22－轮SMS4的差分分析*

张美玲，刘景美，王新梅
（西安电子科技大学综合业务网国家重点实验室，陕西 西安 710071）

摘 要：SMS4是中国官方公布的第一个商用分组密码标准，使用差分方法分析了18轮的SMS4差分特征，并在此基础上攻击了22－轮的SMS4，攻击过程需要2^{117}个选择明文，2^{112}字节的存储空间，而时间复杂度为2^{123}次22－轮加密。此结果是目前对SMS4差分分析的最好结果。

关键词：SMS4；差分分析；时间复杂度；数据复杂度
中图分类号：TP309 **文献标志码**：A **文章编号**：0529－6579（2010）02－0043－05

Differential Attack on 22-Round SMS4 Block Cipher

ZHANG Meiling, *LIU Jingmei*, *WANG Xinmei*
(National Key Lab. of Integrated Service Networks, Xidian University, Xi'an 710071, China)

Abstract: SMS4 is a 128-bit block cipher used in WAPI, the Chinese WLAN national standard. A new 18-round differential characteristic of SMS4 is presented, basing on which the 22-round SMS4 is attacked. The attack requires 2^{117} chosen plaintexts, 2^{112} bytes of memory and 2^{123} 22-round encryptions. And the attack is the best differential cryptanalytic result on SMS4.

Key words: SMS4; differential cryptanalysis; time complexity; data complexity

SMS4[1]是中国国家标准针对目前数字媒体所面临的信息安全问题提出来的运用于无线局域网的一种分组密码标准。SMS4的加密算法和密钥扩展算法都采用了32轮非线性迭代结构，且加解密结构一致，只是轮密钥的使用顺序相反，解密轮密钥是加密轮密钥的逆序。

目前对SMS4算法的分析相对较少，主要有整体攻击[2]，线性攻击[3]，差分攻击[3-4]，不可能差分攻击[5-6]，矩形攻击[3-4]，飞来器攻击[3]，边信道攻击[7]。其中差分分析[8-9]是分析分组密码最有效的方法之一。表1列出了以往对SMS4的各种传统攻击方法最好的结果以及本文的结果。

表1 SMS4的攻击结果汇总

Table 1 Summary of attacks on SMS4

攻击类型	轮数	数据复杂度	存储空间（bytes）	时间复杂度
Integral[2]	13	2^{16} CP	2^{20}	2^{114} Enc.
Differential[3]	22	2^{118} CP	2^{123}	$2^{125.71}$ Enc.
Linear[3]	22	2^{117} KP	2^{109}	$2^{109.86}$Enc. $+2^{120.39}$ A. O.
Boomerang[3]	18	2^{120} ACPS	2^{123}	$2^{116.83}$ Enc.
Rectangle[3]	18	2^{124} CP	2^{128}	$2^{112.83}$ Enc.
Impossible differential[6]	17	2^{103} CP	2^{89}	2^{124} Enc
Differential（this paper）	22	2^{117} CP	2^{112}	2^{123} Enc.

KP－已知明文；CP－选择明文；ACPS－自适应选择明文和密文；Enc. －加密单元；A. O. －算术操作。

* **收稿日期**：2009－03－07
基金项目：国家自然科学基金资助项目（60773002，60903199）；863项目资助项目（2007AA01Z472）；高等学校创新引智基地资助项目（B08038）；信息安全国家重点实验室开放课题，ISN重点实验室开放课题资助项目（ISN10－11）
作者简介：张美玲（1982年生），女，博士生；E-mail：zhangml21@yahoo.cn

1 SMS4 算法简介

1.1 基本记号与符号

Z_2^e：e-比特的向量集，Z_2^{32} 中的元素称为字，Z_2^8 中的元素称为字节。

Sbox（·）：S 盒，固定的 8 比特输入 8 比特输出的置换。

⊕：32 比特的逐比特异或。

$<<<i$：32 比特循环左移 i 位。

*：任意的 32 位的字；N 为任意的 8 位的字节。

$(\varepsilon)_i$：取 ε（$\varepsilon \in Z_2^{32}$）的第 i 个字节，（$i=0,1,2,3$）。

$|x|$：计算 x 的个数。

$(X_i, X_{i+1}, X_{i+2}, X_{i+3}) \in (Z_2^{32})^4$：第 i 轮的输入，（$i=0,1,\cdots,31$）。

