安全隐写系统的信息理论分析*)

吕 欣[1,2] 马 智[1] 冯登国[1]
(中国科学院研究生院信息安全国家重点实验室 北京 100049)[1]
(国家信息中心 北京 100045)[2]

摘 要 本文采用信息论的观点对信息隐藏系统做了分析,给出了评价一个安全隐写系统的一般思路。介绍了用香农熵和相对熵来定义安全隐藏系统的方法,并对两种方法作了比较。得出了设计一个安全隐藏系统应注意的几个问题。

关键词 信息隐藏,信息论,熵,无条件安全

Information-Theoretically Analysis of Secure Steganographic System

LU Xin[1,2] MA Zhi[1] FENG Deng-Guo[1]
(Graduate School of Chinese Academic Sciences, the State Key Laboratory of Information Security, Beijing 100049)[1]
(State Information Center, Beijing 100045)[2]

Abstract Information-Theoretical methods are applied to analysis Secure Steganographic System in this paper. Shannon entropy and relative entropy are introduced to evaluate a secure Steganographic System separately. And the two methods are compared and some conclusions are drawn in the later of the paper, which are helpful to design and analysis a Secure Steganographic System.

Keywords Information hiding, Information theory, Entropy, Unconditional security

1 引言

经典密码学是研究明文消息的屏蔽,而信息隐藏企图掩饰信息的存在性。它是一门古老的技术。由于计算机通信网络的发展,在近几年得到了广泛的重视。Simmons的"囚犯问题"(Prisoner problem)的提出,使得信息隐藏步入了科学研究的轨道。在"囚犯问题"中,包含三个角色:Alice(A)、Bob(B)、Willie(W)。Alice、Bob是两个囚犯,他们通过交换信息来预谋逃跑计划,但他们信息的交换都要通过看守员Willie。所以A和B通过共享某一秘密来把预谋信息隐藏在一与其毫不相干的书信中,方可逃脱W的检查。这就是一个典型的信息隐藏问题。

具体地说,加密(encryption)技术,是采用密钥(key)来加密要保密的消息,发送方将加密后的密文(ciphertext)通过公共信道(public channel)传送给接收方,接收方利用其拥有的特定密钥进行解密,恢复出明文(plaintext)。而信息隐藏,是将秘密消息隐藏在其它消息中。一个隐写系统主要由以下几部分组成(如图1)。

(1) 嵌入消息M,是通信双方(如发送方A和接收方B)想要共享的秘密信息集合。

(2) 载体C,是发送方A选择的一组与消息M无关且不易引起怀疑的消息的集合,它可以是一幅图像,一段文本或一个声音文件。一般C可以由发送方A从载体随机发生器R中随机产生。

(3) 嵌入算法F和提取算法F^{-1},通常消息M的嵌入需要某种算法来实现。F^{-1}是F的逆过程。

(4) 嵌入消息后的载体称为隐文S。

(5) 消息嵌入和提取的过程是在一组密钥的控制下进行的,分别称为嵌入密钥K和提取密钥K′。K和K′如果相同,则称之为一组对称密钥(私钥),否则称之为非对称密钥(公钥)。所对应的嵌入算法F和提取算法F^{-1}则相应称之为私钥算法和公钥算法。本文主要讨论私钥算法。

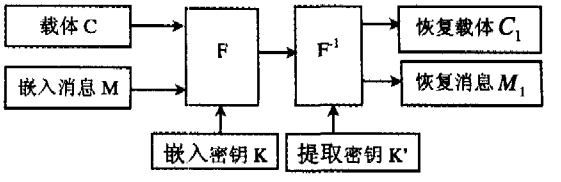

图1 信息隐藏系统示意图

2 信息论和无条件安全

类似于密码系统,衡量一个保密系统的方法有两种,一种是计算安全性,另一种称为无条件安全性[2]。一个安全系统是计算上安全的,是指利用已有的方法破译该系统所需的努力超过了敌手的破译能力(诸如时间、空间和资源)或破译系统的难度等价于解一个公认的数学难题。称一个保密系统是无条件安全的,又称理论安全性,是指一个具有无限资源(诸如时间、空间和资源等)的破译者也无法破译该系统。本文主要针对信息隐藏系统的理论安全性进行讨论。

Claude Shannon在1949发表的"保密系统的信息理论"的论文[1],用信息论的观点来分析信息保密问题,使得信息论

*)基金项目:国家杰出青年科学基金(60025205)和国家自然科学基金项目(60273027,60403004)。吕 欣 博士,助理研究员。马 智 副教授,研究方向为信息安全。冯登国 教授,博士生导师,研究方向为网络信息安全。

