논문 16-41-12-34 The Journal of Korean Institute of Communications and Information Sciences '16-12 Vol.41 No.12
https://doi.org/10.7840/kics.2016.41.12.1892

NIST SP 800-90B 프레딕터를 이용한 잡음원의 엔트로피 추정량에 대한 실험적 분석

박 호 중◆, 배 민 영*, 염 용 진**, 강 주 성°

An Experimental Analysis on Entropy Estimators for the Entropy Sources Using Predictors of NIST SP 800-90B

Hojoong Park◆, Minyoung Bae*, Yongjin Yeom**, Ju-Sung Kang°

요 약

잡음원(Noise source)의 안전성 평가에 사용되는 대표적인 표준으로는 미국 NIST의 SP 800-90B가 있다. 최근 SP 800-90B가 2차 안(Second Draft)으로 개정되면서 Non-IID 트랙의 최소 엔트로피 추정에 프레딕터(predictor)를 이용한 추정 방법이 새롭게 추가되었다. 프레딕터는 잡음원의 주기적인 특성을 검출하기에 용이하다고 알려져 있지만, 그 특성에 대한 구체적인 언급은 하지 않고 있다. 이에 본 논문에서는 프레딕터가 검출해낼 수 있는 잡음원의 주기적 특성을 명확히 밝히기 위한 실험을 진행한다. 먼저 주기적 성질을 갖는 잡음원에 대하여 Non-IID 트랙의 추정을 실시했을 때, 잡음원의 최소 엔트로피가 대체적으로 프레딕터보다는 Non-IID 트랙의 다른 추정 방법에 의해서 결정되고 있음을 실험적으로 확인한다. 다음으로 프레딕터를 이용한 추정법이 검출해낼 수 있는 주기적 특성을 밝혀내기 위한 다양한 실험 결과를 제시함으로써, 프레딕터 추정 방법의 의미와 그 역할을 실험적으로 규명한다.

Key Words : Predictor, SP 800-90B, Entropy estimation, Noise source, Non-IID test

ABSTRACT

NIST SP 800-90B is developed to evaluate the security of entropy sources. As SP 800-90B was updated to Second Draft, Estimators with predictors were added at Non-IID track. Though the predictors are known as detecting periodic property of noise sources, periodic properties which are detected by predictor are not clearly known. In this paper, we experiment to find properties of predictors. Once, by experiments we have a result that the min-entropy of Non-IID noise sources is generally determined by tests except for estimators with predictors. And then through presenting various experimental results for clarifying periodic properties detected by predictor, we experimentally analyze on its meaning and role of predictor estimation.

※ 본 연구는 2016년도 정부(미래창조과학부)의 재원으로 한국연구재단-차세대정보·컴퓨팅기술 개발사업의 지원을 받아 수행되었습니다.(NO. NRF-2014M3C4A7030648)
◆ First Author : Kookmin University Department of Financial Information Security, ruokay@kookmin.ac.kr, 학생회원
° Corresponding Author : Kookmin University Department of Math. / Financial Information Security, jskang@kookmin.ac.kr, 정회원
* Kookmin University Department of Financial Information Security, mypear@kookmin.ac.kr, 학생회원
** Kookmin University Department of Math. / Financial Information Security, salt@kookmin.ac.kr, 종신회원
논문번호 : KICS2016-08-214, Received August 30, 2016; Revised November 8, 2016; Accepted November 25, 2016

Ⅰ. 서 론

난수발생기에서 안전한 난수를 출력하기 위해 잡음원(Noise source 혹은 Entropy source)의 안전성 평가는 필수적이다. 잡음원에 대한 안전성 평가의 대표적인 표준으로는 미국 국립표준기술연구소(National Institute of Standards and Technology, 이하 NIST)에서 2012년도에 공표한 SP 800-90B가 있다[1]. SP 800-90B는 잡음원의 안전성 평가를 위해 최소 엔트로피(Min-entropy)를 측도로 사용하여 잡음원의 엔트로피를 보수적(conservative)으로 측정하고 있다. 또한 암호모듈 검증제도(Cryptographic Module Validation Program, CMVP)에서 난수발생기의 안전성 평가에 활용[2,3] 및 암호학적으로 안전한 난수발생기 설계를 위한 국제 표준에 반영되는 등 중요한 표준문서로 분류되고 있다.

