

一种新的基于混沌映射的文本零水印算法

程玉柱¹, 孙星明², 黄华军²

(1. 湖南大学 软件学院, 湖南 长沙 410082; 2. 湖南大学 计算机与通信学院, 湖南 长沙 410082)
(Vogue21ct@163.com)

摘要:在分析现有文本数字水印方法缺陷的基础上,提出了一种新的基于混沌映射的文本零水印算法。实验结果证明算法具有以下性能:加载水印时对文档信息不作任何修改,水印完全不可见;可以有效地抵制针对文本文档的复制、剪切、格式调整等编辑操作,水印具有较强的鲁棒性;检测水印受口令限制且实现了盲检测。

关键词:数字水印;零水印;混沌映射

中图分类号:TP309.2 **文献标识码:**A

Text zero-watermarking algorithm based on chaotic mapping

CHENG Yu-zhu¹, SUN Xing-ming², HUANG Hua-jun²

(1. School of Software, Hunan University, Changsha Hunan 410082, China;

2. School of Computer and communication, Hunan University, Changsha Hunan 410082, China)

Abstract: A novel text zero-watermarking algorithm based on chaotic mapping was brought forward. The features of presented method are as follows: First, there are no modifications to the text document among the process of watermarking embedding, thus the watermarking is perceptually invisible. Second, the watermarking is robust against copying, cutting and formatted modifying. Third, the watermarking can be detected only by using a secret key and doesn't need the original text.

Key words: digit watermarking; zero-watermarking; chaotic mapping

0 引言

目前文本数字水印算法按编码方式大致分为两类:

1) 基于文本格式或字符特征编码

基于文本格式编码算法通过改变文本字间距或行间距来嵌入水印^[1],而基于字符特征编码水印算法一般是通过修改文本字符特征或在文档中附加空格以嵌入水印^[2],以上两种算法均是基于HVS的视觉掩蔽特性。但由于文本没有足够的冗余信息,在文本中嵌入信息极易被阅读者发现;同时,一些字处理软件在有意无意间也会破坏原始文件,所以这些方法存在一定的局限,很难同时保证水印的鲁棒性和不可见性。

2) 基于文本语义编码

基于文本语义编码算法则是利用本体论知识分析文本语义,在不改变文本语义的条件下调整自然语言句子的结构,有效地将水印嵌入到文本内容之中。如文献[3]提出的基于TMR(Text Meaning Representation)的水印算法。

相对基于文本格式或字符特征编码算法来说,基于文本语义编码算法在提高水印鲁棒性上有了较大突破,水印不再依赖于文本格式或字符特征,对于英文文本已取得了一定成果。但由于汉语的多义性,该算法目前尚不能很好地应用于中文文本。基于此,本文提出了一种新的基于混沌映射的文本零水印算法,结合零水印基本原理,提取文本特征构造水印,水印发送给可靠第三方来保存和验证。实验结果证明,该算法实现的水印达到了鲁棒性和不可见性的较好统一。

1 混沌系统理论基础

混沌是指在确定性系统中出现的一种貌似无规则的、类似随机的现象^[4],对一维离散动力系统: $x_{n+1} = f(x_n, v)$ ($n = 0, 1, 2, \dots$), 设 V 是一个度量空间, 如果映射 $f: V \rightarrow V$ 满足以下三个条件, 则称 f 在 V 上是混沌的: 1) 对初值的敏感依赖性; 2) 拓扑传递性: 任意一点的邻域在 f 的作用下将“遍布”整个度量空间 V ; 3) f 的周期点集在 V 中稠密。由混沌系统的这些特性, 要对系统进行长期的预测是不可能的, 将这个性质应用到数字水印算法中, 可以提高水印的随机性、增大攻击难度。本文选取 Logistic 混沌映射:

$$x_{n+1} = \mu x_n (1 - x_n) \quad x_n \in [0, 1] \quad \mu \in [0, 4] \quad (1)$$

作为水印加密理论基础。由该映射产生的混沌序列有几个优点: 1) 混沌序列的产生非常方便, 只需给出一个参数 μ 和初值 x_0 , 便可产生数量众多的混沌序列; 2) 混沌序列是一个类似随机的过程, 而且从混沌序列的值很难推出原始的参数和初值, 因此具有很好的保密性; 3) 混沌序列具有良好的相关特性, 只有使用相同参数和初值产生的序列相关性较好, 而采用不同参数和初值产生的序列相关性近似为零, 这有利于水印信号的检测。

