通信情報保護學會誌
第4卷 第2號, 1994. 6

비선형함수와 해쉬함수

조 한 혁*, 황 석 근**

요 약

함수 $f(x)$: $GF(2)^n \rightarrow GF(2)^n$은 부울함수인 n개의 좌표함수를 사용하여 $f=(f_1, f_2, \ldots, f_n)$과 같이 표현될 수 있다. 이때, 함수 $f(x)$의 암호론적 안전성은 $f(x)$의 좌표함수들의 안전성과 깊은 관계가 있다. 이제 주어진 함수의 비선형성을 부울함수들의 비선형성을 통해 살펴보고, 또한 비선형함수와 해쉬함수와의 관계에 대해 살펴본다.

1. 비선형부울함수

부울함수(Boolean function)는 본래 부울대수 위에서 정의되는데, 논리적인 연산을 대수적인 연산으로 바꿀 수 있기에 보통 유한체인 $GF(2)=\{0, 1\}$에서 정의된다. 이제 m개의 변수를 갖는 부울함수를 m비트의 벡터를 1비트의 벡터로 대응시키는 함수, 즉 $f: GF(2)^m \rightarrow GF(2)$인 함수로 이해하자. 이때, 부울함수 $f(x)$는 유한기하의 측면에서 $GF(2)^m$의 부분집합으로 이해될 수 있으며, 또한 $f(x)$의 진리표(Truth table) $v(f)$는 길이가 2^m인 벡터이기에 부울함수 $f(x)$는 길이가 2^m인 $(0, 1)$-벡터 $v(f)$로도 이해될 수 있다. 이제 m개의 변수를 갖는 모든 부울함수들 전체를 B_m이라 하면, B_m은 부울함수들의 진리표 벡터를 모아 놓은 벡터공간 $GF(2)^n$ (여기서 $n=2^m$)과 동일하게 볼 수 있다. 이제 B_m의 원소 중에 비선형성이 높은 부울함수에 대해 살펴보고, 이를 m-비트에서 k-비트로 가는 비선형 함수 $f(x)$: $GF(2)^m \rightarrow GF(2)^k$로 확장해보자.

B_m의 원소중에서 아핀함수인 부울함수들의 진리표를 모아 놓은 벡터공간을 1차 리드-뮬러코드(First-order Reed-Muller Code)라 하고, $R(1, m)$으로 표시한다. 따라서 $n=2^m$이라면 $R(1, m)$은 B_m 즉 $GF(2)^n$의 부분공간이 된다. $GF(2)^n$에 속한 두 벡터의 거리는 서로 다른 두 벡터의 성분의 개수로 주어지며 이러한 거리는 해밍거리(Hamming distance)로 알려져 있다. B_m에 속한 부울함수 $f(x)$, $g(x)$의 해밍거리 $d(f(x), g(x))$는 이것들의 진리표 벡터들의 해밍거리로 정의하며, 부울함수 $f(x)$의 비선형성거리 $\delta(f)$는 $\min\{d(f(x), l(x)) \mid l(x) \in R(1, m)\}$으로 정의된다. 이제 B_m을 $R(1, m)$으로 나눈 coset space C_m을 생각하자. 이때, C_m의 원소인 coset S의 weight를 $\min\{wt(f) \mid f(x) \in S\}$로 정의하며, 이러한 weight를 갖는 S의 원소를 S의 coset-leader라 한다. C_m의 원소 S를 잡았을 때, S에 속한 부울함수 $f(x)$, $g(x)$를 잡자. 이때 $f(x)-g(x)$는 $R(1, m)$의 적당한 원소인 $l(x)$가 된다. 이때 아핀 부울함수의 집합인 $R(1, m)$에 속한 임의의 원소 $h(x)$에 대해 $d(f(x), h(x)) = d(f(x)$

* 정회원, 서울대학교 수학교육과
** 정회원, 경북대학교 수학교육과

$-l(x)$, $h(x)-l(x)$)이기에 S에 속한 모든 부울함수의 비선형거리는 같음을 볼 수 있다. 또한 S의 coset leader $g(x)$에 대해 $wt(g(x))=d(g(x), 0)$이고 $R(1, m)$의 임의의 원소 $h(x)$에 대해 $d(f(x)-h(x), 0)=d(f(x), h(x))$이므로 다음의 정리를 얻을 수 있다.

