基于数据随机性特征和极速学习机的加密数据流识别

周宇欢[1] 蒋大伟[2] 龚 勇[1] 陈 聪[3]
(解放军理工大学指挥信息系统学院 南京 210007)[1] (解放军国际关系学院 南京 210007)[2]
(上海宝信软件股份有限公司南京分公司 南京 210007)[3]

摘 要 为了在不解密加密数据的前提下获取加密数据流的类型信息,提出一种基于数据随机性特征和模式识别的加密数据流识别方法。该方法利用加密数据与非加密数据,或者不同类型加密数据 0,1 分布的随机性特性作为分类特征,再利用模式识别方法对不同数据进行建模,从而实现对不同类型数据的自动识别。首先利用 NIST 随机性测试方法对数据流进行分析,将得到的 15 类随机性测试得分作为分类特征;然后对不同类型的数据流分别建立分类模型;最后利用训练好的数据模型对未知数据流进行识别。仿真实验显示,与仅用单个随机性特征进行明密数据识别相比,采用模式识别方法可以将错分率由原来的 60%以上下降到 30%左右;进一步利用滤波器方法对 15 类随机性特征进行优化降维,平均错分率进一步下降到 15%左右。

关键词 加密数据,数据随机性,模式识别,极速学习机
中图法分类号 O235 **文献标识码** A

Encrypted Data Stream Recognition Based on Data Randomness and ELM

ZHOU Yu-huan[1] JIANG Da-wei[2] GONG Yong[1] CHEN Cong[3]
(Institute of Command Information System, PLA University of Science and Technology, Nanjing 210007, China)[1]
(PLA International Studies University, Nanjing 210007, China)[2]
(Shanghai Baosight Software Co., Ltd. Nanjing Branch, Nanjing 210007, China)[3]

Abstract This paper presented the identification method of encrypted data stream based on data randomness and pattern recognition without decrypting the encrypted data. The method uses the randomness distribution characteristics of different data as classification features, and then uses the pattern recognition method to classify different types of data. Firstly, the randomness test method NIST is used to analyze the data stream, getting the 15 kinds of randomness test values as the classification feature. Then the method creates classification models for different types of data streams. Finally, the method uses the trained model to identify the unknown data stream. Simulation results show that using the 15 kinds of features, the proposed method can effectively classify the different types of data stream, and the error rate decreases from 60% to 30%. Using the feature optimization method, the error rate drops to 15%.

Keywords Encrypted data, Data randomness, Pattern recognition, ELM

1 引言

加密代理系统的一个关键环节就是在用户终端和境外代理之间建立一个加密的信道,从而规避安全审查。因此,通过分析数据的随机性特性,提取能够表征数据加密特点的随机性特征,就可以在不解密加密数据的前提下,分析密数据来源。

本文将 NIST 随机性测试方法对数据的随机性测试得分作为特征,利用模式识别的方法建立相应的数据类型模型,最后将建立的模型运用到明密数据类型的识别中。从识别结果可以看出,运用极速学习机(ELM)的方法可以有效区分加密代理数据的类型,相比直接应用随机性测试得分来区分明密数据,其识别率有大幅度的提升。

2 加密数据流的随机性度量

目前所有的加密代理软件都通过对数据内容进行加密来绕过基于内容过滤的防火墙,本文采用了目前比较流行的 NIST 随机性测试方法进行随机性测试[1-3],包括频率测试、组内频率测试、游程测试、最大 1 游程测试等共 15 种随机性测试方案。对于这 15 种测试方法的描述如表 1 所列。

本文受中国博士后科学基金第八批特别资助项目(2015T81081),第 54 批中国博士后面上项目一等资助(2013M542425),江苏省自然科学基金青年基金面上资助项目(BK20140075),江苏省博士后科研资助计划项目(1401001A)资助。
周宇欢(1980—),男,博士,主要研究方向为语音识别、说话人识别、数据流识别等,E-mail:52761263@qq.com。

