

基于雅克比符号的文本信息隐藏

左祥建, 杨晓莉

(陕西师范大学 计算机科学学院, 陕西 西安 710119)

摘要: 文本信息隐藏是版权维护的一种重要手段。通过轻微缩放字符大小比例来实现信息隐藏的方法鲁棒性差, 不能抵抗格式攻击。结合密码学中大整数因子分解困难性假设, 提出了基于雅克比符号的改进算法, 通过用雅克比符号判断文本载体中的元素是否为二次剩余。若为二次剩余, 即雅克比符号为“1”, 将秘密信息每个字节的 ASCII 码值, 分别替换该字符的缩放比例值; 若为非二次剩余, 则将文本载体元素的缩放比例保持标准型。该方法能将秘密信息分散到文本载体中。改进后的算法具有更好的隐藏性, 提高了抵抗添加和删除攻击的能力, 增强了文本信息隐藏的鲁棒性。与传统的文本信息隐藏算法相比具有更好的可靠性, 而且算法的实现简单, 具有灵活性, 可以适用于各种不同语言文本。

关键词: 文本载体; 信息隐藏; 雅克比符号; 二次剩余; 秘密钥

中图分类号: TP301

文献标识码: A

文章编号: 1673-629X(2016)02-0073-04

doi: 10.3969/j.issn.1673-629X.2016.02.017

Text Information Hiding Based on Jacobi Symbol

ZUO Xiang-jian, YANG Xiao-li

(School of Computer Science, Shaanxi Normal University, Xi'an 710119, China)

Abstract: Text information hiding is an important way of copyright maintenance. Text information hiding algorithms that hide information through slightly zooming in or zooming out the characters are not robust, and nor can they resist reformat attacks. In this paper, based on the assumption that factoring a large integer is hard, propose an improved algorithm which hides or not hides text information by determining whether a number is a quadratic residue module a big integer or not. Judging by Jacobi symbol text vector element is quadratic residue. If a quadratic residue, which Jacobi symbol is "1", the secret information of each byte ASCII code value, scaling values are replaced with the characters. If the non-quadratic residue, scaling text vector elements will keep the standard. This method can disperse into the text secret information carrier. The improved algorithm can better hide text information and can resist adding or deleting attack. It is simple and flexible, and can be applied to different language texts.

Key words: text carrier; information hiding; Jacobi symbol; quadratic residue; secret key

1 概述

随着 Internet 的日益普及, 多媒体信息的交流已经达到前所未有的深度和广度。人们可以在互联网上进行网络资源的共享和信息的交互, 但互联网给人们带来便利的同时, 也面临着信息在传输过程中的安全问题。当要传输一些机密资料信息时, 通常将这些信息进行加密再传输。但加密的信息会形成一堆激发非法者破解机密资料动机的乱码^[1]。为了保证信息的安全传输, 在选择信息安全加密算法的同时, 也可以将信息隐藏在文本中进行传输。这样非法攻击者很难发现信息的存在, 也很难猜出信息是以何种方式被隐藏的。因此文本信息隐藏技术是保证信源和信道均安全的信

息安全技术^[2]。如何安全地隐藏文本信息是文中研究的目的。

相对于视频信息隐藏、图像信息隐藏、音频信息隐藏等领域的研究^[3-5], 文本信息隐藏起步较晚、成果少, 这和文本的特征有关系。文本相对于其他多媒体来说, 具有冗余空间少、嵌入率低、文本的修改容易检测等缺点。但也有自身的优点, 人们的很多创意和表达都是以文本的方式进行存储和传输的, 其应用大大超过了视频、图像、音频的适用范围, 它潜在的价值仍然不可低估。在未来的信息安全保障体系中仍将发挥着重要的作用。目前文本信息隐藏主要有以下几种方法:

收稿日期: 2015-05-22

修回日期: 2015-08-25

网络出版时间: 2016-01-26

基金项目: 国家中央高校基本科研业务费专项资金项目 (GK201504017); 包头市科技局项目 (2014S2004-2-1-15)

作者简介: 左祥建 (1990-), 男, 硕士研究生, 研究方向为密码学与信息安全。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20160126.1521.064.html>

(1) 基于文本格式的信息隐藏^[6]。文本格式的信息隐藏就是对文本的内容不进行修改,只是对文本格式进行微弱的调整,如微调文本的显示格式和文字间距,改变字体颜色和增加格式标记符号等。美国贝尔实验室最早提出过基于字移、行移的信息隐藏方法^[7],该方法的优点是隐藏的内容没有发生改变,缺点是一旦文本的格式发生改变,隐藏的信息也会消失。

