数字水印的攻击方法 [1]

陈明奇 钮心忻 杨义先

(北京邮电大学信息安全中心 北京 100876)

摘 要 数字水印是近两年来出现的数字产品版权保护技术，目的是保护数字产品的合法拷贝和传播，是当前国际学术界的研究热点．数字水印包括算法设计和攻击技术两方面，它们是互相促进的对立面．本文简单讨论了数字水印的概念和应用，全面分析总结了当前数字水印的各种攻击方法，提出了数字水印下一步的发展方向，对改进和设计水印算法具有重要的指导作用．

关键词 知识产权保护，数字水印，水印攻击

中图号 TN911.72, TP391.41

1 前 言

Internet 和 Intranet 以多种新的手段和运作方式为商业、科研、娱乐等带来了许多机会。然而，这也使盗版者能以低廉的成本复制及传播未经授权的数字产品内容，出于对利益的考虑，数字产品的版权所有者迫切需要有办法解决知识产权 (Intellectual property rights) 保护的问题。依靠密码学的加密或置乱技术能保证数字产品内容的安全传送，同时，可作为存取控制和征收费用的手段。因为，只有用户或版权所有者才有密钥，只有他们可以解密或恢复置乱从而得到数字产品的内容。但是，仅采用密码技术的重大缺点是所加密的数字内容在解密之后就无有效的手段来保证其不被非法拷贝、再次传播和盗用。为了防止这种情况的发生，人们提出了数字水印的概念。

数字水印 (Digital watermarking) 是近两年来出现的数字产品版权保护技术，可以标识作者、所有者、发行者、使用者等，并携带有版权保护信息和认证信息，目的是鉴别出非法复制和盗用的数字产品，作为密码学的加密或置乱技术的补充，保护数字产品的合法拷贝和传播。随着网络化信息化进程的加速发展，对数字产品的版权保护技术的要求也是迫在眉睫。因此，数字水印一经提出就迅速地成为了热点问题，出现了许多数字水印方案，也有许多公司已推出了数字水印的产品。

同密码学分为密码设计学和密码分析学类似，数字水印的研究目前一般也主要集中在两个方面：水印算法设计和水印算法攻击。有关数字水印的绝大多数文献都是讨论如何设计数字水印方案或如何攻击数字水印，两方面的研究是互相依存，互相促进的，好的攻击方案能促进人们设计出更好的水印算法，而好的水印算法出现，也促使人们考虑对它的攻击以验证安全性和稳健性。

本文简单介绍数字水印的概念和应用，分析现有的水印攻击方法，提出了水印的下一步的发展方向。本文的安排如下：第 2 节介绍数字水印的概念和应用；第 3 节对数字水印攻击方法进行分类，讨论各种方案的特点；第 4 节提出数字水印的发展方向。

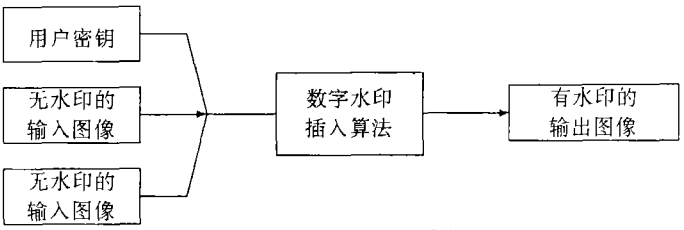
2 数字水印

2.1 数字水印的定义

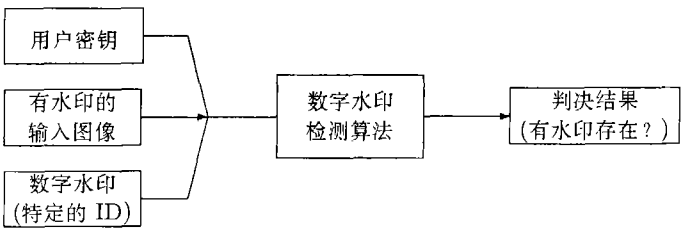
目前虽然有许多文献讨论数字水印技术的许多方面的问题，但对数字水印还没有给出一个明确的统一的定义。通过综合一些学者提出的定义和分析已有的数字水印方案，我们可给出如

