LI Chao, WANG Wenling, HU Pengsong

Differential attack on nonlinear combined sequences

Abstract By using the coding properties and statistic properties of the plaintext, the differential properties of the key stream sequences generated by a nonlinear combined generator were analyzed. Then a differential attack algorithm on the nonlinear combined sequences was proposed. At last, an attack example adopting the differential attack algorithm was presented.

Keywords combined generator, differential attack, differential position set, differential validity

1 Introduction

Nonlinear combined generators based on linear feedback shift register (LFSR) and nonlinear combined function have wide applications in the design and analysis of stream ciphers. Currently, two kinds of the most efficient attack approaches to such systems are correlation attacks and linear cryptanalysis, both of which make use of certain correlation properties between the input and output of the combined function. A correlation attack called “divide-and-conquer” was proposed in Ref. [1], and a fast correlation attack by using low-density parity codes to conduct iterative probability decoding was presented in Ref. [2]. Some improvements on correlation attacks have been discussed in Refs. [3,4].

The differential attack proposed by Biham and Shamir in the early 1990s has wide applications in the research of block ciphers [5]. However, it has rarely been used in the analysis of stream ciphers for a long time [6–8]. The differential attack proposed in this paper is a kind of cipher-text-only cryptanalysis. By using the coding properties and the statistic properties of the plaintext, the differential properties of the key stream sequences which are generated by a nonlinear combined generator are analyzed. Then a differential attack algorithm on the nonlinear combined sequences is proposed, and an attack example using the differential attack algorithm is presented.

2 Basic concepts of difference

Definition 1 Let \( A = \{a_1, a_2, \ldots, a_r\} \) be a binary sequence of length \( r \), and \( S = \{(i_1, j_1), (i_2, j_2), \ldots, (i_t, j_t)\} \), where the elements of the set \( S \) are two-dimensional arrays in the set of positive integers. If \( 1 \leq i_s, j_s \leq r \) holds for any \( 1 \leq k \leq t \), then the set \( \{a_{i_s} \oplus a_{j_s} \mid (i_s, j_s) \in S\} \) is called a differential set of \( A \) with respect to \( S \), and \( S \) is called a differential position set of \( A \). The whole differential position sets of \( A \) is denoted as \( U_a \).

Definition 2 Let \( A = \{a_1, a_2, \ldots, a_r\} \) be a binary sequence of length \( r \). \( S \) is a differential position set of \( A \) and \( \Delta A \) is a differential set of \( A \) with respect to \( S \). Let \( P_A = P\{x = 0 \mid x \in \Delta A\} \). If \( |S| > 0 \), then \( A \) is called \( \delta \)-differential validity with respect to \( S \). Otherwise, it is called differential invalidity with respect to \( S \). Similarly, differential validity can be defined by \( P_A = P\{x = 1 \mid x \in \Delta A\} \).

First, the coding properties and statistic properties of the plaintext should be briefly introduced. As is well known, the plaintext needs to be encoded into binary stream data before it is encrypted. The plaintext stream data always reflect some designated information in a certain circumstance and consequently it displays some statistical features. For example, word-based stream data can reflect the distinct style or manner of writing. The stream data based on voice signals and pictures also have obvious statistical features. As a result, it is supposed that the plaintext stream data \( M = \{m_1, m_2, \ldots, m_t\} \) has the following property \( T \): plaintext stream data \( M \) is the sum of all the efficient information expressed in some certain coding modes. The coding mode of the plaintext is not emphasized for the sake of convenience.

Definition 3 Let plaintext stream \( M = \{m_1, m_2, \ldots, m_t\} \) possess the property \( T \) and \( U_{\alpha} \) be the whole differential position set of \( M \). For any positive integer \( i \), \( S \) is called the plaintext differentially optimal for \( i \). If \( S \) is a subset of \( U_{\alpha} \), then \( |S| = i \) and \( |P_S M - 1/2| = \max \{1/2 \mid P_S M - 1/2 \mid S \in U_{\alpha} \} \).
3 Differential attack

The conditions and models for cryptanalysis will be given at first. For a nonlinear combined generator (Fig. 1), suppose that the characteristic polynomials of those LFSRs are known, and they are primitive polynomials with degree $r_i (1 \leq i \leq n)$ respectively. Let nonlinear combined function $f(x_1, x_2, \ldots, x_n)$ be non-correlation-immune balanced function. The cipher-text stream is denoted by $C = \{c_1, c_2, \ldots, c_n\}$. The output signals for the former $N$ steps of LFSR are denoted by $\{x'_1, x'_2, \ldots, x'_N\} (1 \leq i \leq n)$ and the corresponding initial state is $X'$, which means $X' = \{x'_1, x'_2, \ldots, x'_n\} \in F_2^n$. The output of $f(x_1, x_2, \ldots, x_n)$ (i.e. key stream sequence) is denoted as $Z = \{z_1, z_2, \ldots, z_N\}$.

