Revealing Information with Partial Period Correlations (extended abstract)

Andrew Klapper Mark Goresky
Northeastern University, College of Computer Science, Boston, MA 02115, U.S.A.

1 Introduction

In several applications in modern communication systems periodic binary sequences are employed that must be difficult for an adversary to determine when a short subsequence is known, and must be easy to generate given a secret key. This is true both in stream cypher systems, in which the binary sequence is used as a pseudo-one-time-pad [11], and in secure spread spectrum systems, in which the sequence is used to spread a signal over a large range of frequencies [10]. While theoreticians have long argued that such security can only be achieved by sequences satisfying very general statistical test such as Yao's and Blum and Micali's next bit test [12,2], practitioners are often satisfied to find sequences that have large linear complexities, thus ensuring resistence to the Berlekamp-Massey algorithm [8]. Linear feedback shift registers are devices that can easily generate sequences with exponentially larger period than the size of their seeds [6], though with small linear complexity. Thus much effort has gone into finding ways of modifying linear feedback shift registers so that the sequences they generate have large linear complexities, typically by adding some nonlinearity.

Chan and Games [4] have suggested using a class of sequences called geometric sequences for these purposes. Geometric sequences are derived from m-sequences over a finite field $GF(q)$ by applying a (nonlinear) map from $GF(q)$ to $GF(2)$. Chan and Games showed that for q odd, geometric sequences have high linear complexities. For this reason, these sequences have been employed in commercial applications, typically with enormous periods. More recently Chan, Goresky, and Klapper [5] derived formulas for the periodic autocorrelation function of a geometric sequence and, in some cases, (with q even) for the periodic cross-correlation function of a pair of geometric sequences with the same period. Knowledge of these correlation function values is essential for applications involving spread spectrum systems. Furthermore, Brynielsson [3] derived a formula for the linear complexity of a geometric sequence when q is even, showing that such sequences can be constructed with moderately large linear complexities.

The purpose of this paper is to show that a certain statistical attack, partial period autocorrelation attack, can be used to obtain critical information about geometric sequences from knowledge of a relatively small subsequence when q is odd. Specifically, for a sequence of period $q^n - 1$, with $q \geq 5$, q odd, and $n \geq 10$, if q^8 bits of the sequence are known, then q can be determined with high probability (n must be at least 11 if $q = 3$). While this does not as yet allow us to determine the sequence, it does render their security questionable.

Suppose a small subsequence is known. We show that if a small shift of this subsequence is taken, then the correlation between the original subsequence and its shift is close to q^2 when q is odd. More specifically, we show that the expected correlation (letting the starting point of the shift vary and keeping the size of the subsequence fixed) is approximately q^2, and that under the hypotheses on q and n given above, the variance is sufficiently small that the correlation is close to its expectation with high probability (this is a consequence of Chebyshev's inequality [1]). One way to view our resluts is that we have introduced a new statistical test that a sequence must satisfy in order to be secure - the variance of the partial period autocorrelation must be high for

small subsequences.

Most of the difficulty in acquiring these results lies in calculating the variance. A critical part of this calculation involves understanding how uniformly each orbit of the action of $GL_2(GF(q))$ on $GF(q^n)$ by fractional linear transformations is distributed in $GF(q^n)$. In some cases we make estimates that we expect can be improved, thus decreasing the number of bits required for success. It is unlikely, however that we can get by with fewer than q^6 bits.

A second application of our results on the partial period autocorrelation of geometric sequences is to spread spectrum CDMA systems [10]. In such systems a signal is distributed over a large number of channels using a pseudo-random binary sequence to determine parity. The signal is then recovered by computing an (unshifted) autocorrelation. To avoid phase shift interference it is desirable that the shifted partial period autocorrelation values be small. Partial period correlations are generally quite difficult to compute, so in practice families of sequences are found that have low full period autocorrelations. These families are then searched for sequences with low partial period autocorrelation. Such searches are slow and may be unsuccessful. The expected partial period autocorrelation always differs from the full period autocorrelation by a factor depending only on the size of the subsequence used, so is low if the full period autocorrelation is low. This is only useful, however, if the variance is low enough to ensure the partial period autocorrelation is low with high probability. Our results show that this is the case for geometric sequences, making them strong candidates for use in spread spectrum systems.