❖ 정리 1 C_m의 원소 S를 잡았을 때, S에 속한 모든 부울함수의 비선형거리는 같다. 또한 coset S의 weight는 S에 속한 임의의 원소의 비선형거리와 같다.

이제 C_m의 원소들 중에 가장 큰 weight를 갖는 원소들은 무엇인가? 그러한 coset에 속하는 부울함수들을 극대비선형(Maximum nonlinear)함수라 하자. 또한 m이 짝수일 때의 극대비선형 부울함수를 bent function[11] 또는 완전비선형 함수(Perfect nonlinear)[8]라 하자. 극대비선형 함수는 이상적인 비선형성의 성질을 갖고 있기에, 어떤 암호학적 성질을 갖는 함수가 극대비선형 함수에 가까울수록 그 함수는 바람직하다고 말할 수 있다. 극대비선형 함수의 비선형거리는 coset들의 weight중에 가장 큰 값인 $R(1, m)$의 covering radius γ_m 값과 같음을 쉽게 증명할 수 있다. $R(1, m)$의 covering radius γ_m은 $R(1, m)$은 각각의 원소에서 반경이 r인 원을 그렸을 때 이것들이 B_m을 덮게하는 가장 작은 수이다. 일반적인 m에 대해 γ_m의 값은 알려져 있지 않지만, m이 짝수일 때 극대 비선형 부울함수 $f(x)$는 $\delta(f)=\gamma_m=2^{m-1}-2^{\frac{m}{2}-1}$를 만족시킨다. 다음은 γ_m이 만족시키는 식인데, 9이상의 홀수인 m에 대한 γ_m의 값은 아직 미해결 문제로 남아있다.

(1):n이 짝수이면 γ_n은 $2^{n-1}-2^{n/2}$이다.

(2):n이 홀수이면 $2^{n-1}-2^{(n-1)/2} \le \gamma_n \le 2^{n-1}-2^{(n-1)/2}$이다.

m-비트에서 1-비트로 가는 비선형 부울함수의 성질을 응용하여 이제 m-bit를 k-bit로 보내는 일반적인 완전비선형 함수 $f(x): GF(2)^m \to GF(2)^k$에 대해 살펴보자. 우선 $w \in GF(2)^m$에 대해 $f_w(x)$는 $f(w+x)-f(x)$으로 정의하자. 이제 모든 $w \in GF(2)^m$에 대해, $f_w(x)$가 균형함수(balanced function)일 때 $f(x)$를 일반적인 완전비선형 함수라 한다. 다음은 비선형 부울함수와 일반적인 완전비선형 함수와의 관계를 나타내는 정리이다.

❖ 정리 2 함수 $f(x): GF(2)^m \to GF(2)^k$가 완전비선형 함수일 필요충분조건은 0-벡터가 아닌 $w \in GF(2)^k$에 대해 $x \in GF(2)^m$를 $w \cdot f(x) \in GF(2)$에 대응시키는 함수가 완전비선형인 것이다.

해쉬함수(Hash function)는 임의의 길이의 비트에 일정한 k 비트 값을 대응시키는 함수로서, 해쉬함수 $h(x)$는 $GF(2)^* = GF(2) \cup GF(2)^2 \cup GF(2)^3 \cup \ldots$라고 할 때 $h(x): GF(2)^* \to GF(2)^k$과 같이 표현될 수 있다. 그런데 임의의 길이의 스트링을 일정한 길이로 축약하여 대응시키는 법칙을 만들기는 매우 어려우므로, 해쉬함수 $h(x): GF(2)^* \to GF(2)^k$는 보통 $m+k$-비트를 k-비트로 보내는 하나의 기초해쉬함수 $g(x): GF(2)^{m+k} \to GF(2)^k$를 바탕으로 만들어진다[3]. 그런데 이때의 기초해쉬함수가 완전비선형 함수라면 좋을 것이다. 함수 $f(x): GF(2)^m \to GF(2)^k$가 완전비선형 함수라면 $m \ge 2 \times k$ 임이 알려져 있으며, 특히 $f(x): GF(2)^{k+k} \to GF(2)^k$인 완전비선형을 많이 만들 수 있다[10]. 해쉬함수는 인증과 서명에 매우 중요한 암호학적 도구이다.

