数据库敏感数据加密算法的研究与改进

刘爱华,陈 钧,解 芳

(南京工程学院 计算机工程学院,江苏 南京 210067)

[摘要] 根据数据库中敏感数据加密的相关要求,研究了常用的数据加密技术,提出了一种能对数据库中敏感数据进行加密解密,并且不产生冗余数据的加密改进算法.

[关键词] 敏感数据,数据加密,冗余数据

[中图分类号] TP311 [文献标志码] A [文章编号] 1672-1292(2012)03-0068-03

Research and Improvement of Sensitive Data Encryption Algorithm in Database

Liu Aihua, Chen Jun, Xie Fang

(Department of Computer Engineering, Nanjing Institute of Technology, Nanjing 210067, China)

Abstract: According to the relevant requirements of encrypting sensitive data in database, this paper first studies the frequently-used data encryption technology, then puts forward an improved encryption algorithm. This algorithm can encrypt and decrypt sensitive data in database, and do not produce redundant data.

Key words: sensitive data, data encryption, redundant data

随着计算机技术和通信技术的不断发展,网络已成为社会发展的重要保证. 网络信息涉及国家政治、经济、军事、文化等重要领域,在网络中存储和传输的信息多有涉及政府文件、商业经济信息、军事机密等敏感内容,难免出现信息泄漏、信息窃取、数据篡改等问题,威胁到信息的安全.

一般认为那些不应公开的数据为敏感数据. 例如,职工数据库中姓名、性别、年龄、职位可能有最低的敏感性,邮箱密码、工资则最敏感,而电话、住址的敏感性介于两者之间. 数据加密技术被誉为信息安全的核心技术,密码技术在保证通信保密、信息鉴别和数字签名等方面具有强大的功能. 选择数据库加密技术时需要充分考虑数据库特点,如加密的强度要保证长时间不被攻击者破译; 加解密不能影响合法用户对数据库的正常操作; 加解密所需要的时间要足够短,不能让用户感到明显的延迟; 具备灵活的密钥管理机制,定期更换密钥; 加密以后的数据存储量没有明显的变化等[1].

通过研究几种常用的加密方法,本文对经典的数据库加密方法进行了改进,提出了一种更加适合于数据库中敏感数据的加密算法.

1 常用的加密算法

目前已有很多加密技术应用于信息安全领域,主要有:

1.1 基于“消息摘要”的算法

“消息摘要”(Message Digest)是一种能产生特殊输出格式的算法,该加密算法的特点是: 无论什么长度的原始数据,经过计算后输出的密文都是固定长度的. 算法的原理是根据一定的运算规则对原数据进行某种形式的提取,这种提取就是“摘要”. 消息摘要算法是“不可逆”的,理论上无法通过反向运算取得原数据内容,因此它通常只能被用来做数据完整性验证,而不能作为原数据内容的加密方案使用. 消息摘要

收稿日期: 2012-03-28.
基金项目: 南京工程学院校级科研基金(QKJC2010005).
通讯联系人: 刘爱华,讲师,研究方向: 网络数据库. E-mail: liuah@njit.edu.cn

算法最经典的 MD5 算法,是最强健的加密算法之一,常用于文件完整性保护和用户的密码保护[2].

1.2 对称/非对称密钥加密算法

由于"摘要"算法加密的数据仅仅能作为一种身份验证的凭据使用,若要对数据库中的数据进行查询和更新等操作,则需要对数据进行解密,就不能采用这种"不可逆"的算法.

1.2.1 对称加密

数据可逆的加密算法需要通过一个"密钥"进行数据加密处理,接收方通过一个"密钥"进行解密. 若双方持有的"密钥"相同(对称),就是"对称密钥"的概念. 也称这种加密算法为秘密密钥算法或单密钥算法。加密密钥和解密密钥可以是相同的,或者可以相互推算出来. 基于"对称密钥"的加密算法主要有 DES、TripleDES、RC2、RC4、RC5 和 Blowfish 等. DES 算法被各个领域广泛采用,包括 ATM 柜员机、POS 系统、收费站等. DES 以其高强度的保密性能为大众服务,其缺陷是密钥长度短,密钥仅有 56 位二进制,在现有的技术条件下用穷举搜索法进行攻击,来获取正确密钥已趋于可行[3].

1.2.2 非对称加密

数据加密和解密的密钥是不同的,即算法产生两个密钥,一个是公钥,对其他用户都公开; 另一个是私钥,仅为自己所有. 经用户公钥加密的信息只能通过私钥来解密. 基于"非对称密钥"的加密算法主要有 RSA、Diffie-Hellman. RSA 密钥的加密长度使得加密速度变慢,且运算复杂[4].

根据以上分析,经典的 MD5 算法虽然具有速度快、加密程度强的优点,但是数据库中的敏感数据需要涉及查询、更新等操作,需要能"可逆",也即能解密; 对称和非对称加密都需要保存密钥,数据库中又增加了需要进一步保密的数据,即密钥,增加了数据的冗余. 本文利用 MD5 算法的速度快、加密强度高的优势,对算法进行了改进,实现了对数据库中敏感数据既可以解密(即可逆),又不需保存密钥(即无冗余).

