

文章编号:1001-9081(2005)02-0297-04

## 利用 Sniffer 分析数据库应用系统的数据传输安全性

陈夕华, 李生红

(上海交通大学 信息安全工程学院, 上海 200030)

(hua\_cx@sjtu.edu.cn)

**摘 要:**分析了几种常见数据应用系统结构的特点,同时介绍了 Sniffer 的工作原理,提出一种运用 sniffer 技术来量化分析数据传输的安全性的方法。依据该方法建立测试环境,通过实验来说明方法的应用和分析实验数据对应用的指导作用。最后总结了该研究的应用,适用于量化评估数据应用系统数据传输的安全性。

**关键词:**数据库; Sniffer; 安全; Client/Server; Browser/Server; 数据传输

**中图分类号:** TP311.13 **文献标识码:** A

## Security analysis of data transporting in database application system with sniffer

CHEN Xi-hua, LI Sheng-hong

(Institute of Information Security, Shanghai Jiaotong University, Shanghai 200030, China)

**Abstract:** This paper first showed respective characteristics of typical database application systems and introduced the theory of sniffer technology. It put forward a new method of analyzing data transfer security numerically with sniffer technology. Then it told how to make an experiment with this method and showed results of the experiment by using this method. At last, the paper emphasized the advantage of using this method to analyze data transfer security of database application system.

**Key words:** database; Sniffer; security; Client/Server; Browser/Server; data transporting

### 0 引言

数据库应用系统是基于数据库的应用软件系统,其安全性是应用系统好坏的一个重要评价标准。如果应用系统的安全性得不到有效的保证,那么所建立的数据库应用系统与其说是资料数据的管理,不如说是对数据的破坏和泄露。随着 Internet 技术的迅速发展,数据库与 Web 的结合日趋紧密,数据传输的安全性日趋重要,这也为数据库应用系统的安全性带来了新的挑战<sup>[1]</sup>。

数据库应用系统的安全性主要包括三个方面:客户端、服务器端和数据传输部分<sup>[2]</sup>。客户端的安全性通过操作系统的授权访问和用户限制来保障,任一用户要进入系统,必须通过用户名、口令登录和身份验证核实后才允许进入,未经授权的用户不允许进入。服务器端的安全性则主要是数据库本身的安全性维护,数据库的安全性是指保护数据库以防不合法的使用所造成的数据泄密、更改或破坏。数据库系统安全机制的核心问题是:作为用计算机保存数据和信息记录的数据库管理系统,应当提供对信息的安全存取的服务。主要手段有:身份验证、授权机制和采用视图、触发器和存储过程维护数据库安全性,另外数据库本身设计的安全性也是加强该部分安全的保证;传输部分的安全性与具体的网络环境和应用系统的本身的设计结构有关,目前常用的数据应用系统的设计结构主要有 Client/Server、Client/Server + 存储过程、Client/Server 三层结构和 Browser/Server 结构。目前对于这些结构

的传输安全性分析只是一些定性的说明,主观因素比较多,缺乏客观性<sup>[3]</sup>。本文提出了一种运用 sniffer 技术抓包分析的方式来量化分析系统的安全性,为数据库应用系统传输安全性提供一种行之有效的量化分析方法,该方法适用于评估数据库应用系统数据传输的安全性。

### 1 数据库应用系统常用结构

#### 1.1 Client/Server 结构

Client/Server(C/S)结构是一种两层结构的分布式系统,数据存放在数据库服务器上,用户通过客户端对数据库直接进行操作。服务器端完成 DBMS 的核心功能,执行数据库的存储逻辑和事务逻辑。应用程序存放在客户端,完成数据处理、数据表示和用户接口等功能,执行应用逻辑并提供用户界面。优点:这种 C/S 结构可以充分发挥计算机网络的优点,最大限度地利用客户端和服务端上的资源,提高计算效率和降低网络传输量;对网络的带宽和延迟时间要求较低。缺点是:在这种模式下,客户方软件不是按照开放标准设计的,因此不同人开发的程序,其结构、功能、通信方式、数据结构差别较大,系统集成困难,维护工作量也大。

