Multi-mark: Multiple Watermarking Method for Privacy Data Protection in Wireless Sensor Networks

1Baowei Wang, 2Xingming Sun, 1Zhiqiang Ruan and 1Heng Ren
1College of Information Science and Engineering, Hunan University, No. 252, Lushan South Road, Changsha, 410082, China
2Institute of Computer and Software, Nanjing University of Information Science and Technology, No. 219, Ningliu Road, Nanjing, 210044, China

Abstract: To achieve a more comprehensive and sustained privacy control and tamper detection for the data in wireless sensor networks, we propose a novel multiple watermarking method, called Multi-mark, which consists of an annotation part and a fragile part. On the one hand, encrypted user’s personal information is embedded into the routine monitoring data, as an annotation watermark, which can be extracted when needed. On the other hand, tampering is detected using fragile watermark. The former can resist various manipulative attacks, while the latter can detect any malicious modifications. Multi-mark not only provides privacy and security, but also saves data transmission amount and storage space. The experimental results show that Multi-mark can reduce 30% of data traffic and only introduces very low computation cost. Multi-mark is a network structure-free scheme, which can be easily and efficiently applied to the resource limited sensor networks.

Key words: Wireless sensor network, privacy data protection, multiple watermarking, annotation watermark, data integrity

INTRODUCTION

Wireless sensor networks become increasingly ubiquitous and more people-centric. They are being gradually applied to the daily life of human beings. Wireless sensor network technology has the potential to change the way we live, work, and do business (Li and Liu, 2009; Yang and Liu, 2010). It can offer viable solutions for a variety of people-centric applications, such as health care, smart space and public safety.

At the same time, privacy data protection issues are becoming more prominent. People expect to enjoy the convenience of WSN services without revealing privacy. People might consider the WBAN technology as a potential threat to freedom, if the applications go beyond secure medical usage. Social acceptance would be the key to this technology finding a wider application. There has been a wealth of researches focused on the privacy problems in WSNs. There are two main types of privacy concerns, context-oriented concerns and data-oriented. Context-oriented concerns concentrate on the contextual information, such as the location and timing of traffic flows in a WSN. There are four existing techniques include flooding, Random walk (Kamat et al., 2005), dummy injection (Shao et al., 2007) and fake data sources (Mehta et al., 2007) against the disclosure of the location of data source. Privacy-preserving technique for the location of the base station is proposed in literature (Jian et al., 2007). Temporal privacy in WSN is formulated by Kamat et al. (2007). Data-oriented concerns focus on the privacy of data collected from, or query posted to. There are three existing techniques to protect the privacy of data being collected (He et al., 2007; Sheng and Li, 2008; Zhang et al., 2008) and the privacy of queries being posed to a WSN is discussed in (Zhang, et al., 2009). More researches are focused on data secure transmission in WSNs (Ruan et al., 2010; Ren et al., 2011).

Watermarking technologies have been extensively studied in the multimedia domain. However, rarely has this technique been used in WSNs. Feng and Potkonjak (2003) used digital watermarking into WSNs in conjunction with non-linear programming, where the additional constraints are used to embed watermarks. But the special formulation may limit its application to many different domains. Guo and Li (2007) used the watermarking method to verify the integrity of streaming data. But it was not designed for sensors with limited computing and power.

Although many efforts have been carried out to mitigate the risks of privacy exposure, privacy control and tamper detection for people-centric WSN data remains to
be a challenging issue (Stankovic et al., 2005). For example, in a wireless sensor network system for health care or condition monitoring of infectious diseases, people’s body physiological parameters are the routine monitoring data and have to be real- timely reported (Zhou et al., 2006). In most cases, this is enough for monitoring. But when these parameters are detected abnormally, the time and the location of the history sensor reading and even the people’s identities of these data could be traced. That is, the collection of people’s privacy data is inevitable. How to ensure reliable, efficient and privacy-preserving data collection and store is a new challenge. People’s private data must be transmitted and stored in an anonymous style to prevent any unauthorized disclosure and use (Horey et al., 2007). Furthermore, these data must preserve stringent data integrity in all their life-cycle and any malicious modification could be detected.

