通信情報保護學會誌
第4卷 第1號, 1994. 3

# (벡터)부울 함수의 비선형성에 관한 연구

정 하 봉*

## ABSTRACT

본 논문에서는 벡터 부울함수의 비선형성의 척도인 선형 함수군까지의 거리에 대해 알아본다. 특히 S-Box의 내부 함수로 이용될 수 있는 출력의 균등 분포성을 만족하는 벡터 부울함수의 선형함수군까지의 거리의 하계를 유도한다. 더불어 DES에서의 S-Box 내부함수의 비선형도를 산출하고 보다 비선형도가 높은, 즉 선형 함수군까지의 거리가 더 먼 새로운 S-Box의 존재를 밝힌다.

## 1. 서 론

암호 시스템에서 자료 보안방법으로 가장 일반적으로 사용되고 있는 암호 알고리즘에는 미국IBM에서 개발한 DES(Data Encryption Standard)를 들 수 있다. DES는 기본적으로 S-Box(대체상자), 비트 재배열(bit permutation), Modulo-2 (exclusive OR) 연산을 반복적으로 사용하여 암호의 강도를 높이는 반복 암호시스템(iterative cryptosystem)이다. 이러한 기본적인 연산 중 비선형적 특성을 갖는 것은 S-Box에 불과하고 따라서 이러한 암호시스템의 보안성은 상당부분이 S-Box의 선택에 달렸다고 할 수 있다.

원래 S-Box는 입, 출력 공히 $n$비트를 가져 $2^n$개의 출력벡터가 $2^n$개의 입력벡터의 자리바꿈(permutation)이 되는 시스템을 의미한다. 다시 말해 S-Box의 내부 변환 함수는 정의역과 치역이 공히 $[GF(2)]^n$인 일대일 대응관계의 이진 벡터 부울함수가 된다. DES에 쓰이는 S-Box처럼 입력 비트수가 출력비트수보다 많도록 S-Box를 확장 정의할 수도 있다. 이때 DES S-Box처럼 입력 6비트 중 특정 2비트를 4개의 서로 다른 4-bit input/4-bit output의 S-Box 중 하나를 결정하는 선택비트로 사용하여 6-bit input/4-bit output의 S-Box처럼 보이게 하는 경우도 있을 수 있겠으나 일반적으로는 출력의 균등 분포 조건(예컨대, 6-bit input/4-bit output S-Box의 경우 16개의 4-bit 출력 조합 모두가 공히 4번씩 출력된다.)만을 만족하면 S-Box라고 확장 정의할 수도 있다.

일반적으로 S-Box의 내부 변환 함수는 비선형이다. 선형 함수는 외부로부터의 공격에 대단히 취약하므로 S-Box의 내부 변환 함수는 그 비선형성이 강할수록 외부로부터의 공격에 강하다고 믿어지고 있다. 여기서 우리가 생각해야 할 것은 함수의 비선형성의 척도를 어떻게 정해야 할 것인가 하는 문제이다. 예를 들어 S-Box에 대한 가장 대표적인 공격 방법으로 들 수 있는 Biham과 Shamir가 제안한 Differential Cryptanalysis의 경우를 보자. Differential Cryptanalysis는 평문쌍

* 정회원, 홍익대학교 전자공학과

(plaintext pair)의 XOR이 어떤 특정한 값을 갖고 입력될 때 이에 상응하여 발생되는 암호문쌍(ciphertext pair)의 XOR 값이 높은 확률로 발생하는다는 데 기초를 둔 chosen-plaintext 공격방법이다. 따라서 DES, 곧 S-Box가 Differential Cryptanalysis에 견디기 위해서는 S-Box의 XOR 분포가 비교적 균일해야 한다고 말할 수 있다[13]. 따라서 Differential Cryptanalysis 관점에서의 함수의 비선형성의 척도는 출력의 XOR 분포의 균일 정도로 볼 수 있을 것이다.(참고로 또 다른 연구에서는 S-Box의 XOR 분포가 완전히 균등하게 되면 그렇지 않을 때에 비해 훨씬 더 취약하다는 사실도 증명되었다[12].)

본 논문에서는 "(벡터) 부울함수의 비선형성이란 그 함수가 선형 함수군으로부터 얼마만큼 멀리 있는가."라는 관점에서 어떤 (벡터) 부울함수가 선형함수군으로부터 떨어져 있을 수 있는 최대 해밍거리(Hamming distance)를 그 함수의 비선형성의 척도로 잡았다. 이진 부울함수의 경우, 선형함수군으로부터 가장 멀리 떨어져 있는 함수(이러한 부울함수는 bent 함수라 불린다.)는 그 출력의 XOR이 균등 분포한다는 사실은 적어도 (스칼라) 부울함수에 대해서는 본 논문에서 정한 비선형성의 척도가 Differential Cryptanalysis의 관점에서는 타당하다는 점을 말하고 있다 하겠다.

