文章编号：2095-4980(2015)05-0816-05

一种新的消息摘要算法

李安志[1]，姜　鹏[2]

(1.中国工程物理研究院 培训中心，四川 绵阳 621999；2.四川九洲电器集团有限责任公司，四川 绵阳 621000)

摘　要：消息摘要算法在电子商务及信息安全领域有广泛应用，单向性和抗冲突性是消息摘要算法的关键。利用整数向量的拆分求和与数据截取的不可逆性以及非线性变换矩阵的可扩展性，构造一个新的消息摘要算法。对算法进行了描述和讨论，设计测试用例对算法进行了验证。测试结果表明，算法能够达到预期效果，具有简洁、灵活、逻辑清晰、应用面广等优点。

关键词：消息摘要；数字签名；密钥；散列函数；信息安全

中图分类号：TN29　　**文献标识码：**A　　**doi：**10.11805/TKYDA201505.0816

A novel message-digest algorithm

LI Anzhi[1]，JIANG Peng[2]

(1.Institute of Technology，China Academy of Engineering Physics，Mianyang Sichuan 621999，China；
2.Sichuan Jiuzhou Electric Group Co., LTD.，Mianyang Sichuan 621000，China)

Abstract：Message-digest algorithm is widely applied in the field of e-commerce and information security，and its unidirectionality and collision resistance are the keys to the algorithm. By using splitting and summing of integer vector，combining with the irreversibility of data interception and extensibility of nonlinear transform matrix, a new message-digest algorithm is constructed. The algorithm is described，discussed and verified through designed test cases. Test results show that the algorithm can achieve the desired effect. with the advantages of being concise，flexible，logical and widely applicable.

Key words：message-digest；digital signature；key；Hash function；information security

随着以 Internet 为基础的电子商务技术的迅猛发展，以公钥密码[1]、数字签名[2]为代表的加密[3-4]安全技术已成为研究热点。消息摘要是数字签名中的关键环节，可以大大缩短签名时间并提高安全性。另外，在消息完整性检测、消息认证[5]，内存的散布分配及操作系统中帐号口令的安全存储方面，消息摘要算法也有重要应用。消息摘要算法的主要特征是加密过程不需要密钥，并且经过加密的数据无法被解密，只有输入相同的明文数据经过相同的消息摘要算法才能得到相同的密文。消息摘要算法不存在密钥的管理与分发[6]问题，适合于分布式网络上使用。本文提出了一种基于整数运算的消息摘要算法，对算法进行了详细描述和讨论。

1　预备知识

消息摘要函数是将任意长度的消息 M 映射成一个固定长度消息摘要 h 的函数 H:$h=H(M)$。

该函数必须满足如下特性：

1) 给定 M，易于计算 h；

2) 单向性：给定 h，很难根据 $H(M)=h$ 反推 M；

3) 抗冲突性[7]：给定 M，要找到另一消息 M'并满足 $H(M)=H(M')$很难。

著名的消息摘要算法有 RSA 公司的 MD5 算法和 SHA-1 算法及其大量变体[8]。

无论输入的消息有多长，计算出来的消息摘要的长度总是固定的。例如，应用 MD5 算法的消息摘要有 128 bit，用 SHA-1 算法的消息摘要有 160 bit，SHA-1 的变体可以产生 192 bit 和 256 bit 的消息摘要。一般认为，摘要的最终输出越长，该摘要算法就越安全。

MD5 和 SHA-1 算法都是基于位运算的消息摘要算法。本文讨论的是一种基于整数运算的消息摘要算法，算

收稿日期：2014-10-15；修回日期：2014-11-15

法命名为 SMD(Secure Message Digest)。

2　SMD 算法描述

假定用 $\boldsymbol{M}$ 表示要进行消息摘要处理的消息字符串，用 $\boldsymbol{Z}$ 表示消息摘要字符串，用 SMD 表示该摘要算法。

