Digital Watermarking Method for Data Integrity Protection in Wireless Sensor Networks

Xingming Sun [1,2] , Jianwei Su [1,2] , Baowei Wang [1,2] and Qi Liu [1]

[1] *Jiangsu Engineering Center of Network Monitoring, Nanjing University of Information Science & Technology, Nanjing, 210044, China*
[2] *School of Computer & Software, Nanjing University of Information Science & Technology, Nanjing, 210044,China*

sunnudt@163.com, sunuist@163.com, wbw.first@163.com, qrankl@163.com

Abstract

Wireless sensor networks are self-organized and data-centric, which have been well applied in many practical areas. But in these applications, data integrity from sensors has not been verified. Thereby in this paper, we have proposed a novel data integrity protection strategy based on digital watermarking technologies, mwhere 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. Compare to other watermarking methods, our algorithm does not increase extra data storage space and remain data accuracy. According to the results of extensive experiments, our algorithm can effectively protect the integrity of the data and has more application values.

Keywords: *Wireless sensor networks, digital watermarking, data integrity protection*

1. Introduction

Wireless Sensor Networks (WSNs) are composed of static and/or mobile sensor nodes in a monitored area. Sensor nodes cooperatively sense information in the covered region, so that perceived data can be collected and sent to the base station. At present, WSNs have been widely used in the fields of environmental monitoring, medical care, military, *etc.* The sensor nodes are usually deployed in unattended areas and limited by energy, computing power, storage space and communication due to their individual capacity. Under such circumstances, traditional security techniques can not meet the needs of communication between these nodes. Data being transmitted are vulnerable to external or internal attacks, such as forgery, tampering, replay and selective forwarding. It has become critical to protect data integrity in wireless sensor networks.

Digital watermarking is currently one of the popular technologies in wireless sensor networks security [1, 2, 12, 13]. It can effectively prevent sensored data being intercepted, and precisely detect whether the data have been tampered. In the last few years, comprehensive studies on digital watermarking techniques for normal data types, such as images, videos and texts, including research activities in the wireless sensor networks field.

In [1], J. Feng *et al.*, have developed the first system of watermarking techniques to embed cryptologically encoded authorship signatures into data and information acquired by wireless

embedded sensor networks. The key idea is to impose additional constraints during the data acquisition or sensor data processing. Constraints corresponding to an encrypted signature are selected with consideration of tradeoffs between the accuracy and the strength of proof of the authorship.

In [2], Guo *et al.*, have proposed a novel fragile watermarking algorithm to verify the integrity of streaming data at the application layer. The data are divided into groups based on synchronization points, so each group can be synchronized, whereas modifications made to one group only affect up to two groups. In each group, a unique watermark is embedded directly into the Least Significant Bit of data to save communications bandwidth. The embedded watermark can detect as well as locate any modifications made to a data stream. To ensure the completeness of data streams, watermarks are chained across groups so that data deletion can be correctly detected.

In [3], Wang *et al.*, have designed a fragile watermark algorithm for preserving stringent data integrity, which is a chain scheme using dynamic group size. Firstly, they convert the sensor data into a character, so that the blank character based embedded scheme can be used. Second, in the data group partition scheme, the dynamic group size is adopted, so the scheme is able to detect any malicious modification. Thirdly, a hash value generated in each group as their own fragile watermark, which is uniquely and closely related to the data of the group, is embedded in the data in each group. Modifications made to the group can thus not affect data integrity authentication in other groups.

Research work above attempts to apply digital watermarking techniques into a wireless sensor network to solve data forgery and integrity problems. However, present studies suffer from increasing extra data storage space or affecting data accuracy. In addition, direct exposure of data has greatly weakened its security. How to keep the balance of data storage efficiency, data accuracy, as well as data concealment becomes a critical challenges.

