to the hybrid scheme and will be justified analytically in later sections.

$$log_2\left({}^{n}/_{e}\right) = log_2\left({}^{n!}/_{e!\,(n-e)}\right) = \sum_{i=n-e+1}^{n}\log_2 i - \sum_{i=1}^{e}\log_2 i \qquad \text{(Eq. 1)}$$

3. Network Architecture and Security Requirements

The overall network is an integration of terrestrial wireless sensor networks and the backbone network along with IP capabilities on the network components. The proposed network architecture can be divided into three sub-networks; namely the terrestrial WSN, the backbone network, and the integrating gateway. A remote monitoring station (RMS) is situated at the end of the integrated network which acts as the receiving end and the network controller. The primary role of the gateway component is to bridge the communication between the WSN and the backbone network. In this case the backbone network is considered as the Internet and the gateway is mainly responsible for transporting IP-based sensor data over the Internet to the RMS. The architecture of the gateway node for the integration with the backbone network is beyond the scope of this paper and it is assumed that the gateway is able to transform data received by the CH nodes into IP data wherein the entire data packet of the CH node acts as the IP payload in the IP data packet. The overall network architecture is as shown in Fig. 1 wherein the various network components are highlighted along with the expected flow of data within the integrated network.

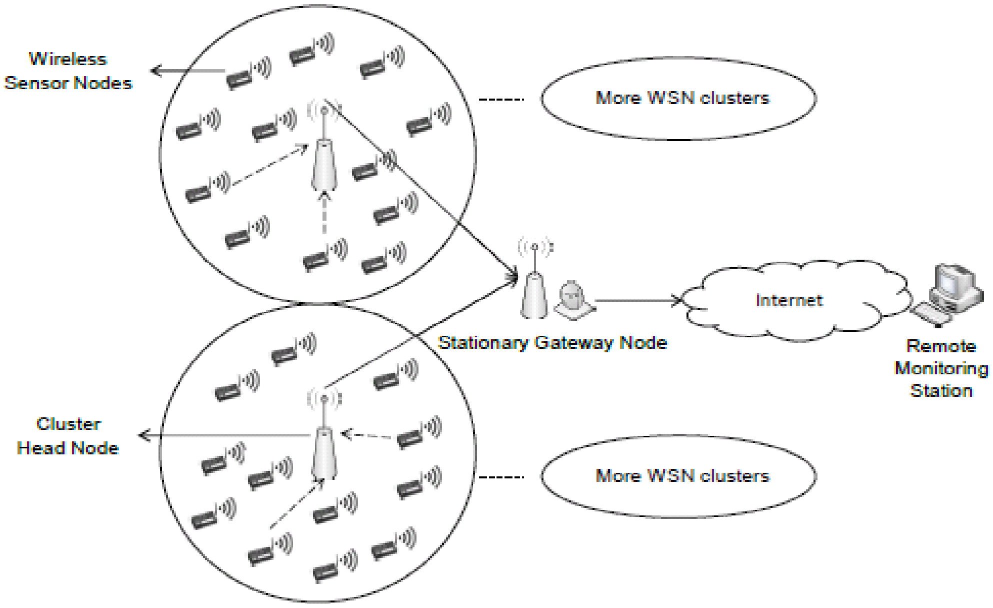


Figure 1. Overall network architecture of the WSN-satellite integrated network used for analysis

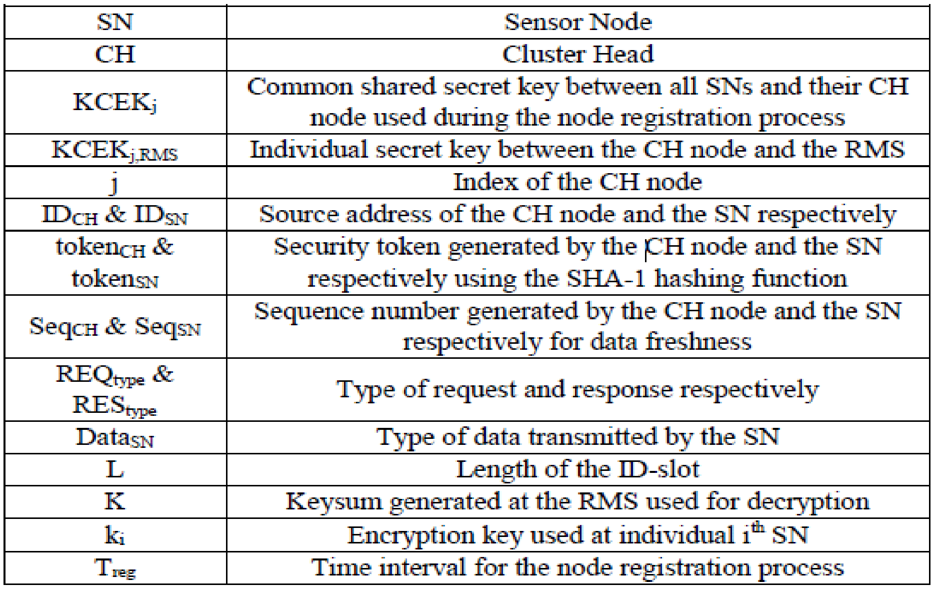
Typically, the WSN consists of several wireless sensor nodes which are deployed within the sensing region in a star topology cluster-based node layout scheme [18], as shown in Fig. 1, wherein the sensing region is divided into multiple clusters and each Sensor Node (SN) within the cluster is randomly distributed within the transmission range of the Cluster Head (CH) node. This proves to be more efficient than a non-cluster based node layout scheme where all the sensor nodes are randomly deployed within the sensing region and directly communicate with the gateway node, since the overall load is being evenly distributed among various clusters thereby increasing the operating lifetime of the network components within the WSN [13].

The SNs are responsible for encrypting the sensor data prior to data transmission towards the CH nodes which are mainly responsible for the in-processing technique of data aggregation. It is important to note that there are no intermediate decryptions at the CH nodes or any other nodes and that the CH nodes perform data aggregation over ciphertext sensor data. The primary security requirements of the integrated network considered in this paper include data confidentiality and data integrity. Data confidentiality is achieved with the implementation of a symmetric homomorphic encryption scheme whereas data integrity can be achieved with the implementation of a Message Authentication Code (MAC) of the ciphertext sensor data. The secondary security requirements within the integrated network include counter against replay attacks and traffic analysis which can be achieved by appending a frame counter or sequence number for data freshness and header encryption to avoid traffic analysis.

4. Proposed Slot-based CMT Scheme (SCMT)

This section presents the proposed energy-efficient additive privacy homomorphic encryption scheme which is highly scalable in WSNs for multi-sensor data. The proposed slot-based CMT (SCMT) scheme mainly implements a slotting algorithm in order to counter the ID-issue of the homomorphic encryption scheme proposed in [12]. Table 1 shows the notations used to describe the proposed scheme.