表 1　NIST 随机性测试方法

序号	名称	简介
1	频率测试	频率测试最基本的随机性测试之一，对于密数据而言，如果频率测试的随机性不通过，则说明该加密数据不随机，因为好的加密算法的序列中的 0 和 1 应大致相等。
2	组内频率测试	组内频率测试是在频率测试的基础上发展起来的，目的是更加细致地刻画序列局部的随机性，组内频率测试将序列分成若干块，当块数为 1 时，即为频率测试。当每一个分块内 0 和 1 的数目大致相等时，说明该序列的局部随机性良好。
3	游程测试	游程测试的主要目的是计算序列中 0 或 1 的连续状态，以及 0 和 1 交替的状态。对于好的加密算法，其游程的性质也应该满足随机性的要求。
4	最大 1 游程测试	最大 l 游程测试是测试序列的若干子块中最长的 1 游程的规律性，由此判断序列的随机性。
5	二元矩阵秩测试	二元矩阵秩测试是通过计算序列中各个子矩阵的秩来测试子矩阵间是否线性独立。通过这一统计结果判断序列的随机性。
6	离散傅里叶变换测试	离散傅里叶变换测试通过计算序列的傅里叶变换，观察其峰值来测试序列的周期性，又叫频谱测试。
7	非重叠匹配测试	非重叠匹配测试是通过统计序列中特定数据流的出现频率来判断序列的随机性。
8	重叠匹配测试	重叠匹配测试也是通过统计序列中特定数据流的出现频率来判断序列的随机性。但与非重叠匹配测试不同，无论何种情况，数据统计窗口都是向后移动一位，而非重叠匹配测试在发现特定数据流后，数据统计窗口会跳过已经测试的数据序列，从原窗口最后一个数据开始重新统计。
9	全局通用测试	全局通用测试通过测试数据被压缩后信息是否损失的情况，来判断序列的随机性，如果压缩后信息有很大的数据损失，则说明序列是随机的。
10	线性复杂度测试	线性复杂度测试通过测试序列是否具有一个较长的 LFSR(线性反馈移位寄存器)阶数，如果足够长，则可以认为数据的复杂度高，随机性强；否则认为数据不随机。
11	串行测试	串行测试通过统计各种比特长度数据流在序列中出现的次数，以此与随机序列的统计结果进行比较，如果两者相当，则认为该序列随机，否则认为不随机。
12	近似熵测试	近似熵测试类似于串行测试，也是统计序列中比特长度相差为 1 的数据块交替的频率，观察其是否与真正随机的序列统计结果近似，如果相似，则认为该序列随机。
13	累积和测试	累积和测试将原序列中的 0 换成 −1，然后计算序列的累积和，观察累积和的最大最小值是否与真正随机序列相似，如果相似，则认为该序列随机。
14	随机偏移测试	随机偏移测试是在累积和测试的基础上，观察累积和出现特殊状态的循环次数，并与真正随机的序列进行比较，如果相似，则认为该序列随机，否则不随机。
15	随机偏移测试变体	随机偏移测试变体与随机偏移测试类似，是其统计累积和出现特殊状态的次数与期望值的偏离程度，偏离越小，序列越随机。

3　加密代理数据流识别

传统方法测试数据的随机性通常是采用一种或几种随机性测试方法，测试数据是否能够通过这些方法的测试，由于不能很好地综合这些方法的测试结果，因此效果并不理想，特别是直接用于加密代理软件的密数据检测时，效果很差。本文利用模式识别的方法，将 NIST 的 15 种随机性测试方法得到的测试数值作为数据的特征值，分别建立密数据和非密数据的模型，在自动识别加密数据的工作中取得了较好的效果。

本文方法分为训练和识别两大部分，其流程如图 1 所示。

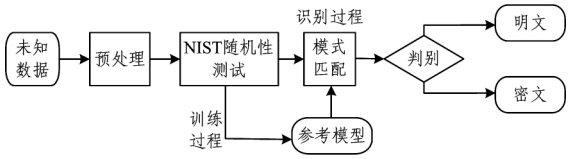

图 1　基于模式识别的密数据检测流程

从图 1 可以看出，整个识别流程主要包含了 5 大模块：数据预处理模块、NIST 随机性测试模块、模型训练模块、模式匹配模块、判决模块。其中各组成模块的功能如下。

