Copyright © 2019 Tech Science Press CMC, vol.58, no.3, pp.679-695, 2019

A Dual-Chaining Watermark Scheme for Data Integrity Protection in Internet of Things

Baowei Wang[1, 2, *], Weiwen Kong[1], Wei Li[1] and Neal N. Xiong[3]

Abstract: Chaining watermark is an effective way to verify the integrity of streaming data in wireless network environment, especially in resource-constrained sensor networks, such as the perception layer of Internet of Things applications. However, in all existing single chaining watermark schemes, how to ensure the synchronization between the data sender and the receiver is still an unsolved problem. Once the synchronization points are attacked by the adversary, existing data integrity authentication schemes are difficult to work properly, and the false negative rate might be up to 50 percent. And the additional fixed group delimiters not only increase the data size, but are also easily detected by adversaries. In this paper, we propose an effective dual-chaining watermark scheme, called DCW, for data integrity protection in smart campus IoT applications. The proposed DCW scheme has the following three characteristics: (1) In order to authenticate the integrity of the data, fragile watermarks are generated and embedded into the data in a chaining way using dynamic grouping; (2) Instead of additional fixed group delimiters, chained watermark delimiters are proposed to synchronize the both transmission sides in case of the synchronization points are tampered; (3) To achieve lossless integrity authentication, a reversible watermarking technique is applied. The experimental results and security analysis can prove that the proposed DCW scheme is able to effectively authenticate the integrity of the data with free distortion at low cost in our smart meteorological Internet of Things system.

Keywords: Dual-chaining, reversible watermark, integrity authentication, Internet of things.

1 Introduction

The rapid development of Internet of Things (IoT) has caused great changes in the national security, industry and people’s lives [Zhang, Sun and Wang (2016); Wang, Gu and Zhou (2017)]. It has the potential to become a vital part of our infrastructure to enrich

[1] Jiangsu Engineering Centre of Network Monitoring, Jiangsu Collaborative Innovation Center on Atmospheric Environment and Equipment Technology, School of Computer and Software, Nanjing University of Information Science & Technology, Nanjing, 210044, China.

[2] Engineering Research Center of Electronic Information and Control from Fujian Provincial Education Department, Fuzhou, 350108, China.

[3] Department of Mathematics and Computer Science, Northeastern State University, Oklahoma, USA.

[*] Corresponding Author: Baowei Wang. Email: wbw.first@163.com.

CMC. doi:10.32604/cmc.2019.06106 www.techscience.com/cmc

lives and make processes easier. It offers economically viable solutions for a variety of people-centric applications, such as, health care, working space automation, public safety and smart space. Smart campus is exactly the most typical application scenario of Internet of Things. As the core part of the smart IoT system perception layer, wireless sensor networks (WSNs) are organized by a great number of sensors with limited computational capacity and power to form a self-organizing network in wireless communication. In the smart campus system, the massive real-time sensory data is continuously collected by the source nodes and sent to the sink node through multi-hop relays. It is used to make decisions and discover deep levels of knowledge [Wang, Gu, Ma et al. (2017); Xia, Wang, Sun et al. (2014)]. Due to the limited resources, traditional data security solutions based on cryptography and Message Authentication Code (MAC) cannot be applied to sensor nodes. Sensor nodes are susceptible to a variety of attacks, such as data forgery attack, data tampering attack, and packets selective forwarding attack. So, verifying the integrity of the transmitted data is critical for most smart space IoT applications.

Lightweight digital watermarking techniques are gradually introduced into the wireless sensor networks and smart Internet of Things to verify the integrity of the data [Feng and Potkonjak (2003); Wang, Yan, Li et al. (2015); Zhang, Liu, Das et al. (2008); Wang, Sun, Ruan et al. (2011); Dong and Li (2009)]. Chaining watermarks are considered to be the most effective method [Guo, Li and Jajodia (2007)]. The core ideas of the existing single chaining watermark schemes are as follows: The data collected by a source node is defined as a data stream. It is divided into multiple groups using the synchronization points or the group delimiters, and the fragile watermark generated by the data in the current group is embedded into the next (or previous) group to form a watermarking chain. Any tamper or attack on the watermarked data would corrupt the watermarking chain. It can efficiently detect and locate any modifications made to the data and authenticate the data integrity [Juma, Kamel and Kaya (2008); Kamel and Juma (2010); Kamel and Juma (2011); Shi and Xiao (2013); Liu, Ge, Zhu et al. (2014)].

But all existing single chaining watermarking schemes still have the following bottlenecks: (1) The tampering of the indispensable synchronization points may lead to a high false negative rate of 50 percent and result in completely meaninglessness of data integrity authentication; (2) The additional fixed group delimiters, which are usually special data elements or fixed packet segments, can be detected easily by the adversaries. (3) It is unacceptable that the embedding method leads to certain irreversible changes to the original data.

In this article, an effective dual-chaining watermark scheme is proposed to ensure the data integrity in the perception layer of smart IoT system. The proposed DCW scheme has the following three characteristics:

(1) Fragile watermarks are generated and embedded into the data stream using a dynamic chaining method to verify the data integrity.

(2) Instead of the additional fixed group delimiter, a chained watermark one is designed to synchronize the both transmission sides in case of the synchronization points failure.

