中国科学 F辑: 信息科学 2009 年 第 39 卷 第 4 期: 397～404
www.scichina.com info.scichina.com
《中国科学》杂志社
SCIENCE IN CHINA PRESS

数字媒体水印非协议模糊性

康显桂 ①*, 黄继武 ①*, 曾文军 ②, 施云庆 ③

① 中山大学信息科学与技术学院, 广州 510275;
② Department of Computer Science, University of Missouri-Columbia, MO 65211, USA;
③ Department of Electrical and Computer Engineering, New Jersey Institute of Technology, NJ 07102, USA
* E-mail: isskxg@mail.sysu.edu.cn, isshjw@mail.sysu.edu.cn

收稿日期: 2008-06-01; 接受日期: 2008-10-26
国家自然科学基金 (批准号: 90604008, 60633030, 60403045)、广东省自然科学基金 (批准号: 04009742) 和国家重点基础研究发展计划 (批准号: 2006CB303104) 资助项目

摘要 抵抗协议模糊攻击对于一个安全的数字媒体版权管理系统来说至关重要. 文中基于安全伪随机序列发生器 (PRSG) 序列与真随机序列计算上不可区分性来研究数字媒体水印的非协议模糊问题. 针对一个使用 PRSG 序列作为水印的算法的攻击可以看成是针对一个加噪 PRSG 序列的攻击. 提出和证明了二进制加噪 PRSG 序列安全定理和广义加噪 PRSG 序列安全定理. 选择合理的检测阈值 $\mathrm{Th} = a\sqrt{n}$ (n 是水印长度和 a 是归一化检测阈值) 和 $n \geqslant 1.39 \times m/a^2$($m$ 是密钥长度) 可以使得协议模糊攻击成功概率 (F_{amb}) 和水印漏检率随水印长度 n 增大迅速变小成为可忽略的, 这使得实际的量化水印算法和扩频盲水印算法可以同时具有非协议模糊性和鲁棒性. 文中得到了水印算法非协议模糊性的充要条件和安全性计算的分析解, 可以用来指导设计一个实际的不可逆水印算法和定量评价水印算法的非协议模糊性.

关键词
数字水印
协议模糊攻击
PRSG 序列
计算上不可区分性
数字版权管理

1 引言

数字水印可以用于多媒体版权保护、数字指纹、篡改检测、认证、访问控制、标注等 [1,2]. 大多数数字水印方法采用两种基本嵌入方法: 扩频水印嵌入方法 [2,3] 和关系嵌入方法 [4]. 基于量化的嵌入水印方法 (本文简称为量化水印方法) 属于关系嵌入方法中的一种主要方法 [4,5].

数字水印用于数字版权管理 (DRM, digital rights management) 是当前数字水印研究的一个热点, 能够对抗协议模糊性攻击 (或者称可逆协议攻击) 对于一个安全的数字版权管理系统来说是一个基本要求. Craver 等 [6] 第一个提出并研究了水印的协议模糊性攻击. 下面将首先回顾一下此后人们对这一问题的许多研究. 遗憾的是这一问题至今没有得到很好地解决 [7~12].

模糊攻击是: Alice (所有者) 拥有原始媒体作品 I 和原始的版权标志水印 W_{A}. 它向公众发布加水印的媒体作品 $\hat{I} = \varepsilon(I, W_{\mathrm{A}})$, 这里 $\varepsilon(\ )$ 是水印嵌入函数. 尽管不知道 Alice 的水印, 攻击者 Bob 想从一个媒体 $\tilde{I} = \hat{I} + N$ (N 代表随机噪声) 中找到伪造水印, 从而挫败 Alice 的版权主张. 攻击成功的必要条件是 Bob 能从媒体 $\tilde{I}$ 中找到一个伪造水印 W_{B} . 我们在第 2 节还会给出模糊攻击的严格定义.

一种对抗模糊攻击的方法是采用不可逆(非模糊)的嵌入、检测方法 [10,13~15], 如采用一些密码学的函数–DES (data encryptial standard) 加密和单向 Hash 函数, 生成水印或者水印密钥. Craver 等 [6] 提出从计算原始媒体 I 的单向 Hash 函数来生成水印密钥, 然后用伪随机数字发生器 (PRG, pseudo randam generator) 生成水印 W, 也就是 $W = \mathrm{PRG}(\mathrm{Hash}(I))$. Qiao 等 [13] 提出将原始媒体的 DES 加密作为水印.

