비선형 로직의 통계적 검정

성 둘 옥† · 신 상 욱†† · 이 경 현††

요 약

최대 주기를 갖는 n차 선형 쉬프트 레지스트(m-LFSR)를 비선형 논리로 결합하여 키 출력 수열을 발생시키는 이진 난수 발생기의 출력 수열과 입력 m-LFSR 출력 수열과의 입·출력 상관관계를 이용한 통계적 검정법을 제안한다. 제안된 검정법은 비선형 함수의 출력과 입력 변수들 간의 상호 정보량으로 분할표의 동질성을 이용하며 χ^2-검정을 수행한다. 또한 제안된 검정법을 대표적인 몇몇 비선형 암호 시스템에 적용한 컴퓨터 시뮬레이션 결과를 기술함으로서 제안된 검정법이 강력한 알고리즘 설계 기준의 평가 척도로써 이용될 수 있음을 보인다.

A Statistical Test for the Nonlinear Combiner Logic

Dul-ok Sung† · Sang-uk Shin†† · Kyung-hyune Rhee††

ABSTRACT

We propose a statistical test for the nonlinear combiner logics which are usually combined with two maximal Linear Feedback Shift Registers and generate pseudorandom bit sequences. This test uses the mutual information between the output and a set of inputs which will be a random variable and its distribution is obeyed to an approximate χ^2-distribution. We adopt this statistic to a χ^2-test of independence by using contingency table. We also apply a proposed test to some non-linear cryptosystems and show that this test is useful to evaluate the strength of the cryptosystems.

1. 서 론

스트림 암호 시스템은 비트나 문자와 같은 데이타 단위를 시간에 따라 변화를 주어 암호화하는 방식으로 스트림 암호 시스템의 대부분은 이진 난수 발생기(Pseudo-Random Bit Generator)로 구성된다. 이러한 발생기들은 짧은 키를 입력해서 긴 의사 난수열을 만들어 평문과 비트별로 XOR 함으로써 암호문을 만들어 낸다. 이상적인 이진 난수 발생기는 Coin-tossing 수열의 실현으로 볼 수 있다. 한편 대부분 이진 난수 발생기의 설계는 한개 또는 그 이상의 최대 주기를 갖는 선형 쉬프트 레지스터(m-LFSR : maximal Linear Feedback Shift Register)를 비선형 논리(Nonlinear Logic)로 결합하여 출력 수열을 발생시키는 방법을 택하고 있다[7,9,10]. 이때 결합함수 설계시 각각의 입력 LFSR 수열이 출력 수열로 누수되는 정보가 최소로 되어야 하며, 출력 수열과 어떤 LFSR 수열이 서로 상관관계 (correlation)가 있는 경우 이 LFSR에 대한 상관관계 공격이 분할 통합법(divide and conquer)으로 가능함이 잘 알려져 있다[7,11,12,18]. 한편, 스트림 암호 시스템의 이진 난수 발생기 설계원칙에 따

† 준 회 원:부산수산대학교 전산정보학과
†† 준 회 원:부산수산대학교 전자계산학과
논문접수:1995년 11월 23일, 심사완료:1995년 12월 26일

된 암호학적 안전성 평가 척도로써 여러가지 방안을 고려할 수 있으나, 최근까지 주로 사용되는 비도 요인 평가 방법은 정량화가 가능한 이진 수열의 랜덤 특성에 대한 통계적 검정법들이다[3,8].

본 논문에서는 위에서 언급한 입·출력 상관관계를 정량화하여 출력 수열이 입력수열에 얼마만큼 의존하여 정보가 누수되는지를 검정하는 통계량을 통계분석 관점에서 분할표의 동질성을 이용하여 제시하며, 이 통계량이 근사적으로 χ^2-분포를 따른다는 사실을 사용한다. 제시된 통계적 검정법은 기존에 블럭 암호 시스템의 주종을 이루는 substitution-permutation 네트워크 암호 시스템에 있어서 S-Box의 입·출력 사이의 의존도를 검정하는 엔트로피 관점의 입·출력 의존도 검정법[4]에 비해 보다 낳은 정량적 평가 통계량을 산출할 수 있어 실제적인 상관관계 검정법에 유용하게 적용할 수 있다.

