一种 Word 2007 中可无源检测的信息隐藏方法

李兵兵[1],王衍波[1],徐 敏[2],周 杰[1]
(1.解放军理工大学 通信工程学院,江苏 南京 210007;
2.西南电子电信技术研究所,四川 成都 610041)

摘 要:研究基于文本的信息隐藏技术,对使用文本传输秘密信息,以及版权维护等方面有很大的实用价值。文本,作为信息隐藏的载体,相对于视频、图像和声音,其冗余度更小,所以基于文本的信息隐藏技术发展相对较慢。新的文本文档格式的出现对基于文本的信息隐藏提供了更多的方法和可能。文中研究分析了一种新的文本文档格式——Word 2007 采用的 Microsoft Office Word XML 格式,提出并实现了一种基于 Word 2007 字体大小微调的可无源检测的信息隐藏方案和算法。理论分析和实验表明算法有效可靠,具有良好的隐蔽性。

关键词:信息隐藏;Word 2007 文档;无源检测
中图分类号:TP301 **文献标识码**:A **文章编号**:1673-629X(2010)05-0154-04

An Information Hiding Method with "Blind" Detection Based on Word 2007

LI Bing-bing[1], WANG Yan-bo[1], XU Min[2], ZHOU Jie[1]
(1. Institute of Communication Engineering, PLA UST, Nanjing 210007, China;
2. Southwest Electronics and Telecommunications Research Institute, Chengdu 610041, China)

Abstract: The research on information hiding techniques based on text is of great use in transmitting secret information with text and protecting copyright of text. Compared with video, image and voice, text, as the carrier of information hiding, has lower redundancy, so the technology of information hiding based on text developed tardily. The presented new text document format provides more methods and possibility for information hiding based on text. In this paper, a new text document format—Microsoft Office Word XML of Word 2007 is researched and analyzed, and a scheme and algorithm of information hiding with "Blind" detection is presented by little modifying the value of character size, then realized. The theoretical analysis and experiment show the algorithm is effective, reliable and possessed of favorable crypticness.

Key words: text information hiding; Word 2007 document; "Blind" detection

0 引 言

信息隐藏技术研究如何将某一信息隐藏于另一公开的信息中,然后通过公开信息的传输来传递隐藏信息[1]。基于文本格式文档的信息隐藏技术是近年来发展细化出来的一个新分支,其中基于 Word 文档的信息隐藏已经取得了一系列成果。Word 作为目前使用最广泛的文本处理软件,发展到现在先后出现了 Word 95、Word 2000、Word 2003 等多个版本,目前最新版本为 Word 2007。Word 软件,在得到不断加强改善的同时,其自身文档格式也在不停演化,从早期版本采用的 Word 97 格式到 Word 2007 采用的 Word XML 格式。新文档格式的出现,对基于 Word 的信息隐藏提供了更多的方法和可能。文中通过对 Word XML 格式进行研究分析,提出了一种可无源检测的信息隐藏方法。

收稿日期:2009-09-18;**修回日期**:2009-12-23
基金项目:江苏省自然科学基金项目(BK2008090)
作者简介:李兵兵(1984-),男,硕士研究生,研究方向为信息隐藏;王衍波,教授,研究方向为密码学及信息安全。

1 Microsoft Office Word 2007 格式简介

Microsoft Office Word 2007 提供了一种新的默认文件格式,叫做 Microsoft Office Word XML 格式(Word XML 格式)[2]。这种新的文件格式由一个压缩的 ZIP 包组成,包中包含了文档中的所有内容。通过这种包格式,因为使用 ZIP 压缩技术,所以可以减少 Office 文档文件的大小。图 1 所示为一个典型的 Word 2007 文档的层次化文件结构。从图 1 中可以看出,这个 ZIP 包中大部分都是 XML 文档,这些文档承载着 Word 文

