Improved Differential Fault Analysis of Trivium

Mohamed Saied Emam Mohamed\textsuperscript{1}, Stanislav Bulygin\textsuperscript{2}, and Johannes Buchmann\textsuperscript{1}

\textsuperscript{1} TU Darmstadt, FB Informatik
Hochschulstrasse 10, 64289 Darmstadt, Germany
\{mohamed,buchmann\}@cdc.informatik.tu-darmstadt.de
\textsuperscript{2} Center for Advanced Security Research Darmstadt (CASED)
Stanislav.Bulygin@cased.de

Abstract. Combining different cryptanalytic methods to attack a cryptosystem became one of the hot topics in cryptanalysis. In particular, algebraic methods in side channel and differential fault analysis (DFA) attracted a lot of attention recently. In [13], Hojsík and Rudolf used DFA to recover the inner state of the stream cipher Trivium which leads to recovering the secret key. For this attack, they required 3.2 one-bit fault injections on average and 800 keystream bits. In this paper, we give an example of combining DFA attacks and algebraic attacks. We use algebraic methods to improve the DFA of Trivium [13]. Our improved DFA attack recovers the inner state of Trivium by using only 2 fault injections and only 420 keystream bits.

Keywords: Differential Fault Analysis, algebraic attack, SAT-Solvers, Trivium

1 Introduction

Instead of attacking a cryptosystem with only one method, an attacker can gain more power by combining several different methods. We present as an application of such a combination, an improvement of differential fault analysis using algebraic techniques.

At INDOCRYPT 2008, M. Hojsík and B. Rudolf introduced an efficient differential fault analysis of the stream cipher Trivium [13]. In this attack they improved their first DFA attack on Trivium [12] by using the floating representation of Trivium instead of the classical representation of the cipher. The basic idea of this attack is to inject a one-bit fault into the inner state of Trivium. In this case an attacker can generate, in addition to the equations system that represents a set of keystream bits, some lower degree polynomial equations that relate a set of keystream bits generated after the injection performed. Using this method one needs 3.2 fault injections on average and 800 keystream bits to recover the inner state of Trivium at certain time $t = t_0$, which leads to recovering the secret key of the cipher.
In this paper, we improve this attack by using a SAT solver to speed up
the solving part of the attack, as well as we improve an equation preprocessing
phase.
In this case, attacker needs exactly 2 one-bit fault injections and 420 keystream
bits to recover the inner state of Trivium.

This paper is organized as follows. In Section 2 we describe the floating
representation of Trivium. In Section 3 we briefly explain the differential fault
analysis of Trivium and in Section 4 we explain the generation of the polynomial
equation system that represent the inner state of Trivium and the DFA of it. Our
attack description and the results are presented in Section 5 and Section 6
respectively. Finally, we conclude the paper in Section 7.

2 Algebraic Description of Trivium

Trivium is a hardware-oriented stream cipher [5]. It was selected in phase three
of profile two of the eSTREAM project [16]. Trivium generates a sequence of
keystream bits from an 80-bit secret key and an 80-bit initial vector (IV). The
inner state of Trivium consists of 288 bits which are stored in three shift registers
respectively as explained in Figure 1. The so called floating representation [13]
In the initialization phase, Trivium loops 1152 times without producing any keystream outputs. Let at time $t = 0$ (directly after the initialization phase), the inner state of Trivium be

$$(a_1, \ldots, a_{93}, b_1, \ldots, b_{84}, c_1, \ldots, c_{111}).$$

The keystream output bits $z_i$ and the new inner state bits $a_{i+93}, b_{i+84}, c_{i+111}$ of Trivium’s registers A, B, C are generated as follows.

$$z_i = a_i + a_{i+27} + b_i + b_{i+15} + c_i + c_{i+45}, \quad i \geq 1$$

(1)

$$a_{i+93} = a_{i+24} + c_{i+45} + c_i + c_{i+1} \cdot c_{i+2}$$

(2.1)

$$b_{i+84} = b_{i+6} + a_{i+27} + a_i + a_{i+1} \cdot a_{i+2}$$

(2.2)

$$c_{i+111} = c_{i+24} + b_{i+15} + b_i + b_{i+1} \cdot b_{i+2}$$

(2.3)

By using equations (2.1)-(2.3), we can clock Trivium forward to generating new inner states. Proposition 1 explains how Trivium can be also clocked backwards to recover previous inner states.