본 논문에서는 "벡터 부울함수가 선형 함수로부터 최대 얼마 만큼 떨어져 있을 수 있는가?"하는 문제를 제기하려 한다. 제2절에서는 함수의 비선형성의 척도로 정한, 함수의 선형함수군까지의 Hamming 거리를 정의하고 그 최대 거리에 대한 하계(lower bound)를 유도한다. 제3절에서는 대상 함수를 출력의 균등분포성을 갖는, 즉 S-Box의 내부 변환 함수로 쓰일 수 있는 함수로 국한하여 역시 최대 거리에 대한 하계를 알아본다. 제4절에서는 DES에서 쓰이고 있는 8개의 S-Box 각각에 대해 내부 변환 함수의 선형함수군까지의 거리를 산출하고 이보다 더 큰 거리를 갖는 새로운 S-Box 내부함수의 설계 방법에 대해서 알아본다. 마지막으로 결론에서는 앞으로의 연구 문제들을 언급하였다. 언급된 명제나 정리에 대한 증명은 생략하였다. 제2절 이후 쓰인 주요 notation 중 $V^n$은 이진 $n$-차원 벡터공간을 의미하고 이진 벡터공간의 원소인 이진 벡터는 $X$, $Y$ 등으로, 그리고 벡터 $X$, $Y$의 성분들은 각각 $x_i$, $y_i$ 등으로 표시하였다. 또 함수 $f$, $g$ 등은 각각 (스칼라) 부울함수를, 함수 $F$, $G$ 등은 각각 벡터 부울함수를 지칭하는 것으로 구별했고, 부울함수 $f$의 출력 중 1의 갯수는 $wt[f]$로 나타내었다. 그리고 $n/m$ S-Box란 입력비트수가 $n$, 출력비트수가 $m$인 S-Box를 의미한다.

## 2. 벡터 부울함수의 선형 함수군까지의 해밍 거리

같은 정의역을 갖는 이진 부울함수 $f$와 $g$간의 해밍 거리 $d(f,g)$는 다음과 같이 정의한다.

$$d(f,g) = wt[f+g] \tag{2.1}$$

정의역이 $V^n$, 치역이 $V^m$인 벡터 부울함수란 $V^n$상에 정의된 이진 부울함수 $f_i$들을 성분으로 가지고 있는 함수 $F = (f_1, f_2, \cdots, f_m)$를 의미하며 식 (2.1)에서 정의된 거리의 개념을 벡터 부울함수 $F$와 $G$에 대해서 확장하여 보면 정의역이 $V^n$인 두 벡터 부울함수 $F$와 $G$간의 해밍거리 $d(F,G)$는 다음과 같이 정의할 수 있다.

$$d(F,G) = |\{X \in V^n \mid F(X) + G(X) \neq 0\}| \tag{2.2}$$

이제 정의역이 $V^n$, 치역이 $V^m$인 선형함수군(set of affine mappings) $L_{n,m}$을 정의하자. 행렬 $A$를 $(m \times n)$ 이진행렬, $B$를 $(m \times 1)$ 이진 열벡터라고 할 때 $V^n$상의 $(n \times 1)$ 열벡터 $X$에 대해서 $L(X) = [AX + B]^t$로 표시되는 함수를 선형함수라고 정의하고 정의역이 $V^n$, 치역이 $V^m$인 이러한 선형함수들의 집합을 $L_{n,m}$이라고 정의한다. 함수 $F: V^n \to V^m$의 $L_{n,m}$까지의 거리 $D_F$는 다음과 같이 정의할 수 있고

$$D_F = \min_{L \in L_{n,n}} \{ d(F,L) \} \tag{2.3}$$

본 논문에서는 함수 $F$의 비선형성의 척도로 식 (2.3)의 $D_F$를 차용한다. 정의역과 치역이 각각 $V^n$과 $V^m$인 함수중 가장 비선형성이 큰 함수 $F$의 $D_F$를 $d(n,m)$이라고 정의하자. 즉 $F_{n,m}$을 정의역과 치역이 각각 $V^n$과 $V^m$인 벡터 부울함수의 집합이라고 할때 $d(n,m)$은 다음과 같이 정의된다.