档本身可见的文本(包括大小、颜色、段间距、字间距等)以及不可见的整个文档的关联框架。

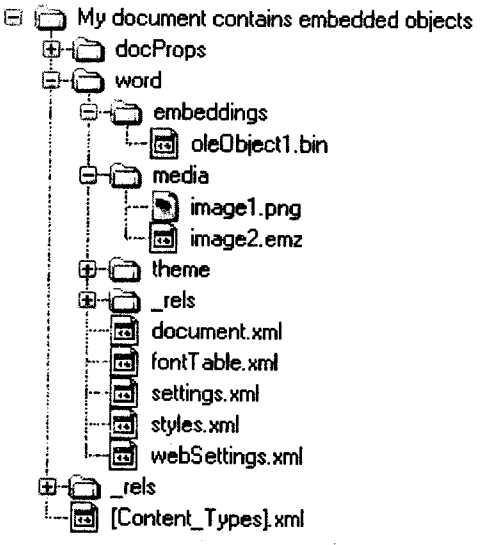

图 1　一个典型的 Word 2007 文档的层次化文件结构

Word 文档的可见文本部分全部存在于/word/目录下的 ducument. xml 文档中,其中文本的字体大小属性,除默认值保存于/word/styles. xml 外,其他所有文本的字体大小属性都存在于 ducument. xml 文档中。在 Word 2007 文档中,对文本字体大小属性的标注并不是以单个文本为单位的,而是以文本域为单位的,即若相连两个或多个文本其属性(包括字形、字体大小、颜色、字间距等)完全相同时,这两个或多个文本就形成一个文本域,此域的标注属性对域中所有文本都适用。这样的好处是可以免去相同属性的重复标注,并且能节约存储空间减少文档的大小。

Word 2007 中文本的字体大小可设置范围是 1～1638,字号调整的最小幅度是 0.5。在 ducument. xml 文档中,字体大小属性值是以字体实际大小的双倍值存储的,如字体大小为 10.5,其字体大小属性值为 21,字体大小为三号(15 号字),其字体大小属性值为 30。

2　一种基于字体大小微调的可无源检测的信息隐藏算法

要在 Word 2007 文档中隐藏信息必须满足两个条件:

1)隐藏信息后的文档必须满足 Word XML 格式的要求;

2)隐藏信息后的文档必须能正常显示。

基于以上要求设计了一种通过改变 Word 2007 字体大小且可以实现无源检查[3]的信息隐藏算法。

2.1　算法思想

微调文本文档中文本的大小是不易被发现的。由于 Word 2007 中文本属性的标注是以文本域为单位的,若对单个文本的大小进行微调,会生成新的文本域,势必大大增加文档本身的大小,所以以文本域为单位进行微调是比较合适的方案。

文献[4]中提出的方法,虽然隐藏信息比较安全可靠,但在隐秘信息检测提取时必须有原始的载体文档做参考,即不能进行无源检查,这就对隐秘信息检测提取造成了不方便,所以在设计信息隐藏方案时必须兼顾考虑检测提取信息的方便性和正确性。对字体大小属性值进行奇偶性微调即可满足上述要求,方法为对文本域字体大小属性值的奇偶性进行格式化,若要隐藏信息 1,则格式化为奇数,若要隐藏信息 0,则格式化为偶数,提取隐藏信息时直接判断文本域字体大小属性值的奇偶性即可提取相应的信息。此方法不但隐藏简单易行,提取也方便快捷。字体大小属性值加 1 可以保证字体大小的变化不容易被察觉。具体隐藏算法如下,假如当前文本域的字体大小属性值为 $k(k \geq 1$ 且 $k \in N)$:

$$\text{隐藏信息 0}\begin{cases} k = k & k \bmod 2 = 0 \\ k = k+1 & k \bmod 2 = 1 \end{cases} \tag{1}$$

$$\text{隐藏信息 1}\begin{cases} k = k+1 & k \bmod 2 = 0 \\ k = k & k \bmod 2 = 1 \end{cases} \tag{2}$$

例如,在文本"我们应该每天都用一个积极向上的心态来面对生活中的困难和挫折"中隐藏信息"1A"。

秘密信息:1A

十六进制 ASCII 码:31 61

二进制码:00110001 01100001

隐藏信息前文本域字体大小如表 1 所示。

表 1　隐藏信息前文本域字体大小

文本域	我们	应该	每天	都用	一个	积极向上	的	心态	来	面对	生活	中	的	困难	和	挫折
大小	21	22	21	23	22	21	22	23	22	21	22	21	22	23	22	21

