autocorrelation for all $n > 1$ and this conjecture has been verified for $3 \le n \le 23$. In this note, we present an interesting formula for the Hadamard transform of the sequence A or, equivalently, for the function $f(x)$. The Hadamard transform of $f(x)$ is defined by

$$\hat{f}(\lambda) = \sum_{x \in \mathrm{GF}(2^n)} (-1)^{f(x)+\mathrm{Tr}(\lambda x)}, \qquad \lambda \in \mathrm{GF}(2^n). \tag{1}$$

Conjecture: For $n \ge 5$, the Hadamard transform $\hat{f}(\lambda)$, $\lambda \in \mathrm{GF}(2^n)$, of $f(x)$ is three-valued with the values $0, \pm 2^{m+1}$. Specifically,

$$\hat{f}(\lambda) = 0 \Leftrightarrow \mathrm{Tr}(\lambda^t) = 0$$
$$\hat{f}(\lambda) = \pm 2^{m+1} \Leftrightarrow \mathrm{Tr}(\lambda^t) = 1$$

where $t = \frac{2}{3}(2^{2l}-1)+1$ for $m = 2l$ or $2l+1$ where $l \ge 1$.

The result has been verified by computer on the range of $5 \le n \le 23$.

II. Discussion

From $3t = 2(2^{2l}-1)+3 = 2^{2l+1}+1$, which is relatively prime to 2^n-1, we have $\gcd(t, 2^n-1) = 1$. Therefore, the Hadamard transform of $f(x)$ is determined by the m-sequence $\{\mathrm{Tr}(\alpha^{it})\}_{i \ge 0}$. This is similar to the case of Gold sequences [2], as follows. Let $g(x) = \mathrm{Tr}(x^{1+2^k})$ where $(k, n) = 1$. The Hadamard transform of $g(x)$ is

$$\hat{g}(\lambda) = \sum_{x \in \mathrm{GF}(2^n)} (-1)^{g(x)+\mathrm{Tr}(\lambda x)}, \qquad \lambda \in \mathrm{GF}(2^n).$$

From [2], $\hat{g}(\lambda)$ is three-valued with values $0, \pm 2^{m+1}$, and

$$\hat{g}(\lambda) = 0 \Leftrightarrow \mathrm{Tr}(\lambda) = 0$$
$$\hat{g}(\lambda) = \pm 2^{m+1} \Leftrightarrow \mathrm{Tr}(\lambda) = 1.$$

For example, the Hadamard transform of $g(x)$ is determined by the m-sequence $\{\mathrm{Tr}(\alpha^i)\}_{i \ge 0}$. Thus we see that the three-term function $f(x)$ behaves like the monomial (one-term) function $\mathrm{Tr}(x^{1+2^k})$ with respect to the Hadamard transform.

References

[1] J. S. No, S. W. Golomb, G. Gong, H. K. Lee, and P. Gaal, "New binary pseudo random sequences of period 2^n-1 with ideal autocorrelation," *IEEE Trans. Inform. Theory*, vol. 44, pp. 814–817, Mar. 1998.

[2] R. Gold, "Maximal recursive sequences with 3-valued recursive cross-correlation functions," *IEEE Trans. Inform. Theory*, vol. IT-14, pp. 154–156, Jan. 1968.

Linear Span of Extended Sequences and Cascaded GMW Sequences

Habong Chung and Jong-Seon No

***Abstract*—In this correspondence, the linear span of extended sequences of period $2^{em}-1$ which are constructed from the ideally correlated sequences of period of 2^m-1 is derived. It is also shown that the linear span of cascaded GMW sequences can be derived in the same context. As examples, the linear span of extended Legendre sequences and cascaded GMW sequences with triple trace are computed.**

***Index Terms*—Cascaded sequences, extended sequences, Legendre sequences, linear span, m-sequences, pseudorandom sequences.**

I. Introduction

Balanced binary sequences with ideal autocorrelation, the so-called pseudorandom sequences, have many applications in spread-spectrum communication systems [5]. Among known sequences having period 2^n-1, there are i) m-sequences for all $n = 1, 2, \cdots$; ii) residue sequences, such as Legendre sequences for a Mersenne prime 2^n-1, Hall's sextic residue sequences for $2^n-1 = 31, 127,$ and 131071, etc; iii) extended sequences for a composite integer n, such as Gordon–Mills–Welch (GMW) sequences, extended Legendre sequences, etc; iv) multitrace sequences such as cascaded GMW sequences; v) sequences of miscellaneous types, such as "3-trace" sequences, "5-trace" sequences, etc. [1]–[3], [5], [6], [8], [9], [11], and [12].

The linear span of a pesudorandom sequence $s(t)$ of period 2^n-1 can be defined as the smallest integer L in the following linear recursion:

$$s(t) = \sum_{i=1}^{L} a_i \cdot s(t-i) \tag{1}$$

where a_i's are in $\{0, 1\}$ with $a_L = 1$. Alternately, the linear span of a pseudorandom sequence of period 2^n-1 can also be defined as the minimum number of terms in its expression by the sum of elements in $\mathrm{GF}(2^n)$. For example, the GMW sequence of period 63 given by $g(t) = \mathrm{tr}_1^3(\{\mathrm{tr}_3^6(\alpha^t)\}^3)$ can be expanded as [14]

$$\begin{aligned} g(t) &= \mathrm{tr}_1^6(\alpha^{3t}) + \mathrm{tr}_1^6(\alpha^{5t}) \\ &= \alpha^{2^0\cdot 3t} + \alpha^{2^1\cdot 3t} + \alpha^{2^2\cdot 3t} + \alpha^{2^3\cdot 3t} + \alpha^{2^4\cdot 3t} + \alpha^{2^5\cdot 3t} \\ &\quad + \alpha^{2^0\cdot 5t} + \alpha^{2^1\cdot 5t} + \alpha^{2^2\cdot 5t} + \alpha^{2^3\cdot 5t} + \alpha^{2^4\cdot 5t} + \alpha^{2^5\cdot 5t} \end{aligned}$$

and its linear span is 12. The linear span may be regarded as a measure of complexity of pseudorandom sequences when used in spread-spectrum systems or cipher systems. The well-known m-sequences, Kasami sequences, and Gold sequences have extremely short linear span, which makes the use of those sequence vulnerable in such environments. Recently, No presented a method generalizing GMW

Manuscript received July 24, 1998; revised March 1, 1999. This work was supported in part by the Institute of Information Technology Assessment under Grant 96190-CT-I2 and under Grant 96003-RT-I1 from the Korean Ministry of Information and Communications.