(3) To achieve completely lossless integrity authentication, a reversible watermark algorithm is applied to DCW.

The proposed DCW scheme is evaluated in a real smart campus meteorological Internet of Things system, which is deployed in Nanjing University of Information Science and Technology. The experimental results and security analysis show that the proposed scheme can effectively authenticate the integrity of the data with free distortion and tiny computational overhead. It does not increase the data size nor change its accuracy. Furthermore, DCW can significantly improve the ability to detect and locate the various packet level attacks. Meanwhile, the adversaries can hardly detect the existence of the dual-chaining watermarking based integrity authentication scheme.

The rest of the paper is organized as follows. Section 2 introduces some related works of single chaining watermark schemes. Section 3 presents our proposed dual-chaining watermark scheme. In Section 4, the experimental results and the analysis are introduced. Section 5 includes the work.

2 Related works

Chaining watermark is considered as a most effective way to verify the integrity of streaming data in network environment. The first single chaining watermark scheme, called SGW, is proposed by Guo et al. [Guo, Li and Jajodia (2007)]. The SGW scheme is used to authentication the integrity of streaming data in network environment. The data stream is divided into groups of variable size by synchronization points. Fragile watermarks are chained across data groups. Any modifications can be detected and located. The problem with the SGW scheme is that the data insertion, modification, and deletion attacks may create confusion at the receiver. It is difficult for the receiver to track the synchronization points. The synchronization points have been the biggest bottleneck of existing dynamic grouping-based single chaining watermark methods.

Juma et al. take the lead in applying the chaining watermark technology to wireless sensor networks. In Juma et al. [Juma, Kamel and Kaya (2008)], they presented two single chaining watermarking methods, called S-SGW and FWC, respectively. In Kamel et al. [Kamel and Juma (2010)], the authors proposed the LWC scheme to try to avoid several drawbacks of SGW. However, these three schemes still suffer from the same bottleneck as the SGW. In Kamel et al. [Kamel and Juma (2011)], the FWC-D scheme uses the group delimiters to synchronize the sender side and the receiver side. In FWC-D, the dynamic grouping method is abandoned, and additional data elements are used as the group delimiters. It is very easy to be detected by the adversaries.

Another common disadvantage of the existing schemes is that watermarks are embedded by replacing the least significant bits (LSB) of the data, which can significantly affect the data accuracy. It is unacceptable that the methods used to protect the data destroy its integrity, especially, in some critical applications such as smart campus, medical care and military applications.

Reversible chaining watermarking technologies are adopted to address this problem [Qiu (2017); Yuan, Xia and Sun (2017)]. In Shi et al. [Shi and Xiao (2013)], a prediction-error expansion based reversible watermarking scheme is proposed by Shi and Li. It can avoid any modification to the data for watermark embedding. But the synchronization point is still its bottleneck. Once the synchronization points are attacked, the false negative rate will be up to 50% and the integrity authentication might be completely meaninglessness.

Shi's scheme does not take into account the upper bound of the buffer size, the buffer overflow would occur in extreme cases [Liu, Ge, Zhu et al. (2014)]. To address this problem, Liu et al. proposed a histogram shifting based reversible watermarking scheme for data integrity protection in BSN. Liu's scheme does not use dynamic grouping but adopts a fixed grouping method to avoid the drawback of synchronization points. In the head segment of the system packet, a serial number (*SN*) segment and a group delimiter (*DF*) segment are attached. Not only increase the transmission overhead, but also easily detected by the adversaries.

3 The dual-chaining watermarking scheme

3.1 Overview of the system model

The system model of the proposed DCW scheme for data integrity protection in smart IoT system is illustrated in Fig. 1. There are three types of nodes in the perception layer sensor network, including source sensor nodes, relay nodes and the sink node. Source sensor nodes continually collect data samplings, and send them to the sink using data packets through the multi-hop relay nodes. A data packet is represented as (head-segments, payload), where head segments are predefined parameters, including route, packet length, etc.; the payload is the encoded data samplings. For simplicity, we need not present the data on the packet level. Instead, each data sampling is represented as a data element E_i. All the sensory data of a source node is defined as an infinite numerical *data stream* $S=\{E_1, E_2, \ldots, E_n\}$.

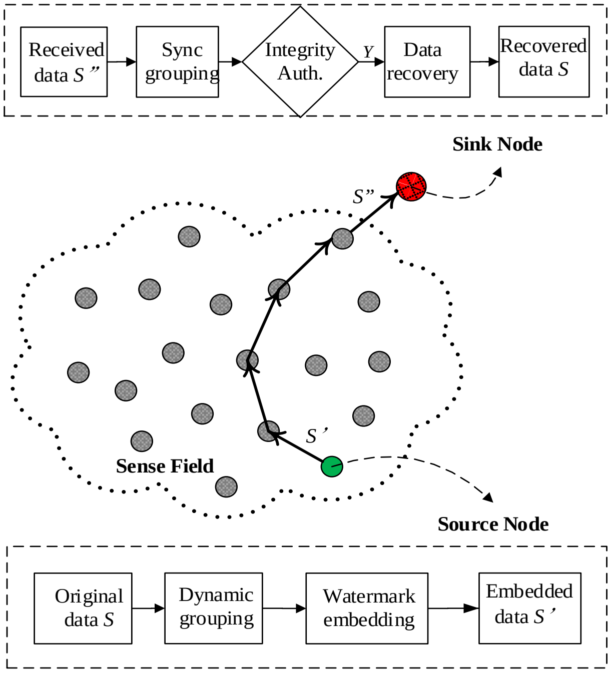


Figure 1: The system model of DCW

To verify the integrity of the data stream, the proposed DCW scheme is adopted in the system's perception layer. The dual-chaining watermarks are generated and embedded on the source node, extracted and verified on the sink node.

