Golomb의 공리를 이용한 Shrinking Generator의 분석

Cryptanalysis of Shrinking Generator by Golomb's Randomness Postulate

김 정 헌* 권 기 호* 박 명 진**
Kim, Jeong-Heon Kwon, Ki-Ho Park, Myung Jin

ABSTRACT

The shrinking generator is simple and scaleable, and known that has good security properties. The bits of one output(R_1) are used to determine whether the corresponding bits of the second output will be used as part of the overall keystream. Two LFSRs consisting the generator generate pseudorandom sequences satisfying Golomb's postulates. We used this property to analyze the stream of LFSR R_1 of the generator.

주요기술용어 : LFSR(선형궤환생성기), 키수열(key stream), 연결다항식(connecting polynomial), 런(run), 블록(block), 갭(gap), 의사난수열(pseudorandom sequences)

1. 서 론

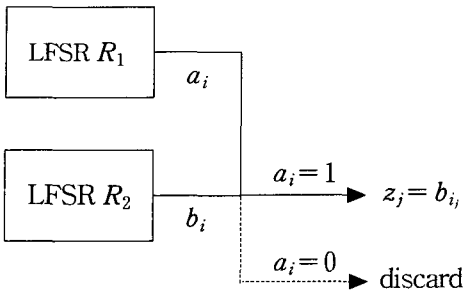
Shrinking generator는 Coppersmith[CKM] 등에 의해 1993년 제안된 키수열 생성기로 구성의 단순함에도 불구하고 높은 안정성을 가지며 처리 속도가 매우 빠른 것으로 알려져 있다.

Shrinking Generator는 두 개의 LFSR R_1, R_2로 구성되어 있다. R_1의 출력수열은 해당 위치의 R_2 출력수열을 키수열의 일부로 포함시킬지의 여부를 결정하는 선택생성기의 역할을 한다. LFSR R_1, R_2의 출력 수열을 각각 $\{a_i\}$, $\{b_i\}$, 최종 키수열을 $\{z_j\}$라고 하면, 키수열 $\{z_j\}$는 $z_j = b_{i_j}$에 의해 결정된다. 여기서 b_{i_j}는 키수열의 j번째 비트로 a_i가 '1'인 경우 이에 대응되는 b_i의 값을 나타낸다.

즉, shrinking generator의 키수열은 LFSR R_1의 i번째 비트 a_i가 '0'이면 대응하는 R_2의 i번째 비트 b_i를 버리고, a_i가 '1'이면 b_i를 출력으로 택한다.

R_1, R_2를 각각 최대주기 L_1, L_2를 갖는 LFSR로 L_1, L_2가 $\gcd(L_1, L_2) = 1$을 만족하면 LFSR R_1, R_2로 구성된 shrinking generator의 키수열 $\{z_j\}$의 주

* 육군사관학교 수학과 교수
** 육군사관학교 물리학과 교수



[그림] Shrinking Generator

기는

$$(2^{L_2}-1)2^{L_1-1}$$

이고, 선형복잡도 $L(z)$는

$$L_2 \cdot 2^{L_1-2} < L(z) \le L_2 \cdot 2^{L_1-1}$$

을 만족한다. 또한 R_1, R_2의 연결다항식이 임의로 선택되면 $\{z_i\}$의 분포 패턴은 균등분포에 가깝다[MOV].

Shrinking generator의 LFSR R_1을 선택생성기라고 하자. 선택생성기의 출력수열 $\{a_i\}$와 키수열 $\{z_j\}$가 알려졌다면 R_2의 출력 수열에 대한 분석은 자연스럽게 행해질 수 있다.

이 논문에서는 shrinking generator의 분석을 위해 R_2의 수열 $\{b_i\}$가 알려져 있을 경우 선택생성기 R_1의 출력수열 $\{a_i\}$를 복원하기 위한 하나의 방법을 제시한다.

2. Shrinking generator의 분석

S. W. Golomb은 어떤 수열이 의사난수열(pseudorandom sequence)이 되기 위한 세 가지 조건을 제시하였다. 의사난수열이란 수열의 일부를 알고 있을 때 이로부터 다음에 오는 수열의 비트를 예측하기 어려운 수열을 말한다.

