

一种基于纠错码的数字指纹方案

贺英英, 张卫党, 王伟, 郑来文

(郑州大学信息工程学院, 河南 郑州 450052)

摘要:数字指纹技术是一种用来辨认非法拷贝, 追踪非授权用户的新技术。它由嵌入拷贝各自不同的数据组成。现在, 数字指纹已发展成为版权保护的一种重要手段。但一个问题也随之出现: 对于数字指纹, 可以通过比较找到它们的不同之处, 于是一些用户可以通过合谋找到数字指纹的位置, 更改数字指纹来隐匿自己。所以设计指纹算法的关键问题是如何防范合谋攻击。在这里, 介绍了一种新的基于纠错码的数字指纹, 并进一步分析了这种方案的性能。

关键词:数字指纹; 维特比算法; 纠错码

中图分类号: TP309.2

文献标识码: A

文章编号: 1673-629X(2006)05-0125-03

A Digital Fingerprinting Scheme Based on Error - Correcting Code

HE Ying-ying, ZHANG Wei-dang, WANG Wei, ZHENG Lai-wen

(Institute of Information Technology, Zhengzhou University, Zhengzhou 450052, China)

Abstract: Digital fingerprinting is a technique for identifying unauthorized copy and tracing back to its user. Fingerprinting consists of uniquely marking and registering each copy of data. Now, digital fingerprinting has appeared as a new technique for copyright protection of digital contents. A problem arises when users collude: for digital data, two different fingerprinted objects can be compared and the differences between them detected. Hence, a set of users can collude to detect the location of the fingerprint. They can then alter the fingerprint to mask their identities. How to defense collusive attack is the key problems for the designing of fingerprinting. In this paper, present a new method for assigning error - correcting code for the purpose of fingerprinting digital data. Furthermore, analyze the performance of this scheme.

Key words: digital fingerprinting; Viterbi algorithm; error - correcting code

0 引言

信息技术的发展以及电子商务的广泛应用使各类音乐、影视等作品通过网络的传播范围空前扩大。同时, 以数字形式存在的产品很容易被非法拷贝, 如何对数字化产品进行版权保护已经成为信息时代产权保护的核心问题之一。数字水印技术和数字指纹技术是近几年发展起来的新颖数字版权保护技术。发行商通过在其所要发售的拷贝中嵌入与购买者有关的数字指纹可以对盗版行为进行跟踪。由于各个用户所得到的拷贝不同, 通过合谋他们就可能发现码字的嵌入位置或内容。因此, 抗合谋攻击是数字指纹研究中需要解决的一个根本问题。对要嵌入的指纹信息进行编码是解决此问题的一种有效手段。在指纹编码的经典文献[1]中, Boneh 和 Shaw 讨论了抗合谋攻击的数字指纹的编码和跟踪问题。他们首先提出了嵌入假设 (marking assumption), 并在此基础上提出了一种指纹码字的长度与用户个数的对数成正比例的编码算法。由

于该嵌入假设要求用户对码元相同的比特位无法作出修改, 一个误发的随机错误就可能导致错误判断。文献[2]指出, 若参与合谋的用户所对应的码字标号有较大间隔, 则会以较高概率产生误判。文献[3, 4]对文献[1]中的弱点进行了进一步的研究, 并提出了更切合实际的嵌入假设。

1 相关编码知识

纠错码的特点就是在接收端收到码字后, 通过检验收到的码字是否满足预先约定的约束关系, 不但能判断是否有错, 还能通过译码直接纠错得到正确的码字。纠错码种类很多, 它包括分组码、卷积码等等。在这些纠错码里, 笔者选用了规律性较强的二进制线性等距码来构造数字指纹。顾名思义, 等距码就是所有码字之间的汉明距离都相等的码, 它还是分组码的一种。相关的一些重要特性如下:

① 如果一种码字中的任意两个不相交的码字子集 $\{a, b\}$ 和 $\{c, d\}$, 满足 $D(a, b) \cap D(c, d) = \emptyset$, 则称这种码为 (2, 2) 隔离码。

