计算机工程
Computer Engineering
ISSN 1000-3428,CN 31-1289/TP

《计算机工程》网络首发论文

题目：基于决策树的 SM4 分组密码工作模式识别
作者：纪文桃，李媛媛，秦宝东
DOI：10.19678/j.issn.1000-3428.0058608
网络首发日期：2020-08-15
引用格式：纪文桃，李媛媛，秦宝东．基于决策树的 SM4 分组密码工作模式识别．计算机工程. https://doi.org/10.19678/j.issn.1000-3428.0058608

网络首发时间：2020-08-15 10:30:03
网络首发地址：https://kns.cnki.net/kcms/detail/31.1289.TP.20200814.1416.001.html

计 算 机 工 程

基于决策树的 SM4 分组密码工作模式识别

纪文桃[1]，李媛媛[1]，秦宝东[1,2]
（1. 西安邮电大学网络空间安全学院，陕西 西安 710121；2. 西安邮电大学无线网络安全技术国家工程实验室，陕西 西安 710121）

摘　要：密码体制识别是密码分析的重要组成部分。现有工作主要针对多种混合密码算法进行识别，而缺乏对分组密码工作模式识别的研究。与应用场景较广、优化能力较强的机器学习相结合，提出了一种基于决策树的分组密码工作模式识别方案，并将其应用于国密 SM4 分组密码算法当中。在训练和测试阶段提出了三种分类模型，即混合分类模型、混合文本大小分类模型和一对一分类模型。实验结果表明：前两种模型的分类结果约为 25%左右，而一对一分类模型中 CBC、CFB、OFB、CTR 之间的正确率高达 90%以上。

关键词：工作模式识别；决策树；SM4 算法；特征提取；工作模式

开放科学标识码(OSID)：

Recognition of SM4 Block-Cipher Modes of Operation Based on Decision Tree

JI Wen-tao[1], LI Yuan-yuan[1], QIN Bao-dong[1,2]
(1. School of Cyberspace Security , Xi'an University of Posts & Communications, Xi'an 710121, China;
2. National Engineering Laboratory for Wireless Security , Xi'an University of Posts & Telecommunications, Xi'an 710121, China)

Abstract: Cryptosystem recognition is an important part of cryptanalysis. Known works mainly focus on the identification of different encryption algorithms, but very few works research on the identification of block-cipher modes of operation. In this paper, we propose a scheme for identification of block-cipher modes of operation based on a popular machine learning algorithm, i.e., decision tree, and apply it to the SM4 block-cipher algorithm. In the training and testing phases, three classification models are proposed, namely hybrid classification model, mixed text-size classification model and one-to-one classification model. The experimental results show that the classification results of the first two models are about 25%, and in the one-to-one classification model, the correct rate among CBC, CFB, OFB, and CTR is more than 90%.

Key words: Recognition of operation modes,decision tree, SM4 algorithm, feature extraction, modes of operation
DOI:10.19678/j.issn.1000-3428.0058608

基金项目：国家自然科学基金资助项目(61872292)；青海省基础研究计划项目（2020-ZJ-701）.

作者简介：（1994）纪文桃，男，硕士研究生，基于机器学习的密码体制识别，1043238267@qq.com；李媛媛，硕士研究生；秦宝东，教授。**E-mail**：1043238267@qq.com

0 概述

密码学包括两个重要的分支：密码编码学和密码分析学。在信息时代，网络安全已成为国家安全的重要组成部分，而密码编码学和密码分析学在其中扮演着不可或缺的角色。目前，网络空间不仅存在众多的数据类型，例如视频、文本、图像等，而且数据量大、冗余度也高。为此，需要研究合理的数据处理方案。机器学习与密码学的结合为处理大量密文数据提供这种可能。机器学习提供了聚类分类算法来处理和分析数据，而密码学为越来越被重视的数据安全保驾护航。