Table 1. Notations and symbols used



SN	Sensor Node
CH	Cluster Head
$KCEK_j$	Common shared secret key between all SNs and their CH node used during the node registration process
$KCEK_{j,RMS}$	Individual secret key between the CH node and the RMS
j	Index of the CH node
ID_{CH} & ID_{SN}	Source address of the CH node and the SN respectively
$token_{CH}$ & $token_{SN}$	Security token generated by the CH node and the SN respectively using the SHA-1 hashing function
Seq_{CH} & Seq_{SN}	Sequence number generated by the CH node and the SN respectively for data freshness
REQ_{type} & RES_{type}	Type of request and response respectively
$Data_{SN}$	Type of data transmitted by the SN
L	Length of the ID-slot
K	Keysum generated at the RMS used for decryption
k_i	Encryption key used at individual i^{th} SN
T_{reg}	Time interval for the node registration process

The scheme mainly consists of the one-time initialization phase followed by the periodic, continuous or event-driven sensor data transmission phase. The initialization phase consists of the following processes; node registration process, ID slot generation and transmission process, and the generation of the Key-ID table process. During the node registration process the individual CH nodes communicate with the SNs within the transmission range of the CH nodes in order to create the cluster and create an ID-table of all the SNs within the cluster at the CH node. The communication between the CH nodes and the SNs during the node registration process is encrypted using the common shared secret key *KCEKj* between all SNs and the CH node where *j* represents the index for the respective CH node. The node registration process is followed by the ID-slot generation and transmission process wherein the proposed slotting scheme is mainly operational. The ID-slots are generated at the CH nodes using the ID-table generated during the node registration process and the generated ID-slots for the respective clusters for different types of sensor data (temperature, humidity, pressure, etc) are transmitted to the RMS via the gateway node. The data exchanged between the CH nodes and the RMS node is encrypted using the individual shared secret key *KCEKj,RMS* between the individual CH node and the RMS where *j* is the index of the respective CH node. The final process of the initialization phase prior to sensor data transmission is the generation of the Key-ID table process wherein the RMS receives the ID information pertaining to the IDs of the SNs and the individual cluster ID-slots for different sensor data types which are then used to generate a Key- ID table. This Key-ID table plays a crucial role in the decryption function of the proposed SCMT scheme and is vital prior to any sensor data transmission. After the execution of the 3 processes of the initialization phase is successfully completed the transmission of encrypted sensor data from the SNs towards the RMS takes place. During this data transmission, the individual encryption keys of the SNs are used for encrypting the sensor data. The same keys are maintained at the RMS for the decryption of the sensor data thereby providing end-to-end data confidentiality. The generation and distribution of the encryption keys are beyond the scope of this paper and it is assumed that all network entities are aware of the security keys. The message exchanges between all network entities during the initialisation phase and data transmission can be summarized in Fig. 2.

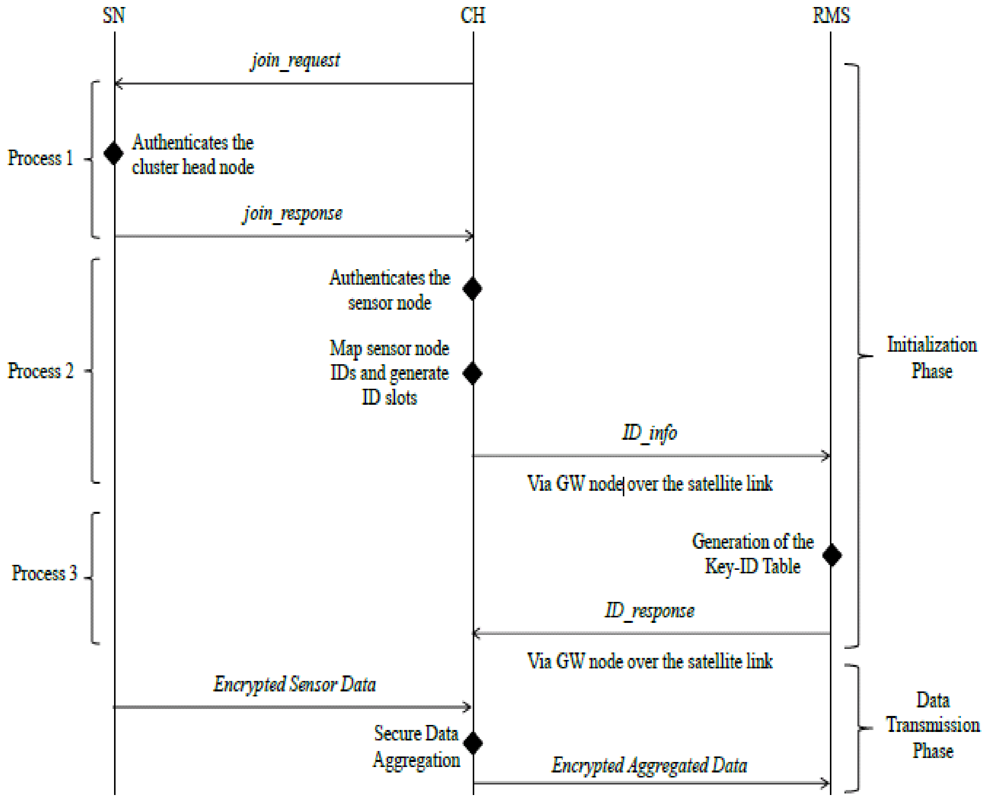

SN
CH
RMS
join_request
Process 1
Authenticates the cluster head node
join_response
Authenticates the sensor node
Process 2
Map sensor node IDs and generate ID slots
Initialization Phase
ID_info
Via GW node over the satellite link
Process 3
Generation of the Key-ID Table
ID_response
Via GW node over the satellite link
Encrypted Sensor Data
Secure Data Aggregation
Encrypted Aggregated Data
Data Transmission Phase

Figure 2. Summary of the Initial Setup Phases of the Proposed Scheme

4.1. Node Registration Process

The node registration phase is mainly responsible for the registration of all SNs within the individual clusters in order to generate their respective ID tables at the CH nodes. The node registration phase mainly consists of the following steps:

- Each CH node broadcasts a *join_request* message to all the SNs within its cluster encrypted using the pre-shared common encryption key *KCEKj* which is shared between all SNs and their CH node. The *join_request* message between the CH node and the SN is encrypted in order to prevent eavesdropping and traffic analysis attacks.

$$join\_request \coloneqq CH \rightarrow SN{:}REQ_{type}||\{ID_{CH}||token_{CH}||Seq_{CH}\}KCEK_j$$

- The sequence number is a monotonically increasing counter value which corresponds to a particular *join_request* message thereby preventing an attacker from transmitting old *join_request* messages. The security token is used by the SNs to authenticate the CH node. It is generated using the SHA-1 cryptographic hash function with {*IDCH*, *SeqCH*, *KCEKj*} as the input parameters.

