

# 基于不可见字符的主副式网页信息隐藏算法

翁 超 周 良 丁秋林

(南京航空航天大学计算机科学与技术学院 南京 210016)

**摘 要** 目前国内外网页信息隐藏算法大多侧重于研究如何提高隐藏效率和扩充隐藏容量,而忽视了信息隐藏算法的安全性。为此,提出了一种基于不可见字符的主副式网页信息隐藏(PSWIH)算法,即以 ASCII 编码下的不可见字符为主方式实现信息隐藏,并以改变标签大小写、嵌入与待隐藏信息不相关的不可见字符、改变标签属性对顺序为副方式实现信息检测干扰。实验结果表明,相对于基于标签大小写、基于属性对顺序、基于不可见字符方法,PSWIH 算法具有更好的不可感知性、抗提取性及鲁棒性。

**关键词** 信息隐藏,不可见字符,主副式,网页

**中图法分类号** TP391.1 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2016.7.024

## Primary and Secondary Webpage Information Hiding Algorithm Based on Invisible Characters

WENG Chao ZHOU Liang DING Qiu-lin

(College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China)

**Abstract** Current domestic and overseas webpage information hiding algorithms mainly focus on how to improve the hiding efficiency and expand the hidden capacity, but they neglect the security of information hiding algorithm. For this reason, the paper put forward a primary and secondary webpage information hiding algorithm based on invisible characters, in which information hiding is realized by the invisible characters based on ASCII coding as the primary method, and interference of information detection is realized by changing the letters of tag, embedding invisible characters unrelated to hidden information, changing the sequence of tag attributes as the secondary method. The experimental result shows that PSWIH algorithm is better than the method based on changing the letters of tag, changing the sequence of tag attributes, embedding invisible characters in non-perceptibility, extractability and robustness.

**Keywords** Information hiding, Invisible characters, Primary and secondary, Webpage

### 1 引言

随着网络传输信息量的增长与黑客技术的日新月异,网络安全问题显得越来越重要,信息隐藏技术成为了许多学者研究的热点。同时,相对于文本、图像等其它传输载体,网页的使得更加广泛普遍。网页信息隐藏技术作为多媒体信息隐藏技术的一个重要分支,其研究得到越来越多学者的关注。目前基于 HTML 标签的网页信息隐藏方法主要包括基于标签大小写、属性对顺序及不可见字符 3 类。

Singh P 等人<sup>[1]</sup>依据 HTML 标签中字母不区分大小写的特点,在不改变显示效果的前提下,对 HTML 标签字母进行大小写变换。通常情况下,大写字母代表比特 1,小写字母代表比特 0。Zhao 等<sup>[2]</sup>在使用主成分分析法(Principal Component Analysis, PCA)进行水印嵌入时,也采用了此种方法。这种方式的优点是操作简便并且隐藏容量相对较大,但其缺点也是显而易见的,因为大小写字母分别仅能表示一种比特,所以在查看 HTML 源码后,稍作分析便可破解。另外,倘若最终需要提取的比特流中一位字母出现大小写篡改,极有可

能导致所提取的信息与原信息不一致。

Garg M<sup>[3]</sup>依据任意调节 HTML 标签中属性对顺序而不影响显示效果的特点,将标签属性分为第一属性和第二属性,隐藏过后,如果第一属性在第二属性之前,则隐藏比特 1,否则隐藏比特 0。此种方法抗检测性较高,但 HTML 标签属性个数有限,一旦获得源网页,对比隐藏后的网页并进行分析,破解隐藏信息就相对容易。

Katzenbeisser S 等人<sup>[4]</sup>最先提出了一种空格标签编码(Space-Tab Coding, STC),即在标签中添加空格 Space 或 Tab 进行信息隐藏。后来,许多学者及专家提出了基于 ASCII 和 Unicode 编码下的不可见字符,这些字符主要包括控制字符及转义字符。Lee IS 等<sup>[5]</sup>采用了 space、&#x20;、&#32;等 8 种不可见字符,并规定每种不可见字符所对应的 3 位比特,将其嵌入到网页中,但此种方法未能充分利用网页标签的独特之处。为此,Zhang 等<sup>[6]</sup>进行了改进,选取 ASCII 编码中值为 0 至 32 的不可见字符嵌入到 HTML 标签中,以</table>标签为例,存在</table>、</table >、</table > 3 种嵌入形式,不仅发挥出了 HTML 标签的优势,同时隐藏量也从 3 位比特

到稿日期:2015-06-03 返修日期:2015-08-23 本文受江苏省产学研联合创新基金项目(SBY201320423)资助。

翁 超(1990-),男,硕士生,主要研究方向为信息系统及集成、信息安全等,E-mail:962297911@qq.com;周 良(1966-),男,博士,副教授,硕士生导师,主要研究方向为信息系统、知识工程;丁秋林(1936-),男,博士,教授,博士生导师,主要研究方向为信息系统、企业信息化。

