International Journal of Technical Innovation in Modern Engineering & Science (IJTIMES)
Impact Factor: 5.22 (SJIF-2017), e-ISSN: 2455-2585
Volume 5, Issue 03, March-2019

SECURE DATA AGGREGATION USING WATERMARKING SCHEME IN WIRELESS SENSOR NETWORKS

[1]Mohit Gambhir, [2]Sapna Gambhir

[1]Verispire Technologies Pvt. Ltd.
[2]J C Bose University of Science & Technology, YMCA, Faridabad, India

Abstract: A wireless sensor network (WSN's) generates a huge amount of data with limited resource constraints which are susceptible to communication failure and various security attacks due to broadcast nature of communication. We use data aggregation techniques to gather data from different sources to eliminate redundant and duplicate data. There are issues in data aggregation such as delay, redundancy elimination, accuracy, traffic load and security. In this paper, we focus on security issues related to data aggregation techniques: data confidentiality, data integrity, node authentication and aggregator node availability. A sensor node which is providing data to the aggregator node must be an authentic node and aggregator node must be capable of detecting false data from unauthorized nodes in order to avoid flooding of unnecessary data which may lead to denial of service attack. In this paper, a secure data aggregation technique in which synopsis is generated with watermarking by sensing nodes and aggregator node to protect data integrity is proposed. In watermarking, each sensor node embeds a unique watermark to sensor data so that base station can verify the data integrity and combating data modification, data deletion and false data insertion attack.

Keywords: secure data aggregation, wireless sensor networks, sensor nodes, aggregator node, watermarking, and data integrity.

I INTRODUCTION

A sensor network is defined under some defined architecture such as clustered network. In such a network, the network is divided into small segments called clusters and each cluster is controlled by a cluster head. All the nodes in the cluster can perform direct communication with the cluster head and cluster head communicates with the Base station. The aggregator node collects data from the sensor nodes then send it to the base station. In this work, a secure data transmission scheme is proposed in WSN based on a watermarking technique. Watermarking technique has already applications in the security of multimedia content and relational databases. However, generation of watermark is based on the utilization of the characteristics of the original data and its application to the data integrity in WSN environment. A watermark is generated on the basis of sensor's own data characteristics, i.e. data length, digit occurrence frequency, and data sensing time of sensor nodes. Each sensor node embeds a unique watermark to sensor data and sends it to the BS along with the data. Then, BS verifies the integrity of data by using the embedded watermark.

Organisation of the paper is as follows: section II discusses some related work in the field of data aggregation; section III explains the proposed work; section IV concludes the paper.

II RELATED WORK

Shih-I Huang et. al. [2] proposed a secure encrypted-data aggregation scheme for wireless sensor networks. Their design for data aggregation eliminates redundant sensor readings without using encryption and maintains data secrecy and privacy during transmission. Conventional aggregation functions operate when readings are received in plaintext. If readings are encrypted, aggregation requires decryption creating extra overhead and key management issues. In contrast to conventional schemes, their proposed scheme provides security and privacy, and duplicate instances of original readings will be aggregated into a single packet.

Dirk Westhoff et. al. [3] show the major threat in WSNs is the presence of an adversary that injects forged data in the network or pretends to be an aggregator. Current mechanisms for authentication are based on complex computations, such as public key cryptography, which are not applicable in WSNs. In most scenarios, an authority issuing shared secrets is not available, as the sensors tend to communicate in a decentralized manner. With the Zero Common Knowledge (ZCK) we provide an authentication protocol that establishes well-defined pair-wise security associations between entities in the absence of a common security infrastructure or pre-shared secrets. We show that with two keyed hash-chains per communication pair, one can establish a certain level of trust within the system; ZCK ensures the re-recognition of a communication partner.

Claude Castellucia et. al. [4] presented an efficient approach is for utilizing the aggregation of data in a Wireless Sensor Network and the assuring of end-to-end encryption of data between the leaves and sink. One of the goals of the paper was to minimize the bit transmission between the sensor nodes and therefore to find an efficient encryption algorithm which is simple to implement and in turn would prolong the life of batteries.

M.Y. Mohamed Yacoab et. al. [5] presented a Compressive Data Aggregation technique which helps to solve the issues of traditional compression techniques. In this technique data is gathered at some intermediate node where size of the data need to be sent is reduced by applying compression technique without losing any knowledge of complete data.