- The individual SNs receive the encrypted *join_request* message and decrypt the message using the security key *KCEKj*. The first step is to verify the sequence number of the *join_request* in order to determine data freshness. The second step is to ensure data integrity by checking if the security token transmitted is the same as the security token generated at the SN using the input parameters {*IDCH, SeqCH, KCEKj*} obtained from the decrypted *join_request* message. If the integrity check fails at the individual SN then the corresponding SN will not register the CH node as its cluster head.

- After the respective checks at the individual SNs, they generate and transmit a *join_response* message encrypted with the same security key *KCEKj*, thus preventing eavesdropping and message modification of the *join_responses* which can lead to the generation of incorrect ID-slots at the CH node. The sequence number is the same as the sequence number of the corresponding *join_request* message and the security token is generated using {*IDCH, SeqCH, IDSN, KCEKj*} as the input parameters.

$$join\_response \coloneqq SN \rightarrow CH: RES_{type} || \{ID_{CH} || ID_{SN} || token_{SN} || Data_{SN} || Seq_{CH}\} KCEK_j$$

- Upon reception of the individual *join_response* messages from the SNs, the CH node verifies if the sequence number of the *join_response* message is the same as the corresponding *join_request* message in order to ensure data freshness. The CH node then proceeds to authenticate the individual SNs by verifying if their transmitted security token in the *join_response* messages are the same as the security tokens generated by the CH node using the corresponding SHA-1 cryptographic hash function with {*IDCH, SeqCH, IDSN, KCEKj*} as the input parameters.
- Upon completion of the two verification processes at the CH node it successfully registers the SNs and proceeds with the next process of ID-slot generation and transmission.

4.2. ID-slot Generation and Transmission Process

This phase is mainly responsible for the generation of the ID-slot which is the novelty of the proposed scheme that addresses the ID-issues of the original CMT scheme in [12]. The ID-slot with respect to the proposed SCMT scheme can be defined as a bit-field that contains information pertaining to the status of the decryption keys used for the respective SN IDs. Each bit-field represents the security-key status of an individual SN where a 1 represents that the security-key of the respective SN was used during the aggregation function of ciphertexts and a 0 represents that the security-key of the respective SN was not used. Upon analysing this ID-slot the RMS can calculate the required keysum *K* in order to decrypt the aggregated ciphertext data. It is important to note that the ID-slot generation occurs before the elapse of a certain time interval *Treg*, which is the time required for every SN within the cluster to successfully register with its CH node after the broadcast of the *join_request* message. The ID-slot generation phase mainly consists of the following steps:

- Once an individual SN has been successfully verified within the time period *Treg*, the ID of the SN is retrieved from the *join_response* message and stored in the ID-table wherein the index of the ID-table forms the ID-slot for a particular sensor data type, as shown in Figure 3.
- When a SN is successfully verified, the ID of the SN is updated into the ID-table followed by the index of the ID-table, *x,* where the value of *x, i.e.* 1 or 0 can be used to denote if the security key of the corresponding SN ID was used during the aggregation process in order to generate the aggregated ciphertext data. The length of the ID-slot is denoted by *L* which is directly proportional to the number of operational SNs within a cluster. For instance, within the time interval of *Treg*, if there are 64 temperature sensor nodes that have been successfully registered then the corresponding ID-slot generated for temperature data will be a 64-bit field where each bit (starting from the LSB) represents the state of the corresponding SN IDs.
- The ID-slot is constantly being generated until the lapse of *Treg*, wherein the ID-slot is finalized and transmitted in the form of an *ID_info* to the gateway node. It is important to note that the transmission of the respective sensor node IDs is a onetime process in order to ensure that all the network components are aware of the participating nodes and the structure of the corresponding ID-slots. In order to protect the ID-slot and the other slot information from eavesdropping and message modification, the *ID_info* is encrypted using the shared secret key *KCEKj,RMS* between the CH node and the RMS.

$$ID\_info \coloneqq CH \rightarrow RMS: REQ_{type}||\{ID_{RMS}||ID_{CH}||L||Data_{SN}||ID_{SN1}||ID_{SN2}||\ldots||ID_{SNL}\}KCEK_{j,RMS}$$

- The *ID_info* is then forwarded to the RMS via the gateway node over the backbone network. The gateway node does not access the slotting information or update the ID-slot information during the initialisation phase.

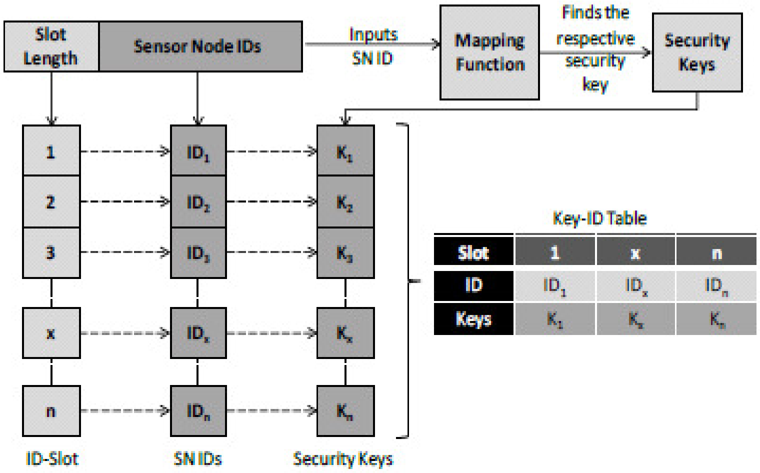
4.3. Generation of Key-ID Table

The primary function of the RMS is to decrypt the aggregated ciphertext data that it receives over the Internet. In order to do this, the RMS must be aware of the security keys used by the individual SNs along with which SNs participated in the aggregation function at the cluster head in order to generate the aggregated ciphertext data. For this, the RMS maintains a Key-ID table of the sensor node IDs within each cluster and their respective security keys. The procedure for generation of this Key-ID table is as follows: each cluster and their respective security keys. The procedure for generation of this Key-ID table is as follows:

- Upon reception of the *ID_info* forwarded from the GW node to the RMS over the Internet, the RMS decrypts the *ID_info* using the security key *KCEKj,RMS*, which is a pre-shared key between the individual jth CH node and the RMS, in order to retrieve the

SN IDs and the corresponding ID-slot for the respective sensor data type. This information is used to create the Key-ID table as shown in Figure 4.

- The sensor node IDs from the ID_info are mapped accordingly with the respective security keys of the SN IDs under the ID-slot. The ID-slot, retrieved from the slot length *L*, is created as a blank slot and used as an index for the Key-ID table.
- Typically, the RMS will access the ID-slot in order to retrieve the length of the ID-slot and determine the total number of SN IDs in the *ID_info* received from the GW node. The RMS will then use a look-up table of wireless sensor node IDs mapped with their respective security keys in order to map the ID-slot and the security keys wherein the ID-slot will be used as an index to determine which SNs were responsible for the generation of the aggregated ciphertext data.
- It is important to note that the order in which the SN IDs are arranged in the ID-table of the CH node is retained at the RMS such that the order of SN IDs is the same. The only inclusion in the Key-ID table is the respective security keys of the SNs. Upon the successful completion of the Key-ID table generation, the RMS transmits an *ID_response* message acknowledging that the ID-slot formats have been successfully mapped with the security keys and that the RMS is ready to accept aggregated ciphertext data.