To achieve a more comprehensive and sustained privacy control and tamper detection, we propose a multiple watermarking method, called Multi-mark, which consists of an annotation part and a fragile part. Encrypted user’s identity information can be embedded into the data as an annotation watermark and tampering can be detected using fragile watermark method. In our design, the former can resist various manipulative attacks, while the latter can detect any malicious modifications. The objectives of Multi-Mark are as follows. (1) It can be easily and efficiently applied to the resource limited sensor networks. (2) It can resist common attacks: packet forgery attack (data insertion attack), selective forwarding (data delete attack), packet replay (data replication attack), packet transfer delay (data rearrangement attack) and packet tampering (data modification attack). (3) Multi-Mark not only provides privacy and security, but also reduces transmission data amount and storage space by introducing very low computation cost.

MULTIPLE WATERMARKING METHOD

Multi-Mark consists of an annotation part and a fragile part. People’s private data (identity and location information) are embedded into the routine monitoring data as annotation watermark. Most of the routine monitoring data is numerical data. Its Last Significant Bit (LSB) is estimated value. Changes of data LSB within a small range are usually acceptable. Base on such an observation, we design a LSB-based robust annotation watermarking embedding algorithm. Different from traditional LSB methods, we define it as Odd-Even Embedding. In which, if the least significant bit is odd, it marked ‘1’. Otherwise, it marked ‘0’. This scheme modifies the original data in the smallest possible range and doesn’t affect the availability of data, so it has strong ability of anti-steganalysis. Even if the eavesdropper monitored the transmission signal, it will be still difficult to discover the existence of the watermark information. As the annotation watermark is usually small, it can be embedded multiple times consecutively to enhance the robustness. The scheme can remain intact with the protected content under various manipulative attacks.

For preserving stringent data integrity, we design a fragile watermark algorithm, which is a chain scheme using dynamic group size. There are the following highlights: Firstly, a lossless fragile watermarking embedding algorithm is designed. We convert the numerical routine monitoring data into character, so that the blank character based embedding scheme can be used. Second, in the data group partition scheme, dynamic group size is adopted. So the scheme is able to detect packet replay attack or data replication attack. Fragile watermark is good for tamper detection and can detect any malicious modification. Thirdly, each group hash value as the fragile watermark, which is unique and closely related with the data itself, is embedded into the data in each group. Any modifications made to one group will have no affect on data integrity authentication in any other groups.

The working model of Multi-mark is shown in Fig. 1. On the data source node, the annotation watermark is embedded into the routine monitoring data. Then the fragile watermark is generated and embedded into the watermarked data. When the final watermarked data are transmitted through the WSN, any mistakes might happen because of the bad network condition or malicious attacks. On the sink, tamperers can be detected by authenticating the fragile watermark. When needed, the annotation watermark can be extracted.

To facilitate the description, we will introduce the notions, definitions and rules used in Multi-mark firstly.

Notations, definitions and rules: We use the notations and expressions in Table 1. As shown in Table 1, the routine monitoring numerical data are regarded as a data set \( D = \{d_1, d_2, ..., d_n\} \), in which \( d_i \) indicates one sampling which is expressed in decimal. They are sorted according to their acquisition time. Accordingly, when the annotation watermark is embedded, the watermarked routine monitoring data is denoted as \( D' = \{d_{1}', d_{2}', ..., d_{n}'\} \). When the fragile watermark is embedded, the final watermarked data one the source node is denoted as \( W = \{w(d_{1}'), w(d_{2}'), ..., w(d_{n}')\} \). The received data set which are reordered by the packet count number in the packets, is denoted as \( W' = \{w'_{1}(d_{1}'), w'_{2}(d_{2}'), ..., w'_{n}(d_{n}')\} \).
The annotation watermark $M_a$ is embedded repeatedly, one bit of $M_a$ [0] can be embedded into each data sample $d_i$, so the data in each $M_a$ embedding period can be called a working period, which is denoted as $D$. The length of $D$ is $m$. So the data set $D$ can be partitioned to a set partition $\{D_0, D_1, ..., D_{m-1}\}$, where $\{D_i = D_{a0}, D_{a1}, ..., D_{a(m-1)}\}$.

