

面向文本的标识分组加解密模式

张通明, 关建峰

(北京邮电大学网络技术研究院, 北京 100876)

摘要: 针对电子密码本模式无法隐藏明文数据模式、分组固定、密钥唯一等问题, 利用文本文件的字符存储特性, 结合分组加密模式, 提出了一种面向文本流的标识分组加解密模式。该模式设计了双 Hash 映射结构, 并创新性地引入了标识库, 一层映射完成对明文的非连续分组映射, 而另一层映射则基于标识库完成对明文分组的密钥映射。该模式成功隐藏了明文数据模式, 并有效解决了电子密码本模式的分块固定和密钥唯一的问题, 提高了密文的安全性。实验表明, 标识分组加解密模式具有较好的安全性、灵活性和扩展性。

关键词: 文本加密与解密; 加密模式; 分组加密; 标识分组加解密

中图分类号: TP309.7

文献标识码: A

doi: 10.11959/j.issn.2096-109x.2017.00149

Block cipher mode research based on identifier of text files

ZHANG Tong-ming, GUAN Jian-feng

(Institute of Network Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China)

Abstract: The electronic code book (ECB) mode cannot conceal the data model of the plaintext and has only one private key to the whole plaintext. A block cipher mode based on identity for text files was proposed, which absorbed the advantages of block cipher mode and the feature of plaintext string. A double Hash mapping structure was designed and one layer of the mapping structure maps the plaintext into discontinuous groups, another one which was based on the concept of identify library produced corresponding private key to the plaintext group. The model overcame the discontinuous group and single private key problem and effectively improved the security of the cipher text. The experiment results show that the new model has better security, flexibility and extensibility.

Key words: text encryption, encryption model, block encryption, identity encryption

1 引言

随着电子化办公的极大流行, 电子文本文档对信息传递起着重要的作用。但电子文本文档给人们办公方式带来极大方便的同时, 也威胁着人们秘密信息的安全。根据 Search Security 网站对 358 个企业进行的信息安全调查, 目前 30%~40%

的企业机密泄露都是由电子文件造成的。《财富》排名前 1 000 名的企业, 每次由电子文档泄露造成的损失约为 400 万美元^[1]。文献[2]针对 DES 和 AES 的各种分组密码模式为主线, 详细介绍了分组密码模式的设计理念、安全模型及研究现状。文献[3]通过分析串行与并行环境的内在不同, 提出了并行密码体制构造的基本理论, 为并行密码

收稿日期: 2016-12-10; 修回日期: 2017-02-03。通信作者: 张通明, zhangtongming2015@163.com

基金项目: 国家重点基础研究发展计划(“973”计划)基金资助项目(No.2013CB329102); 国家自然科学基金资助项目(No.61003283)

Foundation Items: The National Program on Key Basic Research Project (973 Program) (No.2013CB329102), The National Natural Science Foundation of China (No.61003283)

体制奠定了理论基础。文献[4]提出的分组密码并行工作模式,可以达到线性的加速比,提高了分组密码并行效率。文献[5]引入自然语言处理和水印技术,提出了一种基于自然语言处理的文本加密算法。该算法未将明文转化成密文,而是通过转换将明文信息隐藏。这种加密文本方式仅适合小容量文本,对大容量文本则略显薄弱。文献[6]针对蜂窝网络数据流容易被窃听的问题,提出了一种蚁群密钥生成算法来生成加密密钥加密网络通信数据流。该算法减少了被存储和分发的密钥数量,但其更适于网络间的数据加密传输,不适用于电子文档加密。文献[7]针对手机网络服务在网络传输中的安全问题,提出了一种适用于黑莓手机安全网络服务的密文反馈(CFB, cipher feedback) 8位模式分组加密算法。该算法保障手机网络的通信安全,没有结合电子文档的特点设计加密方式。文献[8]根据DNA序列合成蛋白质的生物过程存在的不可逆性,提出了一种sEncrypt序列加密算法。

综上,目前针对电子文本文档的深入研究较少,而电子密码本(ECB, electronic code book)的加密模式算法连续分块、单一密钥的特点,使算法安全性较低。单一密钥形式无法隐藏原有的明文数据模式,同时,连续分块形式使明文块间

相对独立,又容易成为恶意攻击的重点。本文针对电子文本文档的特点,结合分组密码模式,提出了一种面向文本流的基于标识的分组加解密模式。此模式更适合电子文本流的特点,解决了ECB模式的不足,具有更好的安全性和灵活性。

