情報保護學會論文誌
第19卷 第5號, 2009. 10

스트림 암호에 대한 향상된 고속 상관 공격 적용 가능성 연구*

정 기 태,[1†] 이 유 섭,[1] 성 재 철,[2‡] 홍 석 희[1]
[1]고려대학교 정보보호기술연구원, [2]서울시립대학교 수학과

Study of the Improved Fast Correlation Attack on Stream Ciphers*

Kitae Jeong,[1†] Yuseop Lee,[1] Jaechul Sung,[2‡] Seokhie Hong[1]
[1]Center for Information Security Technologies, Korea University
[2]Department of Mathematics, University of Seoul

요 약

Zhang 등은 SAC'08에서 스트림 암호에 대한 향상된 고속 상관 공격을 제안하였다[8]. 이 공격은 Crypto'00에서 제안된 고속 상관 공격에 기반을 두고 FWT(fast Walsh transform)을 적용하여 설계되었다. [8]에서는 다양한 공격 환경에서 공격 알고리즘의 복잡도와 성공 확률이 제시되었지만, 제안된 공격 알고리즘을 실제 구현한 결과, 제시된 결과와 다르게 나타났다. 본 논문에서는 실험 결과를 토대로 [8]에서 제시된 공격 결과의 문제점을 분석하고, 이 공격 알고리즘이 유효하게 적용되는 bias의 threshold를 제시한다.

ABSTRACT

Zhang et al. proposed a improved fast correlation attack on stream ciphers at SAC'08[8]. This attack is based on the fast correlation attack proposed at Crypto'00 and combined with FWT(fast Walsh transform). Given various attack environments, they presented complexities and success probabilities of the proposed attack algorithm. However, we found that our simulation results of the proposed attack algorithm are different from them presented in [8]. In this paper, we correct results of the proposed attack algorithm by analyzing it theoretically. And we propose a threshold of valid bias.

Keywords: Cryptanalysis, Stream Cipher, Fast Correlation Attack

I. 서 론

상관 공격(correlation attack)은 스트림 암호에 대한 가장 대표적인 공격 기법 중 하나이다. 이 분석 기법은 1985년 Siegenthaler에 의해 그 공격의 개념이 소개되었고[6], 1989년 Meier와 Staffelbach에 의해 향상된 버전인 고속 상관 공격(fast correlation attack)이 제안되었다[5]. 이후, 이 공격의 개념을 이용하여 다수의 LFSR 기반 스트림 암호뿐 아니라, 일반적인 스트림 암호의 안전성 분석에 활용될 수 있도록 상관 공격 알고리즘의 효율성 및 공격 복잡도를 개선시킨 다양한 연구 결과가 발표되었다[1-4,7,8]. 최근에는 LFSR 기반 스트림 암호뿐 아니라 일반적인 스트림 암호의 안전성 분석에 기본적인 분석 방법으로 활용되고 있다.

Zhang 등은 SAC'08에서 스트림 암호에 대한 향상된 고속 상관 공격을 제안하였다[8]. 본 논문에서는 이 공격을 편의상 IFCA(Improved Fast Correlation Attack)라 부르기로 한다. IFCA는 Crypto'00에서 Johansson 등이 제안한 고속 상관 공격[4]에 기반

접수일(2009년 3월 5일), 게재확정일(2009년 8월 6일)
* 이 논문은 2009년도 정부(교육과학기술부)의 재원으로 한국과학재단의 지원을 받아 수행된 연구임. (No. 2009-0060420)
† 주저자, kite@cist.korea.ac.kr
‡ 교신저자, jcsung@uos.ac.kr

〔표 1〕 IFCA와 기제안된 고속 상관 공격의 비교

공격	L	p	N	복잡도		
				실계산	메모리	선계산
IFCA	40	0.531	40,000	2^{20} ($\mathbf{2^{43.04}}$)	2^{25} ($\mathbf{2^{45.48}}$)	$2^{30.6}$ ($\mathbf{2^{34.51}}$)
[3]			80,000	2^{31}	$2^{34.1}$	2^{37}
[7]			2^{22}	2^{24}	$2^{32.8}$	2^{27}

※ ()안은 수정된 결과임.

을 둔다. [4]에서 제안된 고속 상관 공격은 다항식 복구 문제(learning a binary linear multivariate polynomial)를 고속 상관 공격에 적용하여 설계되었다. 하지만, 이 공격은 구성된 패리티 검사 방정식에 키스트림 수열을 대입하고 계산하는데 비효율적이라는 단점을 갖고 있다. 그래서 IFCA에서는 이 문제를 FWT(fast Walsh transform)을 적용함으로서 해결하였다. 또한, 기제안된 공격보다 효율적임을 보이기 위해 길이가 61인 LFSR을 사용하는 shrinking generator에 적용하였다.