[의사난수열에 관한 공리]

G1. 주기 안에서 '1'과 '0'들의 개수의 차가 적어야 한다.

G2. 주기 안에서 총 런(run)의 수를 r, 길이 l인 런의 수를 $r(l)$이라고 하면

$$r(l)=2^{-l}r$$

이다.

G3. 주기를 p라고 할 때 자기상관도

$$C_P(\tau)=\frac{1}{p}\sum_{i=1}^{p}x_i x_{i+\tau}$$

는 상수이다($\tau \neq 0$ 또는 p).

LFSR에 의해 생성되는 최대주기의 출력수열은 Golomb의 의사난수열에 관한 공리를 만족한다[Gol]. 특히 주기 안에서 '1'과 '0'의 수를 각각 $n(1)$, $n(0)$이라고 하면

$$n(1)-n(0)=1$$

이고, $\tau \neq p$이면 자기상관도는

$$C(\tau)=-\frac{1}{p}$$

로 동일한 값을 갖는다.

Golomb의 공리로부터 다음의 사실을 기술할 수 있다.

1. 선택생성기에 의한 출력수열은 '0'과 '1'의 구성비가 거의 동일하기 때문에 n비트의 키수열을 얻기 위

해서는 확률적으로 LFSR R_2의 수열 $2n$ 비트가 필요하다.

2. *최대주기 $p=2^r-1$ 인 LFSR의 출력 수열은 $(p+1)/2$ 개의 런을 가지며 이들 가운데 1/2는 크기가 1인 런이고, 1/4는 크기가 2인 런이며, 1/8은 크기가 3(...)인 런이다.*

예를 들어 $p=15$인 LFSR의 출력수열은 반드시 8개의 런을 가지며 이들 가운데 길이가 1인 것이 4개, 길이가 2인 것이 2개, 길이가 3, 4인 것이 각각 1개씩이다. 만일 이 수열이 Golomb의 공리를 정확히 만족한다면 이 수열은 길이가 3인 런 '0, 0, 0', 길이가 4인 런 '1, 1, 1, 1'을 반드시 포함하고 있어야 한다. 즉, 선택생성기 R_1에 의한 출력수열의 주기가 $p=15$라면 선택생성기의 출력에 대응하는 LFSR R_2의 수열 15비트 가운데 3비트는 연속하여 버려지고, 4비트는 연속하여 선택되어 키수열에 나타나야 한다.

이는 키수열 $\{z_j\}$에서 LFSR R_2의 출력수력 $\{b_i\}$와 연속적으로 일치하는 4개의 부호를 찾을 수 있고, 이에 대응하는 선택생성기의 부호가 '1, 1, 1, 1'임을 의미한다. 일반적으로 주기 $p=2^k-1$ 인 LFSR의 출력수열은 길이 k인 블록(block, '1'로 구성된 런)과 길이 $k-1$인 갭(gap, '0'로 구성된 런)을 포함한다.

이 논문에서 언급되는 LFSR의 출력수열은 최대주기를 갖는 것으로 가정한다. 선택생성기에 의한 출력수열의 주기를 알고 있을 경우 다음의 분석방법은 다음 절의 예가 보여 주는 것처럼 매우 효과적인 분석방법이다.

[분석 알고리즘]

1. *선택생성기의 주기 $p=2^k-1$ 에 해당하는 길이의 LFSR R_2의 출력수열 $\{b_i\}$와 키수열$\{z_j\}$에서 길이 $p/2$ 인 부분수열을 선택하여 연속적으로 일치하는 k개의 부호를 찾아 선택생성기에 의한 수열의 해당부분의 부호를 '1'로 한다.*
2. *LFSR R_2의 출력수열 $\{b_i\}$에서 연속적으로 버려진 $k-1$개의 부호를 찾고 주변의 부호를 고려하여 선택생성기에 의한 수열의 해당부분의 부호를 '0'으로 할 수 있는 가능성을 조사한다. 연속된 '0'의 정확한 위치는 결정하기 어려울 수도 있다.*
3. *Golomb의 의사난수열에 관한 공리를 전반적으로 적용하여 선택생성기에 의한 수열을 추정한다.*