2. 통계적 검정

비선형 암호 알고리즘의 모델에서 함수의 각 변수들은 서로 통계적으로 독립이고 균일분포인 확률변수라고 가정한다.

따라서 이러한 비선형 함수의 출력과 비선형 함수의 입력 변수들 간의 상관관계의 척도인 상호 정보량(mutual information)을 계산할 수 있다. 즉, 이러한 상호 정보량은 출력 비트당 입력 변수의 최대 엔트로피 손실 정도를 나타낸다. 본 논문에서는 비선형 결합 함수로써 2변수인 경우만 고려한다. 즉, $w=F(u, v)$, $u \in Z_m$ & $v \in Z_n$, $w \in Z_N$ 이다.

여기서 $m=2^k-1$ (k : m-LFSR1의 차수), $n=2^l-1$ (l : m-LFSR2의 차수), $N=2$를 가정한다.

입력 변수 u와 v에 대한 각 확률변수 U, V는 서로 독립이고 균일분포를 따른다고 하자. 이때 U와 출력 w에 대한 확률변수 W간의 상호정보량(Mutual Information) I(U;W)는 다음과 같이 주어진다.

$$I(U;W)=H(U)-H(U|W).$$

단, H는 엔트로피 함수로써

$H(U)=-\sum_{i=1}^{m} P(a_i)\log P(a_i)$, $P(a_i)=P(U=a_i)$ 이다.

만약 F의 함수값이 Z_N에서 임의의 값을 갖는다면 상호 정보량 I(U:W)는 하나의 확률변수가 되고 I(U;W)는 다음 결과를 만족한다.

〈정리 2.1〉
충분히 큰 n에 대해 확률변수 $2mn-I(U;W)$는 근사적으로 자유도 $(m-1)(N-1)$을 갖는 χ^2-분포를 따른다.

(증명) mn개의 표본이 두 변수 U, W에 대해 다음 표와 같이 분류되었다고 가정하자.

〈표 1〉 분할표
〈Table 1〉 Contingency table

변수	U	W	Total
1	X_{11}	X_{12}	n
2	X_{21}	X_{22}	n
·	·	·	·
·	·	·	·
m	X_{m1}	X_{m2}	n
Toatal	$X_{\cdot 1}$	$X_{\cdot 2}$	mn

X_{ij}를 입력 u=i에 대해 출력이 j가 되는 v의 빈도수라 하고, $\overline{X}_j$을 평균 빈도수라 하자.

즉, $X_{ij}=\#\{v; F(i,v)=j\}$, $\overline{X}_j=\frac{1}{m}\sum_{i=1}^{m} X_{ij}$

이때 상호 정보량은 다음과 같이 얻어진다.

$$\begin{aligned} I(U:W) &= H(W)-H(W|U) \\ &= -\sum_{j=1}^{N} P(W=j)\log P(W=j) \\ &\quad +\sum_{i=1}^{m} P(U=i)\sum_{j=1}^{N} P(W=j|U=i)\log P(W=j|U=i) \\ &= -\sum_{j=1}^{N}\frac{1}{m}\sum_{i=1}^{m}\frac{X_{ij}}{n}\log\frac{\sum_{i=1}^{m} X_{ij}}{mn} \\ &\quad +\frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{N}\frac{X_{ij}}{n}\log\frac{X_{ij}}{n} \\ &= \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{N} X_{ij}\log\frac{X_{ij}}{\overline{X}_j} \end{aligned}$$

$X\log\frac{X}{\overline{X}_j}$ 를 $\overline{X}_j$ 근방에서 Taylor 멱급수 전개한 후 3

항 이상을 무시하고 $X=X_{ij}$를 대입하면,

$$X\log\frac{X}{\bar{X}_j}\approx(X_{ij}-\bar{X}_j)+\frac{1}{2\bar{X}_j}(X_{ij}-\bar{X}_j)^2$$