2 文本零水印算法

2.1 汉字数学表达式基本原理和部件频次统计

1) 汉字数学表达式的基本原理

汉字数学表达式是一种全新的汉字数学表达方法, 即把

收稿日期: 2005-06-14; 修订日期: 2005-08-21

作者简介:程玉柱(1980-), 男, 安徽安庆人, 硕士研究生, 主要研究方向: 数字水印、自然语言处理; 孙星明(1963-), 男, 湖南益阳人, 教授, 博士生导师, 博士, 主要研究方向: 信息安全、自然语言处理、分布式数据库; 黄华军(1978-), 男, 湖南宜春人, 博士研究生, 主要研究方向: 信息安全、数字水印。

汉字表示成由汉字部件作为操作数、运算符号为部件间结构关系的数学表达式^[5]。选定组成国标一、二级汉字的 580 个基本部件,6 种运算符:lr、ud、ld、lu、ru、we,它们依次表示左右、上下、左下、左上、右上、全包含等关系。图 1 是部分所选基本部件及其编号;图 2 是对上述运算符号的一个简要描述,图中 A 和 B 分别表示基本部件。

1	2	3	4	5	6	7	8	9	10
一	乙	之	二	十	丁	厂	卜	人	八
11	12	13	14	15	16	17	18	19	20
又	入	几	九	儿	力	刀	乃	了	七
21	22	23	24	25	26	27	28	29	30
匕	乚	广	丫	于	工	下	土	上	才
31	32	33	34	35	36	37	38	39	40
兀	寸	女	丈	大	干	子	口	门	上
41	42	43	44	45	46	47	48	49	50
千	久	夕	义	川	丸	么	已	马	弓

图 1 部分所选基本部件及其编号

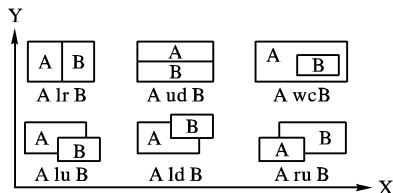


图 2 运算符的直观描述

为进一步说明,表 1 给出了部分汉字的数学表达式实例。

表 1 部分汉字数学表达式实例

袄: 470 lr 108	敖: (335 ud 302) lr 450
芭: 313 ud 122	捌: 430 lr (38 ud 16) lr 416
奥: 507 ud 495 we 195 ud 35	

2) 部件频次统计方法

部件频次统计方法首先将文本内各汉字表示成数学表达式形式;然后通过分析表达式结构特点,统计出各部件出现频次。下面以规则形式给出方法描述:

规则 1 令 T 表示待处理的文本文档, Ω 表示 T 中出现的所有汉字, Θ 表示汉字基本部件的集合。设 $C_i \in \Omega$ 中的任一汉字, $D(C_i)$ 表示 C_i 在 T 中出现的次数; P 为 Θ 中的任一基本部件, $D(P)$ 表示 P 在 T 中出现的次数, $t(p, c_i)$ 表示 P 在 C_i 中出现的次数。则有:

$$D(P) = \sum_{C_i \in \Omega} D(C_i) \cdot t(P, C_i) \quad (2)$$

2.2 文本零水印算法

零水印是指利用载体的特征来构造一个水印,它的最大特点是不改变宿主信息^[6]。就文本零水印技术而言,就是考虑如何利用文本的特征来构造水印,并且该水印能有效地抵制文本编辑操作(如复制、剪切、格式调整等)。零水印构造完毕后,将其发送到注册中心进行注册登记。当文本发生版权纠纷时,通过比较检测得到的水印与已注册水印来验证版权归属。

在描述算法前,先给出如下定义:

定义 1 设 v 为根据文档长度设定的阈值, $No(p)$ 为部件 p 对应的部件号, $D(p) \in v$ 表示部件 p 频次在阈值 v 内。

定义 2 方法 MD5(String) 表示采用 hash 函数 MD5() 对字符串 String 求 hash 摘要。

算法 1: 零水印构造算法

Input: 文本文档 T , 密钥 k_1, k_2

Output: 零水印 w

Begin

① 对 $\forall C_i \in T$, 统计各部件频次 $D(p)$;

② 任选部件 $p(D(p) \in v)$, 取 $k_1 = No(p)$ 并查询 p 在 T 内对应的汉字序列 $List(C)$;