In this paper, we propose a novel watermarking method to protect data integrity in WSNs. Collected data from each source sensors are capsulated into news packages, which contain diverse data fields for particular sensing sources. Different from existing studies which embed watermarking by attaching to or modifying perceived data, no intrusions to the original data are performed. Instead, redundant space of data bytes is employed. The data fields design in a package and bounds checking routines facilitate the selection of redundant bits with no affecting data accuracy. According to the experimental results, our method can effectively verify the data integrity and make sure whether they are forged, manipulated, discarded and/or under repaly attacks during the transmission.

The rest of the paper is organized as follows. Section 2 describes the digital watermarking method, which includes watermarking embedding and extraction process. Section 3 describes the performance evaluation. At last, we conclude the paper in Section 4.

2. Digital Watermarking Method

Digital watermarking can be used to verify data authenticity and reliability in WSNs. The method we designed is based on nondestructive digital watermarking algorithms, where watermarking information is directly associated with sensor data. Redundant space of the data is used for the digital watermarking information. A digital watermark embedding and extraction model is shown in Figure 1. In this model, the watermarking information is generated and embedded at source sensor nodes side, whereas watermarking extraction and its verification are performed at base station side.

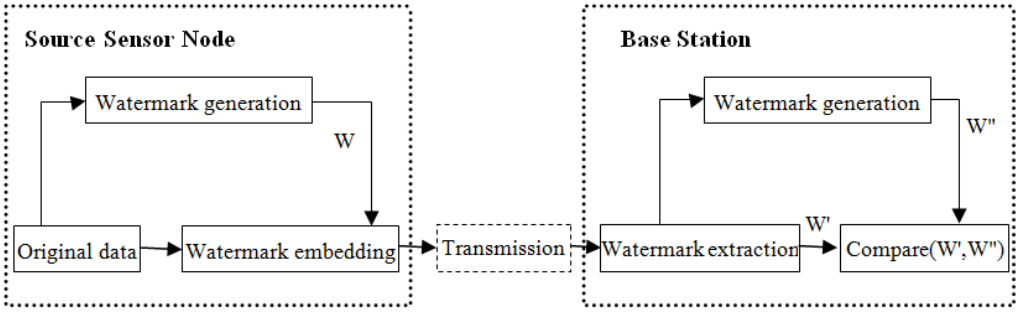


Figure 1. Digital Watermark Embedding and Extraction Model

Symbols and rules used in the algorithm are defined as follows:

Define 1: In a WSN network, sensor data are transmitted in a package including data fields, where sensor data are stored in a pre-decided order. These data fields are denoted as D= $\{d_0, d_1, \ldots, d_n\}$, in which d_i indicates one sampling which is expressed in decimal, and n represents the number of the data field.

Define 2: The size of the data field is in byte for the unit. The range of data acquisition is determined by the data resolution. Taking a Telosb node as an example, which uses humidity and temperature sensor of SHT11. The resolution of humidity is 0.03%RH, so the collected humidity requires 12 bits of storage space. In this case, two bytes are required in the package to store the data, which makes 4 bits as the redundant space. As shown in Figure 2, redundant space is denoted as $R(i) = \{r_0, r_1, r_n\}$, $(0 \leq i \leq n)$, where n represents the number of the data field, r_i represents the size of the redundant space of i^{th} data field, and the total size of the redundant space in a package is N.

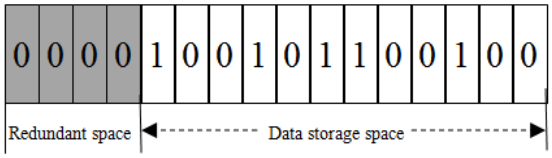


Figure 2. Redundant Space of the Data Field

Rule 1: Generation of digital watermark uses the one-way hash function [11] to calculate each sensed data, denoted as $H_i = \text{Hash}(t_i, d_i)$,$(0 \leq i \leq n)$, where t_i is the time of the i^{th} collected data. Then calculation of the watermarking information W, can be denoted as $W = \text{GroupXOR}(H_0 \oplus H_1 \oplus \ldots \oplus H_n)$, in which "$\oplus$" is the XOR operation. The length of W is N, which is equal to the total size of the redundant space in a package.

