A Comprehensive System Design for Website Secure Protection

Xiaoming Zhang, Cuixia Feng, Junlong Xiao

Department of Computer, College of Information Engineering, Beijing Institute of Petrochemical Technology, Beijing, China
{zhangxiaoming & fengcuixia & xiaojunlong}@bipt.edu.cn

Abstract - The secure monitoring of website is very important for the website information distribution. A kind of comprehensive system is proposed to assure the website operating reliably. It is based on technologies of watermarking and network measuring. The group 5-bit hiding scheme is designed to embed typical invisible characters into the webpages. Meanwhile, the distributed measuring approach is adopted to obtain the statistical value of website visiting delay. The experimental result shows that the system prototype is effective by testing some actual portal websites.

Index Terms - Website security, Comprehensive system, Watermarking technology, Network measuring

1. Introduction

Nowadays, Web pages are used popularly and frequently today because the message in the pages can be fast spreading all over the world in real-time. The security of website always relies on other systems, such as intrusion detection system and intrusion prevention system. Although it spent large of devices and money, the effect is not satisfying. With the rapid development of information hiding technology, some watermarking approaches were studied and adopted successfully in webpage tamper-proof demands. Webpage information hiding is a kind of technology to hide important message into the web pages. Owing to the global distribution of webpage in real time, the secret message can be transmitted and spread quickly without any other channels. Meanwhile, if some watermarks are embedded into the web pages, these pages can be protected by temper-proof technology. Both of the sides become important study contents of webpage information security. There exist several approaches for WIH to aim at the HTML documents [1-5]. Meanwhile, many websites are often attacked by hacker. Serious attack may lead to the website broken down. This problem does not belong to the page changes. However, the website may lose its visiting performance. Researchers often studied the attack of DDoS to the website to form prediction model, or created detection system with large number of data [6-7]. It shows that single solution is difficult to assure the website security.

For these reasons, a kind of secure protection system is proposed with tamp-proof and breakdown-proof features. The watermark is embedded into the appointed webpage to assure its security as the inner protection, while the network measuring approach is designed for website visiting performance from the remote side. The remainder of this paper is organized as follows. In Section II, we state the system design solution. In Section III, we explain the critical technology about the watermarking and network measuring. In Section IV, we conduct experiments using real website operation.

2. Design Solution of the Comprehensive System

The comprehensive system is designed as Fig. 1.

Two kinds of approaches, webpage watermarking and website measuring, are designed independently and in parallel in the system. The system has distributed network management scheme with many measuring sites. All these measuring results are transmitted and stored to the Data Center. Then, through database management and prediction model, the final measuring results about website security are conducted in the Monitoring Center.

For all the multimedia elements in the key web pages, such as image, audio, video, animation and text information, watermark can be embedded easily before these pages are distributed. However, the embedding process and extracting process will consume a lot of time. Cloud platform has strong computing ability and storing space, which can help to solve the problem of watermarking process slowly. After the watermark is extracted through the cloud environment, the watermark will be sent to the Data Center.