隐藏信息后文本域字体大小如表 2 所示。

表 2　隐藏信息后文本域字体大小

文本域	我们	应该	每天	都用	一个	积极向上	的	心态	来	面对	生活	中	的	困难	和	挫折
大小	*22*	22	21	23	22	*22*	22	23	22	21	*23*	*22*	22	*24*	22	21

表 2 中斜体标注的大小属性值即为格式化过的。从上两个表的比较来看,只改变了少部分字体的大小,这样可以较少代价(最少修改原文)达到隐藏信息的目的。

隐藏信息流程如图 2 所示:由于 Word 2007 文档实质上是一个 ZIP 压缩包,所以可以对 Word 文档进行解压压缩处理。对于解压 Word 2007 文档抽取得到的 Document. xml 文件,其中文本域大小属性值与默认值相等的情况下,大小属性是缺省的,就不能在此文本

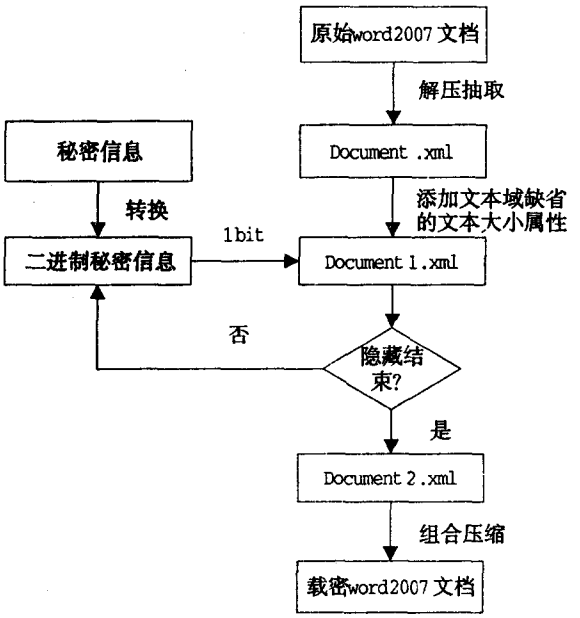

图 2 信息隐藏流程图

域嵌入隐秘信息。对缺省文本大小属性的文本域添加大小属性后，即可在每个文本域中嵌入隐秘信息，提高整篇文档的信息隐藏容量。

检测提取隐秘信息时，不需要原始载体文档，只需通过对文本域字体大小属性值奇偶性进行判断，即可容易得到隐藏的信息。

2.2 算法分析

(1) 由于 Word 2007 文档采用了一种 Word XML 新格式，使得该算法能非常简单的实现。相对于老版本的 Word 文档，Word 2007 本质上是一个 ZIP 压缩包，只用抽出包含文本字体大小属性的 document. xml 单个文档，然后利用成熟的 XML 文档处理技术对其编辑，即可实现该算法。

(2) 容量：隐藏 1bit 的信息需要一个文本域，字符数量不定，至少一个字符。相比文献[5～7]所提的隐藏算法，该算法隐藏容量相对较低，这源于 Word 2007 文档的新格式，即在第一节中所讲，对文本字体大小属性的标注并不是以单个文本为单位的，而是以文本域为单位的。如若按文献[5～8]中提出的算法隐藏信息，势必会大大增加载体文档的大小，透明性降低，安全性减弱。

(3) 透明性：隐藏后的载体文件与原始载体文件相比，正常情况下，视觉上是无差别的，人眼很难觉察得到。算法只改变字体大小属性值的数据，特殊情况下才会添加文本域大小属性，所以载密文档与载体文档大小只可能有细微差距，甚至完全一样，透明性较好。

(4) 鲁棒性：对文档中的字符进行大小设置，如果这个字符大小属性值的奇偶性不变，则对该隐藏算法毫无影响，隐藏信息可以完整准确的提取出来。其他情况下，则很可能使隐藏信息的提取出错而变得毫无意义，所以该算法鲁棒性较脆弱。