본 논문에서는 [8]에서 제시된 공격 결과가 옳지 않음을 보이고, 다양한 공격 환경에서 이 공격이 유효하게 적용되는 bias의 threshold를 제시한다. Zhang 등은 bias와 구성된 패리티 검사 방정식의 수가 작더라도, 효율적으로 LFSR의 초기 상태값을 복구할 수 있다고 주장하였다. 하지만 IFCA를 실제 구현한 결과, 성공 확률이 잘못 계산된 것으로 나타났다. 이는 성공 확률을 계산하는데 사용되는 두 개의 분포인 central chi-square distribution과 noncentral chi-square distribution의 평균값의 차이에서 발생한다. 만약 두 분포의 그래프가 차이가 많이 나면, 높은 성공 확률로 LFSR의 초기 상태값을 복구할 수 있다. 하지만, bias가 작거나 구성된 패리티 검사 방정식의 수가 작으면 두 분포는 거의 차이가 없다. 따라서 잘못 추측된 초기 상태값이 IFCA를 통과할 확률도 증가한다. [8]에서 제시된 공격 환경에서 수정된 공격 결과는 기제안된 고속 상관 공격의 공격 결과보다 비효율적이다 (표 1 참조). 따라서 IFCA가 기제안된 공격보다 효율적이라는 주장은 옳지 않다.

본 논문은 다음과 같이 구성되어 있다. 먼저, 2장에서 IFCA를 소개한다. 3장에서는 [8]에서 제시된 공격 결과의 문제점을 분석하고, 4장에서 다양한 공격 환경에서 이 공격이 유효하게 적용되는 bias의 threshold를 제시한다. 마지막으로 5장에서 결론을 맺는다.

II. IFCA (Improved Fast Correlation Attack)

본 절에서는 IFCA를 소개한다. 본 논문에서는 다음과 같은 표기를 사용한다.

- ⊙ $(a_0, a_1, \cdots)$: LFSR의 출력 수열.
- ⊙ $(z_0, z_1, \cdots)$: 키스트림 수열.
- ⊙ $P(z_i = a_i) = p = 1/2 + \varepsilon$: 상관 확률 $(\varepsilon > 0)$.
- ⊙ L: LFSR의 길이.
- ⊙ $k(< L)$: LFSR의 초기 상태값 중 추측될 비트 수.
- ⊙ t: 패리티 검사 방정식을 구성할 때 사용하는 비트 수.
- ⊙ $q = \frac{1}{2} + 2^{t-1}\varepsilon^t$: t 비트를 이용하여 구성된 패리티 검사 방정식의 bias.
- ⊙ N: 주어진 키스트림 수열의 길이.
- ⊙ $\Omega(\mathbf{v}_{L-k})$: $\mathbf{v}_{L-k}$에 대한 패리티 검사 방정식 개수의 기댓값.
- ⊙ n: $\mathbf{v}_{L-k}$의 개수.

전체 공격 과정은 [표 2]와 같다. 표에서 T는 IFCA의 성공 확률에 의해 결정되는 threshold이다.

LFSR의 출력 수열 $a_i (i = 0, 1, \cdots)$는 LFSR의 초기 상태값 $(a_0, \cdots, a_{L-1})$의 식 (1)과 같은 선형 결합으로 표현 가능하다. 여기서 w_{ij}는 LFSR의 연결 다항식 $g(x) = 1 + g_1 x + g_2 x^2 + \cdots + g_L x^L$에 의해 결정된다.

$$a_i = \bigoplus_{j=0}^{L-1} w_{ij} a_j. \tag{1}$$

$\mathbf{a}_k = (a_0, a_1, \cdots, a_{k-1})$, $\mathbf{a}_{L-k} = (a_k, a_{k+1}, \cdots, a_{L-1})$이고 $\mathbf{1}_t$는 모든 원소가 1인 t차 벡터일 때, 식 (1)을 이용하여 식 (2)와 같은 패리티 검사 방정식을 구성할 수 있

〔표 2〕 IFCA

Parameters: t, k, n
Precomputation
precompute n groups of parity-checks like equation (2) with n different $\mathbf{v}_{L-k}$ values
Input: keystream sequence $(z_0, z_1, \cdots, z_{N-1})$
Processing
let $B_\omega = 0$ for the 2^k possible values of ω
for each group of parity-checks specified by $\mathbf{v}_{L-k}$ do
  let $\mathbf{a}_{L-k}$ take a randomly assigned value
  define a function $h_{\mathbf{v}_{L-k}}(\mathbf{x}_k)$ as in equation (4)
  apply FWT to compute $H_{\mathbf{v}_{L-k}}(\omega)$ for the 2^k possible values of ω
  update $B_\omega = B_\omega + \frac{(H_{\mathbf{v}_{L-k}}(\omega))^2}{4}$ for the 2^k possible values of ω
end for
search for $B_\omega \geq T$ and accept the corresponding ω as a candidate for $\mathbf{a}_k$
Output: $\mathbf{a}_k = (a_0, a_1, \cdots, a_{k-1})$ or a small list of candidates

다. 여기서 '·'은 벡터의 내적이다. 그리고 $\mathbf{a}_t = (a_{i_1}, a_{i_2}, \cdots, a_{i_t})$이고, $i_j (j = 1, \cdots, t)$는 출력 비트 중 임의의 비트를 의미한다. 이때, $\mathbf{v}_{L-k}$는 0이 아닌 임의의 벡터를 의미한다. 따라서 각각의 $\mathbf{v}_{L-k}$에 대해 여러 개의 패리티 검사 방정식을 구성할 수 있다.