(1) On the source sensor, the data stream *S* is cached in two buffers and dynamically divided into different data groups. And two adjacent data groups form an authentication group. The dual-chaining watermarks, including the chaining fragile watermark and the pre-defined chained delimiter watermark, are embedded into the authentication groups in a chaining way.

(2) On the sink, it synchronizes the data groups using the synchronization point and the chained delimiter watermark, and verifies the integrity of the data stream using the chaining fragile watermark. If the data is not temped, the original data can be restored.

3.2 Definitions and rules

Firstly, some system parameters and notations used in the proposed DCW scheme are pre-defined. To simplify the description, all the notations and parameters are shown in Tab. 1, and considered to be known parameters.

Table 1: The system parameters and notions

Symbol	Description
Hash ()	The cryptographic hash function that can be adopted in resource-constrained sensor node.
Key	The secret key used to compute and select the synchronization point.
m	The pre-defined grouping parameter that indicates the expected average group length.
L	The lower bound of the data buffer size
N	The upper bound of the data buffer size
l	The length of the binary delimiter watermark (bits)
W	The binary watermarking

To embed and extract the dual-chaining watermarking, the data elements would firstly be cached and grouped on the both transmission sides, including the source node and the sink node. Two data buffers B_0 and B_1, which are also denoted as Buffer (b_0) and Buffer (b_1), are needed to cache the data groups. The data buffer is defined as follows:

Definition 1: (Data Buffer, *B*) A data buffer *B* is an allocated memory space to cache the data elements. The capacity *N* of a data buffer means that it can cache up to *N* data elements, which is usually a system-defined parameter.

Definition 2: (Synchronization Point, *SP*) A data element E_i is a *SP*, if and only if: Hash $(Key, E_i) \bmod m == 0$.

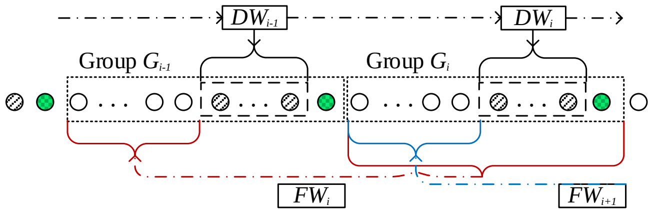
In DCW, the SPs are used for dynamic grouping of data. Since a *SP* is a selected data element using the secret key, and the group length is variable, it is very difficult to find out and destroy the data groups. All the data elements are cached in the two buffers B_0 and B_1, alternately. When each data element is cached into a data buffer, Rule *1* is used to determine if a data group is cached successfully.

Rule 1: (Data Grouping Rule) For each element E_k cached in a buffer: (1) if E_k is not a SP && $k < N$, then continue to buffer the data; (2) if E_k is a SP && $k < L$, where L is defined as the lower bound of the group length, then continue to cache data; (3) if E_k is a SP && $L \leq k \leq N$, then all the data elements cached in the buffer is defined as a Data Group $G_x=\{E_1, E_2, \ldots, E_k\}$; (4) else, that is, when the buffer is full, a suitable SP has not been found, then all the data elements cached in the buffer is defined as a Data Group $G_x=\{E_1, E_2, \ldots, E_N\}$, (no SP is included).

In summary, two adjacent data groups are cached in two buffers. The group size k is a variable value and is bounded by L and N. Buffer overflow would not occur in any cases.

The two adjacent data groups can be defined as an Authentication Group, which is the workspace for embedding and verifying the dual-chaining watermark scheme. To avoid confusion, they are denoted as the previous group (G_{i-1}) and the current group (G_i).

As shown in Fig. 2, the data elements in each data group are divided into three segments, including $S_{F=}\{E_1, E_2, \ldots, E_{k-l-1}\}$, $S_{D=}\{E_{k-l}, E_{k-l+1}, \ldots, E_{k-1}\}$ and $SP=E_k$. The length l of S_D is a fixed value; and the length of S_F is k-l-1. The well-designed fragile watermark and delimiter watermark are embedded into S_F and S_D of the *previous group* G_{i-1} respectively, using a reversible method. The fragile watermark FW_i is generated by all the original data elements in the current group G_i, the delimiter watermark DW_{i-1} is generated using Rule 2.



○Data Element in S_F ▨Data Element in S_D ●Synchronization Point

Figure 2: The proposed Dual-Chaining Watermarking Model

The workflow of the dual-chaining watermarking embedding: When a data group G is cached in the buffer using Rule 1, the following two operations are performed: (1) Firstly, the fragile watermark binary string FW_i of the current group G_i is computed according to Rule 3, and is embedded into the S_F section of the previous group G_{i-1} using Rule 4. A fragile watermarking chain is formed. It is used to verify the data integrity. (2) Meanwhile, the delimiter watermark DW is generated using Rule 2, and is embedded into the S_D section of G_{i-1} using Rule 4. That is the other watermarking chain to keep the both transmission sides synchronized. It can avoid failure in sync grouping when SP confusion happens. Once the dual-watermarking embedding ends, the data in the group G_{i-1} will be sent to the sink node and the next group G_{i+1} will be buffered.