(2) 基于语义的文本信息隐藏^[8]。语义信息隐藏的基本原理就是将一段正常的语言文字修改成另一段正常语言的过程中将秘密的信息隐藏进去。该方法主要是利用同义词或同音词替换嵌入秘密消息。目前相对成熟的算法有基于同义词替换算法^[9]、基于等价规则替换算法^[10]等。该方法的优点是具有很好的不可察觉性,算法简单。缺点是通信双方必须事先约定好替换规则,一旦遇到文本的部分的删除或篡改,会影响秘密信息的恢复。

(3) 基于语法的文本信息隐藏^[11]。语法文本信息隐藏的基本原理是利用概率统计理论和自然语言生成技术,根据语法规则,在句法模板和字典等资源的支持下,将私密的信息编码成类似自然语言的文本内容。为了生成更安全的隐写文本,Chapman 和 Davida 利用概率上下文文法和同义词替换策略,开发出一款基于语法隐藏的隐写软件 Nicetext^[12]。该方法生成的文本在意义上杂乱无章,很难躲过人眼的检查,一般只用于对安全性要求不高的地方。

2 新的隐藏方法

针对将秘密信息隐藏到文本格式中,文献[13]通过轻微缩放字符大小比例来实现信息隐藏。在一篇正常的文档中,字符的缩放比例通常是标准形,即 100%,对需要嵌入秘密信息的字符采用缩放的比例分别设为 101%、102%、103% 和 104% 来进行编码,从而使得每个载体文本的字符可实现 2 位二进制码的隐藏而不易被发觉。此算法在隐蔽性和隐藏容量方面比直接改变字体有很大的提高,但鲁棒性很差,不能抵抗格式攻击。

文中在该算法的基础上,结合密码学中大整数因子分解困难性假设,提出一种基于雅克比符号判断的文本信息隐藏方法。该方法具有更好的隐藏性,提高了抵抗添加和删除攻击的能力,增强了文本信息隐藏的鲁棒性。

该方法的基本思想是利用雅克比符号判断二次剩余问题的基本原理实现在文本中安全地隐藏秘密信息。通过用雅克比符号判断文本载体中的元素是否为二次剩余。若为二次剩余,即雅克比符号为“1”,将秘密信息每个字节的 ASCII 码值,分别替换该字符的缩

放比例值。如将秘密信息转化为二进制比特流,若比特数为 1,则将文本载体元素的比例缩放为 101%;若比特数为 0,则将文本载体元素的缩放比例保持标准型 100%。若为非二次剩余,则将文本载体元素的缩放比例保持标准型 100%。该方法能将秘密信息分散到文本载体中,而且具有很好的不可察觉性。

3 预备知识

定义 1: 设 n 是正整数,若同余式有解,则 a 叫模 n 的二次剩余;否则 a 叫模 n 的非二次剩余。

$$x^2 \equiv a \pmod{n}, (a, n) = 1$$

定义 2: 设 p 是素数, a 是一整数,符号 $(\frac{a}{p})$ 的定义如下:

$$\left(\frac{a}{p}\right) = \begin{cases} 0, & \text{若 } p \mid a \\ -1, & \text{若 } a \text{ 是模 } p \text{ 的非二次剩余} \\ 1, & \text{若 } a \text{ 是模 } p \text{ 的二次剩余} \end{cases}$$

称符号 $(\frac{a}{p})$ 为勒让德 (Legendre) 符号, 计算

$(\frac{a}{p})$ 有一个简单的公式:

$$\left(\frac{a}{p}\right) \equiv a^{(p-1)/2} \pmod{p}$$

设 $n = p_1 p_2 \cdots p_r$ 是奇数 p_i 的乘积。对任意的整数 a , 定义雅克比 (Jacobi) 符号 (记做 $J(a, n)$) 为:

$$\left(\frac{a}{n}\right) = \left(\frac{a}{p_1}\right) \left(\frac{a}{p_2}\right) \cdots \left(\frac{a}{p_r}\right) = J(a, p_1) J(a, p_2) \cdots J(a, p_r)$$