**Proposition 1.** Let the values of the inner state bits at time $t$

$$(a_{t+1}, \ldots, a_{t+93}, b_{t+1}, \ldots, b_{t+84}, c_{t+1}, \ldots, c_{t+111})$$

be given. Then the previous inner state values at the time $t - 1$

$$(a_t, a_{t+1}, \ldots, a_{t+92}, b_t, b_{t+1}, \ldots, b_{t+83}, c_t, c_{t+1}, \ldots, c_{t+110})$$

can also be found.

**Proof.** When we replace $i$ by $t$, the values of the unknown bits $a_t, b_t, c_t$ can be computed from (2.1), (2.2), and (2.3) respectively as follows:

$$a_t = b_{t+6} + b_{t+84} + a_{t+27} + a_{t+1} \cdot a_{t+2}$$

$$b_t = c_{t+24} + c_{t+111} + b_{t+15} + b_{t+1} \cdot b_{t+2}$$

$$c_t = a_{t+24} + a_{t+93} + c_{t+45} + c_{t+1} \cdot c_{t+2}$$

Since all variables on the right-hand sides are bits in the given inner state at time $t$, the inner state at time $t - 1$ can be recovered as well.

3 Preliminaries on the DFA of Trivium

We use the same assumptions as in [13]. Namely, the attacker is able to inject a one-bit fault at a random position within the inner state at $t = 0$. Also, he can obtain the first $n$ keystream bits $z_i$, $1 \leq i \leq n$, before any fault injections and after a fault injection which we call $z'_i$, $1 \leq i \leq n$. The attacker can do the previous steps several times with the same secret key and IV.
Each one-bit fault injection leads to additional equations. We use the difference of keystream outputs ($\triangle z_i = z_i + z'_i$) and shift registers inputs ($\triangle a_i$, $\triangle b_i$, $\triangle c_i$) before and after performing an injection to generate additional polynomial equations as in (3) and (4.1)-(4.3).

$$\triangle z_i = \triangle a_i + \triangle a_{i+27} + \triangle b_i + \triangle b_{i+15} + \triangle c_i + \triangle c_{i+45}, \quad i \geq 0$$ (3)

$$\triangle a_{i+93} = \triangle a_{i+24} + \triangle c_{i+45} + \triangle c_i + \triangle (c_{i+1} \cdot c_{i+2})$$ (4.1)

$$\triangle b_{i+84} = \triangle b_{i+6} + \triangle a_{i+27} + \triangle a_i + \triangle (a_{i+1} \cdot a_{i+2})$$ (4.2)

$$\triangle c_{i+111} = \triangle c_{i+24} + \triangle b_{i+15} + \triangle b_i + \triangle (b_{i+1} \cdot b_{i+2})$$ (4.3)

In this case the difference values of the inner state at $t = 0$, $(\triangle a_1, \ldots, \triangle a_{93}, \triangle b_1, \ldots, \triangle b_{84}, \triangle c_1, \ldots, \triangle c_{111})$, are zeros everywhere except at the fault injection position. For example, suppose that the injected bit is $a_{35}$ then the difference vector will be

$$(0, \ldots, 0, \triangle a_{35} = 1, 0, \ldots, 0)$$

The fault position is not known a priori, but it can be determined by observing the faulty keystream as in [13]. So we assume that we know it. In the next section, we explain how we use this differential model to generate the additional polynomial equations produced after inserting several one-bit fault injections.

## 4 Generating Low Degree Polynomial Equations

We explain the generation of the polynomial equation system that we use in our attack. As defined in Section 2, we let the inner state at time $t = 0$ of Trivium be

$$(a_1, \ldots, a_{93}, b_1, \ldots, b_{84}, c_1, \ldots, c_{111}).$$

Also, we suppose that we have an $n$-bit keystream output vector $Z = (z_1, \ldots, z_n)$.

The TRIV procedure (Alg. 1) constructs the polynomial equation system that describes the generation of $n$ keystream bits using the 288 inner state bits at $t = 0$ as in Algorithm 1. It uses the strategy of generating low degree polynomials. In this case, we represent each new inner state bit generated by one of the shift-registers $A, B, C$ ($a_{i+93}, b_{i+84}, c_{i+111}, \ 1 \leq i \leq n$, respectively) as a new internal variable. By adding them to the 288 initial inner state bit variables, we have totally $3n + 288$ variables ($a_1, \ldots, a_{n+93}, b_1, \ldots, b_{n+84}, c_1, \ldots, c_{n+111}$). We call them the inner state variables.