编码扩充到 5 位比特编码,隐蔽性较强,方法较为成熟。

对于网页信息隐藏,应尽最大可能从不可感知性、完整性、高容量性、鲁棒性等多方面实现隐藏<sup>[7,8]</sup>。不同因素下 3 类方法的性能参数如表 1 所列。

表 1 不同因素下 3 类方法的性能参数

因素 \ 方法	不可见字符	字母大小写	属性对顺序
不可感知性	强	弱	强
完整性	中	中	强
高容量性	强	中	弱
鲁棒性	中	强	中

此外, Yang<sup>[9]</sup> 根据 HTML 标签属性多种等价写法的特点,提出了多种等价属性对替换的网页信息隐藏方法,采用对标签中属性使用单引号、双引号以及两者均不使用的方进行信息隐藏。与 Yang 算法类似,其他学者同期还提出了基于省略某些 HTML 结束标签、标签属性值等价拼写方式等方法。Huang 等<sup>[10]</sup>进一步从公式上对等价替换这种隐藏方法进行了定量分析和深度研究。此种方式简便易行,但等价数量有限,查看 HTML 源码后,很容易破解,其安全性较低。

目前网页信息隐藏方法均存在对比 HTML 源码和 HTML 隐藏代码后易破解、安全性不高的缺点。基于此,本文设计了一种基于不可见字符的主副式网页信息隐藏 (Primary and Secondary Webpage Information Hiding Based on Invisible Characters, PSWIH) 算法,借鉴不可见字符强隐蔽性、标签字母大小写简便易行以及属性对顺序抗检测性的特点,并结合 HTML 4.0 若干特性进行设计,以提高网页信息隐藏的不可感知性、抗检测性和鲁棒性。

## 2 PSWIH 算法设计

### 2.1 算法思想

本文提出的 PSWIH 算法在综合考虑了目前主要的 3 类方法各自优点的基础上,采取主/副方式实现网页信息的嵌入和提取,其中主方式是基于 ASCII 编码下的不可见字符进行信息隐藏,副方式是基于改变 HTML 标签字母大小写、嵌入与待隐藏信息不相关的不可见字符以及改变 HTML 标签属性对顺序进行信息检测干扰。网页信息隐藏主要包括信息嵌入和信息提取两大过程,将 PSWIH 算法应用于网页信息隐藏的嵌入和提取过程中,层次清晰、分工明确,实现网页信息隐藏的同时提高了信息隐藏的不可感知性、抗检测性和鲁棒性。

主方式采用不可见字符进行信息隐藏,目前有空格标签编码下的不可见字符、ASCII 编码下的不可见字符、Unicode 编码下的不可见字符等,通过对比其相关稳定性、可存储容量以及未来编码发展趋势等若干因素,本文采用 ASCII 编码下的不可见字符作为主方式。ASCII 编码下序号为 0 至 32 以及 127 的控制字符和功能字符均属于不可见字符,根据 PSWIH 算法设计需要,本文仅使用前 32 个序号的字符。

副方式采用改变标签字母大小写、嵌入与待隐藏信息不相关的不可见字符以及改变标签属性对顺序三者相结合的方式,使得入侵者即使获得源网页也很难通过对比 HTML 代码进行信息提取,从而提高了网页信息隐藏的安全性。

