第 33 卷 第 6 期 计 算 机 工 程 2007 年 3 月
Vol.33 No.6 Computer Engineering March 2007
·安全技术· 文章编号：1000—3428(2007)06—0101—03 文献标识码：A 中图分类号：TP393

# 基于 Hash 函数的报文鉴别方法

肖皇培[1]，张国基[2]
(1. 华南理工大学计算机科学与工程学院，广州 510640；2. 华南理工大学数学科学学院，广州 510640)

**摘 要：**基于当前网络通信中对报文鉴别码(MAC)的需求，介绍了 Hash 函数在密码学上的安全性质，分析了 Hash 函数在报文鉴别中的应用和针对 Hash 函数的主要攻击。在此基础之上，提出一种基于 Hash 函数的报文鉴别码——伪报文鉴别码(PMAC)。利用当前现有的 Hash 函数来构造 MAC，而不改变原有的 Hash 函数的内部结构。在没有利用任何现有加密算法的基础上，仅应用一个密钥不仅对报文提供了鉴别，而且也提供了机密性。对该伪报文鉴别算法的安全性进行了初步分析。

**关键词：**Hash 函数；报文鉴别码；伪报文鉴别码

# Message Authentication Method Based on Hash Function

XIAO Huangpei[1], ZHANG Guoji[2]
(1. College of Computer Science and Engineering, South China University of Technology, Guangzhou 510640;
2. School of Mathematical Sciences, South China University of Technology, Guangzhou 510640)

**【Abstract】**Based on the requirements of message authentication code (MAC) in current network communication, this paper introduces the cryptographic secure properties of a Hash function, and analyses its applications in message authentication and the main attacks against a Hash function. It presents a new message authentication method based on a hash function, the pseudo-message authentication code (PMAC). The proposed method constructs a MAC by an existing Hash function, and doesn't change the inner structure of this Hash function. The proposed method, which doesn't apply any existing encryption arithmetic, can provide message confidentiality as well as authentication by only a secret key. This paper gives a basic security analysis of PMAC.

**【Key words】**Hash function; Message authentication code; Pseudo-message authentication code

在计算机网络通信中，报文鉴别（也叫报文认证）是一个证实收到的报文来自可信源点且未被篡改的过程，报文鉴别也可用于证实报文的序列编号和及时性，因此利用报文鉴别码可对抗以下的攻击[1]：(1)伪造消息：攻击者伪造消息发送给目的端，却声称该消息源来自一个已授权的实体，或攻击者以接收者名义伪造假的确认报文。(2)内容篡改：以插入、删除、调换或修改等方式篡改消息。(3)序号篡改：在依赖序号的通信协议中（如TCP协议），对通信双方报文序号进行修改，包括插入、删除、重排序号等。(4)记时篡改：篡改报文的时间戳以达到报文延迟或回放的目的。

产生报文鉴别符的方法可归结为 3 种：(1)对报文进行加密，以整个报文的密文作为鉴别符；(2)用MAC，该算法使用一个密钥，以报文内容为输入，产生一个较短的定长值作为鉴别符；(3)用Hash函数，也叫散列函数或杂凑函数，是一个将任意长的报文映射为定长Hash值的公共函数，以Hash值作为鉴别符。常规的加密技术已十分成熟，但出于多种原因，常规加密技术没有被简单地用作报文鉴别符而采用独立的报文鉴别码[2]。目前，用避免加密的方法提供报文鉴别越来越受到重视，Tsudik G[3]指出有一些原因需避免对报文加密，如加密软件或硬件的开销及加密算法的专利保护等。

在最近几年，报文鉴别研究的热点转向由Hash函数导出MAC，如HMAC[4]就是一个成功的例子，这样的目的是多方面的，如现有的Hash函数MD5 和SHA-1 等软件的执行速度比对称分组密码DES快，而且Hash函数也没有其他国家的出口限制而很容易获得其库代码。