2 标识分组模式

2.1 电子密码本模式

针对文献[2]提到的分组密码工作模式的研究已有一段历史,早在1980年,美国国家标准局公布的4种DES工作模式就包含ECB密码模式。NIST 2001年在文件800-38A中公布的AES用于保密性的5种工作模也包含ECB密码模式。所以,电子密码本模式是分组密码模式的基础模式和典范。分组密码,也叫块加密(block cipher),一次加密明文中的一个块。将明文按一定的位长分组,明文组经过加密运算得到密文组,密文组经过解密运算,还原成明文组。经典电子密码本模式如图1所示。

符号说明如下。

P_i : 表示第*i*块明文,即明文块1为 P_1 ,明文块2为 P_2 ,明文块*i*为 P_i 。

C_i : 表示第*i*块明文通过加密模式转化后生成

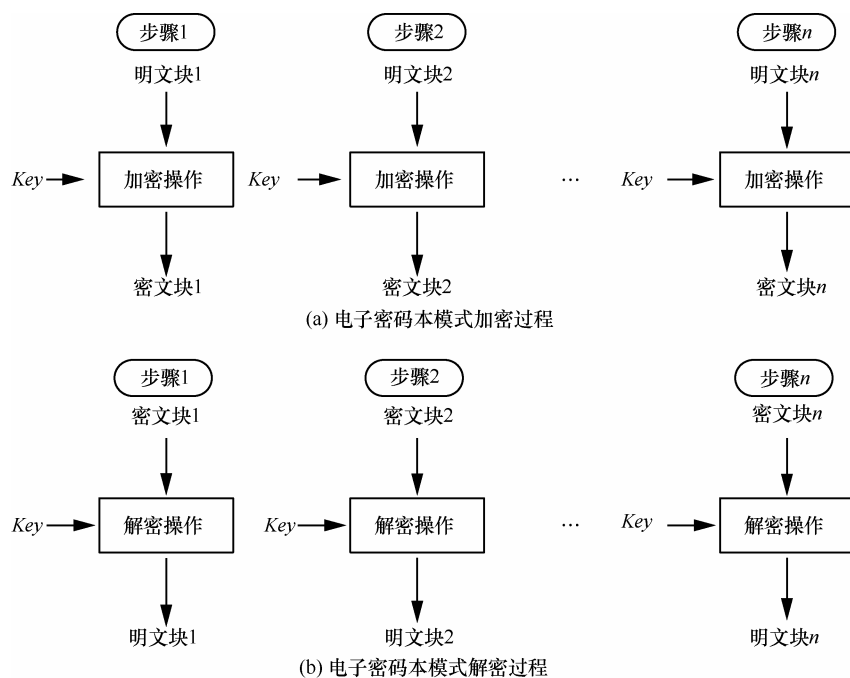


图1 电子密码本模式

的对应密文,即明文块 P_1 转化后生成 C_1 ,明文块 P_2 转化后生成 C_2 ,明文块 P_i 转化后生成密文块 C_i 。

Key: 表示在加解密过程中用到的加解密密钥。

$E_{Key}(P_i)$: 在密钥 Key 作用下加密算法对明文块 P_i 的加密。

$E_{Key}^{-1}(C_i)$: 在密钥 Key 作用下解密算法对密文块 C_i 的解密。

电子密码本模式加密和解密转化过程可以用式(1)和式(2)表示。

$$C_i = E_{Key}(P_i) \quad (1)$$

$$P_i = E_{Key}^{-1}(C_i) \quad (2)$$

分析电子密码本模式的加解密过程可得几点特征,即明文被分成大小相同的若干块、组成各分块的内容连续、所有明文块使用相同的唯一密钥。此模式优点是可并行计算,不存在误差的传递,加解密速度快,且易于标准化。但缺点是这种分组加密方式会使相同的明文组经过加密后得到相同的密文组,因此无法隐藏数据模式;另外,加密明文块相互独立,可以对明文块进行主动攻击(如组的嵌入、删除、重放等);并且,由于固定分组的特点,加密长度必须调整到分组的整数倍。

