抗协议攻击的数字版权保护安全方案*

刘立刚, 陈晓苏, 胡 蕾
(华中科技大学计算机科学与技术学院, 湖北 武汉 430074)

摘 要: 随着数字水印技术在数字版权保护方面的应用, 对数字水印的攻击也日益严重, 早期的水印攻击主要是针对水印算法鲁棒性的攻击, 近几年来出现了针对数字版权保护过程中协议的攻击。通过对协议攻击的描述、定义和分析, 将数字签名等技术应用到数字版权保护方案中, 提出一个基于 PKI (public key infrastructure) 的数字版权保护安全方案, 分析讨论了该方案的安全性。

关键词: 数字版权保护; 数字水印; 协议攻击; 数字签名; PKI (public key infrastructure)
中图分类号: TP391.41 文献标识码: A 文章编号: 0529-6579 (2004) S2-0083-04

随着计算机和网络技术的飞速发展, 数字图像、音频和视频等多媒体数字作品越来越需要一种有效的版权保护方法, 此外通信系统在网络环境下的信息安全问题也日益显露出来。数字水印技术为上述问题提供了一个有效的解决方案。

为了避免水印在对数字作品正常或恶意操作中被损坏或移除, 对数字水印技术的研究主要集中在水印算法的鲁棒性上。鲁棒的数字水印算法能够抵抗各种针对算法的攻击, 但是在版权保护过程中, 鲁棒的数字水印算法是不能够抵抗针对数字水印的协议攻击。

针对数字水印的协议攻击是在协议层破坏版权保护, 它不是移除水印, 而是添加一个伪造的水印来干扰版权的认证。显然, 要实现数字版权保护, 不仅要求数字水印能够抵抗各种针对算法的攻击, 而且要求能够抵抗各种协议攻击。本文讨论了在数字作品中引入数字签名来抵抗协议攻击, 并将 PKI (public key infrastructure) 应用到版权保护过程中, 构建安全的数字版权保护方案。

1 数字水印与协议攻击

1.1 通用数字水印系统

用于数字作品版权保护的数字水印方案能够将附加的信息嵌入到一个数字作品 (载体) 中, 并且这些附加信息能够被检测或提取出来, 不会被移除。一个典型的数字水印方案主要包括密钥生成 GenKey, 水印嵌入 Embed 和水印检测 Detect 3 个部分。

GenKey () $\rightarrow K_{Emb}$ / K_{Dec};

Embed (W, WM, K_{Emb}) $=W'$;

Detect (W', W, WM, K_{Dec}) =TRUE

或 Detect (W', WM, K_{Dec}) =TRUE

其中, W 是原数字作品, WM 是水印, K_{Emb} 是水印嵌入的密钥, W' 是含有水印的数字作品, 它与 W 在感官上是一样的, K_{Dec} 是检测水印的密钥, 当 K_{Emb} 与 K_{Dec} 相同时, 这个水印系统是一个对称密钥水印系统; 当 K_{Emp} 与 K_{Dec} 不同时, 这个水印系统就是一个非对称密钥水印系统。

1.2 拷贝攻击

Kutter 等[1] 首先引入了术语 "拷贝攻击", 并叙述了一种实现该协议攻击的方法。给定一个合理的水印化作品 W_1' 和一个未加水印的目标作品 W_2, 首先对 W_1' 进行预处理获得原始作品的近似作品 W_1', 然后通过比较 W_1' 和 W_1 来估算嵌入的水印:

Compare (W_1', W_1) =WM

将估算出的水印嵌入到未加水印的作品 W_2 中从而得到水印作品 W_2':

Embed (W_2, WM, K) $=W_2'$

使得

Detect (W_2', W_2, K, WM) =TRUE

为了防止拷贝攻击, 一个有效的方法是使水印和作品的内容相关, 即使攻击者成功地进行了拷贝攻击, 即在 W_2 中检测到了攻击者后加入的水印, 但是攻击者加入的水印和 W_2 不相关, 由此就可以判断该水印是伪造的。

1.3 歧义攻击

歧义攻击又称为混淆攻击, 主要是指攻击者由

* 收稿日期: 2004-08-11
作者简介: 刘立刚 (1972 年生), 男, 博士生; E-mail: 2000_cs@163.com

水印化数字作品 W′构造出一个水印 WM′，同时也构造出所谓的原作品 W″和密钥 K′，即

Construct (W′) = (WM′, W″, K′)