(1)预处理模块

通过对输入的数据包进行预处理，滤除掉其中不重要的信息，比如协议信息等，只留下进行 NIST 随机性测试的有效的数据部分，这一部分还包括对未知数据进行划分，如将不同长度的数据截取成若干段相等长度的数据，或者在模型训练时给数据打上标记，明文数据为一类标记，密文数据为一类标记，甚至可以为不同加密方式的密文数据打上不同的标记。

(2)NIST 随机性测试模块

将预处理之后的数据进行 NIST 随机性测试，一共包括 15 种测试，如表 1 所列，并将测试数值用来表征该数据的加密特征。在这一部分还可以进行特征的优化，因为 15 种测试各不相同，对于特定的加密数据，表征效果也不尽相同，因此可以对 15 类特征进行优化，一方面降低特征维数，另一方面提高模型训练和匹配的精度。

(3)模型训练模块

主要包括模型结构的表示和模型参数的估计两方面内容。这一部分将在后面的内容中详细叙述。

(4)模式匹配模块

密数据识别方法与密数据的模型结构相互对应。目前，可用于密数据识别的方法有很多，主要的识别方法包括动态时间规整、矢量量化、隐马尔可夫模型、高斯混合模型、人工神经网络、支持向量机等。

(5)判决模块

将提取的密数据的特征参数与参考模型进行匹配，根据特定的相似性准则来计算结果，最终由系统判决密数据是明文还是密文，或者是哪种加密方式产生的密文。

3.1　特征提取

首先获取网络数据，并提取网络层数据。提取网络层数据的方法有：根据网络数据来源，即传输链路的类型，判断链路层协议；根据链路层协议处理网络数据，丢弃与链路层以上协议无关的内容，提取出网络层数据。如果网络层数据为非 IP 协议报文，则丢弃；如果网络层数据为 IP 协议报文，则去除 IP 协议报文头部，然后将该 IP 协议报文保留下来做进一步处理。根据源 IP 地址、目的 IP 地址、源端口、目的端口和协议汇聚数据流，将一系列具有相同源地址、相同的目的地址、相同的源端口、相同的目的端口和相同的协议的 IP 协议报文组成一个数据流。从数据流中提取有效数据，如果 IP 协

议报文的载荷是 TCP 协议报文或 UDP 协议报文，则去除 TCP 协议报文或 UDP 协议报文的头部，剩余数据即为有效数据；如果 IP 协议报文的载荷不是 TCP 协议报文或 UDP 协议报文，则直接将 IP 协议报文的载荷作为有效数据。

对有效数据进行 NIST 随机性检验，利用 NIST 一共可提取 15 类随机性特征，共 188 维特征，分别是：

1)近似熵检验(可提取 1 维特征，特征代码 1)

2)分组组内频数检验(可提取 1 维特征，特征代码 2)

3)累加和检验(可提取 2 维特征，特征代码 3 和 4)

4)离散傅里叶变换检验(可提取 1 维特征，特征代码 5)

5)单比特频数检验(可提取 1 维特征，特征代码 6)

6)线性复杂度检验(可提取 1 维特征，特征代码 7)

7)组内最长游程检验(可提取 1 维特征，特征代码 8)

8)非重叠模式匹配检验(可提取 148 维特征，特征代码 9-156)

9)重叠模式匹配检验(可提取 1 维特征，特征代码 157)

10)随机游动检验(可提取 8 维特征，特征代码 158-165)

11)随机游走变体检验(可提取 18 维特征，特征代码 166-183)

12)二进制矩阵秩检验(可提取 1 维特征，特征代码 184)

13)游程检验(可提取 1 维特征，特征代码 185)

14)串行检验(可提取 2 维特征，特征代码 186-187)

15)Maurer 通用统计检验(可提取 1 维特征，特征代码 188)