The proofs of all lemmas are omitted and will later appear in a full version of the paper.

2 Geometric Sequences and Correlations

In this section we recall the definition of the geometric sequences and some of their basic properties, and the definition of full and partial period correlation functions of periodic sequences. Throughout this paper, q will denote a fixed power of a fixed prime p, and $GF(q)$ will denote the Galois field of q elements. See Lidl and Niederreiter's or McEliece's book [7,9] for background on finite fields.

Definition 1 (Chan and Games [4]) *Let n be a positive integer and let α be a primitive element of $GF(q^n)$. The sequence $\mathbf{U}_i = Tr_q^{q^n}(\alpha^i)$ is a q-ary m-sequence. Let f be a (possibly nonlinear) function from $GF(q)$ to $GF(2)$. The binary sequence $\mathbf{S}$ whose ith term is*

$$\mathbf{S}_i = f(Tr_q^{q^n}(\alpha^i)).$$

is called a geometric sequence.

Geometric sequences with q odd have been used in commercial applications where easily generated sequences with large linear complexities are needed. Recall that the m-sequence $\mathbf{U}$ can be generated by a linear feedback shift register over $GF(q)$ of length n, so the geometric sequence $\mathbf{S}$ is easy to generate if the feedforward function f is easy to compute. Such a geometric sequence is a (q^n-1)-periodic binary sequence. The periodic autocorrelation function $\mathcal{A}_{\mathbf{S}}(\tau)$ of $\mathbf{S}$ is the function whose value at τ is the correlation of the τ-shift of $\mathbf{S}$ with itself.

$$\mathcal{A}_{\mathbf{S}}(\tau) = \sum_{i=1}^{q^n-1} (-1)^{\mathbf{S}_{i+\tau}} (-1)^{\mathbf{S}_i}$$

The partial autocorrelation of a sequence is defined by limiting the range of values in the sum to a fixed window. It is parametrized by the start position k and length D of the window, as well as the shift τ. Precisely

$$\mathcal{A}_{\mathbf{S}}(\tau, k, D) = \sum_{i=k}^{D+k-1} (-1)^{\mathbf{S}_{i+\tau}} (-1)^{\mathbf{S}_i}$$

We next recall a result due to Chan, Goresky, and Klapper [5] regarding the autocorrelation of a geometric sequence. We use the notation $F(x) = (-1)^{f(x)}$, $I(f) = \sum_{x \in GF(q)} F(x)$, the imbalance[1] of f, and $\Delta_a(f) = \sum_{x \in GF(q)} F(ax)F(x)$, the short autocorrelation function[2] of f.

Theorem 1 *Then the values for the periodic autocorrelation of a geometric sequence* S *are:*

1. $\mathcal{A}_{\mathbf{S}}(\tau) = q^{n-2} I(f)^2 - 1$, *if* $\alpha^\tau \notin GF(q)$.
2. $\mathcal{A}_{\mathbf{S}}(\tau) = q^{n-1} \Delta_{\alpha^\tau}(f) - 1$, *if* $\alpha^\tau \in GF(q)$.

Note that, letting $\nu = (q^n - 1)/(q - 1)$, the second case of the theorem, $\alpha^\tau \in GF(q)$, occurs exactly when ν divides τ.

If p equals two, f can be chosen to be balanced ($I(f) = 0$) and so that $\Delta_a(f) = 0$ for $a \neq 1$, $\Delta_1(f) = q$. This implies the shifted autocorrelation of S is -1, the minimum possible. Unfortunately, in this case the linear complexity is much smaller. If p is odd, then the imbalance is at least 1, so the autocorrelation is always large. Such sequences are used in commercial systems due to their enormous linear complexities. We will show, however, that their poor autocorrelations render such systems vulnerable. The idea is that if we know the autocorrelation, then we know q^n. Of course we will never be able to compute the full autocorrelation, since we will never see the full period of the sequence. However, if the partial period correlation is sufficiently well behaved for small enough windows, then we can get similar information by seeing only a small part of the sequence – seeing $D + \tau$ bits of the sequence allows us to compute a partial period autocorrelation with window D and shift τ. Such a short subsequence may be discovered, for example, by a known plaintext attack on a pseudo-one-time-pad system. Unfortunately, the partial period autocorrelation may vary considerably as the start position varies. We will show that the expected partial period autocorrelation (averaged over the starting position of the window) is closely related to the full period autocorrelation. We will also show that for certain window sizes the variance of the partial period autocorrelation (with fixed shift τ and window size D) is low enough that an adversary has high probability of discovering q. This is a consequence of Chebyshev's inequality [1] which says that a bound on the variance implies a bound on the probability that a particular partial period correlation is far from the expected partial period correlation.