雅克比符号形式上是勒让德符号的推广,但所蕴含的意义不同,雅克比符号为 -1,可判断 a 是模 n 的二次剩余,但雅克比符号为 1 却不能判断 a 是模 n 的二次剩余。比如 $n = p_1 \times p_2$, a 关于 p_1 和 p_2 都不是二次剩余,即 $x^2 \equiv a \pmod{p_1}$ 和 $x^2 \equiv a \pmod{p_2}$ 都无解,由中国剩余定理知 $x^2 \equiv a \pmod{n}$ 也无解。但是,由于 $(\frac{a}{p_1}) = (\frac{a}{p_2}) = -1$, 所以 $(\frac{a}{n}) = (\frac{a}{p_1}) (\frac{a}{p_2}) = 1$, 即 $x^2 \equiv a \pmod{n}$ 虽无解,但 $J(a, n)$ 却为 1。当 $(\frac{a}{p_i}) = 1 (1 \leq i \leq r)$ 时, a 才是模 n 的二次剩余。

4 方案的实现

4.1 文本载体的生成

在网络上随机下载一份文本,将其转化为一个 $s \times t$ 的文本载体矩阵 D 。

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1t} \\ a_{21} & a_{22} & \cdots & a_{2t} \\ \vdots & \vdots & \vdots & \vdots \\ a_{s1} & a_{s2} & \cdots & a_{st} \end{pmatrix}$$

D 中元素的下标值代表该元素进行雅克比符号判断的数值,即元素 a_{ij} ($1 \leq i \leq s, 1 \leq j \leq t$) 的雅克比符号为 $J(ij, n)$, 其中 ij 都是十进制数。

4.2 密钥的产生

随机选择两个大素数 p 和 q , 满足 $p \equiv q \equiv 3 \pmod{4}$, 即这两个素数形式为 $4k+3$ 。计算 $n = p \times q$, 此时的 p , q 就是秘密钥。

4.3 嵌入算法

输入秘密信息 M , 文本载体 D , 秘密钥 p 。将秘密信息 M 隐藏在文本载体 D 中的算法如下:

(1) 将要隐藏的秘密信息转化成二进制比特流 m_i ($i = 1, 2, \dots$)。

(2) 在文本载体矩阵 D 中顺序读取其中的元素 a_{ij} 的下标值 ij , 运用秘密钥 p 和 q , 判断元素 a_{ij} 的雅克比符号 $J(ij, n)$, 其中 $n = p \times q$ 。

(3) 若 a_{ij} 的雅克比符号为“-1”或“0”, 则将文本载体元素的缩放比例保持标准型 100%。

(4) 若 a_{ij} 的雅克比符号为“1”, 由 $J(ij, n) = J(ij, p) \times J(ij, q)$, 判断 $J(ij, p)$ 和 $J(ij, q)$ 的符号, 若 $J(ij, p)$ 和 $J(ij, q)$ 的符号都为“-1”, 则将文本载体元素的缩放比例保持标准型 100%; 若 $J(ij, p)$ 和 $J(ij, q)$ 的符号都为“1”, 则 ij 为模 n 的二次剩余, 将秘密信息每个字节的二进制值, 分别替换该字符的缩放比例值。若比特数为 1, 则将文本载体元素的比例缩放为 101%, 若比特数为 0, 则将文本载体元素的缩放比例保持标准型 100%。

(5) 重复上述步骤, 直到将秘密信息的比特流 m_i 依次隐藏到 D 中, 生成一个隐藏秘密信息的文本 D' 。

4.4 提取算法

输入隐藏信息的文本载体 D' , 秘密钥 p 和 q 。将秘密信息 M 从文本载体 D' 提取出来的算法如下:

(1) 在 D' 中顺序读取元素 a_{ij} 的下标值 ij , 运用秘密钥 p 和 q , 判断元素 a_{ij} 雅克比符号 $J(ij, n)$ 。

(2) 若 a_{ij} 的雅克比符号为“-1”或“0”, 则不用对该元素进行任何处理。

(3) 若 a_{ij} 的雅克比符号为“1”, 由 $J(ij, n) = J(ij, p) \times J(ij, q)$, 判断 $J(ij, p)$ 和 $J(ij, q)$ 的符号, 若 $J(ij, p)$ 和 $J(ij, q)$ 的符号都为“-1”, 则不用对该元素进行任何处理; 若 $J(ij, p)$ 和 $J(ij, q)$ 的符号都为“1”, 则 ij 为模 n 的二次剩余, 则让该元素的缩放比例与 $J(ij, n) = -1$ 的元素缩放比例进行比较, 若比例增大, 则得到比特 1, 若比例保持不变, 则得到比特 0。

(4) 重复上述步骤, 顺序取出秘密信息的比特流 m_i , 恢复秘密信息 M 。

4.5 改进提取算法

上述方案在隐藏秘密信息和提取秘密信息时, 需

要对文本载体矩阵 D 中的元素依次作雅克比符号的判断, 时间复杂度为 $O(n^2)$, 现可对以上算法作如下改变:

由 n 是两个大素数 p 和 q 的乘积, 则 1 到 $p-1$ 之间有一半的数是模 p 的二次剩余 (记这些数的集合为 Q_p), 另一半是模 p 的非二次剩余 (记这些数的集合为 $\overline{Q_p}$)。对大素数 q 也有类似的结论 (分别记两个集合为 Q_q 和 $\overline{Q_q}$)。若 a 是模 n 的二次剩余, 当且仅当 a 既是模 p 的二次剩余也是模 q 的二次剩余, 即 $a \in Q_p \cap Q_q$ 。

对 $0 < a < n$ 且 $\gcd(a, n) = 1$, 满足该条件的 a 的个数为 $\varphi(n) = (p-1)(q-1)$, 其中有一半满足 $(\frac{a}{n}) = 1$ ($a \in Q_p \cap Q_q$ 或 $a \in \overline{Q_p} \cap \overline{Q_q}$), 另一半满足 $(\frac{a}{n}) = -1$ ($a \in Q_p \cap \overline{Q_q}$ 或 $a \in \overline{Q_p} \cap Q_q$)。而在满足 $(\frac{a}{n}) = 1$ 的 a 中, 有一半满足 $(\frac{a}{p}) = (\frac{a}{q}) = 1$ ($a \in Q_p \cap Q_q$), 这些 a 就是模 n 的二次剩余; 另一半满足 $(\frac{a}{p}) = (\frac{a}{q}) = -1$ ($a \in \overline{Q_p} \cap \overline{Q_q}$), 这些 a 就是模 n 的非二次剩余。

由上分析知, 有 $\varphi(n)/4$ 个元素满足 $(\frac{a}{n}) = 1$ 。在隐藏秘密信息时, 若文本载体矩阵 D 中元素的下标值属于集合 $Q_p \cap Q_q$, 则隐藏一个比特的信息。改进后算法的时间复杂度为 $O(n)$ 。

5 安全性和鲁棒性分析

文中算法的安全性主要体现在以下几个方面:

(1) 隐蔽性: 字符缩放比例的选取是关键, 文献 [14] 中的改变文字大小法中, 载体文本中每个字符的字号大小改变量为 0.5 磅 (word 文档中字号的最小变化量为 0.5 磅)。在这种情况下, 对于较大的字体, 如四号 (14 磅) 字以上的文本文档, 其大小改变比例小于 $0.5/14 \approx 3.57\%$, 人的视觉对此改变感觉不明显; 但是对于较小的字体, 如小五号 (9 磅) 字以下的文本文档, 其大小改变比例大于 $0.5/9 \approx 5.56\%$, 人的视觉对此改变感觉较为明显。对于任意字号大小的字体, 其纵向高度没有改变, 横向大小改变比例都小于或等于 4%, 人的视觉对此改变感觉不明显。文中字符缩放的比例为 1%, 具有更好的隐蔽性。

(2) 安全性: 文中算法的安全性是基于密码学中大整数因子分解困难性难题。在不知道密钥 p 、 q 的情况下, 无法用雅克比符号判断二次剩余问题, 即使攻

击者知道了字符的缩放比例,也无法将秘密信息从文本载体中提取出来。

从这个意义上来说,文中提出的基于雅克比符号文本信息隐藏算法是非常安全的,具有很好的不可察觉性和抗分析能力。

由于文本的特殊性,鲁棒性弱是基于文本载体的信息隐藏技术不可避免的问题。当文本的内容遭到删除或篡改攻击,都会影响到隐秘密信息的恢复。文中基于对雅克比符号的判断隐藏信息,使得秘密信息均匀分散在整篇文档中,大大减少了被删除的秘密信息量,提高了秘密信息的抗攻击性和自恢复性。对于在文本中增加字符或更改文本格式的攻击,也是基于雅克比符号的判断提取信息,只要发现文本中某个元素的雅克比符号和缩放比例不一致,就能发现文本载体被篡改,接收者可以终止算法。对于字符替换攻击,由于文档具有继承性,替换后的字符仍然具有秘密信息的特征,所以可以正确提取秘密信息。

6 结束语

文中在轻微缩放字符大小比例来实现信息隐藏的基础上,提出了基于雅克比符号的判断实现信息隐藏的算法。该算法具有很好的隐蔽性、安全性,增强了文本信息隐藏的鲁棒性。与传统的文本信息隐藏算法相比具有更好的可靠性,而且算法的实现简单。在实际的使用中还需要对密钥的分发和管理作进一步研究。

参考文献:

- [1] 李顺东,王道顺.现代密码学:理论,方法与研究前沿[M].北京:科学出版社,2009.
- [2] Petitcolas F A P, Anderson R J, Kuhn M G. Information hiding - a survey[J]. Proceedings of the IEEE, 1999, 87(7): 1062-1078.
- [3] 胡 洋,张春田,苏育挺.基于 H.264/AVC 的视频信息隐藏[J]. 电子学报, 2008, 36(4): 690-694.
- [4] Wu D C, Tsai W H. A steganographic method for images by pixel-value differencing [J]. Pattern Recognition Letters, 2003, 24(9): 1613-1626.
- [5] Avcibas I. Audio steganalysis with content-independent distortion measures[J]. IEEE Signal Processing Letters, 2006, 13(2): 92-95.
- [6] 曹卫兵,戴冠中,夏 煜,等.基于文本的信息隐藏技术[J]. 计算机应用研究, 2003, 20(10): 39-41.
- [7] Low S H, Maxemchuk N F, Brassil J T, et al. Document marking and identification using both line and word shifting[C]//Proc of fourteenth annual joint conference of the IEEE computer and communications societies. [s. l.]: IEEE, 1995: 853-860.
- [8] 徐迎晖,杨 榆,钮心忻,等.基于语义的文本隐藏方法[J]. 计算机系统应用, 2006(6): 91-94.
- [9] Topkara U, Topkara M, Atallah M J. The hiding virtues of ambiguity: quantifiably resilient watermarking of natural language text through synonym substitutions [C]//Proceedings of the 8th workshop on multimedia and security. Geneva: ACM, 2006: 164-174.
- [10] Lucena N B, Pease J, Yadollahpour P, et al. Syntax and semantics - preserving application - layer protocol steganography [C]//Information hiding. Berlin: Springer, 2005: 164-179.
- [11] 赵敏之,孙星明,向华政.基于虚词变换的自然语言信息隐藏算法研究[J]. 计算机工程与应用, 2006, 42(3): 158-160.
- [12] Chapman M, Davida G. Hiding the hidden: a software system for concealing ciphertext as innocuous text [J]. Information and Communications Security, 1997, 1(1): 335-345.
- [13] 李向辉.基于 Word 文本文档的信息隐藏方法研究[D].南宁:广西大学,2006.
- [14] 刘玉玲,孙星明.通过改变文字大小在 Word 文档中加载数字水印的设计与实现[J]. 计算机工程与应用, 2005, 41(12): 110-112.

(上接第 72 页)


- [8] 朱 强,王慧强,冯光升,等. VNE-ABC: 基于人工蜂群的网络虚拟化映射算法[J]. 北京工业大学学报, 2014, 40(1): 68-73.
- [9] Dong Z. A study on virtual network decomposing mapping algorithm based on network balance [C]//Proc of fourth international conference on computational and information sciences. [s. l.]: [s. n.], 2012: 880-883.
- [10] Di H, Yu H, Anand V, et al. Efficient online virtual network mapping using resource evaluation [J]. Journal of Network and Systems Management, 2012, 20(4): 468-488.
- [11] Huang Tao, Liu Jiang, Chen Jiangya, et al. A topology-cognitive algorithm framework for virtual network embedding problem [J]. China Communications, 2014(4): 73-84.

- [12] Cui Hongyan, Tang Shaohua, Huang Xu, et al. A novel method of virtual network embedding based on topology convergence-degree [C]//Proc of IEEE international conference on communications workshops. [s. l.]: IEEE, 2013: 246-250.
- [13] Butt N F, Chowdhury M, Boutaba R. Topology awareness and reoptimization mechanism for virtual network embedding [J]. Networking, 2010, 6091: 27-39.
- [14] Li Wen, Wu Chunming, Chen Jian, et al. Virtual network mapping algorithm with repeatable mapping over substrate nodes [J]. Journal of Electronics and Information Technology, 2011, 33(4): 908-914.

基于雅克比符号的文本信息隐藏

作者：[左祥建](#)，[杨晓莉](#)，[ZUO Xiang-jian](#)，[YANG Xiao-li](#)

作者单位：[陕西师范大学 计算机科学学院](#)，[陕西 西安](#)，[710119](#)

刊名：[计算机技术与发展](#)

英文刊名：

年，卷(期)：2016 (2)

引用本文格式：[左祥建](#)，[杨晓莉](#)，[ZUO Xiang-jian](#)，[YANG Xiao-li](#) [基于雅克比符号的文本信息隐藏](#)[期刊论文]-[计算机技术与发展](#) 2016 (2)