doi:10.11959/j.issn.1000-436x.2016108

简化版 Trivium 算法的线性逼近研究

马猛，赵亚群
（信息工程大学数学工程与先进计算国家重点实验室，河南 郑州 450001）

摘 要：针对初始化轮数为 288 个时钟的简化版 Trivium 算法（又称 2 轮 Trivium）进行了线性逼近研究，设计了搜索最佳线性近似式算法，并通过对第 1 轮关于密钥、初始化向量和密钥流比特的表达式做非线性逼近，结合该算法，在同等条件下给出了 2 轮 Trivium 16 个偏差为 $2^{-23.42}$ 的线性近似式，使通过多线性攻击去识别 2 轮 Trivium 的一个具有特定比特的密钥所需要的数据量降为 $2^{42.84}$ 个选择 ***IV***，为 Turan 方案所需数据量的 $\frac{1}{2^{19.16}}$，且成功率保持不变。

关键词：流密码；Trivium 算法；多线性密码分析；线性逼近
中图分类号：TN918.1 **文献标识码**：A

Research on linear approximations of simplified Trivium

MA Meng, ZHAO Ya-qun
(State Key Lab of Mathematical Engineering and Advanced Computing, Information Engineering University, Zhengzhou 450001, China)

Abstract: The linear approximations of simplified Trivium with the initialization of 288 clocks (2-round Trivium) was studied. An algorithm was designed to search optimal linear approximations. Moreover, a method was presented to conduct a linear approximation of 2-round Trivium by approximating the first round equation which involved the key bits, ***IV*** bits and the first keystream bit with a nonlinear equation. Based on this method, 16 linear approximations with the same bias $2^{-23.42}$ were found using proposed algorithm. Furthermore, multiple linear cryptanalysis was made on 2-round Trivium. The result shows that it can approach the same success rate with $2^{42.84}$ chosen ***IV***s, that is to say, the data complexity is $\frac{1}{2^{19.16}}$ of that in Turan's scheme.

Key words: stream ciphers, Trivium, mutiple linear cryptanalysis, linear approximation

1 引言

由 Cannière 和 Preneel[1]于 2005 年设计的 Trivium 算法是 2004 年欧洲序列密码计划最终胜选的标准算法之一，该算法面向硬件实现，具有设计简洁、安全、速度快和灵活性好等特点。因此，Trivium 在发布之初就引起了广泛关注，针对 Trivium 的安全性分析一直是流密码的热点研究方向之一。

目前针对 Trivium 的主要攻击方法有线性攻击、区分攻击、代数攻击、滑动攻击、立方攻击、错误引入攻击等。Maximov 和 Biryukov[2]在 2007 年提出了 Trivium 的代数攻击方法，攻击的计算复杂度为 $2^{83.5}c$，其中，常数 c 是求解线性方程组的复杂度。2009 年，Dinur 和 Shamir[3]针对初始化轮数为 767 个时钟的简化版 Trivium 提出了立方攻击，攻击的计算复杂度为 2^{45}。2010 年，Stankovski[4]对初始化轮数为 1 078 个时钟的简化版 Trivium 进行了区分攻击，攻击的计算复杂度为 2^{54}。2011 年，Hu 等[5]通过重复错误注入的方式，提出了一种校验部分原始密钥流和部分错误密钥流的方法，确定错误注入时间及位置，进而说明 Trivium 在故障攻击下的脆弱性，较 Hojsík 等的故障攻击方法[6]，该模型有更弱的假设条件。2014 年，Avijit 等[7]提出了一种故障攻击方法，在已知 117 bit 原始密钥流和

收稿日期：2015-06-25；**修回日期**：2015-12-16
基金项目：信息保障技术重点实验室开放基金资助项目（No.KJ-13-009）
Foundation Item: The Foundation of Science and Technology on Information Assurance Laboratory (No.KJ-13-009)

236 bit 错误密钥流的条件下，攻击的计算复杂度是 $2^{23.85}$。Prakash 等在文献[8]中给出了一种条件很弱的差分错误攻击模型，具有很好的通用性。

Turan 等[9]将 Trivium 的初始化轮数由 1 152 个时钟数降低为 288 个，得到了一个简化的版本（即 2 轮 Trivium），在某种弱密钥条件下，找到了该简化版本的一个偏差为 2^{-31} 的线性近似式，从而可以对其进行单线性密码分析。在此基础上，贾艳艳等[10]找到另一个偏差为 2^{-31} 的线性逼近，基于多线性密码分析理论，提出了具体的攻击算法，从而可以对其进行多线性密码分析，同时指出，若能找到 n 个相同偏差的线性近似式，多线性密码分析所需的数据量只有单线性密码分析的 $\frac{1}{n}$，且攻击成功的概率保持不变。孙文龙等[11]给出了在弱密钥和选择 ***IV*** 攻击的条件下搜索最佳线性近似式的算法，据此算法找到了 3 个偏差为 2^{-25} 的线性近似式，但是由于密钥不同，不能进行多线性密码分析。欧智慧等[12]通过改变 2 轮 Trivium 第一轮的时钟个数，给出了 1 个偏差为 2^{-29} 的线性近似式和 8 个偏差为 2^{-30} 的线性近似式，使多线性密码分析所需要的数据量降为文献[9]的 $\frac{1}{16}$，具体的攻击算法可见文献[10]。

本文的目的是寻找偏差更大、个数更多的 2 轮 Trivium 线性近似式，从而降低对 2 轮 Trivium 进行多线性密码分析所需要的数据量，主要工作体现在如下 2 点：1）在某种弱密钥条件下，通过搜索最佳线性近似式算法给出了偏差为 $2^{-23.42}$ 的线性近似式；2）对 2 轮 Trivium 的第 1 轮关于密钥、初始化向量和密钥流比特的表达式做非线性逼近，找到了 2 轮 Trivium 的 16 个偏差为 $2^{-23.42}$ 的线性近似式。