第 1 步：构造 2 个 $k\times k$ 阶非线性变换函数 F_1 和 F_2

设 $\boldsymbol{X}=(x_1,x_2,\cdots,x_k)'$，$\boldsymbol{Y}=(y_1,y_2,\cdots,y_k)'$，$\boldsymbol{Z}=(z_1,z_2,\cdots,z_k)'$。$\boldsymbol{Y}=F_1(\boldsymbol{X})$，$\boldsymbol{Z}=F_2(\boldsymbol{Y})$。

$$F_1(\boldsymbol{X})=\begin{cases} y_1=a_{11}x_1^2+a_{12}x_2+\cdots+a_{1k}x_k \\ y_2=a_{21}x_1+a_{22}x_2^2+\cdots+a_{2k}x_k \\ \vdots \\ y_k=a_{k1}x_1+a_{k2}x_2+\cdots+a_{kk}x_k^2 \end{cases} \tag{1}$$

$$F_2(\boldsymbol{Y})=\begin{cases} z_1=b_{11}y_1^2+b_{12}y_2+\cdots+b_{1k}y_k \\ z_2=b_{21}y_1+b_{22}y_2^2+\cdots+b_{2k}y_k \\ \vdots \\ z_k=b_{k1}y_1+b_{k2}y_2+\cdots+b_{kk}y_k^2 \end{cases} \tag{2}$$

$$\boldsymbol{A}=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kk} \end{bmatrix} \tag{3}$$

$$\boldsymbol{B}=\begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1k} \\ b_{21} & b_{22} & \cdots & b_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ b_{k1} & b_{k2} & \cdots & b_{kk} \end{bmatrix} \tag{4}$$

第 2 步：对消息 $\boldsymbol{M}$ 规整化处理得 n 个分组

1) 在原消息字符串 $\boldsymbol{M}$ 末尾追加消息长度数字字符+原消息字符串第一个字符+矩阵 $\boldsymbol{A}$ 中对角线元素的数字字符+原消息字符串最后一个字符+矩阵 $\boldsymbol{B}$ 对角线元素的数字字符；

2) 将消息分组，每组长度为 $2\times k$ 个字符；

3) 若最后一个分组长度小于 $2\times k$，则填充 1，使其长度为 $2\times k$。

比如：取 k=6，取非线性变换的系数矩阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 如下：

$$\boldsymbol{A}=\begin{bmatrix} 2\,234 & 6\,723 & 3\,902 & 4\,583 & 5\,523 & 1\,856 \\ 3\,765 & 2\,372 & 1\,395 & 2\,943 & 2\,415 & 7\,649 \\ 5\,479 & 4\,387 & 2\,594 & 4\,285 & 7\,246 & 7\,621 \\ 3\,322 & 4\,466 & 7\,601 & 2\,321 & 5\,462 & 5\,367 \\ 4\,758 & 6\,957 & 7\,968 & 4\,635 & 2\,516 & 5\,362 \\ 1\,357 & 2\,468 & 5\,839 & 2\,019 & 1\,029 & 3\,854 \end{bmatrix} \tag{5}$$

$$\boldsymbol{B}=\begin{bmatrix} 1\,640 & 2\,970 & 3\,610 & 5\,210 & 2\,310 & 1\,231 \\ 3\,902 & 1\,717 & 1\,457 & 2\,901 & 3\,807 & 2\,179 \\ 4\,601 & 5\,702 & 2\,410 & 7\,103 & 4\,619 & 3\,141 \\ 3\,576 & 4\,291 & 2\,030 & 1\,223 & 1\,987 & 2\,783 \\ 4\,271 & 5\,832 & 3\,572 & 3\,598 & 1\,435 & 2\,031 \\ 2\,988 & 3\,777 & 2\,888 & 4\,919 & 5\,783 & 1122 \end{bmatrix} \tag{6}$$

消息 $\boldsymbol{M}$ 为“abcdefg”，则规整化后的字符串为：

“abcdefg7a22342372259423212516385 4g16401717241012231435112211”

第 3 步：迭代过程

迭代算法为：

Begin

$\boldsymbol{Y}$=(0,⋯,0); /*$\boldsymbol{Y}$ 初始化为 0 向量*/

for(i=1; i<=n; i++) /*n 为分组数*/

{

将 $\boldsymbol{M}$ 中第 i 分组的 2×k 个字符转换为 k 个整数给向量 $\boldsymbol{X}$；

转换方法是：$\boldsymbol{X}[p]=256\boldsymbol{M}[2k(i-1)+2p-1]+\boldsymbol{M}[2k(i-1)+2p]$ (p=1,2,⋯,k)