RK_i：第 i 轮的字密钥。

$C_{i,k}$：为 C_i 的第 k 个字节，（$k=0,1,2,3$）。

$_T(\alpha\rightarrow\beta)$：输入差分为 α，经过轮函数 T 后，输出差分为 β，$\alpha, \beta \in Z_2^{32}$。

$(T)_{in}$，$(T)_{out}$：分别为函数 T 的输入差分和输出差分。

Λ_α：当输入差分为 α 时，所有可能的且第一字节为 0 的输出差分的集合。

1.2 SMS4 的加密算法简介

SMS4 是一个具有 32 轮非平衡结构的 Feistel 型密码，其分组长度和密钥长度均为 128 比特。设明文 P 为 $(X_0, X_1, X_2, X_3) \in (Z_2^{32})^4$，$(Y_0, Y_1, Y_2, Y_3) \in (Z_2^{32})^4$ 为明文 P 所对应的密文 C。$RK_i \in Z_2^{32}$ 是第 i（$i=0,1,\cdots,31$）轮的子密钥。SMS4 的加密过程如下（如图 1 所示）：

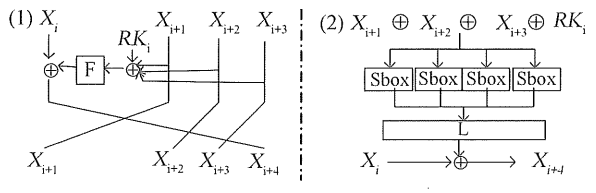

图 1 （1）SMS4 第 i 轮的轮函数，（2）F 函数

Fig. 1 （1）Round function of the i^{th} round，（2）F function

1）输入明文 $P=(X_0, X_1, X_2, X_3)$，

2）For i（$=0,1,\cdots,31$）

$-X_{i+4}=F(X_i, X_{i+1}, X_{i+2}, X_{i+3}, RK_i) = X_i \oplus T(X_{i+1} \oplus X_{i+2} \oplus X_{i+3} \oplus RK_i)$，

3）输出密文 $C=(Y_0, Y_1, Y_2, Y_3) = R(X_{32}, X_{33}, X_{34}, X_{35}) = (X_{35}, X_{34}, X_{33}, X_{32})$。

其中 R 为反序变换，而变换 T 包含了一个非线性变换 S 和一个线性混淆函数 L，即 $T(\cdot) = L(S(\cdot))$。S 变换是将输入的 32 比特分成 4 个字节，并行查 S-box 表（8－进 8－出），然后将输出的 4 个字节合并为一个字作为 L 变换的输入。而线性混淆函数 L：令 $B \in Z_2^{32}$ 和 $C \in Z_2^{32}$ 分别为 L 的输入和输出，则有：$C=L(B) = B \oplus (B<<<2) \oplus (B<<<10) \oplus (B<<<18) \oplus (B<<<24)$；$L$ 的逆变换 L^{-1} 为：$B=L^{-1}(C) = C \oplus (C<<<2) \oplus (C<<<4) \oplus (C<<<8) \oplus (C<<<12) \oplus (C<<<14) \oplus (C<<<16) \oplus (C<<<18) \oplus (C<<<22) \oplus (C<<<24) \oplus (C<<<30)$。

SMS4 的密钥扩展算法与加密算法基本一样，只在线性混淆函数 L 上有一些差别，密钥扩展算法中的线性函数是：$L'(B) = B \oplus (B<<<13) \oplus (B<<<23)$，（详细过程见文［1］）。

2 22－轮的 SMS4 的差分攻击

在这一节中给出了一个新的 18 轮的 SMS4 的差分特征，当遍历所有可能的 2^{13} 个 α，差分链概率的总和约为 $2.700\ 58 \times 10^{-34}$，则平均差分概率为：$2.700\ 58 \times 10^{-34}/2^{13} \approx 2^{-124.5}$，并在此基础上攻击了 22 轮的 SMS4，如图 2 所示。