密码编码学是一种保护信息在传递过程中不被第三方或者敌方所解读、利用和窃取的技术，它解决的主要问题是信息的安全性问题。在现实生活中，总会存在一些用户在未经信息持有者授权或者本就无意愿将信息共享的情况下，对所传输的信息进行非法获取、恶意篡改、以及删除和伪造。密码学者对此采取的措施主要是使用密码分析技术来进一步完善所设计的密码方案并优化方案中所包含的密码算法。目前较为熟悉的密码分析技术都是基于 Kerckhoffs[1]这一原则，即在进行密码分析时分析者是知道具体密码算法的。根据可用于密码分析的信息，密码分析攻击的手段可分为以下几类：唯密文攻击、已知明文攻击、选择明文攻击、选择密文攻击和侧信道攻击等。在现实情况下，密文数据是密码分析者唯一容易获得的信息。已知的密码分析技术大部分是基于某种具体密码算法或者在已知密文所使用的密码算法的前提下设计的。因此，密码分析的首要任务是对获取的密文数据所使用的密码算法进行识别。密码算法识别不仅是进一步开展密码分析的基础，也是发现密码算法是否存在安全隐患的一种重要方式，对增强密码算法的安全性具有重要的影响。

1 相关工作

基于密文特征的密码体制识别往往都同机器学习技术相得益彰[2]。2011 年，Manjula 和 Anitha[3]提出了基于 C4.5 决策树的密码体制识别方案。该方案提取了八种密文特征，对十一种加密算法进行识别，包括古典算法、分组密码算法和非对称加密算法，并从训练文件数量和所用密文文件大小两方面进行试验，得到的识别率为 70%-75%。2012 年，Chou 等人[4]提出利用支持向量机对密码体制进行识别。通过对数据集的分析发现密码的工作模式主导着分类任务的执行。该方案提取了 12 种密文特征，对 AES（高级加密标准，advance encryption standard）和 DES（数据加密标准，data encryption standard）分别在 ECB（电码本，electronic codebook）模式和 CBC（密码分组链接，cipher-block chaining）模式下进行算法识别。实验结果表明，当对每个明文使用 CBC 模式和随机初始向量时，性能较差，而使用 ECB 模式时，某些数据集的性能相对较好。2013 年，Mishra 等人[5]提出并实现了将模式识别和决策树结合从而识别分组密码和流密码的方案。该方案包括三个子模块技术：块长度/流检测、熵/重现分析、基于字典/决策树的方法。通过这三个部分的实现大大提高了密码算法的识别率。2013 年，Willam 等人[6]提出基于神经网络的区分攻击方法。它利用语言学和信息检索方法，从 MARS、RC6、Rijndael、Serpent、和 Twofish 加密的密文中生成分类模型；然后将获取的密文集合提交到“聚类过程”，并将结果输入到分类器中，从而得到以上 5 种加密算法的分类结果。2014 年，Lomte 等人[7]对 Willam 所提出的方案进一步研究。后者在聚类识别过程中使用单一的密钥进行加密的，

而前者在训练和测试过程中设置了不同的密钥，相比较而言，识别率有所下降。2015 年，吴杨和王韬等人[8][9]提出了基于 K-means（k 均值聚类，k-means clustering algorithm）的密码体制分层识别方案，对五种分组密码（AES、Camellia、DES、3DES、SMS4）进行两两识别。在特征提取部分使用了密文随机性度量值的方法，大大提高了典型分组密码的识别率，高达 90%左右。2016 年，Mello 等人[10]在 ECB 模式下，对七种不同语言编写的纯文本文件使用七种密码算法进行编码。这些文件提供了六种数据挖掘算法的信息，以用于识别文本加密的算法。通过大量的元数据和大量的耗时计算，得到了非常高的识别率。2016 年，ChengTan 等人[11]提出一种基于支持向量机的密码体制识别方案，对五种常见的分组密码算法进行识别，分别为 AES，Blowfish，3DES，RC5 和 DES。对这五种算法在四种不同的情况下进行试验，即训练和测试密文的密钥是否相同，在此基础上，用其他四种加密算法与 AES 进行一一识别，当训练和测试阶段的密钥相同时，它的识别率较高。2017 年，Barbosa 等人[12]提出对加密的多媒体文件进行密码算法识别。该过程使用四种加密算法对音频和视频文件进行加密，然后将加密的文件提交给数据挖掘算法，并将其产生的混淆矩阵编译成图表。2017 年，黄良韬等人[13]提出基于随机森林的密码体制分层识别方案，并介绍了三种簇分方式：CM-簇分、CSN-簇分和 CSBP-簇分。首先，将已加密的文件按照大的分类进行归类，然后再从每一类中区分出具体的密码算法。结果表明，加入分层的方案较单分的密码体制识别效果更佳。在以往的密码体制识别中，大多的密码算法都是在 ECB 模式下的，2018 年，Tan 等人[14]则提出基于 CBC 模式的密码体制识别方案。该方案用五种算法进行多类识别和一对一识别，在此基础上考虑了训练和测试时密钥是否相同以及初始向量是否相同的情况。其中一对一识别是将 AES 与其他四种算法进行识别。2019 年，赵志诚等人[15]采用随机性测试进行密文特征的提取，提出了基于随机森林的识别方案。该方案中对六种分组密码进行两两识别，在不同的特征下其识别率各有差异，部分特征下其识别率能达到 80%以上。