$\boldsymbol{X}=\boldsymbol{X}+\boldsymbol{Y}$;

$\boldsymbol{Y}=F_1(\boldsymbol{X})$;

$\boldsymbol{Y}_1=\boldsymbol{Y}/10^6$; /*取 $\boldsymbol{Y}$ 中每个分量除以 10^6 的商存放到向量 $\boldsymbol{Y}_1$ 中*/

$\boldsymbol{Y}_2=\boldsymbol{Y} \bmod 10^6$; /*取 $\boldsymbol{Y}$ 中每个分量除以 10^6 的余数存放到向量 $\boldsymbol{Y}_2$ 中*/

/* 将向量 $\boldsymbol{Y}_1$ 和向量 $\boldsymbol{Y}_2$ 进行混洗求和后给向量 $\boldsymbol{Y}$ */

$\boldsymbol{Y}[i]=\boldsymbol{Y}_1[i]+\boldsymbol{Y}_2[i+1]$ (i=1,2,⋯,k-1);

$\boldsymbol{Y}[k]=\boldsymbol{Y}_1[k]+\boldsymbol{Y}_2[1]$;

$\boldsymbol{Y}=\boldsymbol{Y}+\boldsymbol{X}$ /*将混洗后的结果再与输入消息 $\boldsymbol{X}$ 求和*/

$\boldsymbol{Y}=F_2(\boldsymbol{Y})$

$\boldsymbol{Y}=\boldsymbol{Y} \bmod 10^6+\boldsymbol{Y}/10^6$ /*取 $\boldsymbol{Y}$ 中每个分量除以 10^6 的商和余数并求和*/

$\boldsymbol{Y}=(\boldsymbol{Y}+\boldsymbol{Y}1) \bmod 10^6$ /*保留低 6 位数据，进入下一次迭代*/

}

End

第 4 步：生成消息摘要

1) 将 $\boldsymbol{Y}$ 向量中各分量 $\boldsymbol{Y}[i]$($i=1,2,\cdots,k$)规整化为 6 位整数，若不够 6 位末尾添 0。

2) 将 $\boldsymbol{Y}$ 中每个整数转换为数字字符存放到字符串 $\boldsymbol{Z}$ 中。

算法结束后，字符串 $\boldsymbol{Z}$ 中存放的消息摘要是 $6\times k$ 个数字字符。

3 SMD 算法的讨论

1) 本算法可公开，其保密性不依赖算法保密性。

2) 单向性好。

本算法的单向性是由算法中向量的拆分求和、混洗求和以及数据截取的不可逆性造成的。

假设 $a/10^6+a\%10^6=10^6$，则 a 有 100 万种可能取值。在本算法中，每一轮循环都有 $2\times k$ 个这样的拆分求和，若以每次求和为 100 万计算，要反推原输入值，需进行 $(10^6)^{2\times k}$ 次枚举，以 k 取 6 为例，要 10^{72} 次枚举，经过的循环次数越多，反推原输入值越困难。因此，理论上，由消息摘要不能直接推出原消息。

3) 抗冲突性。

本算法产生的消息摘要是 $6\times k$ 个十进制数，若以 k=6 为例，可能的消息摘要有 10^{36} 个，假设全世界每年产生 10^{24} 个消息摘要，也需要 1 万亿年才可能产生出所有的消息摘要。在本算法中，由于使用了拆分混洗求和和截取运算，并分别由 k 个 6 位整数组成消息摘要，使得消息摘要的分布比较均衡，产生冲突的可能性较小。通过测试发现，即使明文有一点变化，消息摘要也会有较大变化，且这种变化没有规律。关于抗冲突性理论证明较难，还需要大量测试和统计分析，这方面的工作还将继续。