#### 1.2 Client/Server 结构 + 存储过程

存储过程是一组为了完成特定功能的 SQL 语句集,经编译后存储在数据库中。用户通过指定存储过程的名字并给出参数(如果该存储过程带有参数)来执行。存储过程不仅可以用来完成应用系统的逻辑处理,提高应用程序的运行性能,而

收稿日期:2004-07-30 基金项目:国家 863 计划项目(2003AA142160)

作者简介:陈夕华(1979-),男,江苏海安人,硕士研究生,主要研究方向:网络安全; 李生红(1971-),男,辽宁葫芦岛人,副教授,博士,主要研究方向:网络安全、计算机病毒、内容过滤。

且也可用于保证数据的安全性与完整性。为了禁止用户直接更改基表,可通过存储过程来更改基表,以达到保护基表的数据,然后授予用户具有执行该存储过程的权力,这就限制了用户对基表的不当操作,从而保证了数据的安全存储过程。存储过程的优点:允许标准组件式编程;能够实现较快的执行速度;能够减少网络流量;可被作为一种安全机制来充分利用。

Client/Server 结构 + 存储过程的方式是介于两层和三层结构之间,常被称为 2.5 层结构,可以达到三层结构的一些特征,又可以减少开发量。

### 1.3 Client/Server 三层结构

三层模式是传统 C/S 模式的扩展,可以将其看作两个 C/S 模式的结合。客户端向应用服务器发送请求,应用服务器响应请求并进行相应的处理,然后把处理结果返回客户端,这是第一层 C/S 模式;应用服务器运行业务处理程序时,如需访问数据库则向数据库服务器发出请求,数据库服务器把数据处理结果返回应用服务器,这是第二层 C/S 模式。因此三层模式具有传统 C/S 模式的全部优点,且用户接口、业务处理和数据管理的分布更加灵活。三层模式信息系统的数据库存放在数据库服务器中,大部分的业务处理程序在应用服务器中存放和运行,应用程序更新、升级时,只需更新应用服务器中的程序,这可使应用程序的维护对用户是透明的,降低维护成本,提高系统可维护性。由于客户端不存放数据,只要保证传到应用服务器的信息不包含关键程序代码,操作人员就不能看到和修改应用程序,只要各类服务器安全设计合理,就可以保证应用程序和数据不被破坏<sup>[4]</sup>。

三层模式的信息系统运行环境一般是开放式的网络环境,系统信息要在网络上传输,系统的功能全部基于网络实现。要保证信息不被窃取,不仅要保证服务器中数据和程序的安全,还应保证信息传输过程中的安全,这是本文涉及的实验中测试的主要方面。

### 1.4 Browser/Server 结构

Browser/Server(B/S)模式是一种三层或多层结构的分布式系统。在该模式下客户端通过浏览器向 Web 服务器提出请求,由 Web 服务器向数据库服务器提出查询请求,Web 服务器再将查询的数据以超文本文件的形式传给客户机。其主要特点是:客户端与服务器通过 TCP/IP 协议互联;客户端采用与应用无关的超文本信息查询工具——浏览器。因而在该模式下系统集成非常容易、维护工作量小、且又易掌握与升级。相对 C/S 模式,它有以下优点:简化客户端。只需在客户端安装浏览器软件即可,不用在不同的客户端上安装不同的客户应用程序;简化了系统的开发和维护。

在安全性要求高、交互性强、处理数据量大、数据查询灵活的地点固定的小范围内使用 C/S 模式;在安全性和交互性不高,地点灵活的广域范围内使用 B/S 模式。这样能充分发挥各自的长处,开发出安全可靠、灵活方便、效率高的软件系统。

## 2 Sniffer 技术

ISS(因特网安全系统公司)是这样定义 Sniffer 的:Sniffer 是利用计算机的网络接口截获目的地为其他计算机的数据报文的一种工具。Sniffer 工作原理:在以太网中,所有的通讯都是按广播方式进行。对于网卡来说,一般有如下 4 种接收模