[1] 1999-07-28 收到，1999-07-28 定稿

本课题得到国家重点基础研究发展规划项目 (G1999035805)、国家杰出青年基金 (69425001)、国家自然科学基金 (60073049) 及高等学校骨干教师资助计划的资助



数字水印插入过程 (a)



数字水印检测过程 (b)

图 1 数字水印的插入 (加载) 及检测 (提取) 过程

下定义：数字水印是永久镶嵌在其他数据中具有可鉴别性的的数字信号或模式，而且并不影响数据的可用性。

不同的应用对数字水印的要求是不尽相同的，一般认为数字水印应具有如下特点 [1]：(1) 隐形性：数字水印应是不可知觉的，即数字水印的存在不应明显干扰被保护的数据，不影响被保护数据的正常使用。(2) 稳健性：数字水印必须难以 (最好是不可能) 被除去，如果只知道部分水印信息，那么试图除去或破坏水印应导致严重的降质而不可用。数字水印应在下列情况下具有稳健性。一般的信号处理下的稳健性：即使有水印的数据经过了一些常用的信号处理，水印仍应能被检测到，这包括 A/D ， D/A 转换，重采样，重量化，滤波，平滑，有失真压缩等常用的信号处理方法。一般的几何变换 (仅对图像和视频而言) 下的稳健性：包括旋转，平移，缩放及分割等操作。(3) 确定性：水印所携带的所有者等信息能够被唯一的鉴别确定，而且在遭到攻击时，确认所有者等信息的精确度不会劣化许多。

数字水印的加载和检测过程通常如图 1 所示，需要指出的是随着数字水印方案的不同，水印的加载和检测过程是不完全相同的，因此，有一些水印算法的水印插入和检测过程将和图 1 有所不同。

2.2 数字水印的应用

目前，数字水印技术的应用大体上可以分为 [2]：维护所有权；指纹；认证和完整性校验；内容标识；使用控制；内容保护等几个方面。下面我们对这些应用一一作介绍。

维护所有权 (Ownership assertion)：为了表明对数字产品内容的所有权，所有者 A 用私钥产生水印并将其插入原图像 (以图像为例) 中，然后即可公开加载过水印的图像，如果 B 声称对公开的有水印的图像有所有权，那么 A 可以用原图像和私钥证明在 B 声称所有的图像中有 A 的水印，由于 B 无法得到原图像，B 无法作同样的证明。但在这样的应用中，水印必须有足够的稳健性，同时也必须能防止被伪造。

指纹 (Fingerprinting)：为了避免未经授权的复制和分发可公开得到的多媒体内容，作者可在其每个产品中嵌入一个明显的水印 (指纹)。如果发现了未经授权的拷贝，则通过检索指纹来追踪其来源。在此类应用中，水印必须是不可见的，而且能抵抗恶意的擦除、伪造或试图使水印无效的攻击。

认证和完整性校验 (Authentication and integrity verification)：在许多应用中，需要验证数字内容未被改变、修改或造假。尽管多媒体内容的认证可通过传统的密码学技术来完成，但

利用数字水印来进行认证和完整性校验的优点在于：认证同内容是密不可分的，因此简化了处理过程。当对插入了水印的数字内容进行检验时，必须用唯一的与数据内容相关的密钥提取出水印，然后通过检验提取出的水印完整性来检验数字内容的完整性。

内容标识 (Content labeling)：此类应用中，插入的水印信息构成一个注释，提供有关数字产品内容的进一步的信息。例如：在图像上标上拍摄的时间和地点，这个构成可以由照相机中的微处理器自动完成。

使用控制 (Usage control)：在一个封闭的系统中，多媒体内容需要特殊的硬件来拷贝和观看使用，插入水印来标识允许的拷贝数，每拷贝一份，进行拷贝的硬件会修改水印内容，将允许的拷贝数减一，以防止大规模的盗版，DVD 就是这种应用的实例。

