网络数据库在传输过程中的安全研究

熊 江[1,2] 朱 婧[1]

(重庆三峡学院计算机科学系 重庆万州 404000)[1] (重庆大学自动化学院 重庆 400044)[2]

摘 要 本文提出了应用防火墙的包过滤技术与数据加密中的对称加密算法和非对称加密算法相结合,并运用链路加密和端对端加密技术的结合来实现网络数据库在传输过程中的安全。

关键词 网络数据库,数据安全,防火墙,数据加密,身份验证,数字签名

Study on the Security of Network Database in Transmitting

XIONG Jiang[1,2] ZHU Jing[1]

(Department of Computer Science, Chongqing Three Gorges College, Wanzhou 404000)

(Department of Automation, Chongqing University, Chongqing 400044)[2]

Abstract This paper puts forward that combining the packet filtering technology of firewall with symmetrical encryption algorithm and asymmetry encryption algorithm of data encryption, and by combination the chain to encrypt with the end-to-end encryption to achieve the security of network database in transmitting.

Keywords Network database, Database security, Firewall, Data encryption, Identification, Digital signature

1 引言

随着 Internet 的日益普及,以计算机和 Internet 为基础的各种信息管理系统为社会各种事务所应用,各种行业信息存储于数据库中,同时许多数据库服务器连接到 Internet 上,如:金税工程、网上售票、电子商务等领域。信息高速公路的巨大优势是有目共睹的,其优势在于网络上的信息共享,而这也是其脆弱性之所在,给社会带来了一些危害。在信息化时代的今天,信息的传输安全、存储安全、处理安全越来越重要,于是人们不断探索网络数据库(Network Database)安全。

2 现有的网络数据库的安全措施

一般网络数据库的安全性是设置在防火墙(Firewall)的基础之上的,首先利用防火墙监测网络,有可疑的 IP 数据包将报警并加以拒绝该数据包。使用代理服务器可以使内部网络不直接与外部网络连接,而对接受的数据包进行分析,了解了其连接请求,才与内部的网络发出请求。防火墙对于对付超大 ICMP 包、IP 伪装、碎片攻击、端口控制、保护数据库服务器的攻击等方面有较大的作用。

其次是采用身份验证(Identification)。数据库往往对于不同的用户设有相应的权限,用户在连接数据库时,应先进行数据库提出的验证过程实现身份认证,输入用户 ID 及密码,数据库将用户 ID 和密码与系统内的数据库进行对比验证,通过验证就授予相应的权限。若无法通过身份认证就只有最小权限。往往由于数据库存放用户 ID 和密码的文件被攻击,ID 和密码泄露,使非法用户可以拥有数据库使用权限,导致数据库中的数据不安全。所以,可以将数据库的所有存放用户 ID 和密码的文件进行加密,以防止用户 ID 和密码被窃取。而加密的密钥只能被网络数据库的系统管理员知道。

3 改进的网络数据库的安全措施

数据加密(Data Encryption)和数字签名(Digital Signature)是常被用于防止数据在传输过程中被篡改、窃取或假冒而采用的一种主动机制。我们考虑数据加密、数字签名和身份验证三种方式的结合:

这里有四对非对称密钥:用户的非对称密钥,PKU1、PKU2(公钥),SKU1、SKU2(私钥);数据库的两对非对称密钥,PKD1、PKD2(公钥),SKD1、SKD2(私钥)。PKU1,SKU1 用于传输会话密钥(对称密钥);PKD1,SKD1 用于验证用户 ID 和密码;PKD2,SKD2 和 PKU2,SKU2 用于做数字签名。

首先用户将用户 ID 和密码用 PKD1 加密传输,数据库接受后用 SKD1 解密,进行验证。授予权限时,将会话密钥用 PKU1 加密传输,用户用 SKU1 解密就得到会话密钥。这样交换了会话密钥后,再传输数据就使用会话密钥加密数据。为了使用户验证是否是数据库发出的数据,数据库在传输数据时用 SKD2 做摘要函数,做出数字签名,再将数字签名和数据包用会话密钥加密传输,用户得到整个数据包后,先用会话密钥解密,再用 PKD2 还原摘要即还原成明文。两个明文进行对比,相同就通过验证。因为只有该数据库有此密钥,所以用于数字签名后的摘要是其他人无法伪造的,从而确认是来自该数据库的数据。同理,用户用 SKU2 进行数字签名,使用户必须正确使用数据库的数据。通过验证的用户获得一定的权限,但是没有通过验证的用户可以得到根据数据库系统开发者开发时设置的最小权限。