3 实验及结果分析

选取秘密文本信息共 176 个字符，大小为 356 个字节，载体文件为Word2007文档共3806个字符，大

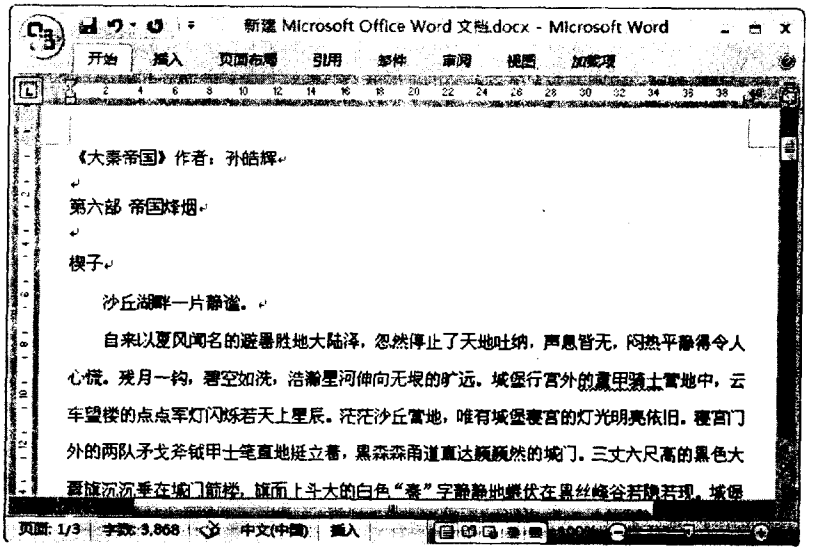

图 3 未嵌入秘密信息的载体文档

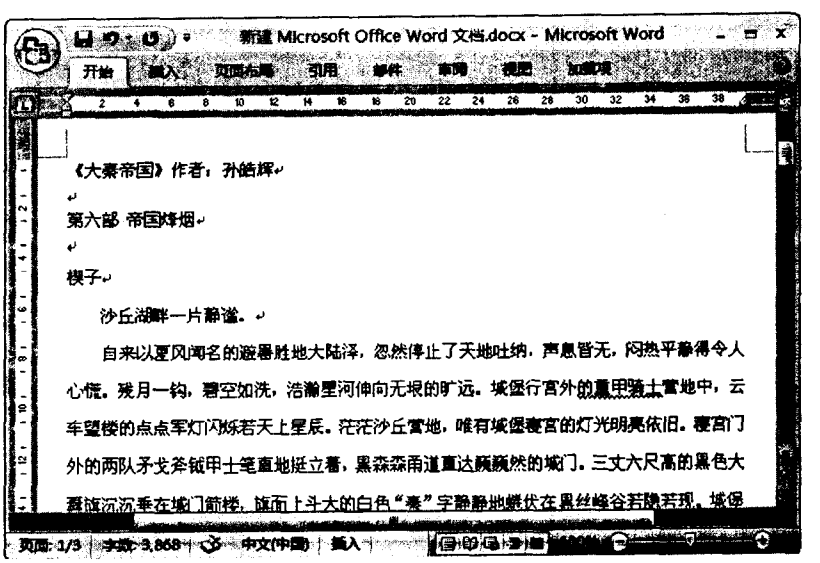

图 4 嵌入秘密信息的载密文档

小为 19187 个字节,嵌入秘密文本信息后的载密文档字符数仍为 3806,大小仍为 19187 个字节。图 3 和图 4 分别为未嵌入秘密信息的载体文档和嵌入秘密信息的载密文档,图 5 为从载密文档中提取的秘密信息。

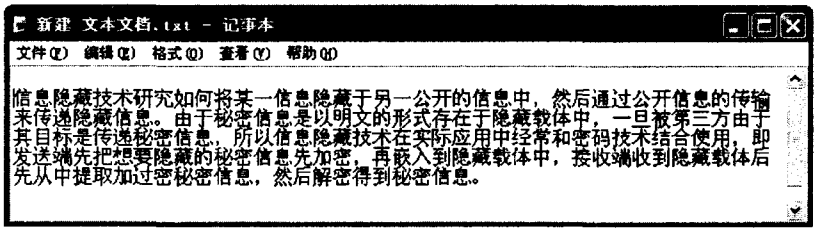

图 5　提取的秘密信息

从实验结果看,嵌入秘密信息后的载密文档与原始的载体文档视觉效果完全一样,感觉不到隐藏秘密信息的存在。

4　结束语

与其他文本文档隐藏方法相比,该隐藏算法具有实现简单、透明性好的优点,缺点是鲁棒性比较脆弱。虽然算法本身比较简单,但通过该算法可以看到,基于 Word 2007 文档的信息隐藏技术和基于 xml 文档的信息隐藏技术是有交叉的,后者的隐藏方法很可能在前者同样适用,这给以后研究基于 Word 2007 文档的信息隐藏提供了一个新思路。

参考文献:

[1] 王炳锡,陈　琦,邓峰森.数字水印技术[M].西安:西安电子科技大学出版社,2003.