We begin by showing that the expected partial period autocorrelation of any sequence can be determined from its full period autocorrelation. We denote the expectation of a random variable X by $\langle X \rangle$. All expectations are taken for fixed window size D and shift τ, assuming a uniform distribution on all start positions k.

[1]The imbalance of f is equal to the number of x for which f is zero minus the number of x for which f is one.

[2]If γ is a primitive element of $GF(q)$, and $a = \gamma^\sigma$, then $\Delta_a(f) - 1$ is the autocorrelation with shift σ of the sequence whose ith term is $f(\gamma^i)$.

Theorem 2 *Let* S *be a periodic binary sequences with period* N. *Then the expectation of the partial period autocorrelation of* S *is given by*

$$\langle \mathcal{A}_{\mathbf{S}}(\tau, k, D) \rangle = \frac{D}{N} \mathcal{A}_{\mathbf{S}}(\tau).$$

Proof: Straightforward.

Suppose S is a geometric sequence of period $q^n - 1$ with feedforward function $f : GF(q) \to GF(2)$, and that S is as balanced as possible, i.e., $I(f) = 1$. Then for shifts $0 < \tau < \nu$, the expected partial period autocorrelation of S is $\langle \mathcal{A}_{\mathbf{S}}(\tau, k, D) \rangle = D(q^{n-2} - 1)/(q^n - 1)$.

We next consider the variance of the partial period autocorrelation. Recall that the variance of a random variable X is defined to be $\langle (X - \langle X \rangle)^2 \rangle = \langle X^2 \rangle - \langle X \rangle^2$, so we must determine the second moment $\langle \mathcal{A}_{\mathbf{S}}(\tau, k, D)^2 \rangle$ of the partial period autocorrelation. We can reduce this determination to the determination of the cardinalities of certain fourfold intersections of hyperplanes (identifying $GF(q^n)$ with n-dimensional affine space over $GF(q)$) asstated in the following lemma. If $s \in GF(q)$, and $A \in GF(q^n)$, then we denote by H_A^s the hyperplane $\{x : Tr_q^{q^n}(Ax) = s\}$.

Lemma 1 *If* S *is a geometric sequence, then*

$$\langle \mathcal{A}_{\mathbf{S}}(\tau, k, D)^2 \rangle = \frac{1}{q^n - 1} \sum_{i,j=0}^{D-1} \Big(\sum_{s,t,u,v \in GF(q)} N_{i,j,\tau}(s,t,u,v) F(s)F(t)F(u)F(v) - 1 \Big)$$

where

$$N_{i,j,\tau}(s,t,u,v) = |H^s_{\alpha^{i+\tau}} \cap H^t_{\alpha^i} \cap H^u_{\alpha^{j+\tau}} \cap H^v_{\alpha^j}|.$$

Thus we must determine the values of $N_{i,j,\tau}(s,t,u,v)$.

3 Intersections of Hyperplanes

There are five possible values for $N_{i,j,\tau}(s,t,u,v)$: q^{n-4}, q^{n-3}, q^{n-2}, q^{n-1}, and 0. The following lemma will be useful in determining which case occurs. If $\{A_m\}$ are elements of a vector space over $GF(q)$, by abuse of notation we denote by $\dim\{A_m\}$ the dimension over $GF(q)$ of the span of $\{A_m\}$.