使得在 W′中能够检测到水印 WM′，

Detect (W′, W″, K′, WM′) =TRUE

在歧义攻击中，同时存在伪造水印 WM′、伪造原始作品 W″、伪造密钥 K′和真实水印 WM、真实原始作品 W、密钥 K。要判定数字作品的所有权，必须在一个水印化媒体中所存在的几个水印里判断出原始水印。一种对策是采用时间戳(timestamps)技术。时间戳由可信的第 3 方提供，由此判断水印化媒体中水印加入的时间顺序，从而可以确定水印的真实性。另一种策略是采用签名技术[6]，使原始水印中包含版权所有者的签名。

1.4 倒置攻击

Craver 定义了另外一种协议攻击——倒置攻击，倒置攻击是歧义攻击的一个特例。如果发生版权所有者的原始作品中包含攻击者的水印，同时攻击者声称的原始作品中包含版权所有者的水印的情况，水印的嵌入顺序无从判断，从而也无法判断版权所有者。倒置攻击就试图制造这种情况。

版权所有者在其数字作品 W 中嵌入水印 WM_1，产生水印化的数字作品 W′，并发布。

Embed (W, WM_1, K) =W′

攻击者用自己的预处理水印 WM_2，加入到版权所有者的 W′中，然后得到 W″。

由于 W′和 W″看起来比较相似，预处理水印 WM_1 在 W″中仍然可以检测到。

然而，攻击者现在声称是 W′的版权所有者，攻击者为了证明他的观点，提供了水印 WM_2，产生 W″，证明 W″是他自己的原始作品，

Exact (Embed (W, WM_1, K), WM_2) =W″

由于提取 WM_2 不会妨碍 WM_1 的检测，可以重写这个等式：

Exact (Embed (W″, WM_2, K), WM_1) =W

表明攻击者的水印在版权所有者的原始作品中也能被检测到。倒置攻击没有任何证据表明数字水印加入的顺序。

2 版权保护安全方案

从对协议攻击的分析可以看出，要实现版权保护仅仅依赖数字水印技术是不够的，还需要结合其它安全技术。Craver[2] 和 Katzenbeisser[3] 分别提出将非对称水印和数字签名技术应用到数字水印中来解决协议攻击。

在开放的网络境下，数字版权保护体系中主要涉及到的实体包括：版权所有者 CH (copyright holder)、购买者 B (buyer)、版权认证中心 CC (copyright center) 和公钥基础设施 PKI。CH 是具有版权的实体，他可以发布和出售他具有版权的作品。B 是向 CH 购买作品的实体。CC 是一个值得 CH 和 B 信赖第 3 方，CC 给 CH 和 B 提供版权保护和水印认证等服务。而 PKI 则支持他们之间的安全的通信信道，密钥的分发，身份认证建立信任关系等等。各部分之间的关系如图 1 所示。

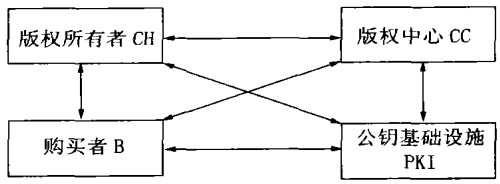

图 1 基于 PKI 的数字版权保护安全模型

Fig. 1 Digital copyright protection security model based on PKI

2.1 符号定义

为了更好地描述该安全方案，定义符号如下：W：原始数字作品 (Work)；W′：水印化数字作品；ID_W：作品标识号；WM：水印；WM′：伪造的水印；K：数字水印嵌入/检测所用密钥；Cert：数字证书 (certificate)；CH：版权所有者 (copyright holder)；ID_{CH}：版权所有者标识号；M：版权信息 (copyright message)；〈SK，PK〉：签名密钥对，其中 SK 为私钥，PK 为公钥；Sig(m, SK) = m_{sig}：对信息用私钥 SK 进行签名，签名为 m_{sig}；Ver (m_{sig}, m, PK) = {TRUE, FALSE}：用公钥 PK 验证信息 m 的签名 m_{sig} 的正确性，输出 {TRUE, FALSE}；B：作品购买者 (buyer)；ID_B：作品购买者标识号 (buyer-ID)；CC：版权中心 (copyright center)；PKI：公钥基础设施。

