情報保護學會論文誌
第14券 第2號, 2004. 4

LFSM 기반의 비선형 필터 모델의 특성

홍 진†, 이 동 훈‡, 지 성 택
국가보안기술연구소

A Characteristic of Nonlinear Filter Models based on LFSMs

Jin Hong†, Dong Hoon Lee‡, Seongtaek Chee
National Security Research Institute

요 약

본 논문에서는 잘 알려진 선형 대수 이론을 이용하여 LFSR을 활용한 LFSM의 구현 방법을 제시한다. 이를 사용하여 셀룰라 오토마타를 LFSR대신 사용하여 Anderson 정보누출과 같은 스트림 암호의 좋지 않은 성질을 제거하고자 하는 시도로는 안전성을 높일 수 없음을 보인다. 또한 LFSM에 기반한 비선형 필터 생성기가 Anderson 정보 누출의 위험성이 있는 형태로 변형 가능한지 확인하는 방법을 제시한다.

ABSTRACT

We present a realization of an LFSM that utilizes an LFSR. This is based on a well-known fact from linear algebra. This structure is used to show that a previous attempt at using a cellular automata in place of an LFSR in constructing a stream cipher did not necessarily increase its security. We also give a general method for checking whether or not a nonlinear filter generator based on an LFSM allows reduction to one that is based on an LFSR and which is vulnerable to Anderson information leakage.

Keywords: *LFSM, LFSR*

Ⅰ. 서 론

선형 피드백 쉬프트 레지스터(Linear Feedback Shift Register, LFSR)는 스트림 암호의 설계에 사용되는 중요한 구성 요소 중 하나이다. LFSR을 사용한 고전적인 스트림 암호 체계로 비선형 필터 모델(Nonlinear Filter model, NF)과 비선형 조합 모델(Nonlinear Combiner model, NC)이 있다. [1,2]의 결과를 바탕으로 하여 Anderson은 NF의 경우 비선형 필터 함수의 입력으로 사용된 셀들이 서로 가까운 경우 출력 수열로부터 LFSR의 초기치에 대한 많은 정보를 얻을 수 있음을 보였다.[3]

Sarkar는 NF의 Anderson 정보누출을 막기 위하여 NF와 NC을 결합한 스트림 암호 시스템을 소개하였다.[4] 이 모델은 LFSR 대신에 셀룰라 오토마타(Cellular Automata, CA)를 사용하여 Anderson 정보누출을 막을 수 있다고 주장하고 있으나, 본 논문에서는 CA를 사용하는 것이 암호 시스템의 안전성을 증가시키지 않음을 보인다.

CA는 선형 유한 상태 장치(Linear Finite State Machine, LFSM)의 특별한 경우이다. 이는 셀들의 1차원 배열로 생각 할 수 있으며 시간의 진행에 따라 셀들의 상태는 선형 변환을 통하여 변하게 된다. 이러한 CA는 그 우수한 랜덤 특성에 의하

접수일: 2004년 1월 2일; 채택일: 2004년 3월 11일
† 주저자, jinhong@etri.re.kr
‡ 교신저자, dlee@etri.re.kr

여 여러 방면에 활용되었다.[5,6,7] 안전성 관점으로 볼 때, CA가 LFSR보다 좋다고 알려졌으나 본 논문에서는 CA와 LFSR은 안전성 관점으로 차이가 없음을 밝히도록 한다.

선형대수의 쉬운 결과들을 사용하여 우리는 LFSM을 LFSR과 선형 변환의 조합으로 구현하는 방법을 제시한다. LFSR이 LFSM에 비하여 간단한 구조이므로, 이는 LFSM을 활용한 암호학적 시스템들의 안전성에 시사하는 바가 크다. 그 구체적인 예를 [4]의 결과를 분석하며 살펴보도록 한다.

본 논문은 다음과 같이 구성되었다. Ⅱ장에서는 이후 논의에 필요한 선형대수의 기초적인 이론을 살펴보고, CA와 LFSR이라는 2종의 대표적인 LFSM에 대하여 알아본다. Ⅲ장에서는 주어진 LFSM을 LFSR과 선형 변환을 결합하여 구현하는 방법을 제시한다. 이를 이용하여 [4]에서 제안되었던 CA를 LFSR 대신 사용하여 안전성을 증가시키고자 한 시도가 의도한 바를 이루지 못했음을 Ⅳ 장에서 보이고, 실제 구체적인 예를 Ⅴ장에서 제시한다. Ⅵ장에서는 주어진 비선형 필터 함수가 Anderson 정보누출이 일어날 수 있는 꼴로 변형이 가능한지 판별하는 방법을 기술하고, 마지막 Ⅶ장에서 결론을 맺는다.

Ⅱ. 기초 지식 및 정의

2.1 선형 대수

F_2를 원소의 개수가 2개인 유한체 $\{0,1\}$이라고 하고, 크기 n의 항등행렬을 I로 표시하자. 주어진 정방행렬 M의 특성다항식은 다음과 같이 정의한다.

$char(M) = \det(xI-M) \in F_2[x]$.