Lemma 2 *Let* $A_1, A_2, A_3, A_4 \in GF(q^n)$ *and* $s_1, s_2, s_3, s_4 \in GF(q)$. *Then* $|\bigcap_{m=1}^4 H_{A_m}^{s_m}| = q^{n-r}$ *if and only if* $\dim\{A_m\} = r$, *and whenever* $\{A_m\}$ *satisfy a linear equation* $\sum_{m=1}^4 a_m A_m = 0$, $\{a_m\} \in GF(q)$, $\{s_m\}$ *satisfy the same relation, i.e.,* $\sum_{m=1}^4 a_m s_m = 0$. *Otherwise the intersection is empty. For a given set* $\{A_m\}$ *whose span has dimension* r, *there are* q^r *sets* $\{s_m\}$ *for which* $\bigcap_{m=1}^4 H_{A_m}^{s_m} \neq \emptyset$.

Let $A = \alpha^i$, $B = \alpha^j$, and $C = \alpha^\tau$ in the sum we derived for the second moment of the partial period correlation. We need to determine $\dim\{A, AC, B, BC\}$, which can be 1, 2, 3, or 4. We consider 4 to be the generic case and consider when the other four cases occur.

3.1 Dimension 1:

This occurs when every element is a $GF(q)$ multiple of every other element. Thus $\alpha^r, \alpha^{j-i} \in GF(q)$. If the window size D is in the range $0 < D < \nu$, then this case cannot occur.

3.2 Dimension 2:

This occurs if AC, A, BC, B satisfy two linearly independent equations, say

$$\begin{aligned} aAC + bA + cBC + dB &= 0 \\ eAC + fA + gBC + hB &= 0, \end{aligned} \tag{1}$$

where $a, b, c, d, e, f, g, h \in GF(q)$ and (a, b, c, d) and (e, f, g, h) are independent vectors.

If $C \in GF(q)$, then the span of $\{AC, A, BC, B\}$ equals the span of $\{A, B\}$. Dim$\{A, AC, B, BC\}$ is two if A/B is not in $GF(q)$, and is one otherwise.

If C is not in $GF(q)$, then we can use each of these equations to write B/A as the result of applying to C a fractional linear transformation with coefficients in $GF(q)$:

$$\frac{B}{A} = -\frac{aC + b}{cC + d} = -\frac{eC + f}{gC + h}.$$

We can use the second equation to find a quadratic equation over $GF(q)$ satisfied by C. This equation is degenerate if and only if $B/A \in GF(q)$. Thus if B/A is not in $GF(q)$, then C is in $GF(q^2) - GF(q)$. Conversely, suppose C is in $GF(q^2) - GF(q)$. If dim$\{A, AC, B, BC\}$ is less than four, then $\{AC, A, BC, B\}$ satisfy a linear equation. The quadratic equation satisfied by C can then be used to produce a second, independent linear equation. Hence dim$\{A, AC, B, BC\}$ is two (it must be at least two since $C \notin GF(q)$).

3.3 Dimension Three:

As a consequence of the preceding subsection, dim$\{A, AC, B, BC\}$ can only be three if C is not a root of a quadratic equation over $GF(q)$. Moreover, we must have a single equation

$$aAC + bA + cBC + dB = 0,$$

or, equivalently,

$$B = \frac{aC + b}{cC + d} A.$$

As before, $(aC + b)/(cC + d)$ is in $GF(q)$ (and hence dim$\{A, AC, B, BC\}$ is two) if and only if $ad - bc = 0$.

There is an action of the general linear group over $GF(q)$ of rank two, $G = GL_2(GF(q))$, on $GF(q^n)$ which we shall make use of. Recall that this group is the multiplicative group of two by two matrices over $GF(q)$ with nonzero determinate. The group acts by fractional linear transformations. That is, the matrix

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

acts on the element $C \in GF(q)$ by

$$C \mapsto \frac{aC+b}{cC+d} = M(C).$$

It is straightforward to check that if $M, N \in GL_2(GF(q))$, then $(MN)(C) = M(N(C))$. Recall that when a group G acts on a set W, the G-orbit of an element $x \in W$ is the set $\{M(x) : M \in G\} = \text{orbit}(x)$. Our analysis of the conditions under which the various dimensions occur is summarized in the following table:

	Dim = 1	Dim = 2	Dim = 3
$C \in GF(q)$	$B/A \in GF(q)$	$B/A \notin GF(q)$	-
$C \in GF(q^2) - GF(q)$	-	$B/A \in GF(q) \cup \text{orbit}(C)$	-
$C \in GF(q^n) - GF(q^2)$	-	$B/A \in GF(q)$	$B/A \in \text{orbit}(C)$

In all other cases $\dim\{A, AC, B, BC\}$ is four.