TRIV takes the keystream bit vector $Z$ as an input and uses equations (1) and (2.1)-(2.3) to generate the polynomial equations. It returns the set of polynomials $P$ that describes $Z$ using the inner state variables. The system of equations $\{p = 0, p \in P\}$ contains $4n$ polynomial equations ($n$ linear and $3n$ quadratic produced by (1) and (2.1)-(2.3) respectively) in the $3n + 288$ inner state variables. We call it the pure system of Trivium without using DFA.
Algorithm 1 TRIV(Z = (z₁, ..., zₙ))

1: P = ∅
2: for i = 1 to n do
3:   P ← P ∪ {aᵢ + aᵢ+27 + bᵢ + bᵢ+15 + cᵢ + cᵢ+45 + zᵢ}  // Eq. (1)
4:   P ← P ∪ {aᵢ+93 + aᵢ+24 + cᵢ + cᵢ+1 : cᵢ+2}  // Eq. (2.1)
5:   P ← P ∪ {bᵢ+84 + bᵢ+6 + aᵢ+27 + aᵢ + aᵢ+1 : aᵢ+2}  // Eq. (2.2)
6:   P ← P ∪ {cᵢ+111 + cᵢ+24 + bᵢ+15 + bᵢ + bᵢ+1 : bᵢ+2}  // Eq. (2.3)
7: end for
8: return P

Now we are going to explain how we generate the additional low degree polynomial equations that are obtained from the faulty key stream. For this we use the differential model that was explained in the previous section. The EQgenerator procedure constructs such equations as described in Alg. 2. It takes as inputs m fault injection positions (l₁, ..., lₘ), the keystream vector Z before any fault injections and m keystream vectors Z⁽¹⁾, ..., Z⁽ᵐ⁾ obtained after each one of the m fault injections, where each keystream vector Z⁽𝑗⁾ = (z⁽𝑗⁾₁, ..., z⁽𝑗⁾ₙ), 1 ≤ j ≤ m.

EQgenerator (Alg. 2) initializes the set of polynomial equations (P) with the set of the pure polynomials returned by the TRIV procedure (line 1). It initializes the arrays a, b, c with the inner state variables (lines 2..4).

For each fault injection j, 1 ≤ j ≤ m, EQgenerator sets the arrays da, db, dc that store the polynomials representing the inner state difference variables

\[(\triangle a_i)_{i=1}^{n+93}, (\triangle b_i)_{i=1}^{n+84}, (\triangle c_i)_{i=1}^{n+111}\]

to zeros (lines 7..9). In line 10 (Inject(da, db, dc, lⱼ)), EQgenerator inserts a fault to the inner state (a₁, ..., a₉₃, b₁, ..., b₈₄, c₁, ..., c₁₁₁) as follows. Let the fault injection position be lⱼ ≤ n. EQgenerator sets da[lⱼ] to 1 when lⱼ ≤ 93. In case of 93 < lⱼ ≤ 177, it sets db[lⱼ] to 1 when lⱼ ≤ 93. Otherwise, it sets dc[lⱼ] to 1.

For each keystream bit zᵢ and the corresponding faulty keystream bit z⁽ᵢ⁾ⱼ, 1 ≤ i ≤ n, EQgenerator evaluates the key stream output difference dz. Then, it uses (3) to generate an additional polynomial and includes this polynomial to P (lines 13,14). Also, by using the fact that

\[\triangle(x \cdot y) = \triangle x \cdot y + x \cdot \triangle y + \triangledown x \cdot \triangledown y,\]

we can reconstruct equations (4.1), (4.2), and (4.3) as follows.

\[
\triangle aᵢ+93 = \triangle aᵢ+24 + \triangle cᵢ+45 + \triangle cᵢ + \triangle cᵢ+1 : cᵢ+2 + cᵢ+1 : \triangle cᵢ+2 + \triangle cᵢ+1 : \triangle cᵢ+2 \tag{5.1}
\]

\[
\triangle bᵢ+84 = \triangle bᵢ+6 + \triangle aᵢ+27 + \triangle aᵢ + \triangle aᵢ+1 : aᵢ+2 + aᵢ+1 : \triangle aᵢ+2 + \triangle aᵢ+1 : \triangle aᵢ+2 \tag{5.2}
\]

\[
\triangle cᵢ+111 = \triangle cᵢ+24 + \triangle bᵢ+15 + \triangle bᵢ + \triangle bᵢ+1 : bᵢ+2 + bᵢ+1 : \triangle bᵢ+2 + \triangle bᵢ+1 : \triangle bᵢ+2 \tag{5.3}
\]
The EQgenerator procedure uses these equations to construct the polynomial entries of the difference polynomial arrays \((da, db, dc)\).