上述工作主要针对不同密码算法进行识别，或者在特定工作模式下进行识别，而缺乏对分组密码工作模式的识别，特别是国密 SM4 分组密码算法的工作模式识别。分组密码的工作模式又对密码算法的识别起着主导作用，这将对舆论分析、互联网审查、电子取证和网络监控具有推进作用。无论是对不同算法识别还是对分组密码工作模式的识别，都利用机器学习将其视为模式分类的问题，使用机器学习的各种分类方法来尝试捕获加密后的密文文件中隐含的行为。通过机器学习算法对大量密文文本进行分析计算，生成对应的分类模型，再将测试文本投入分类器中，进行比较，最后得到分类结果。这样就可以避免人为的去分析论证密码工作模式本身的区别，而是将这一行为交由机器学习去完成，通过与统计学方法的结合，智能的挖掘和分析在同一明文文本集下经由不同工作模式加密后产生的密文文本之间所隐含的不同信息。本文主要研究 SM4 分组密码的工作模式进行识别方法。SM4 算法是我国国家密码管理局提出的一种分组长度和密钥长度均为 16 字节的分组密码算法，适用于无线局域网产品。它的安全性能较高，可以抵抗差分、线性和代数等分析技术。SM4 算法现已实现的工作模式有六种，在本文中用到了 CBC（密文分组链接模式）、CFB（密文反馈模式）、CTR（计数器模式）和 OFB（输出反馈模式）四种。若直接使用 SM4 算法加密消息，它的安全性是非常脆弱的。为了在不同的场景中更好地保护明文的安全性，需要借助不同的工作模式。

对工作模式的识别能够提高密码算法的安全性，从而更好地抵抗密码攻击。

本文提出了基于 C4.5 的分组密码工作模式识别方案并对国密 SM4 算法的四种工作模式进行识别。该方案首先利用加密工具在不同的工作模式下对大量的文本文件进行加密，得到密文文件。然后构造训练阶段和测试阶段所需的特征向量空间，该空间是由特征提取算法对密文文件处理后得到的，每个特征向量中包括五个值，即大写字母数量，小写字母数量，数字字符数量，其他字符数量，标签值。在此基础上，训练阶段通过对特征空间的学习生成决策树。测试阶段根据生成的决策树进行决策，最后将决策值与标签值相比较得到分类结果。实验结果表明，在一对一分类模型中，CBC、CFB、OFB 和 CTR 之间的识别正确率达到 90%以上。上述实验结果表明国密 SM4 算法的工作模式具有一定的可识性。为了增强 SM4 算法在应用中的安全性，需要进一步研究其工作模式的隐藏方法。