引用格式: 康显桂, 黄继武, 曾文军, 等. 数字媒体水印非协议模糊性. 中国科学 F 辑: 信息科学, 2009, 39(4): 397–404

对于这些非盲水印算法, 非法攻击者面临两个问题: 从待检测媒体中检测到一个水印和通过合法性检验–该水印是用原始媒体的 Hash 值作为密钥产生的一个 PRG 序列; 或者是原始媒体的加密函数. 人们以前一般认为这样构造的水印算法是安全和不可逆的, 因为对原始媒体引入了单向函数, 只有原始媒体的所有者才能成功地声称他们的所有权. 但是, 人们在随后的研究中成功地引入了迭代搜索攻击 [7,10,11,15]. Ramkumar 等 [11,16] 和 Zeng 等 [15] 分别独立地对 Craver 提出的上述协议 [6] 引入迭代搜索攻击: 修改待测媒体的低位比特, 得到一个伪造原始媒体 I_f, 并检查相应的伪造水印 $W_f = \mathrm{PRG}(\mathrm{Hash}(I_f))$ 在媒体中是否可检测到. 文献 [15] 对盲扩频水印算法也提出了类似的迭代攻击, 并且分析得出成功的概率与水印的虚警概率成正比, Adelsbach 等 [7] 根据上述研究结果认为这样一个 “不可逆的” 的水印算法可能是不安全的, 他们还构造了针对上述 Qiao 等提出的水印方案的迭代攻击. 作为对抗迭代搜索攻击的改善措施, 一些方案提出对有效水印空间加以严格限制, 从而降低水印算法的虚警概率. 文献 [10, 15] 提出原始媒体不能参与水印的检测过程 (也就是只能采用盲水印算法) 和有意义字符串或者登记的所有者身份识别码作为水印密钥. 第 1 个要求是正确的, 因为水印检测器本身没法正确区分是否是 “原始媒体”. 但是, 第 2 个要求也不能完全克服迭代搜索攻击. Ramkumar 等 [11] 提出有效水印必须要与原始媒体不相关. 但这个方法要求原始媒体参与水印的检测认证过程, 从而要冒将原始媒体信息公开的风险. 最近 Li 和 Chang[9] 采用密码学上安全的伪随机序列作为水印, 提出了一种可抗逆转攻击的扩频水印渐进原理模型. 但他们的证明是基于一种启发性的方法–统计检验, 仅仅提出了定性的渐进原理模型 [17], 并且要求伪随机序列发生器的密钥长度 m 很小, 要求小于 $\sqrt{n}$ (n 为水印长度), 基于他们提出的渐进原理模型难以设计一个实际的不可逆水印算法. 受到他们工作的启发, 我们 [18] 基于伪随机序列与真随机序列在计算上的不可区分性严格证明了盲水印算法的非协议模糊性, 提出了水印算法的非协议模糊性必要条件, 将密钥长度 m 从 $m < \sqrt{n}$ 扩展到 $m < 0.72 \times a^2 \times n$($n$ 是 PRSG (psendo random sequence generator) 序列长度, a 是归一化的水印检测阈值). 但是什么是水印算法的非协议模糊性的充分必要条件? 如何设计一个实际的可证明不可逆的水印算法? 如何定量评价一个实际水印算法模糊攻击的成功率? 这些问题仍然有待研究.

另一种构造抗协议模糊攻击的方法是在生成水印过程中结合可信任第三方 (TTP, trusted third party) 签发数字签名, 水印中包含有对应于原始媒体的密码学签名 [8,19]. 这种方法在攻击者不能轮询 TTP 的被动攻击者假设下是可证明非模糊的. 但这样一个包含 TTP 的水印协议存在一些缺点, 包括需要拥有大量 “秘密” 的 TTP, 这种 TTP 是否能经得起攻击, 水印生成过程需要 TTP 参与所带给用户的不便和成本增加. 更重要的是, 因为 TTP 在区分合法用户和潜在的攻击者时存在困难, 被动攻击者假设往往不符合实际情况. 比如一个可能的阻止攻击的方法 [8] 是让 TTP 记录已经签名的消息, 并阻止相同消息的重复请求签名. 但是, 攻击者仍然可以采用修改原始消息的低位比特的方法来进行迭代攻击 [10,15]. 因此, TTP 必须能够辨别出同一消息的修改版本, 然而在现阶段这是个难点 [12].

本文在既没有 TTP 参与水印协议的全过程 (包括水印生成、嵌入和检测过程), 又没有原始媒体用于水印生成和检测过程的背景下来研究水印算法的非协议模糊性. 我们的研究表明, 正确的选择水印算法的检测阈值和水印长度 n (或者密钥长度 m), 可以使得模糊攻击的成功率 (定义为模糊攻击者找到一个伪造水印的概率) 和漏检概率同时变得可忽略 (量化水印算法的漏检概率可以变为 0), 这样实际的量化水印和扩频水印算法可以同时具有非协议模糊性和稳健性.

本文的结构安排如下. 为了能严格地分析问题, 本文的第 2 节先介绍标记符号和定义一些概念, 然后提出二进制加噪伪随机序列安全定理, 它将会用于证明和定量分析量化水印算法的协议非模糊性; 第 3 节讨论采用合理检测阈值的量化水印算法的不可逆性, 给出量化水印算法的漏检概率和模糊攻击成功概率的定量估计方法; 第 4 节提出广义加噪伪随机序列安全定理, 讨论扩频水印算法的漏检概率和模糊攻击成功概率的定量估计方法; 第 5 节给出结论.

2 概念、符号和加噪伪随机序列安全性定理

这部分先定义概念和介绍一些标记符号, 然后提出二进制加噪伪随机序列安全定理. 为了说明与真随机序列在计算上不可区分的伪随机序列, 先引入对 n 可忽略函数的概念.

定义 1 概率函数 ζ: $\mathbb{N}\to\mathbb{R}^+$ 是一个可忽略的概率函数, 如果存在 n_0, 对所有的正整数 $n>n_0$ (即 n 为充分大), 对任意的正多项式 $p(\cdot)$ 都有

$$\zeta(n)<\frac{1}{p(n)}. \tag{1}$$

例如, 函数 $2^{-\sqrt{n}}$ 和 2^{-n} 都是对 n 为可忽略函数. 可忽略函数有一个性质: 两个可忽略函数 $\zeta_1(n)$ 和 $\zeta_2(n)$ 的相加 $\zeta(n)=\zeta_1(n)+\zeta_2(n)$ 仍然为可忽略函数 [17]. 我们也同时定义所有不是可忽略函数的概率函数均为不可忽略函数.

定义 2 计算上不可区分性 [17]. 对于任意概率多项式算法 A, 任意的正多项式 $p(\cdot)$ 和充分大的 n, 如果两个集合 $V=\{V_n\}_{n\in\mathbb{N}}$ 和 $U=\{U_n\}_{n\in\mathbb{N}}$ 满足:

$$\begin{aligned}\Delta(n)&=|\Pr[A(V_n)=1]-\Pr[A(U_n)=1]|\\&<\frac{1}{p(n)},\end{aligned} \tag{2}$$

则称它们为多项式时间内不可区分或者称为计算上不可区分. 这里 $\Pr[\ ]$ 表示概率算子. $\Pr[A(V_n)=1]$ 或者缩写为 $\Pr[A(V_n)]$ 表示: A 接受输入 V_n, 输出 "1"(true) 来表示 $V_n\in\boldsymbol{V}$ 的概率. 概率在随机变量 V_n(or U_n) 和算法 A 内部随机掷币等条件下计算. n 为 V_n 的长度. 用 U_n(或者缩写为 U) 来表示一个长度为 n 的真随机序列, $U=\{U_n\}_{n\in\mathbb{N}}$ 称为真随机序列集合. U_n 在长度为 n 的序列集合中均匀随机分布.