2.1 新的 18 轮的 SMS4 的差分特征

令 $\alpha, \beta \in Z_2^{32} \backslash \{0\}$ 为 32 比特的非零差分，且 $(\alpha)_0=0$，即 α 形如（0，N，N，N），其中 $N \in Z_2^8 \backslash \{0\}$。选择加密算法的初始输入差分为（$\alpha$，$\alpha$，$\alpha$，0），那么 $(T)_{in}$ 为 0，从而以概率 1 得到下一轮的差分为（α，α，0，α）；同理可以概率 1 得到第 1 轮和第 2 轮的输出差分（α，0，α，α）、（0，α，α，α）；第 3 轮的 $(T)_{in}$ 为 α，设 $(T)_{out}$ 为 β，则此轮的输出差分为（α，α，α，β）；第 4 轮的 $(T)_{in}$ 为 β，设 $(T)_{out}$ 仍为 β，则此轮的输出差分为（α，α，β，$\alpha \oplus \beta$）；第 5、6 轮的 $(T)_{in}$ 为 0，所以轮输出差分分别为（α，β，$\alpha \oplus \beta$，α）和（β，$\alpha \oplus \beta$，α，α）；第 7 轮的 $(T)_{in}$ 为 $\alpha \oplus \beta$，设 $(T)_{out}$ 仍为 $\alpha \oplus \beta$，则此轮的输出差分为（$\alpha \oplus \beta$，α，α，α）；第 8 轮的 $(T)_{in}$ 为 α，设 $(T)_{out}$ 为 β，则此轮的输出差分为（α，α，α，α）；第 9 轮的 $(T)_{in}$ 为 α，设 $(T)_{out}$ 为 α，则此轮的输出差分为（α，α，α，0）；从第 10 轮到第 17 轮重复第 0 轮到第 7 轮的过程；设第 18 轮的输出差分为（α，α，α，θ）；接下来的三轮的输出差分设为（α，α，θ，

