密码体制与分布式 Web 数据库的安全设计

Cryptography and the Security Design of Distributed Database

姚炎炎　陈怀义　郑若忠　宁　洪

（国防科技大学计算机学院　长沙 410073）

Abstract　The security is the most important thing that we should solve before connecting distributed dataase systems to the Internet. In this article, we briefly introduce aspects of the problem, and then, its characteristics and implementation.

Keywords　Databases, Data security, Encryption, Decryption, Authentication

1. 引言

20 世纪 90 年代初，Internet 在全球得到了迅猛的发展，公共和私人部门的一些机构愈来愈多地应用电子数据处理，将数据存储在数据库中，因此防止非法泄露、删除和修改数据等是十分重要的问题。

分布式数据库系统（DDBS）是指物理上分布于多个网络站点而逻辑上统一的数据库系统，其体系结构如 1 图所示。其中，Web 服务界面是一种目前常见的用户接口形式。用户首先要选择一个服务器站点，通过该服务器站点的用户接口登录，系统确认用户的合法身份后接受用户提出的事务处理请求，并把用户事务经用户接口转换后由编译层进行语法分析、语义分析、授权检查、事务分解等操作，而后交事务管理层监督执行。分解得到的访问本地数据的子事务，由本地数据库管理系统具体执行；访问远程数据的子事务，则通过通讯系统交给远程服务器站点的事务管理层，由远程服务器站点的事务管理层监督远程数据库管理系统具体执行。

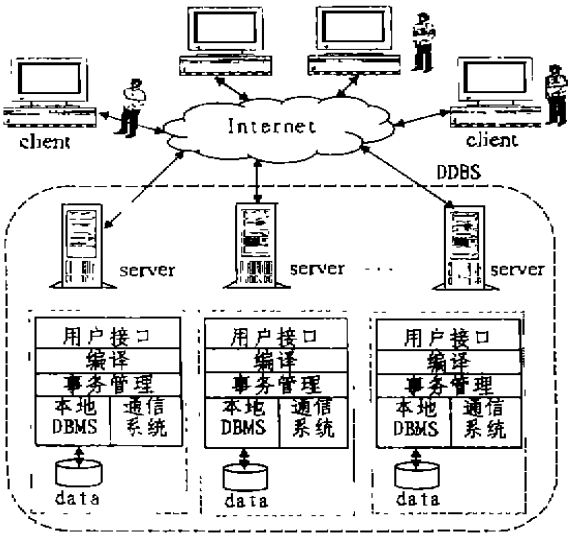

图 1　DDBS 体系结构示意图

DDBS 内的各站点一般通过 Internet 互联，Internet 的发展为各站点之间提供了快捷的通信服务，但是 Internet 的信道是公用、开放的，容易受到恶意攻击，这又威胁着 DDBS 的安全。因此，必须综合采用多种措施来保证它的安全。

2. 数据加密技术

2.1　对称密码体制与加密模式

有两种基本类型：分组密码（block cipher）和序列密码（stream cipher）。前者是在明文分组和密文分组上进行运算，通常分组长为 64 位，但有时更长；序列密码作用在明文和密文数据序列的 1 位或一字节上（有时甚至是一个 32 位的字），利用分组密码，相同的明文用相同的密钥加密永远得到相同的密文；而利用序列密码，每次对相同的明文位或字节加密都会得到不同的密文位或字节。

密码模式通常由基本密码、一些反馈和一些简单运算组合而成。这些运算都很简单，因为信息的安全性依赖于基础的密码，而不依赖于密码模式。

2.1.1　电子密码本模式（Electronic Code Book，ECB）　是使用分组密码的最明显的方式：一个明文分组加密成一个密文分组。它也是最容易的运行模式，因为可以不考虑分组的次序而独立地对每个明文分组进行加密运算。如果一个数据库用 ECB 模式进行加密，那么任意一个记录都可以独立于其他记录被添加、删除或者解密。对于多处理机系统来说，可以并行地对数据进行处理，也就是说，可以并行地对不同的分组同时