모닉 다항식

$$p(x)=a_0+\cdots+a_{n-1}x^{n-1}+x^n \in F_2[x] \qquad (1)$$

의 동반행렬은 다음의 행렬로 정의한다.

$$\begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & & \vdots & \vdots \\ 0 & 0 & 0 & & & 1 & 0 \\ 0 & 0 & 0 & \cdots & \cdots & 0 & 1 \\ a_0 & a_1 & a_2 & \cdots & \cdots & a_{n-2} & a_{n-1} \end{pmatrix} \qquad (2)$$

그러면 다음의 정리는 일반적인 선형 대수의 교재에 소개되어 잘 알려져 있다.[8,9]

[정리 1] 정방행렬 M의 특성다항식을 $p(x)$라 하고, $p(x)$의 동반행렬을 L로 표기하자. $p(x)$가 기약이면, $TMT^{-1}=L$을 만족하는 가역인 (기저변환) 행렬 T가 존재한다.

그런데 위 [정리 1]의 식을 만족시키는 행렬 T가 유일한 것은 아니다. 정방행렬 M의 크기가 n인 경우 최대 2^n-1가지 가능성이 있다.

2.2 선형 유한 상태 장치

앞으로 $\mathcal{L}$로 표기할 n-비트 선형 유한 상태 장치 (linear finite state machine, LFSM)란, $n\times n$ 행렬 M으로 주어진 하나의 쌍 (F_2^n, M)을 말한다. LFSM $\mathcal{L}$의 내부상태는 하나의 n-비트 벡터 $v=(v_0,\cdots,v_{n-1})\in F_2^n$으로 표시한다. 음이 아닌 정수 시간 t의 진행에 따른 $\mathcal{L}$의 내부상태 변화는 $v^{(t+1)}=Mv^{(t)}$을 만족하는 n-비트 벡터 $v^{(0)}, v^{(1)}, \cdots$로 주어진다. 여기서 $v^{(0)}$는 초기상태를 나타낸다. 시간 $t\geq 0$에서의 상태를 좀 더 구체적으로 표현할 때는 $v^{(t)}=(v_0^{(t)}, v_1^{(t)}, \cdots, v_{n-1}^{(t)})$를 사용한다. 행렬 M의 특성다항식이 F_2상에서 원시다항식인 경우 다음의 수열은 각각의 주기가 2^n-1이라는 사실이 잘 알려져 있다.[10]

$$v_i=(v_i^{(t)})_{t\geq 0} \qquad (3)$$

이는 바로 LFSM의 내부상태로 주어진 이와 같은 수열로 얻을 수 있는 최대 주기이다. LFSM 중 많이 사용되는 것으로는 셀룰라 오토마타와 선형 피드백 쉬프트 레지스터를 들 수 있다.

2.3 셀룰라 오토마타

셀룰라 오토마타(Cellular Automata, CA)는 LFSM의 한 종류로 그를 정의하는 행렬 M이 tri-diagonal인 경우를 말한다. 특별히 M의 대각원 바로 윗쪽과 바로 아랫쪽이 모두 1인 경우를 90/150 CA라 부른다. 좀 더 구체적으로 말하면, 다음 형태의 행렬로 주어지는 LFSM을 90/150

CA라 한다.

$$\begin{pmatrix} c_0 & 1 & 0 & 0 & \cdots & 0 \\ 1 & c_1 & 1 & 0 & \cdots & 0 \\ 0 & 1 & c_2 & 1 & & 0 \\ \vdots & & & & & \vdots \\ & & & & 1 & 0 \\ 0 & & \cdots & 1 & c_{n-2} & 1 \\ 0 & & \cdots & 0 & 1 & c_{n-1} \end{pmatrix}$$

여기서 각각의 c_i는 0 또는 1이다. 본 논문에서의 셀룰라 오토마타는 90/150 CA만을 다루도록 한다. 이러한 CA의 내부상태는 각 $0 \le i \le n-1$와 $t \ge 0$에 대하여 관계식 $v_i^{(t+1)} = v_{i-1}^{(t)} \oplus c_i v_i^{(t)} \oplus v_{i+1}^{(t)}$을 만족한다. 여기서 $v_{-1}^{(t)} = v_n^{(t)} = 0$이다.

2.3 선형 피드백 쉬프트 레지스터

식 (1)로 주어진 모닉 다항식 $p(x)$가 정의하는 선형 피드백 쉬프트 레지스터(linear feedback shift register, LFSR)란, 행렬 (2)로 주어진 $p(x)$의 동반행렬로 정의된 LFSM을 말한다. 따라서 LFSR의 시간 $t \ge 0$에서의 내부상태를 $v^{(t)} = (v_0^{(t)}, v_1^{(t)}, \cdots, v_{n-1}^{(t)})$라 놓으면 각각의 $0 \le i \le n-2$에 대하여 $v_i^{(t+1)} = v_{i+1}^{(t)}$이 성립하며 $v_{n-1}^{(t+1)} = a_0 v_0^{(t)} \oplus a_1 v_1^{(t)} \oplus \cdots \oplus a_{n-1} v_{n-1}^{(t)}$ 또한 성립한다. 즉, LFSR을 이루는 각 셀의 내용물은 시간의 진행에 따라 한 칸씩 왼쪽으로 이동된다.