3.2 极速学习机

为了验证 NIST 随机性测试数据用于加密数据分类的有效性，我们采用极速学习机对几类加密数据的随机性特征建模。

极速学习机(Extreme Learning Machine，ELM)是 Huang G B[4]于 2004 年正式提出的训练前馈型神经网络问题的快速算法，并在文献[5]中给出了详细的理论和实验验证。最新 ELM 综合文献[6-7]总结前几年的理论成果，指出极速学习机与支持向量机具有相同的优化目标，但前者的约束条件更为宽松，且训练效率和泛化能力都优于后者。

Huang 严格证明了可以随机选择单隐藏层前馈神经网络的输入层到隐藏层的权重和隐藏层节点偏差，并在此基础上提出了极速学习机理论。图 2 是一个典型的 3 层网络结构。一次学习的过程是，随机选择输入层到隐藏层的权重和隐藏层节点偏差，隐藏层到输出层的权重由训练样本和上述参数共同决定。

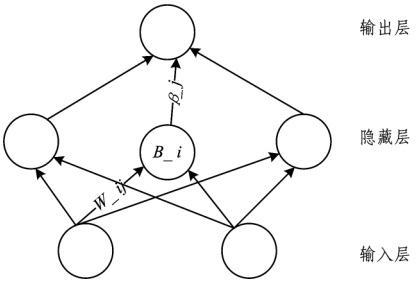

图 2 单隐藏层前馈型神经网络

对于 N 个样本(x_i,t_i)，$x_i=[x_{i1},x_{i2},\cdots,x_{im}]^{\mathrm{T}}\in R^n$，$t_i=[t_{i1},t_{i2},\cdots,t_{im}]^{\mathrm{T}}\in R^m$，一个拥有$\widetilde{N}$个隐藏层节点和激励函数为 $g(x)$的 SLFN(Single-hidden Layer Feed-forward Neural Network)可以表示为：

$$\sum_{i=1}^{\widetilde{N}}\beta_i g_i(x)=\sum_{i=1}^{\widetilde{N}}\beta_i g(w_i x_j+b_i)=o_j,\ j=1,\cdots,N \tag{1}$$

其中，$w_i=[w_{i1},x_{i2},\cdots,w_{im}]^{\mathrm{T}}$ 为输入层到第 i 个隐藏层节点的权重，$\beta_i=[\beta_{i1},\beta_{i2},\cdots,\beta_{im}]^{\mathrm{T}}$ 为第 i 个隐藏层节点到输出层的权重，b_i 是第 i 个隐藏层节点的偏差。o_j 为第 j 个样本的输出。$g(x)$可以采用 Sigmoid 或 RBF(Radial Basis Function)等形式。训练误差是输出值和实际值的误差，用$\sum_{j=1}^{N}\|o_j-t_j\|$来表示，完全拟合即误差为 0 时，式(1)可以改写成下列形式：

$$H\beta=T \tag{2}$$

其中，

$$H=(w_1,\cdots,w_{\widetilde{N}},b_1,\cdots,b_{\widetilde{N}},x_1,\cdots,x_N)=\begin{bmatrix}g(w_1\cdot x_1+b_1) & \cdots & g(w_{\widetilde{N}}\cdot x_{\widetilde{N}}+b_{\widetilde{N}})\\ \vdots & \cdots & \vdots\\ g(w_1\cdot x_N+b_1) & \cdots & g(w_{\widetilde{N}}\cdot x_N+b_{\widetilde{N}})\end{bmatrix}_{N\times\widetilde{N}}$$

$$\beta=\begin{bmatrix}\beta_1^{\mathrm{T}}\\ \vdots\\ \beta_{\widetilde{N}}^{\mathrm{T}}\end{bmatrix}_{\widetilde{N}\times m},\ T=\begin{bmatrix}t_1^{\mathrm{T}}\\ \vdots\\ t_N^{\mathrm{T}}\end{bmatrix}_{N\times m} \tag{3}$$

