文章编号: 1001-0920(2004)02-0191-04

抗主动攻击的隐秘通信方案

刘春庆, 戴跃伟, 王执铨
(南京理工大学 自动化系, 江苏 南京 210094)

摘　要: 针对隐秘通信过程中存在的破坏、篡改、伪造 3 种主动攻击行为进行了比较系统地分析, 使用计算复杂性理论给出了明确的定义, 运用密码学中的密码 Hash 函数和数字签名技术, 提出了两种能够抗击主动攻击行为的隐秘通信方案, 从理论上证明了该方案对主动攻击的安全性.

关键词: 隐秘通信; 主动攻击; 安全性; 认证技术
中图分类号: TP39　　**文献标识码**: A

Securing schemes of steganography against active attacks

LIU Chun-qing, DAI Yue-wei, WANG Zhi-quan
(Department of Automation, Nanjing University of Science and Technology, Nanjing 210094, China. Correspondent: LIU Chun-qing, E-mail: liu-chunqing@sohu.com)

Abstract: The active attacks (namely damage, substitution fraudulent and impersonative fraudulent) to steganographic system are discussed. They are defined from a complexity-theoretic point of view. By applying secure cryptographic Hash function and public-key signature scheme, two novel steganographic systems are constructed. The security of the systems against these attacks is proved in theory.

Key words: steganographic system; active attack; security; authentication technique

1 引　　言

信息隐藏技术的两个主要应用分支为数字水印和隐秘通信[1, 2]. 其中隐秘通信是将所要传送的秘密消息嵌入到隐秘载体(即公开传输的普通多媒体数据)中, 形成与隐秘载体相似的隐秘对象在公开信道上传输, 从而实现数据的秘密传送和安全保护. 另一方面, 隐秘通信系统中的攻击者必然采用一定的技术手段对从公共信道中所接收的隐秘对象进行秘密消息的提取、破坏、篡改或伪造, 以损害通信方的利益. 依据攻击行为的危害程度可将其分为主动攻击和被动攻击两大类. 被动攻击是指攻击者监听公开传播的数据, 从中寻找隐秘对象并进行分析, 提取所嵌入的秘密消息; 而主动攻击除了试图提取秘密消息外, 还要进一步破坏、篡改隐秘对象中的秘密消息, 甚至以发送者的身份伪造隐秘对象进行发送, 使接收者受骗. Simmons 于 1983 年提出的"囚犯问题"[3]可清楚地表明这种情况: Alice 和 Bob 因犯罪被逮捕并关押在不同的囚室内, 他们想策划逃跑, 但是两人之间的所有通信都要在看守 Wendy 的监视下进行. Wendy 不允许他们进行加密通信, 并且她一旦发觉可疑的通信会将他们都送入严密守卫的囚室, 因此, Alice 和 Bob 必须以隐秘的方式进行通信以免引起 Wendy 的疑心. 假设 Alice 和 Bob 在被逮捕前已交换了一些秘密信息, 即密钥 K, 则 Alice 在密钥 K 的控制下将要传送的秘密消息 M 嵌入到随机选取的一个普通载体 C 中形成隐秘对象 S, Alice

收稿日期: 2002-11-27; **修回日期**: 2003-03-10.
基金项目: 博士点基金资助项目(20020288025); 江苏省自然科学基金资助项目(BK2001054).
作者简介: 刘春庆(1965—), 男, 山东昌邑人, 博士, 从事信息隐藏理论与技术研究; 王执铨(1939—), 男, 湖北武汉人, 教授, 博士生导师, 从事信息安全、动态大系统建模与控制、混沌控制等研究.

在公开信道上将S传送给Bob, 希望Wendy不会注意到所嵌入的消息. 因为Bob知道Alice所采用的数据嵌入方法和嵌入过程中所使用的密钥K, 所以他能恢复秘密消息M. 另一方面,Wendy监视到隐秘对象S后进行分析, 确定Alice和Bob是否进行非法通信, 这时她所充当的是被动攻击者的角色. 更进一步地,Wendy出于某种恶意目的, 实施主动攻击, 在提取了M后将自己拟定的消息M′嵌入到隐秘对象S中形成伪造的隐秘对象S′发送给Bob, 或直接伪造一个隐秘对象S′发送给Bob, 这时Bob如果提取了消息M′并误认为是Alice所发送的信息, 就会因受骗而跌入Wendy设计的陷阱中. 因此, 可以设想Alice和Bob为防止在通信时受骗, 在被捕前还约定了一个Alice所独有的任何人都不能冒充的印记,Bob可以通过检测所提取的消息中是否含有Alice的特定印记来判断该消息是否为Alice发送而且是否被篡改过.