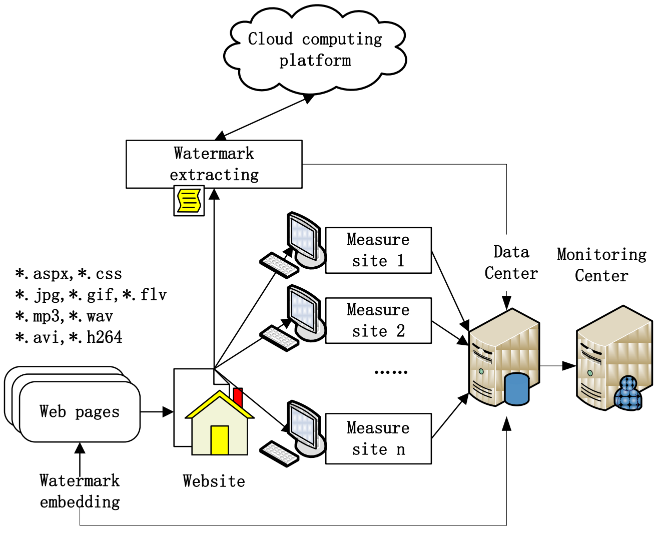


Fig. 1 A comprehensive system for website secure protection

3. Watermarking Technology

A new kind of approach by inserting invisible characters is proposed. Through the bit separation and composition algorithm, the secret message with any character can be embedded into the webpage. Because the inserting information is not bit information but special characters, the hiding capacity increases largely. Next, the hiding idea will be stated firstly. Then, the new algorithm is presented with flow chart and process description. The experiments are carried out for the ordinary portal homepage, while several typical attacks on the hiding algorithm are designed and tested with good effects.

Invisible character (IC): it means the character inserting into the web page but without display sign when page browsing. The invisible character shows special symbol or white space in the page source code.

There are 33 characters in the ASCII table for IC with ASCII coding values of 0-32. The codes of 1-31 show special symbols. While the code 0 is null, and the ASCII code 32 is white space. All these characters are chosen as invisible characters to form a set S_p:

$$S_p = \{s_i | i = [0-32]\}. \tag{1}$$

In order to hide any character from ASCII table, all the secret data have to be converted to the IC. Considering the IC value from 0 to 31, it only need the lower 5 bits in an 8-bit data to represent all the IC value. Therefore, a kind of group scheme is created for the hiding data, as shown in Fig. 2. The initial data are grouped as 5 bits, and then each group is added with 3 bits of 0 to its higher position to form a new 8-bit data. For example, one group with 5 bits of 10110 is chosen for IC data creation. After three bits of 0 are inserted to the head of group, the IC data will have value of 22. Its ASCII character belongs to an IC of device control.

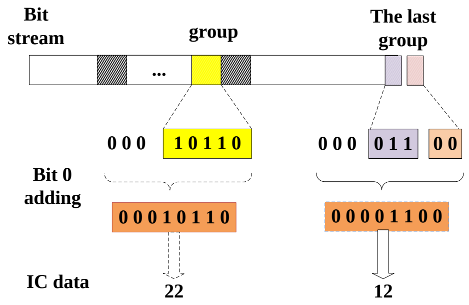


Fig. 2 Grouping 5-bit scheme for IC creation

For the webpage hiding application, the checksum can be used to increase the reliability of secret transmission, and its computation solution is different from that in OSI model.

Definition 2 Hiding Checksum (HC): In the secret hiding process, the hiding checksum is applied to check two kinds of data together. One is the length of secret, and the other is the data of secret.

Definition 3 Hiding Data (HD): It refers to the data which is actually embedded into the webpage. In order to increase the reliability of web information transmission, the hiding data is composed of three kinds of data, that is, the length of secret, the HC and the data of secret, as shown in Fig.3.

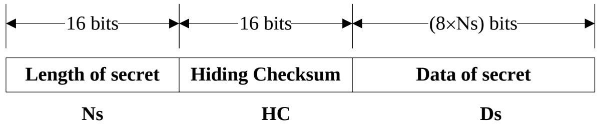


Fig. 3 Composition of hiding data composition in webpage

The length is used to decide the actual number of secret data, while the checksum is for checking the network error and the data consistence. Both of the length and the checksum are 16 bits. Therefore, the final length of hiding data will be as following:

$$L_{HD} = 32 + 8N_S. \tag{2}$$

The process is as shown in Fig. 4.