Slot	1	x	n
ID	ID_1	ID_x	ID_n
Keys	K_1	K_x	K_n

Figure 4. Key-ID table generation at the RMS node

4.4. Data Transmission Phase

The data transmission phase is operational only after the successful completion of the initialisation phases such that the ID tables at the CH node and the RMS are successfully generated for each cluster. Once the initialisation phase is complete the CH node broadcasts a Clear to Send (CTS) signal to all its SNs. During the data transmission phase the individual sensor nodes encrypt the sensor data with their respective encryption key *ki* and forward the encrypted sensor data to its CH node. The CH node then performs data aggregation over the ciphertext data received from several SNs within the cluster followed up the updating of the ID-slots. Based on the source address of the transmitting SNs (ID) the CH node is able to update the ID-slot indicating that the respective SN was responsible for transmitting an encrypted sensor data that was used during the aggregation function. Once the aggregation time period has elapsed the CH node transmits the aggregated ciphertext data along with updated ID-slot to the RMS via the gateway node. Upon reception of the aggregated ciphertext data, the RMS analyses the ID-slot to retrieve information about the security keys it needs to use in order to calculate the keysum K used for the decryption of the aggregated ciphertext sensor data. The status of the individual bit fields of the ID-slot indicates whether the security key of the SN was used during aggregation wherein 1 indicates that the security key for the particular SN was used and 0 indicates that the security key of the particular SN was not used. This phase continues periodically as the SNs continue to transmit encrypted data to the CH node, the CH node performs data aggregation and ID-slot updating followed by data transmission to the RMS and decryption at the RMS. An example implementation of the SCMT scheme in WSNs will shed more information about the working of the proposed scheme. Let us consider that the above mentioned initialization phases are complete and that the CH node has successfully generated an ID-slot for 64 SNs within the cluster measuring ambient temperature and that the RMS has successfully generated the Key-ID table for the 64 SNs. During a given dissemination interval of 2 seconds for sensor data and 2.5 seconds for aggregated sensor data, for the sake of simplicity with respect to this example let us assume that 3 sensor nodes which correspond to the 16th, 29th and 50th bit in the ID-slot are responsible for the generation of the aggregated ciphertext data, *C* as shown in Eq. 2 to Eq.4.

$$c_1 = (m_{16} + k_{16}) modM \qquad c_2 = (m_{29} + k_{29}) modM \qquad c_3 = (m_{50} + k_{50}) modM \qquad \text{(Eq. 2)}$$

$$C = \sum_{i=1}^{3} c_i = (m_{16} + m_{29} + m_{50}) + (k_{16} + k_{29} + k_{50}) modM = \left(m_{aggr} + K\right) modM \qquad \text{(Eq. 3)}$$

$$m_{aggr} = (C - K) modM \qquad \text{(Eq. 4)}$$

First, the ID-slot is generated at the CH node wherein the corresponding states of the bit fields of the ID-slot (16, 29 and 50) are updated to 1. After the ciphertext *C* is generated, the ID-slot information is appended to the ciphertext data and transmitted to the GW node along with the auxiliary security header which contains information about the type of sensor data and frame counter used to counter replay attacks. The data packet is transmitted from the CH node to the RMS via the gateway node which is connected to the backbone network. The RMS receives the ID-slot for the aggregated ciphertext of temperature data and processes the ID-slot (starting from the LSB) in order to determine the SN IDs responsible for the aggregated ciphertext data (16, 29 and 50) as shown in Fig. 5. The security keys (k16, k29 and k50) for the corresponding SN IDs are retrieved from the Key-ID table generated prior to data transmission and the key sum *K* is generated in order to decrypt the aggregated ciphertext data *C* to retrieve the aggregated plaintext summation of temperature sensor data *maggr* at the RMS.

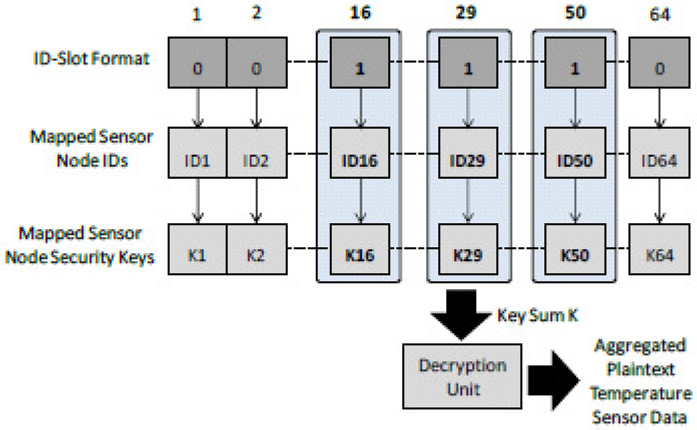


Figure 5. Theoretical Implementation of the Proposed Scheme

5. PERFORMANCE EVALUATION OF THE SCMT SCHEME

5.1. Preliminary Simulation Parameters

For the successful operation of the CMT and SCMT scheme and data aggregation within the WSN it is important that the network components agree upon certain parameters which include the following; a large integer M, the maximum sensor value V_{max} and the maximum number of sensor nodes n, which satisfy the condition $M > V_{max} . n \; \forall \; t \; and \; n$. This is mainly done in order to avoid incorrect decryptions when multiple ciphertext sensor data are aggregated. Due to the resource constraint nature of the WSNs the size of the ciphertext values are generally kept short ranging from 8 to 20 Bytes in length. This corresponds to an approximate integer value of 1.8447*1019 – 1.4615*1048 for the large integer parameter M used for encryption and decryption. The value of M is distributed to all SNs and CH nodes during the key-distribution phase which is beyond the scope of this paper. The analysis was carried out for a selection of as the value of M, which corresponds to a ciphertext size of 8 Bytes in length as this suffices for the range of applications being targeted and also proves to be lightweight for the sensor nodes in terms of processing and transmission. The value of M can be obtained by $2^{log_2(n.V_{max})}$ where $log_2(n.V_{max})$ is the size of the ciphertext 8 Bytes (64 bits). By increasing the value of M the encryption can be made stronger since the security key employed during encryption will belong to a larger set $[0, M - 1]$, thereby decreasing the probability of determining the security key by brute force methods.