定义 3 伪随机序列发生器 PRSG[17], 对于密钥长度 m 和序列长度 n, 伪随机序列发生器 PRSG 是一个确定的多项式时间算法: $g:\{-1,1\}^m\to\{-1,1\}^n$ (本文所有的二进制序列都是双极性的), 它产生二进制的伪随机序列集合 $\boldsymbol{V}=\{V_n=g(s)\}$ 与 $\boldsymbol{U}$ 在计算上不可区分, 其中 s 是伪随机序列发生器 g 的种子–密钥. 我们用 V_n(或者用 V) 来表示一个长度为 n 的 PRSG 伪随机序列.

根据上述定义, PRSG 序列是安全的. 这是本文研究水印安全性的基础和假设前提. 给定一个 PRSG 序列 V, 对于任意正多项式 $p(\ )$ 和任意的多项式时间算法 A, 所有足够大的 n, 成功找到 V 的密钥的概率是 $\Pr[A(V)]$. $\boldsymbol{V}$ 的势 [9,17] 小于或者等于 2^m. 本文假设 $2^m\ll 2^n$, 对于任意序列 $\bar{V}\notin\boldsymbol{V}$, $\Pr[A(\bar{V})]=0$. 由于 U 是均匀分布序列, $U\in\boldsymbol{V}$ 的概率 $\leqslant 2^m/2^n\ll 1$. 根据 (2) 式有

$$\begin{aligned}\Delta(n)&=\Pr[A(V)]-\Pr[A(U)]\\&\approx\Pr[A(V)]-(2^m/2^n)\times\Pr[A(V)]\\&\approx\Pr[A(V)]\\&<1/p(n).\end{aligned} \tag{3}$$

定义 $\Delta(n)$ 为攻击者具有的能将 V 与 U 区分开的能力优势, 所以 $\Delta(n)$ 或者 $\Pr[A(V)]\approx\Delta(n)$ 度量了 PRSG 序列 V 的安全性. 根据 (3) 式, $\Pr[A(V)]$ 和 $\Delta(n)$ 在 n 增大时是可忽略的, 所以 PRSG 序列是安全的. 当给定一个 PRSG, $\Pr[A(V)]$ 或者 $\Delta(n)$ 也就可以给定了. PRSG 的例子有 Blum PRSG, RSA PRSG 和离散对数 PRSG[17].

以上仅讨论了攻击者 A 的输入是无噪的 PRSG 序列, 且定义攻击者 A 只接受无噪 PRSG 序列作为输入的情况. 现将攻击者 A 的概念加以推广, 得到攻击者 B. 攻击者 B 的输入可以是无噪的, 也可以是可能含噪的 PRSG 序列. 后一种情况类似于我们将在第 3 节中提到的对量化水印算法的模糊攻击模式. 给定一个序列 $\tilde{V}\in\{-1,+1\}^n$, 它是通过对 V 加噪得到的, 即 $\tilde{V}=V+N$, 其中 N 是二进制加性噪声, 这里相当于误码序列. 攻击者 B 对 $\tilde{V}$ 进行攻击, 试图将 $\tilde{V}$ 和含噪真随机序列 $\tilde{U}=U+N$ 区分开.

下面分析攻击者 B 的攻击过程. 对 B 而言, 输入序列 $\tilde{W}$ 既可以是 $\tilde{V}=V+N$, 也可以是 $\tilde{U}=U+N$. 给定一个 PRSG 序列发生器 $g:\{-1,1\}^m\to\{-1,1\}^n$, 对于任意的攻击算法 B 和作为 B 的输入序列 $\tilde{W}$, $B(\tilde{W})$ 输出 "1" 表示输入 $\tilde{W}$ 等于 $\tilde{V}$; 反之, 输出 "0" 表示输入 $\tilde{W}$ 等于 $\tilde{U}$. B 要同时通过以下两项检验时, $B(\tilde{W})$ 才输出 "1". 1) 合法性检验: B 能找到一个密钥 s 来生成 PRSG 序列 $W'\{w'_i|1\leqslant i\leqslant n\}$; 2) 可检性检验: 检测统计量 $w_d=\tilde{W}W'/\sqrt{n}=\sum_{i=1}^n\tilde{w}_iw'_i/\sqrt{n}\geqslant a\sqrt{n}$, a 为归一化的检测阈值, 或者说 $\tilde{W}$ 和 W' 的归一化相关值 $\tilde{W}W'/n$ 大于 $a=1-2P_{\rm e}$, $P_{\rm e}$ 表示关于 $\tilde{W}$ 和 W' 间的误码率. 用 $\tilde{\Delta}(n)=\Pr[B(\tilde{V})]-\Pr[B(\tilde{U})]$ 来表示

攻击者 B 具有的能将 $\tilde{V}$ 和 $\tilde{U}$ 区分开来的能力优势. 我们注意到, 噪声不会提高攻击者区分 PRSG 序列 $\tilde{V}$ 和 $\tilde{U}$ 的能力, 所以有 $\tilde{\Delta}(n) \leqslant \Delta(n)$, 这里 $\Delta(n)$ 是 (3) 式中所定义的. 为此提出下面的有关二进制加噪 PRSG 序列的安全定理.