本文设计的 PSWIH 算法的总体流程如图 1 所示。

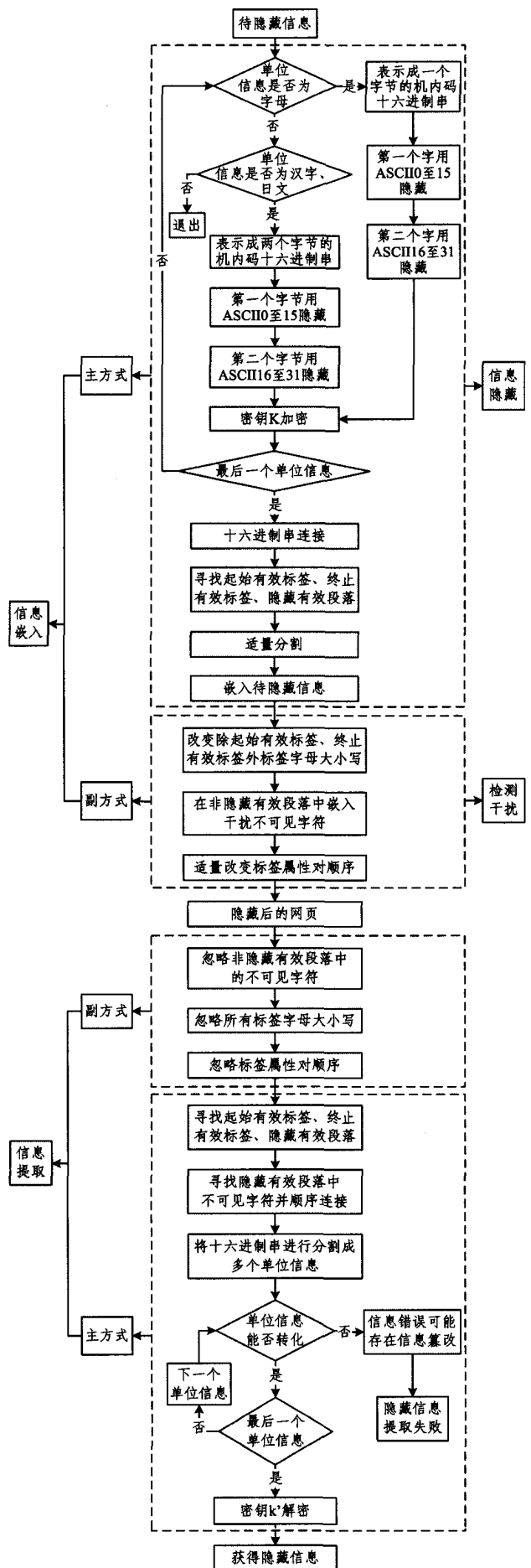


图 1 PSWIH 算法总体流程

## 2.2 PSWIH 嵌入算法

在设计 PSWIH 嵌入算法之前,深入研究了 ASCII 编码方式及 HTML 标签特点,并设计了单位信息,即待隐藏信息中的每一个英文字母、汉字或日文等。例如:待隐藏信息为“String 表示字符串”,字母 S、t 以及汉字“表”、“示”等均称为单位信息。同时还设计了起始有效标签、终止有效标签及隐藏有效段落。

起始有效标签:HTML 标签中至少有两个字母,且第一个字母大写,最后一个字母小写,其余字母可大写可小写,称为有效起始标签,其正则表达式如式(1)所示:

$$\langle \wedge (/)?[A-Z][a-zA-Z]\{0,\}[a-z]\$ \rangle \quad (1)$$

终止有效标签:HTML 标签中至少有两个字母,且第一个字母小写,最后一个字母大写,其余字母可大写可小写,称为有效终止标签,其正则表达式如式(2)所示:

$$\langle \wedge (/)?[a-z][a-zA-Z]\{0,\}[A-Z]\$ \rangle \quad (2)$$

隐藏有效段落:处于起始有效标签和终止有效标签之间的段落称为隐藏有效段落。

PSWIH 嵌入算法将待隐藏信息以 ASCII 编码方式下的不可见字符形式嵌入到网页中,并以改变标签大小写、嵌入与待隐藏信息不相关的不可见字符及改变便签属性对顺序进行检测干扰。针对 PSWIH 嵌入算法中主方式信息隐藏和副方式属性对干扰,分别设计了隐藏约束和属性对约束两条规则。

### 规则 1 隐藏约束

字母式主隐藏方式约束:英文字母一般以 ASCII 码表示,占一个字节。若待隐藏信息为英文字母,可将英文字母表示成一个字节的机内码十六进制串,并对第一个字采用 0 至 15 的 ASCII 不可见字符对应的十六进制串隐藏;第二个字采用 16 至 31 的 ASCII 不可见字符对应的十六进制串隐藏。为了方便表示,用 Hidden()函数表示隐藏后的机内码十六进制串,用 Convert()函数表示隐藏所对应的不可见字符。例如字母 a,转成机内码十六进制串为 61,则 Hidden(a)=0611,Convert(a)=ACK,DC1,即在 HTML 标签中隐藏 ACK 和 DC1 不可见字符便可隐藏字母 a。

汉字式主隐藏方式约束:汉字、日文等一般以 ASCII 码表示,占两个字节。若待隐藏信息为汉字、日文等,可将其表示成两个字节的机内码十六进制串,并对第一个字节采用 0 至 15 的 ASCII 不可见字符对应的十六进制串隐藏;第二个字节采用 16 至 31 的 ASCII 不可见字符对应的十六进制串隐藏。例如汉字“信”,转成机内码十六进制串为 D0C5,则 Hidden(信)=0D001C15,Convert(信)=CR,NUL,FS,NAK,即在 HTML 标签中隐藏 CR,NUL,FS,NAK 不可见字符便可隐藏汉字“信”。