EQgenerator extracts from \(P\) all possible univariate polynomials of the form 
\[x + v,\]
where \(x\) is an inner state variable and \(v\) is 0 or 1 (line 20). If there are such univariates, it simplifies \(P\) by substituting the solved variables in each polynomial \(p \in P\) (line 21). It repeats these two steps as long as it generates more univariates (lines 18...23). After that, EQgenerator uses all univariates produced from the previous loop to simplify the elements of the arrays \(da, db, dc, a, b, c\) (line 24).

Finally, the EQgenerator procedure returns the set of generated polynomials \(P\) together with the set of generated univariates \(S\).

\begin{algorithm}
\textbf{Algorithm 2} EQgenerator\((l_1, \ldots, l_m, Z, Z^{(1)}, \ldots, Z^{(m)})\)
\begin{algorithmic}[1]
\State \(P \leftarrow \text{TRIV}(\mathcal{Z})\)
\State \(a \leftarrow [a_1, \ldots, a_{n+93}]\)
\State \(b \leftarrow [b_1, \ldots, b_{n+84}]\)
\State \(c \leftarrow [c_1, \ldots, c_{n+111}]\)
\State \(S \leftarrow \emptyset\)
\For {\(j = 1\) to \(m\)}
\State \(da \leftarrow [0, \ldots, 0]\) \quad \text{// length}(da) = n + 93
\State \(db \leftarrow [0, \ldots, 0]\) \quad \text{// length}(db) = n + 84
\State \(dc \leftarrow [0, \ldots, 0]\) \quad \text{// length}(dc) = n + 111
\State \(\text{InjectFault}(da, db, dc, l_j)\) \quad \text{// Insert a one-bit fault to one of } da, db, dc \text{ based on the value of } l_j\)
\For {\(i = 1\) to \(n\)}
\State \(S_1 \leftarrow \emptyset\)
\State \(dz \leftarrow z_i + z_i^{(j)}\)
\State \(P \leftarrow da[i] + da[i + 27] + db[i] + db[i + 15] + dc[i] + dc[i + 45] + dz \quad \text{// (3)}\)
\State \(da[i + 93] \leftarrow \text{right hand side of (5.1)}\) \quad \text{// replace } \triangle a_{i+24} \text{ by } da[i + 24], \ldots\)
\State \(db[i + 84] \leftarrow \text{right hand side of (5.2)}\) \quad \text{// replace } \triangle b_{i+6} \text{ by } db[i + 6], \ldots\)
\State \(dc[i + 111] \leftarrow \text{right hand side of (5.3)}\) \quad \text{// replace } \triangle c_{i+24} \text{ by } dc[i + 24], \ldots\)
\Repeat
\State \(S_2 \leftarrow \emptyset\)
\State \(S_2 \leftarrow \text{ExtractUnivariate}(P)\)
\State \(P \leftarrow \text{Substitute}(P, S_2)\)
\State \(S_1 \leftarrow S_1 \cup S_2\)
\Until {\(S_2 = \emptyset\)}
\State \(da, db, dc, a, b, c \leftarrow \text{Substitute}(da, db, dc, a, b, c, S_1)\)
\State \(S \leftarrow S \cup S_1\)
\EndFor
\State \text{return } \(P \cup S\)
\EndFor
\end{algorithmic}
\end{algorithm}

In the EQgenerator procedure, we use the same way of generating polynomial equations as in [13]. However, the authors of [13] substituted by the solved variables only in the higher degree generated polynomials \((P)\), whereas we sub-
stitute solved variables in all the generated polynomials (linear and non-linear) and all constructed \( \Delta a_i, \Delta b_i, \Delta c_i \) polynomials. Moreover, we replace the solved variables by their values in the set of variables to prevent their occurrence in the remaining computation. Our generator does not use any elimination process. In Table 1 we compare the Hojsík-Rudolf (H-R) polynomial system generator [13] with our generator. For this comparison, we have \( n = 800 \) and denote the number of fault injections by \( m \). We report the number of the produced polynomial equations of degree \( \leq 4 \). Clearly, our generator creates systems that contain more linear equations than those created by H-R. This leads to an easier system to solve. We evaluate the average over 1000 experiments.

