用混沌系统实现文本文件的加解密

邓绍江　廖晓峰

（重庆大学计算机学院　重庆400030）

摘　要　电子政务中要处理大量的电子文本文件，如何保证电子文本文件在网络传输中的安全是非常关键的。目前已经有许多传统的加密方法，本文分析了混沌系统和传统加密方法的联系，根据混沌系统对于初始值的高度敏感性，提出了将混沌系统的状态作为密钥信号对文件进行加密的一种方法，用这种方法加密，密文分布均匀，能有效提高文本的保密性，提高信息在传输中的抗破译能力。

关键词　混沌，加密，Logistic映射

Encryption and Decryption for Text Files Using Chaotic Systems

DENG Shao-Jiang　LIAO Xiao-Feng

(College of Computer Science, Chongqing University, Chongqing 400044)

Abstract　A large number of electronic text files should be dealt with in electronic government affairs. So, how to guarantee the security of these text files transmitted in the network is becoming a vital issue. Recently, based on traditional cryptology, many encryption algorithms have been reported. By analyzing the nature relationship between chaotic systems and traditional cryptology, a new algorithm is proposed, which is based on the characteristics of chaotic systems, namely, deep dependence on the initial conditions. Using the states of the chaotic systems as secret key signal, the proposed algorithm can make the resulting cipher signal uniform, and therefore, can improve the security and the ability to resist attack.

Keywords　Chaos, Encryption, Logistic map

1　引言

现在很多的企事业单位都开始使用电子政务系统，在电子政务系统中经常传输的是保密的文件和数据。这些文件和数据一旦被别人截取或篡改，必将导致管理的混乱。所以，如何提高电子政务中信息的安全性，是非常关键的。在电子政务的文件和数据传输中，安全的主要隐患是截取或修改内容，破坏数据的私密性和完整性。解决传输安全的最有效方法是对网络传输的文件和数据进行加密。

1990年以来，混沌通信和混沌加密技术成为国际电子通信领域的一个热门课题。混沌系统是一种高度复杂的非线性动力学系统，具有对初始条件和系统参数非常敏感以及生成的混沌序列具有非周期性和伪随机性的特性。因此，混沌系统近年来被应用于保密通信领域，混沌密码学方法也得到了大量研究。扩散和混淆是由Shannon提出的密码系统设计的两个基本原理，扩散的作用在于将明文的统计特性散布到密文中去，实现方式是使得明文的每一位影响密文中多位的值；混淆则是使密文和密钥之间的统计关系变得尽可能复杂，使敌手无法得到密钥。混沌的混合(mixing)特性(与轨道的稠密性和对初始值的敏感性直接相联系)对应于传统加密系统的扩散性能，而混合特性和对系统参数的敏感性相应于传统加密系统的混淆特性。可见，混沌系统优异的混合特性保证了混沌加密器的扩散和混淆作用可以与传统加密算法一样好。混沌加密器通常工作于连续值方式，而传统加密器工作于离散值方式，这是两者的主要差别。针对混沌加密和传统加密的不同，1998年，M. S. Baptista提出把由混沌系统演化得到的连续值映射成可以用于加密的离散值，来实现混沌加密[1]。该方法利用低维混沌动力学系统的遍历性特点，通过混沌系统的当前初值和迭代次数对组成明文的基元符号进行加密，具有非线性的加密功能。后来K. W-Wong对该方法做了改进，克服了该方法的两大缺陷[3]。主要的改进在于不用参数η来控制加密结果，而是在每个符号加密结果的控制中，引入了迭代的最高限r_{max}和随机的最低迭代次数r。本文在该方法的基础上，仔细分析了汉字加密和英文加密的区别，提出了一种用于汉字加密的新方法。

2　混沌系统加解密算法的实现

2.1　理论分析

目前工作中处理的公文都是一些由字母和汉字组成的文本文件。因此我们可以针对每个字母和汉字实现加密。根据国家标准局公布的汉字标准，可以把6732个两级汉字分成94个区，每个区分成94个位。因此一个汉字的表示可以由两个部分组成：由区和位来表示。这样我们可以针对汉字的区和位来进行加密。具体做法是首先把94个区和94个位映射到唯一的数字区间上(这个区间用ε表示，见表1)。