### 规则 2 属性对约束

基于改变 HTML 标签属性对顺序而不改变网页显示的特性,隐藏信息时可将标签属性对任意排列。由于改变标签属性对顺序作为副方式仅起到干扰作用,不隐藏信息,因此这里不需要规定传统改变标签属性对顺序方法中的主属性和从属性,也不需要利用数据库记录原始标签属性对顺序。对于一个 HTML 标签  $Tag(a_1, a_2, \dots, a_n)$ ,其具有  $n$  个属性,总共有  $n!$  种排列方式,其中任意两种排列方式均等价,可用式(3)表示:

$$Tag(a_1, a_2, \dots, a_n) \Leftrightarrow Tag(\overline{a_1}, \overline{a_2}, \dots, \overline{a_n}) \quad (3)$$

其中,  $Tag(\overline{a_1}, \overline{a_2}, \dots, \overline{a_n})$  是  $n!$  种排列方式中一种不同于  $Tag$

$(a_1, a_2, \dots, a_n)$  的排列。

基于以上所描述的规则,提出了 PSWIH 嵌入算法。算法主要从信息隐藏和检测干扰两方面进行设计,具体如算法 1 所述。

### 算法 1 PSWIH 嵌入算法

输入:

①容量充足的网页载体 H, H 由起始有效标签 S、多段隐藏有效段落  $C_i$ 、终止有效标签 E 以及其他部分 O 组成。

②待隐藏信息 I, I 由多个单位信息  $I_j (1 \leq j \leq n, j \in Z)$  组成。

③加密密钥 k。

输出:隐藏信息后的网页 H'。

说明:为了方便描述,规定 Hex()为转成字节数的函数。

步骤:

①获得 I 中每个单位信息  $I_j$  机内码的十六进制串:  $N = \text{Hex}(I_j) (1 \leq j \leq n, j \in Z)$ 。

②判断单位信息类型并选择隐藏方式:根据规则 1,如果  $N=1$ ,采用字母式主隐藏方式约束并用加密密钥 k 加密;如果  $N=2$ ,采用汉字式主隐藏方式约束并用加密密钥 k 加密;否则,待隐藏信息不适用于 PSWIH 算法,提示并退出。

③将所有隐藏后的十六进制串连接:  $I' = \text{Convert}(I_1) + \text{Convert}(I_2) + \dots + \text{Convert}(I_n)$ ,  $I'$  为所有隐藏信息的十六进制串。

④依据隐藏有效段落个数,对  $I'$  进行分割:  $I' = T_1 + T_2 + T_3 + \dots + T_n$ , 其中  $n$  为隐藏有效段落个数。

⑤将分割后的十六进制串依顺序嵌入到对应的隐藏有效段落中:  $C_1 \supseteq T_1, C_2 \supseteq T_2, \dots, C_i \supseteq T_i, \dots, C_n \supseteq T_n$ 。此时,完成了 PSWIH 算法主方式部分,实现了信息隐藏。

⑥修改对除起始有效标签、终止有效标签外的其他标签字母大小写。

⑦在 O 中嵌入适量与隐藏信息不相关的不可见字符。

⑧根据规则 2 最后调节整个网页载体 H 的 HTML 标签属性顺序,完成后得到 H'。此时,完成了 PSWIH 算法副方式部分,实现了检测干扰。

## 2.3 PSWIH 提取算法

PSWIH 提取算法是嵌入算法的逆过程,首先需要忽略副方式中的 3 种检测干扰;其次利用起始有效标签、终止有效标签迅速查找出隐藏在有效隐藏段落中的不可见字符;最后连接所有查找出的不可见字符并进行转换,获得隐藏在网页中的信息。为此,设计了转化约束规则进行信息转换,采用起始有效标签和终止有效标签约束及范围约束规则进行信息查找。

### 规则 3 转换约束

基于规则 1 隐藏约束,可推知隐藏后有效的十六进制串仅有两种形式,如式(4)所示。

$$\begin{cases} \text{字母: } 0X1X \\ \text{汉字、日文等: } 0X0X1X1X \end{cases} \quad (4)$$

其中, X 为十六进制,  $0 \leq X \leq F$

### 规则 4 起始有效标签和终止有效标签约束

起始有效标签和终止有效标签必须成对出现,不可嵌套、缺失、颠倒。以下采用 S 表示起始有效标签, E 表示终止有效标签,  $C_i (i=1, 2, \dots, n)$  表示隐藏有效段落。