② 所有的线性等距码都是 (2, 2) 隔离码^[5]。

③ 假设有 a, b 两个等长度的二进制码, 当码字 c 中的

收稿日期: 2005-10-19

基金项目: 郑州市科技攻关项目 (052SGYG21120)

作者简介: 贺英英 (1980-), 男, 河南信阳人, 硕士研究生, 研究方向为信道编码; 张卫党, 副教授, 研究方向为信道编码。

每一位码元都是从 a 或者 b 中对应位的码元中选取时,就可认为 $c = D(a, b)$ 。

在性质 3 中, a, b, c 三者的关系实际上就模拟了数字指纹中两人合谋攻击的模型。其中 a, b 可以看作是两位合谋者,而 c 就是他们合谋攻击制造出的数字指纹。而在性质 1 和 2 中,说明了二进制线性等距码就是一种 (2, 2) 隔离码。因为 (2, 2) 隔离码作为数字指纹可以保证二人合谋攻击时不同的合谋集团产生不同的数字指纹,这对于减少误判几率,提高数字指纹的性能有着非常重要的意义。

④ 设 a, b 为 (n, k, d) 二进制线性等距码。 a, b, c 之间的汉明距离有下面 3 种情况:

- * $d(a, b) = d, d(a, c) = d(b, c) = d/2$
- * $d(a, b) = d, d(a, c) > d/2, d(b, c) < d/2$
- * $d(a, b) = d, d(b, c) > d/2, d(a, c) < d/2$

不管合谋用户采用怎样的攻击方式,至少有一个合谋用户对应的数字指纹码字与篡改得到的数字指纹码字距离小于或等于 $d/2$ 。

2 线性分组码的网格图表示法

在文献[6]中,作者论述了将线性分组码表示成网格图形式的一种方法。这种方法的实质就是用一个离散有限状态的马尔科夫过程来描述一个线性分组码的码字。维特比算法是估计离散有限状态马尔科夫过程的最优方法,所以用网格图表示分组码的目的就是用维特比算法^[7]对等距码进行纠错译码。

所谓网格图,是由两种基本元素构成的:节点和连接这些节点的单向线。其中节点代表了一个状态,而单向线代表的是各个状态之间的转换。在这里,把各个节点可能的状态归入 S_t ,其中下标 t 为深度参数,它表示马尔科夫过程进行到了第几个状态。单向线只能从前一级深度的节点指向下一级深度的节点。

对于任何一个 (n, k) 线性分组码,都必定对应有 $n \times (n - k)$ 的校验矩阵 H 。设分组码的码字为 C ,则 C 必定满足

$$C \times H^T = 0 \quad (1)$$

利用矩阵的性质,可把上式转换为

$$c_1 h_1 + c_2 h_2 + c_3 h_3 + \cdots + c_n h_n = 0 \quad (2)$$

在式(2)中,依照码字 C 的 n 个码元把它化为 n 个二进制数;把矩阵 H 按行化为 n 个 $(n - k) \times 1$ 的矩阵。

利用上述性质,就可以画出网格图。具体过程如下:

① 初始化:设 $S_0 = \{S_0^0\} = (0, \cdots, 0)$;

② 利用迭代法算出从深度 1 到深度 $n - 1$ 的各个状态,计算公式为 $S_{t+1}^i = S_t^j + c_j h_{t+1}$ 。

将各个状态放入节点,再用单向线连接 S_t^j 到 S_{t+1}^i 。最后再将在深度 n 处对应状态不为零的路径去掉就得到了最终的网格图。注意,在深度 n 处不为零就表示这个码字不满足式(1)。

3 跟踪方案

首先采用线性二进制等距码作为数字指纹的码字。发行商将每个用户分配到的码字嵌入拷贝,然后将拷贝分发给用户。当发现了非法拷贝后,就从中提取出拷贝中的数字指纹,然后对指纹采取改动后的维特比算法进行“纠错”,从而得到至少一个参与合谋攻击用户的数字指纹。

具体跟踪方案如下:

首先计算在深度 t 计算篡改后的数字指纹码字 R_t 相对于各个用户的数字指纹码字的相似度,称为分支量度 BM。在二进制硬判决情况下, BM 即为汉明距离。然后计算第 t 时刻到达状态 i 的所有汉明距离小于等于 $d/2$ 的路径并全部保存下来,这些路径称之为留存路径 $S^*(i)$ 。最后不断重复上述步骤,计算出所有 i, t 的可能取值下的留存路径,算到深度 $(n - 1)$ 就可以得到所有的参与了合谋攻击的用户指纹。

4 性能分析

在方案中,将合谋用户对指纹的篡改视为码字在通信过程中产生的错误,把得到的数字指纹通过维特比算法进行“纠错”,寻找参与合谋攻击的用户。在这有两点需要注意:第一,文中在进行纠错译码时,对维特比算法稍作改动。原有的维特比算法是找到最大似然路径的过程,在这里只需改成找到所有与合谋得到码字的汉明距小于等于 $d/2$ 的码字。但是对维特比算法的改动也引起了第二个问题:由于在采用改动的算法进行译码时,找到的是所有与合谋码字的汉明距小于或等于 $d/2$ 的码字,所以就增大了误判的可能。无辜用户的数字指纹码字与合谋攻击得到的数字指纹码字的距离小于或等于 $d/2$ 可能性,显然要大于无辜用户与合谋攻击的指纹汉明距恰好是最小的可能性。

在这里可以计算一下误判的可能性。设合谋得到的码字为 a ,一名无辜用户的码字为 b ,考虑简单的情况,假设 a, b 不相关。由于 a, b 都是 n 位的二进制码字,则 a, b 间的汉明距离小于等于 $d/2$ 的概率为:

$$P(d_{a,b} \leq d/2) = C_n^0 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^n + C_n^1 \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^{n-1} + \cdots + C_n^{d/2} \left(\frac{1}{2}\right)^{d/2} \left(\frac{1}{2}\right)^{n-d/2} \quad (3)$$

式中等号右边第一项是 a, b 距离为零的概率,最后一项是 a, b 距离为 $d/2$ 的概率,把所有的项叠加就得到了 a, b 间的汉明距离小于等于 $d/2$ 的概率。在式(3)中,第 k 项就是距离为 $(k - 1)$ 的距离,即 a, b 两个码字有 $(k - 1)$ 项不同,其他的项有相同的概率。在所有 n 个码元中,选择 $(k - 1)$ 位相异的码元,而且每个二进制码元相同或相异的概率都为 $1/2$,最后就可以得到 a, b 的距离取 $(k - 1)$ 的概率为 $C_n^{k-1} \left(\frac{1}{2}\right)^{k-1} \left(\frac{1}{2}\right)^{n-k+1}$

根据相关的概率论知识,从式(3)可以看出:随着 a, b 之间的汉明距的变化,相应的随机概率 P 呈正态分布。从图 1 中看到: a, b 汉明距的概率分布在 a, b 间的距离取

$n/2$ 时值达到最大,而且其概率以 $n/2$ 为中心,向两侧逐渐递减。图中纵轴表示概率,横轴表示 a, b 之间的汉明距离(本图在横轴上应该是离散分布的,但为了便于观察,笔者采用了连续分布图)。

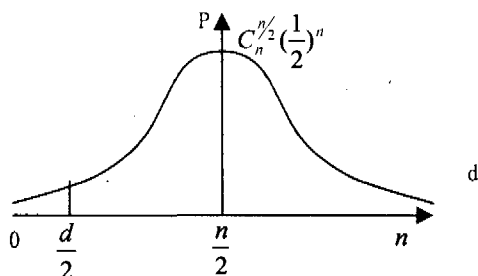


图1 a, b 之间汉明距的概率分布图

算法发生误判的几率就是图中 d 小于等于 $d/2$ 的那部分的积分和。只要选择合适的 n 和 d , 误判的几率可以得到有效控制。就整个指纹方案的跟踪方案来说,产生误判的原因主要有两个:维特比算法本身的差错概率 P_E 和 a, b 间的汉明距离小于等于 $d/2$ 的概率。当 a, b 间的汉明距离小于等于 $d/2$ 时,不管维特比算法本身译码是否正确,跟踪方案都将出错;而当距离大于 $d/2$ 时,出错几率则取决于维特比算法的差错概率。所以得到整个算法的误判率就约为 $P = P_E * (1 - P(d_{a,b} \leq d/2)) + P(d_{a,b} \leq d/2)$