姚炎炎　硕士生，研究方向：数据库安全。陈怀义　教授，研究方向：计算机理论。

进行加、解密运算。

ECB 模式下，相同的明文分组永远被加密成相同的密文分组，这就为密码分析者提供了方便，因为在实际情形中，消息格式趋于重复。另外，格式化的消息头和消息尾也是个致命之处。

2.1.2 密码分组链接模式(Cipher Block Chain，CBC)在分组密码中加入一种反馈机制。前一个分组的加密结果被反馈到当前分组的加密中，换句话说，每一分组被用来修改下一分组的加密。每个密文分组不仅依赖于产生它的明文分组，而且依赖于所有前面的明文分组。用数学语言表示为：

$C_i=E_k(P_i \oplus C_{i-1})$

$P_i=C_{i-1} \oplus D_k(C_i)$

由于 CBC 模式仅在前面的明文分组不同时才能将完全相同的明文分组加密成不同的密文分组，因此两个相同的消息仍将被加密成相同的密文。为了防止这种情况的发生，保证每个消息的唯一性，CBC 模式中用加密随机数据作为第一个分组，这个随机数据被称为初始化向量(IV)。

2.1.3 密码反馈模式(Cipher Feed Back，CFB)

分组密码应用于序列密码，就是密码反馈模式。在 CBC 模式下，整个数据分组在接收完之后才能进行加密，而在 CFB 模式下，数据可以在比分组小得多的单元里进行加密，同时 CFB 模式将明文字符连接起来以使密文依赖所有以前的明文，若算法的分组是 n 位，那么 n 位 CFB 用数学语言可表示为：

$C=P_i \oplus E_k(C_{i-1})$

$P_i=C_i \oplus E_k(C_{i-1})$

2.1.4 输出反馈模式(Output Feedback，OFB)

是将分组密码作为同步序列密码运行的一种方法。同密码反馈模式类似，不同的是 OFB 是将前一个 n 位输出分组送入队列最右端的位置，解密是其逆过程。这种方法有时也叫内部反馈，因为反馈机制独立于明文和密文而存在。OFB 模式有一个很好的特性就是在明文存在之前大部分工作可以离线进行。用数学语言表示为：

$C=P_i \oplus S_i; S_i=E_k(S_{i-1})$

$P=C_i \oplus S_i; S_i=E_k(S_{i-1})$

对称密码体制的优点是具有很高的保密强度，可以达到经受国家级破译力量的分析和攻击。另一方面，它的密钥必须通过安全可靠的途径传输，密钥管理成为管理系统安全的关键性因素，使它难以满足系统的开放性要求，对应的密码算法有：DES，IDEA，RC2 等等。

2.2 非对称密码体制(公开密钥密码体制)

非对称密码体制的出现是现代密码学研究的一项重大突破，主要应用于数据加密和数字签名，其主要优点是可以适应开放性的使用环境，密钥管理问题相对简单，但保密程度目前还远达不到对称密码体制的水平，同时其加密速度也远小于对称密码体制。

非对称密码体制一般都是建立在一些难于求解的问题之上的(尤其是指计算上的不可行)(比如：大数分解问题、旅行商问题、背包问题、三方匹配问题等)。根据问题解法的复杂性，问题可被分为一些复杂性类型。如 2 图所示，在底层，P 类包括所有能在多项式时间解决的问题。NP 类问题则包括所有在非确定型图灵机上可用多项式时间解决的问题。NP 类包括 P 类，因为任何在确定型图灵机上可用多项式时间解决的问题，在非确定型图灵机上也是可用多项式时间解决的。另一方面，P=NP(或 P≠NP)还未被证明。在复杂性等级上较 NP 高的是 PSPACE 类问题。PSPACE 中的问题能在多项式空间内解决，且不一定需要多项式时间。PSPACE 包括 NP，但 PSPACE 中的问题被认为比 NP 还要难解。同 NP 类类似，PSPACE 中有部分被称为 PSPACE 完全问题，它具有这样的性质：如果它们中任何一个在 NP 中，则 PSPACE=NP；如果它们中任何一个在 P 中，则 PSPACE=P。至于 EXPTIME 类，则是指可以在指数时间内解决的问题。

