Binary Additive Counter Stream Ciphers

Cunsheng Ding*, Wenpei Si†

Abstract

Although a number of block ciphers have been designed and are available in the public domain, they are usually used in one of the four modes: the cipher block chaining mode, the cipher feedback mode, the output feedback mode, and the counter mode. In all these cases, a stream cipher is actually used, as any block cipher used in any of these modes becomes a stream cipher. Stream ciphers are preferred, as they can destroy statistical properties of natural languages to some extent. The objective of this paper is to provide the state-of-the-art of a special type of stream ciphers, called binary additive counter stream ciphers, by surveying known results in the literature, deriving design criteria, and presenting experimental results. Two examples of binary additive counter stream ciphers are analysed in details, and are used to illustrate that it is possible to construct a practical stream cipher with many security properties. The security of the two ciphers with respect to known plaintext attacks is proven to be equivalent to the computational complexity of a mathematical problem.

2000 Mathematics Subject Classification: 11T71, 68P25, 94A55, 94A60.
Keywords and Phrases: Additive synchronous stream ciphers, Counter generator, Difference sets, Highly nonlinear functions.

1 Introduction

Ciphers are classified into stream and block ciphers, depending on whether or not the encryption transformation is time-varying. In most applications, stream ciphers are preferred, as they can destroy statistical properties of natural languages to some extent.

The only cipher which is provably secure in the information sense and simple in structure is the one-time pad, which is not practical for real applications. Ciphers employed in real systems are usually complex in structure and it is thus hard to analyse and prove their security. Two open problems in cryptography are the following:

*Department of Computer Science and Engineering, The Hong Kong University of Science and Technology, Clearwater Bay, Kowloon, Hong Kong, China. Email: cd ding@cse.ust.hk.
†Department of Computer Science and Engineering, The Hong Kong University of Science and Technology, Clearwater Bay, Kowloon, Hong Kong, China. Email: si wenpei@ust.hk.
1. Is there a practical cipher with provable security in terms of computational complexity?
2. If there is a practical cipher with provable security in terms of computational complexity, how do we design it?

One simple and natural type of stream ciphers is the binary additive counter stream ciphers depicted in Figure 1, where the keystream generator consists of a cyclic counter with period $N$ and a function $f$ from $\mathbb{Z}_N := \{0, 1, 2, \cdots, N - 1\}$ to $\mathbb{Z}_2 := \{0, 1\}$, where $N$ is a huge integer. The cyclic counter has a memory unit and counts the integers in $\mathbb{Z}_N$ cyclically. The initial content of the memory unit of the cyclic counter is the secret key, which could be any integer between 0 and $N - 1$. If the secret key is $k$, the keystream bit $k_t$ at time unit $t$ is then $k_t = f((t + k) \mod N)$. The encryption of a message bit is the exclusive-or of the message bit and the corresponding keystream bit. The decryption process is the same as the encryption process.

The objectives of this paper are to survey all known results scattered over a number of references and present new ones about the binary additive counter stream ciphers of Figure 1. This is to provide the reader with the state-of-the-art of the binary additive counter stream ciphers. The paper is organized as follows. Section 2 presents a number of design criteria for the binary additive counter stream ciphers depicted in Figure 1. Section 3 documents an example of the binary additive counter stream ciphers, called Legendre cipher, and its security properties. Section 4 describes another example of the binary additive counter stream ciphers, called two-prime cipher, and its security properties. Section 5 provides information on functions $f$ from $\mathbb{Z}_N$ to $\mathbb{Z}_2$ with optimal nonlinearity $p_f$ which may be employed in the ciphers of Figure 1, and concludes this paper.

In this paper, the security of the Legendre and two-prime ciphers with respect to known plaintext attacks is proven to be equivalent to the computational complexity of two number-theoretic problems. This is the first time that the security of a cipher with respect to known plaintext attacks is proved to be equivalent.
to the computational complexity of a mathematical problem.

2 Possible attacks and design criteria

2.1 The linear and sphere complexity attacks and the associated design criteria

2.1.1 The linear complexity attack and the design criterion associated to this attack

Let $z_n = z_0z_1 \cdots z_{n-1}$ be a sequence of length $n$ over the finite field $\mathbb{F}_q$. The linear complexity (also called the linear span) of the sequence $z^n$ is defined to be the smallest nonnegative integer $L$ such that there exist constants $c_1, c_2, \cdots, c_{L-1} \in \mathbb{F}_q$ for which

$$z_j + c_1z_{j-1} + \cdots + c_{L-1}z_{j-L} = 0, \text{ for all } L \leq j < n.$$  

This definition applies also to semi-infinite sequences $z^\infty = z_0z_1 \cdots$ over $\mathbb{F}_q$, where $n = \infty$. For an ultimately periodic sequence $z^\infty$ over $\mathbb{F}_q$, the linear complexity must be a finite number. The corresponding polynomial $1 + c_1x + c_2x^2 + \cdots + c_Lx^L \in \mathbb{F}_q[x]$ is called the minimal polynomial of the sequence. In engineering terms, the linear complexity is the length of the shortest linear feedback shift register that can produce the sequence, where the minimal polynomial is called the feedback polynomial of the linear feedback shift register (LFSR).

If the linear complexity of the output sequence of the counter generator is $L$, then $2L$ consecutive output bits of the counter generator can be used to construct an LFSR of length $L$ that produces the same keystream sequence. The equivalent LFSR can be constructed using the Berlekamp-Massey algorithm or by solving a system of linear equations. Hence, the keystream sequence of an additive synchronous stream cipher must have large linear complexity.

**Design Criterion 1.** The linear complexity of the keystream sequence of the binary additive counter stream cipher in Figure 1 should be large.

2.1.2 The linear complexity stability attack and the associated design criterion

Although the linear complexity of a keystream sequence may be very large, there might be another sequence with very low linear complexity such that the Hamming distance between the two sequences is very small. If this is the case, one can use the sequence with low linear complexity to approximate the original keystream sequence. In other words, in this case one can construct an LFSR with short length to approximate the original keystream generator.

If changing a small number of entries in a sequence decreases the linear complexity of the sequence to a large extent, we say that the linear complexity of the original sequence is not stable. The linear complexity stability issue was observed in 1989 ([7]) and a measure of the linear stability (called weight complexity) was
introduced there. Shortly afterwards, the sphere complexity for both finite and periodic sequences was introduced in the monograph [16], as a measure of the linear complexity stability.

Let $x^n$ be a sequence of length $n$ over $\text{GF}(q)$, and let $\ell$ be any integer with $0 < \ell < n$. The sphere complexity of $x^n$ is defined to be

$$SC_\ell(x^n) = \min_{0 < W_H(y^n) \leq \ell} \text{LC}(x^n + y^n),$$

where $y^n$ is any sequence of length $n$ over $\text{GF}(q)$, $W_H(y^n)$ denotes the Hamming weight of $y^n$, and $\text{LC}(x^n)$ is the linear complexity of the sequence $x^n$.

