

基于LECC和MD5算法的网络身份认证

王小冬¹, 周海燕²

(1. 南京工程学院计算机工程学院, 江苏 南京 211167)

(2. 南京师范大学数学科学学院, 江苏 南京 210023)

[摘要] MD5算法常用于对远程用户进行身份认证,但其抵挡不了查字典或差分攻击. 本文对用户身份进行MD5加密后,再采用LECC对密文进行纠错编码,提高算法抵御防攻击的能力. 理论和实践证明,本算法具有较高的安全性和实用价值.

[关键词] MD5算法, 线性纠错码, 远程身份认证

[中图分类号] TP301.6 **[文献标志码]** A **[文章编号]** 1672-1292(2015)04-0053-05

Authentication in Networking Environment Based on LECC and MD5 Algorithm

Wang Xiaodong¹, Zhou Haiyan²

(1. School of Computer Engineering, Nanjing Institute of Technology, Nanjing 211167, China)

(2. School of Mathematical Sciences, Nanjing Normal University, Nanjing 210023, China)

Abstract: MD5 algorithm is commonly used in authentication for the remote user, but it cannot resist the dictionary or differential attack. This paper improves an ability of the algorithm to resist against the attack by MD5 encryption for user identity and LECC ciphertext with error correction coding. Theory and practice have proved that this method is of high security and practical value.

Key words: MD5 algorithm, linear error correcting code, distanced-identity authentication

随着互联网技术的飞速发展,电子商务、网上银行、网上商城等网络平台的出现,人们的生活方式发生了天翻地覆的变化. 电子商务的年交易额达到10万亿,仅双十一一天淘宝网的销售额就达250.18亿元. 在频繁的网络交易活动中,存在着个人信息泄密等安全问题,进一步影响到人们的财产安全. 在远程身份认证过程中,传统的认证方式为“用户名+口令密码”,这种以明文方式在网上传输的认证方式很不安全. 目前广泛应用MD5加密算法来保护用户口令^[1],然而王小云在2004年国际密码会议上宣告MD5理论上可以破译^[2]. 虽然有很多学者致力于改进MD5算法^[3],但更多的学者研究了实践中破解MD5的算法^[4-6],在网络认证中利用MD5或改进的MD5保护口令根本抵御不了查字典攻击^[6].

1 MD5算法特点

MD5算法是将任意长度的明文经过复杂的变换转换成128位的散列值,该散列值即数字指纹. MD5算法是非常优秀的加密算法,具有下列特点^[1-3]:

(1) 容易计算. MD5算法为线性算法,加密速度相当快.

(2) 单向性. 常用的加密算法是双向加密,即明文经过算法加密为密文,密文也可通过算法解密为明文. 单向加密是相对双向加密而言的,也即被加密的密文不能解密为明文. MD5算法加密后的密文,不能通过有效的方法解密为明文. 因此,MD5算法加密的密文即使被泄露,也不会泄露明文的真正信息,故而

收稿日期: 2015-07-02.

基金项目: 江苏省教育厅自然科学研究面上项目(13KJB110016).

通讯联系人: 王小冬, 实验师, 研究方向: 数据挖掘与模式识别. E-mail: wangxd@njit.edu.cn

具有较高的安全性.

假定集合 A 为明文的集合, 集合 B 为密文的集合, MD5 即为一个从 A 到 B 的映射, 记为 $f: A \rightarrow B$. MD5 在一致性验证等应用中明文非常长, 因此, 集合 A 的元素个数非常大, 很难通过查字典的办法来获取明文. 但当明文为密码时, 用户的密码通常为字母加数字且不长, 此时集合 A 的元素个数很少, 这样即可把所有的明文和对应的密文放入数据库, 搜索数据库就能很快地查找密文的明文. 现存许多在线 MD5 解密网站, 如: www.cmd5.com, 对常规的 6 位密码用时不超过 20 s.