5 结束语

文中介绍了一种利用纠错码构造数字指纹的方案和此方案的基本理念、主要的问题及解决方法。从总体上

(上接第49页)

有很强的抗误码性能。PCAS 算法通过采用更复杂的上下文模型实现系数分类和自适应算术编码取得了更好的编码性能,但增加编解码的复杂性。一般来说,基于集合分裂的算法比基于上下文模型的算法编解码快、复杂性低^[7]。

8 结论

对嵌入式小波编码算法进行了研究,指出了它们的算法原理和改进方案,对比了它们的性能。充分利用小波变换系数的能量聚集特性和子带间、子带内的非线性相关性,并结合人体视觉系统(HVS)有效分类和组织小波系数,以快的排序速度和少的重要图编码来寻找重要图,对不同统计特性的小波系数采用不同的编码策略,设计适于硬件实现的快速编解码算法以及从对重要系数量化方法上进行研究,进一步提高数据的压缩率,仍是此类算法发展的重要方向。

参考文献:

[1] Shapiro J. Embedded image coding using zerotrees of wavelet

看,这个方案适于合谋人数不大于两人(出于保密考虑,合谋集团的人数是很有限的),且用户数量不是太多的情况。至于这种方案的缺点则是随着用户数量的增长,计算量增长的很快,这也是以后需要改进的地方。

参考文献:

- [1] Boneh D, Shaw J. Collusion - Secure fingerprinting for digital data[A]. In: Coppersmith D. Advances in Cryptology: Proceedings of the CRYPTO'95[C]. Berlin: Springer - Verlag, 1995. 452 - 465.
- [2] Liu Z H, Yin P. Techniques and Applications of Information Hiding[M]. Beijing: Science Press, 2002. 178 - 180.
- [3] Guth J, Pfitzmann B. Error - and collusion - secure fingerprinting for digital data[A]. In: Pfitzmann A. Proceedings of the 3rd International Workshop on Information Hiding (IH'99)[C]. Berlin: Springer - Verlag, 2000. 134 - 145.
- [4] Safavi - Naini R, Wang Y. Collusion secure q - ary fingerprinting for perceptual content[A]. In: Sander T. Security and Privacy in Digital Rights Management: Proceedings of the ACM Digital Rights Management Workshop [C]. Berlin: Springer - Verlag, 2002. 57 - 75.
- [5] Cohen G, Encheva S, Schaathun H G. On separating codes [R]. Paris: ENST, 2001.
- [6] Wolf J K. Efficient maximum likelihood decoding of linear block codes using atrellis[J]. IEEE Trans Inform Theory, 1978, 24: 76 - 80.
- [7] 曹雪虹, 张宗橙. 信息论编码[M]. 北京: 北京邮电大学出版社, 2001.
- [8] coefficients [J]. IEEE Transactions on Signal Processing, 1993, 41(12): 3445 - 3462.
- [2] Said A, Pearlman W A. A new, fast and efficient image code based on set partitioning in hierarchical trees[J]. IEEE Transactions on Circuits and Systems for Video Technology, 1996, 6(3): 243 - 250.
- [3] Islam A, Pearlman W A. An Embedded and Efficient Low - Complexity Hierarchical Image Coder[A]. Visual Communications and Image Processing '99, Proceeding of SPIE[C]. [s. l.]: [s. n.], 1999. 294 - 305.
- [4] Taubman D. High performance scalable image compression with EBCOT[J]. IEEE Trans Image Processing, 2000, 9(7): 1158 - 1170.
- [5] Peng K, Kieffer J C. Embedded image compression based on wavelet pixel classification and sorting[J]. IEEE Trans Image processing, 2004, 13(8): 1011 - 1017.
- [6] Taubman D S, Marcellin M W. JPEG2000 图像压缩基础、标准和实践[M]. 魏江力, 柏正亮, 等译. 北京: 电子工业出版社, 2004.
- [7] 朱向军, 朱善安. 基于小波变换的嵌入式图像编码算法综述[J]. 信号处理, 2004, 20(1): 54 - 58.