<table>
<thead>
<tr>
<th>Generator</th>
<th>( m )</th>
<th>degree 1</th>
<th>degree 2</th>
<th>degree 3</th>
<th>degree 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>H-R</td>
<td>0</td>
<td>800</td>
<td>2400</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>H-R</td>
<td>1</td>
<td>825</td>
<td>2466</td>
<td>35</td>
<td>57</td>
</tr>
<tr>
<td>H-R</td>
<td>2</td>
<td>1017</td>
<td>2419</td>
<td>36</td>
<td>57</td>
</tr>
<tr>
<td>H-R</td>
<td>3</td>
<td>1258</td>
<td>2298</td>
<td>37</td>
<td>56</td>
</tr>
<tr>
<td>H-R</td>
<td>4</td>
<td>2402</td>
<td>498</td>
<td>8</td>
<td>12</td>
</tr>
<tr>
<td>our</td>
<td>0</td>
<td>800</td>
<td>2400</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>our</td>
<td>1</td>
<td>994</td>
<td>2394</td>
<td>28</td>
<td>27</td>
</tr>
<tr>
<td>our</td>
<td>2</td>
<td>1212</td>
<td>2362</td>
<td>64</td>
<td>76</td>
</tr>
<tr>
<td>our</td>
<td>3</td>
<td>1619</td>
<td>1990</td>
<td>82</td>
<td>66</td>
</tr>
<tr>
<td>our</td>
<td>4</td>
<td>2688</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 1. Comparison between our generator and Hojsík-Rudolf generator for \( n = 800 \) and \( m \) fault injections.

Table 2 shows the number of generated polynomial equations after performing several fault injections when we have only \( n = 430 \) keystream outputs.

<table>
<thead>
<tr>
<th>( m )</th>
<th>degree 1</th>
<th>degree 2</th>
<th>degree 3</th>
<th>degree 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>430</td>
<td>1290</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>623</td>
<td>1398</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>830</td>
<td>1237</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>1127</td>
<td>988</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>1578</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 2. Number of the produced polynomials when \( n = 430 \) and \( m \) fault injections.

During the process of Algorithm 2, the most important factor that affect the number of solved variables and significantly influence the shape of the generated system is the fault injection position. Table 3 shows a comparison of the generated polynomial equations after inserting two faults to a random positions in
any of \((A,B,C)\) and to a single shift register \((A\) or \(B\) or \(C)\). We fix the inner state at \(t = 0\). Experiments show that when we insert two fault injections to the middle shift register \((B)\), we have the best results on average over 1000 runs.

<table>
<thead>
<tr>
<th>(n)</th>
<th>Register</th>
<th>univariate degree 1</th>
<th>degree 2</th>
<th>degree 3</th>
<th>degree 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>800</td>
<td>(A,B,C)</td>
<td>106</td>
<td>1225</td>
<td>2360</td>
<td>61</td>
</tr>
<tr>
<td>800</td>
<td>(A)</td>
<td>92</td>
<td>1189</td>
<td>2379</td>
<td>63</td>
</tr>
<tr>
<td>800</td>
<td>(B)</td>
<td>120</td>
<td>1261</td>
<td>2348</td>
<td>58</td>
</tr>
<tr>
<td>800</td>
<td>(C)</td>
<td>105</td>
<td>1226</td>
<td>2353</td>
<td>62</td>
</tr>
<tr>
<td>430</td>
<td>(A,B,C)</td>
<td>96</td>
<td>793</td>
<td>1152</td>
<td>4</td>
</tr>
<tr>
<td>430</td>
<td>(A)</td>
<td>84</td>
<td>769</td>
<td>1168</td>
<td>5</td>
</tr>
<tr>
<td>430</td>
<td>(B)</td>
<td>107</td>
<td>824</td>
<td>1137</td>
<td>3</td>
</tr>
<tr>
<td>430</td>
<td>(C)</td>
<td>97</td>
<td>788</td>
<td>1148</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 3. Produced polynomials after inserting 2 fault injections to one of the shift registers.