Rule 2: Digital watermarking embedding rules includes the following steps:

(1). Generate the watermarking information W according to the rule 1;

(2). Embed the W into the redundant space of the data fields R(i), and denote the watermarked data D'.

Rule 3: At the recipient side(*i.e.*, the base station), when receiving the D', extract the watermarking information from the redundant space in turn. The watermarking information expressed as W'. Then D' will be back to D, so that we recalculate the recalculate the

watermark information W" according to D with the same rules. If W′ is equal to W", the data integrity has not been damaged or destroyed during the transmission.

2.1. Watermarking Embedding Algorithm

In WSNs, data collected from a source sensor node is formatted in binary. The watermarking embedding algorithm uses the redundant space by **Define 2** to embed watermark information. The embedding process is shown in Figure 3.

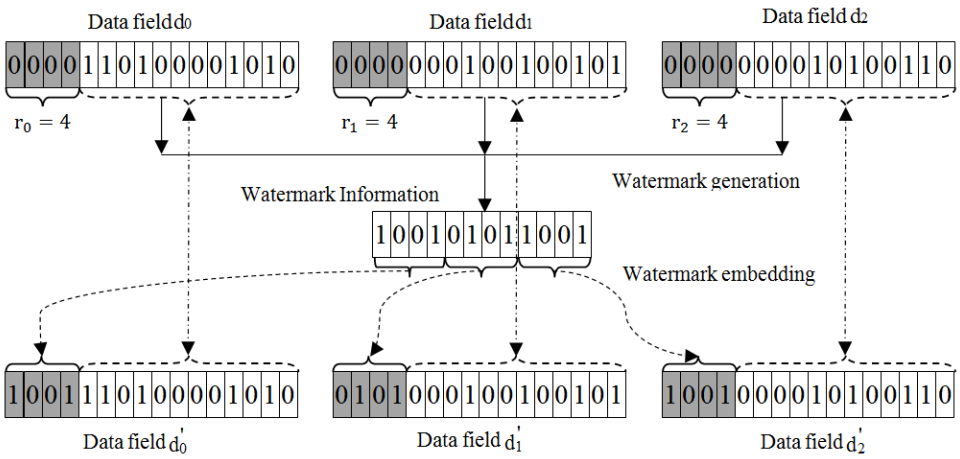


Figure 2. Digital Watermarking Embedding Process

In order to preserve energy and calculating resources, intermediate nodes do not need to extract packages but forward them only. In the watermarking embedding algorithm, the size of redundant space in each data field has been pre-determined according to the type and sensing capacity of the sensor nodes. The watermark embedding algorithm is shown in **Algorithm 1**.

Algorithm 1. Digital Watermark Embedding

Input: Data field D and its number n, the size of the redundant space R(i) and its total size N.

Output: The watermarked data D′.

Steps:

1. For(int i=0; i<n; i++)
2. $H_i = \mathrm{Hash}(t_i, d_i)$;
3. $W = \mathrm{GroupXOR}(H_0 \oplus H_1 \oplus \ldots \oplus H_n)$;
4. i = 0;
5. For（int j= P ; j > (P − r_0) ; j--）
6. d_0 [j]=W[i]; i++;
7. For (j= P ; j > (P − r_1) ; j--)

8. d_1 [j]=W[i]; i++;
9. …
10. For (j= P ; j > (P − r_n) ; j--)
11. d_n [j]=W[i]; i++;
12. Send(D′);

In order to facilitate the description of the algorithm, P represents the highest position of each data field. In **Algorithm 1,** the collected data D and its number n, the size of the redundant space of each data field R(i) and the total length of the redundant space N are inputs. The watermarking is specially encoded as binary strings, which is embedded into the redundant space of each data field. The watermarked data D′ is finally generated.