4 Variance of Partial Period Correlations

Returning to our computation of the variance of the partial period correlation of geometric sequences, we need to know, for given $0 \le i, j < D$, whether $\alpha^{i-j} \in \text{orbit}(\alpha^\tau)$. We break down our analysis depending upon whether α^τ is in $GF(q)$, $GF(q^2) - GF(q)$, or $GF(q^n) - GF(q^2)$.

4.1 $\alpha^\tau \in GF(q)$

In this case $N_{i,j,\tau}(0,0,0,0) = q^{n-2}$ if $\alpha^{i-j} \notin GF(q)$, $N_{i,j,\tau}(0,0,0,0) = q^{n-1}$ if $\alpha^{i-j} \in GF(q)$. We have $\alpha^{i-j} \in GF(q)$ if and only if ν divides $i-j$. Thus for a given i, $0 \le i < D$, the number of j, $0 \le j < D$, such that the i, j term contributes to the sum is the number of j in this range such that ν divides $i-j$. This number is

$$\left\lfloor \frac{D-i-1}{\nu} \right\rfloor + \left\lfloor \frac{i}{\nu} \right\rfloor + 1 \le \frac{D-1}{\nu} + 1.$$

We can bound the second moment as follows

$$\begin{aligned}
\langle \mathcal{A}_S(\tau,k,D)^2 \rangle &= \frac{1}{q^n-1} \sum_{i,j=0}^{D-1} \Big(\sum_{t,v \in GF(q)} N_{i,j,\tau}(\alpha^\tau t, t, \alpha^\tau v, v) F(\alpha^\tau t) F(t) F(\alpha^\tau v) F(v) - 1 \Big) \\
&= \frac{1}{q^n-1} \Big(\sum_{i,j=0}^{D-1} \Big(\sum_{t,v \in GF(q)} q^{n-2} F(\alpha^\tau t) F(t) F(\alpha^\tau v) F(v) - 1 \Big) \\
&\quad + \sum_{\substack{0 \le i,j < D \\ \nu \mid (i-j)}} \Big(q^n - \sum_{t,v \in GF(q)} q^{n-2} F(\alpha^\tau t) F(t) F(\alpha^\tau v) F(v) \Big) \Big) \\
&\le \frac{D^2}{q^n-1} (q^{n-2} \Delta_{\alpha^\tau}(f)^2 - 1) + \frac{D}{q^n-1} \Big(\frac{D-1}{\nu} + 1 \Big) (q^n - q^{n-2} \Delta_{\alpha^\tau}(f)^2).
\end{aligned}$$

The expectation in this case is $\frac{D}{q^{n-1}}(q^{n-1}\Delta_{\alpha^\tau}(f)-1)$. Therefore the variance is

$$\begin{aligned} V(\mathcal{A}_S(\tau,k,D)) &= \langle \mathcal{A}_S(\tau,k,D)^2\rangle - \langle \mathcal{A}_S(\tau,k,D)\rangle^2 \\ &\le \frac{D^2}{q^n-1}(q^{n-2}\Delta_{\alpha^\tau}(f)^2-1)+\frac{D}{q^n-1}(\frac{D-1}{\nu}+1)(q^n-q^{n-2}\Delta_{\alpha^\tau}(f)^2) \\ &\quad -\frac{D^2}{(q^n-1)^2}(q^{n-1}\Delta_{\alpha^\tau}(f)-1)^2 \\ &= \frac{q^{n-2}D}{q^n-1}((\frac{D-1}{\nu}+1)(q^2-\Delta_{\alpha^\tau}(f)^2)-\frac{D}{q^n-1}(q-\Delta_{\alpha^\tau}(f))^2) \\ &\le \frac{q^nD}{q^n-1}(\frac{D-1}{\nu}+1). \end{aligned}$$

In particular, if $D \le \nu$, then the variance is bounded above by $2q^nD/(q^n-1)$.