**Rule 1 (LSB state transition rule):** The least significant bit of each sampling is denoted as $B$. If $0 \leq B \leq 4$, then $B = B+1$, else if $5 \leq B \leq 9$, then $B = B-1$. As shown in Table 1, we use the expression LSB-Trans() to indicate Rule 1.

**Rule 2 (Group size generation rule):** Assume that the $j$th group size is $s_j (x \leq s_j \leq y)$, in which $x$ and $y$ are its lower and upper bounds as shown in Table 1. For each $H_i = HASH^*(k_0), (0 \leq i \leq q)$, $s_i = max (H_i y, y-H_i)$. So the data set $D$ can be partitioned to a set partition $\{D'_0, D'_1, ..., D'_{q-1}\}$, where:

$$D'_i = \left\{ \sum_{j=0}^{s_i} \sum_{l=0}^{s_i} \sum_{m=0}^{s_i} \sum_{n=0}^{s_i} \right\}$$

**Rule 3 (Fragile watermark generation rule):** For each data set $D'_i$, compute $H_{M_a} = (\text{watermark}_i + \text{watermark}_i) + ...$

$$\text{hash}(\text{watermark}_i) +$$

in which "+" is connection operation. Then partition $\text{mark}$ into $L$ equal parts $\{H_n, H_{n-1}, ..., H_0\}$. Then $D_i'$'s fragile watermark $M_f(D_i') = (H_0 \oplus H_{n-1} \oplus ... \oplus H_0)$, in which "\oplus" is XOR operation. The length of $M_f(D_i')$ is $s_i$.

**Annotation watermark:** The annotation-watermarking scheme includes watermark embedding algorithm (Algorithm 1) and extraction algorithm (Algorithm 2). The former runs on the data source node, while the latter runs on the sink of the network.

**Algorithm 1: Annotation watermark embedding**

**Input:** The data set $D$, the annotation watermark $M_a$

**Output:** The data set $D$

**Steps:**

1. for each data set $D_i$ in $D$
2. for each data $d_i$ in $D_i$
3. if $(M_a[i] == '0' \ and \ LSB(d_i) \ is \ odd)$
4. $d_i = \text{LSB-Trans}(d_i)$
5. else if $(M_a[i] == '1' \ and \ LSB(d_i) \ is \ even)$
6. $d_i = \text{LSB-Trans}(d_i)$
7. }
8. }
9. return $D$

**Annotation watermark embedding algorithm:** As shown in Algorithm 1, the routine monitoring data set $D$ and the annotation watermark are inputs. The annotation
watermark, which is specially encoded as binary strings, is embedded into the data one time in each working period. Each bit of watermark is embedded into the corresponding sampling of the data in each period by changing the parity of its least significant bit using Rule 1. We get the watermarked routine monitoring data $D'$.  

**Annotation watermark extraction algorithm:** The annotation watermark extraction algorithm is described in Algorithm 2.

Algorithm 2: Annotation watermark extraction

**Input:** The data set $W$ received at the sink  
**Output:** The extracted watermark $M_w$  
**Steps:**  
1: for each group $W_i$ in $W$  
2: {for each data $d_i$ in $W_i$  
3: if (LSB($d_i$) is odd)  
4: $M_w^i$ = $M_w^i$ + $1$;  
5: else if (LSB($d_i$) is even)  
6: $M_w^i$ = $M_w^i$ + $0$;  
7: }  
8: }  
9: return $M_w$

The input of the annotation watermark extraction algorithm is the data set $W$ received on the sink of the network. By judging the parity of the least significant bit of each data in $W$, one possible copy of annotation watermark $M_w$ can be extracted from the data in each working period. In fact, the annotation watermark extraction operation is always run after the data integrity authentication. As long as the data in one of the periods are not tampered, we can get the annotation watermark information. Even there are mistakes in the data in all working periods, we still have the chance to extract the annotation watermark $M_w$ by analyzing multiple sets of error watermark information.