![Fig. 1 Nonlinear combined sequence cipher](image)

Due to the primitive polynomials of LFSRs, the periods of the output sequences generated by these LFSRs with the non-zero initial states are maximal. Therefore, it can be assumed that the inputs of combined function $f$ are all independent binary random variables with the same distribution. Furthermore, for any $1 \leq i \leq n$ and $1 \leq k \leq N$, $P(x'_i = 0) = P(x'_i = 1) = 1/2$ holds. Generally speaking, a key stream sequence, it should have favorable pseudo-random characteristics. Therefore, the outputs of the combined function $f$ can also be considered as some independent binary random variables with the same distribution for any $1 \leq k \leq N$. Thus $P(z_i = 0) = P(z_i = 1) = 1/2$ is approximately obtained.

**Definition 4** If the sequence $\{x'_1, x'_2, \ldots, x'_N\} (1 \leq i \leq n)$ is generated by the $i$th LFSR with the initial state $X' \in F_2^n$, then a set of two-dimensional vectors $\lambda'_{X'} (1 \leq i \leq n)$ is defined as follows:

$$\lambda'_{X'} = \{(k, l)|x'_k \oplus x'_l = 0, 1 \leq k, l \leq N\} \quad (1)$$

**Lemma 1** Let the key stream sequence $Z = \{z_1, z_2, \ldots, z_N\}$ be generated by the above model with the initial state $X' \in F_2^n (1 \leq i \leq n)$, and the set $\lambda'_{X'}$ defined by Eq. (1). Then there exists $i (1 \leq i \leq n)$ and $\epsilon (0 < \epsilon \leq 1/2)$ such that $P\{z_i \oplus z_i = 0|(k, l) \in \lambda'_{X'}\} = 1/2 + \epsilon$.

**Proof** Because $x'_i (1 \leq i \leq n)$ can be considered as some independent binary random variables with the same distribution, then $P(x'_i = 0) = P(x'_i = 1) = 1/2$ and note that $f(x_1, x_2, \ldots, x_n)$ is a non-correlation-immune balanced function, there exists $i (1 \leq i \leq n)$ and $\xi (0 < |\xi| \leq 1/2)$ such that $P\{f(x'_i, \ldots, x'_n) \oplus x'_i = 0\} = 1/2 + \xi$. Thus

$$P\{z_i \oplus z_i = 0|(k, l) \in \lambda'_{X'}\}$$

$$= P\{f(x'_i, \ldots, x'_n) \oplus f(x'_i, \ldots, x'_n) = 0|\xi, l \in \lambda'_{X'}\}$$

$$= P\{f(x'_i, \ldots, x'_n) \oplus x'_i \oplus f(x'_i, \ldots, x'_n) \oplus x'_i = 0\}$$

$$= P\{f(x'_i, \ldots, x'_n) \oplus x'_i = 0\} \cdot P\{f(x'_i, \ldots, x'_n) \oplus x'_i = 0\}$$

$$+ P\{f(x'_i, \ldots, x'_n) \oplus x'_i = 1\} \cdot P\{f(x'_i, \ldots, x'_n) \oplus x'_i = 1\}$$

$$= \left(1 - \frac{1}{2} + \xi\right) + \left(1 - \frac{1}{2} - \xi\right)^2$$

$$= 1 - 2\xi^2$$

$0 < 2\xi^2 \leq 1/2$ is obtained as $0 < |\xi| \leq 1/2$. Let $\xi = 2\zeta$, then Lemma 1 is proved.

**Lemma 2** Let $Z = \{z_1, z_2, \ldots, z_N\}$ be generated by the above model with the initial state $X' \in F_2^n (1 \leq i \leq n)$. Then for any $i$ and $X' \in F_2^n$, if $X' \neq X'$, $P\{z_i \oplus z_i = 0|(k, l) \in \lambda'_{X'}\} = 1/2$ is obtained, where $\lambda'_{X'}$ is defined by Eq. (1).

**Proof** For any $i$ and $X' \in F_2^n$, let $\{x'_i, x'_i, \ldots, x'_i\}$ be generated by the $i$th LFSR in the above model with the initial state $X'$, so the following formula can be obtained:

$$P\{z_i \oplus z_i = 0|(k, l) \in \lambda'_{X'}\}$$

$$= P\{f(x'_i, x'_i, \ldots, x'_i) \oplus f(x'_i, x'_i, \ldots, x'_i) = 0|\xi', l \in \lambda'_{X'}\}$$

$$= P\{f(x'_i, x'_i, \ldots, x'_i) \oplus x'_i \oplus f(x'_i, x'_i, \ldots, x'_i) \oplus x'_i = 0\}$$

$$= P\{f(x'_i, x'_i, \ldots, x'_i) \oplus x'_i = 0\} \cdot P\{f(x'_i, x'_i, \ldots, x'_i) \oplus x'_i = 0\}$$

$$+ P\{f(x'_i, x'_i, \ldots, x'_i) \oplus x'_i = 1\} \cdot P\{f(x'_i, x'_i, \ldots, x'_i) \oplus x'_i = 1\}$$

$$= \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^2 \frac{1}{2} = \frac{1}{2}$$