关于数字水印系统中的主动攻击行为的研究文献[4,5]比较常见, 研究成果也比较多, 而隐秘通信系统中的主动攻击行为却鲜见提及, 但对其进行研究在许多方面如军事、外交等关系到国家安全的通信领域中无疑有着重要的现实意义.

2 隐秘通信系统中的主动攻击行为

本文假设攻击者只能直接观测到公开信道上传输的数据, 而不能接触到数据的嵌入和提取过程.

2.1 隐秘通信系统的传统模型

参照文献[6]中对数字水印系统的描述方法, 可将隐秘通信系统描述为一个多项式时间的概率算法的四元组

$$\Sigma = \langle G, E, A, D \rangle. \tag{1}$$

1) G为密钥生成算法: 输入1^n(由n个1组成的字符串), G输出1个n bits的密钥K. 因为G是随机的, 所以每次可能产生不同的密钥.

2) E为数据嵌入算法: 对于给定的隐秘载体C, 秘密消息M和隐秘密钥K, E输出隐秘对象S: $S = E(C, M, K)$. 假定C和S保持知觉相似性.

3) A为主动攻击算法: 输入S和A, 输出S'= $A(S)$. 假定S和S'保持知觉相似性.

4) D为数据提取算法: 输入S'和K,D输出M' = $D(S', K)$.

2.2 破坏攻击

攻击者截获隐秘对象后, 在保持知觉相似性的前提下, 利用某种数字处理方法如有损压缩、重采样等对隐秘对象进行处理后再发送出去, 如果嵌入算法的鲁棒性不强, 接收者提取出的秘密消息M'不是原来的秘密消息M, 则认为攻击者进行了一次成功的破坏攻击.

一般地, 可将破坏攻击定义为:

定义1(破坏攻击) 设式(1)为隐秘通信系统, 如果攻击者能够找到一种多项式时间的算法A_D, 当输入为S, 输出为$S' = A_D(S)$,S'与S保持知觉相似性, 使得$M' = D(S', K) \neq M$, 则称攻击者实施了一次成功的破坏攻击.

显然, 信道失真可以作为破坏攻击来处理.

2.3 篡改攻击

篡改攻击是指攻击者不仅能截收到隐秘对象, 而且篡改了秘密消息的内容, 在保持修改后的隐秘对象S'与原隐秘对象S知觉相似性的前提下将S'发送出去. 如果接收者提取出了篡改的隐秘消息M', 则认为攻击者进行了一次成功的篡改攻击.

一般地, 可将篡改攻击定义为:

定义2(篡改攻击) 设式(1)为隐秘通信系统, 如果攻击者能够找到一种多项式时间的算法A_S, 当输入为S和M_A, 输出为$S' = A_S(S, M_A)$,S'与S保持知觉相似性, 使得$M' = D(S', K) = M_A$, 则称攻击者实施了一次成功的篡改攻击.

2.4 伪造攻击

攻击者在没有观察到隐秘对象的条件下, 拟定消息M_A, 运用某种方法伪造一个隐秘对象S', 并冒充合法发送者的名义发送出去. 如果接收者提取出了伪造的消息M_A, 则称攻击者实施了一次成功的伪造攻击. 由此可以给出如下定义:

定义3(伪造攻击) 设式(1)为隐秘通信系统, 如果攻击者能够找到一种多项式时间的算法A_I, 当输入为C',M_A时, 输出为$S' = A_I(C', M_A)$,S'与C'保持知觉相似性, 使得$M' = D(S', K) = M_A$, 则称攻击者实施了一次成功的伪造攻击.