$$\mathbf{a}_t \cdot \mathbf{1}_t = (\mathbf{a}_k \cdot \mathbf{x}_k) \oplus (\mathbf{a}_{L-k} \cdot \mathbf{v}_{L-k}). \tag{2}$$

식 (2)에 키스트림 수열을 대입하여, 식 (3)을 구성할 수 있다. 여기서 $\mathbf{a}'_k$는 $\mathbf{a}_k$의 추측값이고, $\mathbf{z}_t = (z_{i_1}, z_{i_2}, \cdots, z_{i_t})$, $\mathbf{e}_t = (e_{i_1}, e_{i_2}, \cdots, e_{i_t})$는 확률 $P(e_{i_j} = 0) = P(a_{i_j} = z_{i_j}) = 1/2 + \varepsilon$으로 $\mathbf{z}_t = \mathbf{a}_t \oplus \mathbf{e}_t$를 만족하는 에러 벡터이다 $(j = 1, \cdots, t)$. 그리고 $\mathbf{a}''_{L-k}$는 $\mathbf{a}_{L-k}$에 할당된 값이며, ζ는 $\mathbf{a}''_{L-k}$에 따라 0 또는 1이다.

$$\begin{aligned}(\mathbf{z}_t \cdot \mathbf{1}_t) \oplus (\mathbf{a}'_k \cdot \mathbf{x}_k) \oplus (\mathbf{a}''_{L-k} \cdot \mathbf{v}_{L-k}) \\ = ((\mathbf{a}_k \oplus \mathbf{a}'_k) \cdot \mathbf{x}_k) \oplus (\mathbf{e}_t \cdot \mathbf{1}_t) \oplus \zeta\end{aligned} \tag{3}$$

선계산 단계에서는, 각각의 $\mathbf{v}_{L-k}$에 대해 $\Omega(\mathbf{v}_{L-k})$개의 패리티 검사 방정식들을 구성한다. IFCA에서 사용하는 $\mathbf{v}_{L-k}$의 개수는 n이다.

실계산 단계에서는, 식 (3)의 좌변을 계산하고 $(\mathbf{z}_t \cdot \mathbf{1}_t) \oplus (\mathbf{a}'_k \cdot \mathbf{x}_k) \oplus (\mathbf{a}''_{L-k} \cdot \mathbf{v}_{L-k}) = 0$을 만족하는 수를 기록한다. 변수에 값을 대입하고 계산하는데 높은 계산 복잡도를 필요로 하기 때문에, 다음과 같은 방법을 사용한다.

각각의 $\mathbf{v}_{L-k}$에 대한 패리티 검사 방정식 집합에 대해, 식 (4)를 정의한다. 단, 패리티 검사 방정식 집합에서 나타나지 않는 $\mathbf{x}_k$에 대해서는 $h_{\mathbf{v}_{L-k}}(\mathbf{x}_k) = 0$이다.

$$h_{\mathbf{v}_{L-k}}(\mathbf{x}_k) = \sum_{\mathbf{x}_k} (-1)^{(\mathbf{z}_t \cdot \mathbf{1}_t) \oplus (\mathbf{a}''_{L-k} \cdot \mathbf{v}_{L-k})} \tag{4}$$

$h_{\mathbf{v}_{L-k}}(\mathbf{x}_k)$의 Walsh 변환을 식 (5)와 같이 정의한다. 각각의 $\mathbf{v}_{L-k}$에 대해, 2^k번의 Walsh 변환을 계산해야 하는데 FWT(fast Walsh transform)을 이용하여 효율적으로 계산할 수 있다.

$$\begin{aligned}H_{\mathbf{v}_{L-k}}(\omega) &= \sum_{\mathbf{x}_k \in Z_2^k} h_{\mathbf{v}_{L-k}}(\mathbf{x}_k)(-1)^{\mathbf{x}_k \cdot \omega} \\ &= \sum_{\Omega(\mathbf{v}_{L-k})} (-1)^{(\mathbf{z}_t \cdot \mathbf{1}_t) \oplus (\mathbf{a}''_{L-k} \cdot \mathbf{v}_{L-k}) \oplus (\mathbf{x}_k \cdot \omega)}\end{aligned} \tag{5}$$

공격자에게 N-비트 키스트림 수열이 주어졌을 때, IFCA의 선계산 복잡도는 $N^{\lceil t/2 \rceil} \cdot \log_2 N$이고, 실계산 복잡도는 $\sum_{\mathbf{v}_{L-k}} (2^k k + \Omega(\mathbf{v}_{L-k})(t+k))$이다. 그리고 메모리 복잡도는 $c \cdot 2^k + \sum_{\mathbf{v}_{L-k}} (t \lceil \log_2 N \rceil + L)\Omega(\mathbf{v}_{L-k})$ 비트이다.