5.2. Packet Structures

The WSN standard adopted in the numerical analysis to evaluate the performance of the SCMT scheme is the 6LoWPAN standard which implements IPv6 over sensor network data. For the purpose of the proposed SCMT scheme an extension needs to be added to the current 6LoWPAN packet format as shown in Figure 6 for SNs and CH nodes. The IEEE 802.15.4 header is introduced in the MAC layer followed by the 6LoWPAN header introduced in the network layer followed by the auxiliary security header introduced at the network layer and lastly followed by the ciphertext data. The auxiliary header for the SNs mainly contains information pertaining to the type of secure data transmitted (temperature, humidity, pressure, etc.) and a frame counter for countering replay attacks. The auxiliary header for the CH nodes consists of the data type field and frame counter field along with the ID-slot which is generated during data aggregation at the CH node during data transmission of the aggregated ciphertext data towards to the RMS. It is seen from the figure that the allowable data payload field is shown thereby setting a maximum payload length (around 102 Bytes for 16-bit addressing) in order to incorporate the auxiliary header, extended frame control and ID information. The unencrypted information enables the intermediate nodes such as the CH node to determine what security operations need to be carried out on the data packet.

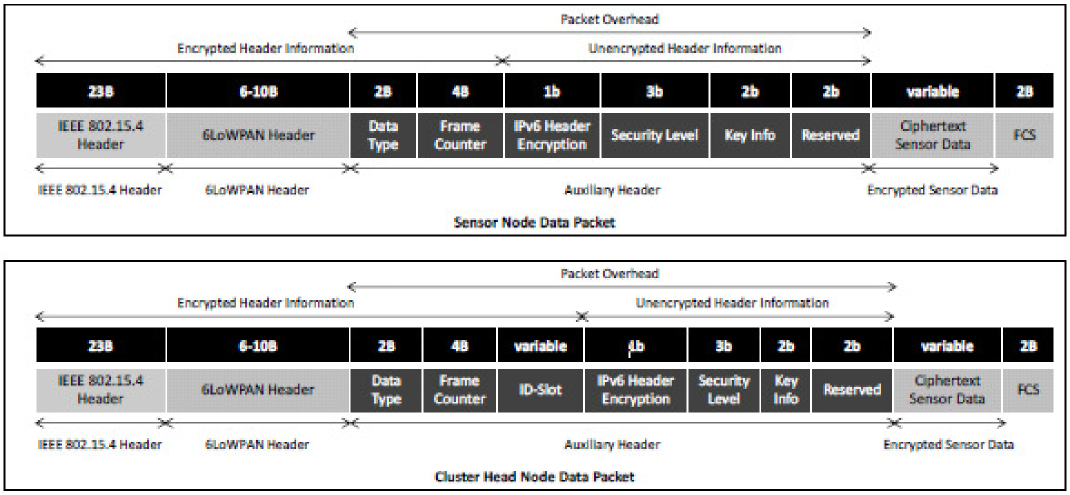


Figure 6. General packet structure of the sensor node and the cluster head node data packet

5.3. Theoretical Analysis

5.3.1. Overhead Analysis

The analysis carried out in this section mainly determines the packet overhead of the original CMT scheme as the number of sensor nodes are increased compared to the proposed SCMT scheme. The overall encrypted packet size which includes the auxiliary header and the ciphertext is also compared for the two schemes. It is evident from the graph in Fig. 7 that the packet overhead and the overall packet size increases linearly along with the increase in the number of sensor nodes participating in the data aggregation process. It is seen that the gradient for the overhead size in the original CMT scheme is much higher compared to the gradient of the proposed scheme. For the increase of every 100 SNs in the original CMT scheme the overhead increases by 200 Bytes which proves to be very expensive for SNs. However, the proposed SCMT scheme shows significant efficiency in terms of overhead wherein for every 100 SNs the overhead is increased by just over 13 Bytes which proves to be 93.5% more efficient than the original CMT scheme, thus proving to be highly scalable and bandwidth efficient for WSNs. Also, it can be seen from Fig. 7 that the proposed SCMT scheme seems to accommodate more number of SNs within the cluster for a single data transmission (below dotted line labelled 1 in Fig. 7) from the CH node whereas for the original CMT scheme the number of transmissions from the CH node increases with the increase in the number of SNs within the cluster. It is seen that for every 50 to 60 SNs the number of transmissions from the CH node increases by 1 thereby consuming more energy for the transmission of the aggregated ciphertext data along with the consumption of more bandwidth.

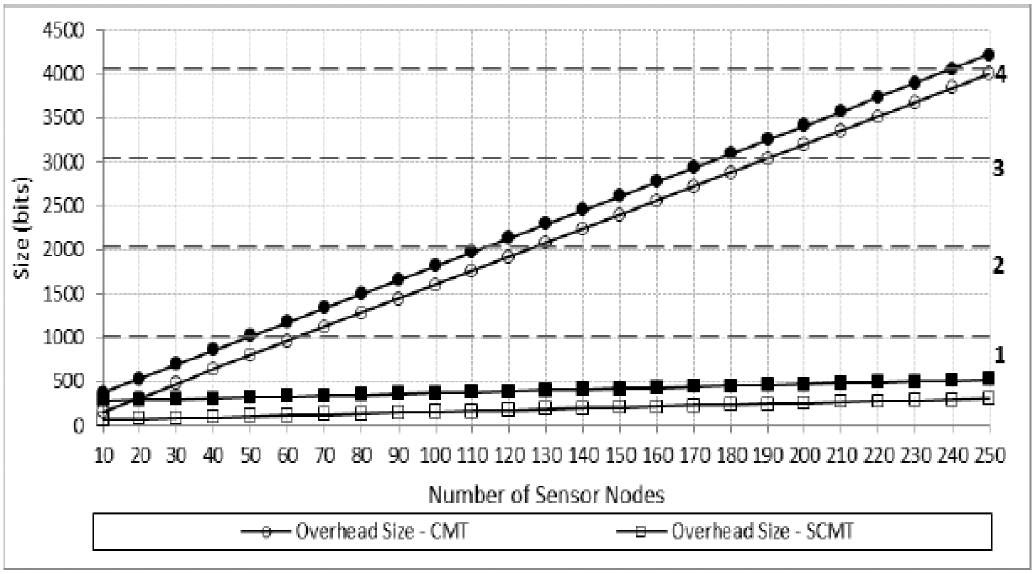


Figure 7. Comparative analysis of the induced packet overhead and total packet size transmitted from the CH node for the original CMT and proposed SCMT schemes

The authors of [17] proposed a solution to counter the ID-issue wherein the IDs of only responding or non-responding nodes (whichever is lesser) are transmitted along with the aggregated ciphertext data. This proved to reduce the overheads by 1/5th which is a significant decrease compared to the original CMT scheme. In this analysis the proposed SCMT scheme is compared with the solution proposed in [17] in order to prove that the SCMT scheme is still significantly more efficient for large number of SNs. From Fig. 8 it is seen that for varying node transmission percentages from 10% to 50% the overhead of the original CMT scheme proves to be slightly lower in some cases compared to the proposed scheme.