正确形式:  $SC_1 ESC_2 ESC_3 E \dots$

错误嵌套:  $SC_1 SC_2 EC_3 E, SSC_1 SEEE, SESC_1 SEE \dots$

错误丢失:  $SC_1 SC_2 E, SC_1 EC_2 E \dots$

错误颠倒:  $EC_1 S, SC_1 C_2 E \dots$

### 规则 5 范围约束

隐藏的信息仅以 ASCII 编码下的不可见字符形式嵌入

到隐藏有效段落中,而非隐藏有效段落中不可见字符不属于待隐藏的信息范围,它起到干扰入侵者、保护隐藏信息的作用。以下采用  $I$  表示待隐藏的信息,  $T$  表示信息转换函数,其可用式(5)~式(7)表示。

$$T(C_i) \subseteq I \quad (5)$$

$$T(\overline{C_i}) \not\subseteq I \quad (6)$$

$$\sum_{i=1}^n T(C_i) = I \quad (7)$$

其中,  $n$  表示所有隐藏有效段落个数。

### 算法2 PSWIH 提取算法

输入:待提取网页  $H'$ , 解密密钥  $k'$ 。

输出:隐藏信息  $I$ 。

步骤:

- ①忽略非隐藏有效段落中所有不可见字符。
- ②忽略除起始有效标签和终止有效标签外的所有标签字母的大小写。
- ③忽略所有标签的属性对顺序。
- ④从隐藏有效段落中提取信息:对  $H'$  从头开始依次按照起始有效标签、有效隐藏段落、终止有效标签的顺序和规则4查找S和E对,并取得S和E之间的  $C_i$ ,根据规则4查找直至  $H'$  结尾处停止。
- ⑤连接  $C_i$  并取得所有隐藏信息的十六进制串  $I'$ :  $I' = C_1 + C_2 + \dots + C_n$ 。
- ⑥信息转化:根据规则3和规则5,将  $I'$  进行分割并按顺序进行信息转化,若转化成功,最后用解密密钥  $k'$  解密得到秘密信息  $I$ ; 否则,提示提取失败,存在信息被篡改的可能。

## 3 实验与分析

实验环境如下: Pentium (R) Dual-Core CPU E5300 2.60GHz & 2.60GHz, 内存 6GB, 操作系统 Microsoft Windows7 Ultimate, 32 位操作系统。使用 C# 语言在 HTML 4.0 及 Microsoft Visual Studio 2010 开发环境下实现算法,并使用 UltraEdit 工具查看不可见字符。

基于研究所某系统的网页实现应用 PSWIH 嵌入算法和提取算法,图2和图3是在 Microsoft Visual Studio 2010 开发环境下嵌入不可见字符的部分 HTML 代码段,除了换行外,其他尚无明显差别。由 HTML 4.0 相关知识可知,信息嵌入前和嵌入后的显示效果从人类视觉上来讲没有任何差别。图4是嵌入后的代码段在 UltraEdit 中的显示,其中  $\langle \text{Tr} \rangle$ 、 $\langle \text{Td} \rangle$  为起始有效标签,  $\langle / \text{td} \rangle$ 、 $\langle / \text{asp:Label} \rangle$  为终止有效标签,在  $\langle \text{Tr} \rangle \langle / \text{td} \rangle$  间和  $\langle \text{Td} \rangle \langle / \text{asp:Label} \rangle$  间用横线标记的为有效不可见字符,其余用横线标记的为干扰不可见字符,同时 HTML 标签相关属性对顺序进行了改变。图5是按网页顺序隐藏的有效段落,其中中间部分用横线标记的为有效不可见字符的十六进制串,即已嵌入隐藏信息的网页内容为 0612-0611-061B-0C0E1C14-0B0C1F1E-0D021D11-0C0A1D15-0B051B1D,根据提取算法得到的结果为 62-61-6B-CEC4-BCFE-D2D1-CAD5-B5BD,通过转化解密后获取的隐藏信息为“bak文件已收到”。

```

<tr>
  <td align="center" class="style8" style="padding-top: 15px;
border-top: 1px solid # 0099cc;">
  <asp:Label ID="Label10" runat="server" Text="数据导入"
Font-Size="medium" CssClass="elb"></asp:Label>
</td>
<td align="center" class="style8" style="padding-top: 15px;
border-top: 1px solid # 0099cc;">
  <table>

```

```

<tr>
  <td>
    <asp:Label ID="Label4" runat="server" Text="上传文件;"
Font-Size="small" CssClass="elb"></asp:Label>
    <asp:FileUpload ID="lookingurl1" runat="server"
Css-Class="input_03" text="上传文件" Width="190px"
Height="16px">
    <asp:ImageButton ImageAlign="AbsMiddle" ID="in_
btn_produce602" runat="server" ImageUrl="~/images/
btn/btn_input.jpg" Height="20px" Width="81px" />
  </td>
</tr>
</table>
</td>
</tr>
<tr>
  <td align="center" class="style8" style="padding-top: 5px;
border-top: 1px solid # 0099cc;">
  <asp:Label ID="Label10" runat="server" Font-Size="medium"
Text="数据导入" CssClass="elb"></asp:Label>
</td>
<td align="center" class="style8" style="padding-top: 15px;
border-top: 1px solid # 0099cc;">
  <table>
    <tr>
      <td>
        <asp:Label ID="Label4" runat="server" Text="上传文件;"
CssClass="elb" Font-Size="small"></asp:Label>
        <asp:FileUpload ID="lookingurl1" runat="server"
Width="190px" text="上传文件" CssClass="input_03"
Height="16px">
        <asp:ImageButton Width="81px" ID="in_btn_
produce602" runat="server" ImageUrl="~/images/
btn/btn_input.jpg" Height="20px" ImageAlign="3AbsMiddle">
      </td>
    </tr>
  </table>
</td>
</tr>