V. Bhoopathy et. al. [6] have proposed Securing Node Capture Attacks for Hierarchical Data Aggregation in Wireless Sensor Networks. During first round of data aggregation, the aggregator upon identifying the detecting nodes selects a set of nodes randomly and broadcast a unique value which contains their authentication keys, to the selected set of nodes. When any node within the set wants to send the data, it sends slices of data to other nodes in that set, encrypted with their respective authentication keys. Each receiving node decrypts, sums up the slices and sends the encrypted data to the aggregator. The aggregator aggregates and encrypts the data with the shared secret key of the sink and forwards it to the sink. In the second round of aggregation, the set of nodes is reselected with new set of authentication keys. By simulation results, they have shown that the proposed approach rectifies the security threat of node capture attacks in hierarchical data aggregation.

Tamer Abu Ahmed et. al. [7] show data aggregation and in-network processing are important techniques for WSN. With the existence of the attacker threat, secure data aggregation scheme with immunity against malicious attacker is crucial for WSN correctness and security. In this paper, they proposed a secure and flexible data aggregation scheme with hierarchical key generation, aggregation, and node join protocols. The key are protected using secret sharing primitive such that the level key will be secure as long as the number of compromised nodes is less than t node.

Wembo He et. al. [8] provided a work on privacy preserving model for route optimization in sensor network. By providing efficient data aggregation and preserving data privacy is a challenging problem in wireless sensor networks. In this paper, they present privacy-preserving data aggregation schemes for additive aggregation functions.

There are data aggregation issues such as delay, redundancy elimination, accuracy, traffic load and security. Various security issues related to data aggregation techniques are presented which are data confidentiality, data integrity, and node authentication and aggregator node availability.

A sensor node which is providing data to the aggregator node must be an authentic node and aggregator node must be capable of detecting malicious nodes in order to avoid flooding of unnecessary data which may lead to Denial of Service attack. The path between the sensing nodes and aggregator node is secured to prevent passive attacks from unauthorized users to increase network throughput. Aggregator node performs registration of nodes using good authentication vector with cryptographic function for encryption.To authenticate and authorize sensor nodes for sending data to the aggregator node as this will help to overcome Denial of Service attack.

Fragile watermarking is used for the data integrity through embedding watermark into original data. Boubiche et al. [9], addressed a security scheme that suggests a new fragile watermarking technique based on a dynamic embedding mechanism and a cross-layer approach which optimizes the data aggregation process on the heterogeneous aggregation nodes. Kamel et al.[10] addresses unauthorized alterations in WSN data streams. The scheme, uses group delimiters to keep the sender and receivers synchronized and help them to avoid ambiguity in the event of data insertion or deletion. The watermark, which is computed using a hash function, is stored in the previous group in a linked-list fashion to ensure data freshness and mitigate replay attacks; FWC-D generates a serial number SN that is attached to each group to help the receiver determines how many group insertions or deletions occurred.

Sun, Xingming[11] presents a data integrity protection strategy based on digital watermarking technologies, where source sensors use a one-way hash function for collected data to create watermark information, and then make it associated with the data by embedding it into the redundant space of the targeted bytes. At the base station side, a watermarking algorithm is designed to extract the watermarking information, which is compared to recalculated watermarking information to verify the integrity of the data during the transmission.

Wei Zhang et al.[12] proposed an end-to-end, statistical approach for data authentication that provides inherent support for in-network processing. Authentication information is modulated as watermark and superposed on the sensory data at the sensor nodes. The watermarked data can be aggregated by the intermediate nodes and data sink is able to authenticate the data by validating the watermark, detecting data if altered. The aggregation–survivable authentication information is only added at the sources and checked by the data sink, without any involvement of intermediate nodes. The simple operation of watermark embedding and complex operation of watermark detection provide a natural solution of function partitioning between the resource limited sensor nodes and the resource abundant data sink.

Boubiche et al. [13] aims to secure the data aggregation process while saving energy. The mechanism uses a lightweight fragile watermarking technique without encryption to insure the authentication and the integrity of the sensed data while saving the energy. The links between the sensor nodes and the aggregation nodes, and also the links between the aggregation nodes and the base station are secured by using the watermarking mechanism.

Khizar Hameed et al. [14] proposed a method to verify the integrity of the sensor data, which is based on a zero watermarking scheme. In watermarking, each Sensor node embeds a unique watermark to sensor data and BS can verify the data integrity. The scheme is robust against different types of data integrity attacks such as data insertion, data modification, and data deletion.