表1

1	2	3	…	92	93	94
0.2min	0.20638	0.21276	…	0.78724	0.79362	0.8max

表1中表示的是：将混沌吸引域的一部分[0.2,0.8]划分成94个等间隔的子区间，每个区间的范围为：$[\min+(i-1)\varepsilon, \min+i\varepsilon]$其中$i\in[1,93]$。在本文中我们取$\min=0.2$，$\max=0.8$，所以$\varepsilon=0.00638$。每个$\varepsilon$区间与每个数字一一对应以形成码表；每个汉字的加密密文由两次加密完成。加密算法中我们使用一维的Logistic映射。

$$X_{n+1}=bX_n(1-X_n) \tag{1}$$

其中：$X_n\in[0,1]$，b是混沌系统的参数，而且满足$0\leqslant b\leqslant 4$。由

*)本文得到重庆大学基础及应用基础基金项目(717411039)资助。邓绍江　博士生研究生，研究方向为信息安全、混沌密码学。廖晓峰　教授，博士生导师，主要研究领域为神经网络。

Lyapunov 指数的计算可知，当3.57≤b≤4时，该映射是混沌的。设置一个映射参数 b（例如 $b=3.87$）使式(1)具有混沌行为，这样就可以对文件和数据进行加密。用系统式(1)的轨道点落在每个数字区间上所需的迭代次数作为相应字符的区号和位号的密文。具体算法描述如下：

2.2 算法描述

密钥 key：初始条件 X_0，控制参数 b。

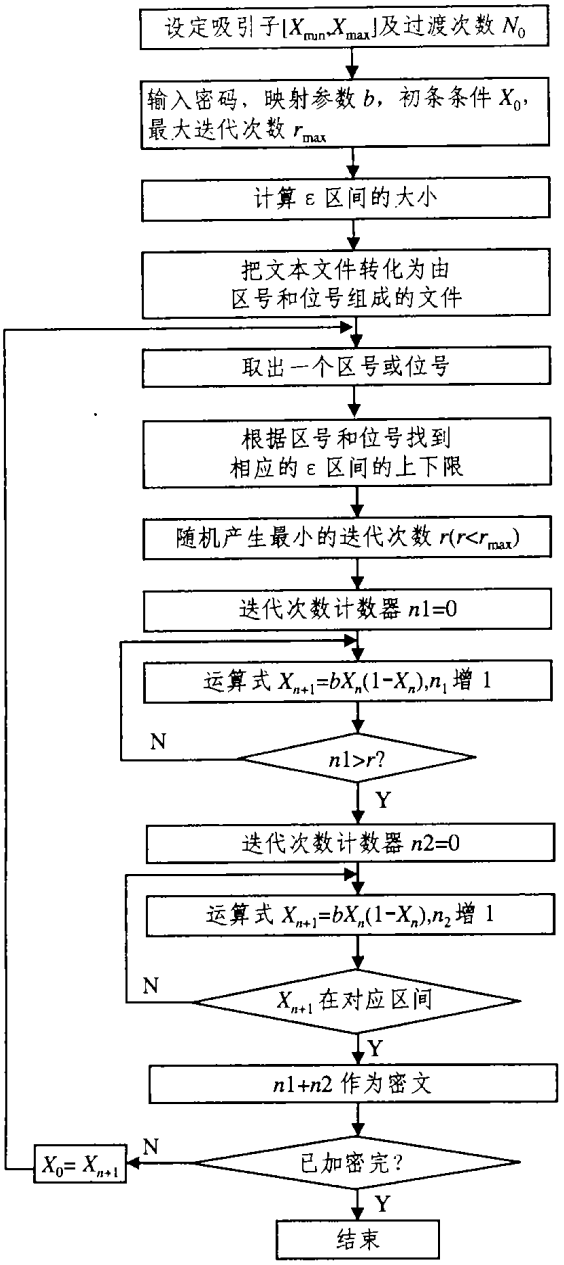

图1 加密程序框图

加密步骤：首先设定一个参数 r_{max}，对每一个要加密的汉字首先选定一个伪随机的初始循环次数 r（250<r<r_{max}），并让混沌系统迭代 r 次。然后继续迭代 c_1 次，$X_0'=F^{c_1+r}(X_0)$，当 X_0'落在明文中的第一个字符的区号所对应的数字区间时，就把循环的总次数(c_1+r)作为第一个字符密文的前半部分。把 X_0'作为新的初始条件迭代 c_2 次，$X_0''=F^{c_2+r}(X_0')$，当 X_0''落在明文中的第一个字符的位号所对应的数字区间时，就把第二次迭代的总次数(c_2+r)作为第一个字符密文的后半部分。然后把 X_0''作为加密第二个字符的初始条件，如此周而复始直至加密结束。