定理 1(二进制加噪 PRSG 序列安全定理) 给定一个 PRSG 序列发生器 $g : \{-1,1\}^m \to \{-1,1\}^n$ 和双极性序列 $\tilde{W}\{\tilde{w}_i|1 \leqslant i \leqslant n\}$. $\tilde{W} = \tilde{V} = V + N$ 或者 $\tilde{W} = \tilde{U} = U + N$, 式中 N 为随机二进制噪声 (误码序列). 对于任意的概率多项式算法 B, $B(\tilde{W})$ 输出 "1" 当且仅当 B 能找到一个密钥 s 来生成 PRSG 序列 W' 且检测统计量

$$w_d = \tilde{W}W'/\sqrt{n} = \sum_{i=1}^{n} \tilde{w}_i w'_i/\sqrt{n} \geqslant a\sqrt{n}. \quad (4)$$

这里 a 为预先定义的归一化检测阈值. 对于任意的正多项式 $p(\,)$ 和充分大的 n 时, 有

$$\begin{aligned}\tilde{\Delta}(n) &= \Pr[B(\tilde{V})] - \Pr[B(\tilde{U})] \\ &\approx \Pr[A(V)] < 1/p(n). \end{aligned} \quad (5)$$

$\Pr[A(V)]$ 为 (3) 式中 PRSG 序列的安全性度量. 如果当 $n \geqslant 1.39 \times m/a^2$ 或者 $m \leqslant 0.72 \times a^2 \times n$, 则

$$\Pr[B(\tilde{V})] \approx \Pr[A(V)]. \quad (6)$$

有关定理 1 的证明请见附录 A. 定理 1 告诉我们当 $n \geqslant 1.39 \times m/a^2$ 时, 加噪 PRSG 序列 $\tilde{V}$ 的安全性和未加噪的 PRSG 序列 V 相等, $n \geqslant 1.39 \times m/a^2$ 是 Alice 声称 $\tilde{W} = \tilde{V} = V + N$ 的非协议模糊性的充分必要条件 (本文默认假设 m 足够大, 能抗对密钥的迭代搜索攻击, 比如 m 大于 80). 只有生成 V 的 Alice 才可以声称她生成了 $\tilde{V}$; 不知道 V 的生成密钥的 Bob 是不可能通过计算来声称他生成了 $\tilde{V}$. 对于许多 PRSG 序列, 满足 $\Pr[A(V)] < 2^{-(m-1)}$. 因而, 当 $\tilde{W}$ 用作水印, (6) 式可以用来估计水印算法的协议模糊攻击成功概率.

定义 4 一般的盲水印算法可以用一个三元组 $W = \langle G, \varepsilon, \partial\rangle$ 来表示, 其中 G, ε, ∂ 分别是多项式时间算法.

1) G 是水印生成过程, 由可公开的 PRSG 序列发生器 g 生成长度为 n 的水印序列 W.

2) $\varepsilon(I, W, \theta)$ 是水印嵌入算法, 以原始媒体作品 I 和水印 W 作为输入, 得到加水印之后的媒体作品 $\hat{I} = \varepsilon(I, W, \theta)$. 沿着某些选定的投影方向上的投影分量, 将水印嵌入到载体作品 I 中, 以保证水印的鲁棒性和不可见性. θ 作为辅助输入, 既可以是投影方向, 也可以是嵌入水印时使用的密钥 (不同于水印的生成密钥) 等.

3) $\partial(\tilde{I}, W, \theta)$ 是检测算法, 用来确定是否可以将给定的水印 W 从待测的随机媒体作品 $\tilde{I}$ 中检测出来. 如果在 $\tilde{I}$ 中检测到水印 W, 则检测算法 ∂ 输出一个 true; 反之, 则输出一个 false. 对于任意的媒体作品 I 和水印 W, 本文要求满足 $\partial(\varepsilon(I, W, \theta), W, \theta)$ =true.

水印的检测采用假设检验. 假设 H_0: $\tilde{I} = I + N$ 或者假设 H_1: $\tilde{I} = \hat{I} + N$, 这里 N 是加性随机噪声. 假设 H_0 时 $\partial(\tilde{I}, W, \theta)$ =true 的概率为虚警率; 假设为 H_1 时 $\partial(\tilde{I}, W, \theta)$=false 的概率为漏警率 (也称漏检率).

4) 要声称拥有加水印媒体作品 $\tilde{I}$ 的版权, 要求 Alice 同时通过两项检验: i) 利用 G 生成序列 W, 也就是证明 W 确实是由公开发生器 g 生成的, 这是合法性检验; ii) 利用检测函数 $\partial(\,)$ 来证明 $\partial(\tilde{I}, W, \theta)$ 的输出是一个 true, 这是可检性检验.

定义 5 协议模糊攻击 (也叫可逆协议攻击) $W = \langle G, \varepsilon, \partial\rangle$ 是符合上述定义的水印算法, $\tilde{I}$ 为加水印的任意媒体作品. 概率多项式时间算法 [17] $B(\tilde{I}, \theta)$ 是对 W 的协议模糊攻击. 当 B 调用 G 生成一个序列 W'(可能不同于嵌入水印 W) 来通过合法性检验, 且调用 $\partial(\tilde{I}, W', \theta)$ 输出一个 true 来通过可检性检验时, 则 B 输出 "1", 其概率称为协议模糊攻击成功概率 F_{amb}; 否则 B 输出 "0", 其概率等于 $1 - F_{\text{amb}}$. 本文默认假设检测算法 $\partial(\,)$ 中的 θ 是已知的, 因而在下文将其忽略.

协议模糊攻击成功概率 $F_{\text{amb}} = \Pr[B(\tilde{I})]$ 是指 B 同时通过合法性检验和可检性检验的概率. 概率 F_{amb} 在 B 选取的所有水印 W', 所有的可能输入 $\tilde{I}$ 和算法 B 内部随机掷币等条件下计算. 用 F_{amb} 表示水印算法 $W = \langle G, \varepsilon, \partial\rangle$ 的可逆转概率. 当 F_{amb} 是可忽略时, 我们就认为水印算法是不可逆的或者是非协议模糊的.

在第 3, 4 节中, 我们将会证明和定量分析扩频水印和量化水印方法的非协议模糊性.

3 量化水印算法的不可逆转性和鲁棒性

一般基于量化的水印嵌入和提取方法 (或广义的关系嵌入嵌入和提取方法)$W=\langle G,\varepsilon,\partial\rangle$ 可以描述如下 [4,20,21]. 下述非协议模糊水印框架是基于这个一般描述而言.