4.2 $\alpha^\tau \in GF(q^2) - GF(q)$

If $x \in GF(q^2)$, and $M \in G$, then $M(x) \in GF(q^2)$. If, moreover, $x \notin GF(q)$, then x is a generator for $GF(q^2)$ over $GF(q)$, that is, every element of $GF(q^2)$ can be written in the form $(ax+b)/(cx+d)$ for some $a,b,c,d \in GF(q)$. It follows that $N_{i,j,\tau}(0,0,0,0)$ is q^{n-2} if $\alpha^{i-j} \in GF(q^2)$, i.e., if $\nu_2 = (q^n-1)/(q^2-1)$ divides $i-j$, and is q^{n-4} otherwise. As above, we can bound the second moment:

$$\langle \mathcal{A}_S(\tau,k,D)^2\rangle \le \frac{D^2}{q^n-1}(q^{n-4}I(f)^4-1)+\frac{D}{q^n-1}(\frac{D-1}{\nu_2}+1)(q^n-q^{n-4}I(f)^4).$$

The expectation in this case is $\frac{D}{q^{n-1}}(q^{n-2}I(f)^2-1)$. Therefore the variance is bounded by:

$$V(\mathcal{A}_S(\tau,k,D)) \le \frac{q^nD}{q^n-1}(\frac{D-1}{\nu_2}+1).$$

In particular, if $D \le \nu$, then the variance is bounded above by $(q+1)q^nD/(q^n-1)$.

4.3 $\alpha^\tau \in GF(q^n) - GF(q^2)$

Unfortunately, in this case the situation more complex. In general, the G-orbit is not uniformly distributed in $GF(q^n)$, so for a fixed i, the number of j in a window with $\dim\{\alpha^{i+\tau},\alpha^i,\alpha^{j+\tau},\alpha^j\}=3$ is not proportional to the size of the window. We settle here for a cruder estimate, based on the structure of the group G. We first determine the size of the G-orbit of α^τ.

Lemma 3 *If $x \in GF(q^n)-GF(q)$, then the G-orbit of x has cardinality q^3-q.*

We will next decompose the elements of G into the composition of certain simple types of matrices with scalar multiplication. Since scalar multiplication by elements of $GF(q)$ moves elements large distances, this will allow us to bound the number of elements in an orbit that are in a given small window.

For matrices M and N, we write $M \sim N$ if M and N define the same transformation (i.e., the matrices differ by a scalar multiple). Let

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

be an element of G, so $\delta = ad - bc \neq 0$. First suppose $a \neq 0$. Then

$$M \sim \begin{pmatrix} a^2/\delta & ab/\delta \\ ac/\delta & ad/\delta \end{pmatrix} = \begin{pmatrix} a^2/\delta & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & b/a \\ ac/\delta & bc/\delta + 1 \end{pmatrix}.$$

Letting $S_x = \{(x+b)/(cx+bc+1)\}$, we have shown that $M(x)$ is a scalar multiple of an element of S_x.

On the other hand, suppose $a = 0$. Then $b \neq 0$ and $c \neq 0$, so

$$M \sim \begin{pmatrix} 0 & b/c \\ 1 & d/c \end{pmatrix} = \begin{pmatrix} b/c & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & d/c \end{pmatrix}.$$

Let $T_x = \{1/(x+d)\}$. Then in this case $M(x)$ is a scalar multiple of some element of T_x. We have shown that an arbitrary element of the orbit of x is a scalar multiple of some element of $S_x \cup T_x$.

Consider a window of size ν. If y is any element of $GF(q^n)$, then there is a unique $a \in GF(q)$ such that ay is in the given window. Therefore, for each element y of $S_x \cup T_x$, there is a unique scalar multiple of y, i.e., a unique element of the orbit of x, in the given window. Thus we have

Proposition 1 *If $x \in GF(q^n) - GF(q^2)$, then the intersection of the orbit of x with a window of size at most ν has cardinality at most $|S_x \cup T_x| = q^2 + q$.*