III. IFCA 분석

3.1 성공 확률

고정된 $\mathbf{v}_{L-k}$에 대한 패리티 검사 방정식의 개수 $\Omega(\mathbf{v}_{L-k})$는 $\binom{N}{t}2^{k-L}$이다. 따라서 식 (3)으로부터, $\mathbf{a}'_k$가 옳게 추측되었다면 $(\mathbf{z}_t \cdot \mathbf{l}_t) \oplus (\mathbf{a}'_k \cdot \mathbf{x}_k) \oplus (\mathbf{a}''_{L-k} \cdot \mathbf{v}_{L-k}) = 0$을 만족하는 개수가 $\frac{1}{2}\Omega(\mathbf{v}_{L-k})$로부터 $\Omega(\mathbf{v}_{L-k})2^{t-1}\varepsilon^t$만큼 차이가 난다. 그렇지 않으면, 차이가 발생하지 않는다.

B_ω를 갱신하는데 사용되는 $(H_{\mathbf{v}_{L-k}}(\omega))^2/4$는 다음과 같은 식에 의해 유도된다. 여기서 $u(\mathbf{v}_{L-k})$는 추측된 $\mathbf{a}'_k$에 대해, $(\mathbf{z}_t \cdot \mathbf{l}_t) \oplus (\mathbf{a}'_k \cdot \mathbf{x}_k) \oplus (\mathbf{a}''_{L-k} \cdot \mathbf{v}_{L-k}) = 0$을 만족하는 개수를 의미한다.

$$\begin{aligned}
&\sum_{\mathbf{v}_{L-k}} \left(u(\mathbf{v}_{L-k}) - \frac{\Omega(\mathbf{v}_{L-k})}{2} \right)^2 \\
&= \sum_{\mathbf{v}_{L-k}} \left(\frac{|H_{\mathbf{v}_{L-k}}(\omega)| + \Omega(\mathbf{v}_{L-k})}{2} - \frac{\Omega(\mathbf{v}_{L-k})}{2} \right)^2 \\
&= \sum_{\mathbf{v}_{L-k}} \frac{(H_{\mathbf{v}_{L-k}}(\omega))^2}{4},
\end{aligned}$$

그래서 $B_\omega \geq T$는 식 (6)과 같다.

$$\begin{aligned}
&B_\omega \geq T \\
\Leftrightarrow &\sum_{\mathbf{v}_{L-k}} \left(u(\mathbf{v}_{L-k}) - \frac{\Omega(\mathbf{v}_{L-k})}{2} \right)^2 \geq T
\end{aligned} \tag{6}$$

만약 $\mathbf{a}'_k$가 옳게 추측되었다면, $u(\mathbf{v}_{L-k})$는 이항 분포 $B(\Omega(\mathbf{v}_{L-k}), q)$를 따르고 그렇지 않으면 이항 분포 $B(\Omega(\mathbf{v}_{L-k}), \frac{1}{2})$를 따른다. 따라서 $\mathbf{a}'_k$가 옳게 추측되었을 경우, 식 (6)은 아래와 같은 식 (7)과 같이 표현된다. 식 (7)은 $\mathbf{a}'_k$가 옳게 추측되었을 경우 $\sum_{\mathbf{v}_{L-k}} \frac{(u(\mathbf{v}_{L-k}) - \Omega(\mathbf{v}_{L-k})/2)^2}{\Omega(\mathbf{v}_{L-k})q(1-q)}$가 noncentral chi-square distribution을 따름을 의미한다.

이에 반해, $\mathbf{a}'_k$가 잘못 추측되었을 경우, 식 (6)은 식 (8)과 같이 표현된다. 식 (8)은 $\mathbf{a}'_k$가 잘못 추측되었을 경우 $\sum_{\mathbf{v}_{L-k}} \frac{(u(\mathbf{v}_{L-k}) - \Omega(\mathbf{v}_{L-k})/2)^2}{\left(\sqrt{\Omega(\mathbf{v}_{L-k})}/2\right)^2}$가 central chi-square distribution을 따름을 의미한다.

$$\begin{aligned}
&\frac{\Omega(\mathbf{v}_{L-k})n}{4q(1-q)} \geq \sum_{\mathbf{v}_{L-k}} \frac{\left(u(\mathbf{v}_{L-k}) - \frac{\Omega(\mathbf{v}_{L-k})}{2} \right)^2}{\Omega(\mathbf{v}_{L-k})q(1-q)} \geq \frac{T}{\Omega(\mathbf{v}_{L-k})q(1-q)} \\
\Leftrightarrow &\frac{\Omega(\mathbf{v}_{L-k})n}{4q(1-q)} \geq \sum_{\mathbf{v}_{L-k}} \left(\frac{u(\mathbf{v}_{L-k}) - \Omega(\mathbf{v}_{L-k})q + \Omega(\mathbf{v}_{L-k})2^{t-1}\varepsilon^t}{\sqrt{\Omega(\mathbf{v}_{L-k})q(1-q)}} \right)^2 \geq \frac{T}{\Omega(\mathbf{v}_{L-k})q(1-q)} \\
\Leftrightarrow &\frac{\Omega(\mathbf{v}_{L-k})n}{4q(1-q)} \geq \sum_{\mathbf{v}_{L-k}} \left(\frac{u(\mathbf{v}_{L-k}) - \Omega(\mathbf{v}_{L-k})q}{\sqrt{\Omega(\mathbf{v}_{L-k})q(1-q)}} + \frac{\Omega(\mathbf{v}_{L-k})2^{t-1}\varepsilon^t}{\sqrt{\Omega(\mathbf{v}_{L-k})q(1-q)}} \right)^2 \geq \frac{T}{\Omega(\mathbf{v}_{L-k})q(1-q)}
\end{aligned} \tag{7}$$