한편, 2016년 1월 27일에 NIST에서는 기존 SP 800-90B를 보완한 2차 안(Second Draft)을 발표하였다[4]. SP 800-90B가 2차 안으로 업데이트 되면서, 기존의 엔트로피 추정 방법보다 복잡하고 더욱 엄밀하게 개정되었다. 기존 문서와 2차 안의 차이를 간단히 요약하면 표 1과 같다.

표 1. SP 800-90B 기존문서와 개정된 문서의 비교
Table 1. Comparison of 1st Draft with 2nd Draft

	1st Draft	2nd Draft
Entropy source Model	· Noise source · Health tests · (Optional) Conditioning component	· Add a (Optional) Post-processing
Determine IID/Non-IID	· 6 tests → 11 scores	· 11 tests → 11 scores
Estimate Non-IID	· 5 tests	· 10 tests
Additional statistical test	· Goodness of fit test · Independence test	· Add a Longest Repeated Sub -string(LRS) estimate

1.1 프레딕터 특성 분석에 대한 연구 배경

SP 800-90B가 개정되면서 기존과의 가장 큰 변화는 Non-IID 잡음원 엔트로피 추정에 프레딕터(predictor) 추정 방법이 추가된 점이다. 프레딕터를 이용하여 엔트로피를 추정하는 것은 최초 Shannon에 의해 도입되었다[5]. 이후 프레딕터는 잡음원에 대한 엔트로피 추정보다는 데이터 압축[6], 언어 분석[7], 학습 모델[8] 등 다른 용도로 주로 사용되었다. 한편, NIST에서는 프레딕터를 이용하여 잡음원의 엔트로피를 추정하는 논문 “Predictive models for min-entropy estimation”을 2015년에 발표하였다[9]. 논문에서는 프레딕터 추정 방법이 개정될 SP 800-90B에 추가될 것을 언급하였으며, 논문에 제시된 프레딕터의 일부가 개정 문서에 반영되었다.

SP 800-90B의 프레딕터 추정 방법은 Non-IID로 판정된 잡음원의 종속성을 확인하는데 사용된다. 특히 프레딕터는 잡음원의 주기적 성질을 검출하기 위해 추가되었다고 알려져 있지만, SP 800-90B[4]와 논문[9]에서는 프레딕터가 검출하는 주기적 종속성에 대한 구체적인 언급은 하고 있지 않다.

1.2 논문의 주요결과

본 논문은 NIST SP 800-90B Non-IID 트랙의 엔트로피 추정 방법인 프레딕터에 대한 연구로, 프레딕터 추정 방법이 검출해낼 수 있는 잡음원의 주기적인 특성을 실험을 통하여 확인하고, 이를 통해 SP 800-90B에서 프레딕터 추정 방법의 의미와 역할을 규명하였다. 다양한 실험을 통하여 얻은 논문의 주요결과를 요약하면 다음과 같다.

- 프레딕터 추정 방법은 주기적인 특성이 강할수록 잡음원 검출에 유리하다.
- 프레딕터 추정 방법은 주기적인 종속성을 가지는 잡음원 검출보다는 주기적인 값이 반복되는 특성을 검출해내는 데 유리하다.
- 주기적으로 같은 값을 출력하는 잡음원에 대하여 프레딕터 추정 방법은 최소 엔트로피 추정에 기존 방법보다 결정적인 역할을 한다.

또한 주요결과를 바탕으로 프레딕터를 활용한 잡음원 분석 및 암호학적으로 안전한 난수발생기를 설계하는 데 활용되는 등 안전한 난수를 요구하는 환경[10]에 본 연구결과가 활용될 것으로 기대한다.