2.2. Watermarking Extraction Algorithm

When the base station receives watermarked data, it firstly extracts the watermarking information W′ from the data. Then watermarking information W″ is recalculated according to the same rules. Comparison between W′ and W″ decides whether the data integrity has been damaged. The extraction process is shown in Figure 4.

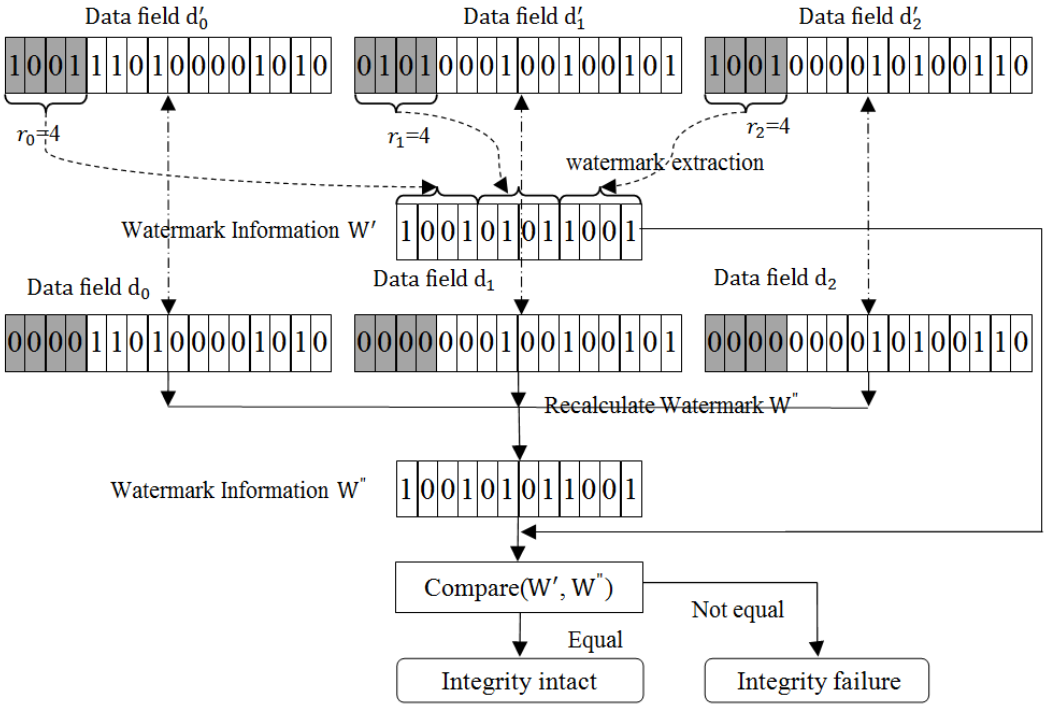


Figure 4. Digital Watermark Extraction Process

The algorithm for watermarking extraction and verification is shown in **Algorithm 2**.

Algorithm 2. Digital Watermark Extraction

Input: The watermarked data D', The size of the redundant space R(i) and its total size N.

Output: The result of data integrity

Steps:

1. int i=0;
2. For(int j= P; j > (P − r_0); j--）
3. { $W'[i] = d'_0[j]$; $d'_0[j] = 0$; i++; }
4. For(j= P; j > (P − r_1); j--)
5. { $W'[i] = d'_1[j]$; $d'_1[j] = 0$; i++; }
6. …
7. For(j= P; j > (P − r_n); j--)
8. { $W'[i] = d'_n[j]$; $d'_n[j] = 0$; i++; }
9. For(i = 0; i < n; i++)
10. $H_i = \mathrm{Hash}(t_i, d_i)$;
11. W''= GroupXOR($H_0 \oplus H_1 \oplus ... \oplus H_n$);
12. If (Compare (W',W'') = = Equal)
13. Return (Integrity intact);
14. Else
15. Return (Integrity failure);

In **Algorithm 2**, extraction of the watermarking information D' turns it into the original data. Then recalculation of the watermarking information and consequent comparison make sure that data integrity issues can be detected.