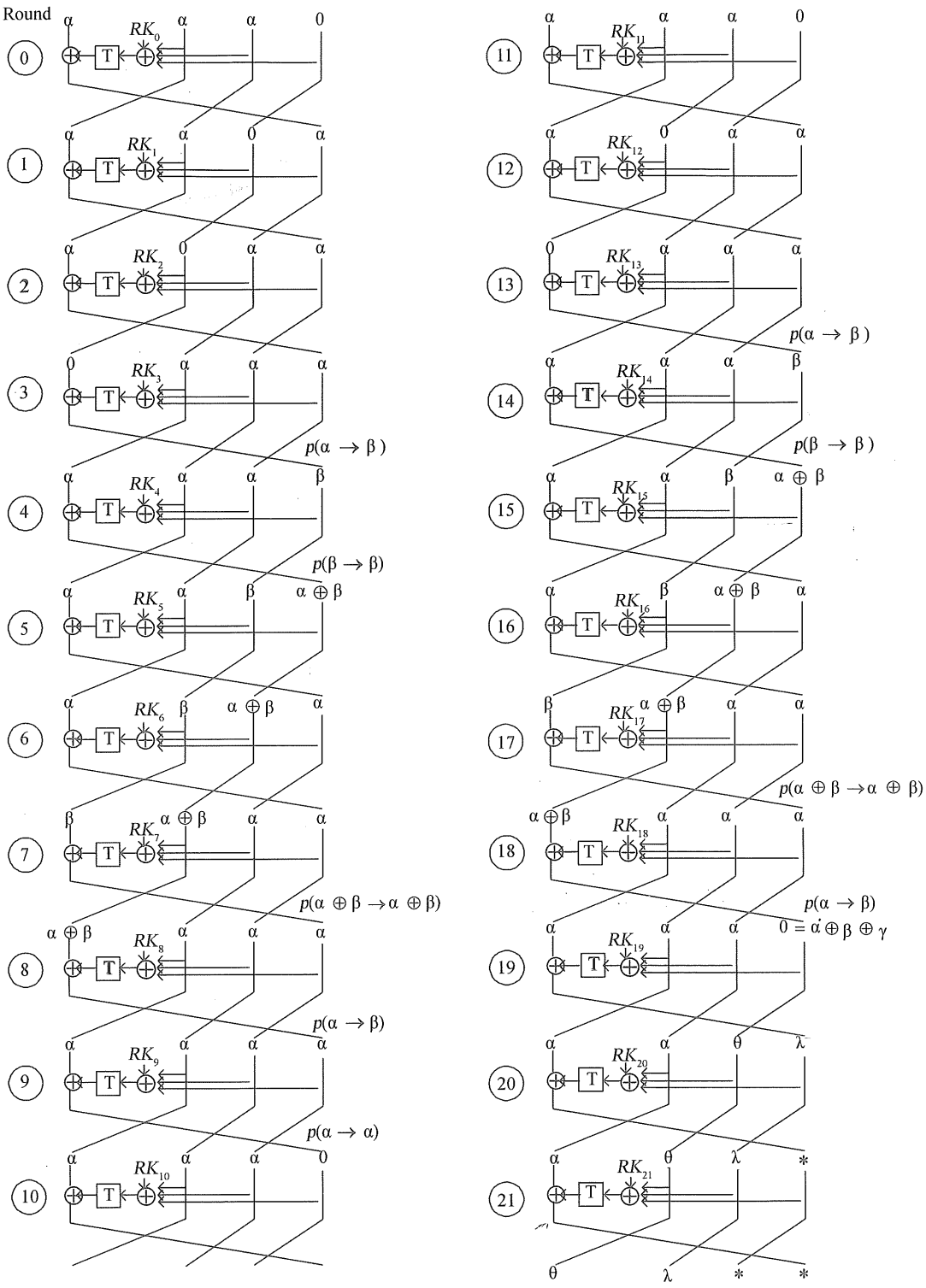

图 2　SMS4 22 轮的差分分析

Fig. 2　Differential attack on 22-round SMS4

λ)、(α, θ, λ, *) 和 (θ, λ, *, *)，其中“*”表示未知的差分值。

性质 1：对于 SMS4 的 S 盒，每一个非零输入差分，均有 127 种输出差分，其中一个出现的概率是 2^{-6}，另外 126 个出现的概率均为 2^{-7}。

定义 1：(分支数 [10]) 令 $W(\cdot)$ 为计算字节重量函数，即计算非零字节的个数。那么线性变换 L ($L: Z_2^{32} \to Z_2^{32}$) 的差分分支数为：

$\min_{\alpha \neq 0, \alpha \in Z_2^{32}} (W(\alpha) + W(L(\alpha)))$

性质 2：SMS4 中的线性变换 L 的分支数为 5。

性质1与性质2很容易经过实验得出。

由性质2可知[4]，若函数 T 的输入和输出差分均为 α 时，由于 S 盒的输入差分与输出差分的非零字节数相等，那么 $(W(\alpha)+W(\alpha))$ 至少为5，则 α 的非零字节数至少为3。所以当 $(\alpha)_0=0$ 时，α 的其它三个字节必然为非零值，且有 $|T(\alpha\rightarrow\alpha)|\approx 2^{13}$。

设 $Diff=\{(\alpha,\beta)\mid \alpha\rightarrow\alpha,\ \alpha\rightarrow\beta,\ \beta\rightarrow\beta,\ \alpha\oplus\beta\rightarrow\alpha\oplus\beta,\ (\alpha)_0=(\beta)_0=0\}$。实验数据显示 $|Diff|\approx 2^{20}$，$|\alpha\approx 2^{13}$，所以对每一个 α 平均约有 2^7 个对应的 β，而对每一个 α 整个差分链的成立的概率平均约为 $2^{-124.5}$。

2.2 22－轮差分攻击过程

差分攻击的思想：首先设计一个高概率的差分路径（如图2），其次寻找符合此路径的明文对（只要存在有轮密钥使得此对满足差分路径就保留此“对”），然后用所有剩下的“对”去过滤密钥（正确的密钥会使更多的“对”满足差分路径）。具体过程如下：