由以上描述的3种主动攻击方式的定义可见, 主动攻击者采用的是以假乱(充)真, 试图用假消息欺骗接收方的策略, 而接收方如果对收到的消息缺乏有效的鉴别方法, 可能会以假为真、上当受骗, 所以传统的隐秘通信系统不能防止对手的主动攻击.

3 能够防止主动攻击的隐秘通信方案

下面将密码学中的认证技术运用于隐秘通信系统, 给出了基于密码Hash函数和数字签名的两种隐秘通信方案, 并从理论上证明该方案能够防止此类主动攻击.

3.1 基于密码 Hash 函数的隐秘通信方案

3.1.1　密码 Hash 函数认证体制[7,8]

密码Hash函数H_H是在密钥K_H的控制下将任意长字符串M映射成一个输出字符串H_M的函数, 称$H_M = H_H(M, K_H)$为字符串M的Hash值或数字指纹. 虽然不能从H_M求出字符串M, 但可以验证任一给定字符串M'是否与M具有相同的Hash值.

将密码Hash函数认证体制记为一个概率算法的三元组$\langle G_H, H_H, V_H\rangle$, 其中$G_H$表示密钥生成过程, H_H表示Hash值的生成过程, V_H表示证实过程.

1) 密钥生成算法G_H: 输入1^k, $G_H(1^k)$输出1个k bits的认证密钥K_H.

2) Hash算法: 输入M和M', 计算M和M'的Hash值, $H_M = H_H(M, K_H)$, $H_{M'} = H_H(M', K_H)$.

3) 证实算法V_H: 输入H_M和$H_{M'}$, 判断M'是否为M.

$$V_H(M' = M) = \begin{cases} \text{True, if } H_{M'} = H_M; \\ \text{False, if } H_{M'} \neq H_M. \end{cases} \tag{2}$$

该体制的安全性在于, 在同一个认证密钥K_H控制下伪造一个消息M'使其具有与给定的消息M相同的Hash值, 或寻找两个不同的消息M'与M, 使其具有相同的Hash值, 在计算上是不可行的.

3.1.2　基于密码 Hash 函数的隐秘通信方案

假设密码Hash函数$H_M = H_H(M, K_H)$对破坏、篡改、伪造攻击是安全的, 在式(1)的基础上构建一个新的隐秘通信方案

$$\bar{\Sigma} = \langle \bar{G}, \bar{E}, \bar{A}, \bar{D}, \bar{J}\rangle. \tag{3}$$

1) $\bar{G}$为密钥生成算法: 输入1^n, $G(1^n)$输出1个n bits的隐秘密钥K; 输入1^k, $G_H(1^k)$输出1个k bits的认证密钥K_H.

2) $\bar{E}$为数据嵌入算法: 输入隐秘载体C, 秘密消息M和隐秘密钥K, 首先计算$H_M = H_H(M, K_H)$, $\bar{M} = M \| H_M$; 然后在K控制下将$\bar{M}$嵌入到C中, $\bar{E}$输出隐秘对象$S = \bar{E}(C, \bar{M}, K)$. 假定$C$和$S$保持知觉相似性. 符号$\|$表示两个字符串的联接.

3) $\bar{A}$为主动攻击算法: 输入S, $\bar{A}$输出攻击后的隐秘对象$S' = \bar{A}(S)$. 假定S和S'保持知觉相似性.

4) $\bar{D}$为数据提取算法: 输入S'和K, $\bar{D}$输出$\bar{M}' = \bar{D}(S', K)$.

5) $\bar{J}$为消息鉴别算法: 输入$\bar{M}'$, 首先将$\bar{M}'$分裂为两部分: $\bar{M}' = M' \| H_{M'}$; 然后计算$H_{M''} = \mathrm{Hash}(M', K_H)$. 若$H_{M'} = H_{M''}$, 则$\hat{J}$输出$M'$为真消息, 否则$\hat{J}$输出$M'$为假消息.