따라서

$$I(U:W)\approx\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{N}(X_{ij}-\bar{X}_j)+\frac{1}{2mn}\sum_{i=1}^{m}\sum_{j=1}^{N}\frac{(X_{ij}-\bar{X}_j)^2}{\bar{X}_j}$$

한편, 위식의 첫째 부분은 $\frac{1}{n}\sum_{j=1}^{N}(\frac{1}{m}\sum_{i=1}^{m}X_{ij}-\bar{X}_j)$로 고쳐 쓸 수 있고 $\bar{X}_j$의 정의에 의해 0 이 된다.

결국,

$$I(U:W)=\frac{1}{2mn}\sum_{i=1}^{m}\sum_{j=1}^{N}\frac{(X_{ij}-\bar{X}_j)^2}{\bar{X}_j}$$

이 된다. 마지막에 구해진 수식은 분할표의 동질성 검정에 의해 근사적으로 자유도가 (m−1)(N−1)인 χ^2-분포를 따른다.

〈따름정리 2.2〉

확률변수 I(U;W)의 기대값 E[I(U;W)]은 근사적으로

$$\frac{(m-1)(N-1)}{2mn}$$

을 갖는다.

(증명) 자유도 n인 χ^2-분포의 기대값은 n이다.

3. 컴퓨터 시뮬레이션 및 결과

2장에서 정의된 통계량 I(U;W)에 대해 귀무가설을 두 개의 변수(입력과 출력변수 상호간)가 서로 독립이라고 가정하고 통계량을 구한 후 5%나 1% 유의수준에서 검정을 수행한다. 적용된 비선형 암호 시스템은 다음과 같다.

3.1 비선형 암호 시스템

먼저 시뮬레이션에 적용된 비선형 암호 시스템을 간략히 소개한다.

3.1.1 J-K 플립플롭

J-K 플립플롭[2]은 m-LFSR 2개의 출력을 J-K 플립플롭에 의해 조합하여 출력수열을 발생하는 시스템이다.

J-K플립플롭을 수식화하면

$q_n=a_n\oplus q_{n-1}(1\oplus a_n\oplus b_n)$ (단, $q_{-1}=0$)이다.

여기서(a_n): m-LFSR1의 출력수열, (b_n): m-LFSR2의 출력수열이다.

J-K플립플롭:

Input: parameters: 2 LFSRs $< L_j, C_j(D) >$,
key: initial states $a_0^{(1)}$, $a_0^{(2)}$ of the 2 LFSRs.
For i = 0, 1, 2, ... do
a. Shift each LFSR
b. Compute kth J-K flip-flop function for corresponding pair of LFSRs
$y_i=a_i^{(1)}\oplus y_{i-1}(1\oplus a_i^{(1)}\oplus a_i^{(2)})$
c. Collect four keystream bits as $z_{i+k}=y_i^{(k)}$
Output: the sequence of z_i, $i=1, 2, ...$

3.1.2 SUMMATION Generator

Summation Generator[2]는 입력 수열은 2진수이며, SUM 함수 $f:Z_N\rightarrow Z$, $Z=\sum_{i=1}^{N}x_i$로 정의하며 최하위 비트에서부터 비트 연산을 한다. 정수 덧셈은 F_2상에서 높은 비선형성을 가지며, 상관관계 공격에 대한 강인한 면역성을 갖는다는 사실에 근간을 두고 연구하였다.