2 系统模型

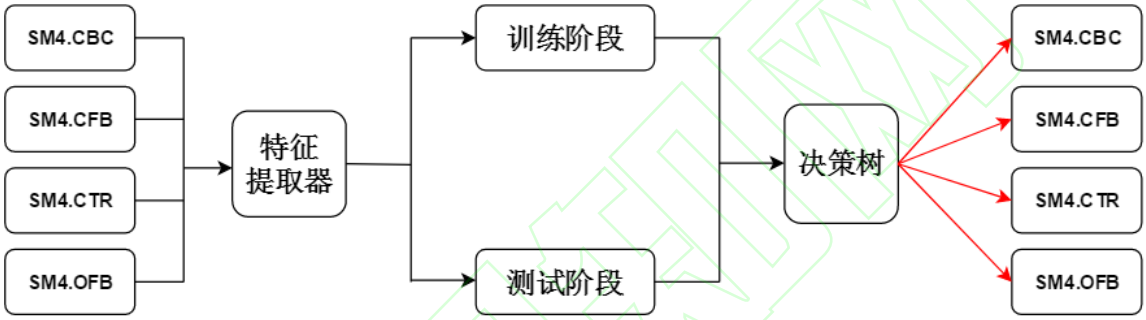

图 1：分组密码工作模式识别模型

Fig.1 Block cipher working pattern recognition model

如图 1 所示，本文设计了基于决策树的密码体制识别系统。整个识别系统包括四个部分：1)关于特定工作模式的密文文件；2)密文特征提取；3）生成决策树模型；4）测试分类。

分组密码的工作模式[16]包括电子密码本（ECB）模式、密文分组链接模式（CBC）、密文反馈模式（CFB）、输出反馈模式（OFB）和计数器模式（CTR）。在该系统模型中，使用了国密 SM4 算法的四种工作模式，电子密码本模式除外。CBC 模式相比较其他四种工作模式较为流行，除第一个明文分组与随机产生的初始向量[17]异或后加密生成密文外，其他明文分组都与前一个密文输出进行异或后生成相应的密文。在信息传输过程中，只要一个明文分组出现错误，则会影响其后的密文分组。因此它的加密过程不能并行化，相反解密过程不受该情况的影响，从而可实现并行计算。CFB 模式首先对随机产生的初始化向量加密，将加密后的结果与明文分组相异或，而后续的过程是对前一个分组加密后的结果进行加密，然后与当前明文相异或，得到对应的密文。与 CBC 模式相比，CFB 模式的错误传播性质更为突出，一个明文分组的错误可能会导致其后所有密文分组出现错误。不仅仅是加密，其解密过程也依赖其他的明文分组，故其加密过程和解密过程均不能实现并行化。OFB 模式的第一步操作与 CFB 相同，不同的是 OFB 把随机初始向量加密后的结果作为下一分组的输入，对该输入加密后再与明文分组进行异或运算得到对应的密文分组。OFB 模式规避了 CFB 模式和 CBC 模式由于明文出错而带来的错误传播的风险，但随之而来的是一旦密文被恶意篡改，

检测工作将变得不再容易。CTR 模式是将计数器进行分组，将每个计数器分组加密后与明文分组进行异或运算，得到对应的密文分组。它每个分组的加密与解密过程相互独立，不会有错误传输的困扰，也因此加解密均可并行。

将明文加密成特定密码工作模式的密文文件后，再对其进行特征提取。特征提取的工作是最重要也是最难的一步，因为其本质是对密数据的处理和分析，而且它对后续模型的生成和分类的结果起着决定性的作用。特征提取的方法很多，如统计学方法、随机性检测[18][19]以及密码学中的熵特性和信号中的频域特征等。在本文中，我们借助统计学方法和信息熵的结合来实现密文数据的特征提取。首先，利用统计学方法对密文文本中的大小写字符、数字字符以及特殊字符的个数进行统计，然后分别计算其对应的熵，以此为基础形成特征向量，进而得到特征向量空间。

对密文数据特征提取后，就进入学习和分类的过程了，这是机器学习算法的本质。机器学习算法包括两个阶段，第一阶段是训练阶段，即将得到向量空间的一部分传递给决策树算法进行学习，并生成分类模型，也称决策树模型。第二阶段为测试阶段，即将向量空间的另一部分投入到机器学习算法中进行测试，依据训练阶段生成的决策树模型来进行分类。训练阶段和测试阶段的向量空间中都包含密文的特征，而决策树算法是有监督学习算法，它的学习标签包含在训练阶段，而在测试阶段将标签隐藏，待测试完成后，用决策树模型预测的结果与测试数据所隐藏的标签进行对比，就可得到实验的结果。