数据库中的数据的安全保护措施是很重要的。但是,在传送过程中的数据也非常重要,因为数据最危险的时候不是在硬盘上,而是在传输过程中。如果数据库的数据在传送过程中能被监听、窃取或篡改,那么我们保护的数据库中的数据也是没有丝毫用处的,因为这些数据关键是它如何才能被正确使用。

4 防火墙与数据加密结合应用于数据库传输中加强安全的构想

4.1 IP 数据报格式

熊 江 副教授,长期从事计算机网络及硬件的教学与科研工作;朱 婧 本科生。

TCP/IP体系中的网络层常称为网际层(Internet Layer)。在网际层上使用的是IP层协议,而在IP层抽象的互联网上,我们看到的只是IP数据报。IP数据报是由应用层数据加上TCP层的首部变成的TCP报文再加上IP层首部(主要含有IP地址)组成的。虽然在IP数据报的首部有源站的IP地址,但路由器只根据目的站的IP地址进行选路。

首先了解一下IP数据报的格式,如图1。

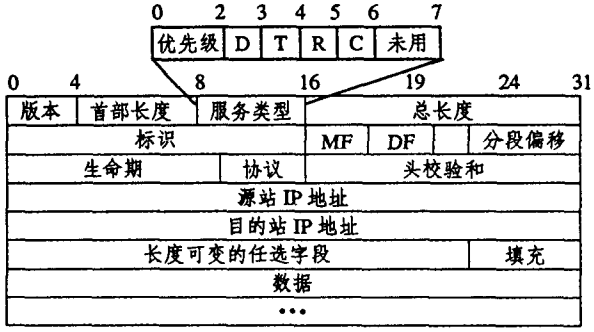

图1 IP数据报格式

一个IP数据报由首部和数据两部分组成。首部的前一部分长度是固定的20字节,后一部分的长度是可变长度。

4.2 提出网络数据库数据在传输过程中的安全措施的构想

图1中可以看见标志位里3bit中只有2bit有意义。第51位是一个空闲位,可以利用这个空闲位解决在网际层上加密数据,或选择怎样加密数据的问题。

一些厂商在应用层实现加密,另一些在IP层实现加密。前者使网络管理员可以选择想要加密的数据或文件,后者是强迫加密给定连接上的所有内容,另一个需要讨论的问题是要选择一种加密后不变长的对称密钥加密算法。不变长的原因:一是由于IP包结构图定,对数据段加密必须前后长度一致,才能保证加密后IP分组能正常地在Internet上传播;二是对称加密算法比非对称密钥加密算法可以有效地减轻网络负荷。

这里我们考虑将所有数据强制性加密传输,可以用于数据安全要求较高的数据库。

图2为一个网络实例,提出的构想建立在该网络实例上。

图2 网络实例

4个模块

模块1:防火墙I的网络管理员和防火墙II的网络管理员是把在一定时间段内确定的对称密钥及算法,通过非对称密钥加密技术(RSA)进行交换。在传输文件数据时就使用对称密钥。而这种是传输对称密钥及算法还是传输的用对称密钥加密的数据是用IP数据报首部的第51位表示,若传输的是对称密钥,则第51位置0,若是加密的数据则置1。

模块2:用户主程序。该程序功能:磁盘浏览功能,以确定需要传送的文件;选择加密技术的功能;发送功能;接受功能。

模块3:对称密钥加密技术。对IP包的数据段进行加密和解密。

模块4:对IP包进行过滤。

步骤:(1) 将防火墙I和防火墙II安装成代理型防火墙。

发送方:(2) 将对称密钥及算法,用先确定的非对称密钥进行加密,并将IP数据包第51位置0,送往防火墙I,传输。

(3) 防火墙I进行包过滤,符合过滤规则的通过,否则丢弃该包。

(4) 检测对方应答信号,若收到应答执行(5),若没有就执行(2)。

(5) 将要通信的数据用对称密钥加密,将IP数据包第51位置1,通过防火墙I传输。

(6) 数据发送完毕后,发送询问信号,是否全部收到,并等待应答。若没收到应答就执行(5),否则传输完毕。

接受方:(2) 通过防火墙II进行包过滤,符合过滤规则的接受该包,否则丢弃该包。

(3) 检测IP数据包第51位,若为0则执行(4),若为1则执行(5)。

(4) 接受对称密钥及其算法,发送应答信号,并将以前的对称密钥进行更新。

(5) 接受数据包,验证并发送应答信号。

(6) 用最近对称密钥及算法解密将接受的密文还原成明文。

这里采取的应答机制为防止网络的应答信号在通信线路上丢失,设置了计时器,在一定的时间内没有接受到对方的应答信号,主机将自动重新传输该数据包。

将通信双方的防火墙设置为代理型防火墙,又具有包过滤型防火墙的特征。这样的设计有以下的作用:内部网络与外部网络没有直接的连接,能进行全面的访问控制;能设置过滤规则确定内部网络和外部网络哪些用户的哪些数据能连接到目的主机上;能使用身份验证进行严格身份检查。