式:广播方式,该模式下的网卡能够接收网络中的广播信息;组播方式,设置在该模式下的网卡能够接收组播数据;直接方式,在这种模式下,只有目的网卡才能接收该数据;混杂模式,在这种模式下的网卡能够接收一切通过它的数据,而不管该数据是否是传给它的。如果将网卡的工作模式设置为“混杂模式”,那么网卡将接受所有传递给它的数据包,这实际上就是 Sniffer 的基本原理:让网卡接收一切它能接收的数据<sup>[5]</sup>。

网络侦听技术是指捕获网络电缆上传输的所有网络报文的技术。一方面,网络侦听技术可以用在网络管理或网络测试中;另一方面,网络侦听在很大程度上危及了局域网的安全。这是因为目前的局域网基本上都采用以广播技术为基础的以太网,任何两个节点之间的通信数据包,不仅为这两个节点的网卡所接收,也同时为处在同一冲突域上的任何一个节点的网卡所截取。因此,黑客只要接入以太网上的任一节点进行侦听,就可以捕获发生在这个冲突域上的所有数据包,对其进行解包分析,从而窃取关键信息。

采用 Sniffer 技术抓取网络环境中的数据包,分析数据包的内容可以知道数据传输过程中的数据泄露情况。Sniffer 技术还可以提供传输过程中数据包的数目和大小,从而为量化分析提供了准确的原始数据。数据应用系统的传输是基于网络传输的,从而可以利用 Sniffer 抓取系统传输的数据包来量化分析系统的安全性。

## 3 实验环境

为了简化数据库应用系统测试的复杂性,并能够量化数据传输部分的安全性,在实验中将应用服务器和数据库服务器集成在一台测试机器上,各种结构的 Demo 均在相同的客户机系统和数据库服务器系统下进行测试。本文在测试中采用了常见的窃听方式,选取了客户端与数据库或应用服务器的数据传输为测试对象,图 1 为实际测试环境的网络结构图。

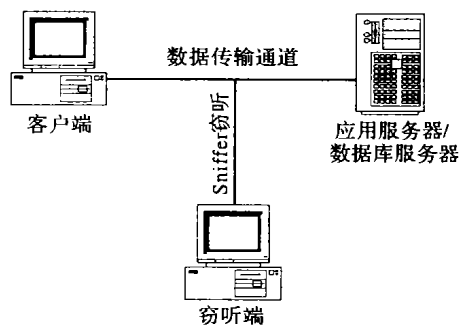


图 1 测试环境的网络结构

客户端安装测试 Demo 的客户端程序或通过浏览器来访问应用服务器,或直接访问数据库。窃听端安装 Sniffer 工具,本实验中采用的是 NAI(美国网络联盟)公司的 Sniffer 工具。测试使用的数据库应用系统的各种结构的样本都通过实验开发的 demo 版本,从而可以保证对系统的本身结构和数据库结构有清晰的了解,利于分析各种结构下数据传输部分的安全性。

## 4 实验数据

为了对传输过程的信息泄露有一个比较精确的量化值,考虑到加密算法的不同会带来数据的差异,在开发测试 demo 时没有进行加密处理,采用的是明文传输的方式。窃听端

Sniffer 抓取到的数据包是明文信息,如果某个数据包包含了数据库的敏感信息,则表明该数据包为信息泄露数据包。对采用 C/S 结构、C/S 结构 + 存储过程、C/S 三层结构、B/S 结构( Websnap 技术开发)、B/S 结构( Asp 技术开发) 5 种结构的开发设计的数据应用系统进行测试,数据泄露的测试结果见表 1。