嵌入: 用基于量化的方法将水印 W 嵌入到原始载体 I 中获得加水印媒体 [4,20~22].

$$\hat{I}=\varepsilon(I,W). \tag{7}$$

提取: 给定加水印的多媒体作品, $\tilde{I}$ 和水印 W, 检测算法 $\partial(\tilde{I},W)$ 输出一个 true 当且仅当检测统计量 w_d 满足如下如下条件:

$$w_d=\frac{\tilde{W}W}{\sqrt{n}}=\frac{\sum_{i=1}^{n}\tilde{w}_i w_i}{\sqrt{n}}\geqslant \mathrm{Th}. \tag{8}$$

这里 $\tilde{W}=\mathrm{extract}(\tilde{I})$ 是从 $\tilde{I}$ 中所提取的水印, w_d 度量了原始水印 W 和提取水印 $\tilde{W}$ 的相关性, Th 是检测阈值. 当 $\tilde{I}=\hat{I}$, $\tilde{W}=W, w_d$ 取最大值 $\sqrt{n}$. 一般来说, $\tilde{W}$ 可能受到加性噪声 N 污染. 假设由此导致的 $\tilde{W}$ 与 W 间的误码率为 P_e. 根据 (8) 式, $w_d=(1-2P_{\mathrm{e}})\sqrt{n}$, 如果 $P_{\mathrm{e}}\leqslant 0.22$, 那么 $w_d\geqslant 0.56\sqrt{n}$, 选 $\mathrm{Th}=0.56\sqrt{n}$, 则漏检率为 $fn=0$, 所以该水印算法可以是很稳健的 [3,20].

检测阈值 Th 是一个很重要的参数, 它决定了水印系统中的漏检概率和虚警概率. 以往研究讨论协议模糊性时, 都假定 Th 是一个固定的常数 [2,10,11], 然后提出一些被认为有效的迭代搜索攻击, 或者对于 Th 的选择不予讨论 [12]. 但是, 假定 Th 是固定的常数是不合理的. 为了对抗模糊攻击, 本文提出取 Th 是一个以水印长度 n 为自变量的函数:

$$\mathrm{Th}=(1-2P_{\mathrm{e}})\sqrt{n}=a\sqrt{n}, \tag{9}$$

这里 $a=(1-2P_{\mathrm{e}})$ 是归一化的检测阈值.

下面分析量化水印算法的协议非模糊性. 给定加了水印的多媒体作品 $\tilde{I}$ 和从中提取出来的 $\tilde{W}$, 并利用 B 去攻击它. 当且仅当 B 利用发生器 g 得到的一个 PRSG 序列 W' 表明 W' 的确是由 PRSG 序列发生器 g 生成的, 且检测统计量为 $w_d=\tilde{W}W'/\sqrt{n}\geqslant a\sqrt{n}$ 时, B 输出才为一个 "1". 根据定理 1 可知, 当 $n\geqslant 1.39\times m/a^2$ 时, 有

$$F_{\mathrm{amb}}=\Pr[B(\tilde{W})]\approx\Pr[A(V)]. \tag{10}$$

根据 (3) 式, B 输出 "1" 的概率 F_{amb} 是可以忽略的, 因而所提出的量化水印算法可证明是非协议模糊的.

在实际应用中, 可以选取合适的参数, 如 m,n 和 a, 然后定量估计算法的可逆转概率. 选用 Blum PRSG, 参数 n=1024, a=0.56, m=80$<0.72\times a^2\times n$=231.5, 那么就有 $F_{\mathrm{amb}}\approx\Pr[A(V)]<2^{-(m-1)}=2^{-79}$, 所以水印算法是计算上不可逆的. 作为例子, 我们用上述参数实现了一个实际的量化水印方法 [18]. 注意因为符号标记的不同, 文献 [18] 中 C_0=0.28 等价于本文的 a=0.56. 水印是不可见的以及加水印图像的视觉质量很好. 实验结果表明水印能对抗 JPEG 压缩, 中值滤波 (3×3, 5×5), 裁剪和幅度缩放攻击 [18].

4 扩频盲水印算法的不可逆转性

一般扩频盲水印算法的水印嵌入和检测方法可以描述如下. 扩频水印算法的水印嵌入函数为 [2,10]

$$\hat{I}_i=I_i+\alpha_i(I_i)w_i. \tag{11}$$

这里 $\alpha_i(I_i)$ 或者下文中简称的α_i, 决定了水印的嵌入强度, 是一个以 I_i 为自变量的函数. $W\{w_i\}$ 是二进制 PRSG 序列水印, $I\{I_i\}$ 根据具体的上下文表示载体作品或者是进行变换之后的系数, 比如 DWT (discrete wavelet transform) 或者是 DFT (discrete Fourier transform) 变换后得到的系数, $\hat{I}\{\hat{I}_i\}$ 是加了水印的媒体或者是它变换后的系数.

对于测试图像 $\tilde{I}$, 当且仅当 $w_d=\frac{\sum_{i=1}^{i=n}Y_i}{V_y\sqrt{n}}=\frac{M_y\sqrt{n}}{V_y}\geqslant\mathrm{Th}$ 时, 检测函数 $\partial(\tilde{I},W)$ 才输出一个 true. 这里 $Y_i=\tilde{I}_i w_i$, M_y 和 V_y^2 分别是对 Y_i 采样所得的均值和方差, 记为 $M_y=\frac{\sum_{i=1}^{i=n}Y_i}{n}$, $V_y^2=\frac{\sum_{i=1}^{i=n}(Y_i-M_y)^2}{n-1}$, Th 是预先定义的检测阈值.