2.2 标识分组加解密模式

2.2.1 模式思路

标识分组加解密模式在电子密码本模式的基础上做了进一步的改进和优化,主要变革了明文块的固定块大小分组方式,引入了 Hash 算法函数,提出了基于 Hash 算法的明文块分组模式,使明文块的分组变得不固定和不连续。另外,引入了标识的概念,引入标识主要克服了电子密码本模式中密钥 Key 的单一性问题,标识可以拥有多个,并且允许动态添加,标识主要决定不同分组块所选用的不同密钥。同时,标识的选取算法和明文块的分组算法相关联。

整个加密模式的流程如图2所示,其执行过程如下。

步骤1 字符流重新组合生成非连续明文块。将待加密的明文文本内容转化为相应的明文字符流(图2中 $ch_1 \sim ch_n$ 分别代表一个明文字符),根据 Hash 映射规则按序将单字符进行映射,每个字符会被映射到其对应的明文块,这样从 $ch_1 \sim ch_n$ 映射结束,就产生了 k 块经 Hash 映射所得到的明文块1到明文块 k 。

步骤2 根据标识库映射得到各明文块 i 所对应的密钥 Key_i 。标识库中存储着带有标识用户信息(用户身份、当前工作环境、设备编号以及

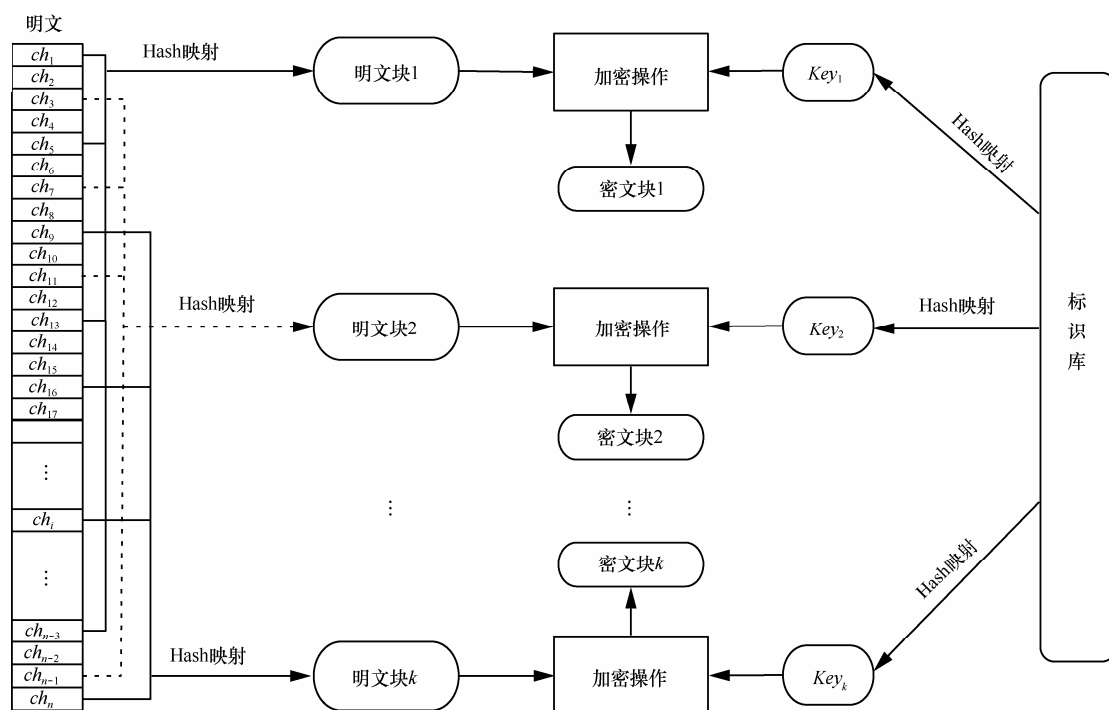


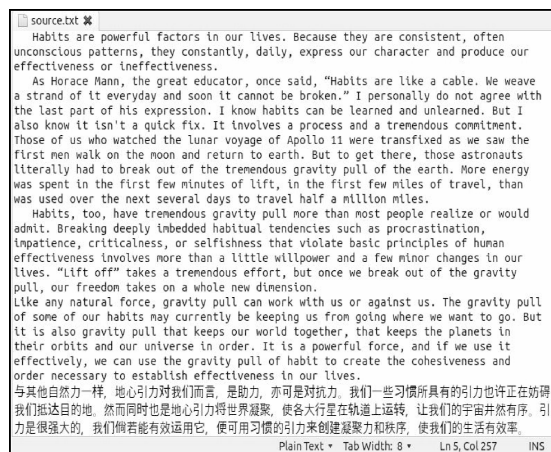
图2 标识分组加密过程

的 $encrypt(pString[i], Key)$ 函数设置了加密函数，将明文通过此函数和密钥 Key 加密为密文。解密过程是加密过程的逆过程，操作流程与加密算法的设计类似，不再详述。