表 1 5 种结构应用系统数据泄露情况

应用系统结构	测试结果描述	安全说明
C/S 结构	泄露信息:数据库类型、连接方式、客户端登录信息、数据库名、表名、字段名、操作方式、操作数据;表名-字段名-数据-操作能够对应	该结构下应用系统操作的数据库信息完全暴露给窃听者,窃听者只需要破解数据库登录的密码(密文信息),获取该密码后窃听者可以通过其他数据库连接方式绕过客户端直接操作数据库。
C/S 结构 + 存储过程	泄露信息:数据库类型、连接方式、客户端登录信息、数据库名、表名(部分)、存储过程名、字段名(少许)、操作方式(少)、操作数据;操作-字段-数据-表名(存储过程)对应困难	该结构下窃听者需要破解数据库登录的密码(密文信息),获取该密码后窃听者也可以通过其他数据库连接方式绕过客户端直接操作数据库。但该结构的好处是数据库的结构没有完全暴露,窃听者只能窃听到少数表的信息。
C/S 三层结构	泄露信息:连接端口、连接 GUD、客户端登录信息、数据提供组件名、操作数据	最大的缺陷用户登录信息的暴露,被窃听到有效者有用的信息有限。传输数据量少。
B/S( Asp 技术)	泄露信息:连接端口、客户端提交方式、客户端登录信息、操作数据。该方式的数据流多,提取有用信息的工作量大	同 C/S 三层结构,与三层结构相比:优点是传送的数据包多,增加了分析的难度;缺点是传输数据量剧增
B/S( Websnap 技术)	同 B/S( Asp 技术)	同 B/S( Asp )结构

传输数据包数据统计分析,为了保证测试数据的可靠性和有效性,抓包测试过程以登录 + 添加一条记录 + 修改一条记录 + 退出四步操作为准,登录指一次成功(用户名和密码一次录入登录成功),添加和修改只进行一次成功操作(没有出错提示情况)。测试结果见表 2(表中记录的条数和 bytes 数以 10 次操作取平均值,不保留小数,比率取百分比,保留两位小数)。

表 2 5 种结构应用系统数据包数据统计

	C/S 结构	C/S 结构 + 存储过程	C/S 三层结构	B/S( Asp 技术)	B/S( Websnap 技术)
传输数据包数目	34	37	33	211	279
泄露信息数据包	10	10	8	11	13
泄露数据包/总数据包	29.41%	27.03%	24.24%	5.21%	4.66%
传输的数据包 byte 数	5218	5450	4591	97796	146672

注:1) B/S 结构中包含了一些图片信息使传输的数据包剧增(这也符合正常 B/S 结构使用,设计实验测试 Demo 时,考虑到了 B/S 结构界面友好和操作方便的要求,添加了一些系统介绍信息);

2) B/S 泄露信息的数据包仅对涉及数据库信息的记录进行统计,不包含各种网页文件信息的泄露。

5 实验结果分析

5.1 传输数据量分析

C/S 结构中:就实验测试的传输数据而言(单客户端 + 单应用服务器 + 单数据库(应用服务器和数据库安装在同一机器上)):C/S + 存储过程最多,C/S 结构次之,C/S 三层结构最少;与 C/S 结构相比,B/S 结构中数据传输量剧增,但泄露信息的数据包数量与 C/S 结构相差不多,Asp 技术和 Websnap 技术在本次测试中相当,Websnap 技术的传输数据量的偏大是因为网页中存在内嵌图片的原因。

5.2 安全性分析

C/S 结构中:C/S 三层结构(客户端 + 应用服务器 + 数据库),用户窃取到的只是客户端和应用服务器之间的数据,故在 C/S 的 3 种模式中传输的被窃听到的数据量最少,只有用户登录信息和操作数据泄露,安全性也最好;其次是 C/S 结构 + 存储过程安全性,但安全性还是很差,泄露了数据库连接

信息对应用系统的安全性会带来的是致命的问题;安全性最差的是简单的 C/S 结构。就 B/S 结构而言,跟 C/S 三层结构相似,也泄露了用户登录信息和数据信息,但由于 Web 方式的特殊性,传输的数据包比较多,从而增加了窃取分析的难度,我们可以从实验测试中泄露数据包/总数据包的比率的数值变化看出,该比率也在一定程度上反映了信息的泄露情况,可称之为信息泄露率,实验测试的 5 种结构的泄露数据包/总数据包(信息泄露率)参见图 2。