## 1 Hash 函数及其性质

Hash函数的基本思想是把其函数值看成输入报文的报文摘要(message digest)，当输入中的任何一个二进制位发生变化时都将引起Hash函数值的变化，其目的就是要产生文件、消息或其他数据块的“指纹”。密码学上的Hash函数能够接受任意长的消息为输入，并产生定长的输出。为了满足报文鉴别的数据完整性需要，Hash函数H()必须满足以下特定的密码学需求[5]：

(1)对任何给定的 x, H(x)要相对易于计算，使得硬件和软件实现实际可行。

(2)对任何给定的值 h，寻找 x 使得 H(x)=h 在计算上是不可行的，即单向性(one-way)。

(3)对任何给定的分组 x，寻找不等于 x 的 y，使得 H(x)=H(y) 在计算上不可行，即弱抗冲突(weak collision resistance)。

(4)寻找对任何的(x,y)，使得 H(x)=H(y)在计算上是不可行的，即抗强冲突(strong collision resistance)。

第 1 个性质可以看作是 Hash 函数用作报文鉴别的实际应用需求，对于后 3 条性质，是针对 Hash 函数在应用中的安全性而特别提出的要求。

## 2 Hash 函数在报文鉴别中的应用

Hash函数可以分为两类：带密钥的Hash函数和不带密钥

**作者简介：**肖皇培(1979－)，男，博士生，主研方向：信息安全；张国基，教授、博导

**收稿日期：**2006-05-12 **E-mail：**xiaohp_2000@163.com

的Hash函数。使用没有密钥的Hash码作为报文鉴别码的体制是不安全的，容易遭受到一些攻击。带密钥的Hash函数通常可用来产生报文的鉴别码，对于通信双方之间传输的任何消息x，用带密钥的Hash函数H()对x做变换，产生H(x)作为MAC附于报文x之后，保证通信双方之间消息的完整性，使双方之间的消息没有被第三方篡改或伪造。目前，利用Hash函数作报文鉴别的最常用方法是在分组密码的密码分组连接(CBC)工作模式和密码反馈(CFB)工作模式中。利用CBC和CFB加密的分组加密方法，一个明文块的改变，在加密时都将会引起相应的密文块以及其后的所有密文块的改变。因此，利用分组密码的CBC和CFB工作模式来构造Hash函数是一个自然的想法[6]。设$E_k$是一个分组密码加密算法，密钥为K，可设$X=X_1X_2\ldots X_L$明文消息的分组，L为其分组数，H()为一个现有的Hash函数，IV为一个初始向量。基于分组密码CBC工作模式的MAC如图 1 所示，基于CFB工作模式的MAC与CBC模式类似。

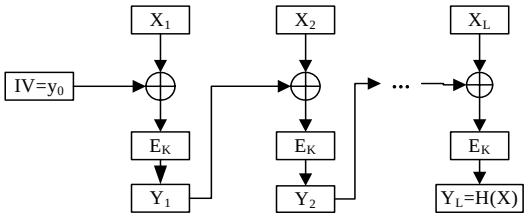


**图 1 基于分组密码 CBC 工作模式的 MAC**

用Hash函数构造报文鉴别码的又一个例子见文献[7]，它通过应用现有的Hash函数与全有全无(AON:all-or-nothing)的性质构造了一系列具有全有全无性质的Hash函数，并用于报文鉴别。全有全无性是一种新的分组加密方式，首先被Rivest[8]提出，其思想是一个攻击者想要得到任何一组明文，他必须先解密全部的密文。因为Hash函数提供的Hash码必须依赖于全部的输入报文，所以全有全无性适合用于构造Hash函数，这意味着对具有全有全无的性质的密码或Hash函数的强行攻击将被削弱。利用全有全无性质的Hash函数，不仅可以为报文提供鉴别，同时也可为报文提供机密性[7]。用Hash函数构造MAC又一个成功的例子是HMAC[4]，HMAC已发表为RFC2104，已经应用到IP安全中，且已被其他的Internet协议如SSL使用。