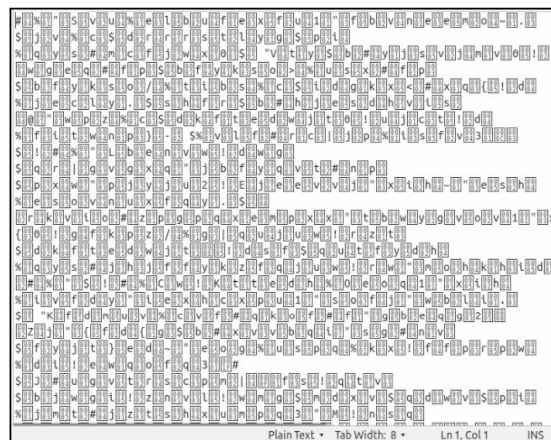
3 实验结果及分析

3.1 实验结果

实验软硬件环境如下。硬件：CPU 为 Intel(R) Core i3 530@2.93GHz 2.93GHz RAM 4 GB。软件：Ubuntu 12.04 LTS 系统 64 bit gcc version 4.4.7 g++ version 4.4.7。本实验测试采用 DES 加密方式来验证电子密码本模式和标识分组模式在不同文本文件大小、文本文件不同分组数目（个）及不同标识库大小（个）条件下 CPU 的用时（ms）及其模式速率（Mbit/s）对比。通过标识分组加密模式的加密明文有效地转化为密文。图 4(a)为操作前的明文效果，图 4(b)为应用本文提出的标识加密模式加密后的效果。



(a) 明文效果



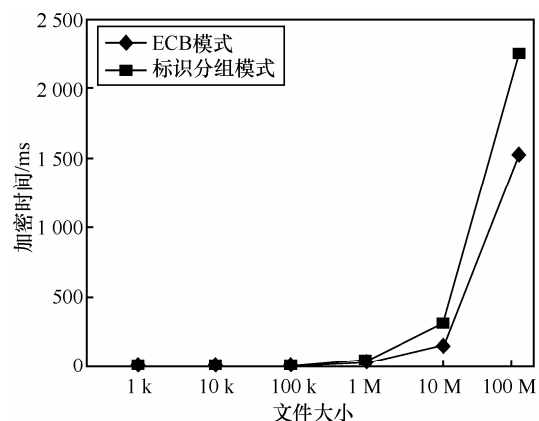
(b) 密文效果

图4 明文密文效果

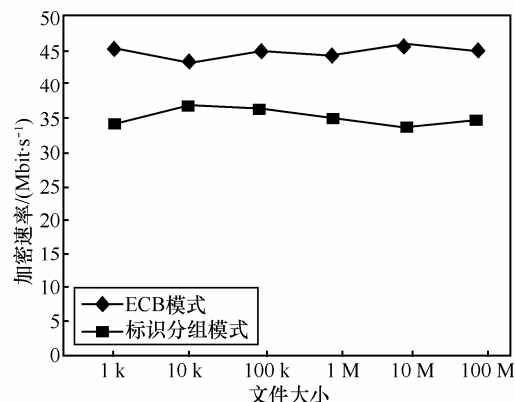
3.2 性能分析

3.2.1 2种模式的加密时间和加密速率分析

实验分别针对不同文件大小的TXT文本文件应用电子密码本模式和标识分组模式加密，所测得的加密时间及加密速率折线如图5所示。在图5(a)中，文件大小10M为临界点，文件大小小于10M时，2种模式的加密时间近似为同一条折线，说明在现有的机器环境下，2种加密模式的时间消耗基本一致。文件大小大于10M时，标识分组模式的时间消耗比ECB模式的时间消耗大，当文件大小为100M时，所消耗的时间差约为700ms。主要原因在于标识分组模型的算法设计中，为了提高安全性应用了2次Hash映射，导致时间消耗有所增加。同时，从图5(b)的分析中可以看出，加密效率上相比ECB模式也有所降低，但在人为感官的可接受范围内。而2种算法在整体加密效率方面均不会随文件大小的变化而波动，总体趋于动态稳定。标识分组模式算法牺牲了小部分效率以提高密文的安全性。