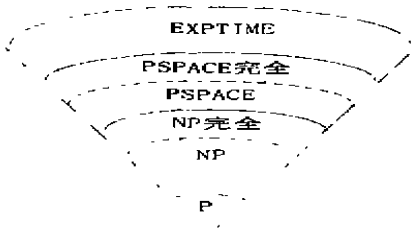

图 2 问题复杂度类

在这些类中密码学家们尤其感兴趣的是 P 类和 NP 类问题。NP 问题与密码学的关系如下：许多对称算法和所有公开密钥算法能够用一个非确定性的多项式时间(算法)攻击。这在理论上很重要，因为它给出了对于这类密码算法，密码分析的复杂性的上限。在 NP 中有些特殊的问题被证明与此类问题中的任何问题一样困难，Steven Cook 证明了可满足性问题是 NP 完全问题。这意味着，如果可满足性问题在多项式时间内是可解决的，那么 P=NP。相反，如果 NP 中的任何问题能够被证明没有一个确定的多项式时间算法，此证明就将给出可满足性问题也没有一个确定的多项式时间算法。如果问题 P=NP 能得到证明，则使用现有的计算能力就可以使许多难题在理论上得到解决。

公开密钥算法并不能取代对称密钥算法，相反，公开密钥算法是对称密钥算法的补充。公开密钥算法首

先应用于对称密钥算法中秘密信道的建立(Diffie-Hellman),并且现在它仍然是公开密钥算法应用的一个重要方面。

3. 数据库安全关键技术

3.1 库文加密

为了防止黑客利用网络协议、操作系统等安全漏洞绕过分布式数据库的安全机制而直接访问库文件,有必要对用户口令表,访问授权表及其它重要库文进行加密。这样即使黑客得到了数据库文件,他们也难以知道明文。对数据库的加密应注意以下几点:

3.1.1 加、解密算法 加密系统应该是实际上不可破的;另一方面为维持数据库系统的性能,必须考虑到加、解密算法的速度。当然系统也可同时提供几种不同安全强度、速度的加/解密算法,这样用户可以根据数据对象的秘密程度来选择适当的算法。加、解密算法选择是数据库加密的核心,一个好的加密算法产生的密文应该频率平衡,随机无重码规律,周期很长而又不可能产生重复现象。窃密者很难通过对密文频率、重码等特征的分析获得成功。

3.1.2 加密的粒度 通常来说,加密的粒度可以是记录、域或者是数据元素。对加密粒度的选择必须根据数据库本身的特点,主要是用户对库的访问方式。比如说一个用户只允许访问数据库中少数记录,那么采用记录加密方式较好。数据元素是数据库加密的最小粒度。这种加密方式的优点是具有更好的灵活性和适应性,完全支持数据库的各种功能。但由于它的加密粒度小,因此加、解密效率较低。

随着数据库加密技术的不断发展,人们提出了许多对数据库加密的技术,如子密钥数据库加密、秘密同态技术等。子密钥数据库加密技术就是一种很有吸引力的方法,该方法按记录加密,而按数据项进行解密。需要哪个记录的某数据项时,就用该数据项的子密钥解密。其简化工作过程如图 3 所示。