前面提到用IP数据包传输必须传输前后长度一致,所以必须使用对称密钥加密,但是使用对称密钥加密数据安全性并不高,容易被破译。所以在这里考虑利用非对称加密密钥能提高数据的安全性的特点,先将要用到的对称密钥用非对称密钥加密再进行传输,不定时地变换对称密钥及算法,提高网络安全性。尽管使用非对称密钥加密所消耗资源大,但不易于鉴定传输数据,所以使用对称加密密钥加密数据的安全性有了一定的保证,并且传输之前可以对加密后的密文的长度和加密前的密文的长度进行比较,长度一致就可以传输。

4.3 用半形式化语言描述

```
send ( )
    {防火墙I对外部网络发来的IP数据包按照预先设定的过滤规
    则进行包过滤;
    If(IP数据包不符合过滤规则)
    丢弃该IP包;
    Else
    {检查IP包第51位;
      If(IP包第51位= =0)
      {利用非对称密钥,启动对应的加密算法对该IP包的数据
    段,也即是对称密钥及算法进行加密;
    让加密后的IP包通过防火墙I进入Internet;
    转发到防火墙II;}
      Else
      {利用该时间段的对称密钥,启动对应的加密算法对IP数据
    包的数据段进行加密;
    让加密后的IP包通过防火墙I进入Internet;
    转发至防火墙II;}
    }
    }
receive ( )
    {当IP包到达防火墙II时,对IP包按预先设定的过滤规则进行
    包过滤;
    If(IP包不符合包过滤规则)
    丢弃该IP包;
    Else
```

```
{ 检查IP包的第51位;
  If(IP包的第51位==0)
    {利用非对称的解密密钥,启动解密算法对IP包得数据段进行解密;
    让解密后的IP包,即是对称密钥及算法通过防火墙II,转发到内部网络的目的站点,并对以前的对称密钥及算法进行更新;
    }
  Else
  {利用该时间段的对称密钥,启动相应的解密算法对IP包的数据段进行解密;
  让解密后的IP包通过防火墙II,转发到内部网络上的目的站点;}
  }
}
```

4.4 传输过程中的加密法

为了进一步提高数据安全性,可以在传输的过程中选择不同的加密方法,这里是对前面的不同内容采取不同的加密方法。先采用端到端加密法,在进入外部网络之前先加密,然后在外部网络中用链路加密法,即在每个节点上将已经在端点加密的密文再加密一次。这样可以防止中间节点的数据是明文。也避免了数据分组头是明文,易受到通信量分析的攻击。

当然这样设计的网络通信提高了数据安全性,但同时,网络资源代价也相当高,硬件设施和软件设施要求高。因此并不能让所有用户的网络通信都这样设置,只适用于对网络数据安全要求相当高的机构使用。也就是那些获得较高权限的用户能采用这种传输机制,没有获得权限的用户就可以直接传送数据。

结束语 本文介绍了防火墙技术和数据加密技术在网络数据库安全中的应用,以及相结合在数据传输过程中的应用的策略。这套方案不仅使内部网络的安全性得到保障,而且还保证了在开放网络中传送敏感数据的安全性,使网络安全得以增强并且在一个更大的范围得以实现。本文提出的防火墙技术和数据加密技术的结合构想,希望能使网络数据库安全性的提高起到抛砖引玉的作用。

参 考 文 献

1 Ziegler R L. Linux 防火墙[M]. 北京:人民邮电出版社,2000
2 罗春荣. 国外网络数据库:当前特点与发展趋势[J]. 中国图书馆学报,2003,3:44~47
3 陈小波. 网络信息安全与防火墙技术[J]. 现代情报,2003,7(7):48~49
4 肖军模,刘军,等. 网络信息安全[M]. 北京:机械工业出版社,2003
5 王睿,林青波. 网络安全与防火墙技术[M]. 北京:清华大学出版社,2000
6 Reed K D. 网络技术[M]. 北京:电子工业出版社,2002
7 袁家政. 计算机网络安全与应用技术[M]. 北京:清华大学出版社,2002
8 李广儒,邹云松. 防火墙技术的几个发展趋势[J]. 河南师范大学学报,2003(2):36~39
9 William T. 防火墙原理与实施[M]. 北京:电子工业出版社,2002
10 雷倩睿,李鹏文. 网络安全防御中数据加密技术的研究[J]. 信息技术,2003(1):6~9
11 孟艳红,秦维佳. 两种密码体制加密技术的研究对比[J]. 沈阳工业大学学报,2003(10):430~432
12 胡予濮. 对称密码学[M]. 北京:机械工业出版社,2002
13 Ford J L. 个人防火墙[M]. 北京:人民邮电出版社,2002
14 Andress M. 计算机安全原理[M]. 北京:机械工业出版社,2001

