【综述与评论】

数字水印的安全性*

温 习,谭月辉

(军械工程学院,石家庄 050003)

摘要:针对近年来出现的新的数字产品版权保护技术——数字水印的安全性问题,总结和分析了现有的对数字水印的主要攻击方法,并对部分攻击方法提出了相应的解决对策,分析了这些对策的优缺点,为新型数字水印系统的研究提供了借鉴作用.

关键词:数字水印;攻击方法;对抗策略;数字水印安全性

中图分类号:TP393.08 **文献标识码**:A **文章编号**:1006-0707(2008)04-0085-03

数字水印安全性主要是指数字水印抵抗恶意攻击的能力[1],它是数字水印系统的最重要的指标之一[2].尤其在版权保护领域,如果水印安全性很差,就容易遭受攻击进而失去版权保护的能力,给实施侵权者造成可乘之机.所以研究数字水印安全性问题是十分必要的.本研究将系统地对攻击技术和现有对策等水印安全性问题进行论述,并将攻击进行特定的归类.

1 常见的水印攻击方法与对抗策略

与水印嵌入技术的发展类似,水印攻击技术也经历了一个快速发展的过程.目前,已有的数字水印的攻击方法主要有4类[3]:去除攻击(removal attacks)、表达攻击(representation attacks)、协议攻击(protocol attacks)、合法攻击(legal attacks).前3类主要利用水印设计上的弱点,通常可以归类为技术攻击,是本研究的重点;而合法攻击主要利用水印使用上的弱点,本文中不加讨论.

1.1 **去除攻击及其对策**.这类攻击方法主要是破坏水印的嵌入环节.此类攻击攻击方法有2种:像素值失真攻击和分析攻击.像素值失真攻击是指对水印图像进行某种操作,通过破坏水印图像的像素值以削弱或删除嵌入的水印,例如有损压缩图像及滤波、gamma校正等[4].分析攻击是通过分析水印图像来估计图像中的水印,然后将水印从图像中分离出来.以这个条件为标准,将敏感性分析攻击和统计平均分析攻击统一归为此类.其中敏感性分析攻击是使用相关水印检测器寻找从水印检测区域到区域边缘的捷径,而该捷径可由检测区域表面的法线近似表示,并且该法线在检测区域的绝大部分是相对恒定的,敏感性分析攻击流程见图1.

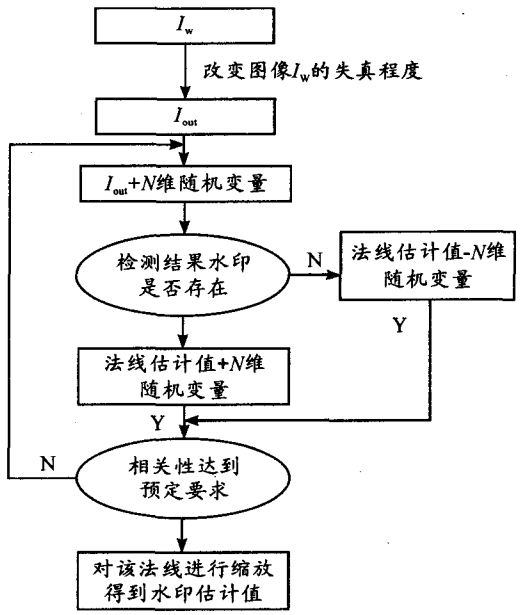

图1 敏感性分析攻击步骤流程

例如I_w是水印图像,I_{out}是接近相关检测区域边界的图像.改变图像I_w的失真程度而获得图像I_{out}的方法有3种:减少图像I_w的对比度或亮度的幅值,使用无水印图像与水印图像I_w的线性组合,以及使用水印图像I_w的平均值来代替采样值.根据文献[5]找出图像I_{out}检测区域表面法线方向的近似值,这是进行水印敏感性分析攻击的核心.而法线方向估计值与图像I_w中水印的相关性是迭代次数的单调递增函数.最后将水印的近似值从图像I_w中减

* **收稿日期**:2008-04-06

作者简介:温习(1982—),男,辽宁沈阳人,硕士研究生,主要从事军事信息安全研究.