2 基础知识

定义 1[13] 称 $P(\{X:X=0\})-0.5$ 是随机变量 X 的偏差，称 $P(\{x:f(x)=0\})-0.5$ 是函数 $f(x)$ 的偏差。若 $f(x)\oplus g(x)$ 的偏差为 ε，则称函数 $f(x)$ 以偏差 ε 逼近 $g(x)$。

引理 1[14] 堆积引理。设 X_1,X_2,L,X_k 是 F_2 上 k 个独立随机变量，$\varepsilon_i(1\leqslant i\leqslant k)$ 表示 X_k 的偏差。ε 表示随机变量 $X_1\oplus X_2\oplus\mathrm{L}\oplus X_k$ 的偏差，则 $\varepsilon=2^{k-1}\prod_1^k\varepsilon_i$。

引理 2[15] 最佳仿射逼近定理。设 $f(x)$ 是 n 元布尔函数，如果 $|S_{(f)}(\omega^*)|=\mathrm{Max}\left\{|S_{(f)}(\omega)|\,\middle|\,\omega\in F_2^n\right\}$，那么当 $S_{(f)}(\omega^*)>0$ 时，$a\oplus\omega^*X$ 为 $f(x)$ 的最佳仿射逼近，当 $S_{(f)}(\omega^*)<0$ 时，$1\oplus\omega^*X$ 为 $f(x)$ 的最佳仿射逼近。

引理 3[12] 设 $f(x)=x_1\oplus x_2\oplus\mathrm{L}\oplus x_k\oplus x_{k+1}x_{k+2}\oplus x_{k+3}x_{k+4}\oplus\mathrm{L}\oplus x_{k+2t}x_{k+2t-1}$ 为 $k+2t$ 元布尔函数，$\omega\in F_2^{k+2t},\omega x=x_{i_1}\oplus\mathrm{L}\oplus x_{i_m}\ (i_1<i_2<\mathrm{L}<i_m)$。如果 $\omega\neq0$ 或 $\{1,2,\mathrm{L},k\}\not\subset\{i_1,i_2,\mathrm{L},i_m\}$，则 $S_{(f)}(\omega)=0$。如果 $\{1,2,\mathrm{L},k\}\subseteq\{i_1,i_2,\mathrm{L},i_m\}$，则 $S_{(f)}(\omega)=(-1)^s2^{-t}$，其中，$s$ 为 $f(x)\oplus\omega x$ 中具有 $x_{k+2n-1}\oplus x_{k+2n}\oplus x_{k+2n-1}x_{k+2n}(1\leqslant n\leqslant t)$ 形式的个数。

3 Trivium 算法描述

Trivium 算法结构由 3 个移位寄存器构成，级数分别为 93、84 和 111，设 $\boldsymbol{K}=(k_1,\mathrm{L},k_{80})$、$\boldsymbol{IV}=(iv_1,\mathrm{L},iv_{80})$ 分别表示算法的密钥和初始化向量。Trivium 算法初始化阶段首先将 80 bit 的密钥 ***K*** 和 80 bit 的初始化向量 ***IV*** 载入到 3 个移位寄存器内部状态中，93 级布态方式为：$(s_1,\mathrm{L},s_{93})\leftarrow(k_1,\mathrm{L},k_{80},0,\mathrm{L},0)$，84 级布态方式为：$(s_{94},\mathrm{L},s_{177})\leftarrow(iv_1,\mathrm{L},iv_{80},0,\mathrm{L},0)$，111 级布态方式为：$(s_{178},\mathrm{L},s_{288})\leftarrow(0,\mathrm{L},0,1,1,1)$，然后运行密钥流生成算法，空跑 4×288 个时钟，此阶段不输出密钥流比特，只为 3 个寄存器布态，输出为 Trivium 的内部状态 (s_1,L,s_{288})。Trivium 算法密钥流生成阶段的算法和初始化阶段算法相同，只是此时输出密钥流 (z_1,L,z_M)，输入为内部状态 (s_1,L,s_{288}) 和密钥数 M，算法描述如下。

1) for $i=1$ to M do
2) $z_i\leftarrow s_{66}\oplus s_{93}\oplus s_{162}\oplus s_{177}\oplus s_{243}\oplus s_{288}$;
3) $t_1\leftarrow s_{66}\oplus s_{91}s_{92}\oplus s_{93}\oplus s_{171}$;
4) $t_2\leftarrow s_{162}\oplus s_{175}s_{176}\oplus s_{177}\oplus s_{264}$;
5) $t_3\leftarrow s_{243}\oplus s_{286}s_{287}\oplus s_{288}\oplus s_{69}$;
6) $(s_1,\mathrm{L},s_{93})\leftarrow(t_3,s_1,\mathrm{L},s_{92})$;
7) $(s_{94},\mathrm{L},s_{177})\leftarrow(t_1,s_{94},\mathrm{L},s_{176})$;
8) $(s_{178},\mathrm{L},s_{288})\leftarrow(t_2,s_{178},\mathrm{L},s_{287})$;
9) end for

贾艳艳等[10]指出，将 Trivium 算法可以看作是函数 $F_i:F_2^k\times F_2^v\rightarrow F_2\,(i=1,2,\mathrm{L})$ 的集合，其中，F_i 是由 k bit ***K*** 和 v bit ***IV*** 生成第 i 个密钥流比特 z_i 的函数，

如果找到了 F_i 的一个偏差为 $\varepsilon > 2^{-\frac{k}{2}}$ 的线性近似式，将密码再同步 ε^{-2} 次，则一定可以得到密钥 $\boldsymbol{K}$ 的 1 bit 信息。文献[9~12]的思想是：对 Trivium 算法进行多线性密码分析时，可以将其初始化阶段像分组密码一样分成 n 轮，将每一轮的线性近似式有效组合，最终得到 Trivium 算法一个关于($\boldsymbol{K}$, $\boldsymbol{IV}$)和密钥流比特的线性近似式，其偏差可利用堆积引理求得。2 轮 Trivium 的初始化轮数为 288 个时钟，文献[9~11]选择对称分轮，将 144 个时钟作为一轮来分析 2 轮 Trivium；文献[12]选择不对称分轮，分别设定第一轮所占时钟个数为 141、154 来分析 2 轮 Trivium。

