doi:10.11959/j.issn.1000-436x.20150107

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

吴杨，王韬，邢萌，李进东
（军械工程学院 信息工程系，河北 石家庄 050003）

摘 要：在研究现有加密算法识别方案局限性的基础上，提出了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。首先，基于码元频数检测、块内频数检测及游程检测对 AES、Camellia、DES、3DES 及 SMS4 密文的随机性度量值取值个数进行了统计分析，采用 *k*-means 算法对其进行了初始聚类划分。其次，针对相同聚类中的分组密码算法识别问题，基于降低特征向量间相似度的原则，求解了码元频数检测、块内频数检测及游程检测对应的密文随机性度量值特征向量维数。最后，对 AES、Camellia、DES、3DES 及 SMS4 算法的实验结果表明，提出方案在已知密文条件下，实现了对以上典型分组密码算法的识别，相关成果可为进一步探索基于密文随机性度量值分布特征的加密算法识别提供参考。

关键词：随机性度量；密文随机性；加密算法识别；游程检测；特征向量
中图分类号：TP 309 **文献标识码**：A

Block ciphers identification scheme based on the distribution character of randomness test values of ciphertext

WU Yang, WANG Tao, XING Meng, LI Jin-dong
(Dept. of Information Engineering, Ordnance Engineering College, Shijiazhuang 050003, China)

Abstract: By researching deficiency of current encryption algorithms identification schemes, a block ciphers identification scheme is proposed based on the distribution character of randomness test values for ciphertext. Firstly, the numbers of randomness test values for AES, Camellia, DES, 3DES, SMS4 are respectively calculated based on the frequency test, frequency test in block, run test and originally clustered by the *k*-means algorithm. Secondly, in order to identify the block ciphers in each clustering, the dimensions of eigenvectors to the frequency test, frequency test in block, run test are calculated on the principle of reducing the comparability between eigenvectors. Eventually, the experimental results of AES, Camellia, DES, 3DES, SMS4 demonstrate that the proposed scheme effectively identified the above representative block ciphers, and the correlative research can promote the further encryption algorithms identification research.

Key words: randomness test; randomness of ciphertext; encryption algorithms identification; run test; eigenvector

1 引言

在已知密文条件下，开展加密算法识别及其密钥恢复是 2 项重要而又富有挑战性的研究内容。在加密算法分析领域，广大研究者关注的重点主要是对各类加密算法攻击技术展开研究，意在研究破解获取其密钥之道，先后提出了针对典型分组密码、公钥密码算法、流密码算法的诸多攻击方案，而加密算法识别可为加密算法攻击技术研究提供支持。当前，加密算法识别主要基于统计学及机器学习方法。基于统计学的加密算法识别技术主要通过度量不同加密算法输出的密文中各字符的出现频率，以实现对加密算法的识别，而基于机器学习方法的加密算法识别可归结为模式识别问题。

1993 年 Spliiman 等[1]率先将遗传算法应用到对简单置换密码的分析过程中。1998 年，Ramzan 等[2]将神经网络成功应用于对加密算法的分析识别过程中。

收稿日期：2014-03-15；**修回日期**：2014-09-30
基金项目：军内科研基金资助项目(YJJXM12033)
Foundation Item: The Military Scientific Research Support Project (YJJXM12033)

2006 年，Dileep[3]同时结合统计学与机器学习方法提出了基于支持向量机的分组密码识别方案，实现了对 AES、DES、3DES 等加密算法的识别，但其方案仍无法用于密钥频繁更新的应用环境，且不同的明文也会对其识别结果造成很大的影响。

2008 年，Meltem[4]针对随机映射问题基于随机性检测理论设计了用于流密码的区分器，并将其运用到了对 eSTREAM 项目第三阶段的候选流密码的安全性评估过程中。

2009 年，陈华等[5]针对分组密码的安全性评估过程中的检测统计问题，提出了一种基于密码分组长度的统计检测方法，分别对 Rijndael 算法、Camellia 算法和 SMS4 算法进行了统计检测，评估了其安全性。Liu 等[6]基于软件逆向工程技术从加密程序的二进制代码中提取出了能够表征加密算法类型的特征码，在建立起相应的加密算法特征库的基础上，采用 Boyer-Moore 匹配算法完成了加密算法特征代码的匹配。

2011 年，Manjula 等[7]提出了基于 C4.5 算法的加密算法识别方案，其方案通过提取密文数据的 8 种特征参数作为加密算法分类标准，采用基于 C4.5 算法的分类器对加密算法进行识别。Gröbert 等[8]则采用细粒度的动态二进制代码分析方法将收集到的代码信息作为进一步挖掘加密算法代码特征的启发式信息，实现了针对完整加密算法或密钥的二进制代码程序的加密原语识别方案。

以上识别方案中，逐渐成为主流分析方法的二进制代码分析方法往往需要对加密程序进行逆向分析，导致其在非合作应用环境下的适用性不强。国外学者从研究密文统计信息角度出发对加密算法识别问题进行了一些研究，但其方案仍存在较大的应用局限性，需进一步探索出更加有效的加密算法识别方案。为此，本文在研究典型分组密码算法的密文随机性度量值分布特征基础上，提出了基于密文随机性度量值分布特征的典型分组密码算法识别技术思想及方案，以弥补传统的基于二进制分析等方法识别分组密码算法的局限性。

2　背景知识

2.1　分组密码

分组密码的加密过程是将明文消息编码表示后的二进制序列划分成固定大小的分组，而每一分组在对应密钥的控制下被变换成等长的二进制序列。同公钥密码算法相比，分组密码算法具有加密速度快、易于标准化等特点，其典型代表有 AES、Camellia、DES、3DES 及 SMS4 等。

多数分组密码为确保其安全性，在设计过程中采用了扩散和混乱的基本原则[9]。扩散的基本要求是每一比特的明文及密钥变化应尽可能多地影响到输出密文序列的比特。从该意义上来说加密算法的明文及密钥均会对其密文输出序列的随机统计特性产生一定的影响，文献[3]的研究成果也印证了该结论。混乱则要求在加密变换过程中明文、密钥及密文之间的关系应尽可能复杂。然而，从解密的过程来看，虽然密码算法在设计过程中采用了扩散及混乱的基本原理，不同的密码算法引入了不同的扩散及混乱方案，但其扩散及混乱及仍是“有规律、可逆的”的变换，否则很难实现其密文的解密。因此，分组密码算法自身所采用的扩散及混淆方案体现了其算法自身的诸多特性，而分组密码算法攻击技术研究者们正是充分利用了其扩散及混淆的特性，实现了对 DES[10]、AES[11]、Camellia[12]、SMS4[13]等算法密钥的恢复。因此，在识别分组密码算法过程中，如何挖掘并利用其扩散及混淆方案体现出的加密算法信息，对成功识别分组密码算法将起到关键作用。