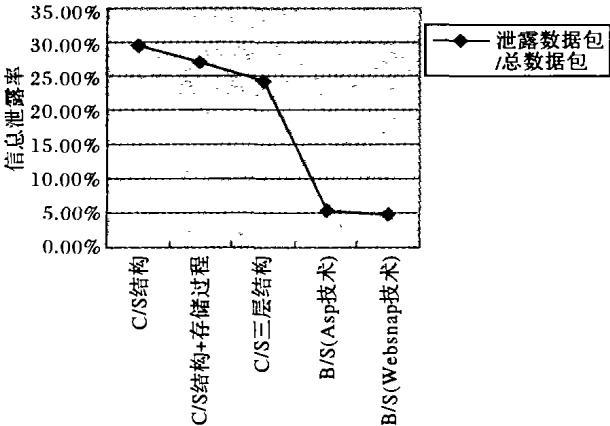


图 2 泄露数据包/总数据包

由图 2 分析可见,信息泄露率符合我们对各种结构安全

性的分析,信息泄露率越低,该结构的应用系统就越安全。C/S 结构中就信息泄露率而言,C/S 三层结构 < C/S + 存储过程 < C/S 结构,且 C/S 与 B/S 结构的信息泄露率成倍数递减关系( $C/S > 20\%$ ,  $B/S < 5\%$ ),从这个角度而言,B/S 结构的安全性优于 C/S 的三层结构;但考虑到 B/S 结构的特点,客户端采用与应用无关的超文本信息查询工具——浏览器,使得任何人访问应用服务器更加简单,从而使得应用服务器受攻击的威胁增大,C/S 三层结构需要有对应得客户端才能连上应用服务器,在这方面进行比较,C/S 三层结构比 B/S 有更好的安全性。

综上所述可见:C/S 和 C/S + 存储过程的结构适用于对安全性不作要求或是对安全要求比较低的应用环境中;B/S 结构适用于安全级别要求中,网络安全性好的应用环境,在互联网环境下使用时,必须要对机密信息进行加密并且应用服务器的安全防护级别要求也需提升;C/S 三层结构适合安全性要求高的应用环境,再辅以加密措施,是一些机密数据库应用系统的最佳选择结构。

从实验中,我们还可以看出,数据库表名和数据字段名被窃听到的概率很高,而这个往往在加密中被忽视,或是考虑到响应的时延要求,没有对这些信息进行加密。解决的思路除了进行加密外,还可以制定一套内部的开发命名代号机制,不采用可理解的命名方式,而是采用代码或代号的进行命名。比如数据库中要建立访问用户表,通常方式是建立数据表:LoginUser { UserID, UserName, Password... },当窃听器抓取

到这类包的时候,也就知道该表保存了系统的用户信息;但如果采用这样的方式进行建表:Tb\_010 { Fd\_001, Fd\_002, Fd\_003 ... },即使被窃听者抓到该数据表的数据,也很难猜测到这是系统的登录用户表。通过这种方式可以降低泄露信息的风险,减少了数据传输的加解密过程,提高了系统的性能。

## 6 结语

本文对数据应用系统的安全性进行研究分析,提出了一种利用 Sniffer 技术量化分析数据传输安全性的方法,通过建立实验环境进行测试来验证方法的可行性,并利用实验的测试结果对目前常用的数据应用系统的传输安全性进行定量评估,客观地分析了系统的安全性。

### 参考文献:

- [1] LOTHIAN P, WENHAM P. Database Security in a Web Environment[J]. Information Security Technical Report, 2001, 6(2): 12 - 20.
- [2] 鲜波,张继棠,陈新安. 数据库应用系统的安全性探讨[J]. 重庆邮电学院学报, 2000, 12(1): 47 - 50.
- [3] WISEMAN S. Database Security: Retrospective and Way Forward [J]. Information Security Technical Report, 2001, 6(2): 30 - 43.
- [4] 罗朝晖,边小凡,刘铁英,等. 三层模式下信息系统的安全[J]. 计算机系统应用, 2000, (8): 17 - 20.
- [5] 张健,李焕洲. 网络嗅探原理及其检测和预防[J]. 四川师范大学学报(自然科学版), 2003, 26(1): 90 - 92.