## 3 基于 Hash 函数的主要攻击

针对 Hash 函数的攻击主要归结为两类:密码分析和强行攻击。

Hash函数的一般结构如图 2，这个结构由Merkle[9]提出，是一个逐层迭代的结构，目前MD5、SHA-1、RIPEMD-160都运用了此结构。由于Hash函数将长度为b比特的报文映射成固定长度为n比特的Hash码，一般b>n，因此Hash函数必定存在冲突，所需要的是使寻找冲突在计算上不可行。在最近几年，针对Hash函数的密码分析已进行了许多工作，也取得了一些成功，主要是针对压缩函数f的内部结构进行密码分析，对Hash函数的密码分析可参考相关文献。

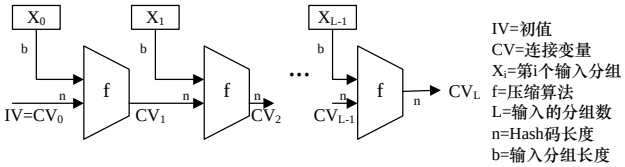


**图 2 Hash 函数的一般结构**

强行攻击即通过穷举所有可能性来获得成功。由于应用于密码学上的Hash函数的单向性、弱抗冲突和强抗冲突性，对于一个长度为n的Hash码，值$2^{n/2}$决定了该Hash码抗强行攻击的强度。对Hash函数比较出名的攻击为生日攻击[2],其主要思想是给定一个Hash函数H()，有$2^n$个可能的输出值H(x)，如果H()有k个随机输入，k必须为多大才能至少存在一个输入y，使得H(y)=H(x)的概率大于 0.5。经过数学计算，只需$k=2^{(n-1)}$上述概率就可满足。这个概率类似于生日悖论的问题，如k个人中，要至少有两个人生日相同的概率$P(365,k)\geqslant 0.5$，k的最小值是多少？计算可得：当k=23 时, $P(365,k)\geqslant 0.5$，这个概率比人们想象的大得多，说明概率结果有时是与人的直觉相违背的。

## 4 伪报文鉴别码(PMAC)

下面提出一种新的报文鉴别码——伪报文鉴别码(PMAC)。该方法利用当前现有的 Hash 函数来构造报文鉴别码，而不改变原有的 Hash 函数的内部结构。该方法在没有利用任何加密算法的基础上，仅应用一个密钥 K，不仅提供报文鉴别，而且也提供机密性。用到的一些符号标记如下:

· n,b：分别为 Hash 函数的输出长度和输入分组长度。

· X,Y：分别为输入报文和输入报文的伪报文。

· H()：任意一个现有的 Hash 函数，如 MD5、SHA-1、RIPEMD-160 等。

· H(x,y)：用 Hash 函数 H()、初始值 x 和输入 y 生成的 Hash 函数值。

· f(x,y)：用 H()算法中的压缩函数 f, 初始值 x 和输入 y 生成的压缩函数值。

· K：通信双方的共享密钥，如果该密钥长度大于 b 比特，用 Hash 函数 H()产生一个 n 比特的密钥，再填充到 b 比特。

· $K^+$：对密钥K进行填充或变换，使其生成b比特的密钥$K^+$。如$K^+=K\oplus i$， i=1,2,...,L；或$K^+=K\oplus i\oplus pad$，其中pad为 01011010 的重复，$\oplus$ 为按比特异或（XOR）。

· ‖：串连。

在通信中，发送方和接受方的操作如下。

发送方:

(1)对密钥K作变换生成b比特的密钥$K^+$。

(2)对输入报文X按n比特分成L组:$X=X_1X_2\ldots X_L$。

(3)计算伪报文$Y=Y_1Y_2\ldots Y_L$: $Y_0=IV$，$Y_i=X_i\oplus f(Y_{i-1},K^+)$，其中IV为初始值，i=1,2,...,L。

(4)计算$PMAC_K=H(IV,Y)$。

(5)发送（Y ‖ $PMAC_K$）。

接收方:

(1)接收（Y ‖ MD）。

(2)计算$PMAC_K=H(IV,Y)$，如果$PMAC_K$不等于MD，证明$PMAC_K$有错，有可能被篡改。

(3)对密钥K的作变换生成b比特的密钥$K^+$。

(4)将伪报文Y按n比特分组$Y=Y_1Y_2\ldots Y_L$。

(5)恢复原报文$Y_0=IV$，$X_i=Y_i\oplus f(Y_{i-1},K^+)$，其中i=1,2,...,L。

计算伪报文的结构如图 3。

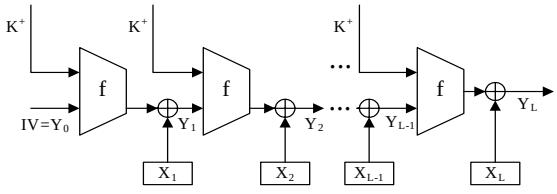


**图 3 计算伪报文结构**

从算法可以看出，PMAC 码利用现有 Hash 函数而没有改变其内部结构，与基于 DES 的报文鉴别相比，Hash 函数的执行速度比分组密码的加密算法更快，程序代码公开更容易获得。从而当出现更快或更安全的 Hash 函数时，可以对算法中的 Hash 函数能轻易地进行替换。

## 5 伪报文鉴别码的安全性分析

伪报文鉴别码是一种基于Hash函数的鉴别码，它的安全性在某种方式下依赖于所使用的Hash函数的安全强度。对于一个n比特输出的Hash函数，生日攻击所需要的计算量是最低的，为发现一个冲突，它需要 $2^{n/2}$次运算量。对本算法中Hash函数的攻击也需要同样的运算量，如果Hash函数的输出是160bit，对于生日攻击目前它已经足够安全了。

MAC 的安全性基于攻击者伪造成功的概率，对 Hash 函数的大多数已知的攻击依赖于对明报文的操作，而伪报文鉴别码传输的是伪报文而不是明报文，攻击者必须想办法截取伪报文来恢复原始的明报文。因为攻击者没有密钥 K，他不能正确地从伪报文中恢复明报文。即使在攻击者获取明文的情况下修改明文，那么伪报文也将相应地被改变，所得的报文摘要不同于原来的报文摘要，从而攻击者在没有正确密钥 K 的情况下无法伪造报文摘要。

对于 PMAC 来说，为了寻找 Hash 函数的冲突报文，攻击者必须寻找独立于 Hash 函数中间链上的有相同报文摘要的报文。PMAC 通过计算第 i-1 组伪报文和第 i 组明报文来产生第 i 组伪报文，任意一组明报文的改变，都将引起本组伪报文和其后所有伪报文的改变。因此，PMAC 可有效地对抗对报文组进行操作的攻击。

## 6 结论

当前网络通信中，对数据完整性的要求越来越高。本文分析了当前网络对报文鉴别的各种需求，介绍了 Hash 函数的密码学安全性质和 Hash 在报文鉴别中的各种应用，总结了对 Hash 函数的主要攻击，最后提出一种基于 Hash 函数的报文鉴别码——PMAC。提出的方法利用当前现有的 Hash 函数来构造报文鉴别码，而不改变原有的 Hash 函数的结构。该算法在没有利用任何加密算法的基础之上，仅应用一个密钥，不仅提供报文鉴别，而且也对报文提供机密性。同时，本文也对该伪报文鉴别算法的安全性进行了初步分析。当然，伪报文鉴别码的安全性还有待于下一步的分析研究，在实际应用中应更加仔细考虑和论证其安全性。

### 参考文献


1 阙喜戎, 孙　锐, 龚向阳, 等. 信息安全原理及应用[M]. 北京: 清华大学出版社, 2003-07: 164-165.