Ⅲ. LFSM의 LFSR으로의 단순화

이제 앞 장의 각 부분이 서로 어떻게 연관성을 가지고 있는지를 살펴보도록 한다. 본 장에서는 LFSR과 선형 변환을 결합하여 LFSM을 구현하는 방법을 제시한다. 행렬 M으로 정의된 LFSM $\mathcal{L}$이 주어졌다고 하자. M의 특성다항식을 $p(x)$로 표기하고 $p(x)$의 동반행렬을 L이라 하자. 이때 행렬 L은 하나의 LFSR을 정의할 수 있으므로 이렇게 정의되는 LFSR을 LFSM $\mathcal{L}$이 정의하는 LFSR이라고 정의한다.

특성다항식 $p(x)$가 기약이라 하자. [정리 1]에 의하여 $TMT^{-1} = L$을 만족하는 가역행렬 T가 존재한다. LFSM $\mathcal{L}$의 초기 상태가 v였다면, 시간 $t \ge 0$에서의 LFSM의 상태 $v^{(t)}$는 다음과 같이 주어진다.

$$v^{(t)} = M^t v$$

여기서 M^t는 M을 t번 적용하였음을 의미한다. 마찬가지로, 행렬 L이 정의하는 LFSR의 초기 상태가 w라면, 시간 $t \ge 0$에서의 LFSR의 상태 $w^{(t)}$는 다음과 같이 주어진다.

$$w^{(t)} = L^t w$$

만일 두 초기상태들이 관계식 $w = Tv$를 만족한다면, $TMT^{-1} = L$의 관계식을 활용하여 우리는 다음의 등식이 사실임을 확인할 수 있다.

$$\begin{aligned} v^{(t)} &= M^t v = (T^{-1} L T)^t v = T^{-1} L^t T v \\ &= T^{-1} L^t w = T^{-1} w^{(t)} \end{aligned}$$

이는 LFSM과 LFSM이 정의하는 LFSR이 밀접히 연관되어 있음을 보여준다.

[정리 2] 초기상태 $v^{(0)}$에서 출발한 LFSM의 시간 t에서의 내부상태 $v^{(t)}$는 같은 LFSM이 정의하는 LFSR의 내부상태 $w^{(t)}$를 활용하여 다음의 간단한 관계식을 통하여 구할 수 있다.

$$v^{(t)} = T^{-1} w^{(t)} \tag{4}$$

여기서 LFSR의 초기상태는 $w^{(0)} = Tv^{(0)}$로 놓으면 된다.

결론적으로 LFSM이 LFSR보다 복잡해 보이기는 하나, 둘의 차이는 결국 하나의 간단한 선형 관계식에 지나지 않는다. 본 장에서 사용된 LFSM에 대한 가정은 대응되는 특성다항식이 기약이라는 것뿐이다. 그러나 LFSM의 암호학적 응용에서는 대부분의 경우에는 최대 주기를 보장하기 위하여 특성다항식이 원시 다항식이라는 더욱 강력한 조건이 사용된다. 따라서 LFSR에 비하여 LFSM이 복잡하다는 생각에 기반한 LFSM의 암호학적 활용은 제고되어야 할 것이다. 더불어, CA 또한 일종의 LFSM이므로 CA에 관하여도 같은 주장을 할 수 있다.

Ⅳ. 셀룰라 오토마타 기반 비선형 필터 모델의 안전성

4.1 NF-CA

[4]에서, Sarkar는 필터-조합 모델(Filter-Combiner, FC)이라는 스트림 암호를 제안하였다. 여기서는 FC 모델 전체를 다루지 않고 그 논문의 주된 논의 중 하나만을 설명하도록 한다.

$\mathcal{L}=(F_2^n, M)$으로 표시되는 CA가 하나 주어졌다고 하고 CA에 대응되는 특성다항식 $p(x)$가 원시다항식이라 하자. 식 (3)으로 주어진 각각의 n개 수열들이 모두 동일한 주기수열임은 잘 알려져 있다. 다만 각 수열의 시작점이 다를 뿐이다. 즉 각 수열은 서로 서로의 상대적 쉬프트로 볼 수 있다.

주어진 CA에 좋은 특성을 가지는 비선형 필터 함수 f를 적용하여 키스트림을 얻는다고 하자. 여기서 좋은 특성이란, 높은 resiliency와 비선형성 등을 말한다. 이렇게 주어진 시스템이 다음의 조건들을 만족한다고 하자.[4]

1. 함수 f의 입력으로 사용된 셀의 수 r은 전체 CA의 크기 n에 비하여 작다. ($r \le \log_2 n$)
2. 함수 f의 입력으로 사용된 셀들이 주는 (동일한) 주기수열들의 시작점들은 전체 주기 안에서 (대략적으로) 균일하게 분포한다.
3. 이 시스템을 사용하여 암호화되는 전체 비트의 양은 $2^n/r$에 비하여 상당히 적다.

우리는 이 (축소된) 모델을 NF-CA라 부르기로 한다. CA를 활용한 비선형 필터 모델이라는 뜻이다. [4]는 위의 조건들을 만족하는 NF-CA는 Anderson 정보누출[3]에 저항성을 가진다고 주장하고 있다. Anderson 정보누출이란 스트림 암호의 일종인 비선형 필터 모델에 대한 결과이다. 사용된 비선형 필터 함수가 LFSR의 셀 중 서로 가까이 있는 것만을 사용할 경우 상당히 높은 확률로 키 스트림으로부터 LFSR의 초기상태에 대한 정보를 얻을 수 있다는 결과이다.[3]

Sarkar는 Anderson 정보누출의 근본적인 원인으로 같은 비트 정보를 필터 함수의 입력으로 여러번 재사용하는 비선형 필터 모델의 특성을 지적했다. 이 같은 생각은 [3]에도 어느 정도 드러나 있다. 따라서 위의 조건들의 기본적인 의도는 주기수열의 어느 부분도 재사용되는 것을 막는 것이었다.

