Some observations on ZUC-256

Alexander Maximov

Ericsson Research, Lund, Sweden
alexander.maximov@ericsson.com

Abstract. In this short paper we find an efficient binary approximation of the FSM of ZUC-256 with high correlation around $2^{-21.1}$ between the keystream words and the LFSR. Thereafter, we make a conjecture on a theoretical complexity of a hypothetical correlation attack on ZUC-256.

Keywords: 5G, ZUC-256

1 Introduction

ZUC-256 [ZUC18] is a stream cipher with the target to be used in 5G as one of the 256-bit security algorithms for confidentiality and integrity. At the moment, ZUC-256 is under evaluation by ETSI SAGE and understanding it's strength is therefore important.

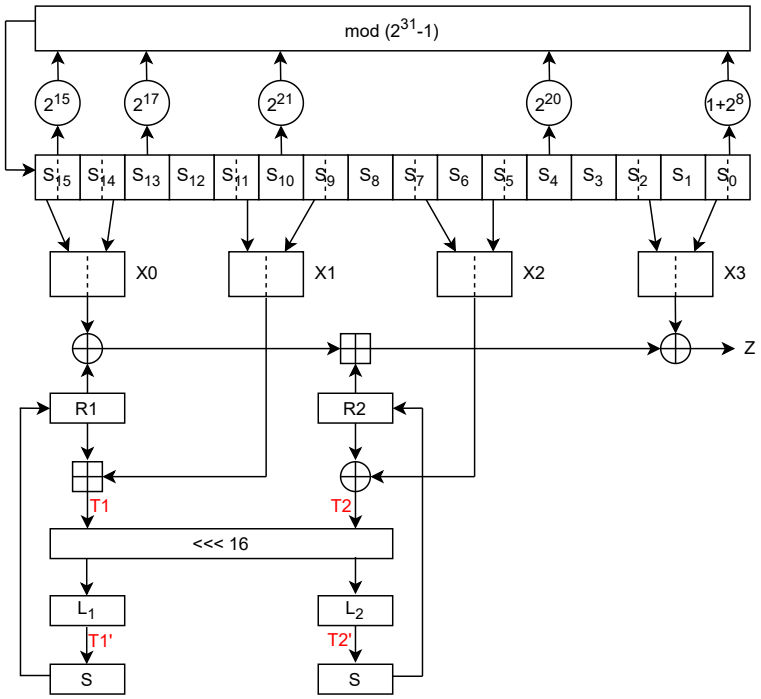

Figure 1: The keystream generation phase of the ZUC-256 stream cipher.

This work was inspired by recent results in [SJZ+21], and we decided to check whether similar methods may be applied to ZUC-256. The keystream generator of ZUC-256 is depicted on Figure 1, and for more details we refer to the original specification of the design [ZUC18]. We start by searching for a good binary approximation of the FSM and

find a strong correlation between the keystream and the LFSR. Then we make a conjecture regarding a hypothetical correlation attack on ZUC-256.

2 FSM approximation and correlation trails

We introduce additional intermediate signals $T1, T2$ and $T1', T2'$ as on Figure 1, then two consecutive keystream words can be expressed as follows:

$$\begin{aligned} z &= (((T1 \boxminus X1) \oplus X0) \boxplus (T2 \oplus X2)) \oplus X3 \\ z' &= ((S(T1') \oplus X0') \boxplus S(T2')) \oplus X3' \\ \text{where } (T1', T2') &= (L_1, L_2) \cdot (T1, T2)_{\lll 16} = \sigma(T1, T2) \end{aligned}$$

There is a linear relation between $(T1, T2)$ and $(T1', T2')$. Thus, if we have a pair of masks (m_1, m_2) for $(T1, T2)$, then corresponding masks for $(T1', T2')$ will be (m_1', m_2') such that $(m_1, m_2) \cdot (T1, T2) = (m_1', m_2') \cdot (T1', T2')$; the masks (m_1', m_2') can be derived linearly from (m_1, m_2), and vice versa.

The steps of the FSM approximation may be derived as follows:

$$\begin{aligned} \alpha \cdot z &= \alpha \cdot [(((T1 \boxminus X1) \oplus X0) \boxplus (T2 \oplus X2)) \oplus X3] \\ &\to \underbrace{m_0 \cdot [((T1 \boxminus X1) \oplus X0] \oplus m_2 \cdot [T2 \oplus X2]}_{\rho_{\boxplus}(\alpha, m_0, m_2)} \oplus \alpha \cdot X3 \\ &\to \underbrace{m_1 \cdot T1 \oplus m_1 \cdot X1}_{\rho_{\boxminus}(m_0, m_1, m_1)} \oplus m_0 \cdot X0 \oplus m_2 \cdot T2 \oplus m_2 \cdot X2 \oplus \alpha \cdot X3 \\ \beta \cdot z' &= \beta \cdot [((S(T1') \oplus X0') \boxplus S(T2')) \oplus X3'] \\ &\to \underbrace{k_1 \cdot (S(T1') \oplus X0') \oplus k_2 \cdot S(T2')}_{\rho_{\boxplus}(\beta, k_1, k_2)} \oplus \beta \cdot X3' \\ &\to \underbrace{k_1 \cdot S(m_1' \cdot T1')}_{\rho_S(k_1, m_1')} \oplus k_1 \cdot X0' \oplus \underbrace{k_2 \cdot S(m_2' \cdot T2')}_{\rho_S(k_2, m_2')} \oplus \beta \cdot X3', \end{aligned}$$

where $\rho_{\boxplus}(s, a, b), \rho_{\boxminus}(s, a, b), \rho_S(k, m)$ are correlation values for approximations of arithmetical additions, subtractions, and S-boxes, given input and output masks.