(上接第 90 页)

该测试集的聚集函数可以选择 SUM、MIN、MAX 进行处理。为了测试方便,所有的测试集只用一个聚集函数做 SUM 处理。图 3 和图 4 分别显示更新时间随着基数不同的变化曲线图。本文基数定义为每个维属性的不同的取值个数。Dwarf 数据集的特点是对基数的不同有明显的效果。

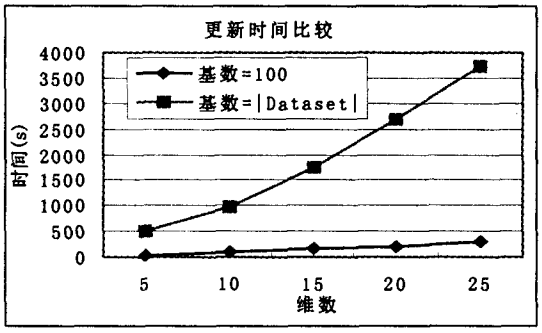

图 4 维数不同更新时间比较

上图分别是对维基数设为 100 的数据集和维基数等于元组数(Dataset)的数据集进行测试得到的更新时间比较图。图 3 是在相同的数据集(200000 条)大小下,更新的数据集分别采用 10000 条、50000 条、100000 条、150000 条、200000 条元组时的算法性能进行测试比较。图 4 是在相同的数据集(200000 条)大小下,在维数采用 5 维、10 维、15 维、20 维、25 维的情况下对 200000 条元组大小的数据集进行更新的性能测试比较。

从实验结果看出:不同的基数对更新性能有很大的影响。基数为 100 的数据集,其响应时间曲线倾斜度无论在数据集变化还是在维数变化的情况下都非常小,而基数等于元组数的数据集的响应时间曲线倾斜度随着维数的增加和元组数的增加急剧上升。从实验结果看出,前缀多的数据集更新性能很好,所以 Dwarf 更新算法在更新性能上对前缀多的数据集有明显的优势,也适用于数据仓库中的高维数据集。

结论 本文主要讨论了 Dwarf 数据立方的增量维护策略,研究并提出了一种新的基于 Dwarf 的增量更新算法。针对更新后聚簇特性被破坏的情况,利用聚簇整理,重新恢复 Dwarf 数据立方的聚簇特性。最后,通过实验给出了在维数不同和元组数不同时,对基数不同的数据集做测试比较的实验结果。实验结果进一步证明了该算法对前缀多的数据集更新性能很好,适用于数据仓库中的高维数据集。

参 考 文 献

1 Agarwal S, Agrawal R, Deshpande P M, et al. On the computation of multidimensional aggregates. In: Proc. of the 22nd VLDB Conf, 1996. 506~521
2 王珊,等. 数据仓库技术和联机分析处理[M]. 北京:科学出版社,1998
3 Li J Z, Rotem D. Aggregation algorithms for Very Large Compressed Data Warehouses. In: Proc. of VLDB Conf, 1999. 51~662.
4 Vitter J S, Wang M, Iyer B. Data cube approximation and histograms via wavelets. In: Proc. of Seventh Intl. Conf. on Information and Knowledge Management, 1998. 96~104
5 Sismanis Y, Roussopoulos N, Deligianannakis A. Dwarf: Shrinking the petacube. In: Proc. ACM SIGMOD, 2002. 464~475
6 Wang W, Lu H Jun, Feng J Lin, et al. Condensed cube: An effective Approach to Reducing Data Cube Size. ICDE, 2002
7 Lakshmanan L V S, Pei J, Han Jia Wei. Quotient cube: How to summarize the semantics of a data cube. In: Proc of VLDB Conf, 2002. 778~789
8 Lakshmanan L V S, Pei J, Zhao Y. QCTrees: An Efficient Summary Structure for Semantic OLAP. In: Proc. ACM SIGMOD 2003
9 Roussopoulos N. Cubetree: Organization of and Bulk Incremental Updates on the Data Cube. In: Proc. ACM SIGMOD, 1997. 88~99