4　研究结果

对于给定的密码算法，线性密码分析的关键是找到它的线性近似式。针对 2 轮 Trivium，一般的做法是首先对每轮进行线性逼近，然后将各个线性逼近有效地组合，得到最终的线性近似式。事实上，也可以通过对第 1 轮求具有较大偏差的非线性近似式，对第 2 轮求线性近似式，有效组合得到 2 轮 Trivium 的线性近似式，这也是本文的思想。

设 $(s_i(1), s_i(2), \mathrm{L}\ , s_i(288))$ 表示第 i 个时钟 Trivium 的内部状态，由 Trivium 算法，输出的密钥流比特 $z_1 = s_{288}(66) \oplus s_{288}(93) \oplus s_{288}(162) \oplus\ s_{288}(177) \oplus\ s_{288}(243) \oplus s_{288}(288)$，进一步得到 z_1 关于 $(s_{144}(1), \mathrm{L}\ , s_{144}(288))$ 的函数表达式如下

$$\begin{aligned} z_1 = {} & s_{144}(6) \oplus s_{144}(33) \oplus s_{144}(57) \oplus s_{144}(84) \oplus s_{144}(96) \oplus \\ & s_{144}(99) \oplus s_{144}(111) \oplus s_{144}(129) \oplus s_{144}(144) \oplus \\ & s_{144}(150) \oplus s_{144}(162) \oplus s_{144}(165) \oplus s_{144}(186) \oplus \\ & s_{144}(192) \oplus s_{144}(210) \oplus s_{144}(231) \oplus s_{144}(237) \oplus \\ & s_{144}(252) \oplus s_{144}(16) s_{144}(17) \oplus s_{144}(31) s_{144}(32) \oplus \\ & s_{144}(82) s_{144}(83) \oplus s_{144}(97) s_{144}(98) \oplus \\ & s_{144}(142) s_{144}(143) \oplus s_{144}(163) s_{144}(164) \oplus \\ & s_{144}(208) s_{144}(209) \oplus s_{144}(235) s_{144}(236) \end{aligned} \tag{1}$$

用非线性二次函数

$$\begin{aligned} z_1 = {} & s_{144}(6) \oplus s_{144}(33) \oplus s_{144}(57) \oplus s_{144}(84) \oplus s_{144}(96) \oplus \\ & s_{144}(99) \oplus s_{144}(111) \oplus s_{144}(129) \oplus s_{144}(144) \oplus \\ & s_{144}(150) \oplus s_{144}(162) \oplus s_{144}(165) \oplus s_{144}(186) \oplus \\ & s_{144}(192) \oplus s_{144}(210) \oplus s_{144}(231) \oplus s_{144}(237) \oplus \\ & s_{144}(252) \oplus s_{144}(82) s_{144}(83) \end{aligned} \tag{2}$$

逼近式(1)。设式(2)以偏差 p_1 成立，注意到式(1)的每一项都是相互独立的，由堆积引理，$p_1 = 2^6 \cdot (2^{-2})^7 = 2^{-8}$。显然，类似于式(2)，以偏差 2^{-8} 成立的非线性二次式有 7 个，而选择式(2)作为第 1 轮的逼近式可使最终的线性近似式偏差最大。由式(2)和迭代算法最终可得 z_1 关于 $\left(s_0(1), \mathrm{L}\ , s_0(288)\right)$ 的函数表达式。其中，$s_0(i) = 0 (81 \leqslant i \leqslant 93)$，$s_0(i) = 1 (286 \leqslant i \leqslant 288)$。$z_1$ 是一个关于 $s_0(i) (1 \leqslant i \leqslant 80, 94 \leqslant i \leqslant 173)$ 的一个三次函数，共 22 个三次项，58 个二次项，25 个线性项，由引理 1~引理 3，消去所有非线性项，即为要找的 2 轮 Trivium 最佳线性近似式。

$$\begin{aligned} z_1 = {} & 1 + s_0(3) \oplus s_0(6) \oplus s_0(15) \oplus s_0(21) \oplus s_0(27) \oplus \\ & s_0(30) \oplus s_0(39) \oplus s_0(54) \oplus s_0(57) \oplus s_0(67) \oplus \\ & s_0(68) \oplus s_0(69) \oplus s_0(72) \oplus s_0(96) \oplus s_0(99) \oplus \\ & s_0(114) \oplus s_0(117) \oplus s_0(123) \oplus s_0(126) \oplus s_0(132) \oplus \\ & s_0(138) \oplus s_0(144) \oplus s_0(165) \oplus s_0(171) \end{aligned} \tag{3}$$

由堆积引理，式(3)由式(2)逼近的偏差为 $p_2 = 2^{79} (0.25)^{58} (0.375)^{22} \approx 2^{-68.13}$。