I.e., the biased binary correlation between two consecutive keystream words and the bits of the LFSR is thus expressed through X-terms, which are composed directly from the bits of the LFSR in the Bit-Reorganisation step of the cipher:

$$\alpha \cdot z \oplus \beta \cdot z' \to m_0 \cdot X0 \oplus m_1 \cdot X1 \oplus m_2 \cdot X2 \oplus \alpha \cdot X3 \oplus k_1 \cdot X0' \oplus \beta \cdot X3'.$$

For an efficient search of the masks and computation of correlation values, we utilised methods similar to [SJZ+21]. In our search for a good approximation trail with high correlation we found, e.g.:

$$\begin{aligned} \alpha &= 0x01860405 \\ m_0 &= 0x01860607 \\ m_1 &= 0x01040405 \\ m_2 &= 0x01010405 \\ m_1' &= 0x00040000 \\ m_2' &= 0x00010000 \\ k_1 &= 0x00300000 \\ k_2 &= 0x00200000 \\ \beta &= 0x00200000 \end{aligned}$$

$$\begin{aligned} \rho_{\boxplus}(\alpha, m_0, m_2) &= +2^{-10.000000} \\ \rho_{\boxminus}(m_0, m_1, m_1) &= +2^{-4.000000} \\ \rho_{\boxplus}(\beta, k_1, k_2) &= +2^{-1.000000} \\ \rho_S(k_1, m_1') &= -2^{-3.415037} \\ \rho_S(k_2, m_2') &= +2^{-3.192645} \\ \rho_{tot} &= -2^{-21.607683} \end{aligned}$$

Simulation results. Because of the reality might be different from the theory, e.g., due to dependencies in the above sequence of approximations, and in order to confirm the above correlation, we run simulations of ZUC-256 and collected 2^{53} samples directly from the keystream generator. The resulting correlation value from simulations is:

$$\rho_{sim}(\text{from } 2^{53} \text{ samples}) \approx -2^{-21.093495},$$

which is the value of a high confidence. These simulations confirm the found correlation, and show that it is actually slightly stronger than the theoretical one ρ_{tot}.

3 A hypothetical correlation attack on ZUC-256

Correlation attacks presented in [GZ21] and [SJZ+21] are quite generic for this class of stream ciphers where LFSR is involved and a biased parity check expression is available. However, the methods are only given for when the LFSR is over a field of characteristic 2, i.e., a *binary LFSR*. In ZUC-256 we, however, have to deal with a *prime LFSR* where the base field has characteristic larger than 2; in case of ZUC-256 it is the prime $p = 2^{31} - 1$.

Intuitively, a correlation attack starts with a system where we have n-bits of entropy (e.g., the length of the LFSR in bits). Then we use the found correlation as a *biased binary parity check*. We then "inject" one such parity check into the system and by this the entropy of the system is reduced. By injecting many enough of such parity checks, the system becomes more and more determined, i.e., the LFSR is then recovered. We believe that the performance of such an attack for *prime LFSRs* should be similar as for *binary LFSRs*, therefore we make the following conjecture:

Conjecture 1. *Given a biased binary parity check on the LFSR state bits, the complexity to recover a prime LFSR with n bits of entropy should be similar to the complexity of recovering a binary LFSR of length n bits.*

At this moment we do not know an exact method how to use a *biased binary parity check* for recovering a *prime LFSR*, and we leave this as an open question. However, if the above conjecture is true, then given the correlation $2^{-21.1}$ and $n \approx 496$ bits ($n \leftarrow 16 \log_2(2^{31}-1)$), there should exist a correlation attack on ZUC-256 with complexities around $T \approx M \approx D \approx 2^{176}$, derived as in [GZ21].

References

[GZ21] Xinxin Gong and Bin Zhang. Resistance of SNOW-V against fast correlation attacks. *IACR Transactions on Symmetric Cryptology*, 2021(1):378–410, Mar. 2021.

[SJZ+21] Zhen Shi, Chenhui Jin, Jiyan Zhang, Ting Cui, and Lin Ding. A correlation attack on full SNOW-V and SNOW-Vi. Cryptology ePrint Archive, Report 2021/1047, 2021. https://ia.cr/2021/1047.

[ZUC18] ZUC design team. The ZUC-256 Stream Cipher, 2018. http://www.is.cas.cn/ztzl2016/zouchongzhi/201801/W020180126529970733243.pdf.