现在, 发送者将秘密消息M连同其数字指纹H_M一起嵌入到隐秘载体C中. 如果一个消息是由攻击者伪造的或攻击者修改处理过, 那么接收者就有充分的理由拒绝它. 因为根据Kerckhoffs假设, 攻击者了解隐秘通信系统式(5)的全部知识, 但不知道密钥K和K_H. 如果实施破坏、篡改和伪造攻击, 就意味着寻找一个多项式时间算法, 使得两个不同的消息具有相同的Hash值或一个消息在不同的密钥下具有相同的Hash值, 这与密码Hash函数的潜在安全性是矛盾的. 基于文献[6]的定理3.1, 本文给出了下面的定理:

定理 1　对于隐秘通信系统(3), 如果密码Hash函数对破坏、篡改、伪造攻击是安全的, 那么Σ对破坏、篡改、伪造攻击也是安全的.

证明　1) 证明Σ对破坏攻击是安全的. 反证法: 假设$\bar{\Sigma}$对破坏攻击是不安全的, 那么根据Kerckhoffs假设和定义1, 攻击者能够找到一种多项式时间的算法$\bar{A}_D$, 当输入为S, 输出为$S' = \bar{A}_D(S)$, S'与S保持知觉相似性, 使得$M' = \bar{D}(S', K) \neq M$, 而$H_{M'} = H_M$, 意味着两个不同的消息具有相同的Hash值, 这与定理的条件矛盾. 因此, Σ对破坏攻击是安全的.

2) 证明$\bar{\Sigma}$对篡改攻击是安全的. 反证法: 假设Σ对篡改攻击是不安全的, 那么根据Kerckhoffs假设和定义2, 攻击者能够找到一种多项式时间的算法$\bar{A}_S$, 当输入为S, 输出为$S' = \bar{A}_S(S, M_A)$, S'与S保持知觉相似性, 使得$M' = \bar{D}(S', K) = M_A \neq M$, 而$H_{M'} = H_M$, 意味着两个不同的消息具有相同的Hash值, 这与定理的条件矛盾. 因此, $\bar{E}$对篡改攻击是安全的.

3) 证明Σ对伪造攻击是安全的. 反证法: 假设$\bar{\Sigma}$对伪造攻击是不安全的, 那么根据Kerckhoffs假设和定义3, 攻击者能够找到一种多项式时间算法$\bar{A}_I$, 输入C', M_A和K_{HA}(攻击者伪造的对M_A的认证密钥), 输出$S' = \bar{A}_I(C', \bar{M}_A)$, 其中$\bar{M}_A = M_A \| H_H(M_A, K_{HA})$; S'与C'保持知觉相似性, 使得$M' = \bar{D}(S', K) = M_A$. 而$H_H(M_A, K_{HA}) = H_H(M', K_H) = H_H(M_A, K_H)$, 意味着一个消息在不同的密钥下具有相同的Hash值, 这与定理的条件矛盾. 因此, Σ对伪造攻击是安全的. □

3.2　基于数字签名的隐秘通信方案

3.2.1　数字签名体制[8,9]

可将数字签名体制记为一个概率算法的三元组$\langle G_S, S_S, V_S\rangle$, 其中: G_S表示密钥生成过程, S_S表示签名生成过程, V_S表示签名证实过程.

1) 密钥生成算法G_S: 输入1^k, $G_S(1^k)$ 输出1对k bits的签名密钥K_P和K_S, 其中: K_P为公钥, K_S为私钥

2) 签名生成算法S_S: 输入M 和K_S, 计算对M的签名$s = S_S(M, K_S)$.

签名证实算法V_S: 输入M, s和K_P, 判断s是否为M 的签名

$$V_S(M, s, K_P) = \begin{cases} \text{True, if } s = S_S(M); \\ \text{False, if } s \neq S_S(M). \end{cases} \quad (4)$$

该体制的安全性在于, 从M 和其签名推出私钥K_S, 或伪造一个M', 使M'和s可被证实为真, 在计算上是不可行的

3.2.2 基于数字签名的隐秘通信方案

假设数字签名体制$\langle G_S, S_S, V_S \rangle$对破坏、篡改、伪造攻击是安全的, 在式(1) 的基础上构建一个新的隐秘通信方案

$$\hat{\Sigma} = \langle \hat{G}, \hat{E}, \hat{A}, \hat{D}, \hat{J} \rangle. \quad (5)$$

1) $\hat{G}$为密钥生成算法: 输入1^n, $G(1^n)$ 输出1个n bits的隐秘密钥K; 输入1^k, $G_S(1^k)$ 输出一对k bits的签名密钥(K_P, K_S), 其中: K_P为公钥, K_S为私钥