③ 设 $List(C)$ 长度为 n , 将方程(1)迭代 n 次得到混沌序列 x_k ($k = 1, 2, \dots, n$), 其中 $x_0 = k_1 \times 0.001, \mu = k_2$, 进而通过定义阈值 σ 对 x_k 进行二值化处理得:

$$\xi_k = \begin{cases} 1 & x_k \geq \sigma \\ 0 & x_k < \sigma \end{cases} \quad k = 1, 2, \dots, n \quad (3)$$

④ 由 ξ_k 映射 $List(C)$ 得文档特征汉字序列 m (运算符“ \circ ”表示按 ξ_k 为 1 或 0 取舍 $List(C_k)$ 中相应汉字):

$$m = \sum_k \xi_k \circ List(C_k) \quad k = 1, 2, \dots, n \quad (4)$$

⑤ 构造零水印 $w = MD5(m)$

End

算法 2: 零水印检测算法

Input: 待检测文档 T , 密钥 k_1, k_2

Output: 零水印 w'

Begin

① 对 $\forall C_i \in T$, 统计各部件频次 $D(p)$;

② 由密钥 k_1 确定部件 $p(No(p) = k_1)$, 并查询 p 在 T' 内对应的汉字序列 $List(C)$;

③ 设 $List(C)$ 长度为 n , 将方程(1)迭代 n 次得到混沌序列 x_k ($k = 1, 2, \dots, n$), 其中 $x_0 = k_1 \times 0.001, \mu = k_2$, 进而由(公式(3))对 x_k 进行二值化处理得 ξ_k ;

④ 根据(公式(4))由 ξ_k 映射 $List(C)$ 得特征汉字序列 m' ;

⑤ 检测零水印 $w' = MD5(m')$

⑥ if ($w = w'$)

文本具有原始水印, 文本版权应属于作者;

else

文本不具原始水印, 文本版权与作者无关

End

3 实验结果与性能分析

实验针对 PDF 文档进行了水印加载和检测。为测试算法性能, 我们分别对文档进行了格式和特征修改以及部分删除。水印加载与检测结果分别如图 3、图 4、图 5 所示。



图 3 加载水印



图 4 对文本进行特征攻击后检测水印

实验结果表明, 本算法较之以往基于文本格式和字符特征编码等水印技术而言, 在不可见性、鲁棒性以及安全性上均有了一定改进:

(1) 不可见性: 不同于“向载体中加入水印信息”的水印

(下转第 2758 页)

过程:在管理站以 keyAdmin 为用户,使用认证且加密安全级别,以错误的认证口令向代理发送一个密钥分配报文。结果:代理不执行请求的操作,并回复错误报告“验证失败”。

(3) 验证消息的防延迟、重放功能

过程:对截获的有效密钥分配报文延迟一段时间(150秒)后重放。结果:代理不执行请求的操作,并回复错误报告“报文超出时间窗”。

(4) 验证消息的保密性

过程:在管理站以 keyAdmin 为用户,使用认证且加密安全级别,向代理发送一个密钥分配报文。使用 sniffer 截获该报文并进行观察分析。结果:sniffer 无法解读截获报文中的变量绑定,只能观察到加密后的“乱码”,如图 4 所示。

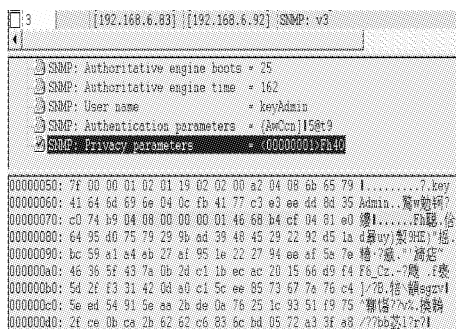


图 4 认证且加密安全级别的密钥分配报文

可见利用 SNMPv3 的 USM 机制,本文所实现的系统能够为密钥分配提供所需的数据完整性认证、数据源认证、防重放和防泄露等安全保证。

4.2 访问控制功能测试

过程:配置一个用于一般网络管理的用户 netAdmin,将

keyMIB 子树排除在其访问权限之外。在管理站以 netAdmin 为用户向代理发送一个密钥分配报文,访问 keyMIB 字树下的管理对象。结果:代理不执行请求的操作,并回复错误报告“对象不存在”。

过程:在管理站以 keyAdmin 为用户向代理发送一般的网络管理操作报文,访问 keyMIB 子树以外的管理对象。结果:代理不执行请求的操作,并回复错误报告“对象不存在”。