H. Chung is with the School of Electronic and Electrical engineering, Hong-Ik University, Seoul 121-791, Korea (e-mail: habchung@wow.hongik.ac.kr).

J.-S. No is with the Department of Electronic Engineering, Konkuk University, Seoul 143-701, Korea (e-mail: jsno@kkucc.konkuk.ac.kr).

Communicated by I. F. Blake, Associate Editor for Coding Theory.

Publisher Item Identifier S 0018-9448(99)06040-X.

sequences, and No *et al.* presented a new construction of binary sequences of a longer period with ideal autocorrelation property, which are extended from a given binary sequence with ideal autocorrelation property.

The technique of using dyadic weight of exponents to determine linear complexity was first studied by Brynielsson [13]. Linear span of cascaded GMW sequences was derived by Klapper, Chan, and Goresky [6] for a special case, and Gong [7] extended their result to the q-ary case.

In this correspondence, we derived the linear span of cascaded GMW sequences and extended sequences of period $2^{em}-1$ which are constructed from the ideally correlated sequences of period 2^m-1. In Section II, we reviewed the linear span of GMW sequences. In Section III, the linear span of extended sequences is derived under certain conditions, and the linear span of extended Legendre sequences is computed as an example. The linear span of cascaded GMW sequences is also derived in Section IV.

II. Preliminaries

The linear span of a pseudorandom sequence of period 2^n-1 can be defined as the minimum number of terms when the sequence is expressed as the sum of elements in GF(2^n). Thus as a natural approach for obtaining the linear span, one can think of expressing the sequences as the sum of decimated m-sequences, in which case the linear span becomes the sum of linear spans of the decimated m-sequences. The linear span of a GMW sequence was derived in [14] by this method. We will briefly review the results, since they are applicable to derivations of linear span in the next section and after.

Definition 1: Let $n=em$, $e>1$. Let r be an integer relatively prime to 2^m-1. The sequence $g(t)$ of period 2^n-1 given by

$$g(t)=\mathrm{tr}_1^m(\{\mathrm{tr}_m^n(\alpha^t)\}^r) \tag{2}$$

is called a GMW sequence.

Proposition 2: Let $n=em$, $e>1$. Let r be an integer less than and relatively prime to 2^m-1 and assume that r can be expressed as

$$r=2^{l_1}+2^{l_2}+2^{l_3}+\cdots+2^{l_w} \tag{3}$$

where $0=l_1<l_2<l_3<\cdots<l_w<m-1$. Then, the GMW sequence $g(t)$ in (2) can be expressed as the sum of decimated m-sequences as

$$\begin{aligned} g(t)&=\mathrm{tr}_1^m(\{\mathrm{tr}_m^n(\alpha^t)\}^r)\\ &=\sum_{a\in I(n,\,m;\,r)}\mathrm{tr}_1^n(\alpha^{at}) \end{aligned} \tag{4}$$

where the index set $I(n,\,m;\,r)$ is given by

$$I(n,\,m;\,r)=\left\{1+\sum_{j=2}^{w}2^{l_j+k_jm}\mid 0\le k_2,\,k_3,\cdots,k_w\le\frac{n}{m}-1\right\}. \tag{5}$$

Proposition 3: All the elements in $I(n,\,m;\,r)$ in (5) belong to distinct cosets of GF(2^n) and their coset sizes are n.

Proposition 4: The linear span of GMW sequence $g(t)$ given in (2) is $n\left(\frac{n}{m}\right)^{w-1}$.

In fact, (4) and (5) in Proposition 2, i.e., expressing the sequences with "double trace" into the sum of decimated m-sequences, hold for not only r which is relatively prime to 2^m-1 but any integer r less than 2^m-1. The only difference is that in the latter case, we cannot guarantee that all the elements in $I(n,\,m;\,r)$ are in the distinct cosets of GF(2^n). For example, if $n=12$, $m=6$, and $r=21$, then $I(12,\,6;\,21)=\{21,\,273,\,1029,\,1281\}$. But, $21, 1029$, and 1281 are in the same coset of GF(2^{12}). Thus $\mathrm{tr}_1^6(\{\mathrm{tr}_6^{12}(\alpha^t)\}^{21})$ is reduced to the sum of two traces. Here, we slightly generalize Proposition 2 as follows.

Theorem 5: Let $n=em$, $e>1$. Let r be an integer less than 2^m-1, such that 2^ir, $0\le i\le m-1$ are all distinct modulo 2^m-1, and assume that r can be expressed as

$$r=2^{l_1}+2^{l_2}+2^{l_3}+\cdots+2^{l_w}$$

where $0\le l_1<l_2<l_3<\cdots<l_w<m-1$. Then

$$\mathrm{tr}_1^m(\{\mathrm{tr}_m^n(\alpha^t)\}^r)=\sum_{a\in I(n,\,m;\,r)}\mathrm{tr}_1^n(\alpha^{at}) \tag{6}$$

where the index set $I(n,\,m;\,r)$ is given by

$$I(n,\,m;\,r)=\left\{2^{l_1}+\sum_{j=2}^{w}2^{l_j+k_jm}\mid 0\le k_2,\,k_3,\cdots,k_w\le\frac{n}{m}-1\right\} \tag{7}$$

and all the elements in $I(n,\,m;\,r)$ belong to the distinct cosets modulo (2^n-1) of size n.