$$d(n,m) = \underset{F \in F_{n,m}}{Max} \{ D_F \} \qquad (2.4)$$

$m = 1$인 경우의 $d(n,1)$을 구하는 것은 선형 부울함수들을 부호어로 가지고 있는 부호, 즉 $[2^n, n+1]$ Reed-Muller code $RM(1,n)$의 최대 weight를 갖는 coset leader를 찾는 문제에 해당하므로 $RM(1,n)$의 covering radius를 구하는 고전적인 문제에 해당한다. $V^n$상의 부울함수 $f$의 Hadamard 변환 $F_H(u)$를 다음과 같이 정의할 때

$$F_H(u) = \sum_{x \subset V^n} (-1)^{f(X) + u \cdot X} \qquad (2.5)$$

함수 $f$의 선형함수 $\sum_{i=1}^{n} u_i x_i$와 $1 + \sum_{i=1}^{n} u_i x_i$까지의 거리는 각각

$$d(f, \Sigma u_i x_i) = \frac{1}{2}\{2^n - F_H(u)\} \qquad (2.6)$$

$$d(f, 1 + \Sigma u_i x_i) = \frac{1}{2}\{2^n + F_H(u)\} \qquad (2.7)$$

와 같이 나타낼 수 있다. 함수 $f$의 Hadamard 변환 $F_H(u)$는 Parseval의 정리에 의해 다음 성질

$$\Sigma F_H(u)^2 = 2^{2n} \qquad (2.8)$$

을 갖게 되므로 모든 $u$값에 대해 $|F_H(u)|^2$의 값이 $2^n$이 되는 함수 $f$가 있다면 식 (2.6)과 (2.7)에 의해서 그러한 함수 $f$의 선형함수군까지의 거리 $D_F$는 $2^{n-1} - 2^{(n/2-1)}$로 주어질 것이다. 이러한 함수를 bent 함수[1]라 하고 bent 함수가 바로 선형 함수군으로부터 가장 멀리 떨어진 함수가 된다. 따라서 $m = 1$인 경우의 $d(n,1)$은 다음과 같다.

$$d(n,1) = 2^{n-1} - 2^{\frac{n}{2}-1} \qquad (2.9)$$

Bent 함수의 구성 방법 및 분류에 대해서는 O.S. Rothaus[1], J.F. Dillon[13] 등에 의해 연구되었고 현재까지 알려진 Bent 함수의 분류는 Maiorana-McFarland's class, Partial-Spread class 등과 그의 변형된 형태들이 있으나 아직 완전한 분류가 되어 있지 않은 상태이다.

일반적으로 $n$이 홀수인 경우의 $d(n,1)$값은 $n$값이 비교적 작은 $n = 3, 5, 7$을 제외하고는 알려져 있지 않다. 그러나 $(n-1)$-차원 이진 벡터공간 상에서 정의된 bent 함수를 $V^n$상의 함수로 간주하게 되면 (예컨대, $f(x_1,x_2,x_3) = x_1x_2$) 그러한 함수의 선형함수군까지의 거리는 $2d(n-1,1)$이 되므로 다음의 하계가

$$d(2k+1,1) \geq 2^{2k} - 2^k \qquad (2.10)$$

알려져 있고 실제로 $n = 3, 5, 7$일때의 $d(n,1)$의 값은 $d(3,1) = 2$, $d(5,1) = 12$[7], $d(7,1) = 56$[8]이 된다. 식 (2.10)에서 등호가 만족될 조건은 다음의 명제가 보여 준다.

**▨ 명제 2.1**

$d(2k+1,1) = 2d(2k,1)$이면 $V^{2k}$상의 임의의 부울함수 $f$와 $g$에 대해 다음이 만족된다.

$$\underset{l \in L_{2k,1}}{Min} \{ \underset{c \in \{0,1\}}{Min} \{wt[f+l] + wt[g+l+c]\} \} \leq 2^{2k} - 2^k$$

한편 1983년 Patternson과 Wiedemann [10]에 의해 $n \geq 15$인 홀수에 대해서는 다음의 개선된 하계가 발표되었다.

$$d(2k+1,1) \geq 2^{2k} - 108 \cdot 2^{k-7}, \; k \geq 7 \qquad (2.11)$$

그리고 $d(2k+1,1)$의 상계(upper bound)로는 다음이 알려져 있다.

$$d(2k+1,1) < 2^{2k} - 2^{k-\frac{n}{2}} \qquad (2.12)$$

이제 $m > 1$인 벡터 부울함수의 경우에 대해서 살펴보자. 정의역과 치역이 각각 $V^n$과 $V^m$인 벡터 부울함수 $F = (f_1, f_2, \cdots, f_m)$와 선형 벡터 부울함수 $L = (l_1, l_2, \cdots, l_m)$과의 거리 $d(F,L)$은 $2^n$에서 $F(X)$와 $L(X)$가 같게 되는 입력 $X$의 갯수를 뺀 것이므로 다음 식이 성립하고

$$d(F,L)=2^n-|\{X\in V^n \mid \prod_{i=1}^{n}(F_i(X)+l_i(X)+1)=1\}| \quad (2.13)$$