1）选择明文：将第0，4，8，12，13，14，15字节（共56比特）固定，遍历其余的字节（共72比特），即共有 2^{72} 个明文，$(2^{72})^2/2=2^{143}$ 对明文对，并将这些明文集合称为一个结构。

2）选取 2^{45} 个结构，则共有 $2^{188}=(2^{45}\cdot 2^{143})$ 对明文对，这些明文对的差分形如：$((0,N,N,N),(0,N,N,N),(0,N,N,N),(0,0,0,0))$。

3）挑选满足明文对差分是 $(\alpha,\alpha,\alpha,0)$，且 α 是选自 $Diff$ 中的 α 的那些对，其中 $|\alpha|\approx 2^{13}$。满足这样的对约有 $2^{188}(2^{13}/2^{72})=2^{129}$。

4）加密剩下的对，得到对应的密文对。选择在第一个字的差分为 $\alpha\oplus\beta\oplus\gamma$ 的那些对，其中 $\gamma\in\Lambda_\alpha$。实验数据得到 $|\Lambda_\alpha|\approx 2^{13}$，所以剩下的对约有 $2^{129}\cdot 2^7\cdot 2^{13}/2^{32}=2^{117}$。

5）猜测密钥 RK_{21} 的第0个字节 $RK_{21,0}$，设密文对为 $(C_{22},C_{23},C_{24},C_{25})$ 和 $(Y_{22},Y_{23},Y_{24},Y_{25})$，计算 $\tau=\text{Sbox}(C_{22,0}\oplus C_{23,0}\oplus C_{24,0}\oplus RK_{21,0})\oplus\text{Sbox}(Y_{22,0}\oplus Y_{23,0}\oplus Y_{24,0}\oplus RK_{21,0})$ 和 $v=L^{-1}(C_{25}\oplus Y_{25}\oplus\alpha)$，测试 τ 和 $(v)_0$ 是否相等，若不是，则删去此对，留下的概率为 2^{-8}，所以剩下的对约有 $2^{117}\cdot 2^{-8}=2^{109}$。

6）猜测密钥 RK_{21} 的第 k（$k=1,2,3$）个字节 $RK_{21,k}$，计算 $\tau=\text{Sbox}(C_{22,k}\oplus C_{23,k}\oplus C_{24,k}\oplus RK_{21,k})\oplus\text{Sbox}(Y_{22,k}\oplus Y_{23,k}\oplus Y_{24,k}\oplus RK_{21,k})$，测试 τ 与 $(v)_k$ 是否相等，若不是，则删去此对。执行完此步后所剩的对约有 $2^{109}\cdot(2^{-8})^3=2^{85}$。且通过猜测 RK_{21} 可得 C_{21} 和 Y_{21}。

7）猜测密钥 RK_{20} 的第0个字节 $RK_{20,0}$，计算 $\tau=\text{Sbox}(C_{21,0}\oplus C_{22,0}\oplus C_{23,0}\oplus RK_{20,0})\oplus\text{Sbox}(Y_{21,0}\oplus Y_{22,0}\oplus Y_{23,0}\oplus RK_{20,0})$ 和 $v=L^{-1}(C_{24}\oplus Y_{24}\oplus\alpha)$，测试 τ 和 $(v)_0$ 是否相等，若不是，则删去此对，留下的概率为 2^{-8}，所以剩下的对约有 $2^{85}\cdot 2^{-8}=2^{77}$。

8）猜测密钥 RK_{20} 的第 k（$k=1,2,3$）个字节 $RK_{20,k}$，计算 $\tau=\text{Sbox}(C_{21,k}\oplus C_{22,k}\oplus C_{23,k}\oplus RK_{20,k})\oplus\text{Sbox}(Y_{21,k}\oplus Y_{22,k}\oplus Y_{23,k}\oplus RK_{20,k})$，测试 τ 与 $(v)_k$ 是否相等，若不是，则删去此对。执行完此步后所剩的对约有 $2^{77}\cdot(2^{-8})^3=2^{53}$，并得到 C_{20} 和 Y_{20}。