Let $x^\infty$ be a sequence of period $n$ (not necessarily the least period) over $\text{GF}(q)$, and let $\ell$ be any integer with $0 < \ell < n$. The sphere complexity of $x^\infty$ is defined to be

$$SC_\ell(x^\infty) = \min_{\text{Per}(y^\infty) = n, W_H(y^\infty) \leq \ell} \text{LC}(x^\infty + y^\infty),$$

where $y^n$ denotes the first periodic segment of the sequence $y^\infty$ over $\text{GF}(q)$, Per($x$) is the period of $x$, and $\text{LC}(x^\infty)$ is the linear complexity of the sequence $x^\infty$.

The sphere complexity was introduced in 1991 in [16], two years earlier than the $\ell$-error linear complexity, which is defined to be $\min\{\text{LC}(x^\infty), SC_\ell(x^\infty)\}$. Clearly, the $\ell$-error linear complexity is nothing new, but the minimum of the two earlier measures: linear complexity and sphere complexity.

Based on the linear complexity stability, the best affine approximation (BAA) attack on certain stream ciphers was developed in [16, Chapter 3]. For the binary additive counter stream ciphers of Figure 1, one can construct an LFSR to approximate the original keystream cipher if the sphere complexity $SC_\ell(s^\infty)$ of the keystream sequence is small for small $\ell$. Hence, another design requirement is that the sphere complexity $SC_\ell(s^\infty)$ of the keystream sequence of the binary additive counter stream ciphers should be large enough for small $\ell$.

**Design Criterion 2.** The sphere complexity $SC_\ell(s^\infty)$ of the keystream sequence of the binary additive counter stream ciphers in Figure 1 should be large enough for small $\ell$.

### 2.1.3 The control of the linear and sphere complexity

The linear complexity and sphere complexity of periodic sequences can be controlled easily as follows [9].

**Proposition 1.** ([9]) Suppose $N = p_1^{e_1} \cdots p_t^{e_t}$, where $p_1, \ldots, p_t$ are $t$ pairwise distinct primes, and $q$ is a power of a prime such that $\gcd(q, N) = 1$. Then for each nonconstant sequence $x^\infty$ of period $N$ over $\text{GF}(q)$,

$$\text{LC}(x^\infty) \geq \min\{\text{ord}_{p_1}(q), \ldots, \text{ord}_{p_t}(q)\},$$

$$SC_k(x^\infty) \geq \min\{\text{ord}_{p_1}(q), \ldots, \text{ord}_{p_t}(q)\}, \text{ if } k < \min\{W_H(x^N), N - W_H(x^N)\},$$

where $W_H(x^N)$ denotes the Hamming weight of the first periodic segment $x^N$ of the sequence $x^\infty$, and $\text{ord}_{p_i}(q)$ is the order of $q$ modulo $p_i$. 

Hence, choosing the period $N$ of the counter is one way to control the linear complexity and its stability for the whole periodic sequence (see [9, 6] for details). However, it is more important to control the linear and sphere complexity for any segment of the keystream sequence. This is in general a hard problem.

2.2 The key approximation attack and the associated design criterion

The idea of this attack is to use a key $k'$ to decrypt the ciphertext obtained from encrypting the plaintext with a key $k$. Let $s^\infty$ denote the keystream sequence of the binary additive counter stream cipher of Figure 1. The rate of correct decryption of this attack with respect to the whole period is given by $(1 + \text{AC}_s(k - k'))/2$, where

$$\text{AC}_s(\tau) = \frac{1}{N} \sum_{i=0}^{N-1} (-1)^{s_{i+\tau} - s_i}$$

is the periodic autocorrelation value at shift $\tau = k - k'$ of the keystream sequence $s^\infty$, i.e., $s_i = f(i)$ for all $i$, and $N$ is the period of the counter.

To thwart this attack, $|\text{AC}_s(\tau)|$ should be as small as possible for all $1 \leq \tau \leq N - 1$. However, we have

$$\sum_{\tau=1}^{N-1} \text{AC}_s(\tau) = (N - 2W_H(s^N))^2/N - 1,$$

$W_H(s^N)$ denotes the Hamming weight of the first period of the keystream sequence.

Note that

$$\text{AC}_s(\tau) = \frac{1}{N} \sum_{i=0}^{N-1} (-1)^{s_{i+\tau} - s_i}$$

$$= \frac{1}{N} \sum_{i=0}^{N-1} (-1)^{f(i+\tau) - f(i)}$$

$$= \frac{1}{N} \left| \{x \in \mathbb{Z}_N : f(x+\tau) - f(x) = 0\} - \{x \in \mathbb{Z}_N : f(x+\tau) - f(x) = 1\} \right|$$

$$= \frac{2}{N} \left| \{x \in \mathbb{Z}_N : f(x+\tau) - f(x) = 0\} \right| - N$$

$$= \frac{N - 2\{x \in \mathbb{Z}_N : f(x+\tau) - f(x) = 1\}}{N}.$$

We have

$$\max_{1 \leq \tau \leq N-1} |\text{AC}_s(\tau)| = \max_{1 \leq \tau \leq N-1} \left| \frac{2}{N} \left| \{x \in \mathbb{Z}_N : f(x+\tau) - f(x) = 0\} \right| - N \right|$$

$$= \max_{1 \leq \tau \leq N-1} \left| \frac{N - 2\{x \in \mathbb{Z}_N : f(x+\tau) - f(x) = 1\}}{N} \right|.$$
It follows that $\max_{1 \leq \tau \leq N-1} |AC_s(\tau)|$ is minimal if and only if
\[
p_f := \max_{a \neq 0} \max_{b \in \{0, 1\}} \frac{|\{x \in \mathbb{Z}_N : f(x + a) - f(x) = b\}|}{N}
= \max_{0 < \tau < N} \frac{\max\{1 + AC_s(\tau), 1 - AC_s(\tau)\}}{2}
\] is minimal.

The quantity $p_f$ is a measure of nonlinearity of $f$. The smaller the $p_f$, the higher the nonlinearity of $f$. It is proved that $p_f \geq \frac{1}{2}$ for any function from $\mathbb{Z}_N$ to $\mathbb{Z}_2$ [4]. To thwart this attack, the underlying function $f$ should have high nonlinearity [4, 5]. So another design criterion is the following.

**Design Criterion 3.** For the underlying function $f$ of the binary additive counter stream ciphers in Figure 1, the nonlinearity $p_f$ should be as close to $1/2$ as possible.

This requirement is the same as that $\max_{1 \leq \tau \leq N-1} |AC_s(\tau)|$ be as small as possible.