선택생성기에 의한 출력수열의 주기가 알려져 있지 않거나 주기가 매우 큰 경우에는 위의 분석방법은 수정되어야 한다. 이 경우에는 키수열의 길이를 임의로 선택하고(7, 15, 31, ···) 이에 해당하는 길이(15, 31, 63, ···)의 LFSR R_2 수열과 비교하여 선택생성기에 의한 출력수열을 추정한다. 임의로 선택된 길이의 LFSR 수열이 Golomb의 공리를 만족하지 않을 수 있으므로 키수열 $\{z_j\}$와 R_2의 수열 $\{b_i\}$가 일치하는 수는 가변적이다. 따라서 여러 개의 분할된 부분수열을 조사하는 것이 효율성을 높일 수 있는 방법이다. 또한 주기가 긴 LFSR수열은 길이가 긴 블록과 갭을 포함한다. 이는 이 논문에서 제시하는 분석방법에 대한 shrinking generator가 갖는 약점이라고 할 수 있다.

3. 분석의 예

Shrinking generator를 구성하는 LFSR이

$R_1 = \langle 4,\ 1+D^3+D^4 \rangle$,

$R_2 = \langle 5,\ 1+D+D^3+D^4+D^5 \rangle$

이고, 초기상태가 각각

$[1,0,1,1]$, $[0,1,0,0,1]$

로 주어졌다고 하면, LFSR R_1, R_2는 각각 주기 $p_1=15$, $p_2=31$ 인 수열

$$\begin{aligned}(3.1)\quad \{a_i\} &= 1,1,0,1,0,1,1,1,1,\\ &\quad 0,0,0,1,0,0,\cdots\\ \{b_i\} &= 1,0,0,1,0,1,0,1,1,\\ &\quad 0,0,0,0,1,1,1,0,0,\\ &\quad 1,1,0,1,1,1,1,1,0,\\ &\quad 1,0,0,0,\cdots\end{aligned}$$

를 출력한다. R_1를 선택생성기로 하면 키수열 $\{z_j\}$는

$$\begin{aligned}(3.2)\quad \{z_j\} &= 1,0,1,1,0,1,1,0,1,\\ &\quad 0,1,0,1,1,1,1,0,\cdots\end{aligned}$$

이다.

이제 (3.1)의 LFSR R_2수열 $\{b_i\}$와 키수열 $\{z_j\}$만을 알고 있다고 가정하고 Golomb의 의사난수열에 관한 공리를 이용하여 선택생성기 R_1의 출력 $\{a_i\}$을 복구하기로 하자.

분석을 위해 알려진 정보 $\{b_i\}$, $\{z_j\}$로부터 선택생성기 수열 $\{a_i\}$의 최초 15항을 복원하기 위해 확률적으로 이에 해당하는 $\{b_i\}$의 최초 15항, $\{z_j\}$의 8항

$$\begin{aligned}(3.3)\quad \{b_i\} &= 1,0,0,1,0,1,0,1,1,\\ &\quad 0,0,0,0,1,1,(1,0)\\ \{z_j\} &= 1,0,1,1,0,1,1,0,\\ &\quad (1,0)\end{aligned}$$

을 선택한다.

선택생성기에 의한 출력 $\{a_i\}$의 주기는 15이므로 길이가 4인 블럭과 길이가 3인 갭을 각각 1개씩 포함한다. 따라서

(A) 키수열 $\{z_j\}$의 연속한 4개의 부호열은 $\{b_i\}$의 연속한 4개의 부호열과 일치하고 $\{b_i\}$의 연속한 3개의 부호열은 키수열에 나타나지 않는다.

생성기 수열의 복구를 위해 부분수열이 최대블록을 포함하는 경우, 최대갭을 포함하는 경우, 그리고 기타의 세 경우로 나누어 분석하고 마지막으로 자기상관도를 조사하기로 한다.