成为研究信息安全的一个重要理论基础。本文从信息论的观点入手来讨论安全隐藏系统。

3 安全隐藏系统的信息论模型

3.1 攻击类型

对于一个信息隐藏系统的攻击，可分为被动(passive)攻击和主动(active)攻击。

(1)如果攻击只为了检测到系统中传输的信息中是否被嵌入了秘密消息，即只需要检测到秘密隐藏信息的存在性，把这种攻击称为被动攻击。本文主要考虑被动攻击。

(2)另一种是，不仅要检测到秘密隐藏信息的存在，而且要恢复出嵌入的信息的内容，这种攻击称为主动攻击。

3.2 香农熵信息论模型

根据信息论的概念。熵是某种不确定性的标志。而条件熵是测度当知道消息 A 后消息 B 仍有的(剩余)的不确定性(信息量)。而联合熵 $H(A,B)$ 则定义为：

$$H(A,B)=H(A)+H(B|A) \tag{1}$$

互信息 $I(A,B)$ 用于已知的 B 取值后所提供的有关 A 的信息。表述如下：

$$I(A,B)=H(A)-H(A|B) \tag{2}$$

定义明文消息的熵为 $H(M)$，载体的熵为 $H(C)$，密钥的熵为 $H(K)$，隐文的熵为 $H(S)$。已知隐文和载体条件下明文消息的条件熵为 $H(M|(S,C))$。下面分情况进行讨论各种攻击的安全性，并做如下假设：

(1)满足 Kerchhoff 假设，即隐写算法已知。

(2)攻击者具有无限的时间和计算资源。

定义 1 研究一个信息隐藏系统 (M,C,S,K)，密钥 K 和明文消息 M，载体 C 可以唯一地确定隐文 S，S 和 K，C 可以唯一地确定明文消息 M 和 C，即满足如下关系：

$$H(S|C,M,K)=0 \tag{3}$$

$$H(M|K,S)=0 \tag{4}$$

如果满足

$$I(M,S)=H(M)-H(M|S)) \tag{5}$$

就称该系统是无条件安全的。这里 $K,M;K,C$ 之间统计独立。其物理意义是，在攻击者知道 S,M 的不确定性不会减小。

定理 1 对于一个无条件安全隐藏系统 (M,C,S,K)，有

$$H(C)>0 \tag{6}$$

证明：由熵的定义知道：$H(C)\geqslant 0$

假定攻击者直到 C。即 $H(C)=0$

但在实际情况下，如果一个攻击者知道 S 和 C 后，很容易通过比较发现是否 S 和 C 相等，来检测隐藏消息 M。而本文考虑攻击类型被动攻击，即如果检测到一个隐藏消息的存在性，那么就说该隐藏系统不安全。所以一个无条件安全隐藏系统 C 的熵必须大于零。

定理 2 无条件安全隐藏系统必有下式成立：

$$H(K|S)\geqslant H(M) \tag{7}$$

该式确保了通过攻击 K 来获取嵌入消息 M 的不可行性。

证明：由于 $I(M,S)=H(M)-H(M|S))$，

因此 $H(M)=H(M|S)\leqslant H(M,C,K|S)=H(K|S)+H(M|S,C,K)$，又 $H(M|S,C,K)=0$。

所以无条件安全隐写系统的必要条件是

$$H(K|S)\geqslant H(M)$$

3.3 基于相对熵的信息理论模型

文[8]借助于相对熵(relative entropy)和假设检验的观点对信息隐藏系统建立安全模型。这里对这种思想加以介绍和分析。

3.3.1 *假设检验和两类错误* 假设检验利用样本的实际资料来检验事先对总体某些数量特征所作的假设是否可信的一种统计方法。当我们把真实的原假设当成假的而加以拒绝，称为第一类错误，也称弃真错误，犯第一类错误的概率就是显著性水平大小；当我们把不真实的原假设当作真的而加以接受，称为第二类错误，也称纳伪错误，犯第二类错误的概率是不确定的。

对于一个信息隐藏系统 (M,C,S,K)，攻击者可以通过假设检验的方法来判断是否有秘密消息的嵌入。设

H_0：表示有秘密消息的嵌入。

H_1：表示没有秘密消息的嵌入。

攻击者可能采取各种攻击方法来判断秘密消息的存在，但有时却检测不到秘密消息的存在犯第一类错误。而有时本来发送方没有嵌入任何的秘密消息，但攻击者却错误地检测到秘密消息的存在而犯第二类错误。