综合式(1)~式(3)，由堆积引理得到式(3)成立的偏差为 $p = 2 p_1 p_2 = 2^{-75.13} < 2^{-\frac{|\boldsymbol{K}|}{2}} = 2^{-40}$，其中，$|\boldsymbol{K}|$ 为密钥 $\boldsymbol{K}$ 的规模。但是这个偏差太小，对于实际的线性分析没有意义。此时，可以通过选择特殊密钥 $\boldsymbol{K}$ 和 $\boldsymbol{IV}$ 增大式(3)成立的偏差，这对于选择 $\boldsymbol{IV}$ 攻击来说并不困难。设 $O_{\boldsymbol{K}} = \{i | k_i = 0, 1 \leqslant i \leqslant 80\}$ 表示为零的密钥比特集合，$O_{\boldsymbol{IV}} = \{i | iv_i = 0, 1 \leqslant i \leqslant 80\}$ 表示为零的 $\boldsymbol{IV}$ 比特集合。$|O_{\boldsymbol{K}}|$、$|O_{\boldsymbol{IV}}|$ 分别表示 $O_{\boldsymbol{K}}$、$O_{\boldsymbol{IV}}$ 的规模。n_1、n_2 分别表示给定 $O_{\boldsymbol{K}}$、$O_{\boldsymbol{IV}}$ 并化简剩余的二次项、三次项的个数，由堆积引理，此时式(3)成立的偏差为

$$\begin{aligned} p &= 2 p_1 p_2 = 2 \cdot 2^{-8} \cdot 2^{(n_1 + n_2) - 1} (0.25)^{n_1} (0.375)^{n_2} \\ &= (0.5)^{n_1 + 8} (0.75)^{n_2} \end{aligned} \tag{4}$$

显然，不同的 $O_{\boldsymbol{K}}$ 和 $O_{\boldsymbol{IV}}$ 会影响偏差 p 的大小。所以，根据一定的准则来选择 $O_{\boldsymbol{K}}$ 和 $O_{\boldsymbol{IV}}$ 得到最佳线性近似式是可行的。下面给出某种弱密钥条件下搜索最佳线性近似式算法。

算法 1　搜索最佳线性近似算法

初始化 $O_{\boldsymbol{K}}$和$O_{\boldsymbol{IV}}$ 为空集。

输入　指定的 $O_{\boldsymbol{K}}$ 规模 $N_{\boldsymbol{K}}$，$O_{\boldsymbol{IV}}$ 规模 $N_{\boldsymbol{IV}}$。

step1　初始化 R 为空集，统计二次项和三次项的频次，分别记为 N_1, N_2。

step2 统计非线性项每个比特 i 涉及到的二次项和三次项的频次，分别记为 n_1^i, n_2^i，计算 $\varepsilon_i = (0.5)^{N_1 - n_1^i} \cdot (0.75)^{N_2 - n_2^i}$，其中，$1 \leqslant i \leqslant 80$ 或 $94 \leqslant i \leqslant 173$，$\varepsilon_i$ 达到最大值时，将比特 i 存入 R：

若 $|R| = 1$,判断 R 中比特类型，密钥存入 O_K，否则存入 O_{IV}；

若 $|R| \geqslant 1$,任选 R 中一个比特，判断类型，密钥存入 O_K，否则存入 O_{IV}。

step3 根据已有的 O_K 和 O_{IV} 重新计算表达式：

如果 $|O_K| < N_K$，$|O_{IV}| < N_{IV}$，则返回 step1；

如果 $|O_K| = N_K$，$|O_{IV}| < N_{IV}$，则停止搜索密钥比特，返回 step1；

如果 $|O_K| < N_K$，$|O_{IV}| = N_{IV}$，则停止搜索 ***IV*** 比特，返回 step1；

如果 $|O_K| = N_K$，$|O_{IV}| = N_{IV}$，输出 (O_K, O_{IV})。

命题 1 指定算法 1 搜索到的比特为 0，式(3)成立的偏差 p 最大。

证明 设 $\{i_1, i_2, \mathrm{L}, i_{20}\} = O_K \cup O_{IV}$，其中，$i_j (1 \leqslant j \leqslant 20)$ 表示第 j 个搜索到的比特。根据算法 1，显然有 $\varepsilon_{i_1} \leqslant \varepsilon_{i_2} \leqslant \mathrm{L} \leqslant \varepsilon_{i_{20}}$。算法 1 基于使偏差达到最大的准则搜索指定为零的比特，所以根据 $\{i_1, i_2, \mathrm{L}, i_{20}\}$ 重新计算表达式后，代入式(4)，式(3)以最大的偏差 $\varepsilon = 2^{-8}\varepsilon_{i_{20}}$ 成立，因此是最佳线性近似式。

输入 $N_K = 10, N_{IV} = 10$，对 z_1 关于 $s_0(i) (1 \leqslant i \leqslant 80, 94 \leqslant i \leqslant 173)$ 的函数表达式执行算法 1,得到

$$O_K = \{4,13,19,38,40,44,46,58,64,65\}$$

$$O_{IV} = \{10,13,25,31,34,37,50,55,56,70\}$$

指定 $k_4 = k_{13} = k_{19} = k_{38} = k_{40} = k_{44} = k_{46} = k_{58} = k_{64} = k_{65} = 0$，$iv_{10} = iv_{13} = iv_{25} = iv_{31} = iv_{34} = iv_{37} = iv_{50} = iv_{55} = iv_{56} = iv_{70} = 0$，从而得到 z_1 新的表达式，具有较少的非线性项如下

$$\begin{aligned} z_1 = 1 + {} & s_0(3) \oplus s_0(6) \oplus s_0(15) \oplus s_0(21) \oplus s_0(27) \oplus \\ & s_0(30) \oplus s_0(39) \oplus s_0(54) \oplus s_0(57) \oplus s_0(67) \oplus \\ & s_0(68) \oplus s_0(69) \oplus s_0(72) \oplus s_0(96) \oplus s_0(99) \oplus \\ & s_0(114) \oplus s_0(117) \oplus s_0(123) \oplus s_0(126) \oplus \\ & s_0(132) \oplus s_0(138) \oplus s_0(144) \oplus s_0(165) \oplus s_0(171) \oplus \\ & s_0(7)s_0(8) \oplus s_0(22)s_0(23) \oplus s_0(34)s_0(35) \oplus \\ & s_0(160)s_0(161) \oplus s_0(52)s_0(53) \oplus s_0(67)s_0(68) \oplus \\ & s_0(16)s_0(17) \oplus s_0(25)s_0(26) \oplus s_0(28)s_0(29) \oplus \\ & s_0(49)s_0(50) \oplus s_0(61)s_0(62) \oplus s_0(151)s_0(152) \oplus \\ & s_0(70)s_0(71) \oplus s_0(154)s_0(155) \oplus \\ & s_0(166)s_0(167) \oplus s_0(134)s_0(146)s_0(147) \end{aligned} \quad (5)$$