4.2 NF-CA의 정보 누출

NF-CA의 초기상태, 좀더 정확히는 사용된 CA, $\mathcal{L}=(F_2^n, M)$ 의 초기상태를 $v^{(0)}$로 표기하자. 필요하면 비선형 필터 함수 f의 입력으로 dummy 변수들을 추가하여 f가 CA 전체 셀을 입력으로 하여 정의되었다고 보자. NF-CA의 t번째 키스트림 비트 c_t는 다음과 같이 주어진다.

$$c_t = f(v^{(t)}) \tag{5}$$

이제 Ⅲ장의 내용을 따라가며 CA가 정의하는 LFSR을 구성하고 식 $TMT^{-1}=L$을 만족하는 T를 찾을 수 있다. 그리고 식 (4)를 위의 식 (5)에 적용하여 $c_t = f \circ T^{-1}(w^{(t)})$를 얻을 수 있다. 이때 LFSR의 초기치는 $w^{(0)} = Tv^{(0)}$로 놓으면 된다. 여기서 T^{-1}는 간단한 선형 변환에 지나지 않음을 생각하면 $g = f \circ T^{-1}$라는 함수가 또 다른 보통의 비선형 필터 함수라는 것을 볼 수 있다. 즉, 우리는 NF-CA의 키스트림을 $c_t = g(w^{(t)})$꼴의 평범한 NF-LFSR 형태로 쓸 수 있는 것이다.

[정리 3] 비선형 필터 함수 f를 사용하고 초기값이 $v^{(0)}$로 주어진 NF-CA는 CA가 정의하는 LFSR에 비선형 필터 함수 $g = f \circ T^{-1}$를 적용하고 그 초기치를 $w^{(0)} = Tv^{(0)}$로 하여 구현할 수 있다.

아직 Anderson 정보누출에 대한 저항성을 측정할 수 있는 일반적인 방법은 없다. 그리고 Anderson의 논문[3]은 랜덤하게 선택한 NF-LFSR들은 상당히 많은 경우 정보를 누출한다고 주장하고 있다. 따라서 식 (5)는 Anderson 정보누출의 위험성이 있는 키스트림을 출력한 가능성을 여전히 가지고 있다.

Ⅴ. 정보 누출을 보여주는 NF-CA의 예

Ⅳ장의 결과를 확인하기 위하여 Ⅳ 장의 3가지 조건을 모두 만족하며 Anderson 정보누출을 보여주는 구체적인 NF-CA를 구성하도록 한다.

5.1 CA 및 관련 LFSR

행렬 M이 식 2.3절의 90/150 CA의 형태로 주어지고 대각원이 $(c_0, c_1, \cdots, c_{22}) = (1,1,1,0,0,1,0,0,1,1,1,1,1,0,0,1,1,1,0,0,1,0,0)$라고 하자. M의 특성 다항식 $p(x)$는 다음과 같다.

$$p(x) = 1 + x^2 + x^4 + x^7 + x^9 + x^{10} + x^{11} + x^{13} + x^{14} + x^{15} + x^{17} + x^{19} + x^{20} + x^{22} + x^{23}$$

이것은 원시다항식이므로 행렬 M으로 정의되는 CA의 23개 셀은 모두 주기 $2^{23}-1$의 수열을 만든다.

다항식 $p(x)$의 동반행렬을 L이라 하자. 이제 T는 다음과 같이 정의한다.

$T_1 = (1,0,0,\cdots,0)$을 T의 첫번째 행으로 하고, 순차적으로 i번째 행 T_i를 $T_i = MT_{i-1}$ $(1 < i \leq 23)$으로 정의한다. 행렬 T의 전체 모습은 다음과 같다.

```
1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
1 0 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
1 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
1 1 0 1 1 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 1 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 1 1 1 1 1 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 1 0 0 0 1 1 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0
1 1 1 0 1 0 1 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0
0 1 0 0 0 0 0 1 0 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0
1 1 1 0 0 0 1 0 0 0 1 0 1 0 0 0 0 0 0 0 0 0 0 0
0 1 0 1 0 1 0 1 0 1 1 0 1 1 0 0 0 0 0 0 0 0 0 0
1 1 0 0 0 1 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0
0 0 1 0 1 1 1 0 0 0 0 0 0 1 1 1 1 0 0 0 0 0 0 0
0 1 1 0 1 1 1 1 1 0 0 0 1 0 0 0 0 1 0 0 0 0 0 0
1 0 0 0 1 1 0 1 1 0 1 1 1 0 0 1 1 1 0 0 0 0 0 0
1 1 0 1 1 0 0 1 0 0 0 1 0 1 1 1 0 1 0 1 0 0 0 0
0 0 0 1 1 1 1 0 1 0 1 1 0 1 1 1 0 1 0 0 1 0 0 0
0 0 1 1 0 1 1 0 1 0 0 0 0 1 1 0 1 1 1 0 1 1 0 0
0 1 0 1 0 0 1 0 1 1 0 0 1 1 1 0 0 1 1 0 1 1 0 0
1 1 0 0 1 1 0 0 0 0 1 1 0 0 1 1 1 0 1 0 0 1 1 1
```