2.2　随机性检测

在信息的传输过程中，保持其内容不被非授权者获取，确保机密性的实现方式可从物理保护到采用让数据无法理解的数学算法[14]。已加密数据主要借助于其在统计方面的随机性实现信息的保密功能，同一般的数据相比，在统计上体现出了更大的随机特性。而不同分组密码密文的随机程度也必然存在一定的差异性，通过度量这些差异也可对加密算法自身的安全性进行评估。随机性检测建立在假设检验的基础上，其基本思想为小概率原理，即如果认为某个假设是真实的，则在这个假设下发生的概率很小的一个事件在一次试验中是不可能发生的，如果这一事件发生，则假设的真实性将受到质疑。目前，已经有了众多的随机性检测项目和方法，主要应用于检测密码算法输出序列的统计特性[15,16]，相关的随机性检测方法更是多达 200 多种[17]。其中，应用较为广泛的是美国商务部国家标准技术协会(NIST)于 2001 年 5 月公布的 FIPS140-2 标准中定义的用于密码系统安全性度量的诸多随机性检测方

案[18]及2010年4月公布的应用于测试随机数及伪随机数生成器性能的随机性检测标准 SP800-22 rev1a(special publication 800-22 revision 1a)[19]。

然而，到目前为止，以上方法均从不同角度对比特序列的随机性进行考察，尚无一种或几种测试方法是测试密文随机性统计特征的充分条件。鉴于目前尚无一种或几种测试方法是测试密文随机性统计特征的充分条件，研究过程主要选择 NIST 标准中最为典型且应用最为广泛的码元频数检测(FT, frequency test)、游程检测(RT, run test)及块内频数检测(FTB, frequency test in block)方法用于评估不同分组密码的密文在码元频数、游程、块内频数方面的随机分布特征，其具体原理及实现过程可参考文献[19]中的相关内容。

3 识别方案

3.1 已有识别方案分析

如前所述，文献[3]提出的基于支持向量机的分组密码算法识别方案，其实现过程首先基于统计及映射方法分别构建定长及变长字段的字典表，并定义 t_i 为字符字典中的第 i 个字符，$tf(t_i,c)$为给定的密文序列中字符 t_i 出现的频率。基于以上定义，可采用向量方式将给定的密文表示为 $\boldsymbol{\phi}(c)=(tf(t_1,c),tf(t_2,c),\cdots,tf(t_N,c))^{\mathrm{T}}$，其中 N 为字典中的字符数，同时也为向量 $\boldsymbol{\phi}(c)=(tf(t_1,c),tf(t_2,c),\cdots,tf(t_N,c))^{\mathrm{T}}$ 的维数。根据建立的字典将给定的密文转换为对应的向量后，在给定的明文及密钥条件下，以建立的向量为支持向量机的训练样本，针对不同的分组密码算法建立对应的识别系统。文献[3]方案中关于密文的向量 $\boldsymbol{\phi}(c)=(tf(t_1,c),tf(t_2,c),\cdots,tf(t_N,c))^{\mathrm{T}}$ 除表征密文中出现字符的频率特征外，密文 c 中比特 0 和 1 出现的频率可根据 $F_a(c)=\sum_{i=1}^{N}K_a(t_i)tf(t_i,c),a=0,1$ 计算获得，而 $Ka(t_i)$ 表示字符 t_i 出现的比特 0 或 1 的频数。因此，$\phi(c)$ 本质上也进一步表征了密文序列中比特 0 和 1 出现的频率信息。然而，文献[3]也进一步指出其识别方案的识别率受明文及密钥的影响较大，并不适用于密钥频繁更新的应用场合。

在密钥变化时造成其识别率降低的主要原因在于：当分组密码算法的密钥发生改变时，相同明文获得的新的密文向量 $\boldsymbol{\phi}'(c')=(tf(t_1,c'),tf(t_2,c'),\cdots,tf(t_N,c'))^{\mathrm{T}}$ 必将发生改变，并导致其识别率降低。同时，文献[3]的方案选择了不同长度的字符参与构建特征向量，当比特序列的划分粒度较大时，向量 $\boldsymbol{\phi}'(c')=(tf(t_1,c'),tf(t_2,c'),\cdots,tf(t_N,c'))^{\mathrm{T}}$ 的变化将更加显著，识别误差也将进一步增大且构建的特征向量维数也更大。因此，要确保其方案在密钥发生改变时仍具有较高的识别率，需降低比特序列的划分粒度。本文将从单比特(码元频数检测、块内频数检测)及多比特(游程检测)的密文序列划分粒度角度出发对不同典型分组密码算法的密文呈现出的随机分布特性展开研究。下面给出具体的实现方案。

3.2 密文随机性度量

在度量密文的随机性过程中，定义如下假设条件。

1) 不同的分组密码 B_i 其密文 $CB_i=(c_{g1},c_{g2},c_{g3},\cdots,c_{gn})$ 存储在不同的集合 CB_i 中，且对应的分组密码算法 B_i 未知，g 为密文分组号、n 为组内的密文条数。

2) 密文集合 C_{Bi} 中的任一密文 c_{gi} 的长度不小于 B_i 的密文长度 L_{Bi}。

3) 密文的随机性度量过程中，B_i 的明文 m 及密钥 k_i 可变。

为从单比特及多比特的密文序列划分角度实现对不同的密码算法的密文随机性进行度量，分别采用频数检测、游程检测及块内频数检测提取密文的随机分布特征。

基于码元频数检测的密文随机分布值统计过程中，由码元频数检测统计分组密码 B_i 的密文 $CB_i=(c_{g1},c_{g2},c_{g3},\cdots,c_{gn})$ 的随机性度量值定义为 $P_{CB_i-FT}=(p_{c_{g1}-FT},p_{c_{g2}-FT},p_{c_{g3}-FT},\cdots,p_{c_{gn}-FT})$；基于游程检测的随机性度量值定义为 $P_{CB_i-RT}=(p_{c_{g1}-RT},p_{c_{g2}-RT},p_{c_{g3}-RT},\cdots,p_{c_{gn}-RT})$；基于块内频数检测的密文随机性度量值定义为 $P_{CB_i-FTB}=(p_{c_{g1}-FTB},p_{c_{g2}-FTB},p_{c_{g3}-FTB},\cdots,p_{c_{gn}-FTB})$。

3.3 基于随机性度量值个数的分组密码初始聚类

实验过程中对典型分组密码(AES、Camellia、DES、3DES 及 SMS4)的密文随机性度量过程发现，在相同的样本量条件下，不同分组密码其密文的随机性度量值取值个数间存在着一定的差异。因此，在识别过程中，首先将采用聚类思想将其密文具有相似的随机性度量值取值个数的分组密码算法划

分到相同的聚类当中，具体步骤如下。

Step1　在输入为随机明文及随机密钥条件下，对分组密码 B_i 的密文 $CB_i=(c_{g1},c_{g2},c_{g3},\cdots,c_{gn})$ 分别进行 FT、RT 及 FTB 检测，得 P_{CB_i-FT}、P_{CB_i-RT}、P_{CB_i-FTB}。

Step2　令 P_{CB_i-FT}、P_{CB_i-RT} 及 P_{CB_i-FTB} 中不同元素的取值个数为 N_{CB_i-FT}、N_{CB_i-RT} 及 N_{CB_i-FTB}。

Step3　对不同的密文分组，重复执行 Step1 和 Step2，得不同取值的 N_{CB_i-FT}、N_{CB_i-RT} 及 N_{CB_i-FTB}，及其均值 $\tilde{N}_{CB_i-FT}$、$\tilde{N}_{CB_i-RT}$ 及 $\tilde{N}_{CB_i-FTB}$。