2) $\hat{E}$为数据嵌入算法: 输入隐秘载体C, 秘密消息M 和隐秘密钥K, 先计算$s = S_S(M, K_S)$, $M = M \| s$, 再在K控制下将M嵌入C中, $\hat{E}$输出隐秘对象$S = \hat{E}(C, M, K)$. 假定C和S保持知觉相似性

3) $\hat{A}$为主动攻击算法: 输入S, $\hat{A}$输出$S' = \hat{A}(S)$. 假定S和S'保持知觉相似性

4) $\hat{D}$为数据提取算法: 输入S'和K, $\hat{D}$输出$M' = \hat{D}(S', K)$.

5) $\hat{J}$为消息鉴别算法: 输入M', 将M'分裂为两部分: $M' = M' \| s$. 若$V_S(M', s, K_P) = \text{True}$, 则$\hat{J}$输出$M'$为真消息, 否则$\hat{J}$输出$M'$为假消息

发送者现将秘密消息M 连同其数字签名s一起嵌入到隐秘载体C 中. 若一个消息是由攻击者伪造的或被攻击者修改处理过, 那么接收者能够识别该消息的虚假性. 因为根据Kerckhoffs假设, 攻击者了解隐秘通信系统(5) 的全部知识, 但不知道密钥K和K_S. 如果实施破坏、篡改和伪造攻击, 就意味着寻找一个多项式时间算法, 使得两个不同的消息具有相同的签名或一个消息在不同的密钥下具有相同的签名, 这与数字签名体制的潜在安全性是矛盾的.

定理2 对于隐秘通信系统式(5), 如果数字签名算法对破坏、篡改、伪造攻击是安全的, 那么$\hat{\Sigma}$对破坏、篡改、伪造攻击也是安全的.

该定理的证明与定理1的证明类似, 略.

4 结 语

本文初步分析了传统的隐密通信系统中存在的破坏、篡改、伪造3种主动攻击行为, 同时基于计算复杂性理论明确定义了这3种攻击模型. 借助于密码学中的认证技术, 通过将秘密消息与其Hash值或数字签名进行捆绑隐藏, 构建了两个能够挫败主动攻击的隐秘通信系统, 并对这两个系统的安全性进行了理论证明, 为进一步研究隐秘通信系统奠定了理论基础. 但它所增加的嵌入数据量, 必然对隐秘通信系统的鲁棒性、不可感知性和信道容量产生或多或少的影响, 而且秘密消息嵌入前的预处理和提取后的鉴别过程增加了系统的成本与工作时间, 这些既是增强系统的安全性所必须付出的代价, 也是今后需要进一步解决的问题.

参考文献(References):

[1] 汪小帆, 戴跃伟, 茅耀斌. 信息隐藏技术——方法与应用[M]. 北京: 机械工业出版社, 2001. 22-27.

[2] Petitcolas F A P, Anderson R J, Kuhn M G. Information hiding: A survey[J]. *Proc of the IEEE*, 1999, 87(7): 1062-1078.

[3] Simmons G J. The Prisoners' problem and the subliminal channel[A]. *Advances in Cryptology, Proc of Crypto* 83[C]. Santa Barbara: Plenum Press, 1984. 51-67.

[4] Qiao L, Nahrstedt K. Watermarking schemes and protocols for protecting rightful ownership and customer's rights[J]. *J of Visual Communication and Image Representation*, 1998, 9(3): 194-210.

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

[6] Katzenbeisser S, Veith H. Securing symmetric watermarking schemes against protocol attacks[A]. *SPIE the Int Society for Optical Engineering*[C]. California, 2002. 260-268.

[7] Jueneman R R, Matyas S M, Meyer C H. Message authentication[J]. *IEEE Communications Magazine*, 1985, 23(9): 29-40.

[8] 王育民, 刘建伟. 通信网的安全——理论与技术[M]. 西安: 西安电子科技大学出版社, 2000. 233-297.

[9] Schneier B. A primer on authentication and digital signatures[J]. *Computer Security J*, 1994, 10(2): 38-40.