在集合 A 的元素个数有限时如何提高 MD5 算法的安全性, 在如今电子商务时代是一个非常迫切且重要的问题. 因 MD5 是单射, 故在数据库中保存的数据记录的条数就是集合 A 的元素个数, 如表 1 所示. 若将加密算法改为多射, 就会增加数据记录的条数. 如表 2 所示, 每个明文映射到 10 个密文, 则数据库记录即为单射记录的 10 倍.

本文采用线性纠错码 (Linear Error-Correcting Code, LECC), 使 MD5 加密算法变为多射, 并检验明文的一致性.

2 LECC 介绍

纠错码起源于通信技术和计算机技术的发展, 纠错码理论已成为信息安全的理论基础^[7-10].

定义 1 设 $a = (a_1, \dots, a_n)$ 和 $b = (b_1, \dots, b_n)$ 是 F_q^n 中的向量, 则向量 a 的 Hamming 权定义为非零向量 a_i 的个数, 而向量 a 和 b 之间的 Hamming 距离是指其相异位的个数^[10].

定义 2 向量空间 F_q^n 的一个 F_q 上的线性子空间 C 叫作 q 元线性码^[10]. C 有 3 个基本参数:

- (1) 码长 n ;
- (2) 码字个数 $K = |C|$ (或用信息位数 $k = \log_q K$), $0 \leq k \leq n$;
- (3) 最小距离 $d = d(C)$ (不同码字之间 Hamming 距离的最小值).

则该纠错码可表示为 $(n, K, d)_q$ 或 $[n, k, d]_q$.

对于线性码利用线性代数工具, 取定线性子空间 C 的一组 F_q -基 $\{v_1, \dots, v_k\}$, 其中 $v_i = (a_{i1}, a_{i2}, \dots, a_{in})$, $a_{ij} \in F_q$ ($1 \leq j \leq n$, $1 \leq i \leq k$). 记

$$G = \begin{pmatrix} v_1 \\ \vdots \\ v_k \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kn} \end{pmatrix},$$

则每个码字 c 可以唯一的表示为:

$$c = b_1 v_1 + \cdots + b_k v_k = (b_1, \dots, b_k) G.$$

则 G 称为线性码 C 的一个生成矩阵.

k 维向量空间 F_q^k 中向量 (b_1, \dots, b_k) 共有 q^k 个, 为了纠错需将其一一映射到 n 维线性子空间 C 中的码字 $c = (b_1, \dots, b_k) G$, 即

$$\phi: F_q^k \rightarrow C \subseteq F_q^n, (b_1, \dots, b_k) \mapsto (b_1, \dots, b_k) G.$$

对于 F_q^n 的 k 维线性子空间 C , 必存在齐次线性方程组:

表 1 MD5 数据表

Table 1 Data table of MD5

id	明文	密文
1	a1	b1
2	a2	b2
3	a3	b3
...
10000	a10000	b10000

表 2 多射数据表

Table 2 Data table of one to many

id	明文	密文
1	a1	b1, 1
2	a1	b1, 2
...
10	a1	b1, 10
11	a2	b2, 1
...
20	a2	b2, 10
...
99991	a10000	b10000, 1
...
100000	a10000	b10000, 10

$$\begin{cases} h_{11}x_1 + h_{12}x_2 + \cdots + h_{1n}x_n = 0, \\ h_{21}x_1 + h_{22}x_2 + \cdots + h_{2n}x_n = 0, \\ \vdots \\ h_{n-k,1}x_1 + h_{n-k,2}x_2 + \cdots + h_{n-k,n}x_n = 0, \end{cases}$$

使得线性空间 C 就是其解空间.

记矩阵 $H=(h_{ij})_{1 \leq i \leq n-k, 1 \leq j \leq n}$, 则对每个 $v \in F_q^n$, 由线性方程组解的定义知

$$v \in C \Leftrightarrow vH^T = 0,$$

从而可用矩阵 H 来检查向量 v 是否为 C 中的码字. 因此, 矩阵 H 称为线性码 C 的一个校验矩阵^[9-10].