在 H_0 假设 (未嵌入水印 W 下), 当 n 充分大时, w_d 的分布近似于零均值和单位方差的正态分布, 记作 $w_d\sim$ Gauss(0,1). 而在 H_1 假设 (已经嵌入水印 W) 下, w_d 的分布近似于均值为 μ 和单位方差的正态分布 [10], 记作 $w_d\sim$ Gauss(μ,1). 因此有

$$M_y\approx E(Y_i)=E(\alpha_i)+E(N_i w_i)=E(\alpha_i), \tag{12}$$

$$V_y^2\approx D(Y_i)=[E(I_i^2)+\sigma_N^2]-E^2(\alpha_i). \tag{13}$$

这里 $N\{N_i\}$ 表示加性随机噪声. 为了不失一般性,

我们假定 $E(N_i w_i)=0$, 则

$$\mu=\frac{E(\alpha_i)\sqrt{n}}{\sqrt{E(I_i^2)+\sigma_N^2-E^2(\alpha_i)}}=\alpha\sqrt{n}, \tag{14}$$

$$\alpha=\frac{E(\alpha_i)}{\sqrt{E(I_i^2)+\sigma_N^2-E^2(\alpha_i)}}. \tag{15}$$

通常我们归一化 $E(I_i^2)=1$.

类似于量化水印算法, 为了对抗协议模糊攻击, 我们令 Th= $a\sqrt{n}$, 这里 a 是一个比 α 小的正常数, α 可根据 (15) 式求得. 例如, 如果令 $E(\alpha_i)$=0.2, 噪声方差 σ_N^2=1, 即噪声的强度和原始媒体作品一样, 则可得 α=0.14. 此时, 水印依然能以很低的漏检率被检测出来 (下文中会证明 fn 是可以忽略的), 可以认为水印算法是很稳健的. 对于 a, 最佳选择是令 $a=0.5\alpha$, 因为此时水印算法总的错误率 (包括 fp 和 fn) 最小. 虚警率 fp 为

$$fp=\frac{1}{a\sqrt{2\pi n}}\mathrm{e}^{-a^2n/2}, \tag{16}$$

当满足 $a\sqrt{n}>4$[23] 时, 上式成立. 漏检率为

$$fn=\frac{1}{(\alpha-a)\sqrt{2\pi n}}\mathrm{e}^{-(\alpha-a)^2n/2}. \tag{17}$$

当满足 $a\sqrt{n}>4$ 时, 上式成立. 当选择检测阈值 $\mathrm{Th}=a\sqrt{n}$ 时, fp 和 fn 都随着 n 的增大而按指数规律迅速减少成为可忽略的. 由此可知, 选择合理的检测阈值可以使得 fp 和 fn 都随着 n 的增大而按指数规律迅速减少, 而这一点在以往的研究中被忽略了.

下面分析扩频盲水印算法的非协议模糊性. 随机给定一个加了水印的多媒体作品 $\tilde{I}$, 利用 B 去攻击它. 当且仅当 B 利用发生器 g 得到的一个 PRSG 序列 W', 且检测统计量为 $w_d=\frac{\sum_{i=1}^{i=n}Y_i}{V_y\sqrt{n}}=\frac{M_y\sqrt{n}}{V_y}\geqslant a\sqrt{n}$ 时, B 才输出一个 "1". 这里 $Y_i=\tilde{I}_i w_i'$. 我们可以根据 (11) 式和表达式 $\tilde{I}=\hat{I}+N$, 把 $\tilde{I}$ 看作是 W 的广义噪声版本 (W 混合了加性噪声或者乘性噪声). 和广义加噪 PRSG 序列 $\tilde{I}$ 的安全性密切相关的问题是: B 的成功概率 $\Pr[B(\tilde{I})]$ 是否可忽略, 因而攻击者 Bob 不能声称对 $\tilde{I}$ 拥有版权?

和定理 1 类似, 很容易证明下面的广义加噪 PRSG 序列安全定理. 定理 2 将定理 1 中的二进制检测序列 $\tilde{W}\in(-1,+1)^n$ 拓宽为随机数序列或者是随机的媒体作品 $\tilde{I}$.

定理 2(广义加噪 PRSG 序列安全定理) 给定一个随机序列 $\tilde{I}\{\tilde{I}_i|1\leqslant i\leqslant n\}$, 和一个用 PRSG 序列发生器 $g:\{-1,1\}^m\to\{-1,1\}^n$ 生成的序列 $W\{w_i|1\leqslant i\leqslant n\}$. $\tilde{I}=\hat{I}+N$, 这里 N 是加性随机噪声. 对于任意的算法 B, 当且仅当 B 利用发生器 g 得到的一个 PRSG 序列 $W'\{w_i'|1\leqslant i\leqslant n\}$, 且检测统计量为 $w_d=\frac{\sum_{i=1}^{i=n}Y_i}{V_y\sqrt{n}}=\frac{M_y\sqrt{n}}{V_y}\geqslant a\times\sqrt{n}$ 时, $B(\tilde{I})$ 才输出一个 "1". 这里 $Y_i=\tilde{I}_i w_i$, M_y 和 V_y^2 分别是对 Y_i 采样所得的均值和方差, 记为 $M_y=\frac{\sum_{i=1}^{i=n}Y_i}{n}$, $V_y^2=\frac{\sum_{i=1}^{i=n}(Y_i-M_y)^2}{n-1}$, a 是预先定义的归一化阈值. 如果 $n\geqslant 1.39\times m/a^2$ 或者 $m\leqslant 0.72\times a^2\times n$, 则有

$$\Pr[B(\tilde{I})]\approx\Pr[A(V)]. \tag{18}$$

这里 $\Pr[A(V)]$ 为 (3) 式中用来度量 PRSG 序列安全性的概率.

定理 2 的证明类似于定理 1. 根据定理 2 和 PRSG 序列安全性可知, $F_{\mathrm{amb}}\approx Pr[A(V)]$ 是可以忽略的, 因此扩频水印算法是可以抵抗协议模糊攻击的.

我们选用参数 n=4096, a=0.165, $m=80<0.72\times a^2\times n$ 实现了一个实际的安全的扩频水印算法. 和上面讨论的量化水印算法很相似, 水印是不可见的以及加水印图像的视觉质量很好. 实验结果表明水印能对抗 JPEG 压缩, 中值滤波 (3×3, 5×5), 裁剪和幅度缩放攻击等.