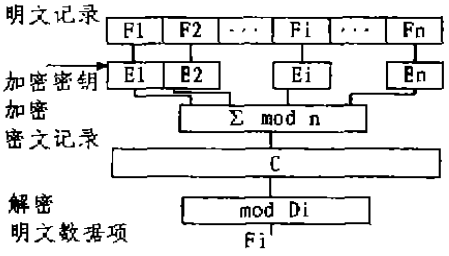

图 3 子密钥数据库加解密工作过程

①加密是把记录由明文变换为密文的过程,其方程为:

$$C=\sum_{i=1}^{n}E_iF_i(\mathrm{mod}\ D)$$

其中,$D=\prod_{i=1}^{n}D_i$;n 是数据库文件中数据项(字段)的个数;D_i 是选定的大于 F_i 的 n 个不同的素数,它是数据项 i 的解密密钥(读子密钥);F_i 为明文记录,E_i 满足:

$$E_i=[D/D_i]Y_i$$

其中:$Y_i[D/D_i]\equiv 1(\mathrm{mod}\ D_i)$

②解密。若需要解密第 i 个数据项,则对 C 用第 i 个数据项的读子密钥 D_i 求模运算即可实现解密,即:

$$F_i=C(\mathrm{mod}\ D_i)$$

3.1.3 密文索引 对数据库的加密不应过分降低对数据库的访问速度。必须对加密的数据对象巧妙地建立加密的数据索引,这样 DBMS 就可以利用密文索引完成快速检索。另一方面,在基于关系模型的数据库中,DBMS 要组织和完成关系运算,参加并、差、积、商、投影、选择和连接等操作。因此,对于参加关系运算的字段必须进行更巧妙的处理,不能明显地降低系统的效能。在不降低系统安全性的前提下,还可以考虑将参与关系运算的字段保持在明文状态。

3.2 密钥管理

密钥管理是加密系统的一个重要部分,同时它也是一个很难解决的问题,特别是数据库系统的密钥管理比网络系统的管理更复杂,并且具有它自己独特的特点,即数据库加密密钥的时间不变和用户不变性。密钥管理系统的设计与实现直接关系到整个数据库系统的安全与否。目前,多数数据库系统都采用多级密钥设置,其目的在于减少单个密钥的使用周期,增加系统的安全性。概念上密钥可分为两大类:数据加密密钥和密钥加密密钥,后者用于保护密钥。

密钥分配目前主要采用两种途径:集中式密钥分配方案和分布式密钥分配方案。后者是指网络中各主机具有相同的地位,它们之间的密钥分配取决于它们自己的协商,不受其他任何方面的限制,如 IBM 的 EMMT 网络密钥管理方案。集中式密钥分配方案则是在数据库中设立密钥管理中心,由密钥管理中心来集中管理系统中的密钥,负责产生密钥和对数据进行加密。在设计和实现上,集中式也相对地较易实现。

3.3 用户身份鉴别

数据库系统的安全性常常依赖于对终端用户身份的正确识别与检验。对数据库系统的访问必须根据访问者的身份施以一定的限制,以防止非法用户的欺诈行为。

目前一种得到广泛使用的鉴别方法是 Kerberos 模型。Kerberos 是为 TCP/IP 设计的可信第三方鉴别协议。它基于对称密码学,与网络上的每个实体分别共

享一个不可问的秘密密钥，因此 Kerberos 可以提供安全的网络鉴别，允许实体对网络上不同资源的访问。其鉴别步骤如图 4 所示。

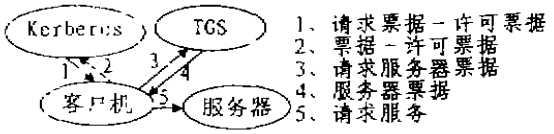

图 4 Kerberos 安全鉴别协议步骤

另一种得到广泛应用的鉴别方法是基于公开密钥密码体制的，就是使用证书和数字签名来实现对用户身份的验证，如 Netscape 公司的 SSL（Secure Socket Layer）协议中服务器端和客户端双方的验证。

3.4 Internet 上的数据传输

Internet 是一种开放性的互连网络，信息在网上的传输是以明文的方式进行的。在这种不安全的网络环境下实现信息安全传输的唯一方法就是采用功能强大的加密技术。分布式数据库系统的安全要求之一就是采用数据加密和解密技术在不安全的 Internet 上提供一条安全的数据通道。即：利用现有的加密、解密算法在数据源和目的地之间进行加、解密，也就是首尾加密。对进入网络的数据加密，然后待数据从网上传送出后，再进行解密。