Proof: If some two elements in $I(n,\,m;\,r)$ are in the same cosets of GF(2^n), this implies that

$$\begin{aligned} &2^{l_1}+2^{l_2+k_2m}+\cdots+2^{l_w+k_wm}\\ &\quad\equiv 2^u(2^{l_1}+2^{l_2+k_2'm}+\cdots+2^{l_w+k_w'm})\bmod(2^n-1). \end{aligned} \tag{8}$$

By reducing (8) modulo 2^m-1, we have

$$r\equiv 2^{u'}r\bmod(2^m-1)$$

where $u'\equiv u(\mathrm{mod}\,m)$, which contradicts the assumption, unless u', the residue of u modulo m, is zero. But this implies that u is a multiple of m, which further implies that $u=n$ and $k_j'=k_j$ for all j. This is from the fact that for (8) to be true

$$\begin{aligned} 2^{l_1}&\equiv 2^u\cdot 2^{l_1}\bmod(2^n-1)\\ 2^{l_2+k_2m}&\equiv 2^u\cdot 2^{l_2+k_2'm}\bmod(2^n-1),\cdots, \end{aligned}$$

etc. Therefore, all the elements in $I(n,\,m;\,r)$ are in the distinct cosets of GF(2^n). By replacing k_i''s by k_i's in the right-hand side of (8), their coset size can be proven to be n in the same way. □

III. Linear Span of Extended Sequences

Recently, No *et al.* [9] presented a closed-form expression for binary sequences of longer period with ideal autocorrelation property constructed from a given binary sequence with ideal autocorrelation. Here, we cite the theorem without proof.

Theorem 6 (No et al. [9]): Let m and n be positive integers such that m divides n. Let β be a primitive element of GF(2^n) and set $\alpha=\beta^T$ where $T=(2^n-1)/(2^m-1)$. Assume that for an index set J, the sequence $\{b(t_1),\,t_1=0,\,1,\cdots,2^m-2\}$ of period 2^m-1 given by

$$b(t_1)=\sum_{d\in J}\mathrm{tr}_1^m(\alpha^{dt_1}) \tag{9}$$

has the ideal autocorrelation property. Then for an integer r, $1\le r\le 2^m-2$, relatively prime to 2^m-1, the sequence $\{c(t),\,t=0,\,1,\cdots,2^n-2\}$ of period 2^n-1 defined by

$$c(t)=\sum_{d\in J}\mathrm{tr}_1^m([\mathrm{tr}_m^n(\beta^t)]^{dr}) \tag{10}$$

also has the ideal autocorrelation property.

Similarly as in GMW sequences, this extended sequence $c(t)$ can be expressed as the sum of decimated m-sequences. Using the same notations as in the previous discussions, $c(t)$ can be rewritten as

$$c(t) = \sum_{d \in J} \sum_{a \in I(n,\, m; \langle dr \rangle_m)} \mathrm{tr}_1^n(\beta^{at}) \tag{11}$$

where the notation $\langle dr \rangle_m$ represents $dr \mod (2^m - 1)$. To obtain the linear span of $c(t)$ from (11), one should make sure that the index sets $I(n, m; \langle dr \rangle_m)$'s are disjoint as d runs over J. The following lemma establishes the condition under which the sets $I(n, m; \langle dr \rangle_m)$ are disjoint.

Lemma 7: Let $n = em$, $e > 1$. Let r and s be integers such that $2^i r$, $0 \le i < m - 1$ are all distinct modulo $2^m - 1$ and $2^i s$, $0 \le i \le m - 1$ are all distinct modulo $2^m - 1$. Assume that r and s can be expressed as

$$r = 2^{l_1} + 2^{l_2} + 2^{l_3} + \cdots + 2^{l_w}$$

$$\text{and } s = 2^{l'_1} + 2^{l'_2} + 2^{l'_3} + \cdots + 2^{l'_w}$$

where

$$0 \le l_1 < l_2 < l_3 < \cdots < l_w < m - 1$$

and

$$0 \le l'_1 < l'_2 < l'_3 < \cdots < l'_w < m - 1.$$

If r and s are not in the same cyclotomic coset modulo $(2^m - 1)$, then any element in $I(n, m; r)$ is not in the same coset modulo $(2^n - 1)$ which any element in $I(n, m; s)$ belongs to.

Proof: Assume that r and s are not in the same cyclotomic coset modulo $(2^m - 1)$ but some element in $I(n, m; r)$ and some element in $I(n, m; s)$ are in the same cyclotomic coset modulo $(2^n - 1)$. Then, for some k's and k''s, we have

$$2^{l_1} + 2^{l_2 + k_2 m} + \cdots + 2^{l_w + k_w m} \equiv 2^u (2^{l'_1} + 2^{l'_2 + k'_2 m} + \cdots + 2^{l'_w + k'_w m}) \mod (2^n - 1). \tag{12}$$

Since $2^{l_i + k_i m} \pm 2^{l_j} \pmod{2^m - 1}$, reducing modulo $2^m - 1$, (12) becomes

$$2^{l_1} + 2^{l_2} + \cdots + 2^{l_w} \equiv 2^{u'} (2^{l'_1} + 2^{l'_2} + \cdots + 2^{l'_w}) \mod (2^m - 1) \tag{13}$$

where $u' \equiv u \mod m$. But (13) implies that r and s are in the same cyclotomic coset modulo $(2^m - 1)$, which contradicts the assumption. □

From Theorem 5 and Lemma 7, we can derive the linear span of an extended sequence.

Theorem 8: Let m and n be positive integers such that m divides n. Let β be a primitive element of $\mathrm{GF}(2^n)$ and set $\alpha = \beta^T$ where $T = (2^n - 1)/(2^m - 1)$. Let the index set J be given by $J = \{d_1, d_2, \cdots, d_Q\}$ and assume the sequence $\{b(t_1), t_1 = 0, 1, \cdots, 2^m - 2\}$ given by

$$b(t_1) = \sum_{i=1}^{Q} \mathrm{tr}_1^m(\alpha^{d_i t_1}) \tag{14}$$

has the ideal autocorrelation property. If $2^j d_i$, $0 \le j \le m - 1$, are all distinct modulo $2^m - 1$ for each i, then the linear span L of the extended sequence $c(t)$ given by

$$c(t) = \sum_{i=1}^{Q} \sum_{a \in I(n,\, m;\, \langle d_i r \rangle_m)} \mathrm{tr}_1^n(\beta^{at}) \tag{15}$$

is given by

$$L = n \sum_{i=1}^{Q} \left(\frac{n}{m}\right)^{\mathrm{wt}\,(\langle d_i r \rangle_m) - 1} \tag{16}$$

where $\mathrm{wt}\,(\langle d_i r \rangle_m)$ is the number of ones in the binary representation of $\langle d_i r \rangle_m$.