It follows that if $D \leq \nu$, then the number of i, j such that $0 \leq i, j < D$ and $N_{i,j,\tau}(s,t,u,v) = q^{n-3}$ or 0 is at most $D(q^2+q)$ (or D^2 if $D < (q^2+q)$). The number of s,t,u,v for which we get q^{n-3} is q^3. Moreover, $N_{i,i,\tau}(s,s,u,u) = q^{n-2}$, and $N_{i,i,\tau}(s,t,u,v) = 0$ if $s \neq t$ or $u \neq v$. In all other cases $N_{i,j,\tau}(s,t,u,v) = q^{n-4}$. Thus if $D \leq \nu$, as above we can bound the second moment:

$$\langle \mathcal{A}_S(\tau,k,D)^2 \rangle \leq \frac{D^2}{q^n-1}(q^{n-4}I(f)^4 - 1) + \frac{D(q^2+q+1)}{q^n-1}(q^n - q^{n-4}I(f)^4).$$

The expectation in this case is $\frac{D}{q^n-1}(q^{n-2}I(f)^2 - 1)$. Therefore the variance is bounded by:

$$V(\mathcal{A}_S(\tau,k,D)) \leq \frac{q^n(q^2+q+1)D}{q^n-1}.$$

Theorem 3 *For any τ, if $D \leq \nu$, then the variance of the partial period autocorrelation of a geometric sequence with shift τ and window D is bounded above by $(q^2+q+1)q^nD/(q^n-1)$.*

5 Cracking Cryptosystems

In this section we describe how the results of the previous sections can be used to obtain critical information on pseudorandom sequences used in stream cyphers. Specifically, using a known plaintext attack, we are able determine q with high probability. We will make use of Chebyshev's inequality [1].

Proposition 2 (Chebyshev's Inequality) *If X is a random variable with expectation ϵ and variance σ^2, then for any k,*

$$Prob\{|X - \epsilon| > k\} < \sigma^2/k^2.$$

First suppose q is odd. Suppose f is a balanced feedforward function (that is, $I(f) = 1$). Geometric sequences are generally taken with $I(f) = 1$ so they are statistically random. For small enough windows ($D \leq \nu$) the expected partial period autocorrelation is $D(q^{n-2} - 1)/(q^n - 1)$, or approximately D/q^2. Thus we can hope to determine q. As we determine bits of the sequence, we can take a small shift and compute a partial period autocorrelation. We must have the value of the partial period autocorrelation close enough to its expectation to unambiguously recover q.

A difficulty is that we must recover q without knowing n. For each q there will be an interval $I_q = (a_q, a_{q-1})$ such that if the computed value of the partial period correlation is in I_q, then we assume that q is being used by the generator of the sequence. In order to have a high probability of success, we must chose I_q so that there is a large k such that for each n, the interval of radius k around the expected partial period autocorrelation lies entirely within I_q. To simplify things, we assume $n \geq 4$. As it turns out, the statistics prevent us from successfully determining q if $n \leq 3$, so this restriction is of no importance.

For a given q, the sequence $d_{q,n} = D(q^{n-2} - 1)/(q^n - 1)$ is an increasing sequence with limit D/q^2. Let c_q be the center of the smallest interval containing all these points ($n \geq 4$). That is

$$c_q = \frac{1}{2}\left(d_{q,n} + \frac{D}{q^2}\right).$$

We then let a_q be the point midway between c_{q+1} and c_q, that is $a_q = (c_{q+1} + c_q)/2$, and $I_q = (a_q, a_{q-1})$ (note that $c_{q+1} < c_q$).

Lemma 4 *Let*

$$k = \frac{D(q-1)}{q^2(q+1)^2}.$$

If $n \geq 4$, then $d_{q,n} \in (a_q + k, a_{q-1} - k)$. Hence $(d_{q,n} - k, d_{q,n} + k) \subseteq I_q$.

The algrithm for determining q is then: compute the partial period correlation. If the result is in I_r, then assume that $q = r$. Applying Chebyshev's inequality, we have

Theorem 4 *If $n \geq 3$, and S is a geometric sequence of period $q^n - 1$, then a partial autocorrelation attack with a window $D \leq \nu = (q^n - 1)/(q - 1)$ will fail to determine q with probability at most*

$$\frac{q^{n+4}(q^2 + q + 1)(q + 1)^4}{D(q^n - 1)(q - 1)^2}.$$

If n is small, this probability will be larger than one, so Chebyshev's inequality does not tell us whether the atack has a positive probability of determining q. However, we have

Corollary 1 *If n is large enough that*

$$\frac{q^{n+4}(q^2+q+1)(q+1)^4}{(q^n-1)(q-1)^2} < \frac{q^n-1}{q-1}$$

then using a window of size

$$\frac{q^{n+4}(q^2+q+1)(q+1)^4}{(q^n-1)(q-1)^2} \sim q^8$$

gives a positive probability of determining q.