G.Prathima E et al.[15] proposes Secure Approximate Data Aggregation (SADA) in which synopsis are generated using primitive polynomial and Message Authentication Codes (MACs) are transmitted along with the synopsis to ensure integrity. SADA provides data freshness and integrity at a lower communication cost. Wang et al. [1] proposed a digital watermarking technique to protect the copyright of data in WSNs. The watermark embedded in original data by using both the LSB and MSRB bits of data field. Both, the original data and watermark are sent to BS for verification of copyright of data.

III PROPOSED WORK

The proposed method is used to secure data in data aggregation process in an efficient manner. Digital watermarking is a mark embedded in data which is used for tracing ownership of the signal. It helps to verify integrity and authenticity of sensed data. These types of watermarking concepts are used by sensing nodes as well as aggregator node to protect data integrity. Here, each sensor node embeds a unique watermark to sensed data so that base station can verify for data integrity.

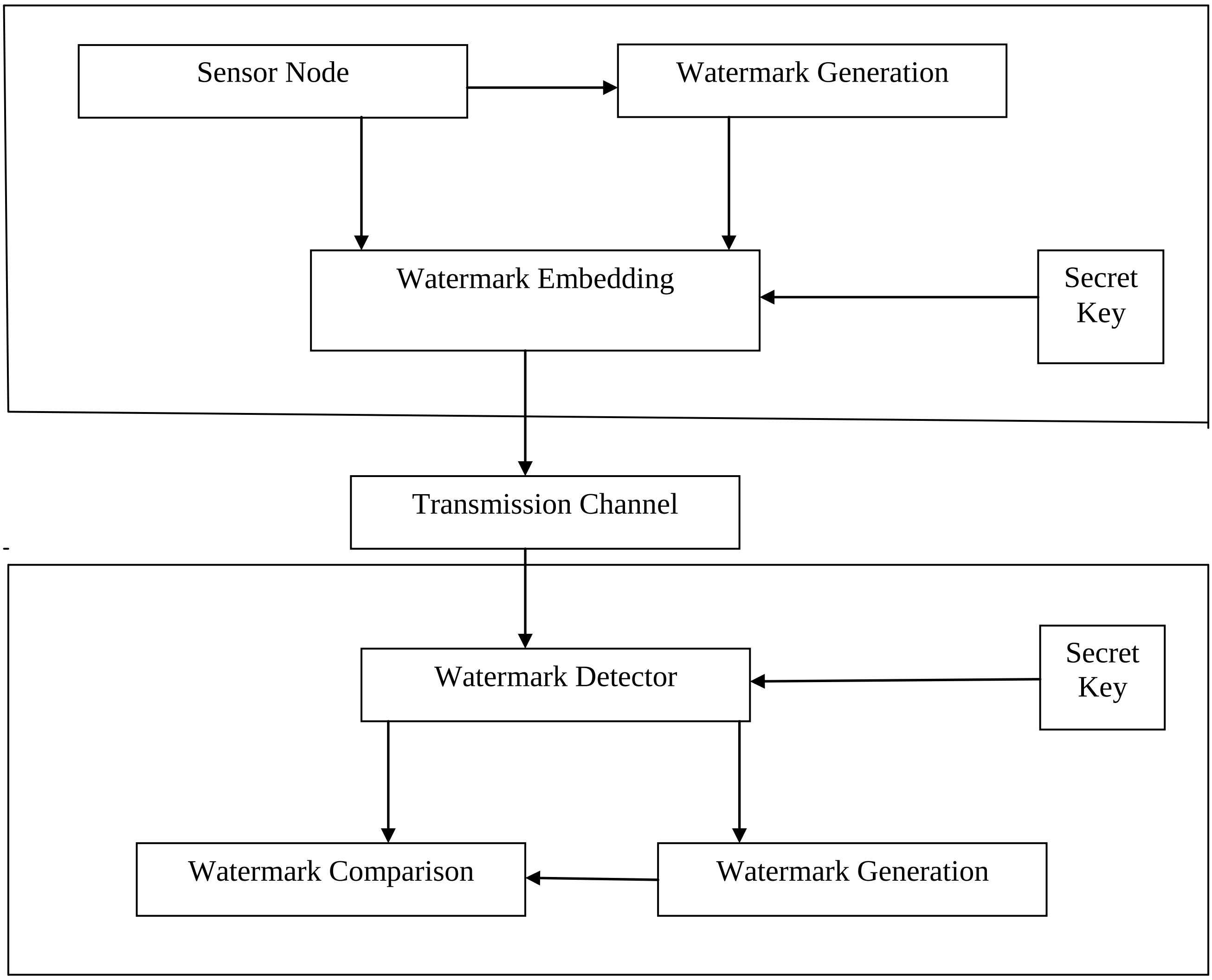

Figure 4.1- Watermark Encoding and Decoding Model

Figure 1 shows working of watermark encoding and decoding model in WSN consists of following steps: In network model, there are different sensor nodes *SN*1, *SN*2,... *SNn* capture sensor data *d*1 , *d*2,... *dn* from the surrounding environment.