For real applications, only a small segment of the whole period of the keystream sequence is used for encryption. If a segment of length $L$ of the keystream is used for encryption, then the rate of correct decryption is given by $(1 + AC_s^{(L)}(k-k'))/2$, where
\[
AC_s^{(L)}(\tau) = \frac{1}{L} \sum_{i=0}^{L-1} (-1)^{s_i + \tau - s_i}.
\]

In general, $\max_{1 \leq \tau \leq N-1} |AC_s^{(L)}(\tau)| = 1$ if $1 \leq L < \log_2 N$ and the sequence is well designed. However, as $L$ approaches to $N$, $\max_{1 \leq \tau \leq N-1} |AC_s^{(L)}(\tau)|$ should approach $1/N$ or $3/N$ if $N$ is odd and to $0$ or $2/N$ if $N$ is even.

**Design Criterion 4.** For the keystream sequence of the binary additive counter stream cipher, the maximum value $\max_{1 \leq \tau \leq N-1} |AC_s^{(L)}(\tau)|$ should be close to $1/N$ when $L$ is close to $N$.

We call the sequence, $(L, \max_{1 \leq \tau \leq N-1} |AC_s^{(L)}(\tau)|)_{\tau=1}^{N-1}$, the periodic autocorrelation profile of the sequence $s^\infty$ of period $N$.

Closely related to the periodic autocorrelation profile are the aperiodic autocorrelation values
\[
AAC_s(u) := (N - u)AC_s^{(N-u)}(u) = \sum_{i=0}^{N-u-1} (-1)^{s_i + u - s_i},
\]
where $0 \leq u < N$. The aperiodic autocorrelation is certainly of cryptographic importance as it is a measure of the correlation between two segments of the keystream sequence.
It is easily seen that
\[ N \times AC_s(u) = AAC_s(u) + AAC_s(N - u), \quad 0 < u < N. \] (5)

Hence, a binary sequence with optimal periodic autocorrelation may not have good aperiodic autocorrelation property. However, a binary sequence with optimal aperiodic autocorrelation should have good periodic autocorrelation.

In general, it is quite difficult to compute the aperiodic autocorrelation values of a binary sequence. As a measure of the aperiodic autocorrelation values collectively, Golay [18] introduced in 1972 the merit factor of a binary sequence \( x \) of length \( n \), which is defined by
\[
MF(x) := \frac{n^2}{2 \sum_{u=1}^{n-1} |AAC_x(u)|^2}.
\] (6)

For any fixed \( n \), the larger the merit factor \( MF(x) \), the smaller the sum \( \sum_{u=1}^{n-1} |AAC_x(u)|^2 \). Hence, the keystream sequence of the binary additive counter stream cipher of Figure 1 should have a high merit factor. For details of the merit factor of binary sequences and its applications in other engineering areas, the reader is referred to the survey paper [22].

2.3 The linear approximation attack and the associated design criterion

Let \( g_1 \) and \( g_2 \) be two functions from a finite Abelian group \( A \) to an Abelian group \( B \). The Hamming distance of the two functions is defined to be \( |\{ x \in A : g_1(x) \neq g_2(x) \}| \). The Hamming weight of \( g_1 \) is defined to be \( |\{ x \in A : g_1(x) \neq 0 \}| \). A function \( g \) from \( A \) to \( B \) is called linear if \( g(x + y) = g(x) + g(y) \) for all \( x \in A \) and \( y \in A \). \( g \) is called an affine function if \( g = g_1 + b \) for a linear function \( g_1 \) and a constant \( b \in B \).

Every well-designed cipher should have both linear and nonlinear building blocks. In the Data Encryption Standard, the only nonlinear building blocks are the eight S-boxes. The basic idea of a linear approximation attack on an additive synchronous stream cipher is to use a linear function to replace a nonlinear function in a cipher. In this way, another decryption transformation is obtained and used to approximate the original decryption transformation, under the assumption that a piece of the keystream is known. To have a high rate of correct decryption, the Hamming distance between the nonlinear function and the linear function should be as small as possible. The best affine approximation (BAA) attacks on two types of stream ciphers developed in 1991 in [16, Chapter 3] were based on this idea.

The best affine approximation attack on the binary additive counter stream ciphers is a known-plaintext attack and is as follows. Assume that \( s_is_{i+1}\cdots s_{i+t-1} \) is a piece of known keystream. Consider all affine functions from \( \mathbb{Z}_N \) to \( \mathbb{Z}_2 \). Choose the one with the smallest Hamming distance with the underlying function \( f \) in
Figure 1. Use it to replace $f$ and use the obtained generator to approximate the original keystream generator. Finally, use the known keystream $s_is_{i+1}\cdots s_{i+t-1}$ to compute a state content for the approximation generator.

To thwart the affine approximation attacks, the Hamming distance between the underlying function $f$ and any affine function from $\mathbb{Z}_N$ to $\mathbb{Z}_2$ should be as large as possible. However, for any affine function $A$ from $\mathbb{Z}_N$ to $\mathbb{Z}_2$ the sum of the Hamming distance between $f$ and $A$ and that between $f$ and $A+1$ is $N$. Hence, the best situation is that the Hamming distance between $f$ and every affine function is $N/2$ for even $N$ and $(N-1)/2$ for odd $N$.

**Design Criterion 5.** For the underlying function $f$ of the binary additive counter stream ciphers in Figure 1, the Hamming distance between $f$ and every affine function from $\mathbb{Z}_N$ to $\mathbb{Z}_2$ should be as close to $N/2$ as possible.

### 2.4 The information stability attack and the associated design criterion

Define $D_i = \{0 \leq x \leq N-1 : f(x) = i\}$ for $i = 0$ and $i = 1$. Given any $t$ consecutive keystream bits $s_ks_{k+1}\cdots s_{k+t-1}$, where $0 \leq k \leq N-1$, we have

$$k \in D_{s_k} \cap (D_{s_{k+1}} - 1) \cap (D_{s_{k+2}} - 2) \cap \cdots \cap (D_{s_{k+t-1}} - (t-1)).$$

So the mutual information between $k$ and $s_ks_{k+1}\cdots s_{k+t-1}$ is

$$\log N - \log |D_{s_k} \cap (D_{s_{k+1}} - 1) \cap (D_{s_{k+2}} - 2) \cap \cdots \cap (D_{s_{k+t-1}} - (t-1))|.$$

So the larger the $|D_{s_k} \cap (D_{s_{k+1}} - 1) \cap (D_{s_{k+2}} - 2) \cap \cdots \cap (D_{s_{k+t-1}} - (t-1))|$, the less the amount of information about the key $k$ given by the keystream segment $s_ks_{k+1}\cdots s_{k+t-1}$.

However, we have the following conservation

$$\sum_{u_0=0}^1 \cdots \sum_{u_{t-1}=0}^1 |D_{u_0} \cap (D_{u_1} - 1) \cap (D_{u_2} - 2) \cap \cdots \cap (D_{u_{t-1}} - (t-1))| = N.$$

So it is necessary to balance the amount of information about $k$ given by all possible consecutive keystream segments of length $t$. The ideal situation is

$$d_{u_0u_1\cdots u_{t-1}}(0, 1, 2, \cdots , t-1) := |D_{u_0} \cap (D_{u_1} - 1) \cap (D_{u_2} - 2) \cap \cdots \cap (D_{u_{t-1}} - (t-1))| \approx \frac{N}{2^t} \tag{7}$$

for every $u_0u_1\cdots u_{t-1}$ and every $1 \leq t < \log_2 N$.