5 结束语

本文基于安全伪随机序列发生器 (PRSG) 序列与真随机序列计算上不可区分性来研究数字媒体水印的非协议模糊问题. 本文的工作和主要成果如下:

1) 针对一个使用 PRSG 序列作为水印的算法的协议模糊攻击可以看成是针对一个加噪 PRSG 序列的攻击, 提出和证明了二进制加噪 PRSG 序列安全定理和广义加噪 PRSG 序列安全定理; 并指出当 $n\geqslant 1.39\times m/a^2$ 或者 $m\leqslant 0.72\times a^2\times n$ 时, 水印算法对抗协议模糊攻击的安全性等于其基于的 PRSG 序列的安全性.

2) 选择合理的检测阈值 Th= $a\sqrt{n}$ 和密钥长度 $m\leqslant 0.72\times a^2\times n$, 可以使得协议模糊攻击成功概率 ($F_{\mathrm{amb}}$) 和水印漏检概率随水印长度 n 增大迅速变为可忽略小的, 这样实际的量化水印算法和扩频盲水印算法就可以同时具有非协议模糊性和鲁棒性.

3) $n \geqslant 1.39 \times m/a^2$ 是水印算法非协议模糊性的充要条件, 得到了如 (10) 和 (18) 式所示的水印算法安全性计算的分析解. 这些可以用来指导设计一个实际的不可逆水印算法和定量评价水印算法的非协议模糊性.

在结合可信任第三方 (TTP) 的数字版权管理系统中, Alice 要想证明对媒体作品 $\tilde{I}$ 拥有版权主张, 可提供一对 ($W_{\rm A}$, $K_{\rm A}$) 给 TTP, 并且证明 $W_{\rm A}$ 是从媒体作品 $\tilde{I}$ 检测出到的. 可信任第三方将通过公用的 PRSG 序列发生器 g 来验证 $W_{\rm A} = g(K_{\rm A})$ 是否成立 (请注意本文提出的水印框架中没有要求 TTP 参与). 然而在没有结合可信任第三方 (TTP) 的数字版权管理系统中, Alice 直接公开一对 $(W_{\rm A}, K_{\rm A})$ 来证明对媒体作品拥有版权主张, 这一行为可能会造成信息泄露, 导致随后的潜在成功攻击. 一种解决方法是利用所谓 "零知识协议" 来交互式的和有效的证明 [24]$W_{\rm A}$ 和 $K_{\rm A}$ 之间的关系, 而同时又使得 ($W_{\rm A}$, $K_{\rm A}$) 对保持秘密. 本文提出的原理也可以推广到采用广义的单向函数 (比如 Hash 等) 作为水印序列发生器的水印算法的非协议模糊性, 这是因为当且仅当单向函数存在时, PRSG 发生器才存在 [17]. 进一步的研究工作将在今后展开.

参考文献

1 Moulin P, Koetter R. Data hiding codes. Proc IEEE, 2005, 93(12): 2079–2082

2 Cox I, Kilian J, Leighton T, et al. Secure spread spectrum watermarking for multimedia. IEEE Trans Image Process, 1997, 6(12): 1837–1687

3 Zhao Y. Geometrically robust video water-marking based on wavelet transform. Sci China Ser F-Inf Sci, 2006, 49(3): 313–327

4 Wu M, Liu B. Data hiding in images and video, Part I: fundamental issues and solutions. IEEE Trans Image Process, 2003, 12(6): 685–695

5 Chen B, Wornell G W. Quantization index modulation: a class of provably good methods for digital watermarking and information embedding. IEEE Trans Inform Theor, 2001, 47(4): 1423–1443

6 Craver S, Memon N, Yeo B L, et al. Resolving rightful ownerships with invisible watermarking techniques: limitations, attacks, and implications. IEEE J Sel Area Comm, 1998, 16(4): 573–586

7 Adelsbach A, Katzenbeisser S, Sadeghi A R. On the insecurity of noninvertible watermarking schemes for dispute resolving. In: Proceedings of International Workshop on Digital Watermarking 2003. LNCS, Vol 2939. Berlin: Springer-Verlag, 2003. 374–388

8 Adelsbach A, Katzenbeisser S, Veith H. Watermarking schemes provably secure against copy and ambiguity attacks. In: Proceedings of ACM Conference on Digital Right Management 2003. New York: ACM, 2003. 111–119

9 Li Q, Chang E C. On the possibility of non-invertible watermarking schemes. In: Proceedings of International Workshop on Information Hiding 2004. LNCS, Vol 3200. Berlin: Springer-Verlag, 2004. 13–24

10 Zeng W, Liu B. On resolving rightful ownerships of digital images by invisible watermarks. In: Proceedings 4th International Conference on Image Processing. Santa Barbara: IEEE, 1997. 552–555

11 Ramkumar M, Akansu A. Image watermarks and counterfeit attacks: some problems and solutions. In: Proceedings of Symposium on Content Security and Data Hiding in Digital Media. Newark: NJIT, 1999. 102–112

12 Sencar H T, Memon N. Watermarking and ownership problem: a revisit. In: Proceedings of ACM Conference on Digital Right Management 2005. Alexandria: ACM, 2005. 93–101

13 Qiao L, Nahrstedt K. Watermarking methods for MPEG encoded video: towards resolving rightful ownership. In: Proceedings of IEEE International Conference on Multimedia Computing and Systems. New York: IEEE, 1998. 276–285

14 Swanson M D, Kobaysshi M, Tewfik A H. Multimedia data embedding and watermarking technologies. Proc IEEE, 1998, 86(6): 1064–1087

15 Zeng W, Liu B. A statistical watermark detection technique without using original images for resolving rightful ownerships of digital images. IEEE Trans Image Process, 1999, 8(11): 1534–1548

16 Ramkumar M, Akansu A N. A robust protocol for proving ownership of multimedia content. IEEE Trans Multimedia, 2004, 6(3): 469–478