Rule 2: (Delimiter Watermarks Definition) The delimiter watermarks, which is denote as

$DW=\{DW_1, DW_1, \ldots, DW_l\}$, are a set of well-defined fixed-size binary strings. They are embedded into fixed position in each group of data, one by one, to form a watermark chain. For simplicity, they can be generated by a fixed delimiter W_D, DW_i is denoted as $W_D\oplus i$, in which "⊕" is the *XOR* operation. *XOR* is a reversible operation. The length of W_D and DW_i is l, which is equal to the length of S_D.

Rule 3: (Fragile Watermark Generation Rule) The fragile watermark FW_i is generated by the hash of the concatenation of all the data elements in the current group and then embedded into the previous group. FW_i is denoted as $GroupXOR\ (H_1\oplus H_2\oplus\ldots\oplus H_k)$, in which "⊕" is the *XOR* operation. The length of FW_i, which is a fixed value, depends on the hash function. If the length of S_F in previous group is larger than the length of FW_i, FW_i will be embedded repetitively.

In DCW, to make the both watermarks reversible, a difference expansion-based reversible embedding algorithm is designed [Chen, Sun, Sun et al. (2013); Alattar (2004)]. The reversible watermarking embedding rules are defined as follows.

Rule 4: (Watermarking Embedding Rule) Given an numerical data group $G_{x=}\{E_1, E_2,\ldots , E_t\}$, and a binary watermark W,(its length is t-1), the embedding rule is as follows: (1) calculate the average value of G_x data, denoted as $u=\lfloor(E_1 +E_2 +\ldots+E_t)/t\rfloor$; (2) compute the differences between each data element E_j $(j=2, 3, \ldots, t)$ and E_1, denoted as $d_j =E_j -E_1$; (3) shift d_j one bit to the left and embed the corresponding bit of W into the vacant least significant bit of d_j, that is $d_j'= 2\times d_j+W[j-2]$; (4) we can get the watermarked data G_x', in which $E_1'=u -\lfloor(d_2'+d_3' +\ldots+ d_t')/t\rfloor$, $E_j'=E_1'+d_j'$ $(j=2, 3, \ldots, t)$.

Since the data elements are collected in a short period of time, and the d_j changes in a small range, the 1-bit left-shifted operation overflow will not happen [Kamel and Juma (2010); Kamel and Juma (2011)].

Rule 5: (Sync Data Grouping Rule) The only difference between the sync grouping rules and the Rule 1 is: the group ends before the buffer is full, if and only when E_k is a *SP* and simultaneously a delimiter watermark can be extracted from the data elements before E_k.

Rule 6: (Watermark Extraction and Data Recovery Rule) Given a watermarked data group $G_x'=\{E_1', E_2', \ldots, E_t'\}$: (1) calculate the average value of G_x', denoted as $u=\lfloor(E_1' +E_2' +\ldots+E_t')/t\rfloor$; (2) compute the differences between each E_j' (j=2, 3, …, t) and E_1', denoted as $d_j'=E_j' - E_1'$; (3) the binary watermarking W_x' can be extracted from the least significant bit of the differences; the corresponding watermark bit is $W_x'\ [j-2] =LSB(d_j')$, and the length of W_x' is t-1; (4) shift d_j' one bit to the right, $d_j= \lfloor d_j' /2\rfloor$; (4) we can get the recovered data G_x, in which $E_1=u -\lfloor(d_2+d_3 +\ldots+ d_t)/t\rfloor$, $E_j=E_1'+d_j$ (j=2, 3,…, t).

The workflow of the dual-chaining watermarking based data integrity authentication scheme is as follows: (1) Two groups of data G_{i-1}' and G_i' are cached correctly using Rule 5. As shown in Fig. 2, data elements in each data group can be divided into three parts, including S_F, S_D and *SP*; (2) the fragile watermarking FW_i' could be extracted from section S_F of G_{i-1}', according to Rule 6; and then the watermarking FW_i'' can be recalculated using Rule 3. Then FW_i' is equal to FW_i'', the data in G_i' is complete; (3) The original data G_i can be regained from G_i' using Rule 6.

3.3 Watermarking embedding algorithm

Without loss of generality, the dual-chaining watermarking generation and embedding schemes, which run on the source sensor node in the perception layer of a smart Internet of Things system, are introduced formally in this section. Two data buffers, which are denoted as Buffer (b_0) and Buffer (b_1), respectively, are needed on each source sensor node. The dual-chaining watermark embedding algorithm is presented in Algorithm 1. It includes the delimiter watermark embedding and the fragile watermark embedding. Algorithm 2 shows the data buffering process on the data source sensor node. The detailed course of the reversible watermarking embedding is designed in Algorithm 3.