Step4　采用 k-means 算法将 $(\tilde{N}_{CB_i-FT},\tilde{N}_{CB_i-RT},\tilde{N}_{CB_i-FTB})$ $(1\leqslant i\leqslant r)$ 划分到对应的 $k(k\leqslant r)$ 个聚类中，并记录每个聚类的聚类中心 $C_i=(c_{FT_i},c_{RT_i},c_{FTB_i})(i\leqslant k)$ 及包含的分组密码算法。

确定聚类中心后，对分组密码算法 B_x 进行聚类划分时，重复执行 Step1 和 Step2，得不同密文分组的随机性度量统计值 N_{CB_x-FT}、N_{CB_x-RT}、N_{CB_x-FTB} 及 $\tilde{N}_{CB_x}=(\tilde{N}_{CB_x-FT},\tilde{N}_{CB_x-RT},\tilde{N}_{CB_x-FTB})$，并根据 $\min\left|C_i-\tilde{N}_{CB_x}\right|$ 原则确定分组密码 B_x 所属的聚类。

3.4　密文随机性度量值分布特征向量提取

基于 FT、RT 及 FTB 随机性测试过程，可分别获得如下的处于同一聚类中的 r 个分组密码的密文随机性度量值集合。

$$\begin{cases}P_{FT}=\left(P_{CB_1-FT},P_{CB_2-FT},\cdots,P_{CB_r-FT}\right)\\P_{RT}=\left(P_{CB_1-RT},P_{CB_2-RT},\cdots,P_{CB_r-RT}\right)\\P_{FTB}=\left(P_{CB_1-FTB},P_{CB_2-FTB},\cdots,P_{CB_r-FTB}\right)\end{cases}\tag{1}$$

在集合(1)的基础上，以分组密码 B_i 为例，其 P_{CB_i-FT}、P_{CB_i-RT} 及 P_{CB_i-FTB} 测试结果中，任一元素 p_i 的取值满足 $p_i\in[0,1]$。对同一聚类中的不同分组密码算法而言，P_{CB_i-FT}、P_{CB_i-RT} 及 P_{CB_i-FTB} 中的元素在区间 $[0,1]$ 内的分布特征也不同，即某固定的 Δp_{FT}、Δp_{RT} 及 Δp_{FTB} 取值区间包含的 P_{CB_i-FT}、P_{CB_i-RT} 及 P_{CB_i-FTB} 中的元素个数 ΔN_{FT}、ΔN_{RT} 及 ΔN_{FTB} 存在一定的差异性。因此，可构建如式(2)所示的特征向量以表征分组密码密文的随机性度量值分布特征。

$$\begin{cases}\boldsymbol{V}_{CB_i-FT}=\left(\dfrac{\Delta N_{FT1}}{n},\dfrac{\Delta N_{FT2}}{n},\cdots,\dfrac{\Delta N_{FTh_{FT}}}{n}\right)\\\boldsymbol{V}_{CB_i-RT}=\left(\dfrac{\Delta N_{RT1}}{n},\dfrac{\Delta N_{RT2}}{n},\cdots,\dfrac{\Delta N_{RTh_{RT}}}{n}\right)\\\boldsymbol{V}_{CB_i-FTB}=\left(\dfrac{\Delta N_{FTB1}}{n},\dfrac{\Delta N_{FTB2}}{n},\cdots,\dfrac{\Delta N_{FTBh_{FTB}}}{n}\right)\end{cases}\tag{2}$$

其中，h_{FT}、h_{RT} 及 h_{FTB} 分别为特征向量 $\boldsymbol{V}_{CB_i-FT}$、$\boldsymbol{V}_{CB_i-RT}$ 及 $\boldsymbol{V}_{CB_i-FTB}$ 的维数。由于 FT、RT 及 FTB 测试过程中，集合 P_{CB_i-FT}、P_{CB_i-RT} 及 P_{CB_i-FTB} 中元素的个数不尽相同。因此，h_{FT}、h_{RT} 及 h_{FTB} 的取值也不一定相同。而 h_{FT}、h_{RT} 及 h_{FTB} 的取值大小对识别率的影响主要体现在以下几方面。

1) 若 h_{FT}、h_{RT} 及 h_{FTB} 取值过小，则易导致向量 $\boldsymbol{V}_{CB_i-FT}$、$\boldsymbol{V}_{CB_i-RT}$ 及 $\boldsymbol{V}_{CB_i-FTB}$ 无法正确反映密文的随机分布特征。

2) 若 h_{FT}、h_{RT} 及 h_{FTB} 取值过大，则易增加向量 $\boldsymbol{V}_{CB_i-FT}$、$\boldsymbol{V}_{CB_i-RT}$ 及 $\boldsymbol{V}_{CB_i-FTB}$ 对某些非特征参数的敏感度，且向量相似性度量阶段的存储及计算量也会随之增加。

3) 对于密文的随机性度量值分布特征有明显差异的分组密码算法而言 h_{FT}、h_{RT} 及 h_{FTB} 的取值，对识别结果不会造成较大的影响。对于密文的随机性度量值分布特征差异较小的分组密码算法而言(如 AES 与 Camellia、DES 与 3DES)，h_{FT}、h_{RT} 及 h_{FTB} 的取值对提高识别率至关重要。

在确定特征向量 $\boldsymbol{V}_{CB_i-FT}$、$\boldsymbol{V}_{CB_i-RT}$ 及 $\boldsymbol{V}_{CB_i-FTB}$ 的维数 h_{FT}、h_{RT} 及 h_{FTB} 的过程中，将采用如下的基于特征向量间最小相似度思想的向量维数确定方案。

Step1　首先，令 $h_{FT}=h_{FT0}$、$h_{RT}=h_{RT0}$ 及 $h_{FTB}=h_{FTB0}$。同时，根据密文的随机性度量值的有效值位数 e_{FT}、e_{RT} 及 e_{FTB} 确定向量最大维数为 $h_{FT\max}=10^{e_{FT}}$、$h_{RT\max}=10^{e_{RT}}$、$h_{FTB\max}=10^{e_{FTB}}$。

Step2　当 $h_{FT}\leqslant h_{FT\max}-1$、$h_{RT}\leqslant h_{RT\max}-1$ 及 $h_{FTB}\leqslant h_{FTB\max}-1$ 时，分别以 $\Delta P_{FT}=1/h_{FT}$、$\Delta P_{RT}=1/h_{RT}$ 及 $\Delta P_{FTB}=1/h_{FTB}$ 为步长值做如下统计

$$\begin{cases}[0,\Delta P_{FT})\rightarrow\Delta N_{FT1}\\[\Delta P_{FT},2\Delta P_{FT})\rightarrow\Delta N_{FT2}\\\cdots\\[1-\Delta P_{FT},1]\rightarrow\Delta N_{FTh_{FT}}\end{cases}\tag{3}$$

$$\begin{cases}\left[0,\Delta P_{RT}\right)\to\Delta N_{RT1}\\ \left[\Delta P_{RT},2\Delta P_{RT}\right)\to\Delta N_{RT2}\\ \cdots\\ \left[1-\Delta P_{RT},1\right]\to\Delta N_{RTh_{RT}}\end{cases}\tag{4}$$

$$\begin{cases}\left[0,\Delta P_{FTB}\right)\to\Delta N_{FTB1}\\ \left[\Delta P_{FTB},2\Delta P_{FTB}\right)\to\Delta N_{FTB2}\\ \cdots\\ \left[1-\Delta P_{FTB},1\right]\to\Delta N_{FTBh_{FTB}}\end{cases}\tag{5}$$