3.3.2 *相对熵* 相对熵，又叫鉴别信息(discrimination information)、交叉熵(cross-entropy)，是由 I. J. Good，L. J. Savage 和 S. Kullback 等提出并发展起来的，后来这一概念在信号处理中得到了应用和推广，成为现代信息论的重要组成部分。

已知随机变量 X 取值为 $\{a_1,a_2,\cdots,a_k\}$，且 X 的分布与假设条件 H_1，H_2 有关。即 X 在假设条件 H_1，H_2 的分布分别为：$P_1(X)$、$P_2(X)$。另外设：

$p_1(a_k)=p(a_k|H_1)$，$p_2(a_k)=p(a_k|H_2)$，表示在假设条件 H_1，H_2 下，X 取 a_k 的条件概率。

随机变量 X 在假设条件为 H_2 下的相对熵定义为：

$$D(p_2,p_1)=\sum_{k=1}^{k}p_2(a_k)\log\frac{p_2(a_k)}{p_1(a_k)} \tag{8}$$

相对熵 $D(h_2,h_1)$ 表示随机变量 X 在假设条件 H_2 下进行观察所平均得到的倾向于 H_2 的信息量。

3.3.2 *基于相对熵信息理论模型*

定义 1 研究一个针对被动攻击的信息隐藏系统 (M,C,S,K)，假设条件同定义 1，P_C，P_S 分别表示载体 C 和隐文 S 的概率分布，如果其相对熵满足：

$$D(P_C,P_S)\leqslant\theta$$

则称该系统是 θ 安全。如果 $\theta=0$，则称该系统是绝对安全的。

同时，可以用假设检验和相对熵相结合来检秘密信息的存在，结论如下：

定理 3 对于一个针对被动攻击的信息隐藏系统 (M,C,S,K)，设 α,β 分别表示攻击者有嵌入信息但检测不到(错误)的概率和没有嵌入信息但错误地检测到一个秘密信息的存在的概率。如果满足如下关系：

$$d(\alpha,\beta)\leqslant\theta \tag{10}$$

则称该隐藏系统是 θ 安全的[8]。

其中：$d(\alpha,\beta)\leqslant\alpha\log_2\frac{\alpha}{1-\beta}+(1-\alpha)\log_2\frac{1-\alpha}{\beta}$

4 两种模型的意义和启示

两类信息隐藏的模型都是从信息论的角度出发来讨论信息隐藏系统的安全性。前者利用了香农信息熵进行讨论，其

方法类似于香农保密系统无条件安全的分析。它便于了解一个安全系统中各个部分之间熵的关系。而后一种方法是利用了相关熵和假设检验的概念，不仅可以用来定义无条件安全隐藏系统，还可以用来分析 θ-安全系统。与香农熵相比，相关熵不方便反映系统各个部分之间的关系。但它们对隐藏的模型的研究和安全隐写系统的设计都有指导意义，都反映出了设计一个无条件安全隐藏系统应该注意的几个方面。

(1)一个安全的隐写系统应该能抵抗被动攻击。即不但不能让攻击者检测到嵌入消息的内容，而且不能检测到秘密消息存在。这才算一个安全的隐写系统。

(2)载体信息 C 不能暴露给攻击者，即必须满足(6)式。

(3)隐写系统的安全性不能依赖于隐写算法。而应该依赖于密钥 K 的保密性。要满足(7)式，避免对密钥进行攻击。

(4)应尽可能地保证隐文消息 S 和载体消息 C，在统计概率分布上的不可区分。对于无条件安全隐写系统，要保证其严格的统计概率分布上的不可区分。

结论 信息隐藏是信息安全领域中一个新兴学科。本文从信息论的观点出发，对信息隐藏系统的安全性进行了定义和分析。指出了设计一个安全信息隐藏系统应注意的几个问题，对隐藏系统的安全性研究和设计有一定的指导意义。

参 考 文 献

1 Simmons G J. The prisoners' problem and the subliminal channel. In: Advances in Cryptology: Proc. of Crypto 83, Plenum Press, 1984. 51～67

2 ollner J Z, Federrath H, Klimant H, et al. Modeling the security of steganographic systems. In: 2nd Intl. Workshop on Information Hiding, LNCS, Springer, 1998. 344～354

3 Anderson R J, Petitcolas F A P. On The Limits of Steganography. IEEE Journal of Selected Areas in Communications, 1998, 16(4): 474～481