5 Attack Description

Algebraic cryptanalysis is based on solving a multivariate polynomial system that describes a cryptosystem. There are several methods for solving such systems. Computing Gröbner basis is one of the standard techniques for solving multivariate polynomial systems including Buchberger’s algorithm [4], \(F_4\) [10] and \(F_5\) [11] algorithms. Other techniques like XL [6], MutantXL [8], and MXL\(_3\) [14] are also used to solve multivariate systems. One of the main problems of all of these techniques is the memory usage when we try to solve large systems even if the systems are sparse. Recently SAT solvers have made a great progress in algebraic cryptanalysis. In [7], Bard and Courtois used a SAT solver combined with the slide attack to break KeeLoq block cipher. Using SAT solvers, Eibach et al. [9] attacked Bivium, a scaled version of Trivium. According to our experiments, SAT solvers yielded superior results to Gröbner basis techniques provided by Magma and PolyBoRi [3]. For our implementation of the MXL\(_3\) algorithm, the best variant of the XL algorithm we have, solving such systems is not feasible since our implementation is based on the dense matrix representation, whereas systems considered in this paper are sparse.

The aim of our attack is to recover the inner state of the stream cipher Trivium at \(t = 0\) that leads to recover the secret key. We assume that the attacker has the following information:

1.) \(m\) fault injection positions \((l_1,\ldots,l_m)\), \(l_i \leq 288\), \(i \in \{1,\ldots,m\}\).
2.) The vector of the output keystream bits before the fault injection, \(Z = (z_1,\ldots,z_n)\).
3.) The vectors of the output keystream bits that are obtained after each fault injection, \((Z^{(1)}, \ldots, Z^{(m)})\).

Algorithm 3 explains the main part of this paper. It describes the steps of our attack. The first step has been explained in the previous section. It is important to note that we used only equations of degree \(\leq 2\) in the attack. In this case, the maximal degree of \(P\) is 2.

SAT solvers can deal with a formula in the conjunctive normal form (CNF), a set of clauses, which is a conjunction (\(\land\)) of disjunctions (\(\lor\)) of some variables or negation of variables. Since we used SAT solving in our attack and the equations from \(P\) are represented using the algebraic normal form (ANF), we need to construct the CNF representation of \(P\). We used the ANF-to-CN converter by Martin Albrecht [1] to convert our generated system to CNF. This converter uses the method of Bard-Courtois-Jefferson [2] for converting The ANF polynomial equations to the SAT problem in CNF.

We briefly explain the ANF-to-CN converter of [2]. In the ANF representation, a Boolean polynomial \(p\) is a sum of terms \((t_1 + t_2 + \ldots + t_m)\), where each term \(t_i\) is a product of variables. For each term \(t_i\) of degree \(\geq 1\), we define a new variable \(b_i\) such that \(b_i = t_i\). Then we generate CNF clauses that equivalent to \(b_i = t_i\). For example, let \(p = (x \cdot y + x \cdot z + y \cdot w + x + z + 1)\). We define three new variables \(b_1, b_2, b_3\), where \(b_1 = (x \cdot y), b_2 = (x \cdot z), b_3 = (y \cdot w)\).

Since the multiplication (\(\cdot\)) of two variables is simply the conjunction (\(\land\)), then \((b_1 = x \cdot y) \equiv (b_1 \iff x \land y)\) which is equivalent to the following clauses

\((\overline{b_1} \lor x) \land (\overline{b_1} \lor y) \land (b_1 \lor \overline{x} \lor \overline{y})\) \hspace{1cm} (6)

In the same way, we construct the equivalent clauses of \(b_2\) and \(b_3\). The constant term 1 can be easily represented by the clause \((b \lor \overline{b})\) which is true in all cases. The addition (+) is equivalent to the logic operation (XOR). In these terms, We can generate the set of clauses of \((b_1 + b_2)\) as

\((b_1 \lor \overline{b_2}) \land (\overline{b_1} \lor b_2) \land (\overline{b_1} \lor \overline{b_2})\) \hspace{1cm} (7)

Using (6) and (7) we can convert a polynomial \(p\) from ANF to CNF. The method that we used is based on splitting the equations that contain more than a certain number (called the cutting number) of terms into shorter equations by adding new variables. This is motivated by the fact with this conversion method the number of variables grows very fast when representing long XOR chains. Then we write each equation in the CNF form as explained above. We found that 4 is the best cutting number for our attack.