**Fragile watermark:** Fragile watermark also includes two parts, a watermark embedding algorithm which works on the data source node and an integrity authentication algorithm which works on the sink node.

**Fragile watermark embedding algorithm:** After the annotation watermark embedding process, the watermarked data set $D'$ is buffered in the buffer on the source sensor node. The buffer size is generated dynamically, which is computed by Rule 2. When the buffer is full, that is we get a group of data $W_i$, the fragile watermark is generated using Rule 3. Using the watermarking embedding rule, the fragile watermark is embedded into the data. The detailed steps are described in Algorithm 3.

Algorithm 3: Fragile watermark embedding

**Input:** the data set $D'$  
**Output:** The data set $W$  
**Steps:**  
1: for each group $D_i'$ in $D'$  
2: {  
3: for each bit $M_i[j]$ in $W(D_i')$  
4: if ($M_i[j]$ == "1")  
5: $W(d_j) = C_d(d'_j) + n$;  
6: else if ($M_i[j]$ == "0")  
7: $W(d_j) = C_d(d'_j)$;  
8: }  
9: }  
10: return $W$

**Data integrity authentication scheme:** When the data is transmitted to the sink node, the data integrity authentication operation is performed. The received data is $W$. Using Rule 2, the group size of $W$ can be re-computed using the same one-way hash chain and the seed key $k_c$ used by the source sensor node. $W$ can be partitioned to a set partition $W' = \{W'_1, W'_2, ..., W'_{p'}\}$. As shown in Algorithm 4, we can extract the embedded fragile watermark $M_i$ from $W'_i$; meanwhile, $W_i$ is converted into numerical set $N_i$, using Rule 3, we can compute its fragile watermark which is denoted as $M_i$. If $M_i$ equals to $M_i$, the data is integrated; otherwise, the data is modified.

Algorithm 4: Integrity authentication

**Input:** The received data set $W$  
**Output:** The data integrity is OK or NOT  
**Steps:**  
1: for each group $W_i$ in $W'$  
2: {for each data $W_i'j$ in $W_i'$  
3: if (LSB($W_i'$) == "")  
4: $M_i = M_i + "1"$;  
5: else  
6: $M_i = M_i + "0"$;  
7: }  
8: $M_i = M_i (N_i)$  
9: if ($M_i$ == $M_i$)  
10: return 1;  
11: else  
12: return 0;

**PERFORMANCE EVALUATION**

The experimental network environment could be simplified as follows: All sensor nodes collect the routine monitoring data continuously; then annotation watermark and fragile watermark are embedded into the data; finally, the final watermarked data are sent to the sink by a Shortest Path Tree. On the sink, it can validate the integrity of the data using the fragile watermark and extract the annotation watermarks when needed.

**Experiment setup:** The experiment setup is divided into two parts.
In the 1st part, we build a real WSN with 10 TelosB sensor nodes. In the environment of Tiny OS, the 802.15 wireless network protocols are specified and the nodes are connected using a Shortest Path Tree routing algorithm. The network takes 34 bytes broadcast packet and the payload packet transmitted contains 26 bytes during the experiment, which includes 6 bytes of packet head and 20 bytes data readings. To simplify the design, in each packet, there is a packet count number to assist the watermark embedding and extraction. This network is named Original WSN here. Then we implement the Multi-mark scheme. As the node feature restrictions, we use the temperature sensor on Telos B to collect the environment temperature data. These data are used to indicate the routine monitoring data in a people-centric WSN. We implement the annotation watermark embedding algorithm and the fragile watermarking embedding algorithm on the TelosB nodes using NesC. Meanwhile the watermark extracting scheme and integrity authentication scheme are implemented on the sink node which is a PC using JAVA. In the experiment, there are 10 samples in each packet. The length of annotation watermark is 16 bits. All source sensor nodes can embed watermark and send data to the sink. On the sink, it can validate the integrity of the data using the fragile watermark and extract the annotation watermark. This network is named Multi-mark WSN.