$$\begin{aligned}
&\Omega(\mathbf{v}_{L-k})n \\
&\geq \sum_{\mathbf{v}_{L-k}} \frac{\left(u(\mathbf{v}_{L-k}) - \frac{\Omega(\mathbf{v}_{L-k})}{2} \right)^2}{\left(\frac{1}{2}\sqrt{\Omega(\mathbf{v}_{L-k})} \right)^2} \geq \frac{4T}{\Omega(\mathbf{v}_{L-k})}.
\end{aligned} \tag{8}$$

따라서 옳게 추측된 $\mathbf{a}'_k$가 $B_{\mathbf{a}'_k} \geq T$를 만족할 확률인 P_{right}와 잘못 추측된 $\mathbf{a}'_k$가 알고리즘을 통과할 확률인 P_{wrong}은 식 (9)와 같다.

$$\begin{aligned}
P_{\text{right}} &= \int_{\frac{T}{\Omega(\mathbf{v}_{L-k})q(1-q)}}^{\frac{\Omega(\mathbf{v}_{L-k})n}{4q(1-q)}+0.5} \phi_2(x)dx, \\
P_{\text{wrong}} &= \int_{\frac{4T}{\Omega(\mathbf{v}_{L-k})}}^{\frac{\Omega(\mathbf{v}_{L-k})}{2}+0.5} \phi_1(x)dx.
\end{aligned} \tag{9}$$

여기서 $x > 0$에 대해 central chi-square distribution의 확률 밀도 함수 $\phi_1(x)$와 noncentral chi-square distribution의 확률 밀도 함수 $\phi_2(x)$는 다음과 같다.

$$\begin{aligned}
\phi_1(x) &= \frac{x^{\frac{n-2}{2}} e^{-\frac{x}{2}}}{2^{\frac{n}{2}}\Gamma(n/2)}, \\
\phi_2(x) &= \frac{e^{-\frac{(x+\delta^2)}{2}}}{2^{n/2}} \sum_{j=0}^{\infty} \frac{x^{j-1+\frac{n}{2}\delta^{2j}}}{\Gamma\left(j+\frac{n}{2}\right)2^{2j}j!},
\end{aligned}$$

$$\Gamma(y)=\int_0^{+\infty} e^{-x}x^{y-1}dx,$$

$$\delta^2=\sum_{\mathbf{v}_{L-k}}\left(\frac{\sqrt{\Omega(\mathbf{v}_{L-k})}\,2^{t-1}\varepsilon^t}{\sqrt{q(1-q)}}\right)^2.$$

$P_{\text{wrong}}<2^{-k}$가 되도록 threshold T를 적절히 선택한다. 이는 잘못 추측된 $\mathbf{a}'_k$는 IFCA를 통과하지 못하고, 옳게 추측된 $\mathbf{a}'_k$는 적당한 확률로 통과함을 의미한다. IFCA의 성공 확률을 P_{right}로 정의한다.

3.2 IFCA 분석

[8]에서는 공격자에게 주어진 다양한 환경에서 $P_{\text{right}}\approx 1$, $P_{\text{wrong}}<2^{-k}$를 만족하도록 매개 변수를 조절하여 공격 복잡도를 조절하였다. 예를 들어, LFSR의 길이가 40인 스트림 암호에 대해 주어진 키스트림 수열의 길이가 40,000일 경우 $2^{30.6}$의 선계산 복잡도와 2^{20}의 실계산 복잡도, 2^{25}의 메모리 복잡도로 LFSR의 초기 상태값을 복구할 수 있다. [표 3]과 [표 4]는 [8]에서 제시된 공격 환경인 $L=40$, $N=40{,}000$, $t=3$, $k=12$일 경우의 공격 결과를 실제 구현한 결과와 비교한 것이다. 구현 결과는 MATLAB R2008a를 이용하여 계산되었다. [표 3]에서는 다양한 상관 확률에 대해 $P_{\text{right}}\approx 1$, $P_{\text{wrong}}<2^{-12}$를 만족하도록 n과 T를 조절하여 복잡도를 계산하였고, [표 4]에서는 [8]에서 제시된 공격 복잡도와 유사한 복잡도를 갖도록 n과 T를 조절하였다. 표에서 알 수 있듯이, [8]에서 제시된 공격 결과와 실제 구현 결과가 많이 다름을 알 수 있다.