Example 9: The Legendre sequence $b(t)$ of period $M = 2^m - 1$ is defined as follows:

$$b(t) = \sum_{i=0}^{\frac{M-1}{2m} - 1} \mathrm{tr}_1^m(\alpha^{u^{2i} t}) \tag{17}$$

where u is a primitive element in the set of integers modulo M, and α is a primitive element of $\mathrm{GF}(2^m)$. Legendre sequences of period M are known to exist when M is a Mersenne prime, and the linear span is $(M - 1)/2$. The extended Legendre sequence $c(t)$ of period $N = 2^n - 1$ extended from the sequences (17) is given by

$$c(t) = \sum_{i=0}^{\frac{M-1}{2m} - 1} \mathrm{tr}_1^m \{[\mathrm{tr}_m^n(\beta^t)]^{r \cdot u^{2i}}\} \tag{18}$$

where β is a primitive element of $\mathrm{GF}(2^n)$, n is some multiple of m, and r is an integer less than and relatively prime to $2^m - 1$. Now, let us apply our theorem. Note that the size of every coset in $\mathrm{GF}(2^m)$ is m, since $M = 2^m - 1$ is a Mersenne prime. From (16), we can conclude that the linear span L of the extended Legendre sequence given in (18) is

$$L = \sum_{i=0}^{\frac{M-1}{2m} - 1} n \cdot \left(\frac{n}{m}\right)^{w_i - 1} \tag{19}$$

where w_i is the Hamming weight of the binary representation of $u^{2i} r$ modulo $2^m - 1$, i.e., $\langle u^{2i} r \rangle_m$. Now, look at the specific example. When $m = 7$, by picking $u = 3$, the Legendre sequence $b(t)$ of period 127 can be expressed as

$$b(t) = \sum_{i=0}^{8} \mathrm{tr}_1^7(\beta^{3^{2i} t}) = \sum_{i=0}^{8} \mathrm{tr}_1^7(\beta^{9^i t}).$$

The sequence $c(t)$ of period $2^{14} - 1$, by using $r = 1$, can be expressed as

$$c(t) = \sum_{i=0}^{8} \sum_{a \in I(n,\, m;\, \langle 9^i \rangle_7)} \mathrm{tr}_1^{14}(\beta^{at}).$$

The weight w_i's are

$$\begin{aligned}
w_0 &= \mathrm{wt}\,(\langle 9^0 \rangle_7) = 1 \\
w_1 &= \mathrm{wt}\,(\langle 9^1 \rangle_7) = 2 \\
w_2 &= \mathrm{wt}\,(\langle 9^2 \rangle_7) = 3 \\
w_3 &= \mathrm{wt}\,(\langle 9^3 \rangle_7) = \mathrm{wt}\,(94) = 5 \\
w_4 &= \mathrm{wt}\,(\langle 9^4 \rangle_7) = \mathrm{wt}\,(84) = 3 \\
w_5 &= \mathrm{wt}\,(\langle 9^5 \rangle_7) = \mathrm{wt}\,(121) = 5 \\
w_6 &= \mathrm{wt}\,(\langle 9^6 \rangle_7) = \mathrm{wt}\,(73) = 3 \\
w_7 &= \mathrm{wt}\,(\langle 9^7 \rangle_7) = \mathrm{wt}\,(22) = 3 \\
w_8 &= \mathrm{wt}\,(\langle 9^8 \rangle_7) = \mathrm{wt}\,(71) = 4.
\end{aligned}$$

Therefore, the linear span L of $c(t)$ is

$$L = 14\left(\frac{14}{7}\right)^{1-1} + 14\left(\frac{14}{7}\right)^{2-1} + 4 \times 14\left(\frac{14}{7}\right)^{3-1} + 14\left(\frac{14}{7}\right)^{4-1} + 2 \times 14\left(\frac{14}{7}\right)^{5-1} = 826.$$

IV. Linear Span of Cascaded GMW Sequences

The generalization of GMW sequences is first done by No [14] and independently by Klapper *et al.* [6], who named it the cascaded GMW sequences, and the general expression of cascaded GMW sequence is as follows.

Definition 10: Let $n = e_1 m_1$, $m_i = e_{i+1} m_{i+1}$ for $1 \le i \le K-1$ and $e_i > 1$ for $1 \le i \le K$. Let r_j be an integer less than and relatively prime to $2^{m_j} - 1$, respectively, for $1 \le j \le K$. Then, the sequence $A(t)$ of period $2^n - 1$ defined by

$$A(t) = \mathrm{tr}_1^{m_K}(\{\mathrm{tr}_{m_K}^{m_{K-1}}(\{\mathrm{tr}_{m_{K-1}}^{m_{K-2}}(\{\cdots \cdots \{\mathrm{tr}_{m_2}^{m_1}(\{\mathrm{tr}_{m_1}^{n}(\alpha^t)\}^{r_1})\}^{r_2}\cdots\}^{r_{K-1}})\}^{r_K}) \quad (20)$$

is called a cascaded GMW sequence.

When $K = 1$, the sequence $A(t)$ in (20) is called a GMW sequence.

The linear span of cascaded GMW sequences was derived by Klapper, Chan, and Goresky [6] for the special case when r_i has $2^{m_{i+1}}$-adic weight two and four, and Gong [7] extended their result to the q-ary case with no weight restrictions.

We derive the linear span of cascaded GMW sequences for any values of r_i's as in the case of GMW sequences [14]. To obtain it, Theorem 5 should be generalized to the multitrace case. In the following lemma, we express a cascaded GMW sequence as the sum of decimated m-sequences.