이렇게 정의한 행렬 T가 $TMT^{-1} = L$을 만족하는 것은 쉽게 확인할 수 있다. 물론 이 식을 만족하는 T가 유일한 것은 아니다. 임의의 0이 아닌 초기 벡터 T_1으로 시작하여 위와 같은 방식을 통하여 정의한 가역인 행렬은 식 $TMT^{-1} = L$을 만족시킨다. 여기서 정한 T에 대한 역원 T^{-1}는 다음과 같다.

```
1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 1 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
1 1 0 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
1 1 1 0 1 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
1 0 1 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
1 0 1 1 1 1 1 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 1 0 0 1 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0
1 1 0 1 0 0 1 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0
1 1 1 1 0 0 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0
0 1 0 1 1 0 0 0 0 0 1 0 1 0 0 0 0 0 0 0 0 0 0 0
1 0 0 0 0 1 1 1 1 1 0 0 1 1 0 0 0 0 0 0 0 0 0 0
0 0 0 1 1 0 1 1 1 1 1 0 0 1 1 1 0 0 0 0 0 0 0 0
1 0 0 0 1 0 1 0 0 0 1 0 1 0 1 1 0 0 0 0 0 0 0 0
1 1 0 1 0 1 0 0 1 1 1 1 0 0 0 0 1 0 0 0 0 0 0 0
0 0 1 1 0 1 0 0 1 0 1 0 0 0 1 1 1 1 0 0 0 0 0 0
1 1 1 1 1 0 1 0 0 0 0 0 0 0 1 0 1 0 1 0 0 0 0 0
0 1 0 0 1 0 0 1 1 0 1 0 0 0 1 0 1 0 0 1 0 0 0 0
1 1 0 1 1 1 1 0 1 1 0 1 0 0 1 1 1 1 1 1 0 1 0 0
1 1 1 1 1 0 0 0 0 0 0 1 1 0 0 0 1 0 0 0 1 1 1 0
1 0 1 0 0 0 0 1 0 1 1 1 0 1 1 1 1 1 1 1 0 1 0 1
```

5.2 CA 셀들 간의 쉬프트와 비선형 필터 함수

각각의 $1 \leq i \leq 23$에 대하여 $m_i^{(t)}$를 CA의 i번째 셀이 출력하는 (시점을 제외하고는 동일한) 주기 수열이라 하자. $m_1^{(t)}$와 $m_i^{(t)}$ 사이의 상대적인 쉬프트는 순서대로 다음과 같다.

0, **1988170**, **8388605**, 5964510, **4125305**, 3763873, **6190462**, 6778815, ...

이것은 [11]의 내용을 구현한 프로그램으로 계산된 값이다. 이를 통하여 2, 3, 5, 7번째 셀의 출력들 사이의 상대적인 쉬프트 값들이 모두 2^{21}이나 2^{22}에 가까움을 확인할 수 있다.

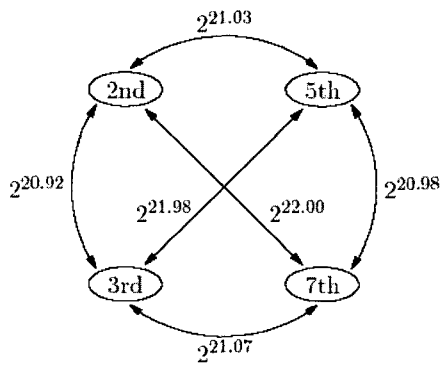


그림 1. 2,3,5,7 번째 셀의 상대적 쉬프트

예를 들어 $m_3^{(t)}$과 $m_2^{(t)}$사이의 쉬프트는 1988170 - 8388605 = 1988172 mod $2^{23}-1 \simeq 2^{20.92301}$ 이다. 따라서 이 CA에 $f=m_2 \oplus m_3 \oplus (m_5 \cdot m_7)$로 주어진 비선형 필터 함수를 적용하면 4.1 절의 2번 규칙이 만족된다. 또한 함수 f가 1-resilient하다는 것을 쉽게 보일 수 있어서 f는 그리 나쁘지 않은 비선형 필터 함수임을 받아들일 수 있다. 다음의 계산을 통하여 1번 규칙 또한 만족됨을 알 수 있다.

$$\log_2 23 = 4.52356\cdots \geq 4$$

규칙 3은 2^{21}보다 적은 개수의 비트를 사용하도록 권고하고 있으나, 우리는 대략 20비트의 키 스트림을 사용할 것이므로 이는 쉽게 만족하게 된다.

5.3 동치 NF-LFSR의 구성

$f(v^{(t)}) = f \circ T^{-1}(w^{(t)})$로부터 LFSR의 상태 $(l_1, l_2, \cdots, l_{23})$에 대하여 비선형 필터 함수 $g=f \circ T^{-1}$를 사전 계산된 T^{-1}의 2, 3, 5, 7번째 행의 구체적인 값을 활용하여 나타내면 다음과 같다.

$$\begin{aligned} g &= f \circ T^{-1} \\ &= (l_1 \oplus l_2 \oplus l_3) \\ &\quad \oplus (l_2 \oplus l_4 \oplus l_5) \cdot (l_1 \oplus l_2 \oplus l_3 \oplus l_5 \oplus l_7) \end{aligned}$$

NF-CA에 사용된 4개의 셀을 표현하는데 사용된 LFSR의 셀들은 왼쪽 7개 셀에 지나지 않음을 확인할 수 있다.