P_{wrong}의 분포인 central chi-square distribution의 평균은 자유도로서 IFCA의 경우는 n이다. 이에 반해 P_{right}의 분포인 noncentral chi-square distribution의 평균은 $n+\delta^2$이다. 따라서 $P_{\text{right}}\approx 1$, $P_{\text{wrong}}<2^{-k}$를 만족하려면, δ^2이 큰 값이어야 한다. δ^2는 n과 ε의 영향을 많이 받으므로, δ^2이 큰 값이 되려면 이 두 값이 큰 값이 되어야 한다. 그림 1과 그림 2는 $L=40$, $N=40{,}000$, $k=12$, $t=3$, $\varepsilon=0.031$일 때 $n=10$[표 4]인 경우와 $n=2^{23.74}$[표 3]인 경우의 그래프를 나타낸 것이다. 그림에서 점선의 그래프가 noncentral chi-square distribution을 의미한다. 그림에서 알 수 있듯이, $n=10$인 경우 두 그래프

[표 3] IFCA의 구현 결과 1 (기준: $P_{\text{right}}\approx 1$, $P_{\text{wrong}}<2^{-k}$)

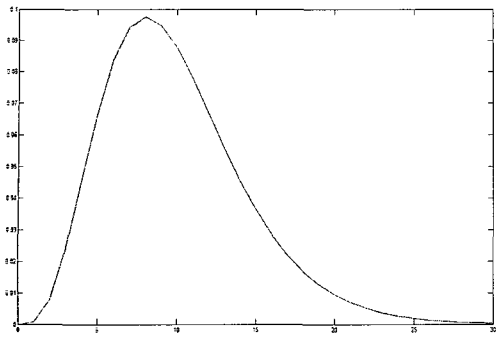
공격	p	N	n	T	P_{right}	P_{wrong}	복잡도		
							실계산	메모리	선계산
[8]	0.531	40,000	·	·	≈ 1	$<2^{-k}$	$2^{20.00}$	$2^{25.00}$	$2^{30.60}$
구현 결과	0.650	40,000	2	$2^{17.34}$	0.9999	2^{-12}	$2^{20.29}$	$2^{22.76}$	$2^{34.51}$
	0.600	40,000	29	$2^{19.26}$	0.9902		$2^{24.16}$	$2^{26.60}$	$2^{34.51}$
	0.550	40,000	$2^{15.46}$	$2^{28.77}$	0.9910		$2^{34.76}$	$2^{37.20}$	$2^{34.51}$
	0.531	40,000	$2^{23.74}$	$2^{37.02}$	0.9934		$2^{43.04}$	$2^{45.48}$	$2^{34.51}$
		50,000	$2^{21.9}$	$2^{36.14}$	0.9960		$2^{42.11}$	$2^{44.60}$	$2^{35.18}$
		100,000	$2^{15.87}$	$2^{33.15}$	0.9942		$2^{39.03}$	$2^{41.62}$	$2^{39.27}$

매개 변수: $L=40$, $t=3$, $k=12$

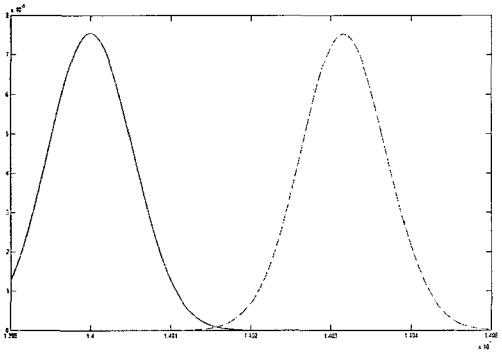
[표 4] IFCA의 구현 결과 2 (기준: 복잡도)

공격	p	N	n	T	P_{right}	P_{wrong}	복잡도		
							실계산	메모리	선계산
[8]	0.531	40,000	·	·	≈ 1	$<2^{-k}$	$2^{20.00}$	$2^{25.00}$	$2^{30.60}$
구현 결과	0.531	40,000	1	133,700	$2^{-11.98}$	2^{-12}	$2^{19.30}$	$2^{21.79}$	$2^{34.51}$
			2	165,247	$2^{-11.97}$		$2^{20.30}$	$2^{22.76}$	$2^{34.51}$
			4	214,254	$2^{-11.97}$		$2^{21.30}$	$2^{23.75}$	$2^{34.51}$
			6	256,132	$2^{-11.96}$		$2^{21.88}$	$2^{24.33}$	$2^{34.51}$
			8	294,479	$2^{-11.96}$		$2^{22.30}$	$2^{24.74}$	$2^{34.51}$
			10	330,608	$2^{-11.96}$		$2^{22.62}$	$2^{25.06}$	$2^{34.51}$
			12	365,170	$2^{-11.96}$		$2^{22.88}$	$2^{25.33}$	$2^{34.51}$

매개변수: $L=40$, $t=3$, $k=12$



〔그림 1〕 $n=10$인 경우



〔그림 2〕 $n=2^{23.74}$인 경우

가 거의 차이가 없기 때문에 P_{right}와 P_{wrong}의 차이가 거의 없다. 이에 반해, $n=2^{23.74}$인 경우 두 그래프가 확연히 차이나기 때문에 P_{right}와 P_{wrong}도 큰 차이가 난다.