3.1 최대블록을 포함하는 경우

이 예의 경우 $\{b_i\}$와 키수열 $\{z_j\}$가 일치되는 비트열은

$$\begin{aligned}(3.4)\quad \{b_6, b_7, b_8, b_9\} &= \{z_4, z_5, z_6, z_7\}\\ &= \{1,0,1,1\}\end{aligned}$$

이다. 따라서 부분수열 $\{b_6, b_7, b_8, b_9\}$에 대응하는 선택생성기의 출력은

(3.5) $\{a_6, a_7, a_8, a_9\} = \{1,1,1,1\}$

이어야 한다. 즉, 선택생성기에 의한 수열은

(3.6) $\{a_i\} = *,*,*,*,*,1,1,1,1,$
$*,*,*,*,*,*,\cdots$

의 형태임을 알 수 있다.

3.2 최대갭을 포함하는 경우

키수열 $\{z_j\}$에 나타나지 않은 $\{b_i\}$의 세 비트는 0, 0, 0. 0, 1, 1 가운데 연속한 3항임을 알 수 있다.

수열 (3.6)의 뒤 부분

(3.7) $\{a_{10}, a_{11}, a_{12}, a_{13}, a_{14}, a_{15}\}$
$= \{*,*,*,*,*,*\}$

는 이에 대응되는 $\{b_i\}$의 부분수열

(3.8) $\{b_{10}, b_{11}, b_{12}, b_{13}, b_{14}, b_{15}\}$
$= \{0,0,0,0,1,1\}$

과 키수열

(3.9) $\{z_8\} = \{0\}$

에 의해 결정된다. 즉, 부분수열 (3.7)은 런의 크기 3인 갭을 포함하고, 수열 (3.8), (3.9)에 의해 a_{10}, a_{11}, a_{12}, a_{13} 가운데 한 비트만 '1'이고 나머지 비트는 모두 '0'이어야 한다.

만일 $a_{10}=1$ 이면, 나머지 비트 $a_{11},\cdots$, a_{15}는 모두 '0'이어야 하므로 $\{a_i\}$가 길이 5인 갭을 갖게 되고, $a_{11}=1$ 이면, 나머지 비트 $a_{12},\cdots,a_{15}$는 모두 '0'이어야 하므로 $\{a_i\}$가 길이 4인 갭을 갖게 되므로 Golomb의 의사난수열에 관한 공리에 어긋난다.

그러므로 a_{10}, a_{11}, a_{12}, a_{13}가운데 '1'이 될 수 있는 비트는 a_{12}, a_{13}가운데 하나이다. 즉, 수열 (3.7)은

(3.10) $\{a_{10}, a_{11}, a_{12}, a_{13}, a_{14}, a_{15}\}$
$= \{0,0,*,*,0,0\}$

의 형태를 갖는다.

3.3 기타의 경우

수열 $\{a_i\}$의 최초 5항 $a_1, a_2,\cdots a_5$는 이에 대응되는 $\{b_i\}$의 부분수열

(3.11) $\{b_1, b_2, b_3, b_4, b_5\}$
$= \{1,0,0,1,0\}$

와 $\{z_j\}$의 부분수열

(3.12) $\{z_1, z_2, z_3\} = \{1,0,1\}$

에 의해 결정되므로

(3.13) $\{a_1, a_2, a_3, a_4, a_5\}$
$= \{1,*,*,1,0\}$

의 형태를 갖는다.

각 경우에 대한 분석으로부터 얻은 부분수열 (3.5), (3.10), (3.13)을 종합하면 선택생성기에 의한 수열은

(3.14) $\{a_i\}=\{\ 1, a_2, a_3, 1, 0, 1, 1, 1, 1, 0, 0, a_{12}, a_{13}, 0, 0\}$

의 형태를 갖는다. 수열 (3.10), (3. 13)으로부터 (3.14)의 연속한 미지의 두 비트 $\{a_2, a_3\}$와 $\{a_{12}, a_{13}\}$는 서로 다른 부호를 가짐을 알 수 있다.