3. Performance Evaluation

3.1. Experiments Setup

In this section, all experiments were performed in a real wireless sensor network environment. 50 Telosb nodes were deployed in the networks. The nodes use the TinyOS operating system, IEEE 802.15.4 wireless communication stack and the collection tree routing protocol (CTP). Temperature, humidity and illumination data were gathered and transmitted every three minutes. The size of the packet payload is 80 bits for node ID, node voltage, temperature, humidity and light information. The size of redundant space is 12 bits

and only exists in the temperature, humidity and illumination fields. The nesC language was used to implement the watermarking embedding algorithm; whereas C# was for watermarking extract and match in the base station. No further extraction was performed during data transmission.

3.2. Experiment Results

Embedding capacity analysis: Figure 5 shows that the comparison of embedding capacity among the Least Significant Bit [5-10], add blank character [3] and our proposed algorithm. Accoring to the figure, the Least Significant Bit has the same embedding capacity as the add blank character method, whereas the capacity of our proposed method is significantly higher than the previous digital watermarking algorithms.

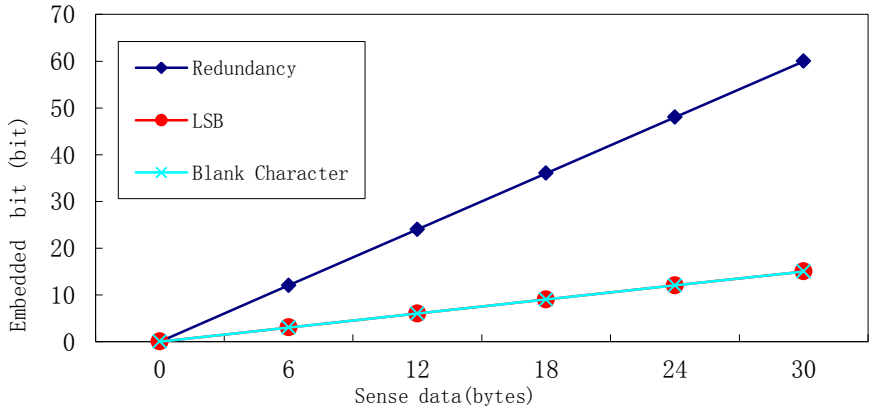


Figure 3. Watermarking Embedded Capacity

Transmission data quantity analysis: Four methods, including three in Figure 5 and the Message Authentication Code (MAC) method are discussed in Figure 6 in terms of data transmission amount . The MAC method adds the message directly into the data for transmission, so the amount of data transmission is relatively large. The add blank character method at the end of the data is according to the corresponding data bit of watermark information is 1 or 0, selective adding blank character, so the data quantity exists wave phenomenon. Although the way of modifying the least significant bit has no effect on the amount of data transmitted, the accuracy of the data can be affected. Our proposed digital watermarking method does not affect fetched data, and is entirely reversible.

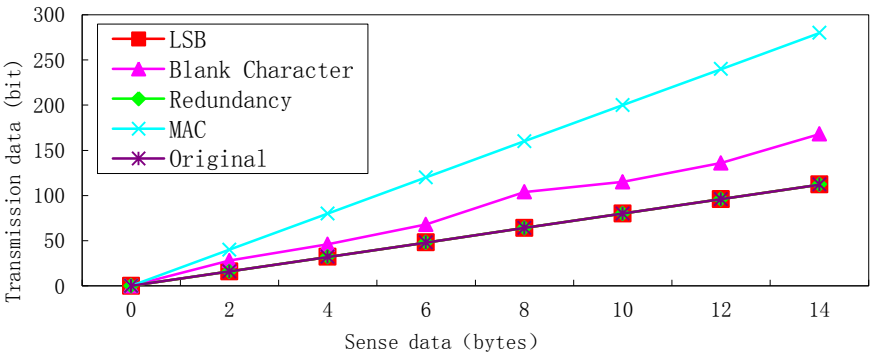


Figure 4. Transmission Data Quantity

Safety analysis: Different types of attacks were tried to verify its safety. Five nodes were randomly selected as attacking nodes to perform following five attacks separately: packet forgery attack, selective forwarding, packet replay, packet transfer delay and packet tampering. Each attack has been tested 100 times, and the experimental results are presented in **Table 1**.