内容保护 (Content protection)：在一些特定应用中，数字产品内容的所有者可能会希望要卖的多媒体内容能被公开自由地预览，以尽可能地多招徕潜在的顾客，但也需要防止这些预览的内容不被其他人用于商业目的，因此，这些预览内容被自动加上可见的但同样难以除去的水印。

对水印技术的要求随着应用的不同而不同，一个水印方案很难满足所有应用的所有要求，因此，数字水印算法往往是针对某类应用而设计的。

3 数字水印的攻击方法

目前，对水印的攻击方法可分为四类 [3]：稳健性攻击 (Robustness attack)，表达攻击 (Presentation attack)，解释攻击 (Interpretation attack)，合法攻击 (Legal attack)。

3.1 稳健性攻击

这类攻击其实是直接攻击，目的在于擦除或除去在标记过的数据中的水印而不影响图像的使用。这类攻击修改图像像素的值，大体上可再细分为两种类型：信号处理攻击法和分析 (计算) 攻击法。

典型的信号处理攻击法包括无恶意的和常用的一些信号处理方法，例如：压缩，滤波，缩放，打印和扫描等。我们对图像经常采取这些处理以适应不同的要求，例如：对图像进行压缩以得到更快的网络传送速度。信号处理攻击法也包括通过加上噪声而有意修改图像，以减弱图像水印的强度，我们用强度这一术语来衡量嵌入水印信号的幅度相对于所嵌入的数据幅度，类似于通信技术中的调制系数这一概念。已经有几种公共软件能成功地除去个别商业软件所嵌入的水印，这些软件进行的攻击所导致的图像质量的下降是可以接受的，证明了将通用的信号处理方法用于水印攻击的可行性。

应该指出，人们通常有这样的误解：一个幅度很小的水印可以通过加上类似幅度的噪声来除去，实际上，相关检测器对随机噪声这类攻击是很稳健的，因此，在实际应用中，噪声并不是严重的问题，除非噪声相对于图像来说幅度太大或者噪声同水印是相关的。

即使是对目前所研制的稳健性很好的基于扩频技术的水印 [1]，也出现了相应的攻击方法，文献 [4] 中，作者发现 3×3 大小的中值滤波器能很有效地除去图像中的扩频水印。如何抵抗类似的非线性的基于信号处理的攻击方法还是非常值得深入研究的问题。

分析 (计算) 攻击法包括在水印的插入和检测阶段采用特殊方法来擦除或减弱图像中的水印。这类攻击往往是利用了特定的水印方案中的弱点，在许多例子中，它证明了分析研究即已足够，不必在真实图像上测试这类攻击。文献 [1] 中提出的共谋攻击 (Collusion attack) 或多重文档攻击 (Multi-document attack) 就是这类攻击，共谋攻击用同一图像嵌入了不同水印后的不同版本组合而产生一个新的“嵌入了水印”图像，从而减弱水印的强度。

首先说明水印的加载与检测过程的数学模型.假设我们从文档 D 中提取出序列 $V=v_1,v_2,\cdots,v$ 在该序列上插入水印 $X=x_1,x_2,\cdots,x_n$ 得到序列 $V'=v'_1,v'_2,\cdots,v'_n(v'_i=v_i+ax_i)$，然后将 V' 替代 D 中的 V 得到加载了水印的文档 D'，检测时，从待检测的文档 D^* 中提取出的水印为 X^*，采用下式衡量 X^* 和 X 之间的相似性：$\mathrm{sim}(X^*,X)=X^*\cdot X/(X^*\cdot X^*)^{1/2}$ 如果

$\mathrm{sim}(X^*, X) > T$, T 是某个预先设置好的阈值，则我们认为 D^* 中提取出的水印 X^* 是原来插入的水印 X 。

许多水印方案是易于遭到多重文档攻击的，为了说明多重文档攻击的危险，我们考虑这样的水印方案：在 v_i 上随机地加上由 1 或 -1 构成的水印而得到 v_i' 。对于这种水印方案，一旦发现两个文档中有不相同的 v_i' ，便可得到 v_i 的值，从而可以完全消除水印的这个分量。如果有采用这种水印方案的 t 个文档，将这 t 个文档求平均就几乎可以擦除所有的水印分量而只剩下 2^{1-t} 部分未擦除。如果加 ± 1 的位置不是如此简单，有更好的插入位置，则对上述简单而直接的多重文档攻击有更好的抵抗能力。注意到上述攻击中，我们并未假定 v_i 服从何种分布，如果假设 v_i 服从均匀分布，则也会有更好的抵抗多重文档攻击的能力，将这 t 个文档求平均后得到的文档 D^* 中的水印 X^* 约为 X 的 $1/t$ ，上述攻击的详细分析可参见文献 [1] 。