(a) 加密时间



(b) 加密速率

图5 2种模式加密时间和速率对比

3.2.2 分组数目大小对标识分组模式的性能影响分析

图6验证了标识分组模式明文块分组数目变化对算法执行时间及效率的影响。从图6(a)中可以看出,文件大小不变的情况下,分组数目从10增加到10 000时,整个文件加密的执行时间基本趋于平稳状态,即分组数目的变化不会影响算法的执行时间,从这点可以分析出,分组数目的变化对算法的执行时间没有影响。图6(b)表示标识分组模式分组数目变化对加密速率的影响,可以看到6条直线均随分组数目的变化趋于平稳且算法执行速率维持在35 Mbit/s。综合来看,明文块分组数目的变化对算法的执行效率基本上没有影响,根据标识分组模式的设计思想,分组数目的变化意味着更高的安全性,即分组数目越多对应明文的加密Key越多,加密形势越复杂,安全性越高。

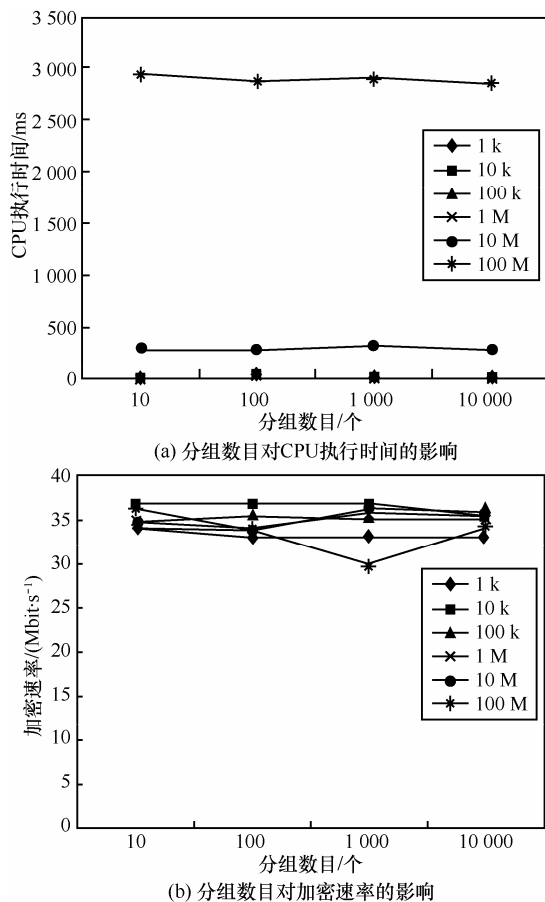


图6 分组数目变化对CPU执行时间及加密速率的影响

3.2.3 标识库大小对标识分组模式的性能影响分析

图7验证了标识分组模式中标识库大小的变

化对算法的执行时间及执行效率的影响。从图7(a)可以看到,标识库大小的变化对不同大小文件的CPU执行时间基本趋于稳定,即对于同一文件的加密时间并不受标识库大小的影响。从图7(b)可以看出标识库的大小变化对标识分组模式加密速率的影响。综合分析,当标识库的大小由5增长到500的过程中,应用标识分组加密模式加密不同的文本文件所得到的加密速率都在35 Mbit/s平衡波动。标识库的大小变化对标识分组模式算法的执行时间及执行效率基本没有影响。同样地,随着标识数目的增多,根据标识分组模式的设计会为明文加密过程提供更丰富的密钥匹配,使加密文本获得更好的安全性。