2. 비선형치환함수

n-비트에서 n-비트로 가는 함수 $f(x): GF(2)^n \to GF(2)^n$ 은 n개의 좌표 함수들에 의해 $f=(f_1, f_2, \ldots, f_n)$과 같이 표현될 수 있다. 그런데 함수 $f(x)$는 다항식 함수로 표현될 수 있고, 또한 $f(x)$는 $n \times 2n$ 인 (0, 1)-행렬로도 표현될 수 있다. 특히 전단사함수인 치환함수 $f(x): GF(2)^n \to GF(2)^n$에 대응하는 다항식을 치환다항식(Permutaion

polynomial)이라 하는데, 이는 n개의 좌표함수에 의해 그 성질이 결정된다. 이 절에서는 부울함수와 치환다항식을 통한 비선형함수에 대해 살펴보자.

n비트에서 n비트로 가는 치환함수 S-Box는 n개의 부울함수로 나타내어지는데, 안전한 S-Box의 제작은 이를 구성하는 부울함수의 모임에 의해 결정된다. 또한 안전한 부울함수는 안전한 S-Box에서 구할 수도 있다. 그런데 수많은 S-Boxes 중에 어떻게 안전한 S-Box를 찾아낼 수 있는가? Adams와 Tavares[3]는 컴퓨터를 사용하여 바람직한 부울함수를 조합하여 안전한 S-Boxes를 찾는 방법을 제안하였다. 그러나 이 방법은 큰 S-Box를 만드는데 엄청난 시간을 들여야 한다는 단점을 갖고 있는데, 이제 이차형의 부울함수들을 조합하여 비선형성이 있는 S-Box를 수학적으로 얻는 방법을 살펴보자. 한편 Pieprzyk와 Finkelstein[12]은 랜덤 치환(Random Permu-tation)을 먼저 잡은 후, 이것을 이루는 부울함수들의 성질을 조사하여 안전한 S-Boxes를 찾는 방법을 제안하였다. 그런데 이 방법은 어떻게 permutation을 잡느냐에 따라 효율성이 좌우되게 되며, 또한 부울함수의 검색에 필요한 시간이 많이 걸리게 되는 단점이 있다. 여기서 Random Permutation 대신에 수학적 성질을 많이 갖고 있는 permutation polynomial을 사용하여 안전한 S-Box를 찾는 방법을 살펴보자. 먼저 주어진 함수 $f(x)$와 이 함수의 $n \times 2^n$인 (0, 1)-행렬 A의 관계를 보여주는 다음의 정리를 보자. 이 정리에 의해 전단사함수를 만드는 문제는 다음과 같은 부울함수 n개를 찾는 문제와 동치가 됨을 볼 수 있다. 여기서 주어진 이진벡터에서 1의 개수를 그 벡터의 Hamming weight라 하며, 주어진 부울함수 $f(x)$의 Hamming weight $wt(f)$는 $v(f)$의 해밍 weight로 정의한다.

❖ 정리 3 $f(x)$가 전단사함수일 필요충분조건은 $f(x)$의 $n \times 2^n$ 행렬인 A에서 임의로 행벡터들을 뽑아도 이들의 일차결합이 항상 2^{n-1} Hamming weight를 갖는 것이다.

이 절에서 다루는 많은 내용은 표수(characteristic)가 p인 유한체 $GF(p^n)$에서도 성립하지만, 암호학과 관련하여 표수가 2인 유한체 $GF(2^n)$에서만 다루자. 이제 $q = 2^n$개의 원소를 갖는 유한체 $F = GF(2^n)$를 생각하자. 이때, 트레이스 함수(trace function)는 $F = GF(2^n)$의 원소 $\alpha \in F$에 대해 $\sum_{i=0}^{n-1} \alpha^{2^i}$를 대응시키는 함수이다. 그런데 이 함수의 치역은 $GF(2) = \{0, 1\}$이기에 트레이스 함수는 $Tr(x) : GF(2^n) \rightarrow GF(2)$이 된다. $GF(2^n)$의 원소 $\alpha_1, \cdots, \alpha_n$이 $GF(2)$위에서 $Tr(\alpha_i \cdot \alpha_i) = 1$이고 $Tr(\alpha_i \cdot \alpha_j) = 0$ $(i \neq j)$인 기저(basis)일 때 이들을 self-dual basis라 한다. 또한 이들이 어떤 한 원소의 2^{2^i} 승꼴로 모두 표현될 수 있을 때 이들을 normal basis라 한다. 표수(characteristic)가 2인 유한체 $GF(2^n)$은 다음과 같이 basis를 갖는다는 사실은 잘 알려져 있다.