攻击者还会采用统计平均攻击，如果攻击者发现了一个通用水印，例如：某水印方案中水印不是非常依赖于图像 I ，此时，统计攻击就非常危险了，攻击者得到水印估计值 u 后，此估计值就可以用来除去任何采用同一水印方案插入水印的图像中的水印 [5] 。这类攻击采用 $I_1 + u, I_2 + u, \cdots, I_N + u$, N 个图像相加得到结果 $Nu + \sum_i I_i$ ，对足够大的 N 和独立的图像 $I_1, I_2, \cdots, I_N$ ，我们可以认为结果是 Nu ，这样就得到了水印 u 。一个抵抗的措施是用至少两个不同的随机水印 u_1 和 u_2 ，各有概率 p_1 和 p_2 ，其中，p_1 和 p_2 满足关系 $p_2 = 1 - p_1$ 。再进行上述攻击仅能得到 $p_1u_1 + (1 - p_1)u_2$ ，无法得到 u_1 和 u_2 的值。但是，该攻击的一个改进是可以计算出加权的平均值，其中，加权因子由一个猜测决定，猜测一个特定的图像中是否含有一个水印或者是另一个水印。假定攻击者将目标图像 (以概率 p_1 和 p_2 随机插入 u_1 或 u_2 的图像) 归类于 $i(i = 1, 2)$ ，如果他认为图像中含有 u_i 。P_e 代表图像分类错误的概率，那么，N_1 个 1 类中的目标图像求和后，得到结果 $S_1 = N_1p_1(1 - P_e)u_1 + N_1(1 - P_e)(P_e)u_2$ ；类似地，N_2 个 2 类中的目标图像求和后，得到结果是 $S_2 = N_2p_1P_eu_1 + N_2(1 - p_1)(1 - P_e)u_2$ 。计算加权差得到：$(S_1/N_1) - (S_2/N_2) = p_1(1 - 2P_e)u_1 - (1 - p_1)(1 - 2P_e)u_2$ 。

因此，对任何 $P_e \neq 0.5$ ，任何选择标准都比一个随机值好，攻击者可以估计总和及 p_1u_1 与 $(1 - p_1)u_2$ 的差值。这样就可解出 u_1 和 u_2 的值，即可擦除水印。抵抗上述各种直接攻击的方法在于：水印的算法是要公开的，因此，算法的安全性应依靠与内容相关或无关的密钥及算法本身特性，攻击者无法得到密钥，就无法擦除水印。需要指出的是这里的密钥同密码学中的严格的密钥概念是不同的，密码学中的密钥在攻击者已知加密算法，已知密文和已知明文攻击下都是非常难以解出的，水印中的密钥尚无如此严格的要求。

3.2 表达攻击

此类攻击有别于稳健性攻击之处在于它并不需要除去数字产品内容中嵌入的水印，它是通过操纵内容从而使水印检测器无法检测到水印的存在。例如：表达攻击可简单地通过不对齐一个嵌入了水印的图像来愚弄自动水印检测器 (如：基于 Web 的智能代理或 Webcrawler 等)，实际上在表达攻击中并未改变任何图像像素值。

剑桥大学计算机实验室的 F. Petitcolas 在文献 [6] 中提出的攻击方法就是一个很好的有效表达攻击的例子，该攻击方法目的是挫败 Webcrawler ，该方法是将一个嵌入了水印的图像切成许多小块，这些小块在 Web 页上按相应的 HTML 标记再组装起来。Webcrawler 只能去查看每个图像小块，但这些小块由于太小而无法容纳任何水印数据，所以 Webcrawler 无法发现水印。该攻击方法实际上并未导致任何图像质量的下降，因为图像像素值被完全保留了。此外，对通过 Java 小件显示图像的软件也可进行类似的攻击。