Lemma 11: Let $n = e_1 m_1$, $m_i = e_{i+1} m_{i+1}$ for $1 \le i \le K-1$, and $e_i > 1$ for $1 \le i \le K$. Let r_j be an integer less than and relatively prime to $2^{m_j} - 1$, respectively, for $1 \le j \le K$. Let the sequence $A(t)$ of period $2^n - 1$ be defined by

$$A(t) = \mathrm{tr}_1^{m_K}(\{\mathrm{tr}_{m_K}^{m_{K-1}}(\{\mathrm{tr}_{m_{K-1}}^{m_{K-2}} (\{\cdots\{\mathrm{tr}_{m_2}^{m_1}(\{\mathrm{tr}_{m_1}^{n}(\alpha^t)\}^{r_1})\}^{r_2}\cdots\}^{r_{K-1}})\}^{r_K}).$$

Then, $A(t)$ can be expressed as

$$A(t) = \sum_{a \in I(n,\, m_1,\, m_2, \cdots, m_K;\, r_1, r_2, \cdots, r_K)} \mathrm{tr}_1^n(\alpha^{at}) \quad (21)$$

where the index set $I(n,\, m_1,\, m_2, \cdots, m_K;\, r_1,\, r_2, \cdots, r_K)$ is recursively given by

$$I(n,\, m_1, m_2, \cdots, m_K;\, r_1,\, r_2, \cdots, r_K) = \{i \in I(n,\, m_1;\, \langle a r_1 \rangle_{m_1}) \mid a \in I(m_1,\, m_2, \cdots, m_K;\, r_2, \cdots, r_K)\} \quad (22)$$

and

$$I(m_j, m_{j+1}, \cdots, m_K;\, r_{j+1}, \cdots, r_K) = \{i \in I(m_j,\, m_{j+1};\, \langle a r_{j+1} \rangle_{m_{j+1}}) \mid a \in I(m_{j+1}, \cdots, m_K;\, r_{j+2}, \cdots, r_K)\} \quad (23)$$

for $j = 1, 2, \cdots, K-2$. Note that in (22) and (23), $\langle a r_1 \rangle_{m_1}$ and $\langle a r_{j+1} \rangle_{m_{j+1}}$ are defined as $a r_1 \bmod (2^{m_1} - 1)$ and $a r_{j+1} \bmod (2^{m_{j+1}} - 1)$, respectively.

Proof: Set $y = \{\mathrm{tr}_{m_1}^n(\alpha^t)\}^{r_1}$. Then the sequence $A(t)$ can be rewritten as

$$A(t) = \mathrm{tr}_1^{m_K}(\{\mathrm{tr}_{m_K}^{m_{K-1}}(\{\mathrm{tr}_{m_{K-1}}^{m_{K-2}}(\{\cdots\{\mathrm{tr}_{m_2}^{m_1}(y)\}^{r_2}\cdots\}^{r_{K-1}})\}^{r_K}). \quad (24)$$

Thus from (21), $A(t)$ can be expressed as

$$\begin{aligned} A(t) &= \sum_{a \in I(m_1,\, m_2, \cdots, m_K;\, r_2, \cdots, r_K)} \mathrm{tr}_1^{m_1}(y^a) \\ &= \sum_{a \in I(m_1,\, m_2, \cdots, m_K;\, r_2, \cdots, r_K)} \mathrm{tr}_1^{m_1}(\{\mathrm{tr}_{m_1}^n(\alpha^t)\}^{a r_1}) \\ &= \sum_{a \in (m_1,\, m_2, \cdots, m_K;\, r_2, \cdots, r_K) b \in I(n,\, m_1;\, \langle a r_1 \rangle m_1)} \sum \mathrm{tr}_1^n(\alpha^{bt}). \end{aligned} \quad (25)$$

By comparing (21) and (25), we have (22) as a result. Similarly, by setting

$$z = \{\mathrm{tr}_{m_j}^{m_{j-1}}(\{\mathrm{tr}_{m_{j-1}}^{m_{j-2}}(\{\cdots\{\mathrm{tr}_{m_2}^{m_1}(\{\mathrm{tr}_{m_1}^n(\alpha^t)\}^{r_1})\}^{r_2}\cdots)\}^{r_{j-1}})\}^{r_j} \quad (26)$$

$A(t)$ can be expressed as

$$\begin{aligned} A(t) &= \mathrm{tr}_1^{m_K}(\{\mathrm{tr}_{m_K}^{m_{K-1}}(\cdots\{\mathrm{tr}_{m_{j+1}}^{m_j}(z)\}^{r_{j+1}}\cdots)\}^{r_K}) \\ &= \sum_{a \in I(m_j, \cdots, m_K;\, r_{j+1}, \cdots, r_K)} \mathrm{tr}_1^{m_j}(z^a). \end{aligned} \quad (27)$$

At the same time, by setting $\omega = \{\mathrm{tr}_{m_{j+1}}^{m_j}(z)\}^{r_{j+1}}$, $A(t)$ can also be expressed as

$$\begin{aligned} A(t) &= \mathrm{tr}_1^{m_K}(\{\mathrm{tr}_{m_K}^{m_{K-1}}(\cdots\{\mathrm{tr}_{m_{j+2}}^{m_{j+1}}(\omega)\}^{r_{j+2}}\cdots)\}^{r_K}) \\ &= \sum_{a \in I(m_{j+1}, \cdots, m_K;\, r_{j+2}, \cdots, r_K)} \mathrm{tr}_1^{m_{j+1}}(\omega^a) \\ &= \sum_{a \in I(m_{j+1}, \cdots, m_K;\, r_{j+2}, \cdots, r_K} \mathrm{tr}_1^{m_{j+1}}(\{\mathrm{tr}_{m_{j+1}}^{m_j}(z)\}^{a r_{j+1}}) \\ &= \sum_{a \in I(m_{j+1}, \cdots, m_K;\, r_{j+2}, \cdots, r_K)} \\ &\quad \times \sum_{b \in I(m_j,\, m_{j+1};\, \langle a r_{j+1} \rangle m_{j+1})} \mathrm{tr}_1^{m_j}(z^b). \end{aligned} \quad (28)$$

Comparing (27) with (28), we have

$$I(m_j, m_{j+1}, \cdots, m_K;\, r_{j+1}, \cdots, r_K) = \{i \in I(m_j,\, m_{j+1};\, \langle a r_{j+1} \rangle_{m_{j+1}}) \mid a \in I(m_{j+1}, \cdots, m_K;\, r_{j+2}, \cdots, r_K)\}. \quad \square$$

Now, we are ready to compute the linear span. To obtain it, we should know the number of a's in $I(n,\, m_1,\, m_2, \cdots, m_K;\, r_1,\, r_2, \cdots, r_K)$ belonging to distinct cosets modulo $(2^n - 1)$, after the pairwise cancellation of those in the same coset, and their coset sizes. The following theorem shows that all the a's in $I(n,\, m_1,\, m_2, \cdots, m_K;\, r_1,\, r_2, \cdots, r_K)$ belong to distinct cosets modulo $(2^n - 1)$, and their coset sizes are n.