❖ 정리 4 임의의 n에 대해 유한체 $GF(2^n)$은 self-dual basis를 갖는다. 또한 n이 홀수일 때 $GF(2^n)$은 self-dual normal basis를 갖는다.

이제 $F = GF(2^n)$에서 F로 가는 함수 전체의 집합 F^F를 생각하자. 이때, 집합 F^F는 계수를 F에서 갖고 차수가 q미만인 다항식의 집합 $F_q[x]$와 본질적으로 같다. $F_q[x]$의 원소중에 전단사함수가 되는 다항식을 우리는 치환 다항식(Permutation Polynomial)이라 하는데, 예를 들어, $F = GF(2^n)$ (단, n은 홀수)에서 $f(x) = x^3$는 치환다항식이 된다. F^F의 원소 즉 $F_q[x]$의 원소들은 다음과 같이 self-dual basis와 트레이스 함수(Trace function)에 의해 좌표함수들을 구할 수 있다.

❖ 정리 5 $\alpha_1, \ldots, \alpha_n$을 $GF(2)$위에서 $F = GF(2^n)$의 self-dual basis라 하자. 이때, 임의의 다항식 $p(x) \in F_q[x]$는 $\sum_{i=1}^{n} p_i(x) \cdot \alpha_i$과 같이 표현될 수 있다. (여

기서 $p_i(x) = Tr(p(x) \cdot \alpha_i)$.

여기서 각각의 i에 대한 $p_i(x)$는 $P(x)$의 i번째 좌표함수가 된다. 구체적으로 임의의 원소 $x \in GF(2^n)$는 $x = \sum_{i=1}^{n} x_i \cdot \alpha_i$(단, $x_i = Tr(x \cdot \alpha_i)$)과 같이 표현될 수 있다. 이제 n이 홀수일때 $GF(2^n)$위의 치환다항식 $f(x) = x^3$를 생각하면, $f(x)$는 $(\sum_{i=1}^{n} x_i \cdot \alpha_i)^3$ (여기서 $x_i = Tr(x \cdot \alpha_i)$)으로 표현될 수 있다. 여기서 $g(x)$를 $Tr(f(x))$라 하면, $g(x) = x_1x_2 \oplus x_2x_3 \oplus \cdots \oplus x_nx_1$이 된다. 이때, 여기서 얻은 부울함수 $g(x)$는 비선형성이 높은 부울함수라 할 수 있다[13]. 따라서 치환함수인 $f(x)$도 암호학적으로 의미있는 S-Box가 된다. 이렇게 permutation polynomial을 그것을 이루는 부울함수의 비선형성과 관련시켜 의미있는 S-Box를 찾을 수 있다. 이렇게 찾아진 S-Box는 해쉬함수의 제작에 사용될 수 있다[2]. 이제 이차형식 부울함수를 사용하여 치환함수를 만드는 것과 주어진 치환다항식의 성질을 좌표함수를 통해 알아보는 방법에 대해 살펴보자.

주어진 부울함수 $f: GF(2)^n \rightarrow GF(2)$과 $w \in GF(2)^n$에 대해 $f_w(x)$를 $f(w+x) - f(x)$으로 정의하자. 만일 $f_w(x)$가 상수함수가 될 때 $w \in GF(2)^n$를 $f(x)$의 아핀방향벡터(affine directional vector)라 하고, $A(f)$를 $f(x)$의 아핀방향 벡터들의 모임이라 하자. 이 때, $A(f)$의 두 원소 u와 v와 $w = u + v$에 대해, $f(u+v+x) - f(x)$는 $f(u+v+x) - f(v+x)$와 $f(v+x) - f(x)$의 차가 된다. 또한 $f(w+x) - f(x)$가 상수라면, 이 상수는 $f(w) - f(0)$이 된다. 따라서 $A(f)$의 두 원소 u와 v에 대해 $f(u+v) + f(v)$는 $f(u) - f(0)$가 되며, 따라서 $f(u+v) - f(v) - f(u)$는 상수 $f(0)$가 된다. 이제 이차형식(Quadratic form) 부울함수 $f(x)$와 이 함수의 $A(f)$에 대해 살펴보자. 이 때, $f(x)$는 크기가 n인 정방행렬 A에 의해 $f(x) = x \cdot A \cdot x^t$과 같이 표현될 수 있다. 이상의 사실을 정리하면 다음의 정리를 얻게된다[11]. 또한 이러한 사실로부터 비선형성이 있는 이차의 부울함수들을 만들고, 정리 3과 그 다음의 사실로부터 비선형성이 있는 전단사함수를 만들 수 있다.