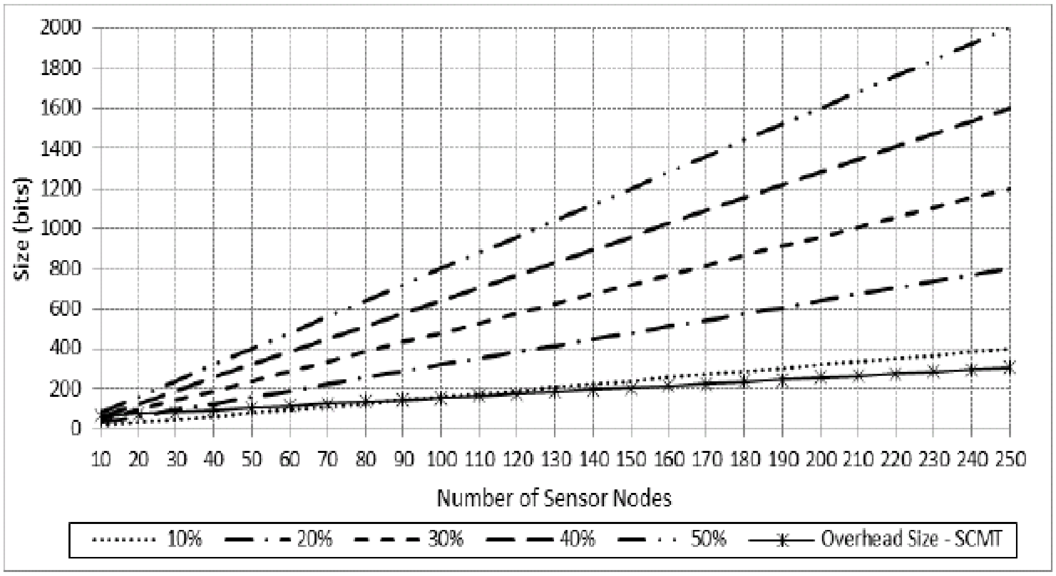


Figure 8. Comparison of overhead size of the original CMT scheme for different percentage ratios with the proposed SCMT scheme

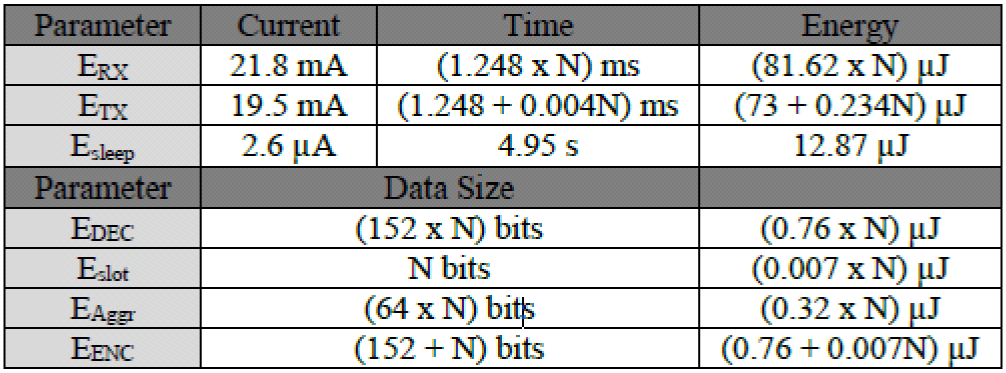
The node transmission percentage in this case can be defined as the percentage of responding or non-responding to the total number of sensor nodes within the cluster. For instance, when the percentage is 10% it may mean that 90% of the sensor nodes are responding and 10% are non-responding or vice-versa. In this case it is more efficient to transmit the IDs of the 10% SNs (responding/non-responding) rather than all the SN IDs. From Fig. 8 the overhead of the hybrid scheme [17] proves to be 75.75% more efficient for a minimum of 10 SNs at node transmission percentage of 10% and 1.37% more efficient for a maximum of 90 SNs at node transmission percentage of 10%. Similarly, for a maximum node transmission percentage of 40% the original CMT schemes proves to be only 3.03% more efficient for a maximum of 10 SNs compared to the proposed SCMT scheme. For the node transmission percentage of 50% the original CMT scheme proves to be 17.5% less efficient compared to the SCMT scheme for a minimum of 10 SNs under a single CH node. It is evident from Fig. 8 that as the node transmission percentage increases the hybrid scheme [17] proves to be less efficient compared to the proposed SCMT scheme for increasing number of SNs under a single CH node. Despite the original CMT scheme being more efficient compared to the SCMT scheme in some cases, it is unlikely that there would be a low node transmission percentage of SNs. Even if the node transmission percentage is low and the number of SNs increase then the hybrid scheme starts to deteriorate in terms of efficiency, thus limiting the number of SNs and not being scalable. Lastly, it is evident from Figure 8 that the proposed SCMT scheme is independent of the node transmission percentage since the ID-slots are formulated according to the total number of successfully registered SNs within the cluster and not based on the responding/non-responding SNs within the cluster.

5.3.2. Numerical Analysis of Energy Consumption at the CH Node

The energy analysis is only carried out for the energy consumption of the cluster head node which implements the slotting scheme. For the original CMT scheme, it is the cluster head node which is mainly affected in terms of energy consumption with the increase in the number of SNs due to the ID-issue which increases the packet overhead and hence the energy required for transmission of multiple data packets. The RMS is considered as a powerful device and hence the energy consumption analysis is not carried out for the RMS node. The main processes which consume energy at the CH node are the header decryptions/encryptions; data aggregation; slotting mechanism; data reception/transmissions; and the sleep period of the CH node. For the analysis a cluster head node dissemination interval of 5 seconds is considered wherein the CH node remains in the active state (rate) for 1% of the total dissemination period thereby leaving the CH node for a 99% sleep period. The energy model for the CH node is derived from [19] and the following current and energy consumption for the sky tmote sensor [20] is as shown in Table 2. The operating voltage of the CH node is 3V and the total energy consumption of the CH node during the 5s dissemination period is given in Eq. 5 wherein $E_{RX,\ N}$ is the energy consumed for the reception of the encrypted sensor data packets from N sensor nodes; E_{DEC} is the energy consumed for the decryption of the encrypted header of the sensor data packet; $E_{Aggr,\ N}$ is the energy consumed for the aggregation of the data from N sensor nodes; E_{slot} is the energy consumed for the generation of the slots; E_{ENC} is the energy consumed for the encryption of the header of the CH node data packet; E_{TX} is the energy consumed to transmit the single CH node data packet; and E_{sleep} is the energy consumed during the sleep period of the CH node. From the study of energy consumption for WSNs in [19] it is seen that for simple data aggregation functions the energy consumed is 5 nJ/bit and the energy consumed for the processing (encryption, decryption and slot generation) is 7 nJ/bit. The total energy is calculated using the energy model for the CH nodes, as given in Eq. 1, for varying number of SNs within the cluster for the CH dissemination interval of 5 seconds is shown in Fig. 9 which compares the original CMT scheme with the proposed SCMT scheme. Fig. 10 on the other hand compares the energy consumption between the scheme proposed in [17] and the SCMT scheme for different node transmission percentages for both the schemes. It is evident from Fig. 9 that as the number of SNs increase within a cluster the energy consumption of the CH node increases and that the proposed SCMT scheme proves to be 11.86% for a maximum of 250 SNs. This is mainly due to the reduced packet overhead introduced by the slotting scheme and the reduced number of transmissions.