以任一分组密码算法候选集合中的算法 B_i 和 B_j 为例，根据式(2)分别构建其密文随机分布特征向量 $\boldsymbol{V}_{B_i}$ 和 $\boldsymbol{V}_{B_j}$，并获得其 g 组密文的随机性度量值分布特征向量均值 $\tilde{\boldsymbol{V}}_{B_i}$、$\tilde{\boldsymbol{V}}_{B_j}$。

构建起向量 $\tilde{\boldsymbol{V}}_{B_i}$、$\tilde{\boldsymbol{V}}_{B_j}$ 后，采用如式(6)所示的相似性度量函数分别度量在 h_{FT}、h_{RT} 及 h_{FTB} 维数下 $\tilde{\boldsymbol{V}}_{B_i}$、$\tilde{\boldsymbol{V}}_{B_j}$ 间的相关性

$$\begin{cases}\lambda_{\tilde{V}_{B_i}\tilde{V}_{B_j}-FT_{h_{FT}}}=\dfrac{4\left(\tilde{\boldsymbol{V}}_{B_i-FT},\tilde{\boldsymbol{V}}_{B_j-FT}\right)}{\left(\left|\tilde{\boldsymbol{V}}_{B_i-FT}\right|+\left|\tilde{\boldsymbol{V}}_{B_j-FT}\right|\right)^2}\\ \lambda_{\tilde{V}_{B_i}\tilde{V}_{B_j}-RT_{h_{RT}}}=\dfrac{4\left(\tilde{\boldsymbol{V}}_{B_i-RT},\tilde{\boldsymbol{V}}_{B_j-RT}\right)}{\left(\left|\tilde{\boldsymbol{V}}_{B_i-RT}\right|+\left|\tilde{\boldsymbol{V}}_{B_j-RT}\right|\right)^2}\\ \lambda_{\tilde{V}_{B_i}\tilde{V}_{B_j}-FTB_{h_{FTB}}}=\dfrac{4\left(\tilde{\boldsymbol{V}}_{B_i-FTB},\tilde{\boldsymbol{V}}_{B_j-FTB}\right)}{\left(\left|\tilde{\boldsymbol{V}}_{B_i-FTB}\right|+\left|\tilde{\boldsymbol{V}}_{B_j-FTB}\right|\right)^2}\end{cases}\tag{6}$$

Step3 取 $h_{FT}=h_{FT}+\Delta h_{FT}$、$h_{RT}=h_{RT}+\Delta h_{RT}$、$h_{FTB}=h_{FTB}+\Delta h_{FTB}$，重复执行 Step2，可获得在不同的 h_{FT}、h_{RT} 及 h_{FTB} 维数下，算法 B_i 和 B_j 的特征向量间的相关性度量值，最终 h_{FT}、h_{RT} 及 h_{FTB} 的取值可根据度量值的变化趋势予以确定，可选择使相似性度量值取得最小值的 h_{FT}、h_{RT} 及 h_{FTB} 作为 FT、RT、FTB 测试的向量维数。确定 h_{FT}、h_{RT} 及 h_{FTB} 的取值后，不同分组密码密文的随机分布特征向量间仍可能存在较大的相似性，按如下方式进一步对其特征向量进行处理。

1) 针对每一加密算法候选集合，将算法 B_i 的 g 组密文平均分为 2 组，以 h_{FT}、h_{RT} 及 h_{FTB} 为密文随机性度量值划分区间数，分别对其 2 组密文的 FT、RT、FTB 测试均值进行统计得 $\tilde{\boldsymbol{V}}_{B_i-FT_1}$、$\tilde{\boldsymbol{V}}_{B_i-FT_2}$、$\tilde{\boldsymbol{V}}_{B_i-RT_1}$、$\tilde{\boldsymbol{V}}_{B_i-RT_2}$、$\tilde{\boldsymbol{V}}_{B_i-FTB_1}$ 及 $\tilde{\boldsymbol{V}}_{B_i-FTB_2}$。

2) 以如下方式计算 B_i 算法密文在 FT、RT、FTB 测试下的随机性度量值变化均值

$$\begin{cases}\Delta\tilde{\boldsymbol{V}}_{B_i-FT}=\left|\tilde{\boldsymbol{V}}_{B_i-FT_1}-\tilde{\boldsymbol{V}}_{B_i-FT_2}\right|\\ \Delta\tilde{\boldsymbol{V}}_{B_i-RT}=\left|\tilde{\boldsymbol{V}}_{B_i-RT_1}-\tilde{\boldsymbol{V}}_{B_i-RT_2}\right|\\ \Delta\tilde{\boldsymbol{V}}_{B_i-FTB}=\left|\tilde{\boldsymbol{V}}_{B_i-FTB_1}-\tilde{\boldsymbol{V}}_{B_i-FTB_2}\right|\end{cases}\tag{7}$$

3) 令 $\Delta T\tilde{\boldsymbol{V}}_{B_i}=\left(\Delta\tilde{\boldsymbol{V}}_{B_i-FT},\Delta\tilde{\boldsymbol{V}}_{B_i-RT},\Delta\tilde{\boldsymbol{V}}_{B_i-FTB}\right)$ 为 B_i 算法在样本量为 g 时的特征向量模板。

4) 根据式(6)，定义特征向量模板 $\Delta T\tilde{\boldsymbol{V}}_{B_i}$、$\Delta T\tilde{\boldsymbol{V}}_{B_j}$ 间的不匹配度为

$$\begin{cases}\bar{\lambda}_{\Delta\tilde{V}_{B_i}\Delta\tilde{V}_{B_j}-FT}=1-\lambda_{\Delta\tilde{V}_{B_i}\Delta\tilde{V}_{B_j}-FT}\\ \bar{\lambda}_{\Delta\tilde{V}_{B_i}\Delta\tilde{V}_{B_j}-FT}=1-\lambda_{\Delta\tilde{V}_{B_i}\Delta\tilde{V}_{B_j}-RT}\\ \bar{\lambda}_{\Delta\tilde{V}_{B_i}\Delta\tilde{V}_{B_j}-FTB}=1-\lambda_{\Delta\tilde{V}_{B_i}\Delta\tilde{V}_{B_j}-FTB}\end{cases}\tag{8}$$

5) 进一步求解初始聚类中加密算法特征向量模板间的平均不匹配度为 $\tilde{\bar{\lambda}}_{FT}$、$\tilde{\bar{\lambda}}_{RT}$ 及 $\tilde{\bar{\lambda}}_{FTB}$。

3.5 密文的随机性度量值分布特征向量匹配

在对某分组密码算法 B_x 的密文 $CB_x=(c_{g1},c_{g2},c_{g3},...,c_{gn})$ 的随机性度量值分布特征进行匹配前，需根据 h_{FT}、h_{RT} 及 h_{FTB} 构建其特征向量 $\Delta\tilde{\boldsymbol{V}}_{CB_x}=(\Delta\tilde{\boldsymbol{V}}_{CB_x-FT},\Delta\tilde{\boldsymbol{V}}_{CB_x-RT},\Delta\tilde{\boldsymbol{V}}_{CB_x-FTB})$。