式(5)中的各个非线性项数量相互独立，有 24 个线性项，15 个二次项，1 个三次项，代入式(4)，式(3)成立的偏差为：$p \approx 2 \cdot 2^{-8} \cdot 2^{-16.42} = 2^{-23.42}$，由上述分析可知式(6)即为偏差最大的 2 轮 Trivium 线性近似函数。

$$\begin{aligned} z_1 = 1 + {} & k_3 \oplus k_6 \oplus k_{15} \oplus k_{21} \oplus k_{27} \oplus k_{30} \oplus k_{39} \oplus k_{54} \oplus \\ & k_{57} \oplus k_{67} \oplus k_{68} \oplus k_{69} \oplus k_{72} \oplus iv_3 \oplus iv_6 \oplus iv_{21} \oplus \\ & iv_{24} \oplus iv_{30} \oplus iv_{33} \oplus iv_{39} \oplus iv_{45} \oplus iv_{51} \oplus iv_{72} \oplus iv_{78} \end{aligned} \quad (6)$$

孙文龙等[11]在指定 10 bit 密钥和 10 bit ***IV*** 为零的前提下，通过算法得到了偏差为 2^{-25} 的最佳线性近似式。而对于同样的目标函数式，执行本文给出的算法 1，输入 $N_K = 10, N_{IV} = 10$，最佳线性近似式以偏差 $2^{-23.42}$ 成立。用类似于式(2)的多个非线性式去逼近式(1)，可以得到多个不同的 2 轮 Trivium 线性近似式。选择非线性式的原则是：来自式(1)的非线性项，或可以被式(1)的非线性项合并，且使 z_1 关于 $s_0(i) (1 \leqslant i \leqslant 80, 94 \leqslant i \leqslant 173)$ 的函数表达式具有较低次数和较少非线性项。多线性密码分析要求识别的是具有特定比特的密钥，所以指定密钥 ***K*** 为零的比特必须是相同位置上的。另外，选取 ***IV*** 时，特定比特尽量在相同位置上，这样选择 ***IV*** 攻击将会易于操作，便于实施。取不同的非线性式逼近式(1)，执行算法 1，令 $O_K = \{4,13,19,38,40,44,46,58,64,65\}$，输入 $N_K = 10, N_{IV} = 10$，仅搜索 ***IV***，得到了 16 个偏差为 $2^{-23.42}$ 的线性近似式，具体的结果如表 1 所示。基于这 16 个线性近似式可对 2 轮 Trivium 进行多线性密码分析，具体算法可见文献[10]。

为了方便表示，给出如下记号

$$\begin{aligned} R = {} & s_{144}(6) \oplus s_{144}(33) \oplus s_{144}(57) \oplus s_{144}(84) \oplus \\ & s_{144}(96) \oplus s_{144}(99) \oplus s_{144}(111) \oplus s_{144}(129) \oplus \\ & s_{144}(144) \oplus s_{144}(150) \oplus s_{144}(162) \oplus \\ & s_{144}(165) \oplus s_{144}(186) \oplus s_{144}(192) \oplus s_{144} \\ & (210) \oplus s_{144}(231) \oplus s_{144}(237) \oplus s_{144}(252) \\ F = {} & 1 \oplus k_3 \oplus k_6 \oplus k_{15} \oplus k_{21} \oplus k_{27} \oplus k_{30} \oplus k_{39} \oplus \\ & k_{54} \oplus k_{57} \oplus k_{67} \oplus k_{68} \oplus k_{69} \oplus k_{72} \oplus iv_3 \oplus \\ & iv_6 \oplus iv_{21} \oplus iv_{23} \oplus iv_{24} \oplus iv_{30} \oplus iv_{33} \oplus iv_{38} \oplus \\ & iv_{39} \oplus iv_{45} \oplus iv_{51} \oplus iv_{68} \oplus iv_{72} \oplus iv_{78} \end{aligned}$$

$$E_1=\{k_i=0|i=4,13,19,38,40,44,46,58,64,65\}$$
$$E_2=\{iv_i=0|i=10,13,25,31,37,50,55,56,70\}$$
$$E=E_1\cup E_2$$

另外，若选取规模为$(|O_K|,|O_{IV}|)=(10,8)$，在表 1 的基础上，令$E_2=\{iv_i=0|i=25,31,37,50,55,56,70\}$，则表 1 中 16 个线性近似式的偏差为$2^{-25.42}$；若选取规模为$(|O_K|,|O_{IV}|)=(10,13)$，在表 1 的基础上，令$iv_{58}=iv_{61}=iv_{73}=0$，则表 1 中 16 个线性近似式的偏差为$2^{-20.42}$。

不妨设第 1 轮所占时钟数$t_1=141$，得到z_1关于$(s_{141}(1),\cdots,s_{141}(288))$的函数表达式如下

$$\begin{aligned}z_1=&s_{141}(3)\oplus s_{141}(30)\oplus s_{141}(54)\oplus s_{141}(66)\oplus s_{141}(81)\oplus\\&s_{141}(93)\oplus s_{141}(96)\oplus s_{141}(108)\oplus s_{141}(126)\oplus\\&s_{141}(141)\oplus s_{141}(147)\oplus s_{141}(159)\oplus s_{141}(162)\oplus\\&s_{141}(171)\oplus s_{141}(183)\oplus s_{141}(189)\oplus s_{141}(207)\oplus\\&s_{141}(228)\oplus s_{141}(234)\oplus s_{141}(249)\oplus s_{141}(205)\\&s_{141}(206)\oplus s_{141}(94)s_{141}(95)\oplus s_{141}(13)s_{141}(14)\oplus\\&s_{141}(28)s_{141}(29)\oplus s_{141}(232)s_{141}(233)\oplus s_{141}(139)\\&s_{141}(140)\oplus s_{141}(160)s_{141}(161)\oplus s_{141}(79)s_{141}(80)\oplus\\&s_{141}(91)s_{141}(92)\end{aligned}\tag{7}$$