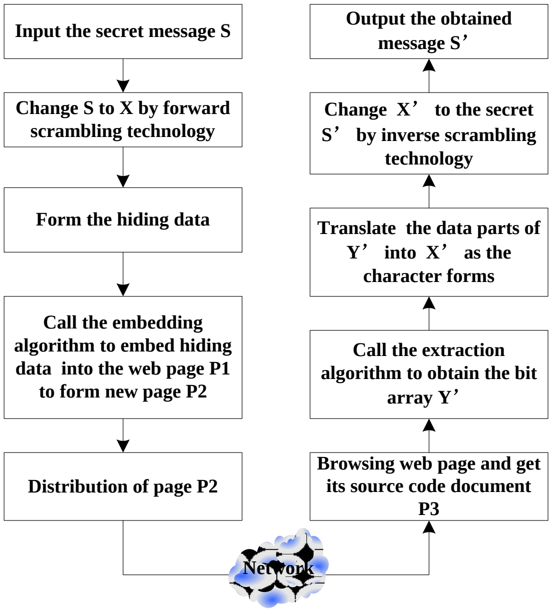

Input the secret message S
Change S to X by forward scrambling technology
Form the hiding data
Call the embedding algorithm to embed hiding data into the web page P1 to form new page P2
Distribution of page P2
Network
Browsing web page and get its source code document P3
Call the extraction algorithm to obtain the bit array Y'
Translate the data parts of Y' into X' as the character forms
Change X' to the secret S' by inverse scrambling technology
Output the obtained message S'

Fig.4 IC-based hiding model

After the hiding data is embedded into the appointed webpage, the new created webpage P2 is distributed through the web environment. Then, the receiver can obtain the important message by browsing the website with a series of extraction algorithm. The browsed webpage is often different from that existed in the web server.

In the embedding process, there are three key technologies which are forward scrambling processing, hiding data forming and the embedding algorithm. As for the information extraction process, the inverse operation with extraction algorithm, character translation and inverse scrambling technology are carried out. Next, the technique implementation and the hiding algorithms are proposed in detail below.

4. Network Measure Technology

Network measuring is often effective for the website operation state detection. Because the network topology may cause the measuring error, single measuring value is often incorrect. This can be solved with distributed measuring methods to get statistical data. With a number of measuring site, the data calculation for the website delay can be more correct. Large number of data helps for the prediction of website breakdown. It adopts protocol ICMP to detection the target website online by suitable interval time.

The flowchart of single measuring process is shown in Fig.5.

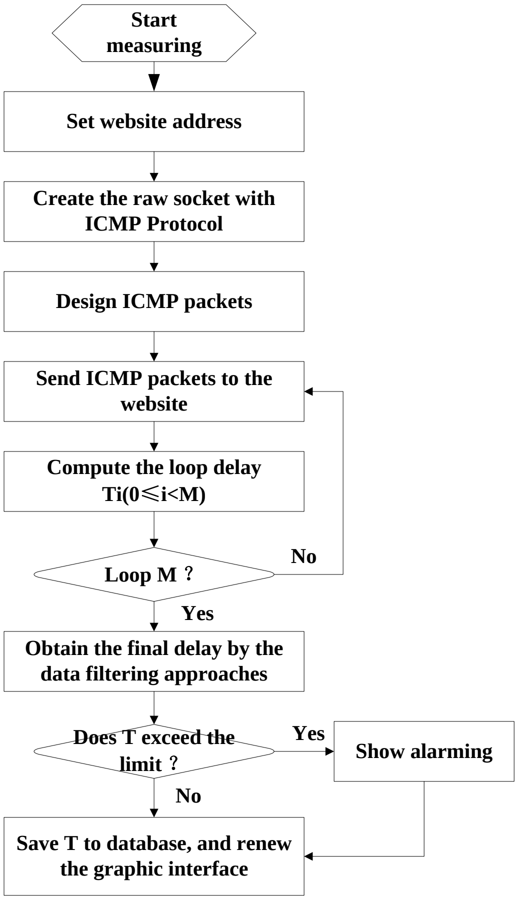

Start measuring
Set website address
Create the raw socket with ICMP Protocol
Design ICMP packets
Send ICMP packets to the website
Compute the loop delay Ti(0≤i<M)
Loop M ?
No
Yes
Obtain the final delay by the data filtering approaches
Does T exceed the limit ?
Yes
Show alarming
No
Save T to database, and renew the graphic interface

Fig 5 Network measuring flowchart for single site

The data from each measuring site are transferred in time to the database in the Data Center. These data are filtered, and composed to form the computational input vector of network measuring model. A series of data with computing window are used to estimate the state of website. Basically, the website can be classified to normal state and abnormal state, along with colours of green and yellow independently.