Algorithm 1: Dual-Chaining Watermark Embedding(*DW*)

```
clear Buffer ($b_0$), Buffer ($b_1$);
$b_0$=0;
Cache_buffer (Buffer ($b_0$)); // Call Algorithm 2
while ($true$)
{
    $b_1$=($b_0$+1) $mod$ 2;
    $FW_i$=Cache_buffer (Buffer ($b_1$)); // Call Algorithm 2
    $k$=elements number in Buffer($b_0$);
    $l$=bits number in $DW$;
    Embedding (Buffer($b_0$), $k-l-1$, $k-2$, $DW$); // Delimiter watermark embedding
    Embedding (Buffer($b_0$), $0$, $k-2-l$, $FW_i$); // Fragile watermark embedding
    send data in Buffer($b_0$);
    clear Buffer($b_0$);
    $b_0$=$b_1$;
}
```

Algorithm 2: Cache_buffer (Buffer(*b*))

```
$k$=0; $W$=0;
while (receive a data element $E_i$)
{
    Buffer($b$) ($k$ ++)=$E_i$;
    $H$=$Hash$ ($Key$, $E_i$);
    $W$=$W \oplus H$;
    if (($H$ mod $m$==0) && ($k$>=$L$) && ($k$<=$N$) || ($k$==$N$))
        return $W$;
}
```

Algorithm 3: Embedding (*buffer E*, *start*, *end*, *W*)

```
$t=end-start$; //Generate $W$ with the needed length
if ($|W|>=t-1$) // $|W|$ is the length of $W$
    $W$=the first $t-1$ bits of $W$;
else
{
    n=$t/|W|$;
    $W$=n $W$ plus the first $t$ % $|W|$- bits of $W$;
}
$u=\lfloor(E_{start}+E_{start+1}+\ldots+E_{end})/t\rfloor$; // $E_i$ is the $i^{th}$ data element of $E$
for ($j=start+1$ to $end$)
{
    $d_j=E_j-E_{start}$;
    $d_j'=2\times d_j+W[j-start-1]$;
}
$E_{start}=u-\lfloor(d_{start+1}'+d_{start+2}'+\ldots+d_{end}')/t\rfloor$;
for ($j=start+1$ to $end$)
    $E_j=E_1'+d_j'$;
```

3.4 Data integrity authentication algorithm

Without loss of generality, the data integrity authentication and data recovery algorithms, which run on the sink node in a smart Internet of Things system, are presented formally in this subsection. In a smart IoT application scenario, the computing resource, storage space and energy of the sink node are not constrained. It can cache data elements into the two buffers as well using Algorithm 6. As shown in Tab. 1, the watermark embedding security *Key*, the secret parameter m, and the lower bound of the buffer size L and the upper bound of the buffer size N are considered to be known parameters. The complete algorithm for data integrity authentication and original data recovery is shown in Algorithm 4. The extracted watermarks include FW_i' and DW'. Here, FW_i' is the fragile watermark that is the group hash value for integrity protection constructed from the previous and the current group; DW' is the chained delimiter watermark, which extracted by Algorithm 5 for the verification of grouping. The data can be recovered using Algorithm 7.

Algorithm 4: Data Integrity Authentication

```
clear Buffer ($0$), Buffer ($1$);
$b_0=0$;
Cache_sync_buffer (Buffer ($b_0$));
$k$=number of data elements in buffer($b_0$);
$l$=number of bits in $DW$;
$FW_i'$=Extraction(Buffer($b_0$), $0$, $k-2-l$); //Call Algorithm 7
while (true)
{
    $b_1=(b_0+1)$ mod $2$;
    if (Cache_sync_buffer (Buffer ($b_1$)) == 1)
    {
        $k$=number of data elements in Buffer($b_1$);
        $FW_{i+1}'$=Extraction(Buffer($b_1$), $0$, $k-2-l$); //Call Algorithm 7
        for ($j=0$ to $k-1$)
            $FW_i''=FW_i''+Hash$ (Buffer($b_1$) ($j$));
        if ($FW_i' == FW_i''$)
            return 1; //Integrity
        else
            return 0; //Tampered
    }
    else
        return 0; //Tampered
    clear Buffer($j_0$);
    $b_0=b_1$;
}
```

Algorithm 5: Delimiter (*buffer E', start, end*)

```
for ($j=start+1$ to $end$)
{
    $d_j'=E_j' - E_{start}'$;
    $W' [j-start-1] =LSB(d_j')$;
}
return $W'$;
```

Algorithm 6: Cache_sync_buffer (Buffer(*b*))

```
$k$=0; $W$=0;
while (receive a data element $E_i$)
{
    Buffer($b$) ($k$ ++)=$E_i$;
    $H_{=}Hash\ (Key, E_i)$;
    $W=W \oplus H$;
    if ((($H$ mod $m$ == 0 && Delimiter (Buffer($b$), $k$- 1-$l$, $k$-1) == $DW$) &&
   ($k$ >= $L$) && ($k$ <= $N$)) || ($k$ == $N$ && Delimiter(Buffer($b$), $k$-$l$, $k$) == $DW$))
        return 1; // Sync Success;
    if ((Delimiter (Buffer($b$), $k$ - $l$, $k$)! =$DW$ && $k$ == $N$)
        return 0; // Sync Failure
}
```