이는 다시 다음과 같이 나타낼 수 있다.

$$d(F,L) = \frac{1}{2^{m-1}}\{\sum_{c_1=0}^{1}\cdots\sum_{c_m=0}^{1} wt[\sum_{j=1}^{m}c_j(f_j+l_j)]\} \quad (2.14)$$

이제 이해의 편이를 위해 $m = 2$인 경우부터 살펴보자. 식 (2.14)는

$$d(F,L) = \frac{1}{2}\{wt[f_1+l_1]+wt[f_2+l_2] +wt[f_1+f_2+l_1+l_2]\}$$

가 되고 따라서

$$D_F = \frac{1}{2}\min_{l_1}\{wt[f_1+l_1]+\min_{l_2}\{wt[f_2+l_2] +wt[f_1+f_2+l_1+l_2]\}\} \quad (2.15)$$

로 표시할 수 있다. 만일 $f_1,f_2,f_1+f_2$ 모두가 bent 함수가 되고 그때

$$wt[f_1+l_1] = wt[f_2+l_2] = wt[f_1+f_2+l_1+l_2] = 2^{n-1}-2^{\frac{n}{2}-1}$$

이 되는 $l_1$, $l_2$가 존재한다면 $D_F = 3(2^{n-2}-2^{\frac{n}{2}-2})$가 될 것이다. 즉 일반적으로 $m$개의 bent 함수 $f_1, f_2,\cdots,f_m$이 있어 그들의 0이 아닌 임의의 선형조합 $\sum_{i=1}^{m}c_if_i$ 역시 bent 함수가 된다면 식 (2.14)에 의해 함수 $F = (f_1,f_2,\cdots,f_m)$와 임의의 선형함수 $L = (l_1, l_2,\cdots,l_m)$과의 거리 $d(F,L)$은 $d(F,L) \geq (2^m-1)(2^{n-m}-2^{\frac{n}{2}-m})$이 되고 만일

$$wt[\sum_{i=1}^{m}c_i(f_i+l_i)] = 2^{n-1}-2^{\frac{n}{2}-1}, \quad \text{for all nonzero } \{c_i\} \quad (2.16)$$

을 성립시키는 $\{l_i\}$가 존재한다면 $D_F$는

$$D_F = \frac{1}{2^{m-1}}\min_{\{l_i\}}\{\sum_{c_1=0}^{1}\cdots\sum_{c_m=0}^{1} wt[\sum_{j=1}^{m}c_j(f_j-l_j)]\} = (2^m-1)(2^{n-m}-2^{\frac{n}{2}-m})$$

이 된다. K. Nyberg[2]에 의하면 $n$이 짝수이고 $m \leq n/2$인 경우, 소위 완전비선형(perfect nonlinear) 함수로 불리우는 $m$개의 bent 함수들이 존재하고 이 $m$개의 함수들의 임의의 (0이 아닌) 선형조합 역시 bent 함수가 된다고 알려져 있다. 이러한 완전비선형 함수를 만드는 가장 간단한 방법으로는 shift register를 이용하는 방법이 알려져 있다. 그리고 그러한 방법으로 $f_i(X)$를 만들었을 때 다음의 명제로 부터 식 (2.16)을 만족하는 $\{l_i\}$가 존재함을 알 수 있게 되어

$$D_F = (2^m-1)(2^{n-m}-2^{\frac{n}{2}-m})$$

이 됨을 보일 수 있다.

### ▩ 명제 2.2

$X = (x_1,x_2,\cdots,x_{2k})$, $X_1 = (x_1,x_3,\cdots,x_{2k-1})$, $X_2 = (x_2,x_4,\cdots,x_{2k})$라고 하자. 길이$(2^k-1)$인 $m$-sequence의 선형궤환 shift register의 상태전이 함수를 $A$라고 할 때 다음 $m(\leq k)$개의 함수 $f_i(X) = A^{i-1}(X_1)\bullet X_2$, $i = 1,2,\cdots,m$에 대해

$$wt[\sum_{i=1}^{m}c_i(f_i+l_i)] = 2^{2k-1}-2^{k-1}$$

이 되는 $m$개의 선형함수 $l_i(X)$, $i = 1,2,\cdots,m$가 존재한다.

따라서 명제 2.2로부터 다음의 하계를 얻을 수 있다.

### ❖ 정리 2.1

$$d(2k,m)\geq(2^m-1)(2^{2k-m}-2^{k-m}) \quad \text{if} \quad m\leq k.$$

한편 다음의 정리 2.2는 어떤 경우에 위 정리 2.1의 등식이 만족되는가를 보여주고 있다.