5.4 Anderson 정보누출

LFSR의 7번째 비트가 출력 키스트림 7개 비트를 뽑는 동안 유효한 효과를 주므로, 우리는 위의 과정을 통하여 얻은 NF-LFSR을 가능한 모든 13-비트 입력에 대한 7비트 출력을 보도록 작동시킨다. 각각의 출력에 대응되는 가능한 13-비트 입력의 수를 〔표 1〕에 정리하였다.

7-비트 출력은 16진수 표현을 사용하여 나타내었고, 개수는 해당 입력 개수를 나타낸다. 16진수 표현의 (8비트 중 첫 번째 0을 제외시킨) 가장 왼쪽 비트가 첫 번째 출력 비트이다. 이상적인 경우라면, 모든 입력 개수들이 $2^6=64$이거나 이에 가까워야 할 것이다.

그러나 〔표 1〕의 결과는 우리가 시험하고 있는 NF-CA/LFSR이 그러하지 않다는 것을 보여준다. 103과 같은 큰 수도 있는 반면, 29와 같이 작은 수도 나타나고 있다. 이는 이 구조의 정보유출 가능성이 높음을 나타낸다.

좀 더 구체적으로 7-비트 출력 수열 0×09=0001001을 주는 다음의 모든 13-비트 입력 값을 살펴보자. 〔표 1〕에 나오듯이 정확히 29개 초기값이 있다.

표 1. 각각의 7-비트 출력에 대응하는 입력 상태의 수

출력	00	01	02	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F
개수	38	51	61	73	51	75	73	89	65	29	67	63	69	69	87	63
출력	10	11	12	13	14	15	16	17	18	19	1A	1B	1C	1D	1E	1F
개수	79	63	45	37	87	67	77	57	49	73	43	59	73	85	59	71
출력	20	21	22	23	24	25	26	27	28	29	2A	2B	2C	2D	2E	2F
개수	73	93	51	71	61	69	39	55	79	83	77	49	75	43	57	49
출력	30	31	32	33	34	35	36	37	38	39	3A	3B	3C	3D	3E	3F
개수	65	49	99	75	57	45	67	55	63	71	69	85	39	59	53	73
출력	40	41	42	43	44	45	46	47	48	49	4A	4B	4C	4D	4E	4F
개수	51	83	65	89	43	55	65	61	77	53	87	71	53	33	71	67
출력	50	51	52	53	54	55	56	57	58	59	5A	5B	5C	5D	5E	5F
개수	87	59	85	57	75	59	41	49	65	101	59	63	61	69	39	55
출력	60	61	62	63	64	65	66	67	68	69	6A	6B	6C	6D	6E	6F
개수	61	61	47	55	69	89	47	83	67	59	57	41	91	79	73	45
출력	70	71	72	73	74	75	76	77	78	79	7A	7B	7C	7D	7E	7F
개수	57	53	59	55	69	53	103	63	47	43	53	81	51	75	73	89

1101000000100	0011000000100	1011000000100
0111000000100	0110101110100	0110100011100
0110100101010	1101000000110	0011000000110
1011000000110	0111000000110	0110101110110
0110100011110	0000011111110	0000011100001
0000000110001	0110101110001	0110101110101
0110100011101	0110101110011	1101000001011
0011000001011	1011000001011	0111000001011
0110010001011	1110010001011	0110100101011
0110101110111	0110100011111	

확률 28/29로 4, 5, 6 번째 비트 중 하나만이 1 임을 알 수 있다. 즉, 이 3개 비트의 XOR은 확률 28/29로 1인 것이다. 이 시스템은 Anderson 정보누출을 보이고 있다. 따라서 [4]에 제시된 방법대로 좋은 암호학적 성질을 가진 비선형 필터 함수를 상대적인 쉬프트를 고려하여 CA에 적용하는 것으로는 Anderson 정보누출을 막을 수 없다.

Ⅵ. 주어진 NF-LFSM의 정보누출 위험성 여부를 확인하는 방법

본 장에서는 주어진 NF-LFSM을 Anderson 정보누출에 위험한 NF-LFSR로 동치 변형하는 것이 가능한지 확인하는 일반적인 방법을 알아본다. CA와 LFSR은 모두 LFSM의 일종이므로 이 방식은 특별히 NF-LFSR에도 적용된다. 즉, NF-LFSR에 사용된 비선형 필터 함수가 정보누출을 줄 수 있는 함수로 변형 될 수 있는지도 밝힐 수 있다는 것이다.

크기 n의 정방행렬 M과 비선형 필터 함수 f로 정의된 NF-LFSM이 하나 주어졌다고 하자. M의 동반행렬을 L이라 표시하기로 하고 L의 centralizer를 $\{Z(L)=Z\in GL(n)|ZL=LZ\}$로 표시한다. 동반행렬 L이 주어지면 $Z(L)$을 좀더 구체적으로 표현하는 것이 가능하다. 다음의 보조정리는 관계식 $ZL=LZ$의 적용만으로도 증명할 수 있다.