**Design Criterion 6.** For the underlying function $f$ of the binary additive counter stream ciphers in Figure 1, the quantity

$$|D_{u_0} \cap (D_{u_1} - 1) \cap (D_{u_2} - 2) \cap \cdots \cap (D_{u_{t-1}} - (t-1))|$$

should be as close to $N/2^t$ as possible for each possible binary string $u_0u_1\cdots u_{t-1}$ and for each $1 \leq t < \log_2 N$. 

The information stability attack is a known-plaintext attack, and works as follows. Assume that a cryptanalyst has got a piece of keystream \( s_k s_{k+1} \cdots s_{k+t-1} \), where \( 0 \leq k \leq N - 1 \). His task is to find out the integer \( k \). He can carry out the following steps:

1. Compute \( D_0 \) and \( D_1 \) using the function \( f \).
2. Compute \( D_{s_k} \cap (D_{s_{k+1}} - 1) \cap (D_{s_{k+2}} - 2) \cap \cdots \cap (D_{s_{k+t-1}} - (t-1)) \).

If \( D_{s_k} \cap (D_{s_{k+1}} - 1) \cap (D_{s_{k+2}} - 2) \cap \cdots \cap (D_{s_{k+t-1}} - (t-1)) \) has only one element, then the state content \( k \) of the register is determined. Otherwise, the known keystream \( s_k s_{k+1} \cdots s_{k+t-1} \) cannot determine \( k \).

Naturally, the computational feasibility of this attack depends on the computational feasibility of the two steps above. It depends on the size of \( N \) and the design of the underlying function \( f \).

3 Example 1: the Legendre cipher

3.1 The design of the Legendre cipher

Let \( N \) be a prime, and let

\[
f(i) = \left[ 1 - \left( \frac{i}{N} \right) \right]/2, \tag{8}
\]

where

\[
\left( \frac{x}{N} \right) = \begin{cases} 
1, & \text{if } x \text{ is a square modulo } N, \\
-1, & \text{otherwise}, 
\end{cases}
\]

which is the same as the Legendre symbol except that \( \left( \frac{0}{N} \right) := 1 \).

Then \( f \) is a function from \( \mathbb{Z}_N \) to \( \mathbb{Z}_2 \). The cipher in Figure 1 is called the Legendre cipher for the \( N \) and \( f \) defined above.

3.2 The security of Legendre with respect to the linear complexity attack

The linear complexity of the keystream sequence of the Legendre cipher was computed in [31], rediscovered in [15], and is stated below.

**Proposition 2.** Let \( s^\infty \) denote the keystream sequence of the Legendre cipher.

1. If \( N = 8t - 1 \) for some \( t \), then \( \text{LC}(s^\infty) = (N + 1)/2 \).
2. If \( N = 8t + 1 \) for some \( t \), then \( \text{LC}(s^\infty) = (N - 1)/2 \).
3. If \( N = 8t + 3 \) for some \( t \), then \( \text{LC}(s^\infty) = N \).
4. If \( N = 8t + 5 \) for some \( t \), then \( \text{LC}(s^\infty) = N - 1 \).

So the linear complexity of the keystream sequence is very large, compared with its least period \( N \). Proposition 2 describes the linear complexity of the whole
periodic keystream sequence of the Legendre cipher. In real applications, only a segment of the periodic sequence is used. For any finite sequence \( u = u_0u_1 \cdots u_{n-1} \), we use \( u^i \) to denote its subsequence \( u_0u_1 \cdots u_{i-1} \), where \( 1 \leq i \leq n \). The sequence

\[
\text{LC}(u^1), \text{LC}(u^2), \cdots, \text{LC}(u^n)
\]

is called the linear complexity profile of the sequence \( u \). Rueppel [29] proved that the mean of the linear complexities of all binary sequences of length \( n \) is about \( n/2 \). A binary sequence \( u^n = u_0u_1 \cdots u_{n-1} \) is said to have perfect linear complexity profile if

\[
\text{LC}(u^i) = \left\lceil \frac{i}{2} \right\rceil, \quad 1 \leq i \leq n,
\]

where \( \lceil x \rceil \) denotes the ceiling function. We say that a binary sequence \( u^n = u_0u_1 \cdots u_{n-1} \) has good linear complexity profile if \( \text{LC}(u^i) \) is very close to \( \lceil i/2 \rceil \) for every \( i \).

The linear complexity profile of the first period of the keystream sequence of the Legendre cipher is not known. But experimental data implies that the sequence has very good linear complexity profile. Table 1 describes the linear complexity profile for the \( N = 29 \) case.

<table>
<thead>
<tr>
<th>( i )</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \lceil i/2 \rceil )</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>6</td>
<td>7</td>
<td>7</td>
</tr>
</tbody>
</table>

Since the linear complexity of the keystream sequence and its segments are about half of their length, the linear complexity attack on the Legendre cipher does not work.

### 3.3 The security of Legendre with respect to linear complexity stability attacks

If we choose the period \( N \) of the counter to have primitive root 2, by Proposition 1 we have the following.

**Proposition 3.** Let the prime \( N \) have primitive root 2, and let \( s^\infty \) denote the keystream sequence of Legendre. Then we have

\[
\text{SC}_\ell(s^\infty) \geq N - 1, \quad \text{if } \ell < \frac{N - 1}{2}.
\]

Proposition 3 means that the linear complexity stability of the whole periodic keystream sequence is the best possible. However, the stability of linear complexity
profile is more important, as in practice only a small segment of the whole period is used for encryption. While it may be hard to obtain theoretical results on the profile of the sphere complexity $SC_\ell$, experiments indicate that it is very good for small $\ell$. Table 2 documents the profile of the sphere complexity $SC_1(s^m)$ for the $N = 29$ case. Experimental data indicates that the linear complexity stability of the periodic keystream sequence and its segments are very good not only for primes $N$ having primitive root 2, but also for primes $N$ in general.

Table 2: The profile of the sphere complexity $SC_1(s^m)$ of the keystream of Legendre.

<table>
<thead>
<tr>
<th>$i$</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>$SC_1(s^i)$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>$[i/2]$</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>6</td>
<td>7</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>$i$</td>
<td>16</td>
<td>17</td>
<td>18</td>
<td>19</td>
<td>20</td>
<td>21</td>
<td>22</td>
<td>23</td>
<td>24</td>
<td>25</td>
<td>26</td>
<td>27</td>
<td>28</td>
<td>29</td>
</tr>
<tr>
<td>$SC_1(s^i)$</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>12</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td>$[i/2]$</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>10</td>
<td>10</td>
<td>11</td>
<td>11</td>
<td>12</td>
<td>12</td>
<td>13</td>
<td>13</td>
<td>14</td>
<td>14</td>
<td>15</td>
</tr>
</tbody>
</table>

According to Proposition 3 and the experimental data, any attack based on the linear complexity stability should not work.