### ❖ 정리 2.2

$$d(2k,m)=(2^m-1)(2^{2k-m}-2^{k-m}) \quad \text{if} \quad d(2k+1,1)=2d(2k,1).$$

앞에서 언급한 바와 같이 $k = 2$와 3인 경우 $d(2k+1,1) = 2d(2k,1)$이 되므로 정리 2.2로부터 다음을 알 수 있다. $d(4,2) = 9$, $d(6,2) = 42$.

$d(6,3) = 49$.

$n$이 홀수인 경우는 부울함수의 경우와 마찬가지로 $V^{2k}$상의 임의의 벡터 부울함수 $F$를 그대로 $V^{2k+1}$상의 함수로 생각하면 $V^{2k+1}$에서의 선형함수군까지의 거리는 $V^{2k}$상에서의 선형함수군까지의 거리의 두배가 된다. 따라서 다음의 정리를 얻을 수 있다.

❖ 정리 2.3 $d(2k+1,m) \geq 2d(2k,m)$

## 3. 균등 분포성을 갖는 벡터 부울함수의 선형함수군까지의 최대 거리

서론에서 언급한 바와 같이 S-Box의 내부 변환함수는 반드시 균등분포성을 가져야 한다. 다시 말해 정의역과 치역이 각각 $V^n$과 $V^m$인 함수 $n/m$ S-Box의 내부 변환함수라 할 때 $F$가 균등분포성을 갖는다는 말은 임의의 출력 $Y \in V^m$에 대해 $F(X) = Y$가 되는 입력 $X \in V^n$가 정확히 $2^{n-m}$개 있어야 한다는 말이다. 이 절에서는 이러한 균등분포성을 갖는 S-Box 내부 변환함수의 선형함수군까지의 거리에 관하여 알아보겠다. 우선 앞서 정의한 최대 거리 파라미터를 다시 정의해야 한다.

▩ 정의 3.1 최대 거리 파라미터 $D(n,m)$은 정의역과 치역이 각각 $V^n$과 $V^m$이며 균등분포성을 갖는 함수가 가질 수 있는 선형함수군까지의 최대 거리이다. 즉,

$$D(n,m) = \underset{F}{Max}\{\underset{L}{Min}\{d(F,L)\}\}$$

이때 선형함수군까지의 거리가 $D(n,m)$인 함수 $F: V^n \to V^m$를 S-Box bent 함수라고 명명한다.

$m = 1$, $n = 2k$인 경우에 함수 $f(x_1,x_2,\cdots,x_{2k}) = \sum_{i=1}^{k-1} x_{2i-1}x_{2i}$의 선형함수 $l(X) = c_0 + \sum_{i=1}^{2k} c_i x_i$까지의 거리를 계산하면 다음과 같다.

$$d(f,l) = \begin{cases} 2^{2k-1} \pm 2^k & \text{if } c_{2k-1} = c_{2k} = 0 \\ 2^{2k-1} & \text{otherwise} \end{cases}$$

따라서 예컨대 함수 $g(X) = x_{2k} + \sum_{i=1}^{k-1} x_{2i-1}x_{2i}$는 균등분포성을 가지며 $D_g = 2^{2k-1} - 2^k$가 된다.

고로 다음의 하계를 얻을 수 있다.

❖ 정리 3.1 $D(2k,1) \geq 2^{2k-1} - 2^k$ (3.1)

실제로 [16,5] Reed-Muller code와 [64,7] Reed-Muller code의 coset들의 weight 분포는 완전히 밝혀졌고 그로부터 식 (3.1)에서 등호는 $k = 2$와 3의 경우 성립함을 알 수 있다. 즉, $D(4,1) = 4$, $D(6,1) = 24$.

다음으로 $n = 2k + 1$인 홀수의 경우는 다음의 함수 $f(x_1,x_2,\cdots,x_{2k+1}) = x_{2k+1} + \sum_{i=1}^{k} x_{2i-1}x_{2i}$가 균등분포성을 갖고 $D_f = 2^{2k} - 2^k$임을 알 수 있게 된다. 따라서 역시 다음과 같은 하계를 얻을 수 있고

❖ 정리 3.2 $D(2k+1,1) \geq 2^{2k} - 2^k$ (3.2)

$d(n,m) \geq D(n,m)$이라는 사실과 $d(5,1) = 12$, $d(7,1) = 56$이라는 사실로부터 $k = 2$와 3인 경우는 식 (3.2)의 등호가 만족됨을 알 수 있다. 즉, $D(5,1) = 12$, $D(7,1) = 56$.