H 称为隐藏层输出矩阵，它的第 i 列代表第 i 个隐藏层节点的输出。现权重 w 和偏差 b 已经通过随机生成的方法得到，目标是找到满足式(2)的权重 β。通常 $\widetilde{N}\ll N$，不定方程(2)没有满足要求的解，根据误差最小原则，即 $\min_{\beta}\|H(w_1,\cdots,w_{\widetilde{N}},b_1,\cdots,b_{\widetilde{N}},x_1,\cdots,x_N)\hat{\beta}-T\|$，由广义逆矩阵的相关理论可求得误差最小且范数最小的解：

$$\hat{\beta}=H^{\dagger}\ T \tag{4}$$

其中矩阵 $H^{\dagger}$ 是矩阵 H 的广义逆矩阵，可以通过奇异值分解等方法得到。也可以用如下公式：

如果 $H^{\mathrm{T}}H$ 是非奇异的，

$$H^{\dagger}=(H^{\mathrm{T}}H)^{-1}H^{\mathrm{T}} \tag{5}$$

或者如果 HH^{T} 是非奇异的，

$$H^{\dagger}=H^{\mathrm{T}}(HH^{\mathrm{T}})^{-1} \tag{6}$$

如果上述两个矩阵是奇异的，根据岭回归理论，原矩阵对角线加上一个极小正值可以增加求解几率和解的稳定性。则式(5)、式(6)变为：

$$H^{\dagger}=(\frac{I}{\lambda}+H^{\mathrm{T}}H)^{-1}H^{\mathrm{T}} \tag{7}$$

或者

$$H^{\dagger}=H^{\mathrm{T}}(\frac{I}{\lambda}+HH^{\mathrm{T}})^{-1} \tag{8}$$

Huang 指出通过式(7)和式(8)联合式(4)解答得到的 β，与下面构造的正则化式的求解是一致的。

$$\min_{\beta}\|H\beta-T\|^2+\lambda\|\beta\|^2 \tag{9}$$

至此，完成一次学习的过程。

4 识别结果和分析

为了检测本文提出的基于模式识别的密数据检测方法的性能，本节进行了密数据识别实验。密数据选用的是通过加密代理软件上网时抓取的密数据流，同时也抓取了同等数量的正常上网数据流，分别从中选取 1000 段，每段长度均为 10K 字节。训练样本为随机的 500 段数据；识别时每种类型的数据流为另外的 500 段，共 1000 段。

数据特征包括 15 种 NIST 随机性测试数值，测试的内容

分别如表 1 所列。NIST 随机性测试数值 pvalue 在(0，1)之间，用于表示该数据的随机性优劣，传统的方法一般是设定一个阈值，大于该阈值的表示数据具有随机性，比如设定阈值为 0.1，当某段数据的随机性测试 pvalue 值小于 0.1 时，表明该数据段的随机性较差，因此可以认为是非加密数据；如果某段数据的随机性测试 pvalue 值大于 0.1 时，表明该数据段的随机性较好，因此该数据是加密数据的可能性较大。利用这一点进行密数据检测最大的难点是如何确定判决阈值，另外有些加密软件的数据的随机性非常差，pvalue 值几乎与明文数据相当，因此无法找到这样一个判决阈值，表 2 是 4 类数据(密数据 1、明文数据 1、密数据 2 和明文数据 2)的非重叠匹配测试的 pvalue 值的分布。

表 2　4 类数据非重叠匹配测试的 pvalue 值的分布

非重叠匹配测试	密数据 1	明文数据 1	密数据 2	明文数据 2
[0.9,1)	22042	14465	3093	10718
[0.8,0.9)	18466	13304	2848	10241
[0.7,0.8)	22087	17095	3262	12247
[0.6,0.7)	22696	18076	3437	13066
[0.5,0.6)	21399	17550	3237	12889
[0.4,0.5)	21139	17631	3352	13619
[0.3,0.4)	20786	18379	3454	15029
[0.2,0.3)	19600	18261	3413	16100
[0.1,0.2)	20086	19912	4157	20143
(0.0,0.1)	26447	55327	179747	85948