更多的关于表达攻击的例子包括旋转，放大及通常的仿射变换。该类攻击的主要思想是在检测水印之前，水印方案要求嵌入了水印的图像被正确地对齐。例如：一种基于统计的水印，其加载方法是任意选择 N 对图像点，在增加一点亮度的同时，降低另一点的亮度值。我们用 I 和 W 表示原图像和水印，而 I' 是插入了水印的图像。假定 I'_S 是 I' 平移一个像素后的结果，I_S

和 W_S 代表了 I 和 W 的相似的平移结果。那么，$I'_S \bullet W = I_S \bullet W + W_S \bullet W$，由于 Patchwork 法中，$I_S \bullet W$ 中 + / - 是独立随机选择的，$I_S \bullet W$ 和 $W_S \bullet W$ 都对消掉了。所以，$I'_S \bullet W$ 的幅度很小，水印就不会被检测到 [3]。这样，仅仅通过平移像素，实际中并未改变任何图像像素值，就使水印失去了作用。

现有的一些图像及视频水印方案中，图像中除嵌入水印外还需嵌入一个登记模式以抵抗几何失真，但在应用中，这个登记模式往往成了水印方案的致命弱点，如果正常的登记过程被攻击者所阻止，那么，水印的检测过程就无法进行而失效。

对一个成功的表达攻击而言，它并不需要擦除或除去水印。为了战胜表达攻击，水印软件应有同人的交互才能进行成功的检测。或者，更聪明的检测算法应设计成为能容纳通常的表达模式，尽管在工程上实现这样的智能仍是非常困难的。

3.3 解释攻击

在一些水印方案中，可能存在对检测出的水印的多个解释。例如，一个攻击者试图在同一个嵌入了水印的图像中再次嵌入另一个水印，该水印有着与所有者嵌入的水印相同的强度，由于一个图像中出现了两个水印，所以导致了所有权的争议。在解释攻击中，图像像素值或许被改变或许不被改变。此类攻击往往要求对所攻击的特定的水印算法进行深入彻底的分析。

文献 [5] 中提出的攻击实例可以视为解释攻击的一个例子，该攻击方法对一类水印技术都有效。不失一般性，此类水印技术的嵌入过程可用公式表达为：$I_w = I + W$，此公式的意义是一幅图像 I 加载了由足够小的值构成的水印 W 而图像无明显的降质，得到的嵌入了水印的图像是 I_w。嵌入过程中，可以根据不同的规则插入水印，或者在图像的不同的位置空间上和不同的频率域上有着不同的插入强度，而不仅仅是在图像中插入固定强度的水印。假定创造者或所有者 A 将原图像 I 和水印 W 秘密存储起来，仅发布嵌入了水印的 I_w。这样，给定可疑图像 I'，我们可以从一个可疑的图像中减去 I 而提取出水印，即 $W' = I' - I$。然后对提取出的水印同原水印 W 作相关以衡量它们之间的差异，$P = C(W, (I' - I))$，其中 $C(W, W')$ 是某个衡量两个水印间相似性的度量。这种类型的水印方案要求在检测水印时有原图像，许多现有的水印技术都是这种类型。

在进行解释攻击时，攻击者 B 将水印插入过程逆过来运用，即他的攻击是减去一个水印。B 计算伪造的原图像 $I_B = I'_B - W_B$，声称 I_B 是他的“原图像”（即未插入水印的图像，由于 W_B 足够小，I_B 和 I'_B 之间无明显的差异），W_B 是所插入的水印，I'_B 是插入水印后的 I_B，而实际上，I'_B 是 B 从某途径得到的未对其授权的图像（其中或许插入了别人的水印）。为了区别清楚，我们将所有者 A 的原图像和水印分别用 I_A 和 W_A 来表示。现在，A 和 B 都可以声称他们用各自的原图像和水印而产生了插入了水印的图像 I_w。A 和 B 还可以用他们各自的原图像和水印来检测对方所产生的插入了水印的图像 I_w。A 所进行的检测过程可用公式表达：$N_A = I_B - I_A = W_A - W_B$, $P_A = C(N_A, W_A) = C(W_A - W_B, W_A)$。而 B 所进行的检测过程类似也可用公式表达：$N_B = I_A - I_B = W_B - W_A$, $P_B = C(N_B, W_B) = C(W_B - W_A, W_B)$。