用非线性二次式

$$\begin{aligned}z_1=&s_{141}(3)\oplus s_{141}(30)\oplus s_{141}(54)\oplus s_{141}(66)\oplus\\&s_{141}(81)\oplus s_{141}(93)\oplus s_{141}(96)\oplus s_{141}(108)\oplus\\&s_{141}(126)\oplus s_{141}(141)\oplus s_{141}(147)\oplus\\&s_{141}(159)\oplus s_{141}(162)\oplus s_{141}(171)\oplus\\&s_{141}(183)\oplus s_{141}(189)\oplus s_{141}(207)\oplus\\&s_{141}(228)\oplus s_{141}(234)\oplus s_{141}(249)\oplus\\&s_{141}(91)s_{141}(92)\end{aligned}\tag{8}$$

以偏差$p_1=2^{-9}$逼近式(7)。进一步得到z_1关于$s_0(i)(1\leqslant i\leqslant 80,94\leqslant i\leqslant 173)$的函数表达式，在指定 10 bit 密钥和 10 bit ***IV*** 为零的前提下，对该表达式执行算法 1，指定$k_{13}=k_{19}=k_{38}=k_{40}=k_{44}=k_{46}=k_{58}=k_{64}=k_{65}=k_{67}=0$，$iv_{10}=iv_{13}=iv_{25}=iv_{31}=iv_{34}=iv_{37}=iv_{50}=iv_{55}=iv_{56}=iv_{70}=0$，重新计算表达式，有 23 个线性项，14 个二次项，1 个三次项。

$$\begin{aligned}z_1=&1+s_0(3)\oplus s_0(6)\oplus s_0(15)\oplus s_0(21)\oplus s_0(27)\oplus\\&s_0(30)\oplus s_0(39)\oplus s_0(54)\oplus s_0(57)\oplus s_0(68)\oplus\\&s_0(69)\oplus s_0(72)\oplus s_0(96)\oplus s_0(99)\oplus s_0(114)\oplus\\&s_0(117)\oplus s_0(123)\oplus s_0(126)\oplus s_0(132)\oplus\\&s_0(138)\oplus s_0(144)\oplus s_0(165)\oplus s_0(171)\oplus\end{aligned}$$

表 1　　**t_1=144 时 z_1 的 16 个偏差为 $2^{-23.42}$ 线性近似**

第 1 轮非线性逼近式	指定为零的比特	z_1 的线性近似
$z_1=R\oplus s_{144}(82)s_{144}(83)$	$E,iv_{34}=0$	$z_1=F\oplus iv_{23}\oplus iv_{38}\oplus iv_{68}$
$z_1=R\oplus s_{144}(82)s_{144}(83)\oplus s_{144}(163)$	$E,iv_{34}=0$	$z_1=F\oplus k_{61}\oplus iv_4\oplus iv_{23}\oplus iv_{38}\oplus iv_{68}$
$z_1=R\oplus s_{144}(82)s_{144}(83)\oplus s_{144}(164)$	$E,iv_{34}=0$	$z_1=F\oplus k_{20}\oplus k_{62}\oplus iv_5\oplus iv_{23}\oplus iv_{38}\oplus iv_{68}$
$z_1=R\oplus s_{144}(82)s_{144}(83)\oplus s_{144}(32)$	$E,iv_{67}=0$	$z_1=F\oplus k_{26}$
$z_1=R\oplus s_{144}(82)s_{144}(83)\oplus s_{144}(163)\oplus s_{144}(32)$	$E,iv_{67}=0$	$z_1=F\oplus k_{26}\oplus k_{61}\oplus iv_4$
$z_1=R\oplus s_{144}(82)s_{144}(83)\oplus s_{144}(164)\oplus s_{144}(32)$	$E,iv_{67}=0$	$z_1=F\oplus k_{20}\oplus k_{26}\oplus k_{62}\oplus iv_5$
$z_1=R\oplus s_{144}(163)s_{144}(88)\oplus s_{144}(82)$	$E,iv_{34}=0$	$z_1=F\oplus k_7\oplus iv_{23}\oplus iv_{38}\oplus iv_{68}$
$z_1=R\oplus s_{144}(163)s_{144}(88)\oplus s_{144}(83)$	$E,iv_{34}=0$	$z_1=F\oplus k_8\oplus iv_{23}\oplus iv_{38}\oplus iv_{68}$
$z_1=R\oplus s_{144}(163)s_{144}(88)\oplus s_{144}(163)\oplus s_{144}(83)$	$E,iv_{34}=0$	$z_1=F\oplus k_8\oplus k_{61}\oplus iv_4\oplus iv_{23}\oplus iv_{38}\oplus iv_{68}$
$z_1=R\oplus s_{144}(163)s_{144}(88)\oplus s_{144}(163)\oplus s_{144}(82)$	$E,iv_{34}=0$	$z_1=F\oplus k_7\oplus k_{61}\oplus iv_4\oplus iv_{23}\oplus iv_{38}\oplus iv_{68}$
$z_1=R\oplus s_{144}(163)s_{144}(88)\oplus s_{144}(83)\oplus s_{144}(32)$	$E,iv_{67}=0$	$z_1=F\oplus k_8\oplus k_{26}$
$z_1=R\oplus s_{144}(163)s_{144}(88)\oplus s_{144}(82)\oplus s_{144}(32)$	$E,iv_{67}=0$	$z_1=F\oplus k_7\oplus k_{26}$
$z_1=R\oplus s_{144}(163)s_{144}(88)\oplus s_{144}(163)\oplus s_{144}(82)\oplus s_{144}(32)$	$E,iv_{67}=0$	$z_1=F\oplus k_7\oplus k_{26}\oplus k_{61}\oplus iv_4$
$z_1=R\oplus s_{144}(163)s_{144}(88)\oplus s_{144}(163)\oplus s_{144}(83)\oplus s_{144}(32)$	$E,iv_{67}=0$	$z_1=F\oplus k_8\oplus k_{26}\oplus k_{61}\oplus iv_4$
$z_1=R\oplus s_{144}(164)s_{144}(88)\oplus s_{144}(164)\oplus s_{144}(83)\oplus s_{144}(32)$	$E,iv_{67}=0$	$z_1=F\oplus k_8\oplus k_{20}\oplus k_{26}\oplus k_{62}\oplus iv_5$
$z_1=R\oplus s_{144}(164)s_{144}(88)\oplus s_{144}(164)\oplus s_{144}(82)\oplus s_{144}(32)$	$E,iv_{67}=0$	$z_1=F\oplus k_7\oplus k_{20}\oplus k_{26}\oplus k_{62}\oplus iv_5$