2.2 抗协议攻击的版权保护安全方案描述

2.2.1 证书分发

CH→PKI：申请 Cert；

PKI→CH：颁发 Cert，CH 获得一对签名密钥〈SK，PK〉，及证书中相应的 CH 的 ID_{CH} (证书 ID)。

B→PKI：申请 Cert；

PKI→B：颁发 Cert，B 得到 ID_B (证书 ID)；

根据 X.509 标准，用户数字证书主要包括主体信息，证书序列号，私钥对应的公钥，颁发者，有效期等，数字证书提供了对 CH 和 B 的身份认证功能。

2.2.2　水印嵌入机制　为了防止协议攻击，Adelsbach 提出“非可逆水印方案”[4]，由于这个方法的计算量很大，往往被误认为是“安全的”，该方案已被证明是不安全的了。

为了有效抵抗协议攻击，将水印嵌入过程与数字签名相结合，以达到更高的安全性，其过程如下：

CH：GenKey（）→K；

CH：Sig（$ID_{CH}+ID_B+M+W$，SK）$=M_{Sig}^{CH}$；

CH→CC：W，$WM=M_{sig}^{CH}+ID_{CH}+ID_B$，K；

CC：Ver（$(ID_{CH}+ID_B+M+W)_{sig}$，

$ID_{CH}+ID_B+M+W$，PK）＝｛TRUE，FALSE｝；

CC：W'＝Embed（W，WM，K）；

CC→CH：W'；

CH 创作完成作品后，就到 CC 去请求版权保护，即嵌入数字水印。首先，CH 生成嵌入水印的密钥 K，然后生成水印 WM。水印 WM 的生成过程如图 2 所示，基于安全的考虑 CH 用自己的私钥 SK 对 W、ID_{CH} 和 ID_B 一起签名，得到 M_{Sig}^{CH}。最后，CH 将 W、WM 和 K 发送给 CC，由 CC 完成水印的嵌入工作，CC 返回给 CH 已嵌入水印 WM 的数字作品 W'。CH 在对作品 W' 进行发布时，可以把水印中的组成部分 ID_B 略去，在对作品 W' 进行交易时，则必须将 ID_B 包含进去，用于实现版权跟踪。CC 与 CH 之间的通信安全均可以在 PKI 上实现。

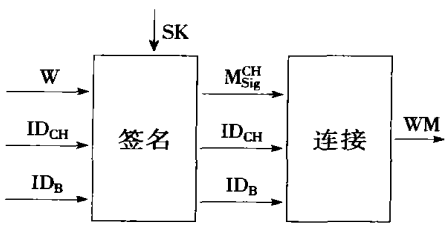

图 2　水印的构造过程

Fig. 2　The construction of watermarking

2.2.3　在线交易机制　当购买者 B 想要向 CH 购买作品 W' 时，B 首先要了解 CH 的身份以及他是否具有 W' 的版权，然后再进行交易。其过程如下：

CH→B：CH 的 Cert；

B→PKI：CH 的 Cert；

PKI→B：｛TRUE，FALSE｝；

if　TRUE

B→CC：W'，ID_{CH}；

CC→CH：请求 W、WM 和 K；

CH→CC：W、WM 和 K；

CC：验证水印的有效性，｛TRUE，FALSE｝；

if　TRUE

B→CH：ID_B；

CH→CC：W、WM、ID_{CH} 和 ID_B；

CC：W'＝Embed（W，WM，K）；

CC→CH：W'；

CH→B：W'；

B：停止交易；

else

B：CH 的 Cert 无效，身份验证失败，停止交易；

其中水印的构造过程和水印嵌入机制中的水印构造过程一样。

2.2.4　水印验证机制　水印验证是基于数字水印版权保护的关键，水印验证的正确性直接影响到版权的正确判断。

CH→CC：W、WM 和 K；

CC

Detect（W'，W，K，WM）＝｛TRUE，FALSE｝；

if TRUE

CC→PKI：ID_{CH}；

PKI→CC：PK；

CC：Ver（$(ID_{CH}+ID_B+M+W)_{sig}$，

$ID_{CH}+ID_B+M+W$，PK）＝｛TRUE，FALSE｝；

if　TRUE

ID_{CH}为合法版权所有者；

else

ID_{CH}为非法版权所有者；

else

水印检测失败；

在水印的验证过程中，不仅要检测水印的存在性，还要在检测出水印的基础上验证水印的合法性，即需要进一步 CH 对的身份进行验证，通过验证水印中 CH 的签名 $(ID_{CH}+ID_B+M+W)_{sig}$ 是否为 TRUE，如果为 FALSE，则可判定 CH 是盗版者。