In the 2nd part, we evaluate the performance of Multi-mark using a custom simulator. In the simulation setup, we randomly deploy 1000 sensor nodes into a (600×600 m²) square sensing field. The node parameters are the same as the real TelosB node and the real communication situation. And we use the data trace from a real sensor network established in China Ocean University as the routine monitoring data. We implemented four data transmission schemes: (1) Scheme without watermarking (No-Mark), in which the annotation information is transferred separately; (2) Scheme with annotation watermarking only (Annotation-only); (3) Scheme with fragile watermarking only (Fragile-only) and (4) Multi-mark Scheme.

Experimental results
Feasibility authentication: We use the 10 nodes real sensor network, which is described in the first part of experiment setup, to evaluate the feasibility of Multi-mark. Under the circumstance, which means no packet loss or bit error, we can see that both the annotation watermark and the fragile watermark can be embedded and extracted triumphantly. And there is no significant affect on the transmission. The experiment results show that Multi-mark can work well on the resource-constrained sensor nodes. More importantly, compared to the Original WSN, the only change in Multi-mark WSN is to add watermark embedding codes in the program, which runs on the source node. It does not need to make any changes to the original network structure. So Multi-mark is network structure-free and can be easily deployed on any sensor networks. The authentication experimental results confirm our original design intentions of Multi-mark.

Anti-attack ability: We will evaluate how the network works in environments with different kind of attacks using the 10 nodes real sensor network. One sensor node is randomly selected as the source node and a PC as the sink node. The other nodes are used as the medium nodes, selecting one of the medium nodes as a malicious node. We designed the following five attacks separately: packet forgery attack (data insertion attack), selective forwarding (data delete attack), packet replay (data replication attack), packet transfer delay (data rearrangement attack) and packet tampering (data modification attack). We run each type of experiments 10 times in different attack intensity. The experimental results are presented in Table 2 which includes integrity authentication results and the annotation watermark extraction results.

In Table 2, the integrity authentication experimental results show that Multi-mark can resist selective forwarding, packet forgery, packet replay and packet tampering attacks. The detection rates for these attacks all reach 100% and confirm our original design intentions.

<table>
<thead>
<tr>
<th>Attacks</th>
<th>Integrity authentication</th>
<th>Annotation watermark extraction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No. of experiments</td>
<td>Detection rate (%)</td>
</tr>
<tr>
<td>Forgery</td>
<td>10</td>
<td>100</td>
</tr>
<tr>
<td>Selective forwarding</td>
<td>10</td>
<td>100</td>
</tr>
<tr>
<td>Replay</td>
<td>20</td>
<td>100</td>
</tr>
<tr>
<td>Transfer delay</td>
<td>20</td>
<td>80</td>
</tr>
<tr>
<td>Tampering</td>
<td>10</td>
<td>100</td>
</tr>
</tbody>
</table>
But as shown in Table 2, it fails to resist the packet transfer delay attacks sometimes. The reason is as follows. On the source node, the data needs to be buffered in the buffer before fragile watermark is embedded. In the buffer the packets are reordered according to the packet count number. So if the packet transfer delay time is less than the maximum buffer time, the attacks are eliminated. So this is not a detection failure, but a more successful defence. The experimental results commendably support our objective of the fragile watermark scheme design in Multi-mark.

The annotation watermark extraction experimental results are shown in the column named annotation watermark extraction in Table 2. It shows that Multi-mark can resist packet forgery, packet replay, packet transfer delay and packet tampering attacks. The extraction success rates all reach 100%. These can confirm our original design intentions. But it fails to resist selective forwarding attacks in one experiment. The reason is that in the failed experiment, the packet loss rate in the selective forwarding attack reaches 60%. All of the copies of embedded annotation watermark are severely damaged, there are no any chances to extract the annotation watermark by analyzing multiple sets of error watermark information. In fact, in this case, all of the transmitted data has been severely damaged. These results also commendably support the design objective of the annotation watermark in Multi-mark.