$$E_{total} = E_{RX} + E_{DEC} + E_{Aggr} + E_{slot} + E_{ENC} + E_{TX} + E_{sleep} \qquad \text{(Eq.5)}$$

Table 2: Energy parameters used for the implementation of the SCMT energy model



Parameter	Current	Time	Energy
E_{RX}	21.8 mA	(1.248 x N) ms	(81.62 x N) µJ
E_{TX}	19.5 mA	(1.248 + 0.004N) ms	(73 + 0.234N) µJ
E_{sleep}	2.6 µA	4.95 s	12.87 µJ
Parameter	**Data Size**		
E_{DEC}	(152 x N) bits		(0.76 x N) µJ
E_{slot}	N bits		(0.007 x N) µJ
E_{Aggr}	(64 x N) bits		(0.32 x N) µJ
E_{ENC}	(152 + N) bits		(0.76 + 0.007N) µJ

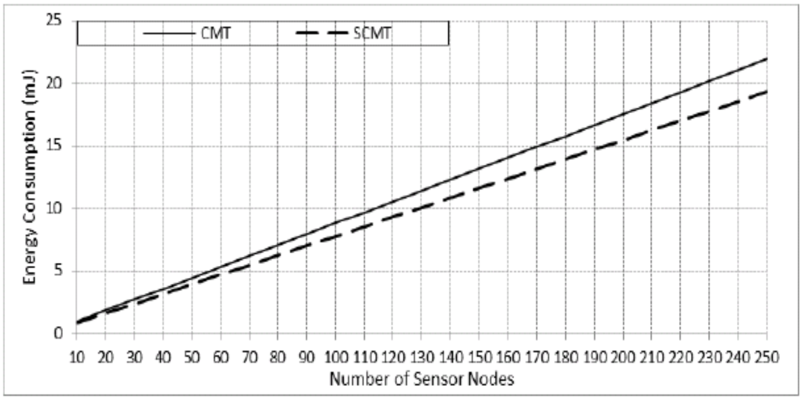
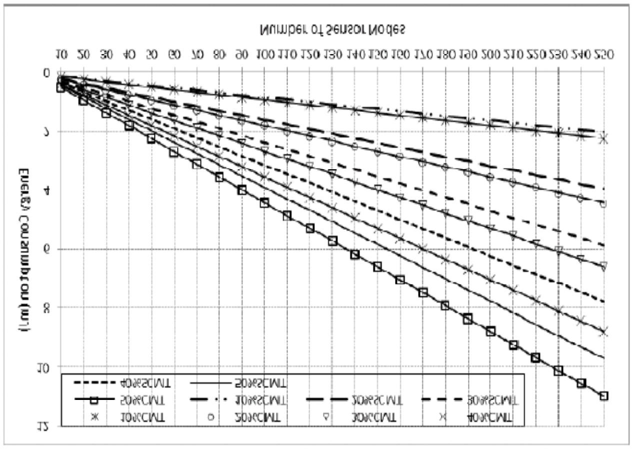


Figure 9. Comparative analysis of energy consumption between the original CMT and the proposed SCMT schemes

From Fig. 10 it is evident that as the node transmission percentage increases the energy consumption of the original CMT scheme increases compared to the proposed SCMT scheme; thus proving that SCMT scheme is 11.57% more energy efficient compared to the CMT scheme for a maximum of 250 SNs at the node transmission percentage of 50%. A possible drawback of the proposed SCMT scheme could be the initial setup phases as mentioned above since the increase in the number of SNs within the cluster increases the number of transmissions from the CH node during the setup phase wherein the IDs of the participating SNs within the cluster are transmitted within the time interval *Treg*. The total energy consumption of the proposed SCMT scheme at the CH node during the setup phase includes broadcasting of the *join_request*, reception of the individual *join_responses*, transmission of the *ID_info* and reception of the *ID_response*. The energy consumption by the CH node for a maximum of 250 SNs was calculated at 24.65 mJ; however this is only the one-time energy consumption during the setup phase.



5.3.3. Number of Transmissions during the Setup Phase

Fig. 11 shows the number of transmissions during the setup stage against varying number of SNs within the cluster. It is seen that with the increase in the number of sensor nodes, the number of transmissions from the CH node increases. This is because the SN IDs have to be appended to the *ID_info* and transmitted to the GW node using the IEEE 802.15.4 MAC frame of 127 Bytes. However, it is important to note that the setup phase is a one-time and the energy costs of the setup phase can be considered to be acceptable as it helps saves energy in the subsequent data transmission phase.

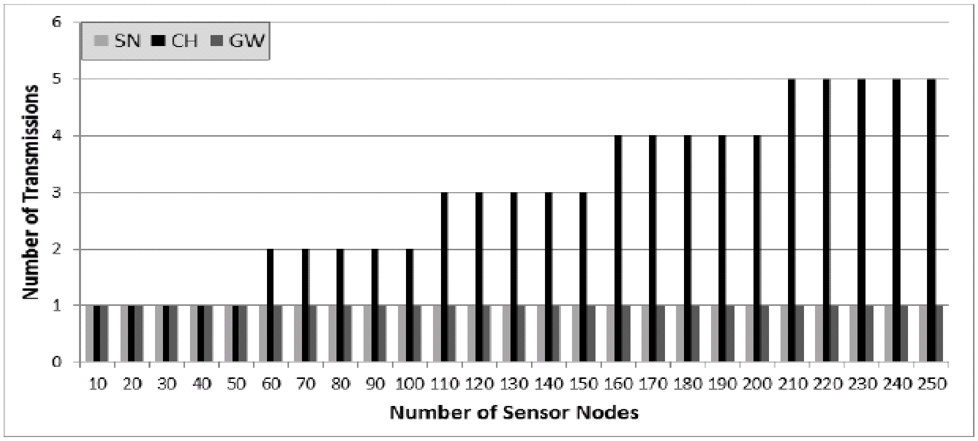


Figure 11. Number of transmissions from individual network entities during setup phase of the proposed scheme

6. FUTURE WORK AND CONCLUSION

The proposed SCMT scheme mainly addresses the ID-issue of the CMT algorithm presented in [12]. However, in order to increases the security level we propose to integrate the proposed SCMT scheme with probabilistic homomorphic encryption schemes which introduces some sort of randomness in order to increase the security level. Nevertheless, most symmetric homomorphic encryption schemes transmit the IDs of responding/non-responding sensor nodes which increase the overhead and energy consumption for data transmission thus making the implementation of security expensive in WSNs. Firstly, it is seen that the proposed SCMT scheme is 17.537% lighter in terms of the induced overhead compared to the original CMT scheme for a minimum percentage rate of transmission of 10% and a maximum of 250 SNs. Secondly, it is seen that the proposed SCMT scheme is 14.21% more energy efficient for a maximum of 250 SNs compared to the original CMT scheme at 10% percentage rate of transmission which is a significant conservation of valuable sensor node energy which can further increase the lifetime of the WSN. Lastly, it is seen that the proposed SCMT scheme is highly scalable since the increase in the number of sensor nodes within the cluster does not significantly increase the energy consumption or size of the ciphertext sensor data unlike the original CMT scheme.

