Distinguishing Attack on CryptMT

Shahram Khazaei Elham Shakour

Zaeim Electronic Industries, Tehran, Iran
{khazaei, shakour}@zaeim.com

Abstract. In this paper we propose a distinguishing attack on CryptMT, one of the ECRYPT stream cipher candidates, which needs 2^{50} bits of the output sequence with the same computational complexity. This is the first attack on this cipher.

Keywords. Stream Cipher, Distinguishing Attack, CryptMT, ECRYPT, Security Evaluation.

1 Introduction

Stream ciphers are widely used for fast encryption of sensitive data. Lots of old stream ciphers that have been formerly used can no longer be considered secure, because of their vulnerability to newly developed cryptanalysis techniques. In particular, the NESSIE project [1] did not select any of the proposed stream ciphers for its portfolio, as it was felt that none of the submissions was sufficiently strong. In order to create a portfolio of secure stream ciphers, the ECRYPT project [2] made a call for designs of new stream ciphers which led to submission of 35 proposals to the project by April 2005.

In [3] two stream ciphers called CryptMT and Fubuki has been proposed as ECRYP stream cipher [2] candidates. CryptMT is very simple but with a huge amount of internal state. It works with 32-bit words and produces an 8-bit byte in each step. The key and IV of the cipher accept any length up to 2048 bits (64 words). One of the generic attacks on stream ciphers is distinguishing attack whose aim is to distinguish the output sequence of a given stream cipher from a purely random sequence, with small error probability, faster than exhaustive search of the key space. In this paper we show that the LSB's (Least Significant Bits) of every two consecutive output bytes are equal with probability $\frac{1}{2}(1 + 2^{-24})$ which shows that the output sequence of CryptMT is distinguishable from a purely random one using about 2^{50} bits with the same computational complexity. Both kinds of error probability of the distinguisher are about 0.15.

2 Outline of CryptMT

The main part of CryptMT is the so-called MT (Mersenne Twister) generator, a linear finite state machine with 19937 bits internal state. The output sequence of MT has period equal to $2^{19937} - 1$ and produces a 32-bit word in each step. The initial state of MT is determined thorough a specified key-IV set up. Then the accumulative product of the output sequence of MT is computed modulo 2^{32}. The initial value of the accumulator is set to a non-zero value determined by the key-IV set up. The LSB of the MT output word is set to one before multiplication with the accumulator to ensure that the content of the accumulator will not be zero. The 8 most significant bits of the accumulator is considered as one byte of the output sequence of CryptMT. Let denote the output sequence of MT, the state sequence of the accumulator and the output sequence of CryptMT itself by $\{w_t\}_{t=1}^{\infty}$, $\{p_t\}_{t=0}^{\infty}$ and $\{z_t\}_{t=0}^{\infty}$ respectively, where $w_t, p_t \in \mathrm{GF}(2)^{32}$ and $z_t \in \mathrm{GF}(2)^8$. A complete description of CryptMT can be given by the following recursive equations

$$w_{t+1} = ((w_{t-622} \,\&\, c_1) \times c_2) \oplus w_{t-226} \oplus \big((w_{t-622} \,\&\, c_3) \,|\, (w_{t-623} \,\&\, c_4)\big) >> 1$$

$$p_{t+1} = p_t \times (w_t \,|\, c_1)$$

$$z_t = (p_t >> 24) \mod 256\,,$$

for $t \geq 0$ where the symbols &, ⊕, | and >> denote bitwise AND, XOR, OR and right-shift operations of 32-bit words. The symbol × denotes multiplication modulo 2^{32} and the constants c_1 to c_4 are defined by: c_1 = 0x00000001, c_2 = 0x9908b0df, c_3 = 0x7fffffff and c_4 = 0x80000000. The expression "mod 256" has been used to emphasis that $z_t \in \mathrm{GF}(2)^8$.The initial values of $\{w_t\}_{t=-623}^{-1}$ and p_0 are determined thorough a certain key-IV set up. The size of the key and the size of IV are up to 2048 bits (64 words). The designers have not directly claimed any security level for the cipher, but it seems that the cipher has been designed to provide 256 bits of security [4].