在得到结果之前，要借助决策树算法[20]来生成模型以及进行分类。决策树算法包括 ID3 和 C4.5 算法，ID3 算法主要借助信息熵和信息增益作为测试属性的衡量标准，而 C4.5 是在 ID3 的基础上对信息增益加以调节得到分裂信息，用分裂信息和信息增益率来划分属性。ID3 在对属性的划分中往往偏向于选择样本较多的，而 C4.5 采用信息增益率规避了这一点。并且 C4.5 算法在树的构造过程中，不仅能够将连续数据进行离散化处理，而且还可以对树进行剪枝处理，从而得到更优的树。本文中选用的是 C4.5，其具体操作如下：

第 1 步：假设有包含 N 种属性的数据集 S 和包含 K 种属性的子集合 A。令p_i表示每个目标属性的概率，$|S_i|$表示子集S_i的样本数，$|S|$表示数据集S的样本数。将数据集按照每一个属性进行划分，并计算其对应的信息熵、分裂信息和信息增益率，如下

信息熵：$H(s)=-\sum_{i=1}^{N} p_i \log_2 p_i$；$H_A(S)=\sum_{i=1}^{k}\frac{|S_i|}{|S|}H(S_i)$

分裂信息：$SplitE\ (A)=-\sum_{i=1}^{k}\frac{|S_i|}{|S|}\log_2\frac{|S_i|}{|S|}$

信息增益：$IG(S,A)=H(S)-H_A(S)$

信息增益率：$IGRatio\ (A)=\frac{IG(S,A)}{SplitE\ (A)}$

第 2 步：将上一步中计算出的每一个属性的信息增益率进行比较，选择最大的一个作为决策树的节点。

第 3 步：在子节点上利用剩余的属性继续执行第 1 步和第 2 步，直到该节点为纯叶子结点。

第 4 步：对生成的决策树进行剪枝处理，包括先剪枝和后剪枝两种方法，以防止过拟合。

该算法结合密文特征提取的结果得到上文描述的信息熵、信息增益和信息增益率，对数据集进行合理高效的划分，在本地生成用于测试阶段判断分类的树结构，根据这个树结构得到所需的分类结果。

3 实验结果与分析

在整个实验过程中，本文用到的实验工具包括 GMSSL、VS2010 和 VS Code。GMSSL 同 OpenSSL 一样，都是密码工具箱。它实现了本文所需的 SM4 加密算法，即在 CBC 模式、CFB 模式、CTR 模式和 OFB 模式下的 SM4 密码算法对 1000 份明文文本加密，得到四种工作模式下各 1000 份的密文文本。明文是随机选取的大小在 1KB~200KB 之间的文本文档，都是分组规模的倍数。密文特征提取的工作是由 VS2010 所提供的 C 语言编写的特征提取算法所完成，得到大小写字符、数字字符和特殊字符的统计个数。4000 份密文文本都会产生一组包含四个特征值的特征向量，从而构成所需的特征空间。决策树算法是在 VS Code 上用 python 语言实现，进而为训练阶段学习模型的生成以及测试阶段分类的实现提供了方法。在训练阶段和测试阶段所用的密钥空间是相同的，所有的密钥都是随机值,并且测试阶段和训练阶段文本的大小是相同的。在同一密钥空间下，对同一工作模式下的不同文本文档的密钥是不同的。本文提供了三种实验模型，分别为混合分类模型、混合文本大小分类模型和一对一模型。混合分类模型是将四种工作模式的样本直接投给分类器进行区分。混合文本大小分类模型是在混合分类的基础上，对密文文本进行甄选，将不同文本大小作为分类的影响因子。一对一分类模型是将四种工作模式两两组合，每组单独进行测试。在混合分类模型中，所用到的训练样本数和测试样本数是相同的并且都包含于同一个密钥空间。此外，该模型将样本数量作为测试结果的影响因子。在这三种模型中，正确率指的是测试数据根据训练模型的预测与其标签是否一致的数量（即预测结果正确的数量）与参与测试的数据数量的比值，用公式表示如下：

$$\text{正确率} = \frac{\text{预测结果正确的数量}}{\text{测试数据的总数量}}.$$

表 1 给出了测试结果随着样本数量增加的变化情况。在每次测试中，四种工作模式的样本数都是平均的。当样本总数从 8 个增加到 100 个时，其正确率也逐渐增加，从 16.7%增加至 26.5%。当样本数较大时，测试结果的正确率更接近于稳定值 25%。