### 3.4 The security of Legendre with respect to linear attacks

To evaluate the security of the Legendre cipher with respect to linear approximation attacks, we need to find out all the affine functions from $\mathbb{Z}_N$ to $\mathbb{Z}_2$. The following proposition is proved in [4].

**Proposition 4.** Let $N$ be an odd prime. The only affine functions from $\mathbb{Z}_N$ to $\mathbb{Z}_2$ are the two constant functions

$$A(x) = b,$$

where $b \in \mathbb{Z}_2$.

Hence, the Hamming distance between the underlying function of (8) and any affine function from $\mathbb{Z}_N$ to $\mathbb{Z}_2$ is $(N - 1)/2$ or $(N + 1)/2$. So any linear attack on the Legendre cipher should not work.

### 3.5 The security of Legendre with respect to the key approximation attack

As for the autocorrelation property of the keystream sequence of the Legendre cipher, the following is well known.

**Proposition 5.** For the keystream sequence $s^\infty$ of the Legendre cipher, the autocorrelation function $AC(\tau)$ is given by

$$AC(\tau) = -3/N \text{ or } 1/N, \text{ for each } 0 < \tau < N$$
if \( N \equiv 1 \pmod{4} \), and by

\[
AC(\tau) = -\frac{1}{N}, \quad \text{for each } 0 < \tau < N
\]

if \( N \equiv 3 \pmod{4} \).

The following proposition follows from (2) and Proposition 5.

**Proposition 6.** Let \( f(x) \) be the function of (8).

1. If \( N \equiv 1 \pmod{4} \), then

\[
p_f = \frac{1}{2} + \frac{3}{2N}
\]

2. If \( N \equiv 3 \pmod{4} \), then

\[
p_f = \frac{1}{2} + \frac{1}{2N}
\]

So if a key \( k' \neq k \) is used to approximate the original secret key \( k \), the rate of correct decryption is one of the values in the set

\[
\left\{ \frac{1}{2} \pm \frac{1}{2N}, \frac{1}{2} \pm \frac{3}{2N} \right\}.
\]

This rate of correct decryption is computed, under the assumption that the whole periodic segment of the keystream sequence was used for decryption.

It is hard to obtain good estimation on the autocorrelation profile of the keystream sequence of the Legendre cipher. But experimental data indicates that \( \max_{1 \leq \tau \leq N-1} |AC^{(L)}(\tau)| \) approaches to \( 1/N \) or \( 3/N \) when \( L \) approaches to \( N \), where the partial autocorrelation function \( AC^{(L)}(\tau) \) is defined in (3). Table 3 documents the autocorrelation profile for the case \( N = 29 \).

**Table 3:** The autocorrelation profile of the keystream sequence of Legendre.

<table>
<thead>
<tr>
<th>( \tau )</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \max_{1 \leq \tau \leq N-1}</td>
<td>AC^{(L)}(\tau)</td>
<td>)</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.75</td>
<td>0.78</td>
</tr>
<tr>
<td>( \max_{1 \leq \tau \leq N-1}</td>
<td>AC^{(L)}(\tau)</td>
<td>)</td>
<td>0.63</td>
<td>0.50</td>
<td>0.38</td>
<td>0.43</td>
<td>0.46</td>
<td>0.50</td>
<td>0.41</td>
<td>0.33</td>
</tr>
<tr>
<td>( \max_{1 \leq \tau \leq N-1}</td>
<td>AC^{(L)}(\tau)</td>
<td>)</td>
<td>0.33</td>
<td>0.27</td>
<td>0.30</td>
<td>0.25</td>
<td>0.28</td>
<td>0.23</td>
<td>0.19</td>
<td>0.14</td>
</tr>
</tbody>
</table>

As made clear at the end of Section 2.2, the keystream sequence of the binary additive counter stream cipher should have high merit factor. It was proved in [21] that the Legendre sequence has merit factor 6. This is the only known family of binary sequences with the highest merit factor.

The theoretical results and the experimental data above show that the key approximation attack should not work.
3.6 The security of Legendre with respect to information stability attacks

In this section, we present information stability attacks on the Legendre cipher. Let \( u_0u_1\cdots u_{t-1} \) be a piece of the keystream, where

\[
u_j := f((i + j) \mod N)\]

for \( j = 0, 1, \cdots, t - 1 \). Let \( I(i; u_0u_1\cdots u_{t-1}) \) denote the amount of information in bits about \( i \) provided by the known piece of keystream \( u_0u_1\cdots u_{t-1} \). Let \( d_{u_0u_1\cdots u_{t-1}} \) be defined as in (7). Then

\[
I(i; u_0u_1\cdots u_{t-1}) = \log_2 N - \log_2 d_{u_0u_1\cdots u_{t-1}}.
\]

**Proposition 7.** ([6, p.207]) If \( N \equiv 3 \pmod{4} \), then

\[
d_{u_0u_1}(0, 1) = \begin{cases} 
\frac{N - 3}{4}, & \text{for } (u_0, u_1) = (1, 1), \\
\frac{N + 1}{4}, & \text{for } (u_0, u_1) \neq (1, 1).
\end{cases}
\]

If \( N \equiv 1 \pmod{4} \), then

\[
d_{u_0u_1}(0, 1) \in \left\{ \frac{N - 5}{4}, \frac{N - 1}{4}, \frac{N + 3}{4} \right\}.
\]

This proposition means that the information \( I(i; u_0u_1\cdots u_{t-1}) \) is optimally stable when \( t = 2 \).

**Proposition 8.** ([11]) Let the symbols be as before. For any integer \( t \) with \( \lceil \log_2 N \rceil \geq t \geq 3 \), we have

\[-A \leq d_{u_0u_1\cdots u_{t-1}}(0, 1, \cdots, t - 1) - \frac{N}{2t} \leq A,
\]

where

\[
A = \frac{\sqrt{N}(2^{t-1}(t - 3) + 2) + 2^{t-1}(t + 1) - 1}{2^t}.
\]

Note that \( A \) is essentially \( \sqrt{N} \frac{t - 3 + t + 1}{2} \). The bounds on \( d_{u_0u_1\cdots u_{t-1}}(0, 1, \cdots, t - 1) \) described in Proposition 8 are tight for small \( t \), and may not be tight for large \( t \). Nevertheless, they mean that the information \( I(i; u_0u_1\cdots u_{t-1}) \) is more or less stable.

The bounds in Proposition 8 mean that the piece of keystream \( u_0u_1\cdots u_{t-1} \) provides enough information about \( i \) if and only if \( t \) is approximately \( \log_2 N \). The question now is how to extract the information and find out \( i \).