식 (9)에서 P_{right}의 $\frac{T}{\Omega(\mathbf{v}_{L-k})q(1-q)}$와 P_{wrong}의 $\frac{4T}{\Omega(\mathbf{v}_{L-k})}$은 실험 결과 거의 같은 값으로 나타났다. 따라서 〔8〕에서 제시하는 성공 기준인 $P_{\text{right}}\approx 1$, $P_{\text{wrong}}<2^{-k}$을 만족하기 위해서는 큰 δ^2이 필요하다. 한편, T를 P_{right}의 $T/(\Omega(\mathbf{v}_{L-k})q(1-q))$와 P_{wrong}의 $4T/\Omega(\mathbf{v}_{L-k})$가 두 그래프의 평균의 평균인 $(n+(n+\delta^2))/2$가 되도록 설정하면, $P_{\text{right}}\approx 1$, $P_{\text{wrong}}<2^{-k}$을 거의 만족하는 것으로 나타났다.

IV. IFCA의 적용 가능성

본 절에서는 IFCA가 유효하게 적용되는 상관 확률의 threshold를 제안한다. 먼저, 앞 절에서 고려한 공격 환경($L=40$, $N=40{,}000$, $k=12$, $t=3$)에서의 적용 가능한 상관 확률을 살펴본 후, 길이가 61인 LFSR을 사용하는 shrinking generator에 적용 가능한 상관 확률을 살펴본다.

4.1 $L=40$, $N=40{,}000$, $k=12$, $t=3$

〔표 5〕는 $L=40$, $N=40{,}000$, $k=12$, $t=3$ 환경에서 다양한 상관 확률에 따른 공격 복잡도를 나타낸 것이다. 여기서 T는 앞 절에서 언급하였듯이 P_{right}의 $T/(\Omega(\mathbf{v}_{L-k})q(1-q))$와 P_{wrong}의 $4T/\Omega(\mathbf{v}_{L-k})$이 두 그래프의 평균의 평균인 $(n+(n+\delta^2))/2$가 되도록 설정하였다.

표에서 알 수 있듯이, $\varepsilon\le 0.10$일 경우 IFCA가 유효하게 적용되기 위한 n이 급격하게 커진다. 따라서 이 공격 환경에서 IFCA는 $\varepsilon\ge 0.10$일 경우에만 유효

〔표 5〕 $L=40$, $N=40{,}000$, $k=12$, $t=3$ 환경에서의 적용 가능성

ε	n	δ^2	T	P_{right}	P_{wrong}	복잡도		
						실계산	메모리	선계산
0.15	2	57.97	307,800	0.9831	$2^{-22.35}$	$2^{20.29}$	$2^{22.76}$	$2^{34.51}$
0.14	3	57.47	315,230	0.9821	$2^{-20.68}$	$2^{20.88}$	$2^{23.34}$	$2^{34.51}$
0.13	4	49.11	283,660	0.9719	$2^{-16.67}$	$2^{21.3}$	$2^{23.75}$	$2^{34.51}$
0.12	5	37.97	238,270	0.9484	$2^{-12.16}$	$2^{21.62}$	$2^{24.07}$	$2^{34.51}$
	7	53.16	333,570	0.9749	$2^{-15.57}$	$2^{22.11}$	$2^{24.55}$	$2^{34.51}$
0.11	5	22.53	161,550	0.8798	$2^{-7.35}$	$2^{21.62}$	$2^{24.07}$	$2^{34.51}$
	14	63.08	452,350	0.9814	$2^{-14.87}$	$2^{23.11}$	$2^{25.55}$	$2^{34.51}$
0.10	5	12.72	112,820	0.7824	$2^{-4.48}$	$2^{21.62}$	$2^{24.07}$	$2^{34.51}$
	29	73.75	654,360	0.9842	$2^{-13.17}$	$2^{24.16}$	$2^{26.6}$	$2^{34.51}$
0.09	5	6.76	83,228	0.6727	0.14	$2^{21.62}$	$2^{24.07}$	$2^{34.51}$
	75	101.36	1,248,400	0.9899	$2^{-12.13}$	$2^{25.53}$	$2^{27.97}$	$2^{34.51}$

〔표 6〕 길이가 61인 LFSR을 사용하는 shrinking generator에 대한 적용 가능성

공격	ε	δ^2	T	P_{right}	P_{wrong}	복잡도	
						실계산	메모리
[8]	0.0195281	-	$8.6*10^8$	0.9742	$2^{-32.16}$	$2^{35.86}$	$2^{36.23}$
구현 결과	0.0195281	$4.8*10^{-6}$	$8.6*10^8$	$2^{-32.16}$	$2^{-32.16}$	$2^{35.85}$	$2^{36.23}$
	0.10482	95.34	$8.6*10^8$	0.9742	$2^{-32.16}$	$2^{35.85}$	$2^{36.23}$

매개 변수: $L=61$, $N=10{,}000$, $n=12$, $t=5$, $k=27$

하게 적용된다고 볼 수 있다.

4.2 길이가 61인 LFSR을 사용하는 shrinking generator

Zhang 등은 기제안된 고속 상관 공격과 효율성을 비교하기 위해 [8]에서 길이가 61인 LFSR을 사용하는 shrinking generator에 대해 IFCA를 적용하였다. [8]에서는 상관 확률이 0.5195281이고 주어진 키스트림 수열의 길이가 10,000일 때, $2^{35.86}$의 실계산 복잡도로 LFSR의 초기 상태값을 복구할 수 있다고 주장하였다. 여기서 $P_{right}=97.42\%$, $P_{wrong}=2^{-32.16}$이다.