[보조 정리 4] 식 (2) 에 의하여 주어진 동반행렬 L에 대하여 그 centralizer $Z(L)$은

$$
\begin{aligned}
z_{i+1,0} &= a_0 z_{i,n-1} \\
z_{i+1,1} &= a_1 z_{i,n-1} \oplus z_{i,0} \\
z_{i+1,2} &= a_2 z_{i,n-1} \oplus z_{i,1} \\
&\vdots \\
z_{i+1,n-1} &= a_{n-1} z_{i,n-1} \oplus z_{i,n-2}
\end{aligned}
$$

을 만족하는 원소 $Z=(z_{i,j})_{i,j=0}^{n-1}\in GL(n)$으로 주어진다.

[보조정리 4]가 의미하는 중요한 결과는 $Z\in Z(L)$의 임의의 항을 일반적인 방식으로 그 첫 행에 나타나는 항들의 선형 결합으로 표현할 수 있다는 것이다. 예를 들어, 위의 보조정리는

$$
Z\left(\begin{pmatrix}0&1&0\\0&0&1\\a&b&c\end{pmatrix}\right)
=\left\{\begin{pmatrix}x & y & z\\ az & x\oplus bz & y\oplus cz\\ ay\oplus acz & az\oplus by\oplus bcz & x\oplus bz\oplus cy\oplus cz\end{pmatrix}\right\}
$$

임을 알려준다.

이제 $\overline{T}M\overline{T}^{-1}=L$을 만족시키는 임의의 행렬 $\overline{T}$를 하나 고정하자. 식 $TMT^{-1}=L$을 만족하는 임의의 T는 $\{Z(L)\overline{T}=Z\overline{T}|Z\in Z(L)\}$로 주어짐은 쉽게 보일 수 있다. 그리고 $Z\in Z(L)$과 $Z^{-1}\in Z(L)$이 서로 동치이므로 우리는 다음의 정리를 얻게 된다.

[정리 5] 식 $TMT^{-1}=L$을 만족하는 T^{-1}의 집합은 $\{\overline{T}^{-1}Z(L)=\overline{T}^{-1}Z|Z\in Z(L)\}$로 주어진다.

이 결과를 활용하여 주어진 임의의 NF-LFSM이 위험한 NF-LFSR로 변형 가능한지 확인하는 방법을 제시한다. 어떠한 LFSR에 적용된 비선형 필터 함수가 서로 가까이 놓인 셀들만을 사용한다면 그러한 NF-LFSR은 Anderson 정보누출의 위험성을 가지고 있다.

반대의 경우라면 Anderson 정보누출의 가능성이 적다고 할 수 있다. 따라서 우리의 문제는 주어진 NF-LFSM에 대하여 [정리 5]에 의한 필터 함수 $g=f\circ T^{-1}$가 가까이 몰려있는 셀들만을 사용하도록 하는 T가 존재하는지 확인하는 것과 동치이다.

우선 작은 수 $s<n$을 고정한다. T를 선택하여 대응되는 g가 사용하는 모든 변수들이 LFSR의 연속된 s개의 셀들로 주어지도록 할 수 있으면 Anderson 정보누출에 위험하다고 판단하고 그렇지 않으면 안전하고 판단하도록 하겠다.

Anderson 정보 누출의 판정 알고리즘

1. LFSM을 정의하는 크기 n의 정방행렬 M으로

부터, 그 특성다항식을 계산하고 이로 정의된 LFSR L을 구한다.

2. 관계식 $\overline{T}M\overline{T}^{-1}=L$을 만족하는 임의의 $\overline{T}$를 하나 찾고 고정한다.
3. [보조 정리 4]의 방식으로 $Z(L)$을 적는다. 즉 아래쪽 행의 모든 항들이 첫 번째 행에 나타나는 항들의 선형 결합으로 표현되도록 한다. 첫 행의 항들을 $x_0, \cdots, x_{n-1}$로 표기하기로 한다.
4. 전 단계에서 구한 $Z(L)$의 일반적인 원소에 $\overline{T}^{-1}$을 곱하고 [정리 5]를 사용하여 T^{-1}를 일반적으로 표현한다. 결국 T^{-1}의 모든 항은 다시 x_j의 일차결합으로 나타난다. 함수 f가 사용한 행의 수를 r이라 표시하자. x_j의 일차결합들의 $n\times n$ 배열인 T^{-1}의 일반적인 꼴 중에서 비선형 필터 함수 f에 사용된 r개의 행들만이 우리의 관심이다.
5. 함수 f에 사용되지 않는 모든 행을 제거한다. 이제, x_j들의 어느 특정 자명하지 않은 구성에 대하여 (남아 있는 행들에 대하여) 처음 s개 열에 포함된 항들을 제외한 모든 항의 값이 0이 된다고 가정해보자. 그러한 경우에 $g=f\circ T^{-1}$는 그 특정 x_j들의 구성으로 주어진 T^{-1}에 대하여 s개 LFSR 셀만을 사용하게 될 것이다.
6. T^{-1}의 남아있는 항들이 이루는 행렬에서 임시로 처음 s개의 열을 제거한다.
7. 남은 나머지 항들을 0으로 놓고 이 연립 방정식이 자명하지 않은 해를 가지는지 확인한다.
8. 자명하지 않은 해가 발견되면 주어진 NF-LFSM은 위험한 NF-LFSR로의 변형이 가능하다고 결론 내린다.
9. 발견되지 않으면, 5단계에서 얻은 행렬을 다시 가져온다.
10. 가능한 모든 연속된 s개의 열을 시도한 것이 아니라면 다음의 연속된 s개 열을 (일시적으로) 제거하고 7단계로 돌아간다.
11. 자명하지 않은 해가 발견되지 않았으면 주어진 NF-LFSM는 Anderson 정보누출에 안전하다고 결론 내린다.