定义分组密码 B_i 密文的随机性度量值特征向量模板 $\Delta T\tilde{\boldsymbol{V}}_{B_i}$ 与待识别分组密码 B_x 密文的随机性度量值特征向量 $\Delta\tilde{\boldsymbol{V}}_{B_x}$ 间的相关系数 $\phi_{B_iB_x}$ 为

$$\phi_{B_iB_x}=\alpha\lambda_{B_iB_x-FT}+\beta\lambda_{B_iB_x-RT}+\gamma\lambda_{B_iB_x-FTB}\tag{9}$$

其中，系数 $\alpha\geqslant 0,\beta\geqslant 0,\gamma\geqslant 0,\alpha+\beta+\gamma=1$，取值如式(10)所示。

$$\begin{cases}\alpha=\dfrac{\tilde{\bar{\lambda}}_{FT}}{\tilde{\bar{\lambda}}_{FT}+\tilde{\bar{\lambda}}_{RT}+\tilde{\bar{\lambda}}_{FTB}}\\ \beta=\dfrac{\tilde{\bar{\lambda}}_{RT}}{\tilde{\bar{\lambda}}_{FT}+\tilde{\bar{\lambda}}_{RT}+\tilde{\bar{\lambda}}_{FTB}}\\ \gamma=\dfrac{\tilde{\bar{\lambda}}_{FTB}}{\tilde{\bar{\lambda}}_{FT}+\tilde{\bar{\lambda}}_{RT}+\tilde{\bar{\lambda}}_{FTB}}\end{cases}\tag{10}$$

1) 当 $Num(\max(\phi_{B_1B_x},\phi_{B_2B_x},\cdots,\phi_{B_rB_x}))=1$，则待识别分组密码 B_x 的密文随机分组特征符合分组密码 B_i 的密文随机性度量值分布特征，判别 B_x 为 B_i。

2) 当 $Num(\max(\phi_{B_1B_x},\phi_{B_2B_x},...,\phi_{B_rB_x}))>1$，即分组密码候选算法不唯一时，则进一步通过计算相关系数 $\lambda_{B_iB_x-FT},\lambda_{B_iB_x-RT},\lambda_{B_iB_x-FTB}$ 的方差 $\mathrm{Var}_{B_iB_x}(\lambda_{B_iB_x-FT},\lambda_{B_iB_x-RT},\lambda_{B_iB_x-FTB})$，若存在分组密码 B_i 使得 $\min(\mathrm{Var}_{B_1B_x},\mathrm{Var}_{B_2B_x},\cdots,\mathrm{Var}_{B_rB_x})=1$，则判定待识别分组密码为 B_i。若仍出现 $\min(\mathrm{Var}_{B_1B_x},\mathrm{Var}_{B_2B_x},\cdots,\mathrm{Var}_{B_rB_x})>1$的情况，则应增加密文的输入量。

4　典型分组密码算法识别结果与分析

4.1　实验对象及相关设置

为验证提出方案的有效性，实验过程选择以 OpenSSL 密码库中的 AES、Camellia、DES、3DES，以及在 VC 6.0 环境下编程实现的 SMS4 算法为识别分析对象。码元频数检测、游程检测、块内频数检测及相关数据的统计分析识别等过程则在 Matlab 仿真环境中实现。实验过程中，以随机明文及随机密钥为分组密码算法输入，获得不同的分组密码在随机明文及随机密钥输入条件下的密文。实验过程中，待分析的 AES、Camellia、DES、3DES、SMS4 的密文样本量均为 2 000 000 条，并按存储顺序将其平均分为 200 组数据(每组 10 000 条密文)。

4.2　分组密码的密文随机性度量值个数统计与分析

实验以 10 000 条密文为分组标准，分别对 AES、Camellia、DES、3DES、SMS4 的 2 000 000 条密文进行码元频数检测，其 200 组测试数据的码元频数检测的度量值个数统计结果如图 1 所示。

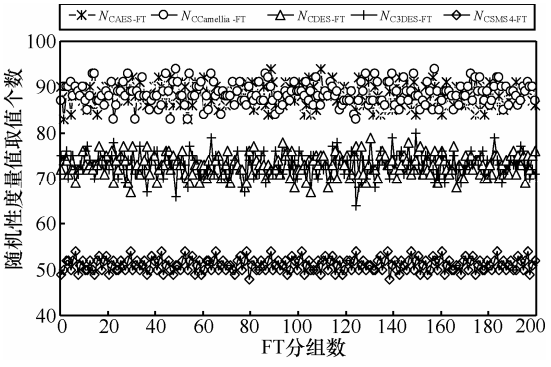


图 1　码元频数检测随机性度量值取值个数

AES 与 Camellia 的 200 组密文数据经 FT 检测后，其随机性度量值个数的取值范围分别为 $N_{CAES-FT}\in[83,94]$、$N_{CCamellia-FT}\in[83,94]$，平均取值数则分别为 $\tilde{N}_{CAES-FT}=88.095$ 和 $\tilde{N}_{CCamellia-FT}=88.450$。DES 与 3DES 的 200 组密文数据经 FT 检测后，其随机性度量值个数取值范围分别为 $N_{CDES-FT}\in[67,79]$ 和 $N_{C3DES-FT}\in[64,80]$，而平均取值数则分别为 $\tilde{N}_{CDES-FT}=72.860$ 和 $\tilde{N}_{C3DES-FT}=72.880$。SMS4 的 200 组密文数据经 FT 检测后，其随机性度量值个数取值范围为 $N_{CSMS4-FT}\in[48,54]$，而平均取值数则为 $\tilde{N}_{CSMS4-FT}=51.120$。

如图 2 所示，FTB 测试过程中(分块长度为 20 bit)，AES 与 Camellia 的随机性度量值个数分别为 $N_{CAES-FTB}\in[413,443]$、$N_{CCamellia-FTB}\in[415,443]$，而平均取值数则分别为 $\tilde{N}_{CAES-FTB}=429.750$ 和 $\tilde{N}_{CCamellia-FTB}=429.845$。DES 与 3DES 的随机性度量值个数取值范围分别为 $N_{CDES-FTB}\in[384,418]$ 和 $N_{C3DES-FTB}\in[386,419]$，而其平均取值数则分别为 $\tilde{N}_{CDES-FTB}=402.090$ 和 $\tilde{N}_{C3DES-FTB}=402.025$。SMS4 的随机性度量值个数取值范围为 $N_{CSMS4-FTB}\in[259,267]$，而其平均取值数则为 $\tilde{N}_{CSMS4-FTB}=263.175$。

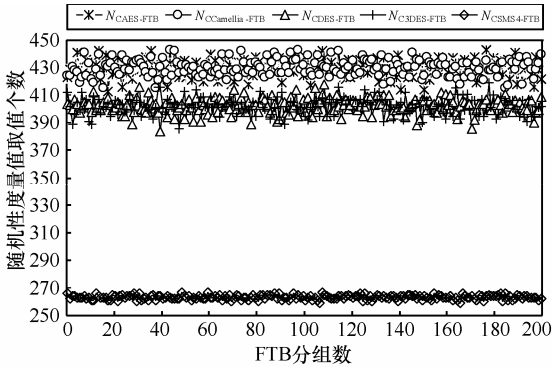


图 2　块内频数检测随机性度量值取值个数

RT 测试过程中，AES 与 Camellia 的随机性度量值个数的取值范围为 $N_{CAES-RT}\in[2\,322,2\,467]$、$N_{CCamellia-RT}\in[2\,327,2\,462]$，而平均取值数则分别为 $\tilde{N}_{CAES-RT}=2\,397.720$ 和 $\tilde{N}_{CCamellia-RT}=2\,395.255$。DES 与 3DES 的随机性度量值个数取值范围分别为 $N_{CDES-RT}\in[2\,133,2\,271]$、$N_{C3DES-RT}\in[2\,128,2\,276]$，而其平均取值数则分别为 $\tilde{N}_{CDES-RT}=2\,195.310$ 和 $\tilde{N}_{C3DES-RT}=2\,195.270$。SMS4 的随机性度量值个数取值范围为 $N_{CSMS4-RT}\in[1\,961,2\,022]$，而其平均取值数则为 $\tilde{N}_{CSMS4-RT}=1\,988.275$，结果如图 3 所示。