由表 2 可以看出只有密数据 1 的 90%能正确通过非重叠匹配随机性测试，而密数据 2 虽然是加密数据，但其随机性测试结果的 pvalue 值分布大部分在(0，0.1)之间，如果以 0.1 作为判决阈值，90%的数据都不能通过随机性测试，也即 90%的数据会判断为正常数据，而明文数据 1 和明文数据 2 的数据却有 60%到 80%的数据能通过随机性测试，会判断为密数据，总体错分率达到 60%以上，因此直接利用 NIST 随机性测试的结果进行明密数据的判别是不可行的。表 3 和表 4 也说明了这一问题。

表 3　4 类数据离散傅里叶变换测试的 pvalue 值分布

离散傅里叶变换测试	密数据 1	明文数据 1	密数据 2	明文数据 2
[0.9,1)	21928	11229	4629	18361
[0.8,0.9)	18643	10502	4173	16705
[0.7,0.8)	22127	14080	6119	20341
[0.6,0.7)	22631	15110	7614	21283
[0.5,0.6)	21493	15044	8424	20000
[0.4,0.5)	21070	15915	10384	19821
[0.3,0.4)	20949	17736	13439	20108
[0.2,0.3)	19738	18486	20345	19056
[0.1,0.2)	19860	21818	49246	20012
(0.0,0.1)	26309	70080	85627	34313

表 4　4 类数据游程测试的 pvalue 值分布

游程测试	密数据 1	明文数据 1	密数据 2	明文数据 2
[0.9,1)	21976	16978	14391	21437
[0.8,0.9)	18530	15104	12417	18430
[0.7,0.8)	22264	20528	15915	22422
[0.6,0.7)	22746	21693	19361	23306
[0.5,0.6)	21149	20307	15628	21288
[0.4,0.5)	20997	20823	13768	20566
[0.3,0.4)	20878	21590	12121	20483
[0.2,0.3)	19365	21166	10149	18771
[0.1,0.2)	19950	21945	11101	18893
(0.0,0.1)	26893	29866	85149	24404

由表 3 可以看出，4 类数据的离散傅里叶变换测试结果与非重叠匹配测试结果相比变化很大，明文数据 2 与密数据 1 的 pvalue 值的分布特点相似，密数据 2 与明文数据 1 的 pvalue值的分布特点相似，简单设定阈值进行明密数据判断也是不可行的。由表 4 可以看出，4 类数据的游程测试结果与前两种测试结果又有很大不同：密数据 1、明文数据 1 与明文数据 2 的 pvalue 值的分布特点相似，而密数据 2 的 pvalue 值的分布有所不同。但我们能发现，从数值上看，密数据 2 在多种随机性测试中，其 pvalue 的分布有其固有特点。

综上所述，尽管有很多随机性测试方法，但是错分率都在 60%以上，因此没有一种方法可以用于明密数据的检测，而且在每种测试方法下，4 类数据的 pvalue 值的分布也不尽相同，因此我们考虑综合利用这些数据进行明密数据识别。

本文将以上数据随机性测试的 pvalue 值作为数据特征，利用模式识别的方法自动挖掘出不同数据的明密特点，并构建相应的模型，再利用这些模型进行明密数据识别，大大提升了明密数据的识别率。首先不对 15 类特征做任何处理，直接用于训练分类器，并测试基于 ELM 的密数据检测性能，实验结果如表 5 所列。

表 5　基于 ELM 的密数据检测结果(15 维特征)

数据流长度/K	测试次数	密数据错识率	明文数据错识率	平均错识率
10	1000	0.34	0.32	0.33
20	500	0.31	0.29	0.30
50	200	0.28	0.28	0.28
100	100	0.28	0.30	0.29

由表 5 可以看出，虽然直接用 15 类特征进行检测，错识率已经大大下降，但仍然有 30%左右，大部分明文数据都被误检测为加密数据，这说明并不是所有的随机性测试 pvalue 值都对区分明密数据有贡献，因此本文利用滤波器方法对特征进行优化[8-9]，比如利用主成分分析(PCA)取得了更优的分类性能，不仅可以区分明密数据，对数据的来源也能够准确识别，平均错识率下降到 15%左右，实验结果如表 6 所列。