4) 算法简洁，逻辑清晰。

本算法主要用了 2 个非线性变换以及向量的拆分求和与整数截取，涉及到长整数的加、减、乘、整除和求余运算，算法流程简单、清晰。

5) 算法灵活。

消息摘要的长度由分组长度 k 决定，即长度为 $6\times k$。只要调整 k 值或调整每个整数的位数，就可改变消息摘要的长度。非线性变换系数矩阵容易构造也容易修改。

6) 应用面广。

本算法可直接用于口令存储和消息完整性检测，与非对称加密配合可用于数字签名，将非线性变换的系数矩阵作为密钥可直接用于消息认证。

关于该算法的理论分析以及第 2)项、第 3)项和第 6)项将另文讨论。

4　测试用例

取 k=6，构造 2 个 6×6 阶矩阵如下：

$$\boldsymbol{A}=\begin{bmatrix}2\,234 & 6\,723 & 3\,902 & 4\,583 & 5\,523 & 1\,856\\ 3\,765 & 2\,372 & 1\,395 & 2\,943 & 2\,415 & 7\,649\\ 5\,479 & 4\,387 & 2\,594 & 4\,285 & 7\,246 & 7\,621\\ 3\,322 & 4\,466 & 7\,601 & 2\,321 & 5\,462 & 5\,367\\ 4\,758 & 6\,957 & 7\,968 & 4\,635 & 2\,516 & 5\,362\\ 1\,357 & 2\,468 & 5\,839 & 2\,019 & 1\,029 & 3\,854\end{bmatrix} \tag{7}$$

$$\boldsymbol{B}=\begin{bmatrix}1\,640 & 2\,970 & 3\,610 & 5\,210 & 2\,310 & 1\,231\\ 3\,902 & 1\,717 & 1\,457 & 2\,901 & 3\,807 & 2\,179\\ 4\,601 & 5\,702 & 2\,410 & 7\,103 & 4\,619 & 3\,141\\ 3\,576 & 4\,291 & 2\,030 & 1\,223 & 1\,987 & 2\,783\\ 4\,271 & 5\,832 & 3\,572 & 3\,598 & 1\,435 & 2\,031\\ 2\,988 & 3\,777 & 2\,888 & 4\,919 & 5\,783 & 1\,122\end{bmatrix} \tag{8}$$

表 1 给出了 SMD 算法的测试情况，表中左边为输入的明文，右边为用 SMD 算法生成的消息摘要，可以看出，即使明文有一点变化，消息摘要都有较大变化，且这种变化没有规律。

表 1 SMD 算法测试结果

Table1 Test results of SMD algorithm

message	message digest
1	251 676 988 971 840 701 329 856 067 631 338 318
2	522 326 478 499 244 106 479 462 083 126 284 846
3	306 229 958 908 493 462 970 297 006 011 231 384
11	387 465 220 926 124 384 816 403 695 211 395 479
12	397 114 177 632 478 567 619 649 773 874 214 086
13	698 021 923 433 236 287 853 094 764 541 382 747
aa	944 111 324 969 397 436 356 726 200 139 875 893
aA	685 439 225 171 276 244 374 455 343 829 617 163
ac	065 095 788 023 909 069 942 913 210 222 851 124
abcdefghi	276 518 630 211 168 537 414 478 395 914 741 195
abcdefgh	655 795 288 946 096 474 621 196 008 535 394 211

5　结论

除上面给出的测试用例外，还针对不同 k 值和不同非线性变换矩阵设计了测试用例进行测试，并对测试结果进行了比较分析，限于篇幅不一一列出。测试表明，该算法能够达到预期效果。关于该算法的理论分析、时间空间复杂度讨论以及与传统算法的比较将另文讨论。

参考文献：

[1] 张小波,程良伦. PKI 在虚拟专用网络中的应用[J]. 计算机工程, 2011,37(15):114–115. (ZHANG Xiaobo,CHENG Lianglun. Application of PKI in virtual private network[J]. Computer Engineering, 2011,37(15):114–115.)

[2] 肖丰霞,刘嘉勇,李国勇,等. 基于电子钥匙的双向身份鉴别方案[J]. 太赫兹科学与电子信息学报, 2007,5(6):473–475. (XIAO Fengxia,LIU Jiayong,LI Guoyong,et al. Mutual authentication scheme using USB key[J]. Journal of Terahertz Science and Electronic Information Technology, 2007,5(6):473–475.)