Theorem 12: Let $m_i = e_{i+1} m_{i+1}$ for $1 \le i \le K-1$ and $e_i > 1$ for $2 \le i \le K$. Let r_j be an integer relatively prime to $2^{m_j} - 1$, respectively, for $1 \le j \le K$. Let

$$I(m_j,\, m_{j+1}, \cdots, m_K;\, r_{j+1}, \cdots, r_K), \qquad j = 1, 2, \cdots, K-2$$

be defined as in (23). Then any two elements in

$$I(m_j,\, m_{j+1}, \cdots, m_K; r_{j+1}, \cdots, r_K)$$

are not in the same cyclotomic cosets of $\mathrm{GF}(2^{m_j})$. Moreover, the size of the cyclotomic coset in $\mathrm{GF}(2^{m_j})$ of any element in $I(m_j,\, m_{j+1}, \cdots, m_K;\, r_{j+1}, \cdots, r_K)$ is m_j.

Proof: The proof will be done by induction. First, consider the initial case, the case when $j = K-2$. From (23), we have

$$I(m_{K-2}, m_{K-1}, m_K; r_{K-1}, r_K) = \{i \in I(m_{K-2}, m_{K-1}; \langle a r_{K-1}\rangle_{m_{K-1}}) \mid a \in I(m_{K-1}, m_K; r_K)\}.$$

Pick any two elements i_1 and i_2 from $I(m_{K-2}, m_{K-1}, m_K; r_{K-1}, r_K)$. Without loss of generality, we can say that

$$i_1 \in I(m_{K-2}, m_{K-1}; \langle a_1 r_{K-1}\rangle_{m_{K-1}}) \quad \text{and} \quad i_2 \in I(m_{K-2}, m_{K-1}; \langle a_2 r_{K-1}\rangle_{m_{K-1}})$$

where a_1 and a_2 are in $I(m_{K-1}, m_K; r_K)$. From Proposition 3, a_1 and a_2 belong to the distinct cyclotomic cosets of GF$(2^{m_{K-1}})$. Thus from Lemma 7, i_1 and i_2 are not in the same cyclotomic coset of GF$(2^{m_{K-2}})$. Also, again from Lemma 7 and Proposition 3, we know that the size of cosets to which i_1 and i_2 belong is m_{K-2}. Thus the statement in Theorem 12 holds when $j = K-2$. Second, assume that the statement is true for $j = l+1$, i.e., any two elements in $I(m_{l+1}, m_{l+2}, \cdots, m_K; r_{l+2}, \cdots, r_K)$ are not in the same cyclotomic coset of GF$(2^{m_{l+1}})$ and the coset size is m_{l+1}. From (23), we have

$$I(m_l, m_{l+1}, \cdots, m_K; r_{l+1}, \cdots, r_K) = \{i \in I(m_l, m_{l+1}; \langle a r_{l+1}\rangle_{m_{l+1}}) \mid a \in I(m_{l+1}, \cdots, m_K; r_{l+2}, \cdots, r_K)\}.$$

Pick any two elements i_1 and i_2 from $I(m_l, m_{l+1}, \cdots, m_K; r_{l+1}, \cdots, r_K)$ such that

$$i_1 \in I(m_l, m_{l+1}; \langle a_1 r_{l+1}\rangle_{m_{l+1}}) \quad \text{and} \quad i_2 \in I(m_l, m_{l+1}; \langle a_2 r_{l+1}\rangle_{m_{l+1}})$$

where a_1 and a_2 are in $I(m_{l+1}, \cdots, m_K; r_{l+2}, \cdots, r_K)$. Since a_1 and a_2 are in distinct cosets of GF$(2^{m_{l+1}})$ from the assumption, i_1 and i_2 are in the distinct cosets of GF(2^{m_l}). Now, check the size of the cosets. Without loss of generality, assume the size of the coset to which i_1 belongs is less than m_l, i.e., for some s $(< m_l)$

$$i_1 \equiv 2^s i_1 \mod (2^{m_l} - 1). \tag{29}$$

Identity (29) implies that s is some divisor of m_l. By reducing (29) modulo $2^{m_{l+1}} - 1$, we have

$$a_1 r_{l+1} \equiv 2^{s'} a_1 r_{l+1} \mod (2^{m_{l+1}} - 1) \tag{30}$$

where $s' \equiv s \pmod{m_{l+1}}$. Identity (30) implies that s' must be zero, i.e., s must be a multiple of m_{l+1}. By setting $a_1 r_{l+1}$ and i_1 as

$$a_1 r_{l+1} = 2^{v_1} + 2^{v_2} + 2^{v_3} + \cdots + 2^{v_w}$$
$$i_1 = 2^{v_1} + 2^{v_2+k_2 m_{l+1}} + 2^{v_3+k_3 m_{l+1}} + \cdots + 2^{v_w+k_w m_{l+1}}$$

(29) can be rewritten as

$$(2^{v_1} + 2^{v_2+k_2 m_{l+1}} + \cdots + 2^{v_w+k_w m_{l+1}}) = 2^s (2^{v_1} + 2^{v_2+k_2 m_{l+1}} + \cdots + 2^{v_w+k_w m_{l+1}}) \mod (2^{m_l} - 1). \tag{31}$$

In (31), it is easy to see that the only s which is a divisor of m_l and a multiple of m_{l+1} is m_l, since the only possible term in the right-hand side of (31) which is congruent to $1 \mod 2^{m_l} - 1$ is 2^s. Therefore, the size of the coset to which i_1 belongs is m_l, i.e., the statement is true for $j = l$, which ends the proof.