9）猜测密钥 RK_{19} 的第0个字节 $RK_{19,0}$，计算 $\tau=\text{Sbox}(C_{20,0}\oplus C_{21,0}\oplus C_{22,0}\oplus RK_{19,0})\oplus\text{Sbox}(Y_{20,0}\oplus Y_{21,0}\oplus Y_{22,0}\oplus RK_{19,0})$ 和 $v=L^{-1}(C_{23}\oplus Y_{23}\oplus\alpha)$，测试 τ 和 $(v)_0$ 是否相等，若不是，则删去此对，留下的概率为 2^{-8}，所以剩下的对约有 $2^{53}\cdot 2^{-8}=2^{45}$。

10）猜测密钥 RK_{19} 的第 k（$k=1,2,3$）个字节 $RK_{19,k}$，计算 $\tau=\text{Sbox}(C_{20,k}\oplus C_{21,k}\oplus C_{22,k}\oplus RK_{19,k})\oplus\text{Sbox}(Y_{20,k}\oplus Y_{21,k}\oplus Y_{22,k}\oplus RK_{19,k})$，测试 τ 与 $(v)_k$ 是否相等，若不是，则删去此对。执行完此步后所剩的对约有 $2^{45}\cdot(2^{-8})^3=2^{21}$，并得到 C_{19} 和 Y_{19}。

11）猜测密钥 RK_{18} 的第 k（$k=1,2,3$）个字节 $RK_{18,k}$，计算 $\tau=\text{Sbox}(C_{19,k}\oplus C_{20,k}\oplus C_{21,k}\oplus RK_{18,k})\oplus\text{Sbox}(Y_{19,k}\oplus Y_{20,k}\oplus Y_{21,k}\oplus RK_{18,k})$ 和 $v=\text{L}^{-1}(C_{22}\oplus Y_{22}\oplus\alpha\oplus\beta)$，测试 τ 与 $(v)_k$ 是否相等，若不是，则删去此对。对于每一个 α，其平均约有 2^7 个对应的 β，所以 v 的个数也约有 2^7 个。则执行完此步后所剩的对约有 $2^{21}\cdot 2^7/2^{24}=16$。

2.3 差分攻击的复杂度分析

由以上分析可知，对于错误密钥，剩下的密文约有16对，另一方面，若猜测的密钥是对的，那么最后剩下的密文对约有 $2^{129}\cdot 2^{-124.5}\approx 23$ 对。即对于所猜测的密钥，若剩下的密文对大于或等于23对时，则认为所猜测的密钥 $RK_{18,1}$，$RK_{18,2}$，$RK_{18,3}$，RK_{19}，RK_{20}，RK_{21} 是正确的。剩下的8比特的密钥 $RK_{18,0}$ 通过穷搜索的方法得到。

此攻击的数据复杂度是需要 2^{117} 个明文和 2^{122}（$=2^{117}\cdot 16\cdot 2$）字节的存储空间来存储明密文

对。而攻击的时间复杂度主要依赖于第 5 步，第 7 步，第 9 步和第 11 步，具体分析如下：

第 5 步，猜测 8 比特的 $RK_{21,0}$，需要部分解密 2^{117} 个密文对，时间复杂度约为 $2^{119.54}$（$\approx 2^{117} \cdot 2 \cdot 2^8 \cdot (1/22) \cdot (1/4)$）；第 6 步的每一次猜测密钥都是相互独立的，与第 5 步的猜测密钥也是相互独立的，所以其攻击复杂度可以忽略不计；第 7 步需要部分解密 2^{85} 个密文对，时间复杂度也为 $2^{119.54}$；同理可得第 9 步的时间复杂度；第 11 步 $k=1$ 时，需要部分解密 2^{21} 个密文对，时间复杂度同为 $2^{119.54}$ 次 22 - 轮加密，以及 2^{132}（$=2^{21} \cdot 2^{96+8} \cdot 2^7$）次异或比较，其时间复杂度约等同于 2^{122} 次 22 - 轮加密。所以总的时间复杂度约为 2^{123} 次 22 - 轮加密。