SUMMATION Generator:
Input: parameters: 2 LFSR $< L_j, C_j(D) >$
Key: initial states of the N LFSRs and carry C_0
For i = 1, 2, ... do
1. Step each shift register once to produce
$x_{1i}, x_{2i}, ..., x_{Ni}$.
2. Compute the integer sum
$S_i=\sum_{k=1}^{N}x_{ki}+C_{i-1}$

3. Set

$z_i = S_i \bmod 2$

$C_i = [\frac{S_i}{2}]$

Output : the sequence z_i, $i = 1, 2, ...$

3.1.3 MUX 시스템

MUX 시스템[2]은 멀티플렉서와 2개의 m-LFSR로 구성된다. m-LFSR1과 m-LFSR2의 차수가 각각 m, n이고 m-LFSR1의 단이 $A_0, A_1, ..., A_{m-1}$, m-LFSR2의 단이 $B_0, B_1, ..., B_{n-1}$, 그리고 m-LFSR1의 출력 수열을 (a_n), m-LFSR2의 출력 수열을 (b_n)이라고 하자. 먼저 $1 < h < m$인 정수 h를 선택한 후, m-LFSR2의 n단 중에서 2^h단을 선택한다. 시간이 t 일때 MUX 시스템의 출력 u_t는 m-LFSR1의 h 단의 내용에 의해 m-LFSR2의 2^h단 중 한 단의 내용으로 결정된다.

Mutiplexer generator :

Input : parameter : 2 LFSRs $< L_j, C_j(D) >$,
h and control vector $j = (j_0, j_1, ..., j_{h-1})$ such that $o \le j_0 \le j_1 < ... < j_{h-1} \le L_1$.
key : initial states $s_0^{(1)}$, $s_0^{(2)}$ of the 2 LFSRs.

For i = 1, 2, ... do

1. Shift $LFSR_1$, $LFSR_2$
2. Compute the integer
$a_t = \sum_{k=0}^{k-1} 2^k s_i^{(1)}(j_k)$
3. Extract
$z_i = s_i^{(2)}(\theta(a_i))$ (θ는 {0, 1, ..., 2^h-1}에서 {0, 1, ..., L_2-1}로의 대응 함수)

Output : the sequence of z_i, i = 1, 2, ...

3.1.4 BRM 시스템

BRM 시스템[2]은 2개의 m-LFSR과 BRM(Binary Rated Multiplexer)으로 구성된다.

Input : parameter : 2 LFSRs $< L_j, C_j(D) >$,
k and control vector $j = (j_0, j_1, ..., j_{k-1})$ such that $o \le j_0 \le j_1 < ... < j_{h-1} \le L_1$.
key : initial states $s_0^{(1)}$, $s_0^{(2)}$ of the 2 LFSRs.

For i = 1, 2, ... do

1. Shift $LFSR_1$, $LFSR_2$
2. Compute the integer
$a_t = \sum_{l=0}^{k-1} 2^l s_i^{(1)}(j_l)$
3. For j = 0, 1, ... a_t do
Shift LFSR2

Output : the sequence of $s_i^{(2)}$, i = 0, 1, 2, ...

3.2 결 과

BRM의 경우 m-LFSR1에서 뽑아내는 단의 수 k는 N이 5, 6, 7인 경우 k = 4이고, N이 9, 10, 11인 경우는 k = 7이다. MUX의 경우 m-LFSF2의 차수 M이 5, 6, 7인 경우 h = 2, 그렇지 않으면 h = 3이다.

h = 2인 경우의 대응함수는

M = 5일때 θ (0, 1, 2, 3) → (1, 0, 4, 2),
M = 6일때 θ (0,1,2,3) → (3,2,0,5),
M = 7일때 θ (0,1,2,3) → (6,4,3,1),

h = 3인 경우의 대응함수는

M = 9일때 θ (0, 1, 2, 3, 4, 5, 6, 7) → (5, 7, 6, 1, 0, 3, 2, 4),
M = 10일때 θ (0, 1, 2, 3, 4, 5, 6, 7) → (5, 3, 8, 1, 0, 7, 9, 4),
M = 11일때 θ (0, 1, 2, 3, 4, 5, 6, 7) → (10, 5, 8, 7, 0, 6, 2, 1)

이다.

χ^2-검정에서 자유도 n이 30이상인 경우 보통 중심극한정리에 의해 표준정규분포표를 이용할 수 있으나 좀더 근사적인 검정통계량으로 Fisher[19]에 의한

$$\sqrt{2*I(U;W)} - \sqrt{2*(n-1)(N-1)-1}$$

를 통계량 값으로 사용하였고 이 결과를 〈표 3-1〉에 나타내었다.