由以上测试结果可知，SMS4 密文的 FT、FTB 及 RT 测试值取值个数均小于 AES、Camellia、DES、3DES 密文的 FT、FTB 及 RT 测试值取值个数。AES、Camellia、DES、3DES 的密文比 SMS4

的密文能够覆盖更多不同的随机性度量值。综合以上实验数据，采用 *k*-means 算法对如下数据进行聚类划分

$$\begin{cases}\left(\tilde{N}_{CAES-FT},\tilde{N}_{CAES-FTB},\tilde{N}_{CAES-RT}\right)\\ \left(\tilde{N}_{CCamellia-FT},\tilde{N}_{CCamellia-FTB},\tilde{N}_{CCamellia-RT}\right)\\ \left(\tilde{N}_{CDES-FT},\tilde{N}_{CDES-FTB},\tilde{N}_{CDES-RT}\right)\\ \left(\tilde{N}_{C3DS-FT},\tilde{N}_{C3DS-FTB},\tilde{N}_{C3DES-RT}\right)\\ \left(\tilde{N}_{CSMS4-FT},\tilde{N}_{CSMS4-FTB},\tilde{N}_{CSMS4-RT}\right)\end{cases} \tag{11}$$

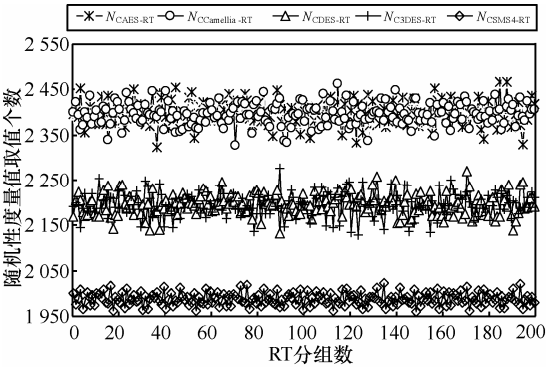


图 3 游程检测随机性度量值取值个数

聚类划分结果表明，AES 与 Camellia 被划分到相同的聚类中 C_1 中，DES 与 3DES 则被划分到相同的聚类中 C_2 中，而 SMS4 则被单独划分到聚类 C_3 中。在密文数据的样本为 10 000 条件下，C_1、C_2 及 C_3 的聚类中心为 $(88.27,429.80,2\,396.49)$、$(72.87,402.09,2\,195.29)$、$(52.17,263.48,1\,989.61)$。后续实验过程中根据 $\min\left|C_i-\tilde{N}_{CB_x}\right|$ 原则，对 AES、Camellia、DES、3DES、SMS4 的 10 000 条密文数据进行处理后，AES 与 Camellia 算法均能被划分相同的聚类当中，DES 与 3DES 也处于相同的聚类，SMS4 则被单独划分到同一聚类中。综上所述，基于 AES、Camellia、DES、3DES、SMS4 密文的 FT、FTB 及 RT 度量值取值个数，采用聚类思想可实现将 AES、Camellia、DES、3DES 进行初始聚类划分，并识别出 SMS4 算法。因此，下面仅需进一步对 AES、Camellia、DES、3DES 密文的随机性度量值分布特征进行进一步挖掘。

4.3 特征向量维数确定

基于 FT 的密文随机性度量值分布特征向量提取及相似性度量过程中，根据 FT 测试下的随机性度量值个数，设定特征向量的最大维数为 500，令 $\Delta h_{FT}=5$。在对 AES、Camellia、DES、3DES 进行了初始聚类划分的基础上，为提高实验分析效率，仅统计了 AES-AES、AES-Camellia、Camellia-Camellia、DES-DES、DES-3DES、3DES-3DES 特征向量间的相似度。AES-AES、AES-Camellia、Camellia-Camellia、DES-DES、DES-3DES、3DES-3DES 间的相似性度量值随着向量维数的增加逐渐趋于稳定。当 $h_{FT}\geqslant 100$ 时，AES、Camellia、DES、3DES 间特征向量的相似度均值如表 1 所示。

表 1 基于码元频数检测的特征向量相似度均值

	AES	Camellia	DES	3DES
AES	0.999 990 01	0.999 958 79	—	—
Camellia	0.999 958 79	0.999 986 90	—	—
DES	—	—	0.999 992 012	0.999 9658 89
3DES	—	—	0.999965889	0.999992148

基于 FTB 的密文随机性度量值分布特征向量提取及相似性度量过程中，根据 FTB 测试下的随机性度量值个数，设定特征向量的最大维数为 1 000，令 $\Delta h_{FTB}=10$。实验同样仅统计了 AES-AES、AES-Camellia、Camellia–Camellia、DES-DES、DES-3DES、3DES-3DES 特征向量间的相似性度量值。AES-AES、AES-Camellia、Camellia–Camellia、DES-DES、DES-3DES、3DES-3DES 间的相似性度量值同样随着向量维数的增加逐渐趋于稳定。当 $h_{FTB}\geqslant 300$ 时，AES、Camellia、DES、3DES 间特征向量的相似度均值如表 2 所示。

表 2 基于块内频数检测的特征向量相似度均值

	AES	Camellia	DES	3DES
AES	0.999 932 70	0.999 697 43	—	—
Camellia	0.999 697 43	0.999 928 46	—	—
DES	—	—	0.999 913 93	0.999 735 67
3DES	—	—	0.999 735 67	0.999 930 15

基于 RT 的密文随机性度量值分布特征向量提取及相似性度量过程中，根据 RT 测试下的随机性度量值个数，设定特征向量的最大维数为 10 000，令 $\Delta h_{RT}=100$，AES-AES、AES-Camellia、Camellia-Camellia、DES-DES、DES-3DES、3DES-3DES 间的相似性度量值随着向量维数的增加逐渐趋于稳定。当 $h_{RT}\geqslant 4\,000$ 时，AES、Camellia、DES、3DES

间特征向量的相似度均值如表 3 所示。

表 3　基于游程检测的特征向量相似度均值

	AES	Camellia	DES	3DES
AES	0.999 430 75	0.997 735 57	—	—
Camellia	0.997 735 57	0.999 452 85	—	—
DES	—	—	0.999 550 09	0.998 109 93
3DES	—	—	0.998 109 93	0.999 551 59