表 6　基于 PCA 的密数据检测结果/%

	测试数据集 1	测试数据集 2	测试数据集 3	测试数据集 4
密数据 1	10	16.29	10.77	10
密数据 2	12.31	10.67	15.39	19.72
明文 1	12.31	11.72	10	10.58
明文 2	19.24	18	15.39	24.43
平均	13.47	14.17	12.89	16.18

结束语　本文提出了基于模式识别的密数据检测方法，利用 NIST 提供的 15 类数据随机性检测方法对数据进行随机性测试，将测试得分作为密数据特征训练分类器，进而进行明密数据的识别和数据类型的识别，在分类器 ELM 下进行测试，明密数据错分率由原来的 60%以上下降到 30%左右；利用滤波器方法对特征进行优化降维，进一步提升了密数据的分类性能，能够对密数据的来源进行识别，平均错分率进一步下降到 15%左右。

参 考 文 献

[1] SOTO J. Statistical Testing of Random Number Generators [OL]. http://www.nist.gov.

[2] RUKHIN A，SOTO J. A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applica-

tions [DB/OL]. http://www.nist.gov.

[3] SOTO J. Randomness Testing of the AES Candidate Algorithms [OLL]. http://www.nist.gov.

[4] HUANG G B, ZHU Q Y, SIEW C K. Extreme Learning Machine: A New Learning Scheme of Feedforward Neural Networks[C]// Proceedings of International Joint Conference on Neural Networks (IJCNN2004). Budapest, Hungary, July, 2004: 25-29.

[5] HUANG G B, ZHU Q Y, SIEW C K. Extreme learning machine: Theory and applications[J]. Neurocomputing, 2006, 70(1-3): 489-501.

[6] HUANG G B, ZHOU H, DING X, et al. Extreme learning machine for regression and multiclass classification [J]. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, 2012, 42(2): 513-529.

[7] MICHE Y, SORJAMANN A, BAS P, et al. OP-ELM: optimally pruned extreme learning machine [J]. IEEE Transactions on Neural Networks, 2010, 21(1): 158-162.

[8] KIM M S, YU H J, KWAK K C, et al. Robust text-independent speaker identification using hybrid PCA & LDA[C]// MICAI'06. Mexico, Nov. 2006: 1067-1074.

[9] LIU C. Gabor-based kernel PCA with fractional power polynomial models for face recognition[J]. IEEE Trans, Pattern Analysis and Machine Intelligence, 2004, 26(5): 573-581.

(上接第 360 页)

本文的加密算法分为两步：置乱和混淆。在置乱阶段混沌映射的初值与明文图像的像素值的总和有关；在混淆阶段，对下一块加密时，混沌映射需迭代 $H+70$ 次后，取后 70 次的混沌序列作为所需的密钥流，而 H 是与上一块的密文相关的。基于以上两点，加密不同的图像，加密系统所产生的置乱阶段的密钥流和混淆阶段的密钥流都不同，所以加密一些特殊的图像并不能得到目标密文的密钥流，因而本文提出的加密算法能抵抗选择明(密)文攻击。

结束语 本文提出了一种参数扰动下混沌的图像加密方案，与已有的加密算法相比，该算法具有以下特点：1)所用混沌系统是参数扰动下的混沌系统，且扰动参数是另一混沌系统所产生的随机序列，克服了由于计算机有限数字精度的限制，混沌序列将退化为周期序列的缺陷。2)已有的加密算法都是先用混沌系统生成密钥流，然后用密钥流与明文图像做混淆或扩散操作，而本文的算法是根据密文动态产生密钥流，这样加密不同的明文图像所用密钥流不同，达到“一次一密”的效果。实验模拟和安全性分析表明，该算法具有对密钥敏感，密钥空间大，密文图像统计特性良好，密文对明文敏感，抵抗选择明(密)文的攻击及抗噪声、抗压缩、抗剪切攻击的优点。

参 考 文 献

[1] AHMAD J, HWANG S O. A Secure Image Encryption Scheme Based on Chaotic Maps and Affine Transformation[J]. Multimedia Tools Applications, 2015, 75(21): 1-26.