❖ 정리 6 $A(f)$는 $GF(2)^n$의 부분공간이 된다. 또한 $A(f)$위에서 $f(x)$는 아핀함수가 된다. $f(x)$가 이차형식이고 정방행렬 A에 의해 $f(x) = x \cdot A \cdot x^t$이면 $A(f)$의 차원은 $n\text{-}rank(A + A^t)$가 된다.

❖ 정리 7 $f = (f_1, f_2, \ldots, f_n)$을 $GF(2)^n \rightarrow GF(2)^n$인 함수라 하고, f_i들을 f의 좌표함수로 이차의 부울함수들이라 하자. 이 때, f가 $N(f) = N(f^{-1}) \geq 2^{n-1}(2^{r-1} - 2^{\frac{r+2}{2}})$인 전단사함수가 될 필요충분 조건은 좌표함수 f_i들의 선형결합이 rank가 r 이상인 0/1 균형잡힌 이차부울함수인 것이다.

참 고 문 헌

[1] 황 석근, 조 한혁, 디지털 서명과 해쉬함수, 통신정보보호학회지, 2권 1호 (1992), 23-29.

[2] 조 한혁, 황 석근, S-Boxes와 해쉬함수, 통신정보보호학회지, Vol. 3 (1993), 78-85.

[3] C. Adams and S. Tavares, The structured design of cryptographically good S-Boxes, J. Cryptology 3 (1990), 27-41.

[4] I. B. Damgard. A design principle for hash function, Proc. of CRYPTO' 89 (1990), 416-427.

[5] R. Forre, The strict avalanche criterion: spectral properties of Boolean functions and an extended definition, in Advances in Cryptology: Proc. of CRYTPTO' 88 (1989), 450-468.

[6] R. C. Merkle, A fast software one-way hash function, J. Cryptology 3 (1990), 43-58.

[7] R. C. Merkle, one way hash functions and DES, proc. of CRYPTO' 89 (1990), 428-445.

[8] W. Meier and 0. Staffelbach, Nonlinear criteria for cryptographic functions, Proc. of EUROCRYPT' 89, 549-562.

[9] K. Nyberg, Constructions of bent functions and difference sets, Proc. of EUROCRYPT' 90 (1990), 151-160.

[10] K. Nyberg, Perfect nonlinesr S-Boxes, Proc. of EUROCRYPT' 91 (1991), 378-386.

[11] K. Nyberg, On the construction of highly nonlinear permutations, Proc. of EUROCRYPT' 92 (1992), 92-98.

[12] J. Pieprzyk and G. Finkelstein, Toward effective nonlinear cryptosystem design, IEE Proc. 135 (1988), 325-335.

[13] J. Pieprzyk, Bent Permutation, In Finite fields, coding theory, and advances in communications and computing (1992), 173-181.

[14] O. S. Rothaus, On "Bent" functions, J. Combi. Theory (A) 20 (1976), 300-305

[15] A. F. Webster and S. E. Tavares, On the design of S-boxes, Proc. of CRYPTO' 85 (1986), 523-534.

□ 著者紹介

조 한 혁(正會員)

1979年 2月 서울大學校 師範大學 數學敎育과 卒業(理學士)
1981年 2月 서울大學校 大學院 數學科 卒業(理學碩士)
1988年 8月 Univ. of Wisconsin-Madison 大學院 數學科 卒業(Ph. D.)
1988年 9月 ~ 1989年 2月 Bowling Green State Univ. 專任講師
1989년 3月 ~ 현재 서울大學校 副敎授

황 석 근(正會員)

1972年 2月 慶北大學校 師範大學 數學敎育과 卒業(理學士)
1977年 2月 慶北大學校 大學院 數學科 卒業(理學碩士)
1985年 8月 Univ. of Wisconsin-Madison 大學院 數學科 卒業(Ph. D.)
1979年 4月 ~ 1990年 2月 慶北大學校 專任講師-助敎授
1990년 3月 ~ 1991年 3月 成均館大學校 副敎授
1991年 3月 ~ 現在 慶北大學校 副敎授