3　安全性分析

攻击者要伪造盗版，需要对水印方案和签名系统进行攻击。对水印安全方案和签名系统的安全性分别进行分析。根据 Kerckhoff 原则，对攻击的分析是在假设攻击者对整个水印安全方案的所有知识都了解的基础上进行的。

3.1　抗拷贝攻击分析

假设设计的水印方案对拷贝攻击是不安全的，攻击者对所要拷贝的水印具备足够的知识，一种简单有效的攻击假设就是攻击者是数字媒体 W_1 的合法版权所有者，他仿照在对 W_1 嵌入水印的过程中

对数字媒体 W_2 进行拷贝攻击，其攻击步骤如下：

①生成水印嵌入/检测密钥 K；

②选取 W_1，具有 W_1 的合法版权；

③选取 W_2，攻击作品；

④生成嵌入 W_1 的水印

$WM_1 = Sig\ (ID_{CH} + M + W,\ SK) + ID_{CH}$；

⑤在 W_1 中嵌入水印

$W_1' = Embed\ (W_1,\ WM_1,\ K)$；

⑥对 W_2 进行拷贝攻击

$W_2' = Embed\ (W_2,\ WM_1,\ K)$。

攻击者只有在水印验证 $Ver\ ((ID_{CH} + ID_B + M + W_2,\ SK)_{sig},\ ID_{CH} + ID_B + M + W_2,\ PK) = TRUE$ 的情况下才攻击成功。由签名算法的安全性可知，该攻击方法是不可能成功的。

3.2 抗歧义攻击分析

攻击者选择一个水印化数字媒体 W'，假设存在算法 G 在输入 W' 后，输出伪造的原始数字媒体 W'' 和水印 WM'，使得对于任何密钥 K，$Detect\ (W',\ W'',\ WM'K) = TRUE$。

进一步，WM' 表示为：

$WM' = Sig\ (ID_{CH}' + ID_B + M + W'',\ SK) + ID_{CH}' + ID_B$

并且，

$Ver\ ((ID_{CH}' + ID_B + M + W'',\ SK)_{Sig},\ ID_{CH} + ID_B + M + W'',\ PK) = TRUE$

歧义攻击能成功。只有攻击者找到算法 G 和一个相关存在的伪造签名算法 FS，并使得伪造的水印检测和验证成功，才能完成攻击。显然，攻击者要得到 G 和 FS 是不可能的。

同理，倒置攻击是歧义攻击的特例，抗歧义攻击的方法同样适用于倒置攻击。

4 总 结

本文通过对歧义攻击，倒置攻击和拷贝攻击的描述、定义和分析，指出了已有的协议攻击解决方案的不足，提出了一种基于 PKI 的数字版权保护安全方案。并从抗协议攻击方面，分析了该方案的安全性。

参考文献：

[1] KUTTER M, VOLOSHYNOVSKIY S, HERRIGEL A. The watermark copy attack[C]. Security and Watermarking of Multimedia Contents Ⅱ, Proceedings of the SPIE, 2000: 371－380.

[2] CRAVER S, KATZENBEISSER S. Copyright protection protocols based on asymmetric watermarking: the ticket concept[M]. Communications and Multimedia Security Issues of the New Century, Kluwer Academic Publishers, 2001: 159－170.

[3] KATZENBEISSER S, VEITH H. Securing symmetric watermarking schemes against protocol attacks [C]. Security and Watermarking of Multimedia Contents Ⅳ, Proceedings of the SPIE, 2002: 260－268.

[4] ADEKSBACH A, KATZENBEISSER S, SADEGHI A R. On the insecurity of non-invertible watermarking schemes for dispute resolving [C]. International Workshop on Digital Watermarking (IWDW 2003), 2004: 355－369.

Digital Copyright Protection Security Schemes Against Protocol Attacks

LIU Li gang, CHEN Xiao su, HU Lei

(Department of Computer Science and Technology,
Huazhong University of Science and Technology, Wuhan 430074, China)

Abstract: The attacks on digital watermarking increasingly become serious, with the application of digital watermarking on digital copyright protection. The attacks focus on the robustness of watermarking algorithm early, but in the recent year, the attacks focus on the protocol about the process of digital copyright protection. Through the description, definition and analysis on the protocol attacks, digital signature is applied to digital copyright protection schemes. A digital copyright protection security scheme based on PKI is presented with the analysis of security.

Key words: digital copyright protection; digital watermarking; protocol attacks; digital signature; public key infrastructure (PKI)