去,得到质量良好的检测不到水印的近似图像.

与敏感性分析攻击不同,统计平均分析攻击的基本思想是:当攻击者可以得到大量的含有水印的图像时,从中选择一些互不相关的图像,然后对它们进行统计平均,就可以获得水印的估计,从而将其从水印图像中删除生成一个近似的不含水印的原始图像[6],来达到去除水印的目的.研究表明,攻击者只需掌握少量不同的拷贝份数就可以成功移除水印[7].

针对像素值失真攻击主要就是建立鲁棒性好的水印模型.首先可以通过不同的攻击特性合理选择嵌入位置(如对于针对有损压缩的攻击方法,可将水印嵌入图像中关键的低频分量),其次可以通过增加嵌入的力度,或采用多次(冗余)嵌入水印的办法,具体来讲在空间域上,可以将同一个水印信号多次嵌入一幅图像的不同位置,采用大多数投票制实现水印提取,来增加水印的强度.其中对JPEG压缩和对信道噪声、滤波具有较好抵抗能力的算法分别在文献[8-12]及文献[13-14]具体给出.而针对敏感性分析,由于水印敏感性分析攻击的成功,依赖于检测区域边界的法线可用于寻找越出检测区域的捷径.如果检测区域边界的曲率使在每一点的法线仅提供关于该捷径方向的极少信息,则敏感性分析攻击在计算上是不可行的.因此构造具有这种性质的水印检测区域是解决的方法.对抗统计平均分析攻击的方法是嵌入多个水印,并让它们在图像中相互独立,也可以在水印生成部分引入随机密钥进行加密,可以有效增加消除这种攻击的计算复杂度,从而使攻击难以实现.

由去除攻击的本质可以知道去除攻击将对水印造成实质性的损害,如果不能抵抗这类攻击,那么水印将难以检测或恢复的.所以关键在于预防,也就是要建立鲁棒性高的水印系统.

1.2 **表达攻击及其对策**.这类攻击方法主要是破坏水印的检测环节.这种攻击主要分为几何攻击[15]和马赛克攻击[16-17].几何攻击原理为是利用旋转、剪切、水平翻转、行(或列)删除等几何变形的方法破坏水印检测器和内嵌水印之间的同步,使水印信号错位,从而使检测器无法识别水印信号,以达到攻击的目的.马赛克攻击原理为:图像越大越容易嵌入一定量的比特信息,反过来图像越小能嵌入的信息越少,小到一定程度,就无法再嵌入信息了,从而无法隐藏一个有意义的标记.利用此原理,首先将图像分成许多个小图像,然后将每个小图像放在HTML页面上拼凑成一个完整的图像,从而使得探测器无法从中检测到侵权行为.

对抗几何攻击的对策有很多,总体来说,一是考虑几何不变量,如早期利用Mellin-Fourier变换,以及后期出现的奇异值分解、Radon变换和Ridgelet变换;二是利用辅助模板嵌入的方法跟踪含水印载体所经历的各种攻击,然后进行逆变换可消除同步攻击;三是采用基于块的检测算法,利用几何形变在局部几乎都是线性这个事实;四是在水印检测算法中,对嵌入水印的位置采用相对的位移地址,而不是采用绝对的存储地址;五是在嵌入水印的同时,也嵌入对准信息或易损水印,这样在检测时就能查出几何形变,并在检测前恢复.而对抗马赛克攻击目前最有效的措施除了上述方法就是保证水印能嵌入到足够小的图像中.

表达攻击并不需要削弱或除去水印,因此它几乎不影响图像质量,这是区别于其他攻击的也是很难对付的一个重要方面.但这也正是表达攻击的致命弱点,当使用更复杂、更智能化的水印检测器时,表达攻击便会失败,所以在水印检测器上下功夫是关键.