可见利用 SNMPv3 的 VACM 机制,本文所实现的系统能够对不同的用户进行访问控制,从而为密钥分配和一般的网络管理功能分别授权,进一步保证了系统的安全性。

5 结语

SNMPv3 是 SNMP 协议的最新版本。它为 SNMP 定义了完整的体系结构,并提供了以 USM 和 VACM 为主的安全机制,既能够满足安全网络管理的需要,又保持了简单易用的特点。基于 SNMPv3 的安全机制,本文提出了一种为网络管理系统增加密钥分配功能的方案。这种方案还可以进一步扩展,例如为支持 SNMPv3 的网管系统增加安全策略管理、服务策略管理等功能,从而实现多功能安全网络管理系统。

参考文献:

- [1] 陈妍,卢泽新,冯艳玲. SNMPv3 新增安全机制的研究与实现[J]. 计算机工程与应用,2004,(16):137-139,164.
- [2] ZELTSEMAN D. SNMPv3 与网络管理[M]. 潇湘工作室译. 北京:人民邮电出版社,2000.
- [3] 金鹏,郝平. SNMPv3 中的安全机制[J]. 通信技术,2002,(4):77-79.
- [4] DAWSON ED, CLARK A, LOOI M. Key management in a non-trusted distributed environment[J]. Future Generation Computer Systems, 2000,(16):319-329.

(上接第 2754 页)

方法,零水印技术不对文本格式、字符特征等作任何修改,原始文本与水印文本在视觉上无丝毫差别,实现了水印的完全不可见性。

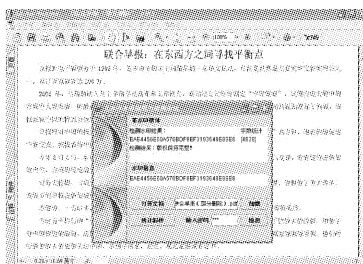


图 5 对文本进行部分删除后检测水印

(2) 鲁棒性:对文本而言,理想的鲁棒性水印应该对合法用户的正常编辑操作是允许的,且水印不因这些操作而轻易丢失;同时,水印对非法用户的恶意攻击(如篡改文本内容)也有一定的抵抗能力。以往基于格式和特征编码的水印技术很难保证这一点,简单的格式重排或特征修改往往会使水印信息轻易丢失;该算法通过提取文档内某汉字部件所对应的部分汉字序列来构造水印,水印不再依赖于文本格式或字符特征,对文本常规编辑操作等有较好的鲁棒性,针对文本的恶意篡改也有一定的抵抗能力。

(3) 安全性:水印的安全性基于 Logistic 混沌映射的强随机性和初值敏感性,非法用户在不知道版权所有者的口令的情况下无法确定水印特征汉字集合。同时,从版权所有者发布的水印信息推导文档特征汉字信息也是不可能的。此外,水印检测时不需要原始文档。

4 结语

随着网络技术的发展,文本数字水印技术作为一种版权保护技术,正得到越来越多的应用。零水印的提出,为文本水印如何加强鲁棒性和安全性提供了一个新的思路。下一步的研究目标是探讨如何利用零水印机制和公钥机制对文本提供更好的认证保护。

参考文献:

- [1] BRASSIL JT, LOW SH, MAXEMCHUK NF, et al. Electrical Marking and Identification Techniques to Discourage Document Copying[J]. IEEE Journal on Selected Areas in Communications, 1995, 13(8):1495-1504.
- [2] LOW SH, MAZEMCHUK NF, BRASSIL JT, et al. Document Marking and Identification Using Both Line and Word Shifting[A]. Proceedings of the Fourteenth Annual Joint Conference of the IEEE Computer and Communication Societies[C]. Boston, USA, 1995, 853-860.
- [3] ATALLAH MJ, RASKIN V, HEMPELMANN CF, et al. Natural Language Watermarking and Tamperproofing[A]. Proceedings of the Fifth International Information Hiding Workshop[C]. Netherlands, 2002. 196-212.
- [4] 王宏霞,何晨,丁科. 基于混沌映射的鲁棒性公开水印[J]. 软件学报, 2004,15(8):1245-1251.
- [5] SUN XM, CHEN HW, YANG LH, et al. Mathematical Representation of a Chinese Character and its Applications[J]. International Journal of Pattern Recognition and Artificial Intelligence, 2002, 16(8):735-747.
- [6] 温泉,孙铁锋,王树勋. 零水印的概念与应用[J]. 电子学报, 2003,31(2):214-216.