定理 1 设向量空间 F_q^n 的一个 F_q 上的线性子空间 C 是参数 $[n, k]$ 的 q 元线性码, $H=(u_1, u_2, \cdots, u_n)$ 是 C 的一个校验矩阵. 若 u_1, u_2, \cdots, u_n 当中任意 $d-1$ 个均 F_q -线性无关, 且存在 d 个列向量是 F_q -线性相关的, 则 C 的最小距离为 d ^[10].

定理 2 设 C 为参数 $[n, k]$ 的 q 元线性码, 若非零合法码字的最小 Hamming 权为 d , 则 C 的最小距离也为 d ^[10].

定理 3(线性码的纠错译码算法) 设向量空间 F_q^n 的一个 F_q 上的线性子空间 C 为参数 $[n, k, d]$ 的 q 元线性码, $l = \left\lfloor \frac{d-1}{2} \right\rfloor$, 且 C 有校验矩阵 $H=(u_1, u_2, \cdots, u_n)$, 其中, $u_i (1 \leq i \leq n)$ 都是 F_q 上的 $n-k$ 维列向量. 若码字 $c \in C$ 在传输过程中错位个数 $\leq l$, 即收到的向量 $y = c + \varepsilon$ ($w(\varepsilon) \leq l$, 其中 $w(\varepsilon)$ 为向量 ε 的 Hamming 权), 则用以下算法可以纠错^[10]:

- (1) 计算 $v = Hy^T$;
- (2) 如果 $v = 0$, 则 $\varepsilon = 0$, $y = c$ (无错误);
- (3) 如果 $v \neq 0$, 则 v 必可以表示为向量组 u_1, \cdots, u_n 中不超过 l 个列向量的线性组合:

$$v = a_{i_1}u_{i_1} + \cdots + a_{i_l}u_{i_l} \quad (1 \leq i_1 < i_2 < \cdots < i_l \leq n),$$

其中 $1 \leq l \leq l$, 此时,

$$\varepsilon = (\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n),$$

其中 $\varepsilon_{i_t} = a_{i_t}$, $\cdots, \varepsilon_{i_l} = a_{i_l}$. 而当 $i \neq i_1, \cdots, i_l$ 时, $\varepsilon_i = 0$, 则可通过 $c = y - \varepsilon$ 计算 c .

定理 4(Golay) 设

$$P = \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 \end{bmatrix},$$

则以 $G=(I_{12} : P)$ 为生成矩阵的二元线性码 G_{24} 具有参数 $[n, k, d]=[24, 12, 8]$ ^[10].

设 $E_m(i, j)$ 表示对调单位矩阵 E_m 的第 i, j 两行 (或列) 得到的初等矩阵, 对定理 4 中的生成矩阵 $G=(I_{12} : P)$ 进行行与行之间或列与列之间的交换得到的矩阵 G' , 则 G' 可表示为 $E_{12}(i_k, j_k)E_{12}(i_{k-1}, j_{k-1}) \cdots E_{12}(i_1, j_1)GE_{24}(i'_1, j'_1) \cdots E_{24}(i'_l, j'_l) = AGB$.

定理 5 对定理 4 中的生成矩阵 $G=(I_{12} : P)$ 进行行与行之间或列与列之间的交换得到的矩阵 G' , 以 G' 为生成矩阵的二元线性码 G_{24}' 也具有参数 $[n, k, d]=[24, 12, 8]$, $H'=(P : I_{12})(B^{-1})^T$ 为 G_{24}' 的一个校验矩阵.