- Each sensor node passes their data to watermark generation process to generate watermark.

- Watermark is encrypted and is embedded with Sensor data by use of watermark embedding to generate watermarked data and hence sent to CH.

- After receiving all watermarks, CH performs aggregation process and send final watermark *dw* to BS.

- At Base Station, WD detects watermark in data and separate it. A comparison is performed between two watermarks that are generated by BS watermark generator *W* and detected watermark *W*. On basis of comparison, BS checks integrity of data that it is tampered or not.

TABLE 1 ALGORITHM NOTATIONS

Symbol	Description
d	Sensory Data
s^e	Secret Key for Encryption
s^d	Secret Key for Decryption
w	Final Watermark
d^w	Watermarked Data
$\|$	Concatenation
w^e	Encrypted Watermark
w_l	Data Length
w_o	Data Occurrence Frequency
w_t	Data Capturing Time
w_b	Watermark Generated by Base Station

The secure data aggregation algorithm using watermarking scheme comprises of following steps:

- Watermark Generation: It is done on the basis of sensed data characteristics i.e length, time of sensed data, and time of synopsis generation.
- Watermark Embedding: It takes sensed data and the final watermark generated and uses secret key to generate final watermarked data by sensor node. It uses secret key to encrypt watermark and produce a final watermarked data.
- Data Aggregation: The watermarked data is received from the sensor nodes participating in the cluster. CH will aggregate received data and pack it in one packet with watermark of aggregator node and retaining the original watermarks by sensor nodes and send it to the base station**.**
- Watermark Extraction and Verification: It is done by the base station it uses decryption key to decrypt the watermark. It accepts watermarked data and with its decryption key it will separate the watermark from the data. BS produces a new watermark by use of received data to verify data integrity. If decrypted watermark is same as generated watermark then data is not tampered.

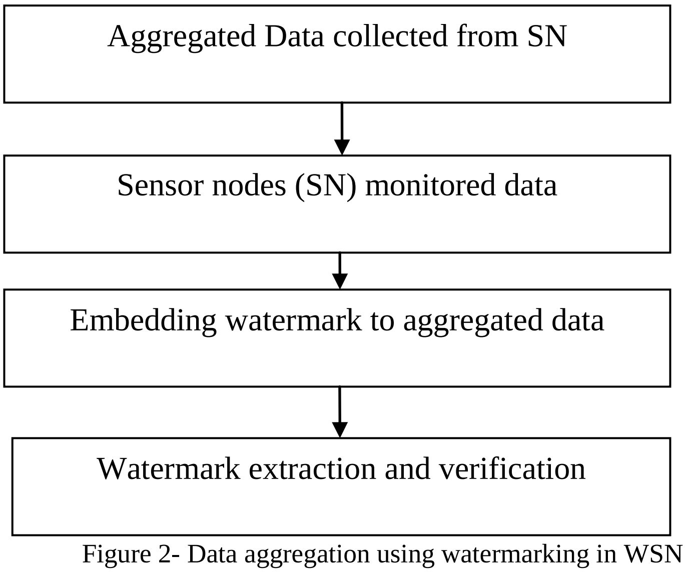
Aggregated Data collected from SN
Sensor nodes (SN) monitored data
Embedding watermark to aggregated data
Watermark extraction and verification

Figure 2- Data aggregation using watermarking in WSN

Figure 2 gives an overview of functions of Sensor node, Base Station, and Aggregator node. At sensor node level, aggregator node collects the data from the sensor nodes. At aggregator node level, the aggregator nodes directly communicate with the base station. Aggregator node is elected based on the energy level of cluster members. After the aggregator node is elected, it collects the data from cluster members. Each aggregator performs the watermarking technique for the verification of data. The aggregator performs three main steps to generate the watermark. The first step is to perform the data aggregation process. The second step is to use aggregation function. The last step is to integrate the outcome to the data packet along with the aggregation result.