**Data transmission amount analysis:** Figure 2a shows the transmission amount difference between Multi-mark and the scheme No-Mark. We can see that the total transmission amount of Multi-mark is significantly decreased. This comparison results confirm our original design intentions. This is because that in Multi-mark, the annotation watermark information is embedded using the LSB-based method and no longer need to be transmitted separately. Fig. 2b shows the transmission amount comparison of the Fragile-only Multi-mark scheme, MAC-based scheme and the scheme without any protection (No-Multi-Mark). We can see that comparing to the No-Multi-Mark Scheme, Fragile-only scheme introduces some additional data transmission. This is because the fragile watermarking information is embedded using a blank symbol based method, in which “1” introduces an extra blank symbol and “0” does not introduce any symbols. Thus, in the average case, one bit additional data is increased for every two data samplings. In order to compare with other integrity protection schemes, we also implement a MAC (Message Authentication Code) based scheme. There are many
MAC-based schemes have been proposed by researchers. We have implemented a SHA1-based one (Bellare et al., 1996). The length of SHA1 is 160 bits. In Fragile-only Multi-mark scheme, we also use SHA1 algorithm. As shown in Fig. 2b, in Fragile-only scheme, the data transmission amount is decreased than MAC-based scheme. Thus, the Fragile-only scheme provides better data integrity protection by less data transmission overhead. This comparison results confirm our original design intentions. Figure 2c shows the comparison of Multi-mark scheme with the full data transmission scheme in which the annotation information is transmitted separately and MAC-based method is used to protect the data integrity (normal+MAC). We can see that Multi-mark can reduce 30% of data traffic. All of the comparison results commendably support our analysis and objective of the study.

**Energy consumption:** When we estimate energy consumption, it is assumed that the cost in data storing, broadcast, communication and routing are equal to convention. So the only considered point is the energy consumed in data embedding and transmitting. Usually, trading computation for transmission can conserve energy and typically on the order of 3000 instructions can be executed for the energy cost required to transmit 1 bit data. Let Le denote the 1 instruction energy consumed and the transmission energy consumed will be 3000e. In our transmission data experiments, the transmitting traffic can be reduced 236.385 bytes for per 320 bytes sense data. But all the costs for this reducing are 320 bits data embedding operations. Therefore, the extra energy consumed in watermark embedding is far less than the energy saved in data transmission reducing. The analysis results again well support the design goal of Multi-mark.

**CONCLUSION**

Multi-mark is suitable for privacy data protection in people-centric WSNs. The annotation watermark can be used to embed user information in private and secure manner, while the fragile watermark offers tamper detection. Multi-mark not only offers privacy and security, but also saves data transmission amount and storage space. The experimental results show that Multi-mark can reduce 30% of data traffic and only introduces very low computation cost. Multi-mark is a network structure-free scheme. It can be easily and efficiently applied to the resource limited sensor networks.

Future work on Multi-mark will lead in two directions. One is to design watermarking schemes without packet count number. The other one is that we will improve Multi-mark to protect multi-attribute privacy data which might change over time. So the users can control which kinds of privacy data at which time can be read.

**ACKNOWLEDGMENTS**

This study was partially supported by National Basic Research Program of China (973 Program) under Grant No. 2009CB326202 (2009-2011.8, Hunan University) and 2010CB334706 (2010-2013.3, Hunan University), Key Program of National Natural Science Foundation of China under Grant No. 60736016 (2008-2011.12, Hunan University), National Natural Science Foundation of China under Grant No.60873198 (2009-2011.12, Hunan University), 60973126 (2010-2012.12, Hunan University), 60973113 (2010-2012.12, Hunan University), 61073191 (2011-2013.12, Hunan University), 61070196 (2011-2013.12, Hunan University).

**REFERENCES**