解密方法：取出密文中的内容，把混沌系统迭代相应的次数，通过得到的值在表1中查出相应区间对应的区号或位号。再根据区号和位号查出相应的字符。

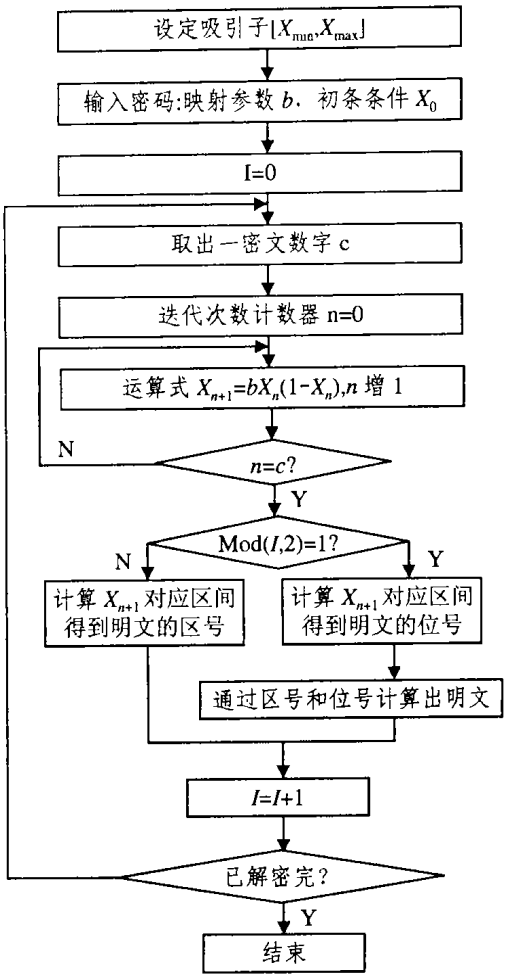

图2 解密程序框图

3 实例

我们用此算法来加密一个文本文件，明文和密文如下所示。

加密前明文

中国在酒泉卫星发射中心进行了我国首次载人航天发射。10月15日上午9时整，我国神舟五号载人宇宙飞船由“长征”二号 F 型火箭成功发射升空。乘坐神舟五号飞船进入太空的是来自中国人民解放军的航天员杨利伟。

加密后密文

337 18826 6677 26626 19617 15843 11928 30000
27073 24511 5979 28223 24143 16872 10406 546
3277 12004 4917 5908 32464 14890 4244 284
1304 12438 17422 19179 20089 19924 5693
22496 14935 11589 2209 20921 26345 26443
17330 10180 28812 24184 31548 30568 18039
4778 15414 7783 28259 6878 26290 27645 32897
32878 20049 12922 9765 9928 27600 1393 12525
3069 22515 2149 538 30270 9335 9422 19683
24928 27888 23875 15938 7245 24427 15453
15665 31366 24508 3731 19650 12662 24087
20443 18989 13042 5008 7394 13965 26413
17249 32496 24668 11385 5620 21591 16121
2155 22958 16698 4871 31140 4716 29743 22815
10831 14492 2497 32255 22579 5303 29052
27147 19292 6368 5996 26936 15732 5143 16655
24135 13661 9675 18681 22674 24839 23662
15607 4105 12187 27352 1595 17049 22109
13992 3544 31311 30250 18347 11962 15709
12727 27846 10784 15769 9017 32226 9950
24281 18747 6658 24954 27509 13183 16582
29348 953 32952 18820 2038 17548 7026 28075
20650 21622 6638 27737 4074 3912 24492 10446
31071 9440 11125 4980 24087 27627 23296
19697 24752 8457 4839 139 2054 26523 27971
7154 3898 18180 1016

结语 本文分析了传统的加密方法和混沌之间的联系，

一种基于RPUC的Web文档索引库的更新算法*)

熊海灵 伍 胜 余建桥 李 航
(西南农业大学信息学院 重庆400716)

摘 要 为提高搜索引擎文档索引库有效性验证的效率,本文提出了一种综合考虑网页更新频度、用户兴趣度及其内容重要程度诸因素相结合以确定文档索引库更新队列的算法。算法将用户的检索率、点击率、网页的Page Rank值和更新频度作为一个特征向量,与不同种类的网页的特征权值组成的矩阵相乘,求得网页的类型向量,依据类型向量实现对文档索引库更新队列的优化,算法改进了统一更新策略周期长、单一更新策略可能产生改变频繁而非常重要的网站长期又得不到更新的问题。