5. Experimental Analysis

The system is composed of 2 servers, 5 measuring sites with Personnel computers. The software prototype is programmed with C# language under VS2010 development platform, along with SQLSERSER 2008 as the database management system. There are two important modules as webpage watermark hiding and website performance measuring.

A. *Performance calculation*

Firstly, the character hiding effects is required to calculate.

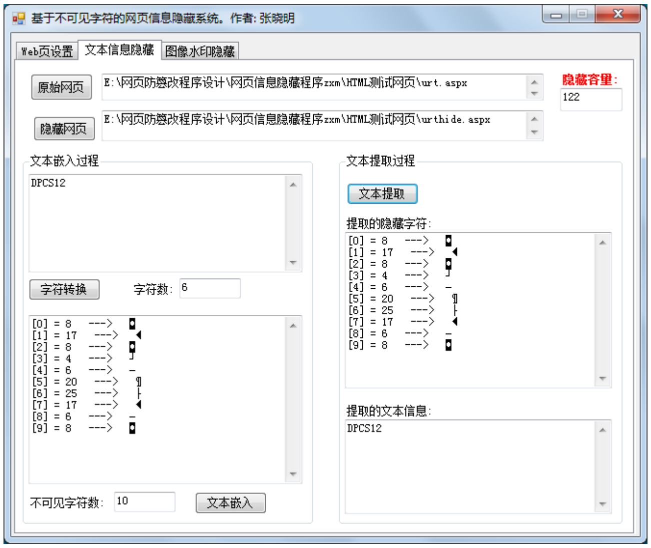
Definition 3 Character Error Rate (CER): Suppose the secret S with ASCII code $X = x_1 x_2 ... x_m$ and the extraction message $X^{'} = x_1^{'} x_2^{'} ... x_m^{'}$, the character error rate is defined as the hiding error:

$$CER = \frac{\sum_{i=1}^{m} \left| x_i^{'} - x_i \right|}{\sum_{i=1}^{m} x_i} \times 100\% \qquad (3)$$

B. *Webpage hiding experiments*

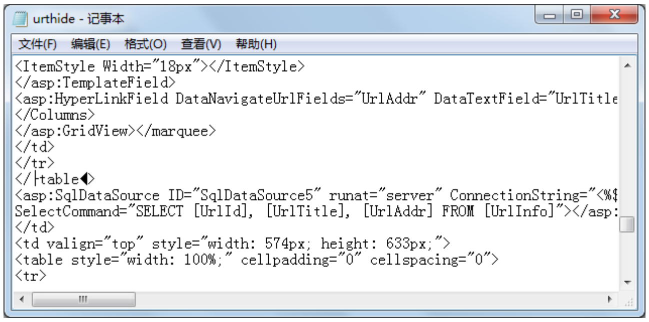
A testing website with mast page, named as "urt.aspx", is used as watermark hiding process. The operating interface is shown in Fig. 6.

It shows watermark "DPCS12" of the dynamic webpage source code in Fig. 7. We can identify the invisible character embedded in the two sides of the end label </table>.



```
E:\网页防篡改程序设计\网页信息隐藏程序zxm\HTML测试网页\urt.aspx
E:\网页防篡改程序设计\网页信息隐藏程序zxm\HTML测试网页\urthide.aspx
隐藏容量: 122

DPCS12
字符数: 6

[0] = 8    --->  ◘
[1] = 17    --->  ◄
[2] = 8    --->  ◘
[3] = 4    --->  ┘
[4] = 6    --->  –
[5] = 20    --->  ¶
[6] = 25    --->  ├
[7] = 17    --->  ◄
[8] = 6    --->  –
[9] = 8    --->  ◘
不可见字符数: 10

提取的隐藏字符:
[0] = 8    --->  ◘
[1] = 17    --->  ◄
[2] = 8    --->  ◘
[3] = 4    --->  ┘
[4] = 6    --->  –
[5] = 20    --->  ¶
[6] = 25    --->  ├
[7] = 17    --->  ◄
[8] = 6    --->  –
[9] = 8    --->  ◘

提取的文本信息:
DPCS12
```