컴퓨터 시뮬레이션 결과 일반적으로 BRM과 MUX는 Summation Generator나 J-K 플립플롭보다 통계적 랜덤 특성이 우수한 비선형 암호 알고리즘으로 알려져 있으나 〈표 2〉에서 알 수 있듯이 대부분의 입·출력 상관관계에서의 독립성을 기각하므로(유의수준 5% 또는 1%) 본 논문에서 제시된 통계적 검정법이 m-LFSR을 이용한 비선형 암호 시스템 설계 및 분석시 비도 평가 요인으로 사용할 수 있슴을 알 수 있다.

한편 3변수 이상의 상호 정보량에 대한 통계량 산출 및 이에 대한 검정법 연구가 앞으로 추진되어야 할 연구과제 중의 하나이다.

〈표 2〉 검정 통계량 결과
〈Table 2〉 Resalts of statistical test

N	M	BRM	MUX	SUM	J-K
5	5	−1.46202	−4.00674	2.12924	−3.84709
	6	−1.59632	−5.4462	11.54360	−0.04816
	7	−1.10360	−6.30431	22.60290	3.91123
	9	−6.98028	−6.98028	57.72990	15.31170
	10	−6.67777	−7.39547	85.99350	24.82990
	11	−6.91735	−7.39557	125.64200	38.44210
6	5	−0.47053	−1.61739	0.06182	−3.21434
	6	0.51023	−5.30255	−3.98463	−5.56617
	7	7.01255	−9.05062	31.04720	7.12281
	9	−5.01750	3.71502	−9.18362	−10.27690
	10	20.87490	50.12910	120.30300	40.63080
	11	52.84650	−10.6499	175.99700	62.26730
7	5	0.10237	−1.69272	−0.08644	−5.26835
	6	−1.14252	−9.21944	13.02140	−0.29631
	7	2.35949	−3.77503	−8.46703	−9.99472
	9	2.48916	0.41376	78.24280	17.30080
	10	−13.94450	−14.63060	164.16600	49.68180
	11	−13.55190	−14.55360	240.55700	76.75840
9	5	−3.30028	−4.021630	9.51589	−7.67306
	6	0.87528	−18.76170	30.04890	−4.20503
	7	0.49363	−14.35130	59.73900	5.85942
	9	−10.06060	−17.39580	108.92500	46.68780
	10	−3.77740	−12.83550	191.21500	39.74300
	11	−6.76153	−28.45220	438.60600	140.05500
10	5	−3.24132	−16.39810	3.83060	−14.32570
	6	0.79309	−25.76640	37.86160	−5.27481
	7	0.04485	−18.47030	91.39340	12.94890
	9	22.57830	57.78870	256.56700	67.45290
	10	22.45110	112.83000	325.83700	87.92350
	11	0.04456	−15.04400	−16.39970	−20.47730
11	5	−2.33174	−17.11868	3.20204	−17.52320
	6	−1.83953	31.97650	32.26240	−15.67040
	7	−2.66712	−27.12290	68.52490	−5.76628
	9	3.74604	−15.10380	235.49900	44.96600
	10	−20.34190	−30.01740	493.35400	137.85000
	11	−0.17287	−14.59190	404.90900	74.14420

참 고 문 헌

[1] M.H.Dawson and S.E.Tavares, "An expanded set of s-box design criteria based on information theory and its relation to differential like attack," Advances in Cryptology, Proc. of EUROCRYPT-'91, Springer-Verlag, pp.352-367, 1992.

[2] Gustavus J.Simmons, Contemporary Cryptology, the Science of Information Integrity, IEEE Press, New York, 1992.

[3] G.Xiao, J.L.M.assey, "A spectral characterization of correlation immune combining function," IEEE Trans. Information Theory, Vol. 34, No. 3, pp.569-571, 1988.