$$s_0(4)s_0(5)\oplus s_0(22)s_0(23)\oplus s_0(34)s_0(35)\oplus s_0(160)s_0(161)\oplus s_0(52)s_0(53)\oplus s_0(16)s_0(17)\oplus s_0(25)s_0(26)\oplus s_0(28)s_0(29)\oplus s_0(49)s_0(50)\oplus s_0(61)s_0(62)\oplus s_0(151)s_0(152)\oplus s_0(70)s_0(71)\oplus s_0(154)s_0(155)\oplus s_0(166)s_0(167)\oplus s_0(134)s_0(146)s_0(147) \quad (9)$$

式(9)以偏差 $p_2=2^{-15.42}$ 逼近式(8)，式(8)以偏差 $p_1=2^{-9}$ 逼近式(7)。由堆积引理可得式(10)以偏差 $p=2p_1p_2=2^{-23.42}$ 成立。

$$z_1=1+k_3\oplus k_6\oplus k_{15}\oplus k_{21}\oplus k_{27}\oplus k_{30}\oplus k_{39}\oplus k_{54}\oplus k_{57}\oplus k_{67}\oplus k_{68}\oplus k_{69}\oplus k_{72}\oplus iv_3\oplus iv_6\oplus iv_{21}\oplus iv_{24}\oplus iv_{30}\oplus iv_{33}\oplus iv_{39}\oplus iv_{45}\oplus iv_{51}\oplus iv_{72}\oplus iv_{78} \quad (10)$$

式(10)即为最终的 2 轮 Trivium 线性近似式。

类似于 $t_1=144$ 时，利用多个不同的非线性二次式去逼近式(7)，找到了 6 个偏差为 $2^{-23.42}$ 的线性近似式，如表 2 所示。给出如下记号。

$$R=s_{141}(3)\oplus s_{141}(30)\oplus s_{141}(54)\oplus s_{141}(66)\oplus s_{141}(81)\oplus s_{141}(93)\oplus s_{141}(96)\oplus s_{141}(108)\oplus s_{141}(126)\oplus s_{141}(141)\oplus s_{141}(147)\oplus s_{141}(159)\oplus s_{141}(162)\oplus s_{141}(171)\oplus s_{141}(183)\oplus s_{141}(189)\oplus s_{141}(207)\oplus s_{141}(228)\oplus s_{141}(234)\oplus s_{141}(249)$$

$$F=1\oplus k_3\oplus k_6\oplus k_{15}\oplus k_{21}\oplus k_{27}\oplus k_{30}\oplus k_{39}\oplus k_{54}\oplus k_{57}\oplus k_{67}\oplus k_{68}\oplus k_{69}\oplus k_{72}\oplus iv_3\oplus iv_6\oplus iv_{21}\oplus iv_{23}\oplus iv_{24}\oplus iv_{30}\oplus iv_{33}\oplus iv_{38}\oplus iv_{39}\oplus iv_{45}\oplus iv_{51}\oplus iv_{68}\oplus iv_{72}\oplus iv_{78}$$

$$E_1=\{k_i=0 \mid i=4,13,19,38,40,44,46,58,64,65\}$$

$$E_2=\{iv_i=0 \mid i=10,13,25,31,37,50,55,56,70\}$$

$$E=E_1\cup E_2$$

对比表 1 和表 2 发现，当 $t_1=141$ 时得到的 6 个偏差为 $2^{-23.42}$ 的线性近似式在取 $t_1=144$ 时出现过，这表明取 $t_1=144$ 时，用上述方法对 2 轮 Trivium 的线性逼近效果不会比 $t_1=141$ 差。本文算法与已有算法的结果对比如表 3 所示。

表 2　　t_1=141 时 z_1 的 6 个线性近似

第 1 轮非线性逼近式 G_1	指定为零的比特	z_1 的线性近似
$z_1=R\oplus s_{141}(91)s_{141}(92)$	$E, iv_{34}=0$	$z_1=F\oplus iv_{23}\oplus iv_{38}\oplus iv_{68}$
$z_1=R\oplus s_{141}(91)s_{141}(92)\oplus s_{141}(80)$	$E, iv_{34}=0$	$z_1=F\oplus k_8\oplus iv_{23}\oplus iv_{38}\oplus iv_{68}$
$z_1=R\oplus s_{141}(91)s_{141}(92)\oplus s_{141}(79)$	$E, iv_{34}=0$	$z_1=F\oplus k_7\oplus iv_{23}\oplus iv_{38}\oplus iv_{68}$
$z_1=R\oplus s_{141}(91)s_{141}(92)\oplus s_{141}(29)$	$E, iv_{67}=0$	$z_1=F\oplus k_{26}$
$z_1=R\oplus s_{141}(91)s_{141}(92)\oplus s_{141}(80)\oplus s_{141}(29)$	$E, iv_{67}=0$	$z_1=F\oplus k_8\oplus k_{26}$
$z_1=R\oplus s_{141}(91)s_{141}(92)\oplus s_{141}(79)\oplus s_{141}(29)$	$E, iv_{67}=0$	$z_1=F\oplus k_{20}\oplus k_{26}\oplus k_{62}\oplus iv_5$