这样，P_B 代表了 A 的水印出现在攻击者 B 声称的原图像和水印所产生的 I_w 中，而 P_A 则代表了 B 的水印出现在所有者 A 声称的原图像和水印所产生的 I_w 中。这个对称性是显然的：在 I_B 中有 W_A，而在 I_A 中有 W_B。这样，形成了死锁，无法判断 I_w 是由谁产生的。采用不同的水印方案得到的实际证据表明没有一个水印能充分地证明谁是造假者。文献 [8] 中提出的 SWICO(Single Watermarked Image Counterfeit Original) 攻击和 TWICO(TwinWatermarked Image Counterfeit Original) 攻击实质上同上述攻击相同。

解释攻击中，攻击者并没有除去水印而是在原图像中“引入”了他自己的水印，从而使水印失去了意义，尽管他并没有真正地得到原图像。在这种情况下，攻击者同所有者和创造者一样拥有发布的图像的所有权的水印证据。

对统计水印技术同样可进行解释攻击，尽管在统计水印技术的检测阶段不需要原图像。这种独特的攻击利用了水印方案的可逆性，这个特性使攻击者可以加上或减去水印。潜在的补救解决方法包括在插入水印过程中，加入一个原图像的单向 HASH 函数，使攻击者除去水印而不产生视觉上可察觉的降质是不可能的，可逆性及其应用可见文献 [9] 。

3.4 合法攻击

这类攻击同前三类攻击都不同，前三类可归类为技术攻击，而合法攻击则完全不同。攻击者希望在法庭上利用此类攻击，它们的攻击是在水印方案所提供的技术优点或科学证据的范围之外进行的。合法攻击可能包括现有的及将来的有关版权和有关数字信息所有权的法案，因为在不同的司法权中，这些法律有可能有不同的解释。合法攻击还可能包括所有者和攻击者的信用，攻击者使法庭怀疑数字水印方案的有效性的能力，除了这些之外，可能还和其他一些因素紧密相关，如：所有者和攻击者的金融实力的对比，专家的证词，双方律师的能力等。

理解和研究合法攻击要比理解和研究技术上的攻击要困难的多。作为一个起点，我们首先应致力于建立一个综合全面的法律基础设施，以确保正当的使用水印和利用水印技术提供的保护，同时，避免合法攻击导致降低水印应有的保护作用。合法攻击是难以预料的，但是一个真正稳健的水印方案必须具备这样的优点：攻击者使法庭怀疑数字水印方案的有效性的能力降至最低。

理解这些攻击有助于我们提出更好的数字水印方案，它们将不仅仅依靠提高水印强度来增加稳健性，而且也通过增强它们抵御这些攻击的能力来增加稳健性。认识到现有的水印技术的缺点和局限性，再结合对这些技术面临攻击时的性能所做的透彻的分析和评估，最终将导致对该领域的更广泛更深入的研究。关于数字水印攻击问题，还大有进一步讨论研究的余地。

4 数字水印的发展方向

尽管数字水印技术还面临许多问题，比如：许多水印算法无法抵抗攻击，必须配合其它技术手段才能达到应用目的，而且水印技术缺少相关的标准或协议，这些都妨碍了数字水印技术的进一步推广应用，但我们相信如果没有恰当地提出在潜在的攻击范围内的问题及论点，数字水印技术也决不会迅速跨越实验阶段而得到使用，水印技术现在得到了一定程度的应用是同水印攻击的研究分不开的。