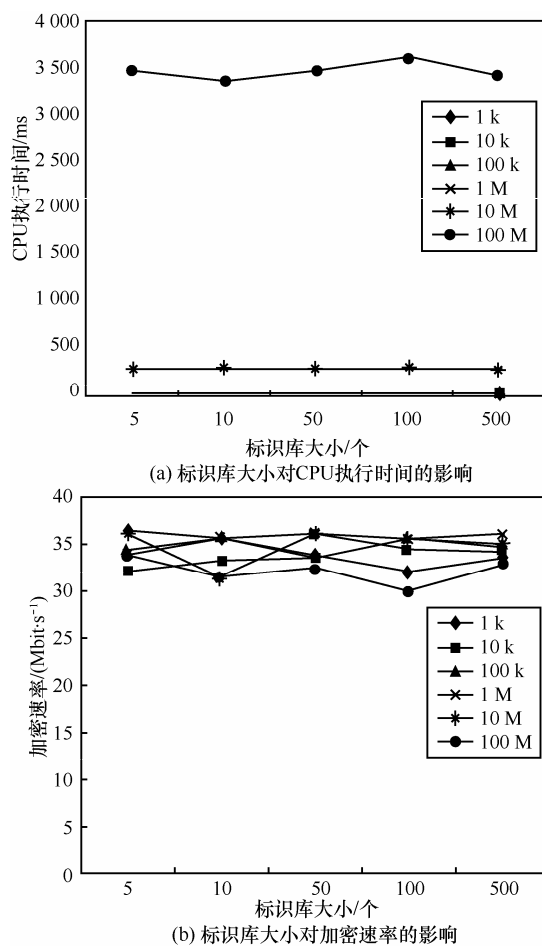


图7 标识库大小变化对CPU执行时间及加密速率的影响

3.3 安全性及扩展性分析

3.3.1 安全性分析

根据 Lockdown.co.uk 近日公布的一份“暴力字母破解”方式获取密码的“时间列表”^[9]。表1中给出了96字符按照密码长度、密码组合数量以

表 1

破解时间

密码长度	密码组合数量	攻击模式 A 级	攻击模式 B 级	攻击模式 C 级	攻击模式 D 级	攻击模式 E 级
4	85 M	14 min	1 min	$8^{\frac{1}{2}}$ s	instant	instant
5	8 B	$22^{\frac{1}{2}}$ h	$2^{\frac{1}{4}}$ h	$13^{\frac{1}{2}}$ min	1 min	8 s
6	782 B	90 day	9 day	22 h	2 h	13 min
7	75 T	24 year	$2^{\frac{1}{2}}$ year	87 day	$8^{\frac{1}{2}}$ day	20 h
8	7.2 Q	2 287 year	229 year	23 year	$2^{\frac{1}{2}}$ year	$83^{\frac{1}{2}}$ day

及破解攻击模式进行破解所需要的时间。

密码组成模式和攻击模式解释如下。

1) 密码组成模式

10 数字：0123456789；26 字母：a~z；52 字母：a~z 和 A~Z；62 字符：10 数字+52 字母；96 字符：10 数字+52 字母+标点符号。

2) 攻击模式解释

A 级：100 000 Password/s—Pentium 100 破解 Windows Password Cache 的速度。

B 级：1 000 000 Password/s—Pentium 100 破解 ZIP 或 ARJ 压缩文档的速度。

C 级：10 000 000 Password/s—双核处理器 PC。

D 级：100 000 000 Password/s—工作站级破解或多个 PC 协同破解。

E 级：1 000 000 000 Password/s—大型计算机网络/超级计算机。

根据上述分析结合标识分组密码模式和电子密码本模式的结构设计，电子密码本模式设计密钥个数为 1，标识分组模式密钥取决于标识库的大小最少为 1。

假设标识库用式(1)表示。

$$S = \begin{Bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{Bmatrix} \quad (1)$$

标识库大小数为 N ，即 $N=n^2$ (n 为大于 0 的整数)，分组数目为 k ，密钥数目为 K ，由标识分组模式的结构设计可知，每个分组对应一个密钥，即密钥数 $K=k$ 。

那么对于一次加密过程根据映射规则，可能的密钥组合数就变成一个排列组合问题，即在 N 中选取 k 个值的可能组合数。假设组合数为 NUM ，

则有

$$NUM = C_N^k = \frac{P_N^k}{k!} = \frac{N!}{k!(N-k)!} \quad (2)$$

于是 NUM 的量级为 $O(N^k)$ 。

电子密码本模式破解时间为 T_{ecb} ，则 T_{ecb} 符合表 1 的破解时间。而对于标识分组模式所需要的破解时间 T_{iden} 为

$$T_{iden} = NUM T_{ecb} \quad (3)$$

由于 NUM 的量级为 $O(N^k)$ ，所以强力破解标识分组模式加密内容需要消耗的时间将指数级增长，安全性更高。

另外，综合分析标识分组模式和电子密码本模式的设计及实现方式，本文提出的基于标识分组加解密模式明文分组方式通过 Hash 映射的方式完成，克服了 ECB 模式的固定分组模式的限制，使明文分组方式可依据不同的 Hash 映射函数设计而不同。同时，引入了标识库，标识库的大小可动态调节，库中元素可以动态更换，而库中的元素又根据 Hash 映射作为不同明文或密文分组的加解密密钥，这使整个加解密流程更具有动态性和灵活性，从另一个方面保证了算法设计的安全性。另外，从密文形势上考虑，标识分组模式由于对不同的分区块采用了不同的密钥，则有效地避免了 ECB 模式单一密钥加密条件下明文数据模式暴露的问题，这样有效地隐藏了明文的数据模式。双 Hash 映射结构的设计，能够构建可行的 Hash 算法，抵抗组的嵌入、重放和删除等攻击行为。