表 1：混合分类模型测试结果

Table 1 Hybrid classification model test results

样本总数	各工作模式样本数	正确率(%)
8	2	16.7
20	5	22.2
100	25	26.5
200	50	24.7
500	125	25.7

1000	250	25.4

在混合分类模型的基础上，混合文本大小分类模型将文本大小也考虑为其影响因素。测试结果如表 2 所示。在该表中，第 1 行给出了文本大小的取值范围，第 2 行给出了每一次所对应的样本数。该模型既有文本大小的影响，也有样本数目的作用，其测试结果随着二者逐渐增大有着微小的变化。与混合分类模型相比，其测试结果更为稳定。

通过上述分类结果可以看出，以上两种分类模型对 SM4 算法的四种工作模式，即密文链接模式、密文反馈模式、输出反馈模式和计数器模式，混合时的分类效果并不显著，其识别正确率范围在 16%至 26%之间。如果样本总数超过 100 时，其识别正确率在 20%以上。

表 2：混合文本大小分类模型测试结果

Table 2　Test results of mixed text size classification model

文本大小(KB)	size<=20	20<size<=60	size>60
样本总数	100	200	500
正确率(%)	25	25.5	26

表 3 给出了一对一分类模型的分类结果。该模型将 CBC、CFB、OFB 和 CTR 进行两两组合，得到六组分类样本。在每组分类样本中，训练阶段和测试阶段的样本总数都为 1000 例。其中，CBC vs CFB 的识别率最高，测试结果达到 97.61%。而 CFB vs OFB 的识别率最低，其测试结果达到 97.36%。从该表还可以看到，四种工作模式分类的结果是非常显著的，其正确率高达 95%以上，只有 OFB 与 CTR 识别的结果为 91.67%。通过表 1 至表 3 的分类结果可以看出，将四种工作模式混合递给分类器进行分类的结果并不明显，而当对其进行一对一分类时，分类结果良好。

表 3：一对一分类模型测试结果

Table 3　One-to-one classification model test results

密文工作模式	样本数	识别率(%)	错误率(%)
CBC vs CFB	1000	96.55	3.45
CBC vs OFB	1000	97.61	2.39
CBC vs CTR	1000	97.56	2.44
CFB vs OFB	1000	97.36	2.64
CFB vs CTR	1000	95.83	4.17
OFB vs CTR	1000	91.67	8.33

4　结束语

本文提出一种基于决策树 C4.5 算法的 SM4 分组密码工作模式识别方案，同时从三种情况去测试该方案的可行性和有效性。在本方案中，训练阶段和测试阶段所用的样本数目是一样的，并且对明文文本加密时

两阶段使用相同的密钥空间。较为成功地是进行一对一分类时，CBC 模式与 CFB 模式、OFB 模式、CTR 模式之间的区分正确率都在 90%以上，不足之处在于将四种工作模式混合后的区分率并不高。在接下来的工作中，可以从以下三个方面改进本文提出的方案：首先，可以结合一些优化算法对 C4.5 算法本身进行优化，以便更好地应用于本文的方案；然后对密文进行特征提取时，可以增加提取的特征数量，进而完善特征向量空间；最后可以尝试其他的机器学习算法，近一步去提高方案的有效性。

参考文献：

[1] Kerckhoffs A. Kerckhoffs, la cryptographie militaire[J].Journal des Sciences Militaires 1883: 9-38.

[2] Dileep A D,Sekhar C C. Identification of Block Ciphers Using Support Vector Machines[C].Proceeding of the International Joint Conference on Neural Networks(IJCNN'06),Gulf Islands,Canada,2006: 2696-2701.

[3] R. Manjula and R. Anitha. Identification of Encryption Algorithm Using Decision Tree[J]. Advanced Computing Communications in Computer and Information Science 2011 Volume 133, Part 3, 237-246.

[4] Jung Wei CHOU, Shou De LIN, Chen Mou CHENG. On the effectiveness of using state-of-the-art machine learning techniques to launch cryptographic distinguishing attacks[C]. Proceedings of the 5th ACM workshop on Security and artificial intelligence. New York, U.S.A., 2012: 105-110.