The problem of determining \( i \) given \( u_0u_1\cdots u_{t-1} \) is equivalent to solving the following set of equations

\[
\left( \frac{x + i}{N} \right) = (-1)^w, \quad i = 0, 1, \cdots, t,
\]

(9)
where $t$ is approximately $\log_2 N$, and which is called the Computational Legendre Problem.

Hence, we have reached the following conclusion.

**Proposition 9.** The security of the Legendre cipher with respect to known plain-text attacks is equivalent to the computational complexity of the Computational Legendre Problem.

**Open Problem 1.** If $t$ is approximately $\log_2 N$, is there any polynomial-time algorithm for the Computational Legendre Problem?

If one could prove that the Computational Legendre Problem has complexity $\Omega(tN)$, the Legendre cipher will then be computationally secure, provided that $N$ has over 128 bits.

### 3.7 Performance of the Legendre cipher

Let $N = 2^{127} - 1$, which is a prime. Then the secret key of the Legendre cipher has 127 bits. We tested the performance of Legendre using the Magma software. The build-in function `Legendre` in Magma is used. An IBM Think Pad (Z60m) with Magma is used for the test. The encryption/decryption rate is 444 bits per second. Hence its performance is low. However, it is practical for off-line applications.

### 4 Example 2: The two-prime cipher

#### 4.1 The design of the two-prime cipher

Let $p$ and $q$ be two distinct primes such that $\gcd(p - 1, q - 1) = 2$, and let $N = pq$. Define for each $0 \leq i \leq N - 1$,

$$f(i) = \begin{cases} 0, & i \in \{0, q, 2q, \cdots, (p - 1)q\}, \\ 1, & i \in \{p, 2p, \cdots, (q - 1)p\}, \\ \left(1 - \left(\frac{i}{p}\right)\left(\frac{i}{q}\right)\right)/2, & \text{otherwise}, \end{cases}$$

(10)

where $\left(\frac{\cdot}{p}\right)$ denotes the modified Legendre symbol described before. Then $f$ is a function from $\mathbb{Z}_N$ to $\mathbb{Z}_2$.

The cipher in Figure 1 is then called the *two-prime cipher* for the $N$ and $f$ defined above, and the *twin-prime cipher* if $p$ and $q$ are twin primes. Note that the keystream sequence of the two-prime cipher is different from the Jacobi sequence, which is defined by

$$\text{Jacobi}(i) = \left(1 - \left(\frac{i}{p}\right)\left(\frac{i}{q}\right)\right)/2, \ 0 \leq i < pq.$$  

(11)
4.2 The security of the two-prime cipher with respect to the linear complexity attack

Proposition 10. \([\text{[10]}]\) Let \(L\) denote the linear complexity of the keystream sequence of the two-prime cipher.

1. If \(p \equiv 1 \pmod{8}\) and \(q \equiv 3 \pmod{8}\) or \(p \equiv -3 \pmod{8}\) and \(q \equiv -1 \pmod{8}\), then \(L = pq - 1\).
2. If \(p \equiv -1 \pmod{8}\) and \(q \equiv 3 \pmod{8}\) or \(p \equiv 3 \pmod{8}\) and \(q \equiv -1 \pmod{8}\), then \(L = (p-1)q\).
3. If \(p \equiv -1 \pmod{8}\) and \(q \equiv -3 \pmod{8}\) or \(p \equiv 3 \pmod{8}\) and \(q \equiv 1 \pmod{8}\), then \(L = pq - p - q + 1\).
4. If \(p \equiv 1 \pmod{8}\) and \(q \equiv -1 \pmod{8}\) or \(p \equiv -3 \pmod{8}\) and \(q \equiv 3 \pmod{8}\), then \(L = \frac{pq + p + q - 3}{2}\).
5. If \(p \equiv -1 \pmod{8}\) and \(q \equiv 1 \pmod{8}\) or \(p \equiv 3 \pmod{8}\) and \(q \equiv -3 \pmod{8}\), then \(L = \frac{(p-1)(q-1)}{2}\).
6. If \(p \equiv -1 \pmod{8}\) and \(q \equiv -1 \pmod{8}\) or \(p \equiv 3 \pmod{8}\) and \(q \equiv 3 \pmod{8}\), then \(L = \frac{(p-1)(q+1)}{2}\).

If \(|p - q|\) is small compared with \(\max\{p, q\}\), the linear complexity of the keystream sequence is very large, compared with its period \(N = pq\).

The linear complexity profile of the first period of the keystream sequence of the two-prime cipher is not known. But experimental data suggests that the sequence has very good linear complexity profile when \(|p - q|\) is very small. Table 4 describes the linear complexity profile for the case \(p = 5\) and \(q = 7\).

<table>
<thead>
<tr>
<th>(i)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\text{LC}(s_i))</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>(\lceil i/2 \rceil)</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(i)</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
<th>21</th>
<th>22</th>
<th>23</th>
<th>24</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\text{LC}(s_i))</td>
<td>6</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>(\lceil i/2 \rceil)</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>10</td>
<td>10</td>
<td>11</td>
<td>11</td>
<td>12</td>
<td>12</td>
</tr>
</tbody>
</table>

Since the linear complexity of the keystream sequence and its segments are about half of their length, the linear complexity attack does not work.
4.3 The security of two-prime cipher with respect to linear complexity stability attacks

By Proposition 1 we have the following.

**Proposition 11.** Let $s^\infty$ denote the keystream sequence of the two-prime cipher. Then we have

$$SC_\ell(s^\infty) \geq \min\{\text{ord}_p(2), \text{ord}_q(2)\}, \text{ if } \ell < \min\left\{\frac{pq + p - q + 1}{2}, \frac{pq - p + q - 1}{2}\right\}.$$  

By controlling $\min\{\text{ord}_p(2), \text{ord}_q(2)\}$, we can control the linear complexity stability of the whole periodic keystream sequence of the two-prime cipher. As made clear before, we are more interested in the profile of the sphere complexity $SC_\ell$. Experiments indicate that it is very good for small $\ell$. Table 5 documents the profile of the sphere complexity $SC_1(s^m)$ for the case that $p = 5$ and $q = 7$.

Table 5: The profile of the sphere complexity $SC_1(s^m)$ of the keystream of the two-prime cipher.

<table>
<thead>
<tr>
<th>$i$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>$[i/2]$</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>$SC_1(s^i)$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>$[i/2]$</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>10</td>
<td>10</td>
<td>11</td>
<td>11</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>$SC_1(s^i)$</td>
<td>5</td>
<td>6</td>
<td>6</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>$[i/2]$</td>
<td>13</td>
<td>14</td>
<td>15</td>
<td>16</td>
<td>17</td>
<td>18</td>
<td>19</td>
<td>20</td>
<td>21</td>
<td>22</td>
<td>23</td>
<td>24</td>
</tr>
<tr>
<td>$SC_1(s^i)$</td>
<td>11</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>15</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>$[i/2]$</td>
<td>13</td>
<td>13</td>
<td>14</td>
<td>14</td>
<td>15</td>
<td>15</td>
<td>16</td>
<td>16</td>
<td>17</td>
<td>17</td>
<td>18</td>
<td>18</td>
</tr>
</tbody>
</table>

According to Proposition 11 and the experimental data, any attack on the linear complexity stability should not work.