The conditions of the corollary are satisfied if $n \geq 10$, and $q \geq 5$; or $n \geq 11$, and $q \geq 3$; or $n \geq 14$, and $q \geq 2$.

6 Conclusions

We have shown that geometric sequences based on m-sequences over a finite field $GF(q)$ of odd characteristic exhibit vulnerability to a partial period correlation attack when enough bits are known. Specifically, for sequences of period $q^n - 1$, if between about q^8 and $(q^n-1)/(q-1)$ bits are known, then q can be determined with high probability. If $n \leq 9$, then this condition is vacuous. However for q small the sequence would have small period. We conclude that q must be large. Even for large q, our results cut down on the number of usable bits of a geometric sequence. For example, if we chose $q = 17$ and $n = 16$, so that the sequence has period greater than 10^{18}, then we must see approximately 7 times 10^9 bits. If we received bits at 9600 baud, then it takes about 8 days to have positive probability of determining q.

There are two points where we have used estimates that might be improved. First, in evaluating the summation expression for the second moment, we make a worst case estimate of the smaller sum that each term contributes a plus one. It is possible that this can be improved by recognizing this sum as a higher order autocorrelation of a shorter sequence (of period $q-1$).

Second, based on computer experiments, we belive that smaller estimates can be made of the number of points in the intersection of an orbit of the action of $GL_2(GF(q))$ with a window of size $D < (q^n-1)/(q-1)$. In particular, we conjecture that there is a small constant c (e.g. two or three) such that in a window of size less than q^{n-3}, there are at most cq elements of an orbit. Such a result would imply that only cq^6 bits are required to have positive probability of determining q.

Of course our results only allow the determination of q, not the entire sequence, or even q^n. It remains to be seen whether this information is enough to compromise sytems using geometric sequences, but it should make users wary.

Finally, the moral of this paper is that it is dangerous to rely on linear complexity as a measure of cryptographic security. There are many other statistical tests a sequence must pass – in this case, we have shown that the variance of the partial period autocorrelation must be high.

References

[1] H. BAUER, *Probability Theory and Elements of Measure Theory*, Holt, Rinehart and Winston, New York, 1972.

[2] M. BLUM AND S. MICALI, How to generate cryptographically strong sequences of pseudo-random bits, *SIAM Jo. Comput.* **13** (1984), 850-864.

[3] L. BRYNIELSSON, On the Linear Complexity of Combined Shift Registers, em in "Proceedings of Eurocrypt 1984," pp. 156-160.

[4] A. H. CHAN AND R. GAMES, On the linear span of binary sequences from finite geometries, q odd, *in* "Proceedings of Crypto 1986," pp. 405-417, Santa Barbara.

[5] A. H. CHAN, M. GORESKY, AND A. KLAPPER, Cross-correlations of geometric sequences and GMW sequences, *in* Proceedings of Marshall Hall Memorial Conference, Burlington, VT, 1990 and Northeastern University Technical Report NU-CCS-90-12.

[6] S. GOLOMB, *Shift Register Sequences*, Aegean Park Press, Laguna Hills, CA, 1982.

[7] R. LIDL AND H. NIEDERREITER, *Finite Fields, Encyclopedia of Mathematics vol. 20*, Cambridge University Press, Cambridge, 1983.

[8] J.L. MASSEY, Shift register sequences and BCH decoding, *IEEE Trans. Info. Thy.*, **IT-15** (1969), pp. 122-127.

[9] R. MCELIECE, *Finite Fields for Computer Scientists and Engineers*, Kluwer Academic Publishers, Boston, 1987.

[10] M. SIMON, J. OMURA, R. SCHOLTZ, AND B. LEVITT, *Spread-Spectrum Communications, Vol. 1*, Computer Science Press, 1985.

[11] D. WELSH, *Codes and Cryptography*, Clarendon Press, Oxford, 1988.

[12] A. YAO, Theory and applications of trapdoor functions, *in* "Proc. 23rd IEEE Symp. on Foundations of Comp. Sci.," 1982.