4 Shannon C E. Communication theory of secrecy systems. Bell System Technical Journal, 1949, 28(10): 656～715

5 P--tzmann B. Information hiding terminology. In: First Intl. Workshop on Information Hiding., LNCS 1174, springer, 1996

6 常迥. 信息理论基础. 北京：清华大学出版社，2001

7 冯登国. 密码学导引. 北京：科学出版社，1999

8 Cachin C. An Information Theoretic model for Steganography. In: Proc. of the Second Intl. Workshop on Information Hiding. LNCS 1525, Springer, 1998. 306～318

(上接第 108 页)

以上 TPC-C 测试的结果表明数据库异构集群在 OLTP 处理中仍具有良好的可扩展性，次线性的加速比，而且还能提供高效费比的并行处理服务。同时我们也发现异构节点的计算能力在集群中考虑网络开销时会有一定的影响，因此异构节点间的负载是不均匀的，在一定的网络环境下，异构节点的数量与系统性能存在一个最佳平衡点。

目前，我们只假设机器的 CPU 和内存硬件配置的异构，对于软件包括数据库在内的异构我们需要再做进一步的研究。

参 考 文 献

1 邸烁，郑纬民，王鼎兴，沈美明. 并行 WWW 服务器集群请求分配算法的研究. 软件学报，1999，10(7)

2 Baker M, Apon A, Buyya R, Jin H. Cluster computing and applications. Encyclopedia of Computer Science and Technology. New York: Marcel Dekker, Aug. 2001, 45

3 Gancarski S, Naacke H, Pacitti E, Valduriez P. Parallel processing with autonomous databases in a cluster system. In: Proc. of on the Move to Meaningful Internet Systems, DOA, CoopIS and ODBASE Confederated Intl. Conf. 2002. Oct. 2002

4 Fox A, Gribble S D, Chawathe Y, et al. Cluster-based scalable network services. ACM SIGOPS Operating Systems Review, 1997, 32(5)

5 Carrera E V, Bianchini R. Efficiency vs. Portability in Cluster-Based Network Servers. In: Proc. of the eighth ACM SIGPLAN symposium on Principles and practices of parallel programming, June 2001

6 Shen Kai, Yang Tao, Chu Lingkun. Cluster load balancing for fine-grain network services. In: Proc. of the Intl. Parallel and Distributed Processing Symposium, April 2002

7 Carrera E V, Bianchini R. Evaluating Cluster-Based Network Servers. In: Proc. of the Ninth IEEE Intl. Symposium on High Performance Distributed Computing(HPDC'00), Aug. 2000

8 Pastor L, Orero J L B. An efficiency and scalability model for heterogeneous clusters. In: Proc. of the 2001 IEEE Intl. Conf. on Cluster Computing, 2002. 427～434

9 Cuellar G D, Salzar D A. A parallelization technique that improves performance and cluster utilization efficiency for heterogeneous clusters of workstations. In: Proc. of the IEEE Intl. Conf. on Cluster Computing, 2002. 275～283

10 Kalinov A. Scalability analysis of matrix-matrix multiplication on heterogeneous clusters. In: The Third Intl. Workshop on Parallel and Distributed Computing, July 2004. 303～309

11 Bosque J L, Perez L P. Theoretical scalability analysis for heterogeneous clusters. In: IEEE Intl. Symposium on Cluster Computing and the Grid, April 2004. 285～292

12 Teodoro G, Tavares T, Coutinho B, et al. Load balancing on stateful clustered web servers. In: Proc. of the 15th Symposium on Computer Architecture and High Performance Computing, Nov. 2003

13 Gunther N J. Issues facing commercial OLTP applications on MPP platforms. IEEE 1994

14 蒋江，张民选，廖湘科. 异构集群系统中一种基于资源的负载平衡算法的设计与模拟. 小型微型计算机系统，2003，24(4)

15 Xiao Li, Zhang Xiaodong, Qu Yanxia. Effective load sharing on heterogeneous networks of workstations. In: Proc. of the 14th Intl. Parallel and Distributed Processing Symposium, May 2000. 431～438

16 Wang Min, Ding Weiqun, Li Hong. E-differentiation for analyzing scalability of parallel algorithms on parallel architectures. In: Proc. of Intl. Conf. on Information, Communications and Signal Processing(ICICS), Singapore, Sept. 1997

17 Ji Yongchang, An Hong, Ding Weiqun, Chen Guoliang. A scalability metric for algorithm-machine on NOW and MPP. IEEE 2000