考虑到对称密码加、解密速度快、安全性高的特点，一般都采用对称密码算法来完成对传输数据的加、解密，这样，问题的核心就集中在秘密密钥信道的建立上，非对称密码体制恰好可以做到这点，如 Diffie-Hellman 协议。通常我们把这种加、解密过程叫做混合密码体制。

由 Netscape 公司提出的 SSL 就是这样的一种协议，它可以在 Internet 上提供安全可靠的数据传输。协议由两部分组成，服务器端和客户端，服务器端和客户端使用 RSA 算法通过证书完成加密密钥的协商，然后再使用 RC2，RC4，IDEA，DES，tripe-DES，MD5 等算法完成对所传输数据的加密和数据摘要。

结束语 安全问题是当前网络研究的一大热点。就目前来说，要实现一个安全的分布式 Web 数据库仍是十分困难和复杂的。在设计和实现分布式 Web 数据库上，如何使设计和实现有效、快捷、安全要依赖于以下几个方面的发展。

1）密码技术（算法）的研究。解决数据库安全的根本方法是加密，而能否快速安全地实现解密对于提高数据库的访问效率尤为关键，因此，算法及其实现首先必须是安全的、可靠的；另一方面，算法实现还要有一定的速度，必须能够符合分布式数据库的应用特点。

2）密钥管理方法的研究。集中式的密钥管理中心相比较而言虽实现较为容易，但在实践中也有可能成为系统的瓶颈和安全隐患。

3）安全操作系统的研究。数据库系统是建立在操作系统的基础之上的。操作系统本身的安全级决定了在其基础上的数据库系统不会超过它的安全级别。

我们还必须注意到：安全是相对的，每次密码技术的进步都伴随着密码分析水平的进一步提高。因此，要真正实现数据库安全，必须不断地发展、实践和应用现代密码技术。

参考文献

1 陈爱民，于康友，管海明. 计算机的安全与保密 电子工业出版社

2 Schneier B. Applied Cryptography（Protocols，algorithms，and source code in C）Second Edition

3 Devida G I. A database Encryption System with Subkeys，ACM Transactions On Database Systems. Available at：http：//www.acm.ogr/pubs/journals/tods/1981-6-2

4 Kerberos The Network Authentication Protocol. Available at：http：//web.mit.edu/kerberos/www/

5 Introduction to SSL Available at：http：//developer.netscape.com/docs/manuals/security/

6 RSA Laboratories Frequently Asked Questions About Today's Cryptography

（上接第 36 页）

背景，为异构平台繁多和业务规则不断变化的企业计算环境提出了一种实现模式。系统具有如下优点：①利用软构件技术封装对象，使下层的数据逻辑与业务逻辑尽量独立于上层而局限于具体应用的应用逻辑与表示逻辑，保证了系统的高扩展性；②采用 CORBA 技术中的事件通道机制，在异构平台之间建立了一个松耦合的通道，保证了不同类型计算资源的有效集成；③利用 CORBA 技术中的对象事务处理服务，借助 ORB 来管理、传播事务语言环境，有效地解决了分布对象间事务处理的通讯问题。

参 考 文 献

1 Object Management Group. The Common Object Services Specification Revision 2. 2. July 1998

2 Object Management Group. CORBAServices：Common Object Services Specification，1998

3 Maffies S. Client/Server term definition. In：Encyclopaedia of Computer Science，D Hemmendinger，et al. eds Zurich，International Thomson Computer Publishing，1998

4 Notify Service，Joint Revised Submission，OMG Documents

5 Schmidt D C，Vinoski S. Object Interconnections —— The OMG Events Service（Column 9），SIGS C++ Report Feb. 1997

6 Nehmer J，Mattern F. Framework for the organization of cooperative services in distributed client-server systems. Computer Communication，1992，15（4）：261～269