3.4 자기상관

수열 $\{a_i\}$의 자기상관도는

$$(3.15)\qquad C(\tau)=\frac{1}{p_1}\sum_{i=1}^{p_1} x_i x_{i+\tau}$$

로 정의된다 여기서

$$x_i=\begin{cases} 1 & \text{if } a_i=1 \\ -1 & \text{if } a_i=0 \end{cases}$$

이다. 수열 $\{a_i\}$는 최대주기 $p_1=15$인 LFSR의 출력수열이므로 Golomb의 공리 *G(3)*를 만족하여야 한다. 즉, $\tau\neq p_1$이면

$$C(\tau)=-\frac{1}{p_1}=-\frac{1}{15}$$

이어야 한다.

수열 (3.14)의 a_2, a_3, a_{12}, a_{13}은 다음의 네 가지 경우 가운데 하나이다;

$$(3.16)\quad \{a_2, a_3, a_{12}, a_{13}\}=\begin{cases} 1,0,1,0, \\ 0,1,0,1, \\ 1,0,0,1, \\ 0,1,1,0. \end{cases}$$

그러나 위의 부분수열 (3.16)의 처음 두 경우에서 $\tau=10$ 일 때의 자기상관도는

$$(3.17)\qquad C(10)=\frac{1}{15}\sum_{i=1}^{15} x_i x_{i+10}$$
$$=\frac{1}{5}\neq-\frac{1}{15},$$

또 세 번째의 경우에서 $\tau=2$일 때의 자기상관도는

$$(3.18)\qquad C(2)=\frac{1}{15}\sum_{i=1}^{15} x_i x_{i+2}$$
$$=\frac{1}{5}\neq-\frac{1}{15}$$

이므로 Golomb의 공리 *G(3)*를 만족하지 못한다.

한편, $\{a_2, a_3, a_{12}, a_{13}\}=\{0,1,1,0\}$인 경우에는 $0<\tau<p_1$이면

$$C(\tau)=-\frac{1}{15}$$

이므로 Golomb의 공리 *G(3)*를 만족한다.

따라서 선택생성기의 출력수열 $\{a_i\}$는 $\{a_i\}=\{1,1,0,1,0,1,1,1,1,0,0,0,1,0,0\}$로 완전히 복구되었다.

4. 결 론

Shrinking generator는 두 개의 LFSR로 구성된 간단한 형태로 신속한 처리 속도를 보장한다는 장점이 있다. 그러나 Shrinking generator를 구성하는 두 개의 LFSR은 최대주기를 갖도록 설계되어 있기

때문에 두 출력수열은 Golomb의 공리를 만족한다. 그러므로 Golomb의 공리를 적용하면 키수열과 LFSR R_2 출력수열의 관계로부터 선택생성기에 의해 생성되는 수열이 복원될 수 있는 가능성을 갖고 있다.

이러한 문제점을 해소하기 위해서는 LFSR에 관련된 정보를 공개하지 않거나, 두 LFSR들의 주기를 크게 설정해야한다. 또한 선택생성기에 의해 생성되는 수열의 초반부에 긴 길이를 갖는 블록이 나타나는 것을 피해야 한다.

이 논문에서 제시한 분석 알고리즘이 앞의 3절의 예에서처럼 선택생성기의 출력을 완전히 복구할 수 없는 경우에도 이 분석 알고리즘은 선택생성기 출력의 많은 부분을 복구할 수 있다. 또한 선택생성기에 대한 정보가 공개되지 않은 경우에도 이 분석 방법은 전수조사에 비해 여전히 높은 효율성을 갖는 의미 있는 공격이라 할 수 있다.

참 고 문 헌

[Bla] S.R.Blackburn, The linear complexity of the self-shrinking generator, IEEE Transactions on Information Theory, 45(6), 2073-2077, 1999.

[CKM] D.Coppersmith, H.Krawczyk & Y. Mansour, "The Shrinking Generator", Advances in Cryptology-CYPTO'93 (LNCS773), 22-39, 1994.

[Gol] S.W.Golomb, Shift Register Sequences, Holden-Day, Inc., San Francisco, 1967.

[MOV] A.Menezes, P.Oorschot, S.Vanstone, Handbook of Applied Cryptography, CRC Press, New York, 1997.

[Shp] I.Shparlinski, On Some Properties of the Shrinking Generator, www.comp.mq.edu.au/~igor/Shrink.ps