From Theorem 12, the computation of the linear span of cascaded GMW sequences is straightforward. It is summarized as Theorem 13 below.

Theorem 13: Let $n = e_1 m_1$, $m_i = e_{i+1} m_{i+1}$ for $1 \le i \le K-1$, and $e_i > 1$ for $1 \le i \le K$. Let r_j be an integer less than and relatively prime to $2^{m_j} - 1$, respectively, for $1 \le j \le K$. Let a cascaded GMW sequence $A(t)$ of period $2^n - 1$ be given by

$$A(t) = \mathrm{tr}_1^{m_K}(\{\mathrm{tr}_{m_K}^{m_{K-1}}(\{\mathrm{tr}_{m_{K-1}}^{m_{K-2}} (\{\cdots\{\mathrm{tr}_{m_2}^{m_1}(\{\mathrm{tr}_{m_1}^{n}(\alpha^t)\}^{r_1})\}^{r_2}\cdots\}^{r_{K-1}})\}^{r_K}).$$

Then, the linear span L of $A(t)$ is

$$L = n \cdot |I(n, m_1, m_2, \cdots, m_K; r_1, r_2, \cdots, r_K)| = n\left\{\sum_{a \in I(m_1, m_2, \cdots, m_K; r_2, \cdots, r_K)} \left(\frac{n}{m_1}\right)^{\mathrm{wt}\,(\langle a r_1\rangle_{m_1})-1}\right\} \tag{32}$$

where the index sets $I(m_1, m_2, \cdots, m_K; r_2, \cdots, r_K)$ and $I(n, m_1, m_2, \cdots, m_K; r_1, r_2, \cdots, r_K)$ are recursively obtained as in (23) and $\mathrm{wt}\,(\langle a r_1\rangle_{m_1})$ is number of ones in the binary representation of $a r_1 \mod (2^{m_1} - 1)$.

Example 14: Let $n = 12$, $m_1 = 6$, $m_2 = 3$, $r_1 = 5$, and $r_2 = 3$. The cascaded GMW sequence $A(t)$ is given by

$$A(t) = \mathrm{tr}_1^3(\{\mathrm{tr}_3^6(\{\mathrm{tr}_6^{12}(\alpha^t)\}^5)\}^3).$$

From (19), the index set $I(12, 6, 3; 5, 3)$ is given as

$$I(12, 6, 3; 5, 3) = \{i \in I(12, 6; 5a) \mid a \in I(6, 3; 3)\}$$

and the index set $I(6, 3; 3)$ is given as $I(6, 3; 3) = \{3, 17\}$. Thus

$$\begin{aligned} I(12, 6, 3; 5, 3) &= I(12, 6; \langle 3 \cdot 5\rangle_6) \bigcup I(12, 6; \langle 17 \cdot 5\rangle_6) \\ &= I(12, 6; 15) \bigcup I(12, 6; 11) \\ &= \{15, 519, 267, 771, 141, 645, 393, 897\} \\ &\quad \bigcup \{11, 137, 515, 641\}. \end{aligned}$$

In other words, $A(t)$ can be rewritten as the sum of decimated m-sequences as

$$A(t) = \sum_{a \in I(12, 6, 3; 5, 3)} \mathrm{tr}_1^{12}(\alpha^{at}).$$

From Theorem 13, since $\mathrm{wt}\,(15) = 4$, and $\mathrm{wt}\,(11) = 3$, the linear span L is

$$L = 12\left(\frac{12}{6}\right)^{4-1} + 12\left(\frac{12}{6}\right)^{3-1} = 144.$$

REFERENCES

[1] L. D. Baumert, *Cyclic Difference Sets* (Lecture Notes in Mathematics). Berlin, Germany: Springer-Verlag, 1971.

[2] D. Jungnickel, "Difference set," in *Contemporary Design Theory*, J. H. Dinitz and D. R. Stinson, Eds. New York: Wiley, 1992, pp. 241–324.

[3] S. W. Golomb, *Shift-Register Sequences*, revised ed. San Francisco, CA: Aegean Park, 1982.

[4] R. Lidl and H. Niederreiter, *Finite Fields*, vol. 20 of *Encyclopedia of Mathematics and Its Applications.* Reading, MA: Addison-Wesley, 1983.

[5] M. K. Simon, J. K. Omura, R. A. Scholtz, and B. K. Levitt, *Spread Spectrum Communications*, vol. 1. Rockville, MD: Computer Science, 1985.

[6] A. Klapper, A. H. Chan, and M. Goresky, "Cascaded GMW sequences," *IEEE Trans. Inform. Theory*, vol. 39, pp. 177–183, Jan. 1993.

[7] G. Gong, "Q-ary cascaded GMW sequences," *IEEE Trans. Inform. Theory*, vol. 42, pp. 263–267, Jan. 1996.

[8] J.-S. No, H.-K. Lee, H. Chung, H.-Y. Song, and K. Yang, "Trace representation of Legendre sequence of Mersenne prime period," *IEEE Trans. Inform. Theory*, vol. 42, pp. 2254–2255, Nov. 1996.

[9] J.-S. No, K. Yang, H. Chung, and H.-Y. Song, "On the construction of binary sequences with ideal autocorrelation property," in *Proc. 1996 IEEE Int. Symp. Information Theory and Its Applications* (ISITA '96) (Victoria, BC, Canada, Sept. 17–20, 1996), pp. 837–840.
[10] J.-S. No, "Generalization of GMW sequences and No sequences," *IEEE Trans. Inform. Theory*, vol. 42, pp. 260–262, Jan. 1996.
[11] R. A. Sholtz and L. R. Welch, "GMW sequences," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 548–553, May 1984.
[12] S. W. Golomb, "On the classification of balanced binary sequences of period $2^n - 1$," *IEEE Trans, Inform. Theory*, vol. IT-26, pp. 730–732, Nov. 1980.
[13] L. Brynielsson, "On the linear complexity of shift registers," in *Advances in Cryptology-Eurocrypt '85*. Berlin, Germany: Springer-Verlag, 1985, pp. 156–166.
[14] J.-S. No, "A new family of binary pseudorandom sequences having optimal periodic correlation properties and large linear span," Ph.D dissertation, Univ. So. Calif., Los Angeles, CA, May. 1988.