Fig. 6 Hiding watermark "DPCS12" into the webpage



```
<ItemStyle Width="18px"></ItemStyle>
</asp:TemplateField>
<asp:HyperLinkField DataNavigateUrlFields="UrlAddr" DataTextField="UrlTitle
</Columns>
</asp:GridView></marquee>
</td>
</tr>
</├table◄>
<asp:SqlDataSource ID="SqlDataSource5" runat="server" ConnectionString="<%$
SelectCommand="SELECT [UrlId], [UrlTitle], [UrlAddr] FROM [UrlInfo]"></asp:
</td>
<td valign="top" style="width: 574px; height: 633px;">
<table style="width: 100%;" cellpadding="0" cellspacing="0">
<tr>
```

Fig. 7 Source code analysis after watermark embedded

C. *Network measure experiment and analysis*

Firstly, the network performance is measured by simulation with OPNET. Five cities in China are chosen with different measuring distances to the web site. The http response time is shown in Fig. 8. The http traffic received is analysed as shown in Fig.9. It indicates that the measuring distance and the traffic change can affect the web visiting performance.

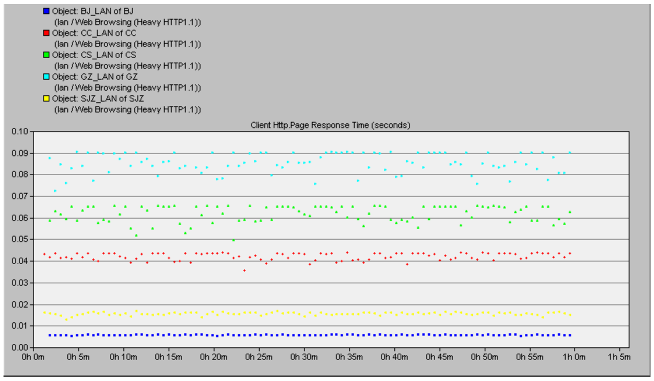


Fig.8 HTTP page response time (seconds) under five cities measuring

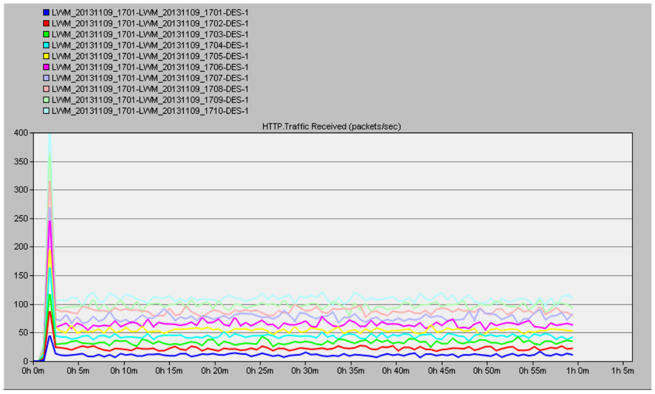


Fig.9 Traffic received (packets/second) under different client number

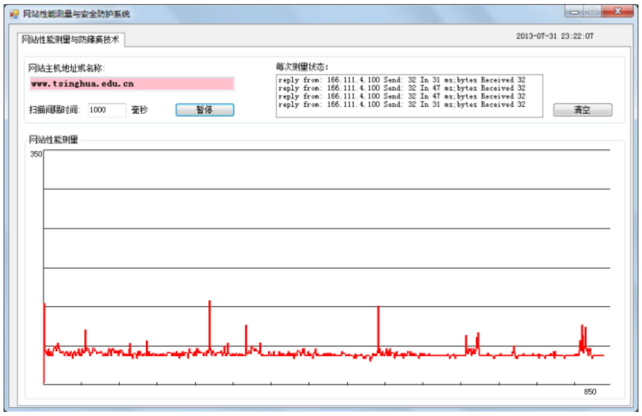
Next, the single site of network measuring is carried out, as shown in Fig. 10 and Fig. 11.