综合数字水印的算法设计和算法攻击的发展现状和趋势，除了需继续研究具有很好的稳健性和安全性的水印算法外，从实际应用的观点看，我们认为数字水印的下列方向应是研究的重点：根据不同的数字产品内容分等级插入水印，即对较重要的内容和对安全性要求高的内容插入强度大，安全性好的水印，而对不太重要的内容和对安全性要求不高的内容插入强度小安全性一般的水印，以适应实际应用的要求，这种分安全等级的水印方案有助于提高效率，也间接增强了水印的安全性；研制出动态水印或具有交互性质的数字水印，可以修改水印内容或者通过水印来实现某些控制如：读取，拷贝。这要求水印中有可执行内容，在网络环境中可以通过在水印中加入 Java 应用小程序或含有特定的 URL 等方法来实现；另外，必须注意到数字水印技术并非是万能的，必须配合密码学技术及数字认证，数字签名或者数字信封等技术一起使用，一个实用的数字水印方案必须有这些技术的配合，才能抵抗各种攻击。

毫无疑问，在数字时代数字水印技术将对保护各种形式的数字产品内容起到重要作用，尽管该领域还是个相对来说非常年轻的领域，但它已经吸引了许多一流的研究者。我们会看到对数字水印攻击方法的研究和水印算法设计的研究将导致更好的水印方案的出现和更成功的数字水印应用。

参 考 文 献

[1] I. J. Cox, J. Killian, F. T. Leighton, T. Shamoon, Secure spread spectrum watermarking for multimedia, IEEE Trans. on Image Processing, 1997, IP-6(12), 1673-1687.

[2] N. Memon, P. W. Wong, Protecting digital media content, Communications of the ACM, 1998, 41(7), 35-43.

[3] S. Crave, B. L. Yeo, M. Yeung, Technical trials and legal tribulations, Communications of the ACM, 1998, 41(7), 45–54.

[4] G. C. Langelaar, R. L. Lagendijk, J. Biemond, Removing spatial spread spectrum watermarks by non-linear filtering, EUSIPCO'98, 1998, 2281–2284.

[5] I. J. Cox, J. M. G. Linnartz, Some general methods for tampering with watermarks, IEEE J. on Selected Areas in Communications, 1998, SAC-16(4), 587–593.

[6] http://www.cl.cam.ac.uk/users/fapp2/.

[7] S. Craver, N. Memon, B. L. Yeo, M. Yeung, Can invisible watermarks resolve rightful ownerships? Proceedings of IS&T/SPIE Electronic Imaging-Storage and Retrieval of Image and Video Databases, San Jose. S, CA: SPIE, Feb. 13–14, 1997, 310–321.

[8] S. Craver, N. Memon, B. L. Yeo, M. Yeung, On the invertibility of invisible watermarking techniques, Proceedings of the IEEE International Conf.on Image Processing, Santa Barbata, CA, IEEE Press, 1997, 540–543.

[9] S. Craver, N. Memon, B. L. Yeo, M. Yeung, Resolving rightful ownerships with invisible watermarking techniques: limitations, attacks, and implications, IEEE J. on Selected Areas in Communications, 1998, SAC-16(4), 573–586.

THE ATTACK METHODS OF DIGITAL WATERMARKING

Chen Mingqi Niu Xinxin Yang Yixian

(*Info. Security Center of Beijing Univ. of Posts and Telecommunications, Beijing* 100876, China)

Abstract Digital watermarking is a technique, which protects digital product copyright and does not appear until two years ago. And its goal is to protect legal copy and distribution of digital product, it has been the hotspot of international academia. Digital watermarking includes algorithm design and attack technique, which are contrarily promoting each other. In this paper the concept and applications of digital watermarking are briefly discussed, the current various attack methods are fully analyzed and summarized. Several directions of development of digital watermarking at next stage are proposed. These are of great importance to improve and design watermarking algorithm.

Key words Intellectual property right protection, Digital watermarking, Watermarking attack

陈明奇： 男，1973 年生，博士生，研究方向：信息伪装及数字水印，信息安全.

钮心忻： 女，1963 年生，北京邮电大学信息安全中心副教授，研究方向：信号与信息处理，信息伪装，软件无线电等.

杨义先： 男，1961 年生，北京邮电大学信息安全中心教授，博士生导师，研究方向：编码密码学，网络与信息安全，信号与信息处理，信息伪装等.