3　总　结

我们提出了一个新的 18 轮 SMS4 的差分特征，并在此基础上攻击了 22 轮的 SMS4。其数据复杂度为 2^{117} 个选择明文和约 2^{122} 字节的存储空间，以及 2^{123} 次 22 - 轮加密的时间复杂度，此结果优于以往对 SMS4 的差分攻击。对一个具体的密码体制，如何选择一条高概率的差分路径、其差分路径的选择是否带有某种规律是我们下一步的目标。

参考文献：

[1] Specification of SMS4, block cipher for WLAN products-SMS4 (in Chinese) [EB/OL]. http://www.oscca.gov.cn/UpFile/200621016423197990.pdf

[2] LIU F, JI W, HU L, et al. Analysis of the SMS4 block cipher [C]. Proceeding of ACISP'07, Springer-Verlag, 2007, 4586: 158 - 170.

[3] KIM T, KING J, HONG S, et al. Linear and differential cryptanalysis of reduced SMS4 block cipher. Cryptology ePrint Archive: Report 2008/281, 2008.

[4] ZHANG L, ZHANG W, WU W. Cryptanalysis of reduced-round SMS4 block cipher [C]. Proceedings of ACISP'08, Lecture Notes in Computer Science, Springer-Verlag, 2008, 5107: 216 - 229.

[5] LU J. Attacking reduced-round versions of the SMS4 block cipher in the chinese WAPI standard [C]. Proceedings of ICICS'07, Springer-Verlag, 2007, 4861: 306 - 318.

[6] 陈杰，胡予濮，张跃宇. 用不可能差分法分析 17 轮 SMS4 算法 [J]. 西安电子科技大学学报：自然科学版，2008，35(3)：455 - 458.

[7] 张蕾 吴文玲. SMS4 密码算法的差分故障攻击[J]. 计算机学报，2006，29(9)：1594 - 1600.

[8] BIHAM E, SHAMIR A. Differential cryptanalysis of DES-like cryptosystem (extended abstract) [C]//MENEZES A, VANSTONE S A (eds.) CRYPTO 1990, Springer, Heidelberg, 1991, 537: 2 - 21.

[9] BIHAM E, SHAMIR A. Differential cryptanalysis of the data encryption standard [C]. Springer, Heidelberg, 1993.

[10] DAEMEN J, RIJMEN V. AES: The advanced encryption standard. [EB/OL]. http://www.nist.gov/aes.

（上接第 42 页）

[18] WARD J J, MCGUFFIN L J, BUXTON B F, et al. Secondary structure prediction with support vector machines [J]. Bioinformatics, 2003, 19(13): 1650 - 1655.

[19] GUO J, CHEN H, SUN Z, LIN Y. A novel method for protein secondary structure prediction using dual-layer SVM and profiles [J]. Proteins, 2004, 54(4): 738 - 743.

[20] VAPNIK V N. Statistical learning theory [M]. New York: Wiley, 1998.

[21] VAPNIK V N. The nature of statistical learning theory [M]. New York: Springer-Verlag, 1995.

[22] STATNIKOV A, ALIFERIS C F, TSAMARDINOS I, et al. A comprehensive evaluation of multicategory classification methods for microarray gene expression cancer diagnosis [J]. Bioinformatics, 2005, 21(5): 631 - 643.

[23] LIU Y H, CHEN Y T. Face recognition using total margin-based adaptive fuzzy support vector machines [J]. IEEE Transactions on Neural Networks, 2007, 1(18): 178 - 192.

[24] HO S Y, SHU L S, CHEN J H. Intelligent evolutionary algorithms for large parameter optimization problems [J]. IEEE Trans Evolutionary Comput, 2004, 8(6): 522 - 541.

[25] CAMPBELL C. Kernel methods: a survey of current techniques [J]. Neurocomputing, 2002, 48 (1 - 4): 63 - 84.

[26] HSU C W, LIN C J. A comparison of methods for multiclass support vector machines [J]. IEEE Transaction on Neural Networks, 2002, 13(2): 415 - 425.

[27] Gems-systems [EB/OL]. http://www.gems-systems.org