Table 1. The Experimental Results of Data Integrity Attacks

Attacks	No. of experiments	Success rate（%）
Packet Forgery	100	100
Selective Forwarding	100	100
Packet Tampering	100	100
Packet Replay	100	100
Transfer Delay	100	94

Our scheme divides the data field into two sections. The lower bits store collected data, whilst the high bits store watermarking information. Watermarking information and the collected data are directly associated. According to the experimental results our method achieved 100% detection on packet forgery attack, selective forwarding, packet replay and packet tampering. It failed six times resisting the packet transfer delay attacks due to the short interval set in the source sensor nodes. If the data transmission delay time is greater than the data collection period, the data freshness was destroyed, so we adopt the principle of discarded directly, not to detect error. If the data transmission delay is less than the data collection cycle, the success rate will also reach 100%. After the watermarking embedding, the data will be immediately sent, the data transmission delay has no influence on the algorithm. Therefore, the experimental results show that our proposed watermarking scheme can effectively verify the integrity of the data, and ensure the authenticity and reliability of the data.

Energy evaluation: When we estimate energy consumption, it mainly includes data storage, watermark embedding, routing，broadcast and data transmission. We assume that the data storage, routing and broadcast of energy consumption have fixed. We only consider the consumption of energy in watermarking embedding and data transmission. Usually, on the order of 3000 instructions can be executed for the energy cost required to transmit one bit over a distance of 100m by radio [4]. So the energy consumption of the data transmission between the nodes is far greater than processing calculation. In our scheme, watermarking information is directly embedded into the redundant space of the data. It does not occupy additional storage space. Compared to the adding blank character method, the energy consumption of data transmission is significantly reduced. In addition, the watermarking is embedded into the binary data instead of the actual value. Compared with the method of adding blank character and modify the Least Significant Bit, the average energy consumption of each node is reduced. So in terms of energy consumption the cost of our proposed watermarking method has incomparable advantages compared to other ones.

4. Conclusion

In this paper, we have proposed a digital watermark scheme for data integrity protection. The proposed approach can prevent data from security attacks such as packet forgery attack, selective forwarding, packet replay, packet transfer delay and packet tampering during the transmission, and can verify the authenticity and reliability of data. Practical experiments have been conducted in a real employed wireless sensor network environment. The results have shown that with regard to energy consumption and data accuracy, our watermarking algorithm is out-performed compared to other algorithms.

In the future, the algorithm can be optimized in two points. First, the digital watermarking model can be refined to detect the whole data lost. Second, in a few sensors, when the resolution of the collected data is equal to the size of the data field, the algorithm does not adapt. So for this type of sensor, the digital watermarking model also needs further improvement.

Acknowledgements

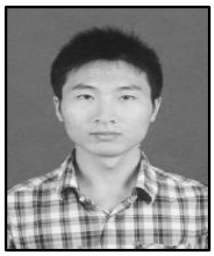
This work is supported by the NSFC (61232016, 61173141, 61173142, 61173136, 61103215, 61070196, 61070195, and 61073191), National Basic Research Program 973 (2011CB311808), 2011GK2009, GYHY201206033, 201301030, 2013DFG12860, PAPD fund and NUIST Research Fund.