证明 设 $c=(b_1, \cdots, b_{12})G \in G_{24}$, 则 $c=(b_1, \cdots, b_{12})A^{-1}(AGB)B^{-1}=(b_1, \cdots, b_{12})A^{-1}G'B^{-1}=(b'_1, \cdots, b'_{12})G'B^{-1}$, 即 $cB=(b'_1, \cdots, b'_{12})G'$, 故 $cB \in G_{24}'$, cB 只是交换 c 中列的位置, 所以 G_{24}' 和 G_{24} 的最小距离相同, 因此, G_{24}'

也具有参数 $[n, k, d] = [24, 12, 8]$.

因 $(P : I_{12})$ 为 G_{24} 的校验矩阵, 所有 $c(P : I_{12})^T = 0$, 从而 $cBB^{-1}(P : I_{12})^T = cB((P : I_{12})(B^{-1})^T)^T = 0$, 因此 $H' = (P : I_{12})(B^{-1})^T$ 为 G_{24}' 的一个校验矩阵.

定理5中的生存矩阵 G' 的数量有 $12! \times 24!$ (约为 3.0×10^{32}) 个, 可保证暴力破解本文的算法.

3 LECC+MD5算法的网络身份认证

3.1 LECC+MD5算法实现多射流程

如图1所示, 先将用户的密码进行MD5加密得到秘密的MD5值, 并将MD5值用LECC编码得到MD5_LECC值, 再将MD5_LECC值映射到LECC即可得到可以纠错的MD5_LECC_FT值.



图1 LECC+MD5算法实现多射流程

Fig.1 LECC+MD5 algorithm realize one to many map

加密流程如下:

- (1) 计算密码的MD5值(16个字符, 其中字母为大写字母), 如密码 wxdhlgl 的MD5值为 00E220A2F558C267.
- (2) 通过函数 C2B(c) 将每个字符转换为6位的二进制数, c 为字符的ASCII码:

$$C2B(c) = \begin{cases} c - 48, & c < 60; \\ c - 55, & c > 60. \end{cases}$$

如 $C2B(3) = 51 - 48 = 000011$, $C2B(R) = 82 - 55 = 32 = 011011$.

然后将MD5值得第 i 位和第 $2i$ 位的C2B值连在一起构成12维二元的向量, 如密码 wxdhlgl 的MD5值为 00E220A2F558C267, 其12维二元的向量为(共8个):

- (0F): (1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0)
- (05): (1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0)
- (E5): (1, 0, 1, 0, 0, 0, 0, 1, 1, 1, 0, 0)
- (28): (0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0)
- (2C): (0, 0, 1, 1, 0, 0, 0, 1, 0, 0, 0, 0)
- (02): (0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)
- (A6): (0, 1, 1, 0, 0, 0, 0, 1, 0, 1, 0, 0)
- (27): (1, 1, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0)

(3) 求生成矩阵 G' , 将生成矩阵 $G = (I_{12} : P)$ (i 从1到12), 第 i 行与16位MD5值得第 i 个字符的ASCII值模12的值的行交换, 如 wxdhlgl 的16位MD5值为 00E220A2F558C267, 则第3行与第9行 (ASCII(E)=69=9 mod 12) 交换; 将生成矩阵 $G = (I_{12} : P)$ (i 从1到12), 第 i 列与32位MD5值得第 i 个字符的ASCII值模24的值的列交换, 如 wxdhlgl 的32位MD5值为 3A4DAF1500E220A2F558C2674C24EC0E, 则第1列与第3列 (ASCII(3)=31=3 mod 24) 交换; 经过行列互换后所得矩阵就是生成矩阵 G' .

则由密码 wxdhlgl 产生的生成矩阵 G' 为:

$$G' = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

(4)将12维二元的向量和生成矩阵 G' 相乘得到24维二元的向量,随机取出3位,让这3位随机出错或不出错,并将新的24维二元的向量转换为5维36元(0,1,...,9,A,B,...,Z)向量,再将8个5维36元向量连接在一起构成40位字符串.这样的字符串有 $(8 \times C_{24}^3)^8$ (约 4.7×10^{33})个.如由密码wxdhlg产生无错位的40位字符串为:5E12T5PSDD5TCJR10P6R15G3M5WZEF5MA2E5MA2E.