Next step at sensor node level is detection and verification mechanism. After the data packet reaches the base station, it extracts the watermark. Then it again calculates the value of watermark using the embedding mechanism. After generating the watermark, compares the generated watermark value with the previously generated value on the sender side. If both values are matched, that means data is authentic and not altered and if they do not, that means, the data is modified. So, the base station will reject the received data and informs the aggregator node that a sender node is compromised.

```
Algorithm 1 Watermark Generation
1: procedure WATERMARK GENERATION
2: Input: d
3: Output: w
4: $w_l \leftarrow$ sensory data length (d)
5: $w_o \leftarrow$ sensory data occurrence frequency (d)
6: $w_t \leftarrow$ sensory data capturing time (d)
7: $w \leftarrow w_l \| w_o \| w_t$
8: end procedure
```

Algorithm1 watermark generation process takes sensory data d as input from sensor node and generates a watermark w on basis of sensor data characteristic such as length, occurrence frequency and time of sensor data when generated as shown in line 4–6 of Algorithm 1. It combines data length w_l, digit occurrence frequency w_o and data sensing time w_t to produce a final watermark w. The data sensing time is the time to capture sensory data at particular moment and assume that the time is not changeable at BS. In Algorithm 2 the watermark embedding mechanism is shown, cluster head and the addresses of all the nodes are used to generate the watermark and combines the generated watermark with the aggregation result. The function is applied to the data after applying the key with the original data, thus obtained the result to generate the watermark that will be integrated with the aggregation result in the data packet and send to the base station.

```
Algorithm 2 Watermark Embedding
1: procedure WATERMARK EMBEDDING

2: Input: d, w, $k^e$

3 Output: $d^w$

4: $w^e \leftarrow$ Encrypt (w , $k^e$ )

5: $d_w \leftarrow d \| w^e$ // (Watermark Embedded)

6: Send watermarked data to Cluster Head ($d^w$)

7: end procedure
```

In watermark embedding process, it takes sensory data *d*, final watermark *w* , and decoding key as inputs and produce a watermarked data w as output. In Algorithm 2, a final watermark *w* is encrypted with secret key k^e (line 4). Afterwards, it embeds sensory data d with encrypted watermark w^e to produce a final watermarked data d_w and send it to the BS through transmission channel (lines 5-6) as show in the Algorithm 2. BS distributes the secret keys to sensor nodes through SHA. It consists of the general watermark generation process. When these compromised nodes are found by aggregator node, it will send the alert message to the base station so that when nodes match the values of the watermark, and it doesn't match then they will again go for the embedding mechanism. Now while repeating the whole mechanism it will use characteristics of the data of the sensor nodes. The new watermark value is generated and sends the data packet along with the watermark value to the neighbor nodes.

Next step at sensor node level is detection and verification mechanism. After the data packet reaches the base station, it extracts the watermark. Then it again calculates the value of watermark using the embedding mechanism. After generating the watermark, compares the generated watermark value with the previously generated value on the sender side. If both values are matched, that means data is authentic and not altered and if they do not, that means, the data is modified. So, the base station will reject the received data and informs the aggregator node that a sender node is compromised.

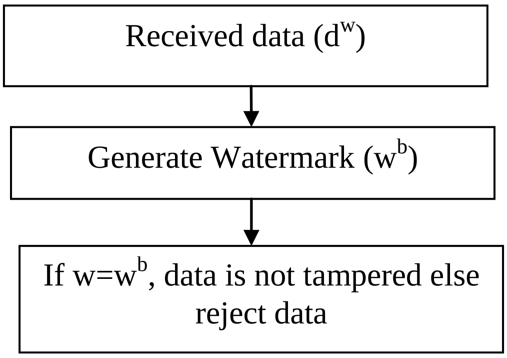
Received data (d^w)
Generate Watermark (w^b)
If $w=w^b$, data is not tampered else reject data

Figure 3- Watermark Extraction and Verification

Algorithm 3 describes the working of watermark extraction and verification algorithm at BS. At BS, watermark extraction and verification algorithm accepts watermarked data d_w and k^d an input and produce a re-generated watermark w_b as output to verify data integrity. BS receives watermarked data dw and extracts into d and encrypted watermark w_e as shown in lines 4-5 of Algorithm 3. At line 6, decryption operation is performed with k^d key on encrypted watermark w_e to get watermark w. In verification phase, watermark generation algorithm is performed again on original sensor data d to re-generate watermark w . A comparison operation is performed to check the integrity of the data by comparing watermarks w_b and w as shown in lines 8 to 11 of Algorithm 3.