$n = 2k$이고 $m \leq k$인 벡터 부울함수의 경우를 살펴보자. 앞에서 언급한 Nyberg의 완전 비선형 함수는 거리 성질은 우수하다고 믿어지나 균등분포성을 만족하지 못한다. 실제로 Nyberg가 제안한 $V^{2k}$상에서 shift register를 이용하여 만들어진 $m$개의 완전 비선형함수 $f_i$을 성분으로 하는 벡터 부울함수 $F = (f_1,f_2,\cdots,f_m)$은 all zero 출력이 $(2^{2k-m} - 2^{k-m} + 2^k)$번, 그리고 그밖의 출력은 정확히 $(2^{2k-m} - 2^{k-m})$번씩 나타난다. 만일 이 함수 $F$의 우수한 거리 성질은 유지한 채로 균등분포성을 만족할 수 있도록 $F$를 변형할 수 있다면 그 새로운 함수는 거리 성질이 우수한 S-Box의 내부 함수로 차용될 수 있을 것이다. 실예로 명제 2.2의 $f_i(X)$에 각각 $x_{2i}\prod_{j=0}^{m-1}(x_{2j+1}+1)$을 더해 만들어진 함수 $f_i'(X)$를 성분으로 하는 벡터 부울함수 $F'(X) = (f_1',f_2',\cdots,$

$f_m'$)를 보면 균등분포성을 만족하고 또 $f_i'$들의 0이 아닌 모든 선형조합 $\Sigma c_i f_i'$들은 선형함수군까지의 거리가 식 (3.1)의 하계를 만족하게 된다. 따라서 이러한 함수를 이용, 다음의 하계를 얻을 수 있고

❖ **정리** 3.3 $D(n,m) \geq (2^m - 1)(2^{n-m} - 2^{k+1-m})$
for $n = 2k,\ m \leq k$

이 결과를 $n = 2k + 1$인 경우로 확장하면 다음의 정리를 유도할 수 있다.

❖ **정리** 3.4 $D(n,m) \geq (2^m - 1)(2^{n-m} - 2^{k+2-m})$
for $n = 2k + 1,\ m \leq k$

앞의 두 정리의 하계를 만족하는 $2m/m$ S-Box의 내부함수의 설계는 다음의 정리 3.5와 같다.

❖ **정리** 3.5 $X = (x_1, x_2, \cdots, x_{2m})$, $X_1 = (x_1, x_3, \cdots, x_{2m-1})$, $x_2 = (x_2, x_4, \cdots, x_{2m})$이라 하자. 길이 $(2^m - 1)$인 $m$-sequence를 발생시키는 선형궤환 shift register의 상태전이함수를 $A$라고 할 때 다음의 함수 $f_i$, $i = 1, 2, \cdots, m$을 성분으로 하는 벡터 부울 함수 $F = (f_1, f_2, \cdots, f_m) : V^{2m} \to V^m$은 출력의 균등분포성을 만족하고

$$f_i(X) = A^{-i}(X_1) \cdot X_2 + x_{2i} \prod_{j=1}^{m} (x_{2j-1} + 1) \qquad (3.3)$$

선형함수군까지의 거리는 $D_F = (2^m - 1)(2^m - 2)$이다.

## 4. DES의 S-Boxes.

DES에는 8개의 6/4 S-Box가 사용되고 있다. S-Box의 입력 6 bit를 $X = (x_0, x_1, x_2, x_3, x_4, x_5)$라고 하고 처음과 마지막 비트를 제외한 부분을 $X_1$이라 하면 $i$번째 S-Box $S_i$의 내부함수 $F_i(X)$는 다음과 같이 분해된다.

$F_i(X) = (x_0 + 1)(x_5 + 1)G_{i1}(X_1) + (x_0 + 1)X_5 G_{i2}(X_1) + x_0(x_5 + 1)G_{i3}(X_1) + x_0 x_5 G_{i4}(X_1)$ 다시말해 입력 6비트 중 처음과 마지막 비트에 따라 하나의 6/4 S-Box는 4개의 서로 다른 4/4 S-Box 중 하나가 되는 것이다. 이 32개의 4/4 S-Box의 내부함수 $G_{ij}(X_1)$, $i = 1, 2, \cdots, 8$, $j = 1, 2, 3, 4$들에 대해서 선형함수까지의 거리 $D_{ij}$를 산출한 결과 32개의 4/4 S-Box 중 10개의 S-Box는 선형함수군까지의 거리 7을 나머지 22개의 S-Box는 거리 8을 가졌음을 알 수 있었다. 본 절에서는 앞에서 제안된 $2m/m$ S-Box를 이용하여 기존의 DES S-Box 내부함수보다 선형함수까지의 거리가 더 먼 새로운 4/4 S-Box가 존재함을 보인다. 식 (3.3)에 의해 4/2 S-Box를 구성하면 그 내부함수 $F = (f_1, f_2) : V^4 \to V^2$는 다음과 같다.

$$f_1(X) = x_1 x_2 x_3 + x_2 x_3 + x_3 x_4 + x_2$$
$$f_2(X) = x_1 x_3 x_4 + x_2 x_3 + x_4$$

이 함수 $F$는 $D_F = 6$을 만족하며 출력 XOR의 분포를 살펴보면 아래의 표와 같다.