REFERENCES

[1] I.F. Akyildiz, W. Su, Y. Sankrasubramaniam and E. Cayirci, “Wireless Sensor Networks: A Survey,” Computer Networks, vol. 38, no. 4, pp. 393-422, 2002.

[2] W.Colitti, K. Steenhaut and N. De Caro, “Integrating Wireless Sensor Networks with the Web,” Presented at IP+SN, 2011.

[3] L.Mainetti, L. Patrono and A. Vilei, “Evolution of wireless sensor networks towards the internet of things: A survey,” Presented at 19th International Conference on Software, Telecommunications and Computer Networks (SoftCOM), 2011.

[4] L.Li, H. Xiaoguang, C. Ke and H. Ketai, “The Applications Of WiFi-based Wireless Sensor Network in Internet of Things and Smart Grid,” 6th IEEE Conference on Industrial Electronics and Applications (ICIEA), 2011, pp. 789-793.

[5] L.Zhang and Z. Wang, “Integration of RFID into Wireless Sensor Networks: Architectures, Opportunities and Challenging Problems,” Fifth International Conference on Grid and Cooperative Computing Workshops, 2006, pp. 463-469.

[6] N.Celandroni, E. Ferro, A. Gotta, G. Oligeri, C. Roseti, M. Luglio, L. Bisio, M. Cello, F. Davoli, A. D. Panagopoulos, M. Poulakis, S. Vassaki, T. De Cola, M. A. Marchitti, Y. F. Hu, P. Pillai, S. Verma, K. Xu, G. Acar, “A Survey of Architectures and Scenarios in Satellite-based Wireless Sensor Networks: System Design Aspects”, International Journal of Satellite Communications and Networking, vol. 31, no. 1, pp. 1-38, 2013.

[7] S.Ozdemir and Y. Xiao, “Secure data aggregation in wireless sensor networks: A comprehensive overview,” Computer Network, vol. 53, no. 12, pp.2022 -2037, 2009.

[8] B.Khaleghi, A. Khamis , F. O. Karray and S. N. Razavi, “Multisensor data fusion: A review of the state-of-the-art,” Information Fusion, vol. 14, no. 1, pp.28-44, 2013.

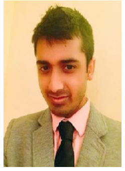
[9] S.Verma, P. Pillai and Y. F. Hu, “Performance Analysis of Data Aggregation and Security in WSN-Satellite Integrated Networks,” IEEE 24th Annual International Symposium on Personal, Indoor, and Mobile Radio Communication (PIMRC), 2013, pp. 3312-3316.

[10] S.Peter, D. Westhoff, and C. Castelluccia, “A Survey on the Encryption of Convergecast Traffic with In-Network Processing,” IEEE Trans. on Dependable and Secure Computing, vol. 7, no. 1, pp. 20-34, 2010.

[11] C.Fontaine and F. Galand, “A Survey of Homomorphic Encryption for Nonspecialists,” EURASIP Journal on Information Security, vol. 1, pp. 41-50, 2009.

[12] C.Castelluccia, E. Mykletun, G. Tsudik. “Efficient Aggregation of encrypted data in Wireless Sensor Networks,” Proceedings of the Second Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services, 2005, pp.109-117.
[13] V.Jariwala and D. Jinwala, “Evaluating homomorphic encryption algorithms for privacy in wireless sensor networks,” International Journal of Advancements in Computer Technology, vol. 3, no. 6, pp. 215-223, 2011.
[14] A.Viejo, Q. Wu and J. D. Ferrer, “Asymmetric homomorphisms for secure aggregation in heterogeneous scenarios,” Information Fusion, vol. 13, no. 4, pp. 285-295, 2012.
[15] J.D. Ferrer, “A Provably Secure Additive and Multiplicative Privacy Homomorphism,” Proc. Fifth Information Security Conf. (ISC’02), 2002, pp. 471-483.
[16] J.Girao, D. Westhoff, M. Schneider, “CDA: Concealed Data Aggregation for Reverse Multicast Traffic in Wireless Sensor Networks,” Proc. IEEE Int’l Conf. Comm. (ICC’05), 2005.
[17] S.Peter, P. Langendorfer, K. Piotrowski, “On Concealed Data Aggregation for Wireless Sensor Networks,” Proc. Fourth IEEE Consumer Comm. and Networking Conf. (CCNC), 2007.
[18] S.Verma, P. Pillai and Y. F. Hu, “Performance Evaluation of Alternative Network Architectures for Sensor-Satellite Integrated Networks,” The 8th International Workshop on the Performance Analysis and Enhancement of Wireless Networks (PAEWN'13), 2013, pp. 120-125.
[19] O.Younis and S. Fahmy, “Distributed Clustering in Ad-hoc Sensor Networks: A Hybrid, Energy-Efficient Approach,” IEEE Trans. on Mobile Computing, vol. 3, no. 4, pp. 366–379, 2004.
[20] Moteiv Corporation, “Tmote Sky Low Power Wireless Sensor Module”, Datasheet, 2006. Available Online: http://www.eecs.harvard.edu/~konrad/projects/shimmer/references/tmote-sky-datasheet.pdf

Authors



Suraj Verma is a PhD student at the University of Bradford (UK). His main research areas include security in wireless sensor integrated networks, homomorphic encryption, key management and next generation networks. He is also actively involved in robotics, eye tracking, embedded systems and web designing and has undertaken several projects at the University of Bradford in these areas. He was also involved in the EU funded SatNEX III project in 2011-2012.

His PhD topic mainly aims at providing an optimized lightweight homomorphic encryption scheme for end-to-end data confidentiality
in wireless sensor integrated networks.

Prashant Pillai (MS’02-PhD’07) is a lecturer in the School of Engineering, Design and Technology at the University of Bradford (UK). His main areas of work are in mobile/ wireless networks such as 2G/3G, WLAN/WiMax and bluetooth and satellite-based networks (DVB and BGAN).

Y. Fun Hu is a professor of Wireless Communications Engineering in the School of Engineering, Design and Technology of the University of Bradford (UK), where she leads the Future Ubiquitous Networks (FUN) Research Group. Her research interests encompass mobile/wireless/satellite networking, protocol design, security, QoS and mobility management.