Algorithm 3 Watermark Extraction and Verification

```
1:    procedure WATERMARK EXTRACTION AND VERIFICATION
2:    Input : $d^w$ , $k^d$
3:    Output: Verified/ Not Verified
4:     Receive Watermarked Data ($d^w$ )
5: Extract Watermarked Data into d and $w^e$
6: w  ← Decrypt ($w^e$, $k^d$)        Extracted Watermark
7:     $w_b$     ← Call Algorithm 1        Re-Generated Watermark
8:     If (w == $w_b$ )
9:     Print “Verified”
10:    Else
11:   Print “Not Verified”
12:    end procedure
```

IV CONCLUSION

In this paper, a scheme to verify the integrity and authenticity of sensor data, based on watermarking technique is proposed. A watermark is generated on the basis of characteristic of sensor data such as length, occurrence frequency, and data sensing time of sensor node. **T**he proposed system will give various benefits to the existing systems like:

- Authentication of sensor node by the Base station using three way handshaking in which SHA based node verification is done before the node participate in the communication. Hence, malicious nodes could not participate in communication as they need authorization from base station.
- Data Integrity is achieved using digital watermarking which helps us to identify data modification attack. So, if data is tampered then BS will reject it.

REFERENCES

[1] Mukesh Kumar, Kamlesh Dutta, "A Survey of Security Concerns in Various Data Aggregation Techniques in Wireless Sensor Networks", in Springer India, 2015.

[2] Shih-I Huang: "Secure encrypted-data aggregation for wireless sensor networks", in Computational Intelligence and Security ,IEEE, PP 848-852, Dec 2007.

[3] Dirk Westhoff, "Security Solutions for Wireless Sensor Networks", in NEC Technical Journal, Volume 1, March 2006.

[4] Claude Castellucia: "Efficient Aggregation of encrypted data in Wireless Sensor Networks", in WS 2009.

[5] M.Y. Mohamed Yacoab: "A Cost Effective Compressive Data Aggregation Technique for Wireless Sensor Networks", in International Journal of Ad Hoc, Sensor & Ubiquitous Computing (IJASUC).

[6] V. Bhoopathy: "Energy Efficient Secure Data Aggregation Protocol for Wireless Sensor Networks", in European Journal of Scientific Research ISSN 1450-216X.

[7] Tamer Abu Ahmed, "A Dynamic Level-based Secure Data Aggregation in Wireless Sensor Network", in Information Security Research Laboratory Graduate School of IT & Telecommunication INHA University.

[8] Wenbo He, "PDA: Privacy-preserving Data Aggregation in Wireless Sensor Networks", IEEE INFOCOM 2007.

[9] Boubiche, Djallel Eddine, Sabrina Boubiche, Azeddine Bilami, "A Cross-Layer Watermarking-Based Mechanism for Data Aggregation Integrity in Heterogeneous WSNs.", in *Communications Letters, IEEE, PP* 823-826, 19.5(2015).

[10] Kamel, Ibrahim, Hussam Juma, "A lightweight data integrity scheme for sensor networks." In *Sensor, PP* 4118-4136, 11.4 (2011).

[11] Sun, Xingming, "Digital watermarking method for data integrity protection in wireless sensor networks.", International Journal of Security and Its Applications 7.4, PP 407-416, 11.4 (2011).

[12] Wei Zhang, "Secure data aggregation in wireless sensor networks: a watermark based authentication supportive approach." In Pervasive and Mobile Computing 4.5, PP 658-680, 2008.

[13] Boubiche, Djallel Eddine, "SDAW: secure data aggregation watermarking-based scheme in homogeneous WSNs." In Telecommunication Systems , 1-12, 2015.

[14] Khizar Hameed, M Saleem Khan, "A zero watermarking scheme for data integrity in Wireless Sensor Networks" 19th International Conference on Network-Based Information Systems, IEEE, PP 2157-0426, 2016.

[15] G.Prathima E, Shiva Prakash T, Venugopal K R, S S Iyengar, L M Patnaik, "SADA : Secure Approximate Data Aggregation in Wireless Sensor Networks" in 2016 IEEE International Conference on Data Science and Engineering (ICDSE), 978-1-5090-1281-7/16, 2016.