References

[1] J. Feng and M. Potkonjak, "Real-time watermarking techniques for sensor networks", SPIE Security and Watermarking of Multimedia Contents, Santa Clara, CA, USA: SPIE Press, **(2003)**, pp. 391-402.
[2] H. Guo, Y. Li and S. Jajodia, "Chaining watermarks for detecting malicious modifications to streaming data", Information Sciences, no. 177, **(2007)**, pp. 281-298.
[3] B. Wang, X. Sun and H. Ren, "Multi-mark: Multiple Watermarking Method for Privacy Data Protection in Wireless Sensor Networks", Information Technology Journal, **(2011)** October, pp. 833-840.
[4] J. Potter and W. J. Kaiser, "Wireless integrated network sensors", Communications of the ACM, **(2000)** May, pp. 51-58.
[5] S. Raja, A. Mikhail and P. Sunil, "Resilient rights protection for sensor streams", Proceedings of the Thirtieth international conference on Very large data bases, vol. 30, **(2004)**, pp.732-743.
[6] W. Zhang, Y. Liu and K. Sajal, "Aggregation Supportive Authentication in Wireless Sensor Networks: A Watermark Based Approach", World of Wireless, Mobile and Multimedia Networks, WoWMoM 2007. IEEE International Symposium, **(2007)**, pp. 1-11.
[7] H. K. Juma and L. I. Kaya, "Watermarking sensor data for protecting the integrity". Innovations in Information Technology, 2008. IIT 2008. International Conference on, **(2008)** December, pp. 598-602.
[8] X., D. X. and Li, "An Authentication Method for Self Nodes Based on Watermarking in Wireless Sensor Networks", Wireless Communications, Networking and Mobile Computing, WiCom '09. 5th International Conference, **(2009)**, pp. 1-4.
[9] R. Xuejun, "A sensitivity data communication protocol for WSN based on digital watermarking", School of Information and Technology, Northwestern University, Xi'an 710127, China, **(2010)**.
[10] I. Kamel and H. Juma, "A Lightweight Data Integrity Scheme for Sensor Networks", Sensors, **(2011)** November, pp. 4118-4136.
[11] W. Jizhi, X. Shujiang, T. Min and W. Yinglong, "The analysis for a chaos-based one-way hash algorithm", International Conference on Electrial and Control Engineering, **(2010)**, pp. 4790-4793.
[12] W. Zhang and S. K. Das, "Secure data aggregation in wireless sensor networks: A watermark based authentication supportive approach", Pervasive and Mobile Computing, **(2008)**, vol. 5, pp. 658-680.
[13] X. Xiao and X Sun, "Secure data transmission of wireless sensor network based on information hiding", Mobile and Ubiquitous Systems: Networking & Services, **(2007)**, pp.1-6.

Authors

Xingming Sun is a professor in the School of Computer and Software, Nanjing University of Information Science and Technology, China from 2011. He received the B.S.degree in Mathematical Science from Hunan Normal University and M.S. degree in Mathematical Science from Dalian University of Technology in 1984 and 1988, respectively. Then, he received the Ph.D degree in Computer Engineering from Fudan University in 2001. His research interests include information security, network security, cryptography and ubiquitous computing security.



Jianwei Su received his B.S. degree in Computer Science from Nanjing University of Information Science and Technology, China in 2011. Currently he is studying for his M.S degree in Meteorological Information Security at the same university. His research interests include steganography, cryptography and network security.

Baowei Wang received his B.S. and Ph.D degrees in Computer Science from Hunan University in 2005 and 2011, respectively. He is currently working as a lecturer in School of Computer and Software, Nanjing University of Information Science and Technology. His research interests include steganography, wireless networks and securing ad hoc networks.

Qi Liu received his BSc degree in Computer Science and Technology from Zhuzhou Institute of Technology, China in 2003, and his MSc and PhD in Data Telecommunications and Networks from the University of Salford, UK in 2006 and 2010. His research interests include context awareness, data communication in MANET and WSN, and smart grid. His recent research work focuses on intelligent agriculture and meteorological observation systems based on WSN.