前面的实验已将 AES 与 Camellia，DES 与 3DES 划分到了对应的聚类当中，虽然采用上述方式确定了 FT、RT 及 FTB 测试的特征向量维数，但同一聚类中的加密算法间其特征向量间的相似性仍然较高，且在多组实验过程中，仍无法得到满意的实验结果。在 FT 测试向量维数为 300，RT 测试的向量维数为 10 000 及 FTB 测试的向量维数为 1 000 条件下(AES 与 Camellia、DES 与 3DES 的特征向量间的相似性较小)，分别对 AES、Camellia、DES 及 3DES 的特征向量 $\Delta\boldsymbol{T}\tilde{\boldsymbol{V}}_{AES}$、$\Delta\boldsymbol{T}\tilde{\boldsymbol{V}}_{Camellia}$、$\Delta\boldsymbol{T}\tilde{\boldsymbol{V}}_{DES}$ 及 $\Delta\boldsymbol{T}\tilde{\boldsymbol{V}}_{3DES}$ 进行了提取，得

$$\begin{aligned}
&\left(\lambda_{\Delta\tilde{V}_{AES}\Delta\tilde{V}_{Camellia}-FT},\overline{\lambda}_{\Delta\tilde{V}_{AES}\Delta\tilde{V}_{Camellia}-FT}\right)=(0.741\,5,0.258\,5)\\
&\left(\lambda_{\Delta\tilde{V}_{AES}\Delta\tilde{V}_{Camellia}-RT},\overline{\lambda}_{\Delta\tilde{V}_{AES}\Delta\tilde{V}_{Camellia}-RT}\right)=(0.625\,5,0.374\,5)\\
&\left(\lambda_{\Delta\tilde{V}_{AES}\Delta\tilde{V}_{Camellia}-FTB},\overline{\lambda}_{\Delta\tilde{V}_{AES}\Delta\tilde{V}_{Camellia}-FTB}\right)=(0.634\,8,0.365\,2)
\end{aligned}\tag{12}$$

$$\begin{aligned}
&\left(\lambda_{\Delta\tilde{V}_{DES}\Delta\tilde{V}_{3DES}-FT},\overline{\lambda}_{\Delta\tilde{V}_{DES}\Delta\tilde{V}_{3DES}-FT}\right)=(0.729\,1,0.270\,9)\\
&\left(\lambda_{\Delta\tilde{V}_{DES}\Delta\tilde{V}_{3DES}-RT},\overline{\lambda}_{\Delta\tilde{V}_{DES}\Delta\tilde{V}_{3DES}-RT}\right)=(0.647\,0,0.353\,0)\\
&\left(\lambda_{\Delta\tilde{V}_{DES}\Delta\tilde{V}_{3DES}-FTB},\overline{\lambda}_{\Delta\tilde{V}_{DES}\Delta\tilde{V}_{3DES}-FTB}\right)=(0.609\,6,0.390\,4)
\end{aligned}\tag{13}$$

4.4　AES/Camellia/DES/3DES 识别结果及分析

1) AES 与 Camellia 的识别结果及分析

在 FT 测试向量维数为 300，RT 测试的向量维数为 10 000 及 FTB 测试的向量维数为 1 000 条件下，根据式(10)及式(12)可获得 $\alpha=0.259\,0$，$\beta=0.375\,1$ 及 $\gamma=0.365\,9$，实验结果如图 4 所示。在 10 次实验过程中，正确识别 AES 及 Camellia 算法的次数均为 9 次。

2) DES 与 3DES 的识别结果及分析

DES/3DES 算法识别过程中，根据式(10)及式(13)可获得 $\alpha=0.267\,1$，$\beta=0.348\,0$ 及 $\gamma=0.384\,9$，实验结果如图 5 所示。在 10 次实验过程中，正确识别 DES 及 3DES 算法的次数分别为 10 次和 8 次。

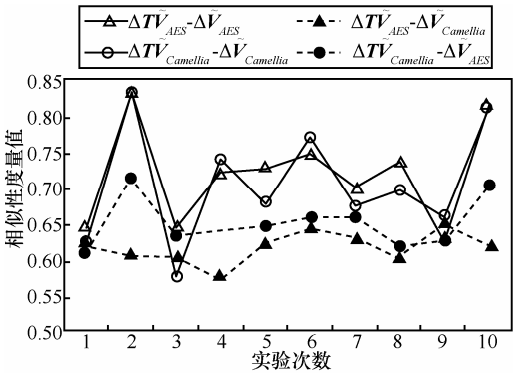


图 4　AES 与 Camellia 的识别结果

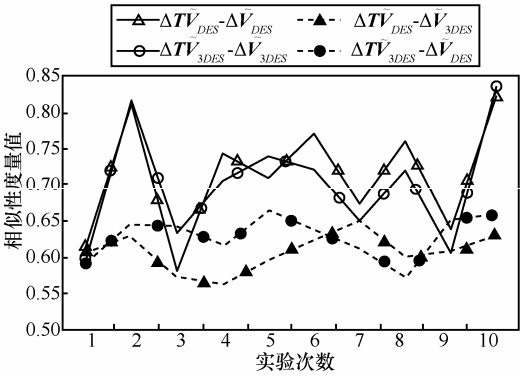


图 5　DES 与 3DES 的识别结果

重复进行 100 次实验，正确识别 AES 与 Camellia 的概率分别为 88%和 89%，正确识别 DES 与 3DES 的概率则分别为 91%和 86%。提出的密文随机性度量值分布特征提取及其相似性度量方案均能以接近 90%的概率实现对 AES 与 Camellia、DES 与 3DES 的识别。

3) FT、RT、FTB 测试的向量维数对识别结果的影响分析

为进一步论证合理的特征向量维数，对于分组密码算法识别有效性。在 FT 测试向量维数为 30，RT 测试的向量维数为 1 000 及 FTB 测试的向量维数为 100 条件下，分别测试了提出方案对 AES、Camellia、DES 及 3DES 的识别率。AES 与 Camellia、DES 与 3DES 的识别结果分别如图 6 和图 7 所示。如图 6 所示，10 次实验正确识别 AES 及 Camellia 的次数均为 6 次。图 7 中 10 次实验正确识别 DES 及 3DES 的次数均为 5 次。重复进行 100 次实验，正确识别 AES 与 Camellia 的概率分别为 59%和 60%，正确识别 DES 与 3DES 的概率分别为 51%和 47%。

同图 4 和图 5 所示的结果相比，当 FT 测试的向量维数降低为 30，RT 测试的向量维数降低为 1 000 及 FTB 测试的向量维数降低为 100 时，正

确识别 AES 与 Camellia、DES 与 3DES 的概率显著降低。原因在于：密码算法密文的 FT、RT、FTB 测试值的向量维数降低时，随机性度量值区间$[0,1]$的划分数也随之减少，各划分区间内样本的累积效应导致向量间的差异性减小，即出现特征向量间的距离趋近现象，导致其可区分度降低，正确识别密码算法的概率也随之降低。以上实验结果表明，不同的特征向量维数会对正确识别密码算法的概率产生较大影响。从确保特征向量匹配速度的角度来看，过大的特征向量维数在增加特征向量建立过程计算量及较多无效数据项的同时，还会降低向量相似性度量的速度，前面已给出确定特征向量维数的方法，不再赘述。