3 Description of the Attack

In this section we present our distinguishing attack on CryptMT. We assume that $\{w_t\}_{t=1}^{\infty}$ and $\{p_t\}_{t=0}^{\infty}$, the MT output sequence and the accumulator state sequence, are uniformly distributed. We define 128 32-bit integers u_0, $u_{1,}$ …., u_{127} where $u_k = k2^{25} + 1$. Let $A_0 = \{u_0\}$ and $A_j = \{u_k \mid k = 2^{7-j} + 2^{8-j} r,\ 0 \leq r < 2^{j-1}\}$ for $1 \leq j \leq 7$. In other words A_j ($1 \leq j \leq 7$) contains all 32-bit integers which their first and $(33\text{-}j)^{\text{th}}$ LSB is one, the $(j–1)$ MSB's are arbitrary and the rest $31–j$ bits are all zero. For every $u \in A_j$ ($0 \leq j \leq 7$), the 8 most significant bits of p_t and $p_t \times u$ are the same in the $8–j$ LSB's and can be either different or the same in the rest j MSB's. It is easy to show that for every $u \in A_j$ ($0 \leq j \leq 7$), the probability that the 8 most significant bits of p_t and $p_t \times u$ are the same is equal to 2^{-j} provided that p_t is uniformly distributed. Moreover, the probability that $w_t \,|\, c_1$ be equal to each one of the u_k's is equal to 2^{-31} for $0 \leq k \leq 127$. If the value of $w_t \,|\, c_1$ is not equal to none of the u_k's, the probability that the most significant 8 bits of p_t and $p_t \times u_k$ are the same is equal to 2^{-8}. Using the Total Probability Theorem, it can be inferred that z_t and z_{t+1} are equal with probability $(1\times2^{-31})\times1 + (1\times2^{-31})\times1/2 + (2\times2^{-31})\times1/4 + \ldots + (64\times2^{-31})\times1/128 + (1-128\times2^{-31})2^{-8} = 2^{-8} + 2^{-29}$; note that $|A_0| = 1$ and $|A_j| = 2^{j-1}$ for $1 \le j \le 7$. This bias can be detected using 2^{52} bytes (2^{55} bits) of the output sequence with error probability of about 0.15. The required computational complexity is $O(2^{52})$.

The data and time complexities can be reduced if we focus on the LSB of the output bytes of the CryptMT generator sequence. It is easy to verify that the LSB of the 8 most significant bits of p_t and $p_t \times u_k$, or equivalently the 25th LSB of them, are the same for each p_t and $0 \le k \le 127$. If the value of $w_t|c_1$ is not equal to none of the u_k's, the probability that the 25th LSB of p_t and $p_t \times u_k$ are the same is equal to ½. Again using the Total Probability Theorem, the LSB of z_t and z_{t+1} are equal with probability $128\times2^{-31} + ½ (1 - 128\times2^{-31}) = ½(1 + 2^{-24})$. This bias can be detected using 2^{50} bits of the output sequence with error probability of about 0.15. The required computational complexity is $O(2^{50})$.

4 Conclusion

In this paper we mounted a distinguishing attack on CryptMT, one of the ECRYPT stream cipher candidates. The result shows that the output sequence of this cipher is distinguishable from a purely random one using 2^{50} bits of the output sequence with the same computational complexity. The possibility of turning this distinguisher to a key-recovery attack remains an open problem.

References

1. NESSIE: New European Schemes for Signature, Integrity and Encryption, `http://www.nessie.eu.org/nessie/`.
2. eSTREAM, the ECRYPT Stream Cipher Project (2005) `http://www.ecrypt.eu.org/stream/`
3. Matsumoto M., Mariko H., Nishimura T. and Saito M.: Mersenne Twister and Fubuki Stream/Block Cipher. eSTREAM, ECRYPT Stream Cipher Project, Report 2005/003 (2005) `http://www.ecrypt.eu.org/stream/`.
4. Bernstein D. J.: Notes on the ECRYPT Stream Cipher project (eSTREAM) `http://cr.yp.to/streamciphers.html`