(上接第 288 页)

性没有任何限制,既可用于纯分类属性或纯数值属性的数据集,也可用于具有混合属性的数据集,且具有更好的时间效率和聚类质量。

## 4 讨论

由于用到的只是引力大小的相对比较,而不是引力的绝对大小,因此省去了对结果没有影响的万有引力定律中的引力常数。前面所述引力可以看成一种特殊相似度,现有文献中的相似度仅是距离的函数,从本文可以看出,将相似度推广成距离以及类大小等因素的函数,可以更准确地度量相似性。在研究过程中我们从几个不同的角度进行了探索。

1) 最初以  $C_n$  作为类  $C$  的质量,但发现当有部分区域非常密集时,会出现类似黑洞的现象,大部分对象会被吸入,改以  $\ln(\ln(C_n + 1))$  作为类  $C$  的质量则达到了比较好的效果。

2) 定义 4 中引力可推广到更一般的情况: $g(C_1, C_2) = \frac{\ln(\ln(C_1 \cdot n + 1)) \cdot \ln(\ln(C_2 \cdot n + 1))}{d(C_1, C_2)^2} (z > 0)$ 。

实验结果表明,  $1 \leq z \leq 4$  而  $r$  在合适的范围内时,检测结果基本稳定。

3) 将  $dissim(C_1, C_2)$  作为  $C_1$  与  $C_2$  间差异程度的度量:

$$dissim(C_1, C_2) = \frac{d(C_1, C_2)}{\sqrt{\ln(\ln(C_1 \cdot n + 1)) \cdot \ln(\ln(C_2 \cdot n + 1))}}$$

用  $dissim(C_1, C_2)$  与用  $g(C_1, C_2)$  度量类  $C_1$  与  $C_2$  间的相似性应该是等价的,但当有  $d(C_1, C_2)$  接近于 0 时,采用  $g(C_1, C_2)$  计算的阈值不稳定,实验结果表明,采用  $dissim(C_1, C_2)$

计算阈值更稳健。

### 参考文献:

- [1] GUHA S, RASTOGI R, SHIM K. ROCK: A robust clustering algorithm for categorical attributes[A]. In proceedings of the 15th ICDE [C], 1999. 512 - 521.
- [2] GANTI V, GEHRKE J, RAMAKRISHNAN R. Cactus —— clustering categorical data using summaries[A]. In Proc 1999 Int Conf Knowledge Discovery and Data Mining [C], 1999. 73 - 83.
- [3] HE ZY, XU XF, DENG SC. Squeezer: an efficient algorithm for clustering categorical data[J]. Journal of Computer Science and Technology, 2002, 17(5): 611 - 624.
- [4] GUHA S, MEYERSON A, MISHRA N, et al. Clustering data streams: Theory and practice[J]. Knowledge and Data Engineering, IEEE Transactions on, 2003, 15(3): 515 - 528.
- [5] PORTNOY L, ESKIN L, STOLFO S. Intrusion Detection with Unlabeled Data using Clustering[A]. In Proceedings of ACM CSS Workshop on Data Mining Applied to Security (DMSA-2001) [C], Philadelphia, PA, 2001.
- [6] ESKIN E, ARNOLD A, PRERAU M, et al. A geometric framework for unsupervised anomaly detection: Detecting intrusions in unlabeled data[Z]. In Data Mining for Security Applications, 2002.
- [7] SHENG YJ, YU MX. An Efficient Clustering Algorithm [A]. In Proc of 2004 International Conference on Machine Learning and Cybernetics [C], 2004. 8.
- [8] MERZ CJ, MERPHY P. UCI repository of machine learning databases[EB/OL]. <http://www.ics.uci.edu/mllearn/MLRRepository.html>, 2000.