하지만 동일한 공격 환경에서 구현한 결과, P_{right}가 97.42%가 아닌 $2^{-32.16}$인 것으로 나타났다 (표 6 참조). 만약 상관 확률이 0.60482이면, [8]에서 제시한 공격 결과와 거의 유사한 것으로 나타났다. 따라서 길이가 61인 LFSR을 사용하는 shrinking generator에 대해 IFCA가 유효하게 적용되려면 $\varepsilon \geq 0.1$을 만족해야 한다.

V. 결 론

본 논문에서는 [8]에서 제안된 향상된 고속 상관 공격의 성공 확률 계산이 잘못되었음을 보이고, 그 원인을 분석하였다. 또한, 이 공격이 유효하게 적용되는 bias의 threshold를 제시하였다. 실험 결과, 이 공격 알고리즘은 $\varepsilon \geq 0.10$인 경우에만 유효하게 적용되는 것으로 나타났다. 이를 통해 [8]에서는 제안된 고속 상관 공격이 기제안된 공격보다 훨씬 효율적이라고 주장하였으나, 이 주장이 옳지 않음을 알 수 있다.

참 고 문 헌

[1] 김현, 홍석희, 성재철, 이상진, 박해룡, 전길수, "스트림 암호에 대한 개선된 다중 경로 고속 상관 공격," 정보보호학회논문지, 17(4), pp. 53-60, 2007년 8월.

[2] 정기태, 성재철, 홍석희, 이상진, 김재헌, 박상우, "Shrinking 생성기와 Self-Shrinking 생성기에 대한 향상된 고속 상관 공격," 정보보호학회논문지, 16(2), pp. 25-32, 2006년 4월.

[3] P. Chose, A. Joux, and M. Mitton, "Fast Correlation Attacks: An Algorithmic Point of View," Eurocrypt'02, LNCS 2332, pp. 209-221, 2002.

[4] T. Johansson and F. Jönsson, "Fast Correlation Attacks through Reconstruction of Linear Polynomials," Crypto'00, LNCS 1880, pp. 300-315, 2000.

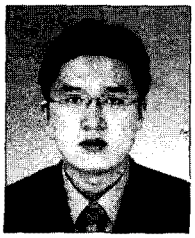
[5] W. Meier and O. Staffelbach, "Fast correlation attacks on certain stream ciphers," Journal of Cryptology, vol. 1, no. 3, pp. 159-176, Oct. 1989.

[6] T. Siegenthaler, "Decrypting a class of stream ciphers using ciphertext-only," IEEE Transactions on Computers, vol. 34, no. 1, pp. 81-85, Jan. 1985.

[7] B. Zhang and D. Feng, "Multi-pass fast correlation attack on stream ciphers," SAC'06, LNCS 4356, pp. 234-248, 2007.

[8] B. Zhang and D. Feng, "An Improved Fast Correlation Attack on Stream Ciphers," SAC 2008, LNCS 5381, 2009.

〈著者紹介〉



정 기 태 (Kitae Jeong) 학생회원
2004년 2월: 고려대학교 수학과 학사
2006년 2월: 고려대학교 정보보호대학원 석사
2006년 3월 ~ 현재: 고려대학교 정보경영공학전문대학원 박사과정
〈관심분야〉 블록 암호, 스트림 암호 및 해쉬 함수의 분석 및 설계

이 유 섭 (Yuseop Lee) 학생회원
2007년 2월: 서울시립대학교 수학과 학사
2007년 3월 ~ 현재: 고려대학교 정보경영공학전문대학원 석박사 통합과정
〈관심분야〉 스트림 암호 및 해쉬 함수의 분석 및 설계

성 재 철 (Jaechul Sung) 종신회원
1997년 8월: 고려대학교 수학과 학사
1999년 8월: 고려대학교 수학과 석사
2002년 8월: 고려대학교 수학과 박사
2002년 8월 ~ 2004년 1월: 한국정보보호진흥원 선임연구원
2004년 2월 ~ 현재: 서울시립대학교 수학과 조교수
〈관심분야〉 암호 알고리즘 설계 및 분석

홍 석 희 (Seokhie Hong) 종신회원
1995년 2월: 고려대학교 수학과 학사
1997년 2월: 고려대학교 수학과 석사
2001년 2월: 고려대학교 수학과 박사
1999년 8월 ~ 2004년 2월: (주) 시큐리티 테크놀로지스 선임연구원
2003년 2월 ~ 2004년 2월: 고려대학교 정보보호기술연구센터 선임연구원
2004년 4월 ~ 2005년 2월: K.U.Leuven, ESAT/SCD-COSIC 박사후연구원
2005년 3월 ~ 2008년 8월: 고려대학교 정보보호대학원 조교수
2008년 9월 ~ 현재: 고려대학교 정보경영공학전문대학원 부교수
〈관심분야〉 대칭키 암호의 분석 및 설계, 컴퓨터 포렌식