2 William S. 密码编码学与网络安全: 原理与实践[M]. 2 版. 杨　明, 胥光辉, 译. 北京: 电子工业出版社, 2001-04: 190-237.

3 Tsudik G. Message Authentication with One-way Hash Function[C]// Proceedings of INFOCOM'92. IEEE Press, 1992-05.

4 Krawczyk H, Bellare M, Canetti R. HMAC: Keyed Hashing for Message Authentication[Z]. IETF RFC 2104, 1997.

5 王张宜, 李　波, 张焕国. Hash 函数的安全性研究[J]. 计算机工程与应用, 2005, 41(12): 18-19.

6 陈鲁生, 沈世镒. 现代密码学[M]. 北京: 科学出版社, 2002: 129-131.

7 Shin S U, Rhee K H, Yoon J W. Hash Functions and the MAC Using All-or-Nothing Property[C]//Proceedings of Public Key Cryptography. Springer-Verlag, 1999-03.

8 Rivest R L. All-or-Nothing Encryption and The Package Transform [C]//Proceedings of Fast Software Encryption Workshop. Springer-Verlag, 1997-01.

9 Merkle R. One Way Hash Function and DES[C]//Proceedings of CRYPTO'89. Springer-Verlag, 1989.


(上接第 100 页)

```
        k = mod((double) ((3 * x_1 * x_1 + a) / y_1) / 0.2e1, p);
        x_(2) = (k * k - 2 * x_1) % p;
        y_(2) = (k * (x_1 - x_2) - y_1) % p;
        k_i = ((y_(i - 1) - y_(1)) / (x_(i - 1) - x_1)) % p;
        x_(i) = (k_i * k_i - x_(i - 1) - x_1) % p;
        y_(i) = (k_i * (x_1 - x_(i)) - y_1) % p;
        Q[0] = x_(i);
        Q[1] = y_(i);          }
    cgret[0] = Q[0];
    cgret[1] = Q[1];   }
```

其次，运行速度上 Maple 语言比 C 语言稍好一点，因为决定加密、解密的运行速度的关键是数乘运算的速度。对于本文中的程序，按运行时间，单位为 s，比较结果见表 1。

**表 1 程序运行时间(s)**

| | Maple 运行时间 | C 语言运行时间 |
|---|---|---|
| coordinate(1,6,18, 153469313) | 17.2s | 21.4s |
| coordinate(1,6,2,48210713) | 47.6s | 58.3s |
| scalarproduct(27521,2,4,1,15469313) | 1.2s | 1.8s |
| scalarproduct(197521,2,4,1, 48210713) | 7.7s | 8.6s |
| Ord(0,2,2,11) | 3.8s | 4.2s |
| Ord(2,4,1,15469313) | 34.2s | 45.8s |
| encrypt(1289,9,1,2,4, 11111046, 3130304,1, 15469313) | 1.2s | 2.1s |
| decrypt(27521, 159375548, 3857563, 177941803, 230905062, 1, 415142263) | 0.9s | 1.4s |

## 4 结束语

由于虚拟内存太小，文中选择较小的素数(30bits 左右)，在大型计算机上，素数可以取得很大。本文利用 Maple 求出了椭圆曲线上点，并且给出点的加法、数乘运算及确定基点的阶数运算，进而利用 Maple 实现了椭圆曲线的加密、解密。同时，Maple 可求出第 i 个素数，省略了数的素性判别。通过效率性分析，Maple 可以方便地转化 C 语言，特别对于数学公式比较多的程序，用 Maple 要比 C 语言简洁很多，且在运行速度上 Maple 语言比 C 语言稍好一点。

### 参考文献


1 何　青, 袁　荣, 王丽芬. Maple 经典[M]. 北京: 高等教育出版社, 2002-07.

2 Alfred M. Elliptic Curve Public Key Cryptography[M]. Kluwer Academic Publishers, 1997.

3 孙淑玲. elliptic curve cryptosystem 应用密码学[M]. 北京: 清华大学出版社, 2004.

4 王继林, 伍前红. 现代密码学理论与实践[M]. 北京: 电子工业出版社, 2004-07.

5 柯　召, 孙　琦. 数论讲义(上册)[M]. 2 版. 北京: 高等教育出版社, 2001-01.