Transform Domain Analysis of DES

Guang Gong and Solomon W. Golomb, *Fellow, IEEE*

***Abstract*— The Data Encryption Standard (DES) can be regarded as a nonlinear feedback shift register (NLFSR) with input. From this point of view, the tools for pseudo-random sequence analysis are applied to the S-boxes in DES. The properties of the S-boxes of DES under Fourier transform, Hadamard transform, extended Hadamard transform, and Avalanche transform are investigated. Two important results about the S-boxes of DES are found. The first result is that nearly two-thirds of the total 32 functions from GF (2^6) to GF (2) which are associated with the eight S-boxes of DES have the maximal linear span 63, and the other one-third have linear span greater than or equal to 57. The second result is that for all S-boxes, the distances of the S-boxes approximated by monomial functions has the same distribution as for the S-boxes approximated by linear functions. Some new criteria for the design of permutation functions for use in block cipher algorithms are discussed.**

***Index Terms*— Block cipher, DES, nonlinear feedback shift register, transform domain analysis.**

I. Introduction

The Data Encryption Standard (DES) is a block cipher involving 64-bit data encryption with a 56-bit key, which was adopted by the U.S. National Institute of Standards and Technology in 1976. DES has been widely used in bank activities and Internet communications. The security of DES has been extensively investigated by many researchers [2], [15], [3], [4], [26], [6], [23], [9], [5], and [20]. DES can be implemented in hardware as well as software.

In this correspondence, we will consider DES as a nonlinear feedback shift register (NLFSR) with input. From this point of view, we will apply the tools for pseudo-random sequence analysis to the S-boxes in DES, i.e., the feedback function when DES is regarded as a NLFSR with input. We will exhibit several new properties of the S-boxes of DES and discuss some new criteria for design of block cipher algorithms.

Concerning the analysis of shift-register sequences, we have three kinds of known transforms: the Fourier transform, the Hadamard transform and the Avalanche transform. Pieprzyk [19], Nyberg [18], and Webster and Tavares [26] discussed the nonlinearity and the Strict Avalanche Criterion (SAC) of a function from Z_2^n to Z_2^m. For a function from Z_2^n to Z_2^m, the Fourier transform represents its linear span property, the Hadamard transform reflects its nonlinearity, and the Avalanche transform shows whether it satisfies SAC. In analysis of shift-register sequences [7], we consider that all m-sequences are equivalent under the decimation operation for elements in a sequence. We apply this idea to approximate the S-boxes in DES, i.e., we use monomial functions instead of linear functions to approximate S-boxes. We call this transform an *extended Hadamard transform* whose definition will be given in Section IV. We found that distributions of the extended Hadamard transform spectra of all S-boxes are the same as the distributions of the Hadamard transform spectra of the S-boxes.

This paper is organized as follows. In Section II, we present some basic concepts that will be used thoughtout this paper. In Section III, we show the Fourier transform spectrum of the S-boxes of DES, which gives that nearly two-thirds of the total of 32 functions from GF (2^6) to GF (2), which are associated with the eight S-boxes of DES, have the maximal linear span value 63. In Section IV, first we introduce the extended Hadamard transform for a function from GF (2^n) to GF (2), then we discuss new criteria for design of permutation functions for use in block cipher algorithms, and we present Hadamard transform spectra, extended Hadamard transform spectra, and Avalanche transform spectra of the S-boxes of DES.

Remark 1: RC5, which was invented by Rivest [21] in 1994, is also a block cipher with parameters that can be easily switched into a mode of 64-bit, or 128-bit, or 256-bit data encryption. RC5 is widely used in Internet communications [24]. The security of RC5 was discussed at recent Crypto conferences [10], [12]. Until now, RC5 has been implemented in software. The approach developed in this correspondence can be applied to RC5, since RC5 has the same NLFSR structure as DES, only differing in their feedback functions. So, RC5 can be easily implemented in hardware in terms of its NLFSR architecture. For RC5, the feedback function is a function from a ring Z_{32} to Z_{32} (here we assume that it is in a 64-bit mode). But it can be transformed into a function from Z_2^{32} to Z_2^{32}, or, equivalently, a function from GF (2^{32}) to GF (2^{32}). Thus an analysis of transform-domain properties of the feedback function of RC5 can be partially done by computation. We will discuss this in a separate paper. Other block ciphers widely used in Internet communications [22], such as IDEA [13] and SAFER K-64 [17], are different modes. They directly use a permutation function from Z_2^n to Z_2^n instead of feedback shift-resister structures. But the transform spectrum analysis techniques used for DES also can be applied to them.

II. Preliminaries

In this section, we will adopt some tools from pseudo-random sequences for the analysis of functions used in the design of conventional cryptosystems.

Let $p = 2^n - 1$ and $f(x)$ be a map from a finite field GF (2^n) to GF (2), i.e., $f(x)$: GF $(2^n) \rightarrow$ GF (2). Let α be a primitive element in GF (2^n). A positive integer r is a coset leader modulo p means that

Manuscript received July 5, 1998; revised February 4, 1999.

G. Gong was with the Communication Sciences Institute, University of Southern California, Electrical Engineering-Systems, EEB # 500, Los Angeles, CA 90089-2565 USA. She is now with the Department of Combinatorics and Optimization, University of Waterloo, Waterloo, Ont., Canada N2L 3G1.

S. Golomb is with the Communication Sciences Institute, University of Southern California, Electrical Engineering-Systems, EEB # 500, Los Angeles, CA 90089-2565 USA.

Communicated by D. R. Stinson, Associate Editor for Complexity and Cryptography.

Publisher Item Identifier S 0018-9448(99)05868-X.