[2] BAPTISTA M. Cryptography with Chaos[J]. Physics Letters A, 1998, 240(1): 50-54.

[3] CHEN G, MAO Y, CHUI C K. A symmetric image encryption scheme based on 3d chaotic cat maps[J]. Chaos Solitons Fractals, 2004, 21(3): 749-761.

[4] LI S, ZHENG X. Cryptanalysis of a Chaotic Image Encryption Method[C]// IEEE International Symposium on Circuits and Systems, 2002(ISCAS'2002), EEE, 2002: 708-711.

[5] WANG Y, WONG K W, LIAO X, et al. A new chaos-based fast image encryption algorithm[J]. Applied Soft Computing, 2011, 11(1): 514-522.

[6] 刘泉，李佩玥，章明朝，等. 基于可 Markov 分割混沌系统的图像加密算法[J]. 电子与信息学报，2014，36(6)：1271-1277.

[7] 置与比特双重置乱的图像混沌加密算法[J]. 通信学报，2014，35(3)：216-223.

[8] 张顺，高铁杠. 基于类 DNA 编码分组与替换的加密方案[J]. 电子与信息学报，2015，37(1)：150-157.

[9] 文昌辞，王沁，黄付敏，等. 基于仿射和复合混沌的图像自适应加密算法[J]. 通信学报，2012，33(11)：119-127.

[10] 李树钧. 数字化混沌密码的分析与设计[D]. 西安：西安交通大学，2003.

[11] ALVAREZ G, LI S J. Some Basic Cryptographic Requirements for Chaos-based Cryptosystems[J]. International Journal of Bifurcation and Chaos, 2006, 16(8): 2129-2151.

[12] KOCAREV L. Chaos-based Cryptography: a Brief Overview[J]. IEEE Circuits and Systems Magazine, 2001, 1(3): 6-21.

[13] 罗启彬，张健. 一种新的混沌伪随机序列生成方式[J]. 电子与信息学报，2006，28(7)：1262-1265.

[14] 韩双霜，闵乐泉，韩丹丹. 一种基于三维离散混沌映射的伪随机数生成器[J]. 华中科技大学学报(自然科学版)，2013，41(8)：16-19.

[15] WANG X, LIU L T. Cryptanalysis of a Parallel Sub-Image Encryption Method with High-Dimensional Chaos[J]. Nonlinear Dynamics, 2013, 73(73): 795-800.

[16] LI C Q, ZHANG L Y, OU R, et al. Breaking a Novel Colour Image Encryption Algorithm Based on Chaos[J]. Nonlinear Dynamics, 2012, 70(4): 2383-2388.

[17] ZHU C, LIAO C L, DENG X. Breaking and Improving an Image Encryption Scheme Based on Total Shuffling Scheme[J]. Nonlinear Dynamics, 2013, 71(1/2): 25-34.

[18] 朱从旭，卢庆. 对结合超混沌序列和移位运算图像密码的攻击[J]. 山东大学学报(理学版)，2016，51(6)：67-71.

[19] ZHU C. A Novel Image Encryption Scheme Based on Improved Hyperchaotic Sequences[J]. Optics Communications, 2012, 285(1): 29-37.

[20] 廖琪男，卢守东，孙宪波. 结合超混沌序列和移位密码的数字图像加密算法[J]. 小型微型计算机系统，2015，36(2)：332-337.

[21] SHI Y M, CHENG G R. Discrete Chaos in Banach Spaces[J]. Science in China Series A: Mathematics, 2005, 48(2): 222-238.

[22] LIAN S. Efficient Image or Video Encryption Based on Spatiotemporal Chaos System[J]. Chaos, Solitons & Fractals, 2009, 40(5): 2509-2519.

[23] BENIA S, AKHSHANI A, MAHMODI H, et al. A Novel Algorithm for Image Encryption Based on Mixture of Chaotic Maps [J]. Chaos, Solitons & Fractals, 2008, 35(2): 408-419.

[24] WANG X Y, TENG L, QIN X. A novel colour image encryption algorithm based on chaos[J]. Signal Processing, 2012, 92(4): 1101-1108.