4.4 The security of the two-prime cipher with respect to linear attacks

To evaluate the security of the two-prime cipher with respect to linear approximation attacks, we need to find out all the affine functions from $\mathbb{Z}_{pq}$ to $\mathbb{Z}_2$.

By Proposition 4, the Hamming distance between the underlying function of (8) and any affine function from $\mathbb{Z}_N$ to $\mathbb{Z}_2$ is $(N - 1)/2$ or $(N + 1)/2$. So any linear attack on the two-prime cipher should not work.

4.5 The security of the two-prime cipher with respect to the key approximation attack

Let $N = pq$ as before. A common primitive root of $p$ and $q$ is a primitive root of both $p$ and $q$. By the Chinese Remainder Theorem a common primitive root $g$
exists since \( p \) and \( q \) are primes. Let \( u \) and \( v \) be a pair of integers satisfying the simultaneous congruences

\[
\begin{align*}
u &\equiv g \pmod{p}, \quad v \equiv 1 \pmod{p}, \\
u &\equiv 1 \pmod{q}, \quad v \equiv g \pmod{q}.
\end{align*}
\]

The existence and uniqueness of such \( u \) and \( v \) are guaranteed by the Chinese Remainder Theorem. Whiteman’s generalized cyclotomic classes of order two are defined by

\[
D_i = \{ g^s u^i : s = 0, 1, \ldots, e - 1 \}, \quad i = 0, 1,
\]

where \( e = (p - 1)(q - 1)/2 \). It is known that \( D_0 \) and \( D_1 \) form a partition of \( \mathbb{Z}_{pq}^* \).

**Proposition 12.** ([12]) Let \( s^\infty \) denote the keystream sequence of the two-prime cipher.

1. If \((p - 1)(q - 1)/4\) is even, then

\[
AC_s(w) = \begin{cases} 
\frac{q - p - 3}{pq}, & \text{if } w \in \{p, 2p, \ldots, (q-1)p\}, \\
\frac{p + 1 - q}{pq}, & \text{if } w \in \{q, 2q, \ldots, (p-1)q\}, \\
-1, & \text{if } w \in \mathbb{Z}_N.
\end{cases}
\]

2. If \((p - 1)(q - 1)/4\) is odd, then

\[
AC_s(w) = \begin{cases} 
\frac{q - p - 3}{pq}, & \text{if } w \in \{p, 2p, \ldots, (q-1)p\}, \\
\frac{p + 1 - q}{pq}, & \text{if } w \in \{q, 2q, \ldots, (p-1)q\}, \\
-3, & \text{if } w \in D_0, \\
\frac{1}{pq}, & \text{if } w \in D_1.
\end{cases}
\]

The following proposition follows from (2) and Proposition 12.

**Proposition 13.** Let \( f(x) \) be the function of (10).

1. If \((p - 1)(q - 1)/4\) is even, then

\[
p_f = \max \left\{ \frac{1}{2} \pm \frac{q - p - 3}{2pq}, \frac{1}{2} \pm \frac{p + 1 - q}{2pq}, \frac{1}{2} + \frac{1}{2pq} \right\}.
\]

2. If \((p - 1)(q - 1)/4\) is odd, then

\[
p_f = \max \left\{ \frac{1}{2} \pm \frac{q - p - 3}{2pq}, \frac{1}{2} \pm \frac{p + 1 - q}{2pq}, \frac{1}{2} + \frac{3}{2pq} \right\}.
\]
By Propositions 12 and 13, the autocorrelation values of the keystream sequence of the two-prime cipher and the nonlinearity $p_f$ of the function $f$ of (10) are very good when $|p - q|$ is very small.

The best case is when $q - p = 2$, i.e., they are twin primes. In this case, $(p - 1)(q - 1)/4$ must be even, and $AC_s(w) = -1$ for all $w$ with $0 < w < N$. In this case, the keystream sequence has optimal periodic autocorrelation property, and the function $f$ of (10) has optimal nonlinearity $p_f = \frac{1}{2} + \frac{1}{2N}$.

Another interesting case is when $q - p = 4$. In this case $AC_s(w)$ is three-valued when $(p - 1)(q - 1)/4$ is even, and two-valued when $(p - 1)(q - 1)/4$ is odd.

It is hard to obtain good estimation on the autocorrelation profile of the keystream sequence of the two-prime cipher. But experimental data suggests that $\max_{1 \leq \tau \leq N - 1} |AC^{(L)}(\tau)|$ approaches to $1/N$ or $3/N$ when $L$ approaches to $N$ and $|p - q|$ is very small, where the partial autocorrelation function $AC^{(L)}(\tau)$ is defined in (3).

Table 6 documents the autocorrelation profile for the case $p = 5$ and $q = 7$.

| $|AC^{(L)}(\tau)|_{\max_{1 \leq \tau \leq N - 1}}$ | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   | 9   |
|-----------------------------------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| $i$                                           | 1.00| 1.00| 1.00| 1.00| 1.00| 0.71| 0.75| 0.78|
| $|AC^{(L)}(\tau)|_{\max_{1 \leq \tau \leq N - 1}}$ | 10  | 11  | 12  | 13  | 14  | 15  | 16  | 17  | 18  |
| $i$                                           | 0.60| 0.63| 0.66| 0.54| 0.43| 0.33| 0.38| 0.41| 0.33|
| $|AC^{(L)}(\tau)|_{\max_{1 \leq \tau \leq N - 1}}$ | 19  | 20  | 21  | 22  | 23  | 24  | 25  | 26  | 27  |
| $i$                                           | 0.37| 0.40| 0.43| 0.36| 0.39| 0.33| 0.28| 0.31| 0.33|
| $|AC^{(L)}(\tau)|_{\max_{1 \leq \tau \leq N - 1}}$ | 28  | 29  | 30  | 31  | 32  | 33  |     |     |     |
| $i$                                           | 0.36| 0.31| 0.27| 0.23| 0.19| 0.15|     |     |     |

As made clear at the end of Section 2.2, the keystream sequence of the binary additive counter stream cipher should have high merit factor. It was proved in [21] that the keystream sequence of the two-prime cipher has asymptotic merit factor 6, which is the largest one known.

The theoretical results and the experimental data above indicate that the key approximation attack should not work.