Algorithm 7: Extraction (*buffer E', start, end*)

```
t=$end$ - $start$;
$u'=\lfloor (E_{start}'+E_{start+1}'+\ldots+E_{end}') / t \rfloor$;
for ($j=start+1$ to $end$)
{
    $d_j'=E_j' - E_{start}'$;
    $W'\ [j-start-1]=LSB(d_j')$;
    $d_{j=}\lfloor d_j' / 2\rfloor$;
}
$E_{start}'=u' -\lfloor (d_{start+1}+d_{start+2}+\ldots+d_{end}) / t \rfloor$; //Data Recovery
for ($j=start+1$ to $end$)
    $E_j'=E_j'+d_j$;
return $W'$;
```

4 Experiments and performance evaluation

In the performance evaluation section, we evaluate the performance of the proposed dual-chaining watermark scheme for data integrity protection from multiple perspectives. We use our own real smart meteorological Internet of Things system which is deployed in Nanjing University of Information Science and Technology, as shown in Fig. 3. The sensor nodes used in this meteorological IoT system are our self-developed products [Wang, Gu, Ma et al. (2017); Wang, Gu and Yan (2018)], which is shown in Fig. 4. The propose DCW scheme is implemented on the node. In this IoT system, the temperature, humidity data are gathered once per minute, and are used for embedding the dual-chaining watermarks.

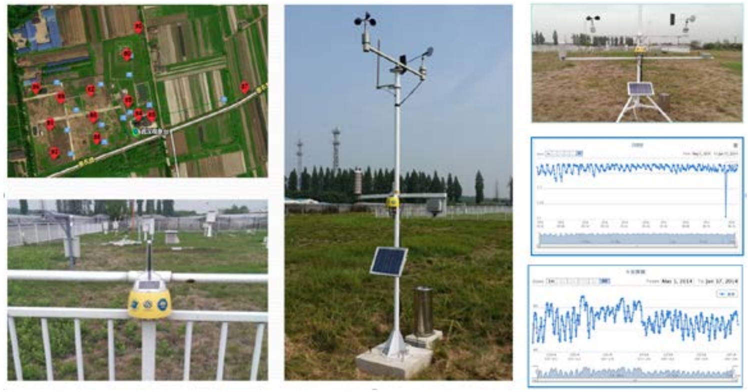


Figure 3: The smart meteorological Internet of Things system deployed in NUIST

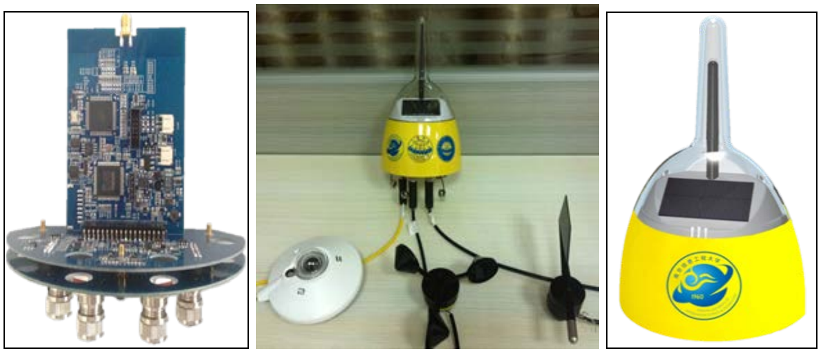


Figure 4: The self-developed sensor nodes used in the experiment

4.1 Data accuracy

In the people centric Internet of Things systems, such as smart campus application, the data accuracy is very important. The watermark embedding method introduces some data changes. Since these changes are usually very small, and insignificant on the perception results, it is hard to be perceived. So, the data accuracy and the invisibility are acceptable under normal circumstances. Tab. 2 shows the statistical comparison in terms of the rate of change of the mean and the variance of data change. Small changes in mean and variance indicate that our proposed dual-chaining watermark scheme has fairly better invisibility. Furthermore, people can recover the original data stream with scarcely error when needed.

Table 2: The statistical comparison of the data change

Attribute	Original mean	Change in mean	Original variance	Change in variance	Restored mean
1	9.76	0.3109	0.0067	1.9523	9.76

2	14.53	0.0012	0.0006	0.0034	14.53
3	23.36	0.0017	0.0018	0.0047	23.36
4	49.45	0.0013	1.0054	0.0946	49.45
5	57.50	0.0026	0.0027	0.0625	57.50

4.2 Data transmission amount

In this subsection, we present the comparison of the data transmission amount among the original data, the existing single chaining watermark scheme (FWC-D) and the proposed dual-chaining watermarking scheme (DCW). In Fig. 5, the x-axis denotes the sense data amount and the y-axis denotes the corresponding data transmission amount. Because the FWC-D scheme directly adds the additional group delimiters into the data groups, the additional amount of data transmission is extremely high. Furthermore, the LSB-based watermarking embedding also destroys the accuracy of the data. Different from FWC-D, DCW takes the chained delimiter watermarks as the GDs. The experimental results show that the DCW scheme doesn't increase the amount of data transmission. It can also recover the original data.