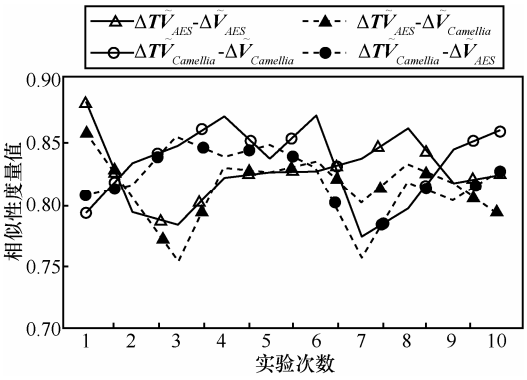


图 6 AES 与 Camellia 的识别结果

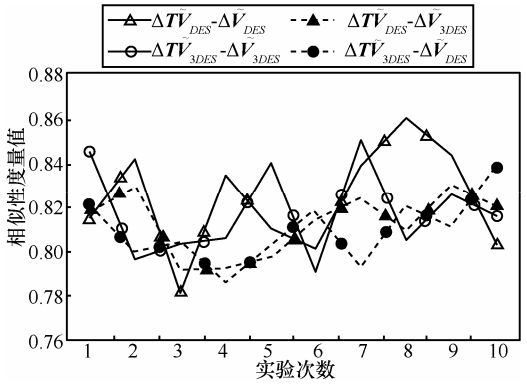


图 7 DES 与 3DES 的识别结果

5 结束语

针对当前基于二进制代码分析等方法识别加密算法存在需进行加密程序逆向分析、适用性不强等问题。在已知密文条件下，以典型分组密码算法 AES、Camellia、DES、3DES 及 SMS4 为识别分析对象，以码元频数检测、块内频数检测、游程检测为密文随机性度量手段，提出了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。基于 OpenSSL 密码库，在 VC 6.0 及 Matlab 环境下进行了实验仿真。结果表明，在一定的样本量条件下，SMS4 密文的随机性度量值取值个数均明显小于 AES、Camellia、DES、3DES 密文的随机性度量值取值个数，基于密文的随机性度量值取值个数的初始聚类过程即可识别出 SMS4 算法。由于 AES 与 Camellia、DES 与 3DES 密文的随机性度量值具有相似的取值个数，AES 与 Camellia 被划分相同的聚类中，而 DES 与 3DES 则被划分到另一聚类中。通过进一步提取 AES、Camellia、DES、3DES 密文的随机性度量值分布特征向量，成功识别 AES、Camellia、DES、3DES 的概率均接近 90%。后续研究工作将进一步评估密文样本量对识别率的影响及验证提出方法对其他密码算法的识别能力。

参考文献：

[1] SPILLMAN R, JANSSEN M, NELSON B, *et al*. Use of a genetic algorithm in the cryptanalysis of simple substitution ciphers[J]. Cryptologia, 1993, 17(1): 31-44.

[2] RAMZAN Z. On Using Neural Networks to Break Cryptosystems[R]. Laboratory of Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139, 1998.

[3] DILEEP A D, SEKHAR C C. Identification of block ciphers using support vector machines[A]. Proceeding of the 2006 International Joint Conference on Neural Networks[C]. Vancouver, Canada, 2006. 2696-2701.

[4] MELTEM S T, ÇAĞDAŞ Ç, NURDAN B S, *et al*. New distinguishers based on random mappings against stream ciphers [A]. Proceeding of the 5th International Conference Lexington[C]. KY, USA, 2008. 30-41.

[5] 陈华,冯登国,范丽敏. 一种关于分组密码的新的统计检测方法[J]. 计算机学报, 2009,32(4): 595-601.
CHEN H, FENG D G, FAN L M. A new statistical test on block ciphers[J]. Chinese Journal of Computers, 2009, 32(4): 595-601.

[6] LIU T M, JIANG L H, HE H Q, *et al*. Researching on cryptographic algorithm recognition based on static characteristic-code[A]. Proceeding of the Future Generation Information Technology Conference[C]. Jeju Island, Korea, 2009.140-147.

[7] MANJULA R, ANTITHA R. Identification of encryption algorithm using decision tree[A]. Proceeding of the First International Conference on Computer Science and Information Technology[C]. Bangalore, India, 2011.237-246.

[8] GRÖBERT F, WILLEMS C, HOLZ T. Automated identification of cryptographic primitives in binary programs[A]. Proceeding of the 14th International Symposium[C]. Menlo Park, CA, USA, 2011. 41-60.

[9] 谷利泽, 郑世慧, 杨义先. 现代密码学教程[M]. 北京: 北京邮电大学出版社, 2009.
GU L Z, ZHENG S H, YANG Y X. Tutorial of the Modern Cryptography[M]. Beijing: Beijing University of Posts and Telecommunica-

tions Press, 2009.

[10] RIVAIN M. Differential fault analysis on DES middle rounds[A]. Proceeding of the 11th International Workshop Lausanne[C]. Switzerland, 2009.457-469.

[11] SAHA D, MUKHOPADHYAY D, ROY C D. A diagonal fault attack on the advanced encryption standard[EB/OL]. http://eprint.iacr.org/2009/581, 2009.

[12] 赵新杰, 王韬, 郭世泽. 一种针对Camellia的改进差分故障分析[J]. 计算机学报, 2011, 34(4): 613-627.
ZHAO X J, WANG T, GUO S Z. An improved differential fault analysis on camellia[J]. Chinese Journal of Computers, 2011, 34(4): 613-627.

[13] 赵新杰, 王韬, 郑媛媛. 针对SMS4密码算法的Cache计时攻击[J]. 通信学报, 2010, 31(6): 89-98.
ZHAO X J, WANG T, ZHENG Y Y. Cache timing attack on SMS4[J]. Journal on Communications, 2010, 31(6): 89-98.

[14] 胡磊, 王鹏等. 应用密码学手册[M]. 北京: 电子工业出版社, 2005.
HU L, WANG P, *et al*. Handbook of Applied Cryptography[M]. Beijing: Publishing House of Electronics Industry, 2005.

[15] RUKHIN A, SOTO J, NECHVATAL J. A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications[R]. SP 800-22, 2001.

[16] FILIOL E. A new statistical testing for symmetric ciphers and hash functions[A]. Proceeding of Information and Communications Security: 4th International Conference[C]. Singapore, 2002.342-353.

[17] 胡俭勇, 苏锦海. 一种随机性实时检测方案[J]. 计算机工程, 2009, 35(9): 136-138.
HU J Y, SU J H. Scheme for real time test of randomness [J]. Computer Engineering, 2009, 35(9): 136-138.

[18] FIPS PUB 140-2-2001 Security Requirements for Cryptographic Modules[S]. Washington, USA: National Institute of Standards and Technology, 2001.

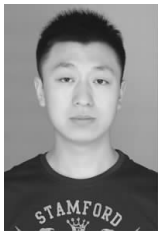
[19] Special Publication 800-22 Revision 1a[S]. Washington, USA: National Institute of Standards and Technology, 2010.

作者简介：

吴杨（1985-），男，四川成都人，军械工程学院博士生，主要研究方向为网络协议识别、网络安全技术、模式识别等。

王韬（1964-），男，河北石家庄人，军械工程学院教授、博士生导师，主要研究方向为网络协议识别、网络安全技术、密码安全技术等。

邢萌（1990-），女，河南濮阳人，军械工程学院硕士生，主要研究方向为网络数据分析。



李进东（1990-），男，新疆石河子人，军械工程学院硕士生，主要研究方向为网络数据分析。