표 3.1 제안된 4/2 S-Box의 출력 XOR 분포

| 입력 XOR | 출력 XOR의 횟수 | | | |
|---|---|---|---|---|
| | 0 0 | 0 1 | 1 0 | 1 1 |
| 0 0 0 1 | | 4 | 2 | 2 |
| 0 0 1 0 | 2 | 2 | 2 | 2 |
| 0 0 1 1 | 2 | 2 | 2 | 2 |
| 0 1 0 0 | | 2 | 4 | 2 |
| 0 1 0 1 | | 2 | 2 | 4 |
| 0 1 1 0 | 2 | 2 | 2 | 2 |
| 0 1 1 1 | 2 | 2 | 2 | 2 |
| 1 0 0 0 | 5 | 1 | 1 | 1 |
| 1 0 0 1 | 1 | 5 | 1 | 1 |
| 1 0 1 0 | 2 | 2 | 2 | 2 |
| 1 0 1 1 | 2 | 2 | 2 | 2 |
| 1 1 0 0 | 1 | 1 | 5 | 1 |
| 1 1 0 1 | 1 | 1 | 1 | 5 |
| 1 1 1 0 | 2 | 2 | 2 | 2 |
| 1 1 1 1 | 2 | 2 | 2 | 2 |

이제 4/4 S-Box의 내부함수를 위해서는 두개의 새로운 함수 $f_3(X)$와 $f_4(X)$를 부가하여 최대의 $D_F$값을 갖도록 하여야 한다. 이선 $f_1$과 $f_2$의 출력을 분석하였다. 그 결

(0,0)가 되는 4개의 입력 {0000, 1000, 0010, 1010}은 2차원 부분공간을 이룸을 알 수 있었고 나머지 3가지 출력들을 내는 각각의 4개의 입력집합들은 2차원 부분공간이 이동한 모습인 flat를 이루었다. 동일출력을 발생하는 입력들의 이러한 체계성을 유지하도록 $f_3(X) = f_1(AX+B)$, $f_4(X) = f_2(AX+B)$가 되도록 해보았으나 그 결과로 만들어진 부울함수는 선형함수와의 거리가 8을 초과하지 못함이 입증되었다. 나머지 경우에 대해 Computer search를 한 결과 선형함수군과의 거리 9가 되는 많은 경우를 찾아 볼 수 있었고 그 중 일부를 다음 표에 수록한다.

표 3.2 함수 $(f_1,f_2)$와 짝을 이뤄 거리 9를 주는 함수 $(f_3,f_4)$

| X | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $(f_3,f_4)$ | 1 | 0 | 0 | 1 | 2 | 0 | 3 | 0 | 3 | 1 | 2 | 1 | 3 | 3 | 2 | 2 |
| | 1 | 0 | 0 | 0 | 2 | 2 | 0 | 1 | 3 | 1 | 2 | 3 | 0 | 3 | 1 | 2 |
| | 1 | 0 | 0 | 0 | 0 | 3 | 3 | 1 | 3 | 1 | 2 | 3 | 2 | 2 | 1 | 2 |
| | 2 | 1 | 0 | 3 | 2 | 2 | 2 | 0 | 3 | 0 | 1 | 3 | 1 | 1 | 0 | 3 |
| | 3 | 3 | 2 | 3 | 1 | 2 | 2 | 0 | 1 | 1 | 0 | 3 | 2 | 1 | 0 | 0 |

## 5. 결 론

본 논문에서는 임의의 정의역과 치역을 갖는 (벡터) 부울함수의 선형함수군까지의 최대거리를 알아보았다. 실제로 이 최대거리가 새로이 밝혀진 경우는 벡터 부울함수의 경우 $(n,m) = (4,2),(6,2),(6,3)$의 경우에 불과하나 일반적으로 $n \geq 2m$인 경우에는 상당히 우수하다고 믿어지는 하계(lower bound)를 유도하였다. 출력의 균등분포성을 만족하는 부울함수, 즉 S-Box의 내부 변환함수의 선형함수군까지의 최대거리에 대한 하계 역시 $n \geq 2m$인 경우에 대해 유도하였고 특히 거리 성질이 우수한 $2m/m$ S-Box를 설계하였다. 또 DES에서 쓰이는 32개의 4/4 S-Box의 선형함수군까지의 거리를 산출하였고 기존의 S-Box들 보다 선형함수군까지의 거리가 더 큰 새로운 S-Box가 존재함을 보였다.