[4] H.Beker & F.Piper, Cipher System:The Protection of Communications. John Wiley & sons, 1982.

[5] Lennart Bryniekesson, "The information leakage through a random generated function," Proc. of CRYPTO'91. pp.552-553, Springer-Verlag, 1992.

[6] R.A.Rueppel, "Correlation immunity and the summation combiner," Proc. of CRYPTO'85. pp. 260-272, Springer-Verlag, 1986.

[7] R.A.Rueppel, Analysis and Design of Stream Ciphers, Springer-Verlag, 1986.

[8] R.Forre, "A fast correlation attack on nonlinearly feedforward filtered shift register sequences," Proc. of EUROCRYPT'89, pp.586-595, Springer-Verlag, 1990.

[9] S.Lloyd, "Properties of binary function," Proc. of EUROCRYPT'90, pp.124-139, Springer-Verlag, 1991.

[10] S.W.Golomb, "Shift Register Sequences," San Francisco, LA, Holden Day, 1967.

[11] T.Siegenthaler, "Correlation immunity of nonlinear combining functions for cryptographic applications," IEEE Trans. Information Theory, Vol. IT-30, pp.776-780, 1984.

[12] T.Siegenthaler, "Decrypting a class of stream ciphers using ciphertext only," IEEE Trans. Computer, Vol. C-34, pp.81-85, 1985.

[13] T.Siegenthaler, "Cryptanalysts representation of

nonlearly filtered ML-sequences," Proc. of EUROCRYPT'85, pp.103-110, Springer-Verlag, 1986.

[14] W.Blaser, P.Heinzmann, "New cryptographic device with high security using public key distribution," Proceedings of IEEE student paper contests, pp.145-153, 1982.

[15] W.Meier, O.Staffelbach, "Fast correlation attacks on certain stream ciphers," Journal of Cryptology, Vol 1, No. 3, pp.159-176, 1989.

[16] W.Meier, O.Staffelbach, "Nonlinearity criteria for cryptographic function," Proc. of EUROCRYPT-'89, pp.549-562, Springer-Verlag, 1990.

[17] W.Meier, O.Staffelbach, "Correlation properties of combiner with memory in stream ciphers," Journal of Cryptology, to appear.

[18] V.Chepyzhov, B.Smeets, "On a fast correlation attack on certain stream ciphers," Proc. of EUROCRYPT'91, pp.176-185, Springer-Verlag, 1992.

[19] V.G, Rohatgi, An Introdution to probability theory and methematical statistics, John wiley & Sons, 1992.

[20] Y.Y.Xian, "On the correlation-immunity of boolean function," Proc. of IEEE Singapore ICCS'88.

[21] 성 수학, "CORRELATION 공격," 통신정보보호학회지 제1권, 제3호, 1991. 12.

[22] 한국전자통신연구소, 현대암호학, 1991.

[23] 성둘옥, 신상욱, 이경현, "비선형 로직의 정보누수도에 대한 통계적 검정," 정보처리학회 춘계학술발표회, 명지 대학교, pp.184-187, 1995. 5.

이 경 현

1982. 경북대학교 사법대학 수학과 졸업(이학사)
1985. 한국과학기술원 응용수학과 졸업(이학석사)
1992. 한국과학기술원 수학과 졸업(이학박사)
1985~1993. 2 한국전자통신연구소 선임연구원
1993. 2~현재 부산수산대학교 전자계산학과 조교수
관심분야: 네트워크 성능분석, 암호 이론, 암호 알고리즘 설계

성 둘 옥

1992. 2 부산공업대학 전산학과 졸업(공학사)
1995. 8 부산수산대학교 전산정보학과 졸업(이학석사)
관심분야: 암호 이론 및 암호 알고리즘 설계

신 상 욱

1995. 2 부산수산대학교 전자계산학과 졸업(이학사)
1995. 3~현재 부산수산대학교 전자계산학과 석사과정
관심분야: 네트워크 성능분석, 암호 이론, 암호 알고리즘 설계