```

图2 原始网页代码

图3 隐藏后网页代码

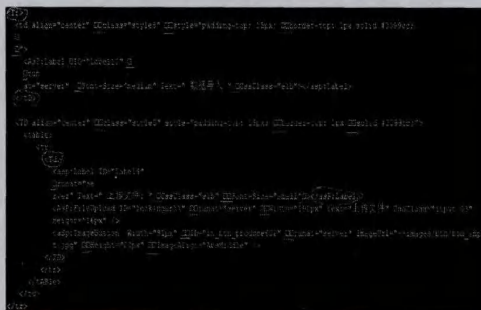


图4 嵌入后代码段在 UltraEdit 中的显示



图5 隐藏有效段落

对网页信息隐藏的安全性,本文主要从抵抗攻击的角度<sup>[16]</sup>对比PSWIH算法与3类方法在不可感知性、抗提取性以及鲁棒性方面的性能。

网页不可感知性是指待隐藏信息嵌入到网页后不会使之产生一定的可感知失真,即人类视觉和机器检测均无法感知、检测。本文采用嵌入率和检测率的关系变化来间接体现网页的不可感知性。有效信息隐藏量占隐藏后整个网页信息量的比例称为嵌入率。由于改变标签大小写和属性对顺序不会增加网页信息量,因此仅需考虑不可见字符隐藏的信息量。用 $F_1$ 表示隐藏有效段落中不可见字符的信息量,用 $F_2$ 表示非隐藏有效段落中不可见字符的信息量,用 $F_w$ 表示嵌入前网页的信息量,用 $\alpha$ 表示嵌入率,计算公式如式(8)所示:

$$\alpha = \frac{F_1}{F_1 + F_2 + F_w} \quad (8)$$

借鉴文献[11]、文献[12]及文献[13]提出的检测方法,对比本文算法与3类方法。随着嵌入率的增长,各方法检测率的变化情况如图6—图8所示。

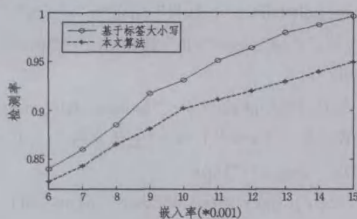


图6 基于标签大小写算法与本文算法的对比

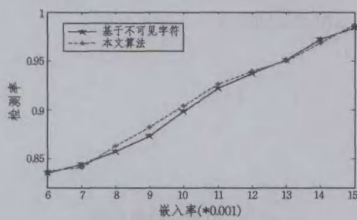


图7 基于不可见字符算法与本文算法的对比

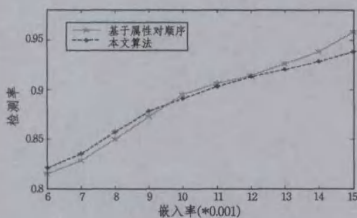


图8 基于属性对顺序算法与本文算法的对比

由图6可以看出,本文算法检测率明显低于基于标签字母大小写算法的;由图7可以看出,本文算法检测率与不可见

字符水平相当;由图8可以看出,本文算法检测率在嵌入率为0.01之后开始逐渐低于基于属性对顺序算法的。由此可看出,PSWIH对比3类方法拥有较强的不可感知性。在实验过程中注意到,对于PSWIH算法副方式中不可见字符和属性对顺序的不同比例,检测率可能有所偏差。

网页抗提取性一般是指在可感知的情况下,攻击者通过分析提取出网页中隐藏信息的能力。实验采用同一个网页作为隐藏载体,隐藏相同的信息。基于标签大小写算法采用大写隐藏比特1,小写隐藏比特0的规则;基于不可见字符算法采用文献[5]中的隐藏规则;基于标签属性对顺序算法采用文献[3]中的隐藏规则。对比PSWIH与3类方法,随着嵌入率的增长,提取隐藏信息的时间变化如图9所示。