앞으로의 연구방향을 제시해 본다면 첫째 벡터 부울함수의 선형함수군까지의 최대거리에 대한 좀더 개선된 하계 및 상계를 얻기 위한 연구가 수행되어야 한다. 특히 $n < 2m$인 경우(DES에서 처럼 $n = 6$, $m = 4$ 또는 $n = 4$, $m = 4$)에 대한 연구가 계속되어야 할 것이다. 둘째, 본 논문에서 그 존재가 입증된 새로운 S-Box, 즉 거리 성질도 우수하면서 동시에 일반적으로 S-Box가 만족해야 할 여러 조건을 모두 충족시킬 수 있는 S-Box의 체계적인 설계 알고리즘의 개발 역시 앞으로의 중요한 연구 방향이다. 마지막으로 본 논문에서는 S-Box의 비선형성의 척도로서 선형함수군까지의 거리를 차용하였으나 이 척도와 실제로 Differential Cryptanalysis 등의 공격하에서의 S-Box의 performance와의 정량적인 상관관계는 입증하지 못하였다. 이러한 관점에서의 연구도 암호학 분야에서의 흥미있는 연구 과제로 생각된다.

## 참 고 문 헌


〔1〕 O. S. Rothaus, "On 'bent' functions," Journal of Combinatorial Theory, 20A, pp. 300-305, 1976.

〔2〕 K. Nyberg, "Perfect Nonlinear S-Boxes," Proceedings of Eurocrypt.

〔3〕 E. Biham and A. Shamir, "Differential Cryptanalysis of DES-like Cryptosystems," Journal of Cryptology, 4. pp. 3-72, 1991.

〔4〕 E. F. Brickell, J. H. Moore, and M. R. Prutill, "Structure in the S-Boxes of the DES," Proceedings of Crypto '86, pp. 3-8, 1987.

〔5〕 F. J. macWilliams and N. J. A. Sloane, The Theory of Error-Correcting Codes, North-Holland, 1977.

[6] T. Helleseth, "All binary 3-error correcting BCH codes of length $2^m$-1 have covering radius 5," IEEE Trans. Inform. Theory, Vol. IT-24, pp. 257-258, 1978.

[7] E. R. Berlekamp and L. R. Welch, "Weight distributions of the cosets of the (32, 6) Reed-Muller code," IEEE Trans. Inform. Theory, Vol. IT-12, pp. 203-207, 1972.

[8] J. Mykkeltveit, "The Covering radius of the (128, 8) Reed-Muller code is 56," IEEE Trans. Inform. Theory, Vol. IT-26, pp. 359-362, 1980.

[9] G. D. Cohen, M. G. Karpovsky, H. F. Mattson, Jr., and J. R. Schatz, "Covering Radius - Survey and Recent Results," IEEE Trans. Inform. Theory, Vol. IT-31, pp. 328-343, 1985.

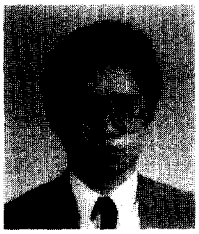
[10] N. J. Patterson and D. H. Wiedemann, "The covering radius of the 215, 16) Reed-Muller code is at least 16276," IEEE Trans. Inform. Theory. Vol. IT-29, pp. 354-356, 1983.

[11] H. Chung, "On the Distance Properties of the s-Box Internal Mappings," presented at the 1993 JW-ISC, Oct. 24-26, Seoul, Korea, 1993.

[12] L. Brown, M. Kwan, J. Pieprzyk and J. Seberry, "Improving Resistance to Differential Cryptanalysis and the Redesign of LOKI," Abstracts of Asiacrypt '91, 1991.

[13] J. F. Dillon, "Elementary Hadamard Difference Sets," Ph.D. Thesis, University of Maryland, 1974.

## □ 著者紹介



정 하 봉(鄭 夏 奉) 정회원

1958년 10월 2일생
1981년 2월 서울대학교 공과대학 전자공학과 학사
1985년 1월 미국 남가주 대학(VSC) 전기공학과 석사
1988년 7월 미국 남가주 대학(VSC) 전기공학과 박사
1988년 8월 ~ 1991년 8월 미국 뉴욕 주립대(SUNY Buffalo) 전기공학과 조교수
1991년 8월 ~ 현재 홍익대학교 공과대학 전자공학과 조교수