3.3.2 可扩展性分析

基于标识的分组加解密模式的扩展性，可以基于下面几点考虑。1) 模式结构上的扩展。双 Hash 结构的特点给算法的实现本身就带来了特有的灵活性，包括分组方式的灵活性和密钥映射

的灵活性。2) 标识库的可丰富性。标识库的可丰富性在于针对标识元素的定义, 可以是数字、字母、时间甚至是具有一定实际代表意义的属性信息等。3) 可引入访问控制功能。由于标识库元素可设计成具有实际意义的标识信息, 这就使引入有效的访问控制成为可能, 可用于本模式的进一步拓展。

4 结束语

传统的电子密码本模式分组方式简单、加密密钥单一的缺点导致其无法有效地隐藏明文数据模式, 并且无法有效地抵御针对明文的主动攻击。本文通过大量的实验对标识分组模式在性能上进行了论证。其模式算法的稳定性高, 双 Hash 映射结构特点, 使标识分组模式具有更好的安全性、灵活性及可扩展性。这样在继承电子密码本模式优势的基础上, 成功改进了分组块固定、加密密钥单一等不安全特性。但标识分组加密模式的双 Hash 结构设计也使文本的加密速率上有一定程度的下降, 但根据实验测试数据结果可知, 其并不影响用户体验。下一步将优化标识分组模式的设计结构, 进一步提高其效率及安全性, 同时, 研究如何将有效的访问控制功能融入该模式。

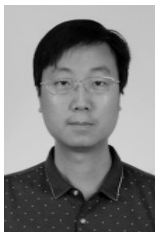
参考文献:

- [1] Search Security. Enterprise information security survey[R].2006.
- [2] 吴文玲, 冯登国. 分组密码工作模式的研究现状[J]. 计算机学报, 2006, 29(1):21-36.
WU W L, FENG D G. The state-of-the-art of research on block cipher mode of operation[J]. Chinses Journal of Computers, 2006, 29(1):21-36.
- [3] 卿斯汉. 并行密码体制的构造[J]. 软件学报, 2000, 11(10): 1286-1293.
QING S H. Establishing of parallel cipher mechanism[J]. Journal of Software, 2000, 11(10): 1286-1293.
- [4] 殷新春, 陈伟鹤, 谢立. 分组密码的并行工作模式[J]. 小型微型计算机系统, 2005, 26(4): 600-603.
YIN X C, CHEN W H, XIE L. Parallel processing model of the block cipher[J]. Mini-Micro Systems, 2005, 26(4): 600-603.
- [5] JING X, HAO Y, FEI H, et al. Text encryption algorithm based on natural language processing[C]//The 4th International Conference on Multimedia Information Networking and Security. 2012: 670-672.
- [6] SREELAJA N K, PAI G A V. Swarm intelligence based key generation for text encryption in cellular networks[C]//The International Conference on Communication Systems Software and Middleware and Workshops. 2008:622-629.
- [7] WANGSADIREDJA M, MUNIR R. Text and file encryption application for blackberry using cipher feedback 8-bit mode[C]//The International Conference on Electrical Engineering and Informatics. 2011:1-6.
- [8] BONHAM-CARTER O, PARAKH A, BASTOLA D. sEncrypt: an encryption algorithm inspired from biological processes[C]//IEEE International Conference on Trust, Security and Privacy in Computing and Communications. 2013:321-327.
- [9] The Home Computer Security Centre. Password recovery speeds [R]. 2009.

作者简介:



张通明(1989-), 男, 河北保定人, 北京邮电大学硕士生, 主要研究方向为网络安全、移动互联网。



关建峰(1982-), 男, 河南巩义人, 博士, 北京邮电大学副教授, 主要研究方向为移动互联网、网络安全、未来网络架构。