[2] Walkthrough: Word 2007 XML Format [EB/OL]. 2008 - 08 - 25[2008 - 10 - 25]. http://msdn.microsoft.com/en-us/library/bb266220.aspx.

[3] Khodami A A, Yaghmaie K. Persian Text Watermarking [C]//PCM 2006. Berlin, Heidelberg: Springer - Verlag, 2006:927 - 934.

[4] Brassil J T, Low S, Maxemchuk N F. Copyright Protection for the Electronic Distribution of Text Documents[J]. Proceedings of IEEE, 1999, 87(7):1181 - 1196.

[5] 李向辉,钟　诚.提高 Word 文本文档信息隐藏容量的方法研究[J].计算机技术与发展,2006,16(9):97 - 99.

[6] 陈　萍,郭水旺,陈华丽.基于字体颜色的文本信息隐藏算法[J].科学技术与工程,2007,7(14):2544 - 2546.

[7] 耿红琴.基于 word 文本文档的信息隐藏技术研究[J].科学技术与工程,2007,7(11):2686 - 2688.

[8] 陈　芳,王　冰.基于文本字体的信息隐藏算法[J].计算机技术与发展,2006,16(1):20 - 22.

(上接第 153 页)

表 3 显示了在这两个方案中生成和验证一个单独签名,一个签名者的多个签名以及多个签名者的多个签名所需的真实时间(t 是签名数,k 是签名人数)。设 *RSA* 的公用指数 e = 3,用中国剩余定理加速签名,但是在任何验证操作中还没有最优化的技术。

表 3　成本比较:验证和签名

		压缩 RSA	BGLS
Sign	1 signature	6.82	3.54
Verify	1 signature	0.16	62
	t = 1000, k = 1	44.12	184.88
	t = 100, k = 10	45.16	463.88
	t = 1000, k = 10	441.1	1570.8

6　结束语

ODB 模式的数据库安全是一个较新的研究课题,笔者在此文中研究提供了有效的 ODB 数据完整性机制模型和安全有效的压缩 RSA 方案,它在单一客户端和多查询者模型中运行良好,只是还不能聚合不同签名者的签名,因此不适用多所有者模式。另一方面,虽然 BGLS 签名方案能把不同用户的签名聚合成一个短签名,但计算复杂度相当高。因此,今后工作的重点就是研究适合多所有者模式的有效实用的签名方案。

参考文献:

[1] Hacigümüs H, Iyer B, Mehrotra S. Providing Database as a Service[C]//In International Conference on Data Engineering. Washington: IEEE Computer Society, 2002:29 - 40.

[2] 赵晓峰,叶　震.几种数据库加密方法的研究与比较[J].计算机技术与发展,2007,17(2):219 - 222.

[3] 徐茂智,游　林.信息安全与密码学[M].北京:清华大学出版社,2007.

[4] 王平水,赵俊杰.多用户环境中签名方案的安全性研究[J].计算机技术与发展,2009,19(1):157 - 160.

[5] Bellare M, Garay J, Rabin T. Fast batch verification for modular exponentiation and digital signatures[C]//In Advances in Cryptology - EUROCRYPT '98, LNCS1403. Berlin: Springer - Verlag, 1998: 191 - 204.

[6] Bellare M, Rogaway P. Random oracles are practical: a paradigm for designing efficient protocols[M]. New York: ACM Press, 1993: 62 - 73.

[7] Boneh D, Gentry C, Lynn B, et al. Aggregate and Verifiably Encrypted Signatures from Bilinear Maps[C]//In Advances in Cryptology - EUROCRYPT '2003, LNCS2656. Berlin: Springer - Verlag, 2003:416 - 432.

[8] OpenSSL Project[EB/OL]. 2009 - 04 - 21. http://www.openssl.org.