表 3　　结果对比

$(\lvert O_K\rvert,\lvert O_{IV}\rvert)$	方法	偏差	线性近似个数	数据量	线性攻击成功率
(10,8)	文献[12]	2^{-29}	1	2^{58}	97.77%
	本文	$2^{-25.42}$	16	$2^{46.84}$	97.77%
(10,10)	文献[9]	2^{-31}	1	2^{62}	97.77%
	文献[10]	2^{-31}	2	2^{61}	97.77%
	文献[11]	2^{-25}	1	2^{50}	97.77%
	文献[12]	2^{-30}	8	2^{57}	97.77%
	本文	$2^{-23.42}$	16	$2^{42.84}$	97.77%
(10,13)	文献[10]	2^{-31}	2	2^{61}	97.77%
	文献[11]	2^{-22}	1	2^{44}	97.77%
	本文	$2^{-20.42}$	16	$2^{36.84}$	97.77%

5　结束语

本文对 2 轮 Trivium 算法进行了线性逼近研究，用非线性式逼近第 1 轮关于密钥、***IV*** 和密钥流比特的表达式，给出了在某种弱密钥条件下搜索最佳线性近似式的算法，据此算法找到了 16 个偏差为 $2^{-23.42}$ 的线性近似式，使对 2 轮 Trivium 算法进行多线性攻击的数据量降低为文献[10]所需数据量的 $\frac{1}{2^{19.16}}$。需要进一步研究的问题：一是算法 1 的时间复杂度和存储复杂度有待降低；二是能否找到 2 轮 Trivium 个数更多、偏差更大的线性近似式；三是如何利用上述思想方法对更多轮数的 Trivium 算法进行线性分析。

参考文献：

[1] CANNIÀRE C, PRENEEL B. Trivium specifications [EB/OL]. http://www.ecrypt.eu.org/stream/p3ciphers/trivium/trivium_p3.pdf, 2007.

[2] MAXIMOV A, BIRYUKOV A. Two trivial attacks on Trivium[C]// c2007:36-55.

[3] DINUR, SHAMIR A. Cubic attacks on tweakable black box polynomials[C]//Advances in Cryptology-EUROCRYPT 2009. c2009: 278-299.

[4] STANKOVSKI P. Greedy distinguishers and nonrandomness detectors[C]//INDOCRYPT 2010. c2010: 210-226.

[5] HOJSÍK M, RUDOLF B. Differential fault analysis of Trivium[C]// FSE 2008. c2008: 158-172.

[6] HU Y P, GAO J T, LIU Q, et al. Fault analysis of Trivium[C]//Designs, Codes and Cryptography. c2011: 289-311.

[7] AVIJIT D, GOUTAM P. Deterministic hard fault attack on trivium[C]//Advances in Information and Computer Security. c2014: 134-145.

[8] PRAKASH D, AVISHEK A. Improved multi-bit differential fault analysis of trivium[C]//Progress in Cryptology. c2014: 37-52.

[9] TURAN M S, KARA O. Linear approximations for 2-round Trivium[C]//Workshop on the State of the Art of Stream Cipher (SASC2007). Bochum, c2007: 22-31.

[10] 贾艳艳, 胡予濮, 杨文峰, 等. 2 轮 Trivium 的多线性密码分析[J]. 电子与信息学报, 2011, 33(1): 223-227.
JIA Y Y, HU Y P, YANG W F, et al. Linear cryptanalysis of 2-round Trivium with multiple approximations[J]. Journal of Electronics & Information Technology, 2011, 33(1): 223-227.

[11] 孙文龙, 关杰, 刘建东. 针对简化版 Trivium 算法的线性分析[J].计算机学报,2012,35(9): 1890-1896.
SUN W L, GUAN J, LIU J D. Linear cryptanalysis of simplified Trivium[J]. Chinese Journal of Computers, 2012,35(9): 1890-1896.

[12] 欧智慧, 赵亚群. 2 轮 Trivium 的线性逼近研究[J].计算机工程,2013,39(11): 31-34.
OU Z H, ZHAO Y Q. Study on linear approximation of 2-round Trivium[J]. Computer Engineering, 2013,39(11): 31-34.

[13] 李世取, 曾本胜, 廉玉忠, 等.密码学中的逻辑函数[M].北京:中软电子出版社,2003: 254-255.
LI S Q, ZENG B S, LIAN Y Z, et al. Logical functions in cryptography[M]. Beijing: Software and Electronic Press, 2003: 254-255.

[14] DOUGLAS R, STINSON. 密码学原理与实践[M]. 北京:电子工业出版社,2010: 123-124.
DOUGLAS R. STINSON. Cryptography theory and practice[M]. Beijing: Publishing House of Electronics Industry, 2010: 123-124.

[15] 丁存生,肖国镇. 流密码及其应用[M].北京:国防科学出版社,1994: 28-29.
DING C S, XIAO G Z. Stream cipher and its applications[M]. Beijing: National Defense Industry Press, 1994: 28-29.

作者简介：

马猛（1986-），男，河南南阳人，信息工程大学硕士生，主要研究方向为流密码分析、概率统计在密码学中的应用。

赵亚群（1961-），女，江苏淮安人，信息工程大学教授、硕士生导师，主要研究方向为密码基础理论及概率统计应用。