关键词 搜索引擎,索引数据库,检索率,Page Rank,更新频度,点击率

A Refreshment Algorithm for Web Indexed Database Based on RPUC

XIONG Hai-Ling WU Sheng YU Jian-Qiao LI Hang
(College of Information, Southwest Agricultural University, Chongqing 400716)

Abstract In order to improve the efficiency of the validity check on the indexed database of search engine, an algorithm for the refreshment of Web indexed database is presented, which is based on the RPUC, i.e. Retrieval ratio, Page Rank, Updated ratio, Click ratio. They constitute the feature vector of a Web page. Cross multiplying the feature vector and matrix, which is consisted of the characteristic values of various Web pages, type vectors of Web pages can be calculated respectively. By means of the type vectors, indexes in refreshing queue can be arranged optimally. Eventually, demerits of the uniform freshness strategy and personal freshness strategy for indexed database are eliminated effectively.

Keywords Search engine, Indexed database, Retrieval ratio, Page Rank, Updated ratio, Click ratio

1 引言

随着Web信息的迅速增加,搜索引擎从1995年开始逐渐发展了起来。查全率和查准率是衡量一个搜索引擎的主要性能指标[7]。搜索引擎文档索引库的内容的获取、组织与更新是提高搜索引擎精度指标和搜索结果的"新鲜性"的关键因素[2,3]。目前对搜索引擎文档索引库的内容的获取、组织都有比较深入的研究,而对文档索引库更新策略的研究还不多见。事实上,搜索引擎收集的网页数量和其文档索引库的更新速度存在着不可调和的矛盾,因此文档索引库更新策略的问题是一个战略性问题,是一个迟早都要面临和解决的问题。

目前,已存在的更新方案大致可归为以下两类[6]:一是统一更新策略,网络蜘蛛以同样的频率访问集合中的所有网页,而不考虑这些网页的改变频率。二是个体更新策略,不同网页其改变频率也不同。直觉上,更多的刷新应该分配给那些更新快的页面,但研究表明,用较高的频率刷新更新快的页面并不一定是明智之举,频繁刷新改变快的网页不能明显提高搜索效率,因为可能产生那些改变频繁的网站长期得不到更新的问题[1,5]。

本文提出了一种Web文档索引库的快速更新算法,该算法综合利用文档索引的检索情况(检索率:R)、文档的页面权值(Page Rank:P)、文档的更新情况(更新频度:U)以及检索出的文档索引的点击情况(点击率:C)(统称RPUC)对文档索引进行分类,并确定其更新周期,按照此更新周期进行信息

*)本论文得到国家自然科学基金(40731061)和重庆市教委科学技术研究项目资助。**熊海灵** 博士生,主要研究方向为信息检索,数据挖掘。**伍 胜** 硕士,主要研究方向为Web技术。**余建桥** 博士,教授,主要研究方向为数据库与人工智能。**李 航** 教授,博导,主要研究方向为分形理论。

找到了一种如何把离散数字空间和连续空间相对应转化的方法。实现了用混沌系统对文本文件的加密。此算法优点十分明显。

1)利用r_{max}控制密文分布和加密时间。引进了参数r_{max},通过此参数可以有效地控制加密时间和密文分布之间的平衡,如果r_{max}取得大密文分布更为广泛,分布更为均匀,但同时会加长加密解密的时间。反之,如果r_{max}取得太小,则密文分布在狭窄区间,分布不均匀,但同时会缩短加密的时间。

2)加密强度提高。每一次的加密中引入了参数$r(r>250)$,r是一个伪随机序列,从而使得每一次加密的轮数都不一样,增强了密文的强度。

3)密文的空间加大。由于一个汉字由两个字节所组成,因此加密后的密文是明文的两倍,所以增加了加密后的传输时间。

参 考 文 献

1 Baptista M S. Cryptography with chaos Phys. Lett. a, 1998, 240: 50
2 Schneier B. Applied Cryptography: Protocols, Algorithms, and Source Code. In: C. Wiley, New York, 1996
3 Wong W-K, Lee L-P, Wong K-W. Comput. Phys. Commun, 2001, 138: 234
4 郝柏林. 从抛物线谈起——空气动力学引论[M]. 上海:上海科技教育出版社,1997