이 과정의 복잡도는 n에 대한 다항식 정도에 지나지 않음을 쉽게 보일 수 있다.

Ⅷ. 결론

본 논문에서는 우리는 LFSM(또는 CA)이 또 하나의 LFSR과 간단한 관계식을 통하여 밀접하게 연관되어 있음을 보았다. 이러한 구조는 LFSR과 선형변환을 통하여 LFSM을 구현할 수 있음을 보여준다. LFSR이 LFSM에 비하여 상당히 간단한 구조이므로 이는 LFSM의 반복적 사용을 기반으로 한 암호학적 시스템의 안전성에 시사하는 바가 크다. 본 논문에서는 CA를 LFSR 대신 사용하여 비선형 필터 모델의 안전성을 높이고자 한 예가 목적한 바를 이루지 못했음을 살펴보았다. 또한 주어진 NF-LFSM이 Anderson 정보누출의 위험성을 가진 NF-LFSR로 변환 가능한지 여부를 확인하는 방법을 보았다.

참고문헌

[1] Jovan Dj. Golic, Correlation via linear sequential circuit approximation of combiners with memory, *Advances in Cryptology - EUROCRYPT'92*, LNCS 658, pp.113-123, Springer-Verlag, 1992.

[2] W. Meier and O. Staffelbach, Fast correlation attacks on certain stream ciphers. *Journal of Cryptology*, vol.1, pp.159-176, 1989.

[3] Ross Anderson, Searching for the optimum correlation attack. *Proceedings of FSE'94*, LNCS 1008, pp.137-143, Springer-Verlag, 1995.

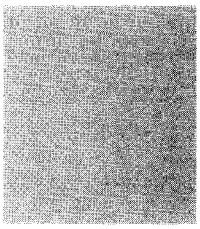
[4] Palash Sarkar, The filter-combiner model for memoryless synchronous stream ciphers. *Advances in Cryptology - CRYPTO 2002*, LNCS 2442, pp.533-548, Springer-Verlag, 2002.

[5] Palash Sarkar, Brief History of Cellular Automata, *ACM Computing Surveys*, vol. 32 (1), pp.80-107, March 2000.

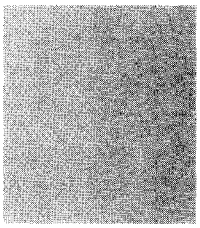
[6] Miodrag Mihaljevic, Yuliang Zheng, and Hideki Imai, A Family of Fast Dedicated One-Way Hash Functions

Based on Linear Cellular Automata over GF(q), *IEICE Trans. Fundamentals*, vol.E82-A(a), pp.1-8, January 1999.
[7] Palash Sarkar, Hiji-bij-bij: A New Stream Cipher with a Self-Synchronizing Mode of Operation, ICAR e-print 2003-014, 2003. http://eprint.iacr.org
[8] T. W. Hungerford. Algebra. GTM 73, Springer-Verlag, 1997.
[9] S. Lang. Algebra. Addison Wesley, 1993.
[10] R. Lidl and H. Niederreiter. Introduction to finite fields and their applications. Cambridge University Press, 1994.
[11] Palash Sarkar, Computing shifts in 90/150 cellular automata sequences. *Finite Fields and their Applications*, vol. 9 (2), pp.175-186, April 2003.

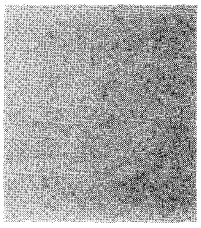
〈著者紹介〉



홍 진 (Jin Hong) 정회원
1994년 2월: 서울대학교 수학과 학사
1996년 2월: 서울대학교 수학과 석사
2000년 8월: 서울대학교 수학과 박사
2000년 9월~2002년 9월: 고등 과학원 연구원
2002년 9월~현재: 국가보안기술연구소 선임연구원
〈관심분야〉 암호론



이 동 훈 (Dong Hoon Lee) 정회원
1994년 2월: 서울대학교 수학교육과 학사
1996년 2월: 한국과학기술원 수학과 석사
2000년 2월: 한국과학기술원 수학과 박사
2000년 2월~2002년 3월: (주)퓨처시스템 선임 연구원
2002년 4월~현재: 국가보안기술연구소 선임연구원
〈관심분야〉 응용 정수론, 암호론, 인터넷 보안



지 성 택 (Seongtaek Chee) 정회원
1985년 2월: 서강대학교 수학과 학사
1987년 2월: 서강대학교 수학과 석사
1999년 2월: 고려대학교 수학과 박사
1989년 10월~1999년 12월: 한국전자통신연구원 정보보호연구본부 선임연구원
2000년 1월~현재: 국가보안기술연구소 책임연구원
〈관심분야〉 암호론, 부울 함수