### 4.6 The security of the two-prime cipher with respect to information stability attacks

In this section, we consider information stability attacks on the two-prime cipher. Let $u_0u_1\cdots u_{t-1}$ be a piece of the keystream, where $u_j := f((i + j) \mod pq)$
for \( j = 0, 1, \cdots, t - 1 \). Let \( I(i; u_0u_1 \cdots u_{t-1}) \) denote the amount of information in bits about \( i \) provided by the known piece of keystream \( u_0u_1 \cdots u_{t-1} \). Let \( d_{u_0u_1 \cdots u_{t-1}} \) be defined as in (7). Then

\[
I(i; u_0u_1 \cdots u_{t-1}) = \log_2 N - \log_2 d_{u_0u_1 \cdots u_{t-1}}.
\]

When \( t = 2 \), the values \( d_{u_0u_1}(0, 1) \) are given directly by the autocorrelation values in Proposition 12, and are optimal when \( p \) and \( q \) are twin primes. When \( |p - q| \) is very small, \( d_{u_0u_1}(0, 1) \) are about \( N/4 \). When \( 3 \leq t < \log_2 N \), we do not have a tight lower bound on \( d_{u_0 \cdots u_{t-1}}(0, 1, \cdots, t - 1) \). But experimental results indicate that they are very close to \( N/2^t \) when \( |p - q| \) is very small.

Hence, a piece of keystream \( u_0u_1 \cdots u_{t-1} \) provides enough information about \( i \) if and only if \( t \) is approximately \( \log_2 N \). The question now is how to extract the information and find out \( i \). This is to solve the following set of equations

\[
f(x + i) = u_i, \quad i = 0, 1, \cdots, t, \tag{13}
\]

where \( t \) is approximately \( \log_2 N \) and \( f \) is the function defined in (10), and which is referred to as the Computational Two-Prime Problem. Note that \( f \) has the optimal nonlinearity \( p_f \) when \( p \) and \( q \) are twin primes.

Hence, we have reached the following conclusion.

**Proposition 14.** The security of the two-prime cipher with respect to known plaintext attacks is equivalent to the computational complexity of the Computational Two-Prime Problem, where \( t \) is approximately \( \log_2 N \).

**Open Problem 2.** If \( t \) is approximately \( \log_2 N \), is there any polynomial-time algorithm for the Computational Two-Prime Problem?

If one could prove that the Computational Two-Prime Problem has complexity \( \Omega(tN) \), then the two-prime cipher will be secure, provided that \( N \) has over 128 bits.

### 4.7 Other comments on the two-prime cipher

The two-prime cipher has about the same performance as the Legendre cipher. As shown before, the best examples of the two-prime cipher are the cases \( q - p = 2 \) or \( q - p = 4 \). There are indeed twin primes with about 120 bits. They can be used to construct twin-prime ciphers.

### 5 Conclusions and concluding remarks

#### 5.1 The selection of the building blocks of the binary additive counter stream ciphers

Let \( (A, +) \) be an Abelian group of order \( v \). Let \( C \) be a \( k \)-subset of \( A \). The set \( C \) is a \((v, k, \lambda)\) difference set (DS) in \( A \) if \( d_C(w) = \lambda \) for every nonzero element \( w \) of
$A$, where $d_C(w)$ is the difference function defined by

$$d_C(w) = |(w + C) \cap C|, \ w \in \mathbb{Z}_N. \quad (14)$$

The complement $\overline{C}$ of a $(v, k, \lambda)$ difference set $C$ in $A$ is defined by $A \setminus C$ and is a $(v, v-k, v-2k+\lambda)$ difference set. The reader is referred to [24, 25] for details of difference sets.

Let $(A, +)$ be an Abelian group of order $v$. A $k$-subset $C$ of $A$ is a $(v, k, \lambda, t)$ almost difference set (ADS) in $A$ if $d_C(w)$ takes on $\lambda$ altogether $t$ times and $\lambda + 1$ altogether $v - 1 - t$ times when $w$ ranges over all the nonzero elements of $A$ [1].

To design a secure additive counter stream cipher of Figure 1, we need to choose a positive integer $N$ and a function $f$ from $\mathbb{Z}_N$ to $\mathbb{Z}_2$ so that the design requirements presented in Section 2 are fulfilled. It is open whether other security requirements for the cipher of Figure 1 exist.

One basic security requirement is that the function $f$ should have optimal nonlinearity $p_f$. This is equivalent to the requirement that the keystream sequence should have optimal autocorrelation values, or that the support $\{x \in \mathbb{Z}_N : f(x) = 1\}$ of the function $f$ is a difference set or almost difference set of $\mathbb{Z}_N$ with special parameters. A number of such functions $f$ with optimal nonlinearity $p_f$ are known and documented in [3], and some of them may be used to construct binary additive counter stream ciphers of Figure 1.

However, another design requirement is that the keystream sequence should have large linear complexity and good linear complexity stability. Therefore, some of the functions with optimal nonlinearity $p_f$ in [3] cannot be employed in the binary additive counter stream cipher of Figure 1. The reader is invited to investigate this issue. The linear complexity of the keystream sequence may be known if some of the functions $f$ with optimal nonlinearity $p_f$ in [3] are employed. Details may be found in [27, 17, 32].

### 5.2 Conclusions

In this paper, we surveyed known results on binary additive counter stream ciphers that are scattered over many references, and derived a number of design criteria. In particular, we systematically analysed the Legendre cipher and the two-prime cipher and showed that their security with respect to known plaintext attacks is equivalent to the computational complexity of the Computational Legendre Problem and the Computational Two-Prime Problem. Experiments on the Legendre and the two-prime cipher were done when no theoretical result is available. At this moment, we are not able to claim the security of the Legendre or the two-prime cipher based on the properties surveyed in this paper, but could make the following conclusions:

1. The Legendre cipher and the twin-prime cipher are secure with respect to a number of attacks. They have more security properties, compared with other stream ciphers in the literature. This puts them into a unique position among all practical ciphers.
2. The Legendre cipher and the twin-prime cipher are secure with respect to known plaintext attacks if and only if the Computational Legendre Problem and the Computational Two-Prime Problem are NP-hard.

3. The performance of the Legendre cipher and the twin-prime cipher is low, compared with other stream ciphers in the literature. But they are practical for offline applications, where security requirement is high.

4. In some types of stream ciphers, there are trade-offs among certain security requirements. In this case, the best one can do is to make a compromise. However, no trade-off among security requirements of the binary additive counter stream ciphers of Figure 1 is known so far. The Legendre cipher and the twin-prime cipher may justify this. This is why they were called natural stream ciphers in [8]. According to the theoretical and experimental results documented in this paper, there is indeed no trade-off among the six design criteria listed before.

In view of the attractive security properties of the Legendre cipher and the twin-prime ciphers summarized before, we invite and urge the cryptographic and the mathematical communities to work out the computational complexity of the Computational Legendre Problem and the Computational Two-Prime Problem.

References


[29] R. A. Rueppel, Analysis and Design of Stream Ciphers, Springer-Verlag,