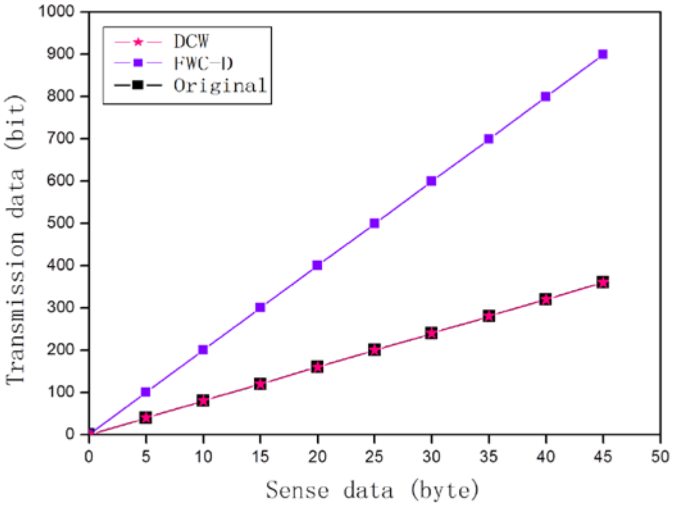


Figure 5: Comparison of the amount of the data transmission among three schemes

4.3 False positive rate comparison

In all the existing single chaining watermark schemes, such as SGW, FWC, when the watermarked data groups cannot be synchronized by the sink node, false positives happen. In the proposed DCW scheme, the chained watermark delimiters can synchronize the both transmission sides in case of the synchronization points are tampered. Fig. 6 shows the comparison of the false positive rate among the SGW scheme, the FWC scheme and the proposed DCW scheme. The false positive rate decreases with the parameter *m*. Meanwhile, compared with the SGW scheme and the FWC scheme, the false positive rate of DCW reduces significantly. That is because the DCW can locate the SP by delimiter watermark.

In total, the false positive performance of the proposed dual-chaining watermark scheme is better than anyone of the previous single chaining watermark schemes.

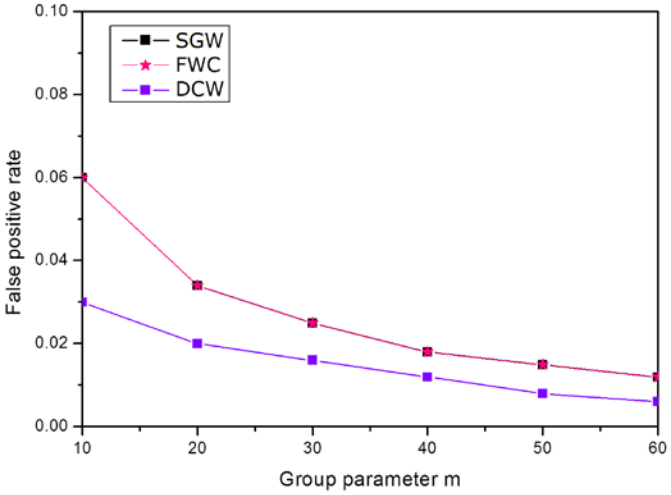


Figure 6: Comparison of the false positive rate

4.4 Anti-attack capability evaluation

In this subsection, we evaluate the anti-attack capability of the proposed DCW scheme. We discuss four most common attacks that can be launched in the wireless smart Internet of Thing application scenario, including packet tampering attack, packet forgery attack, selective forwarding and packet replay attack. We define that an attack is successful if it is not detected by the sink node. In our experiments, we randomly select 5 relay sensor nodes as the malicious nodes, which can launch these four attacks. And each type of attack is evaluated 500 times. The experimental results of anti-attack capability of the proposed DCW scheme are shown in Tab. 3.

Table 3: Anti-attack capability evaluation

Attack types	Experiment Numbers	Detection rate
Packet tampering attack	500	100%
Packet forgery attack	500	100%
Selective forwarding attack	500	100%
Packet replay attack	500	100%

In our proposed DCW scheme, the well-designed dual-chaining watermarks, including the chaining fragile watermark and the pre-defined chained delimiter watermark, are embedded into the data in a dynamic and reversible way. It is difficult for the adversaries to analysis any data information by capturing the data packets without the data grouping secret key. The DCW scheme embeds chained delimiters watermarks which can prevent the adversaries to track and improve the security greatly compared to the single chaining watermarking scheme. According to the experimental results shown in Tab. 3, the prosed

DCW scheme achieved 100% detection rate on all the four types of packet attacks. We can see that the proposed DCW scheme can verify the integrity of the data stream. It can be used to ensure the authenticity of the data.

4.5 Computation overhead and energy consumption analysis

It is difficult to accurately evaluate the energy consumption of the proposed dual-chaining watermark scheme in the real smart meteorological Internet of Things experiment environment. So, we only evaluate energy consumption using the computation overhead analysis according to common practice. Generally, the energy consumed to transmit one bit data over a distance of 100m by radio can execute about 3000 instructions [Kamel and Juma (2011)]. That is, if *1E* represents the energy consumed to execute one instruction, the energy consumption to transmit 1-bit data is 3000E. So, we can see that the energy consumption for data transmission is far greater than processing the sensory data. In DCW, the chained delimiter watermarks and the fragile watermark are both embedded into the data elements instead of being transmitted as additional data. It does not introduce any data transmission overhead. So, the total energy consumption of the perceptual layer of the network is reduced. The network lifetime of the smart Internet of Things system is extended significantly.