The measuring data shows small changes when measuring www.tsinghua.edu.cn, the campus Website of Tsinghua University in China. Most of the data stays at 50 million seconds, while seldom data reaches to 100.

However, the measuring result on Portal Website of www.sohu.com shows different state. There exist some of big data up to maximum limit. It indicates that the website may be affected with suddenly large network crowd.

6. Conclusion

The system is integrated with two typical technologies of watermarking and network measuring. The system prototype can be effective to assure the website secure under normal operation. Next, the data analysis based on the large number of measuring data should be carried out. And the watermark hiding process will be tested through the cloud platform to get much more effective performance for the website security.



2013-07-31 23:22:07

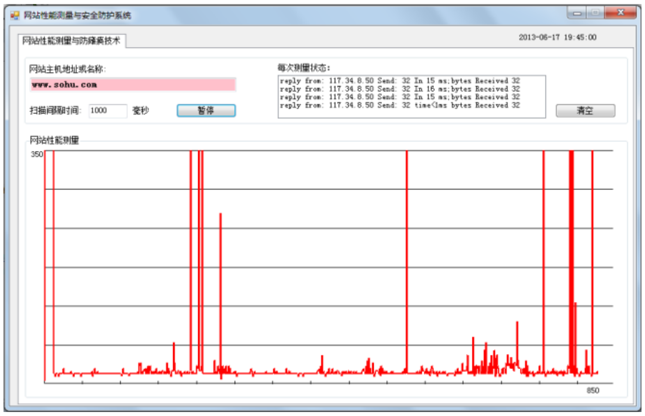
网站主机地址或名称: www.tsinghua.edu.cn

扫描间隔时间: 1000 毫秒

每次测量状态:

```
reply from: 166.111.4.100 Send: 32 In 31 ms;bytes Received 32
reply from: 166.111.4.100 Send: 32 In 47 ms;bytes Received 32
reply from: 166.111.4.100 Send: 32 In 47 ms;bytes Received 32
reply from: 166.111.4.100 Send: 32 In 31 ms;bytes Received 32
```

Fig.10 Measuring effects of website www.tsinghua.edu.cn



2013-06-17 19:45:00

网站主机地址或名称: www.sohu.com

扫描间隔时间: 1000 毫秒

每次测量状态:

```
reply from: 117.34.8.50 Send: 32 In 15 ms;bytes Received 32
reply from: 117.34.8.50 Send: 32 In 16 ms;bytes Received 32
reply from: 117.34.8.50 Send: 32 In 15 ms;bytes Received 32
reply from: 117.34.8.50 Send: 32 time<1ms bytes Received 32
```

Fig.11 Measuring effects of website www.sohu.com

References

[1] Sun, XM., Huang H., Wang B. An Algorithm of Webpage Information Hiding based on Equal Tag. Journal of Computer Research and Development 2007, 44(5):756-760

[2] Zhang XY, Zhang XM. An Algorithm of Webpage Information Hiding Based on the Property of Table. Journal of Beijing Institute of Petro-chemical Technology (in Chinese).2009, 17(1):43-47.

[3] Qin CY, Zhang XM. Zhao GQ. Design of Information Hiding Algorithm for Complicated Webpage Tables. Microelectronics & Computer (in Chinese). 2011, 28(8):132-135.

[4] Zhang XY, Zhang XM. Information Hiding Algorithm Based on Flash Animation. Computer Engineering (in Chinese). 2010,36(1):181-183

[5] Zhang XM, Zhao GQ, Niu PF. A novel approach of secret hiding in webpage by bit grouping technology. Journal of Software, 2012,4(11)

[6] Yi Xie and Shun-Zheng Yu, Monitoring the Application-Layer DDoS Attacks for Popular Websites. IEEE/ACM TRANSACTIONS ON NETWORKING, VOL. 17, NO. 1, FEBRUARY 2009

[7] Wen S., Jia WJ, Zhou W.. CALD: Surviving Various Application-Layer DDoS Attacks That Mimic Flash Crowd. 2010 Fourth International Conference on Network and System Security.