In steps 3 and 4, we may pass the generated CNF file to any SAT solver and extract the values of the inner state \((a_1, \ldots, a_{93}, b_1, \ldots, b_{84}, c_1, \ldots, c_{111})\). Then we clock Trivium backward to recover the secret key \(K\) as explained in Proposition 1. In terms of corresponding the values of variables from CNF to ANF, we identify each 1 with “True” and 0 with “False”.
Algorithm 3

Require: \( m \) fault injection positions \((l_1, \ldots, l_m)\) and the vectors \( Z, Z^{(i)}, 1 \leq i \leq m \)
1: \( P \leftarrow \text{EQgenerator}(l_1, \ldots, l_m, Z, Z^{(1)}, \ldots, Z^{(m)}) \)
2: \( \text{CNF}(P) \leftarrow \text{Converting } P \text{ to a satisfiability problem in the CNF form} \)
3: \( \text{Solution} \leftarrow \text{Solve the satisfiability problem } \text{CNF}(P) \text{ by using a SAT solver} \)
4: \( \text{IS} \leftarrow \text{extract the inner state values } (a_1, \ldots, a_{n+93}, b_1, \ldots, b_{n+84}, c_1, \ldots, c_{n+111}) \)
5: \( \text{Recover the secret key } K \text{ from IS as explained in Proposition 1} \)
6: \( \text{return IS} \)

6 Experimental Results

We present our results to show how advanced solving techniques can improve the differential fault analysis of the stream cipher Trivium. We used our C++ implementation to generate the Trivium equations and the additional equations that are produced from DFA as in Section 4 and Albrecht’s converter [1] as explained in the previous section.

We used the SAT solver Minisat2 [15] to recover the inner state of Trivium at \( t = 0 \). We run all the experiments on an Intel(R) Core(TM)2 Duo CPU, each CPU is running at 2.8 GHz, and we used only one out of the two cores.

<table>
<thead>
<tr>
<th>( n )</th>
<th>( m )</th>
<th>degree 1</th>
<th>degree 2</th>
<th>time (sec.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>800</td>
<td>2</td>
<td>1216</td>
<td>2365</td>
<td>0.261</td>
</tr>
<tr>
<td>700</td>
<td>2</td>
<td>1088</td>
<td>2080</td>
<td>0.356</td>
</tr>
<tr>
<td>600</td>
<td>2</td>
<td>1005</td>
<td>1771</td>
<td>0.414</td>
</tr>
<tr>
<td>500</td>
<td>2</td>
<td>890</td>
<td>1437</td>
<td>0.127</td>
</tr>
<tr>
<td>450</td>
<td>2</td>
<td>831</td>
<td>1230</td>
<td>1.573</td>
</tr>
<tr>
<td>430</td>
<td>2</td>
<td>799</td>
<td>1138</td>
<td>1.936</td>
</tr>
<tr>
<td>420</td>
<td>2</td>
<td>769</td>
<td>1117</td>
<td>138.653</td>
</tr>
</tbody>
</table>

Table 4. Results of using Minisat2.

Table 4 shows the results of solving DFA Trivium equations when we insert several one-bit faults to random positions in any of the three registers (as explained in Alg. 2). For each case \( (n = 800, \ldots, 430) \) in Table 4, we have generated 100 systems and used only the equations of degree \( \leq 2 \). Each system has been solved 100 times by Minisat2. In case of \( n = 420 \), we have generated 10 systems and each system solved 10 times by Minisat2. Starting from \( n \leq 420 \), Minisat2 has taken significantly more time to solve the generated systems. This is due to the fact that the number of low degree equations in the generated system becomes lower. We report the average of these experiments. We have observed that the complexity of solving these systems is based on the number of linear equations generated by the EQgenerator procedure and the heuristic of the SAT solver.
7 Conclusion

We introduce an improvement to the differential fault analysis of Trivium from [13]. By using the SAT solver Minisat2 we could reduce the number of fault injections needed to recover the inner state of the cipher which leads to recovering the secret key. We show that our attack can recover the secret key of Trivium by using only two fault injections and 420 keystream output bits. As a future work, we plan to improve our attack to recover the secret key of Trivium using only one fault injection by applying more advanced conversion techniques and tuning SAT solver parameters.

Acknowledgments

The first author is supported by the BMBF project RESIST. The second author is partially supported by the German Science Foundation (DFG) grant BU 630/22-1. We want to thank the useful comments from Marcel Medwed and anonymous referees of the COSADE workshop on this paper and their valuable suggestions which helped to improve the paper.

References