[3] 李红芳,杨领军,曹三省. 基于密钥长度的数据加密标准算法改进[J]. 太赫兹科学与电子信息学报, 2013,11(1):131–135. (LI Hongfang,YANG Lingjun,CAO Sanxing. Improved data encryption standard algorithm based on the length of key[J]. Journal of Terahertz Science and Electronic Information Technology, 2013,11(1):131–135.)

[4] 周斌,彭应宁,汤俊. 雷达系统超精简流式AES加密器设计和优化[J]. 太赫兹科学与电子信息学报, 2010,8(3):276–280. (ZHOU Bin,PENG Yingning,TANG Jun. Design and optimization of compact AES as stream cipher in radar system[J]. Journal of Terahertz Science and Electronic Information Technology, 2010,8(3):276–280.)

[5] 鄢喜爱,杨金民,常卫东. 基于散列函数的消息认证分析[J]. 计算机工程与设计, 2009,30(12):2886–2888. (YAN Xiai,YANG Jinmin,CHANG Weidong. Analysis on message authentication based on Hash function[J]. Computer Engineering and Design, 2009,30(12):2886–2888.)

[6] 胡运松,单洪. 异构无线传感器网络高效密钥管理方案[J]. 计算机工程, 2010,36(18):160–162. (HU Yunsong,SHAN Hong. Efficient key management scheme for heterogeneous wireless sensor network[J]. Computer Engneering, 2010,36(18):160–162.)

[7] 刘美,王玉柱,何定养,等. SHA–512 算法及其基于生日攻击的安全性分析[J]. 后勤工程学院学报, 2010,26(3):92–96. (LIU Mei,WANG Yuzhu,HE Dingyang,et al. SHA–512 algorithm and its security analysis based on birthday attack[J]. Journal of Logistical Engineering University, 2010,26(3):92–96.)

[8] 戴慧珺,董文瀚,钟世刚. 一类 SHA-x 改进杂凑算法的设计及分析[J]. 计算机工程, 2009,35(6):181–185. (DAI Huijun,DONG Wenhan,ZHONG Shigang. Design and analysis of modified SHA-x Hash algorithm[J]. Computer Engineering, 2009,35(6):181–185.)

作者简介:

李安志(1966–),男,四川省射洪县人,副教授,硕士,主要研究方向为数值计算. email: li_anzhi123@163.com.

姜 鹏(1992–),男,四川省绵阳市人,本科,主要研究方向为管理信息系统与数据库.

(上接第 793 页)

[7] 油气田开发专业标准化委员会. SYT 5346–2005 岩石毛管压力曲线的测定[S]. 北京:石油工业出版社, 2005. (Oil and Gas Field Development Professional Standardization Committee. SYT 5346–2005 Rock capillary pressure measurement[S]. Beijing:Petroleum Industry Press, 2005.)

[8] 王家禄,高建,刘莉. 应用 CT 技术研究岩石孔隙变化特征[J]. 石油学报, 2009,30(6):887–893. (WANG Jialu,GAO Jian, LIU Li. Porosity characteristics of sandstone by X-ray CT scanning system[J], Acta Petrolei Sinica, 2009,30(6):887–893.)

[9] AL-KHARUSI A S,Martin J. Network extraction from sandstone and carbonate pore space images[J]. Journal of Petroleum Science and Engineering, 2007,56(4):219–231.

作者简介:

吴运强(1987–),男,新疆维吾尔自治区阿克苏市人,博士,教授级高级工程师,主要研究方向为油气田开发.email:wyq@petrochina.com.cn.

左 琛(1991–),男,陕西省汉中市人,在读硕士研究生,主要研究方向为数字图像处理.

赵增义(1963–),男,新疆维吾尔自治区奎屯市人,高级工程师,主要研究方向为油气田开发研究与管理.

王子强(1981–),男,山东省青岛市人,硕士,工程师,主要研究方向为油气田开发.

滕奇志(1961–),女,成都市人,博士,教授,主要研究方向为数字图像处理与模式识别.