17 Goldreich O. Foundations of Cryptography—Basic tools. Cambridge: Cambridge university press, 2001. 101–180

18 Kang X, Shi Y Q, Huang, J. Achieving non-ambiguity of quantization based watermarking. In: Proceedings of SPIE: Security, Steganography, and Watermarking of Multimedia Contents 2006. Vol 6072. New York: SPIE, 2006. 60720F-1–60720F-11
19 Katzenbeisser S, Veith H. Securing symmetric watermarking schemes against protocol attack. In: Proceedings of SPIE, Security and Watermarking of Multimedia Content 2002. Vol 4675. New York: SPIE, 2002. 260–268
20 Kang X, Huang J, Shi Y Q, et al. A DWT-DFT composite watermarking scheme robust to both affine transform and JPEG compression. IEEE Trans Circ Syst Vid, 2003, 13(8): 776–786
21 Perez-Freire L, Comesana-Alfaro P, Perez-Gonzalez F. Detection in quantization-based watermarking: performance and security issues. In: Proceedings of SPIE: Security, Steganography, and Watermarking of Multimedia Contents 2005. Vol 5681. New York: SPIE, 2005. 721–733
22 Eggers J J, Bäuml R, Tzschoppe R, et al. Scalar Costa scheme for information embedding. IEEE Trans Signal Proces, 2003, 51(4): 1003–1019
23 Proakis J G. Digital Communications. 4th ed. New York: McGraw-Hill Companies Inc, 2004. 200–306
24 Li Q, Chang E C. Zero-knowledge watermark detection resistant to ambiguity attacks. In: ACM Multimedia Security Workshop. New York: ACM, 2006. 158–163

附录 A

定理 1 的证明.

1) 证明 $\tilde{\Delta}(n) < 1/p(n)$.

用反证法. 假设 $\tilde{\Delta}(n)$ 是不可忽略的, 给定一个 V 和一个 U, 各自用 N 进行加噪, 即可得 $\tilde{V} = V+N$ 和 $\tilde{U} = U+N$. 因为 $\tilde{\Delta}(n)$ 是不可忽略的, 所以必然存在能将 $\tilde{V}$ 和 $\tilde{U}$ 区分开的攻击者 B, 也就能 V 和 U 区分开, 但这与 PRSG 的定义相矛盾, 即假设不成立, 因而有 $\tilde{\Delta}(n) < 1/p(n)$.

2) 证明 $\tilde{\Delta}(n) \approx \Pr[A(V)]$.

由 1) 的证明可知, $\tilde{\Delta}(n)$ 也是用来度量 PRSG 序列的安全性的. 下面我们分析当 $\tilde{W} = \tilde{V} = V + N$ 的情况.

假设存在序列 W' 是一个 PRSG 序列且能通过可检性测试的联合概率为 $\Pr(W')$. 记 $\Delta(n)$ 为攻击者 A 攻击 W' 具有的能力优势, 那么 $\tilde{\Delta}(n) = \Pr(W')\Delta(n) \leqslant \Delta(n)$. 如果噪声 N 在 a 允许的范围之内, 则 $\Pr(W')$=1. 由此可得 $\tilde{\Delta}(n) = \Delta(n) \approx \Pr[A(V)]$.

3) 证明 $\Pr[B(\tilde{V})] \approx \Pr[A(V)]$.

我们考虑两种情况.

情况一, 假设输入的 $\tilde{W}$ 为 $\tilde{U} = U + N$, 则 $\Pr[B(\tilde{W})] = \Pr[B(\tilde{U})]$. $\tilde{W}$ 是由真随机序列 U 加噪得到的. 假设 W' 是一个特殊的 PRSG 序列, 根据中心极限定理, 当 n 充分大时, 两个独立生成的序列 $\tilde{W}$ 和 W' 的内积 $\tilde{W}W'$ 的分布为 Gauss 分布. 则 $w_d = \tilde{W}W'/\sqrt{n}$ 也是 Gauss 分布, 且均值 μ=0 和方差 $\sigma^2 = 1$, 记为 Gauss(μ, σ^2) = Gauss(0,1). 给定的一个特殊的 PRSG 序列 W', 能找到和它的密钥的概率为 $\Pr[A(W')] = \Pr[A(V)]$. 因此存在一个特殊的 PRSG 序列 W' 使得 $w_d > a\sqrt{n}$ 和找到它的密钥的联合概率为

$$\begin{aligned} F_j &= \Pr(w_d > a\sqrt{n}) \times \Pr[A(W')] \\ &\approx \frac{1}{a\sqrt{2\pi n}} \mathrm{e}^{-a^2 n/2} \times \Pr[A(V)]. \end{aligned}$$

又因为集合的势 $\boldsymbol{V} \leqslant 2^m$, 所以概率 $\Pr[B(\tilde{U})]$ 的计算可以用下式表示

$$\begin{aligned} \Pr[B(\tilde{U})] &\leqslant \sum_{j=1}^{2^m} F_j = \frac{1}{a\sqrt{2\pi n}} \mathrm{e}^{(0.693m - a^2 n/2)} \times \Pr[A(V)] \\ &= F(n) \times \Pr[A(V)]. \end{aligned}$$

这里 $F(n) = \frac{1}{a\sqrt{2\pi n}} \mathrm{e}^{(0.693m - a^2 n/2)}$.

情况二, 假设输入的 $\tilde{W}$ 为 $\tilde{V} = V + N$, 则 $\Pr[B(\tilde{W})] = \Pr[B(\tilde{V})]$. 由此可得

$$\begin{aligned} \Pr[B(\tilde{V})] &= \tilde{\Delta}(n) + \Pr[B(\tilde{U})] \\ &\approx \Pr[A(V)] + \Pr[B(\tilde{U})] \\ &\leqslant (1 + F(n)) \times \Pr[A(V)]. \end{aligned}$$

本文我们默认假设 m 足够大能避免对密钥的迭代搜索攻击, 比如大于或等于 80. 当 $n \geqslant 1.39 \times m/a^2$ 即 $m \leqslant 0.72 \times a^2 \times n$ 时, $F(n) \leqslant \frac{1}{a\sqrt{2\pi n}} = \frac{1}{\sqrt{8.71 \times m}} = 0.038$, 可见 $F(n)$ 随 n 增大而迅速变小成为可忽略的, 因而可得 $\Pr[B(\tilde{V})] \approx \Pr[A(V)]$. 证毕.