1.3 **协议攻击及其对策**.协议攻击是使水印检测的结果错误或不明确.协议攻击主要有解释攻击、拷贝攻击.解释攻击又称为IBM攻击,指产生一个伪造水印信息,但不破坏原来水印信息.版权所有者对自己拥有的作品嵌入自己的水印信息,然后将水印作品发布.由于无法区分水印的真伪和嵌入的先后,如果水印强度非常接近的话,就会阻止任何一方确立所有权,从而删除水印效果.拷贝攻击[18]是从嵌入水印的图像中估计出水印并拷贝到目标图像的其他图像中,这既不需要算法知识又不需要水印密钥知识.拷贝攻击分为3步进行:第1步,找出图像中水印的估计值;第2步,处理该估计值,使得水印能量最大化并满足不可感知性要求;第3步,将处理后的水印估计值嵌入目标图像得到伪造的水印图像.

针对解释攻击可以对水印的使用环境加以限制[19],也可对攻击的条件加以破坏.限制使用环境的方法有2种:一是引入时间戳机制.在加密学中,时间戳机制主要用于数字签名[20].在数字水印嵌入过程中,如果合理使用了时间戳机制,就能够轻易判定水印添加的先后顺序,因为后被加入的水印即为伪造水印,也就解决了解释攻击所引起的版权纠纷.二是引入公证机制.当版权所有者向公证机关注册水印序列的同时,也将原始作品进行注册.这样当引起版权纠纷时,如果攻击者的作品中能够检测出作者的水印,而在作者的作品中无法检测到攻击者的水印,则可以证明攻击者的作品是伪造的.另外也可由公证机关提供的注册时间来判定,攻击者的注册时间肯定在原作者的注册时间之后.破坏攻击条件的方法也有2种:一是构造合理的单向化函数嵌入水印,使水印方案非可逆和非对称[21],从而消除水印嵌入过程中的可逆性.如果水印的嵌入机制具不可逆性,那攻击者无法对水印图像进行逆操作来达到伪造目的.二是使用双水印技术[22],先嵌入鲁棒水印,再嵌入脆弱水印.当发生版权纠纷时,版权所有者可以提供只嵌有合法水印的图像,而攻击者在利用无需原始图像存在的检测算法时,无法提取得到其所嵌入的伪造水印信息,从而攻击失败.

而由于拷贝攻击的基础是必须从水印图像中估计出水印,所以将水印与图像联系在一起,使水印依赖于图像,这样就能有效地抵抗拷贝攻击,比如密码签字技术,以及在注册水印序列的同时对原始作品也加以注册的方法.还可以利用盲检测技术以杜绝伪造原始图像的可能性.另外,基于量化的水印方案对拷贝攻击也具有免疫力.

解释攻击和拷贝攻击都属于在协议层上的攻击,会对水印的许多应用造成严重损害.在版权保护方面,解释攻

击的威胁要大得多,因为任何人都可声称他对访问过的任何水印图像拥有所有权;拷贝攻击在这方面的应用是作者盗用某名人的名义,出售自己的作品以牟利.在有关身份认证的应用中,拷贝攻击所造成的威胁是重大的,使得用户无法根据水印的检测结果确定作品来源的真实性.对于协议攻击可以用算法限制克服,破坏可以进行攻击的条件及环境,这样此类攻击便无从下手.

2　结束语

数字水印安全性问题是数字水印技术中一个相当重要的问题,它直接影响了数字水印的最终目的——版权问题.本研究分析了数字水印的一些常见的攻击方案和相应的对策,并以某种角度对某些攻击及对策进行了重新分类比如将敏感性分析攻击和统计平均分析攻击统一归为一类以及把抵抗解释攻击的方法细分等.目前现有的算法对策都只能抵抗较低级的独立的攻击手段,而当各种攻击方法有机组合或同时施展多种攻击时却力不从心.因此,针对能抵抗多种或多种攻击组成的综合攻击的研究是当前一个重要的研究方向.