3.2 LECC+MD5算法实现网络身份认证

3.2.1 设置密码

(1)客户端用户输入新密码,确认后,客户端调用MD5加密算法对密码进行加密,将用户MD5加密后的MD5值发送给服务器;

(2)服务器端收到用户密码的MD5值后,用上述加密流程随机产生可纠错的40位字符串,将字符串作为密码保存.

3.2.2 用户认证

(1)客户端用户输入账号、密码,确认后,客户端调用MD5加密算法对密码进行加密,将用户的账号和MD5加密后的MD5值发送给服务器;

(2)服务器端收到用户账号和密码的MD5后,用上述加密流程产生8个24维无错位的二元向量,将数据库中的40位字符串转化为8个24维有错位的二元向量,然后比较8个对应24维的二元向量的错位数,若都小于等于3,则认证通过,否则认证失败.

4 结语

本文提出基于LECC和MD5算法的网络认证方案,利用用户密码的MD5值构造纠错码,将密码的MD5值多射到约 4.7×10^{33} 个可容错的MD5_LECC_FT值.如采用查字典方法进行暴力破解,则数据库的记录就要增加 4.7×10^{33} 倍,因此本文算法在空间和时间上保证了无法暴力破解,具有较高的实用价值.

[参考文献](References)

- [1] 柏银,李志蜀,朱兴东,等. MD5算法及其在远程身份认证中的应用[J]. 四川大学学报(自然科学版),2006,43(2): 305-309.
BAI Y, LI Z S, ZHU X D. MD5 algorithm and the application in distanced-identity authentication[J]. Journal of Sichuan university(natural science edition), 2006, 43(2): 305-309.(in Chinese)
- [2] WANG X Y. How to break MD5 and other hash functions[J]. Lecture notes in computer science, 2005, 3494: 19-35.
- [3] 毛熠,陈娜. MD5算法的研究与改进[J]. 计算机工程, 2012(24): 111-114.
MAO Y, CHEN N. Research and improvement of MD5 algorithm[J]. Computer engineering, 2012(24): 111-114.(in Chinese)
- [4] LIU F B, LIU Y, XIE T, et al. Fast password recovery attack: application to APOP[J]. Journal of intelligent manufacturing, 2014, 25(2): 251-261.
- [5] WANG L, OHTA K, SASAKI Y, et al. Cryptanalysis of two MD5-based authentication protocols: APOP and NMAC[J]. IEICE transactions on information and systems, 2010, E93/D(5): 1 087-1 095.
- [6] 周林,曾伟,徐良华. MD5差分攻击算法在FPGA上的实现技术[J]. 保密科学技术, 2012(4): 69-73.
ZHOU L, ZENG W, XU L H. Implementation technology of MD5 differential attack algorithm on FPGA[J]. Secrecy science and technology, 2012(4): 69-73.(in Chinese)
- [7] WANG X M. Digital signature scheme based on error-correcting codes[J]. IEEE electronics letters, 1990, 26(13): 898-899.
- [8] WANG X M. Modification of the digital signature scheme based on error-correcting codes[J]. Acta electronica sinica, 2000, 28(2): 110-112.
- [9] 钱建发. 纠错码理论及应用研究[D]. 西安:西安电子科技大学, 2010.
QIAN J F. Research on theory and application of error-correcting codes[D]. Xi'an: Xidian University, 2010.(in Chinese)
- [10] 冯克勤. 纠错码的代数理论[M]. 北京:清华大学出版社, 2005.
FENG K Q. Algebraic theory of error-correcting codes[M]. Beijing: Tsinghua University Press, 2005.(in Chinese)

[责任编辑:严海琳]