5 Conclusion

In this article, a dual-chaining reversible watermarking method for data integrity protection in the perception layer of the IoT system. The proposed reversible scheme can ensure the integrity of the data with free distortion. In addition, DCW scheme takes the chained watermarking delimiters to synchronize the data source node and the sink node. It makes the adversaries difficult to detect the existence of the watermarking and track the data groups. The dual-chaining watermarks can resist various types of attacks such as packet forgery attack, selective forwarding attack, packet replay and tampering attack, and authenticate the integrity of data effectively. Meanwhile, the proposed dual-chaining watermark scheme does not increase the transmission overhead. Experimental results have shown that the DCW has remarkable advantages over the existing single chaining watermark methods not only in terms of data accuracy, but also data security and the lifetime of the wireless sensor network.

Acknowledgement: This work is supported by the Major Program of the National Social Science Fund of China under Grant No. 17ZDA092; by the Electronic Information and Control of Fujian University Engineering Research Center Fund under Grant No. EIC1704; by the National Key R&D Program of China under grant 2018YFB1003205; by the National Natural Science Foundation of China under grant 61173136, U1836208, U1536206, U1836110, 61602253, 61672294; by the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD) fund; by the Collaborative Innovation Center of Atmospheric Environment and Equipment Technology (CICAEET) fund, China.

References

Alattar, A. M. (2004): Reversible watermark using the difference expansion of a generalized integer transform. *IEEE Transaction on Image Processing*, vol. 13, no. 8, pp. 1147-1156.

Chen, X.; Sun, X.; Sun, H.; Zhou, Z.; Zhang, J. (2013): Reversible watermarking method based on asymmetric-histogram shifting of prediction errors. *Journal of Systems and Software*, vol. 86, no. 10, pp. 2620-2626.

Dong, X.; Li, X. (2009): An authentication method for self-nodes based on watermarking in wireless sensor networks. *Proceedings of the 5th International Conference on Wireless Communications, Networking and Mobile Computing*, pp. 1-4.

Feng, J.; Potkonjak, M. (2003): Real-time watermarking techniques for sensor networks. *Proceedings of SPIE 5020, Security and Watermarking of Multimedia Contents V*, pp. 391-402.

Guo, H.; Li, Y.; Jajodia, S. (2007): Chaining watermarks for detecting malicious modifications to streaming data. *Information Sciences*, vol. 177, no. 1, pp. 281-298.

Juma, H.; Kamel, I.; Kaya, L. (2008): Watermarking sensor data for protecting the integrity. *Proceedings of the International Conference on Innovations in Information Technology*, pp. 598-602.

Kamel, I.; Juma, H. (2010): Simplified watermarking scheme for sensor networks. *International Journal of Internet Protocol Technology*, vol. 5, no. 1/2, pp. 101-111.

Kamel, I.; Juma, H. (2011): A lightweight data integrity scheme for sensor networks. *Sensors*, vol. 11, no. 4, pp. 4118-4136.

Liu, X.; Ge, Y.; Zhu, Y.; Wu, D. (2014): A lightweight integrity authentication scheme based on reversible watermark for wireless body area networks. *KSII Transaction on Internet and Information Systems*, vol. 8, no. 12, pp. 4643-4660.

Qiu, Y. (2017): Steganography using reversible texture synthesis based on seeded region growing and LSB. *Computers, Materials & Continua*, vol. 55, no. 1, pp. 151-163.

Shi, X.; Xiao, D. (2013): A reversible watermarking authentication scheme for wireless sensor networks. *Information Sciences*, vol. 240, pp. 173-183.

Wang, B.; Gu, X.; Zhou, A. (2017): E^2S^2: a code dissemination approach to energy efficiency and status surveillance for wireless sensor networks. *Journal of Internet Technology*, vol. 18, no. 4, pp. 877-885.

Wang, B.; Gu, X.; Ma, L.; Yan S. (2017): Temperature error correction based on BP neural network in meteorological WSN. *International Journal of Sensor Networks*, vol. 23, no. 4, pp. 265-278.

Wang, B.; Sun, X.; Ruan Z.; Ren, H. (2011): Multi-mark: multiple watermarking method for privacy data protection in wireless sensor networks. *Information Technology Journal*, vol. 10, pp. 833-840.

Wang, B.; Yan, J.; Li, T., Sun, X; Ma, L. (2015): A packet loss tolerated method for data integrity protection in wireless sensor networks. *International Journal of Multimedia and Ubiquitous Engineering*, vol. 11, no. 7, pp. 1-12.

Wang, B.; Gu, X.; Yan, S. (2018): STCS: a practical solar radiatio based temperature correction scheme in meteorological WSN. *International Journal of Sensor Networks*, vol. 28, no. 1, pp. 22-23.

Xia, Z.; Wang, X.; Sun, X.; Wang, B. (2014): Steganalysis of least significant bit matching using multi-order differences. *Security and Communication Networks*, vol. 7, no. 8, pp. 1283-1291.

Yuan, C.; Xia, Z.; Sun, X. (2017): Coverless image steganography based on SIFT and BOF. *Journal of Internet Technology*, vol. 18, no. 2, pp. 435-442.

Zhang, W.; Liu, Y.; Das, S. K.; De, P. (2008): Secure data aggregation in wireless sensor networks: a watermark based authentication supportive approach. *Pervasive and Mobile Computing*, vol. 4, no. 5, pp. 658-680.

Zhang, Y.; Sun, X.; Wang, B. (2016): Efficient algorithm for K-barrier coverage based on integer linear programming. *China Communications*, vol. 13, no.7, pp. 16-23.