参考文献:

[1] Wolfgang R B, Delp E J. Overview of image security techniques with applications in multimedia system[C]// Proceedings of the SPIE Conference on Multimedia Networks: Security, Displays, Terminals and Gatways. USA Dallas: Texas, 1997:297 - 308.

[2] 刘瑞祯,谭铁牛.数字图像水印研究综述[J].通讯学报,2000,21(8):39 - 48.

[3] 季智,戴旭初. 数字水印攻击技术及其对策分析[J]. 测控技术,2005(5):14 - 18.

[4] Barnett R, Pearson D E. Attack operators for digitally watermarked images[J]. IEEE Proceedings on Vision Image and Signal Processing, 1998, 145(4):271 - 279.

[5] Solachidis V, Tefas A, Tsekeridou S, et al. A benchmarking protocol for watermarking methods[C]//Proc of IEEE Int Conf on Image Processing. Thessaloniki: Institut of Electrical and Electronics Engineers Commputer Society, 2001:1023 - 1026.

[6] Cox I J, Kilian J, Leighton T, et al. Secure spread spectrum watermarking for multimedia[J]. IEEE Transon Image Processing, 1997, 6(12):1673 - 1687.

[7] Cheswick B. An evening with Berferd[EB/OL]. [2007 - 10 - 10]. http://www.trackinghackers.com/papers/berferd.pdf.

[8] 杨恒伏,陈孝威.小波域鲁棒自适应公开水印技术[J].软件学报,2003,14(9):1652 - 1660.

[9] 张军,王能超.数字图像的自适应公开水印技术[J].计算机学报,2002,25(12):1371 - 1377.

[10] Barnia M, Bartolinib F, Furonc T. A general framework for robust watermarking security[J]. Signal Processing, 2003, 83:2069 - 2084.

[11] Van Trees H L. Detection, estimation, and modulation theory[M]. [S.l.]:John Wiley&Sons Inc, 2001.

[12] 黄达人,刘九芳,黄继武.小波变换域图像水印嵌入对策和算法[J].软件学报,2002,13(7):1290 - 1297.

[13] Barnia M, Bartolinib F. Improved wavelet-based watermarking through pixel-wise masking[J]. IEEE Transactions on Image Processing, 2001, 10(5):783 - 791.

[14] 牛夏牧,陆哲明,孙圣和.基于多分辨率的数字水印技术[J].电子学报,2000,28(8):1 - 4.

[15] Joseph J K, Ruanaidh O, Pun Thierry. Rotation, scale and translation invariant spread spectrum digital image watermarking[J]. Signal Processing, 1998, 66(3):303 - 317.

[16] 刘春庆,王执铨,戴跃伟.常用数字图像水印攻击方法及基本对策[J].控制与决策,2004 (6):601 - 606.

[17] 陈明奇,钮心忻,杨义先.数字水印的攻击方法[J].电子与信息学报,2001 (7):705 - 711.

[18] Kutter M, Voloshynovskiy S, Herrigel A. The watermark copy attack[C]//Proc of the San Jose SPIE. San Jose: [s.n.],2000:371 - 380.

[19] 袁中兰,夏光升,温巧燕,等.数字作品著作权保护协议[J].北京邮电大学学报,2005(1):19 - 22.

[20] Katzenbeisser S, Veith H. Securing symmetric watermarking schemes against protocol attacks[C]//Proc of SPIE. San Jose: The Int Society for Optical Engneering, 2002:260 - 268.

[21] Qiao L T, Nahrstedt K. Watermarking schemes and protocols for protecting rightful ownership and customers's rights [J]. Journal of Visual Communication and Image Representation, 1998, 9(3):194 - 210.

[22] 周 军.图像数字水印的攻击方法及对策研究[J].科技与经济,2006(16):100 - 104.