[5] S. Mishra and A. Bhattacharjya. Pattern analysis of cipher text: A combined approach[C].International Conference on Recent Trends in Information Technology, 2013: 393-398.

[6] De Souza William AR, Tonlinson Allan. A distinguishing attack with a neural network[C]. Proceedings of the IEEE 13th International Conference on Data Mining Workshops(ICDMW'13), Dallas, U.S.A., 2013: 154-161.

[7] Vina M. Lomte, Archana D. Shinde. Review of a New Distinguishing Attack Using Block Cipher with a Neural Network. International Journal of Science and Research, VOL. 3, No. 8, August 2014: 733-736.

[8] Yang WU, Tao WANG. Jin Dong LI. Research on a New Method of Statistical Detection of Block Cipher Algorithm Ciphertext[J]. Journal of Ordnance Engineering College. 2015, 27(3): 58-64.
吴杨，王韬，李进东. 分组密码算法密文的统计检测新方法研究[J]. 军械工程学院学报，2015，27(3): 58-64.

[9] Yang WU, Tao WANG, Meng XING, et al. Recognition Scheme of Block Cipher Algorithm Based on Distribution Characteristics of Ciphertext Randomness Measure Value[J]. Journal of Communications, 2015,36(4): 147-155.
吴杨，王韬，邢萌等. 基于密文随机性度量值分布特征的分组密码算法识别方案[J]. 通信学报，2015, 36(4): 147-155.

[10] Mello F L D , Xexeo J A M . Cryptographic Algorithm Identification Using Machine Learning and Massive Processing[J]. IEEE Latin America Transactions, 2016, 14(11):4585-4590.

[11] Cheng TAN and Qing bing JI. An approach to identifying cryptographic algorithm from ciphertext[C]. 2016

8th IEEE International Conference on Communication Software and Networks (ICCSN), Beijing, China, 2016: 19-23.

[12] F.M.Barbosa, A.R.S.F. Vidal, H. L. S. Almeida et al. Machine Learning Applied to the Recognition of Cryptographic Algorithms Used for Multimedia Encryption[J]. Revista IEEE America Latina. v.15, No.7, July 2017.

[13] Liang Tao HUANG, Zhi Cheng ZHAO, Ya Qun ZHAO. Hierarchical recognition scheme of cryptosystem based on random forest[J]. Journal of Computer, 2018,41(2): 382-399.

黄良韬，赵志诚，赵亚群. 基于随机森林的密码体制分层识别方案[J]. 计算机学报，2018,41(2):382-399.

[14] Cheng TAN, Xiaoyan DENG, Lijun ZHANG. Identification of Block Ciphers under CBC Mode[J]. International Congress of Information and Communication Technology. 2018: 65 - 71.

[15] Zhi Cheng ZHAO, Ya Qun ZHAO, Feng Mei LIU. Recognition scheme of block cipher system based on randomness test[J]. Journal of Cryptography, 2019,6(2): 177-190.

赵志诚，赵亚群，刘凤梅. 基于随机性测试的分组密码体制识别方案[J]. 密码学报，2019, 6(2): 177-190.

[16] PUB F.DES Modes of Operation[S]//Federal Information Processing Standards Publication 81,1980.

[17] Menezes A J,Van Oorschot P C, Vanstone S A. Handbook of applied cryptography[M]. Florida: CRC press,1996.

[18] Hong Chao LI. Research on Cipher Algorithm Recognition Based on Ciphertext Features[D]. Xi’an Shaan Xi, Xidian University, 2018.

李洪超. 基于密文特征的密码算法识别研究[D].陕西西安：西安电子科技大学，2018.

[19] Zhi Cheng ZHAO. Research on Cryptographic System Recognition Based on Machine Learning[D]. Zheng zhou, He nan. Information Engineering University, 2018.

赵志诚. 基于机器学习的密码体制识别研究[D].河南郑州：战略支援部队信息工程大学，2018.

[20] Quinlan J R. Induction of Decision Trees[J]. Machine learning,1986,1(1): 81-106.