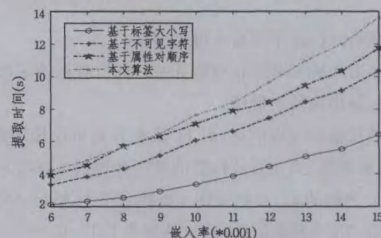


图9 本文算法与3类方法的对比

从图9中可以看出,对比3类方法,本文算法的提取时间最长,拥有较强的抗提取性。

网页鲁棒性一般是指在隐藏信息的网页遭受非法篡改、破坏的条件下仍能提取出隐藏信息的能力。对于基于标签大小写算法的隐藏方式,一旦某个标签大小写被篡改,将会导致整个隐藏信息无法提取,因此基于标签大小写隐藏方式鲁棒性较差;基于不可见字符算法的隐藏方式不易改变其不可见字符,但若插入一个无关的不可见字符,也将导致整个隐藏信息无法提取,因此基于不可见字符算法的隐藏方式鲁棒性一般;基于属性对顺序算法的隐藏方式,只要少量改变标签属性对顺序就会导致整个隐藏信息无法提取,因此基于属性对顺序算法的隐藏方式鲁棒性较差;而本文PSWIH将待隐藏信息分散到网页各个地方,同时采取改变标签大小写、嵌入与待隐藏信息不相关的不可见字符、改变属性对顺序进行干扰,不易察觉和篡改,即使篡改或者添加无关不可见字符,只要其不在隐藏有效段落中,仍可提取出完整的隐藏信息,因此鲁棒性较好。

综上所述,各方法的性能比较情况如表2所列。

表2 性能比较

方法	不可感知性	抗提取性	鲁棒性
字母大小写	弱	弱	弱
不可见字符	强	中	弱
属性对顺序	强	中	弱
算法	强	强	强

PSWIH算法不仅继承了不可见字符在浏览器中不改变显示效果的优点,同时还结合了基于标签字母大小写简便易行和改变属性对顺序抗检测性的特点,具有较强的不可感知性、抗提取性及鲁棒性等,从而提高了信息隐藏的安全性。

**结束语** 本文采用ASCII编码方式下的不可见字符为主方式进行信息隐藏,改变标签字母大小写、嵌入无关不可见字符和改变属性对顺序为副方式进行检测干扰,既提高了传统3类方法的不可感知性,又克服了弱提取性、弱鲁棒性等安全性不高的弊端。实验结果表明,PSWIH算法在不可感知性、抗提取性及鲁棒性等方面有了一定的提高和加强,能够有

(下转第146页)

- [16] Cao L Y. Practical method for determining the minimum embedding dimension of a scalar series [J]. *Physical D: Nonlinear Phenomena*, 1997, 110(1/2): 43-50
- [17] Xu Guo-dong, Song Jia-ning, Li Peng-fei. Pulsar navigation adaptive filtering algorithm based on information quality[J]. *Optics and Precision Engineering*, 2015, 23(3): 827-837 (in Chinese)  
徐国栋, 宋佳凝, 李鹏飞. 基于信息质量的脉冲星导航自适应滤波算法[J]. *光学精密工程*, 2015, 23(3): 827-837
- [18] Ji Shu-jiao, Zhu Ming, Lei Yan-min, et al. Video stabilization with improved motion vector estimation[J]. *Optics and Precision Engineering*, 2015, 23(5): 1458-1465 (in Chinese)  
吉淑娇, 朱明, 雷艳敏, 等. 基于改进运动矢量估计法的视频稳像[J]. *光学精密工程*, 2015, 23(5): 1458-1465
- [19] Yang Gong-liu, Guo Wei-lin, Yuan Er-kai. Compensation of time delay in ship deformation measured by attitude matching[J]. *Optics and Precision Engineering*, 2015, 23(5): 1409-1415 (in Chinese)  
杨功流, 郭蔚林, 袁二凯. 姿态匹配法测量船体变形角中时间延迟的补偿[J]. *光学精密工程*, 2015, 23(5): 1409-1415
- [20] Sen S, Spatscheck O, Wang Dong-mei. Accurate, scalable in-network identification of p2p traffic using application signatures[C]// *Proceedings of the 13th international conference on World Wide Web*. New York, NY, USA: ACM, 2004: 512-521
- [21] Kasera S, Pinheiro J, Loader C. Fast and robust signaling overload control[C]// *Proceedings of Ninth International Conference on Network Protocols*. Riverside, USA: IEEE, 2001: 323-331
- [22] Bauer M. Approximation Algorithms and Decision Making in the Dempster-Shafer Theory of Evidence—an empirical study[C]// *12th Conference on Uncertainty in Artificial Intelligence (UAI 96)*. 1997: 217-237
- [23] Yager R, Liu L. *Classic Works of the Dempster-Shafer Theory of Belief Functions* [M]. Springer-Verlag, Berlin, 2008
- [24] Mruphy C K. Combining belief function when evidence conflicts [J]. *Decision Support System*, 2000, 29(1): 1-9
- [25] Stegink M, Idziejczak I. Detection Of Peer-To-Peer Botnets [R/OL]. <http://staff.science.uva.nl/~delaat/sne-2007-2008/p22/report.pdf>
- [26] Zhaoa D, Traorea I, Sayed B, et al. Botnet detection based on traffic behavior analysis and flow intervals[J]. *Computers & Security*, 2013, 39(4): 2-16
- [27] Kang Jian, Zhang Jun-Yao, Li Qiang, et al. Detecting New P2P Botnet with Multi-chart CUSUM[C]// *International Conference on Networks Security, Wireless Communications and Trusted Computing (NSWCTC 2009)*. Wuhan, China, 2009: 688-691