2 改进的加密算法

改进后的算法的核心思想是利用 MD5 算法和异或运算. 以字符串数据为例进行加密,将原文中的每个字符的低半个字节清零,得到由各高半个字节构成的字符串. 将处理过的字符串进行 MD5 运算得到 16 个字节的初始密钥,再将初始密钥的高半个字节清零,即得到密钥. 将密钥与原文逐个字节异或就得到密文,此时密文的高半个字节构成的字符串和原文的高半个字节字符串相同.

密文的解密过程和加密过程一样,将密文的每个字符的低半个字节清零,得到由各高半个字节构成的字符串. 该字符串和原文的高半个字节构成的字符串相同,所以对该字符串进行 MD5 加密得到的 16 个字节的初始密钥也是一样的,初始密钥高半个字节清零后得到的密钥也相同. 将密钥和密文逐个字节异或就得到了原文.

2.1 对字符串数据加密解密

字符串 s 数据的加密过程用 Delphi 伪代码方式给出,结果为 result 字符串:

```
begin
  for i: =1 to length( s) do
    sKey: = sKey + Chr( Ord( s [i]) and $ F0) ; //取每个字符的高 4 位,作为密钥种子
    D: = StringToMD5( sKey) ; //获得 16 字节的 MD5 数作为初始密钥
    i: =1; //i 作为字符串指针
    j: =0; //j 作为密钥指针
    while i < =length( s) do//开始循环
    begin
      k: = D [j] and $ 0F; //
      if Ord( s [i]) > $ 1F then//如果不是控制字符
        sKey: = Chr( Ord( s [i]) x or k) //加密
      else//对控制字符不加密
        sKey: = s [i];
      end
```

```
        Inc(i); //移动指针
        Inc(j);
        if j>15 then j:=0;
        result:=sKey+result; //反向构成密文,增加加密难度
    end;
end
```

字符串的解密过程和加密过程类似,可以合并为一个过程,通过参数 mode 加以区别是加密还是解密. 解密时应先将密文反向恢复为正常顺序.

2.2 对整形数据加密解密

整形数据 i 的加密解密和字符串有一些差别. 字符串数据不存在符号问题,但会产生控制字符,所以初始密钥是用原文的高半个字节低半个字节清零的字符串进行 MD5 运算. 而整形数据不必考虑该问题,初始密钥也可用原数据的各个低半个字节高半个字节清零构成的数据进行 MD5 运算,也用 Delphi 伪代码方式给出:

```
begin
  k:=i and $0F0F0F0F; //每个字节的低4位
  s:=IntToStr(k); //转换为8个16进制字符
  D:=StringToMD5(s); //求得MD5数组
  s:='$';
  for k:=0 to 3 do//取前4个数据的高半字节为密钥
    s:=s+IntToHex(D[k],2);
    fKey:=StrToInt(s) and $F0F0F0F0;
    result:=i x or fKey; //加密
  end
end;
```

3 结语

采用改进后的加密算法对数据库中的密码、工资等敏感数据加密后,存储的数据是经过 MD5 算法加密且再处理后的密文. MD5 算法非常经典,便于移植和使用,加密速度快、安全性高,即使泄露,破解者看到密文也无任何意义,所以达到了数据库数据安全性要求. 密文处理后又可以得到密钥,不需要额外的空间存储密钥,满足了数据库数据的存储性要求. 改进后的算法克服了 MD5 算法不可逆的弱点,能快速解密数据,不影响合法用户的数据库查询、更新等操作,满足了数据库的操作性要求.

[参考文献](References)

[1] 朱鲁华,陈容良. 数据库加密系统的设计与实现[J]. 计算机工程,2002,28(8):61-63.
Zhu Luhua,Chen Rongliang. Design and implementation of database encryption system[J]. Computer Engineering,2002,28(8):61-63.

[2] 王津涛,覃尚毅,王冬梅. 基于MD5的迭代冗余加密算法[J]. 计算机工程与设计,2007,28(1):41-111.
Wang Jintao,Qin Shangyi,Wang Dongmei. Iterative redundant encryption based on MD5 algorithm[J]. Computer Engineering and Design,2007,28(1):41-111.

[3] 谢志强,高鹏飞,杨静. 基于前缀码的DES算法改进研究[J]. 计算机工程与应用,2009,45(9):92-119.
Xie Zhiqiang,Gao Pengfei,Yang Jing. Research on improved algorithm of DES based on prefix codes[J]. Computer Engineering and Applications,2009,45(9):92-119.

[4] 蒋波. 一种基于三重DES和RSA的综合加密方案[J]. 微计算机信息,2007(3):52-53.
Jiang Bo. A encoding solutions based on triple DES and RSA[J]. Micro Computer Information,2007(3):52-53.

[责任编辑:严海琳]