(上接第 140 页)

效提高网页信息隐藏的安全性。针对即将到来的 HTML 5.0 时代,在仅识别小写字母的情况下,如何更好地定义起始有效标签和终止有效标签是未来算法必须进一步考虑的重点。同时,信息隐藏量与添加干扰不可见字符、改变标签属性顺序对数量之间的关系,即安全性、隐藏容量与干扰量三者之间的关系,也是本文算法实用化关注的方向。

### 参 考 文 献

- [1] Singh P, Chaudhary R, Agarwal A. A Novel Approach of Text Steganography based on null spaces[J]. *IOSR Journal of Computer Engineering*, 2012, 3(4): 11-17
- [2] Zhao Q, Lu H. PCA-based web page watermarking[J]. *Pattern Recognition*, 2007, 40(4): 1334-1341
- [3] Garg M. A novel text steganography technique based on Html documents[J]. *International Journal of Advanced Science and Technology*, 2011, 35: 129-138
- [4] Katzenbeisser S, et al. *Information hiding techniques for steganography and digital watermarking*[M]. Artech House, 2000
- [5] Lee I S, Tsai W H. Secret communication through web pages using special space codes in HTML files[J]. *International Journal of Applied Science and Engineering*, 2008, 6(2): 141-149
- [6] Zhang X, Zhao G, Niu P. A novel approach of secret hiding in webpage by bit grouping technology[J]. *Journal of Software*, 2012, 7(11): 2614-2621
- [7] Sun P, Lu H. An efficient web page watermarking scheme[C]// *2nd IEEE International Conference on Computer Science and Information Technology*, 2009 (ICCSIT 2009). IEEE, 2009: 163-167
- [8] Chou Y C, Liao H C. A Webpage Data Hiding Method by Using Tag and CSS Attribute Setting[C]// *2014 Tenth International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP)*. IEEE, 2014: 122-125
- [9] IIII Yang Y, Yang Y. An efficient webpage information hiding method based on tag attributes[C]// *2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*. IEEE, 2010, 3: 1181-1184
- [10] Huang H, Zhong S, Sun X. An algorithm of webpage information hiding based on attributes permutation[C]// *IIHMSP08 International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, 2008. IEEE, 2008: 257-260
- [11] Huang H, Sun X, Sun G, et al. Detection of hidden information in tags of webpage based on tag-mismatch[C]// *Third International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, 2007 (IIHMSP 2007). IEEE, 2007, 1: 257-260
- [12] Huang H, Sun X, Li Z, et al. Detection of hidden information in webpage[C]// *Fourth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2007)*, 2007, 4: 317-321
- [13] Wang Jian-feng, Huang Liu-sheng, Tian Miao-miao, et al. Web pages hidden information detection based on statistics and the SVM classification [J]. *Chinese Computer Systems*, 2014, 35(6): 1221-1225 (in Chinese)  
王剑锋, 黄刘生, 田苗苗, 等. 基于统计和 SVM 分类的网页隐私信息检测[J]. *小型微型计算机系统*, 2014, 35(6): 1221-1225
- [14] Channalli S, Jadhav A. Steganography an art of hiding data[J]. *International Journal on Computer Science and Engineering*, 2009, 1(3): 137-141
- [15] Rafat K F, Sher M. Innocuous Communication via HTML Hiding Data in Plain Sight[J]. *Arabian Journal for Science and Engineering*, 2014, 39(2): 783-798
- [16] Tang Guang-ming, Wang Ya-di. Information hiding safety study [J]. *Computer Engineering*, 2008, 34(16): 183-185 (in Chinese)  
汤光明, 王亚弟. 信息隐藏安全性研究[J]. *计算机工程*, 2008, 34(16): 183-185
- [17] Huang Hua-jun, Wang Bao-wei, Sun Xing-ming, et al. An Algorithm of Webpage Information Hiding Based on Repeated Importing of the CSS Class Selectors[J]. *Computer Research and Development*, 2009, 46(z1): 138-142 (in Chinese)  
黄华军, 王保卫, 孙星明, 等. 基于 CSS 类选择符重复引入的网页信息隐藏算法[J]. *计算机研究与发展*, 2009, 46(z1): 138-142