본 논문은 2장 SP 800-90B에 대한 개요 및 잡음원의 엔트로피 추정 과정, 3장 프레딕터의 개요 및 이론적 근거, 4장 프레딕터 특성에 분석에 대한 실험 및 결과, 마지막 5장 결론으로 구성되어있다.

Ⅱ. SP 800-90B

2.1 SP 800-90B 개요

SP 800-90B는 잡음원의 안전성 평가를 위한 표준

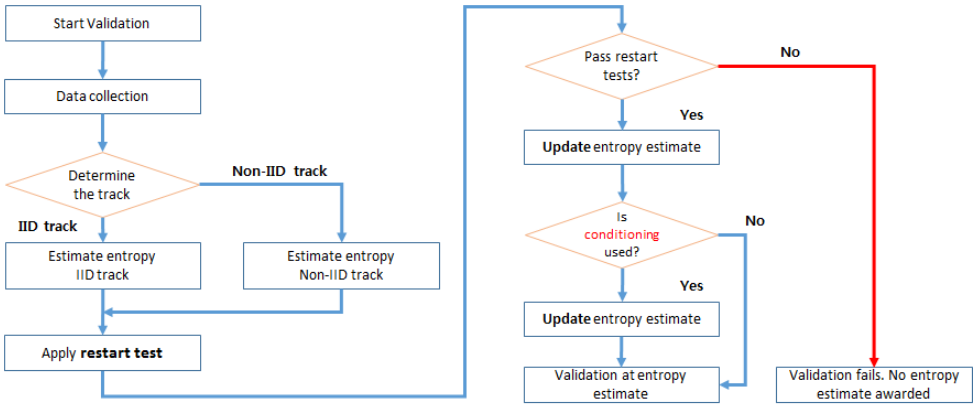

그림 1. 잡음원에 대한 SP 800-90B의 엔트로피 추정 과정
Fig. 1. Process of SP 800-90B Second Draft

으로, 암호모듈 검증제도에서 난수발생기의 안전성 평가에도 활용되고 있다[11,12]. 그림 1은 SP 800-90B의 잡음원에 대한 엔트로피 추정 과정을 도식화한 것으로, SP 800-90B 엔트로피 추정 방법은 가설검정 원리를 이용하여 잡음원의 통계적 특성인 IID(Independent and Identically distributed)/Non-IID를 확인하는 단계(Determine the track), IID 여부에 따라 잡음원의 엔트로피를 추정하는 단계(Estimate entropy IID track 또는 Non-IID track), 잡음원의 엔트로피 과추정(Overestimate)을 막는 단계(Restart test), 엔트로피 출력비율을 결정하는 단계(Conditioning) 순서로 진행된다. 개정된 SP 800-90B에는 기존에 존재하지 않았던 재시작 검사(Restart test)의 추가와 보다 엄격해진 후처리(Conditioning) 기준으로 잡음원의 엔트로피를 더욱 보수적으로 추정하고 있다.

2.2 Non-IID 트랙의 엔트로피 추정

Non-IID로 판정된 잡음원은 Non-IID 트랙에서 엔트로피를 추정하게 된다. SP 800-90B가 개정되면서 Non-IID로 판정된 잡음원의 엔트로피를 추정하는 방법이 두드러지게 변화하였다. 기존에 사용했던 빈도수(Frequency), 충돌(Collision), 마코브(Markov), 부분수집(Partial Collection), 압축(Compression)의 다섯 가지 추정 방법에서, 빈도수와 부분 수집 추정 방법이 제외되고 최빈값(The Most Common Value), t-쌍(t-Tuple), 반복부분난수열(LRS), 4 종류의 프레딕터(predictor) 추정 방법이 추가되었다. 이때 Non-IID 트랙에서는 잡음원의 종속성을 확인하는 10 가지 엔트로피 추정치 중 최솟값을 잡음원의 최소 엔트로피로 추정하고 있다. 특히, Non-IID 트랙의 10 가지 엔트로피 추정 방법 중 새롭게 추가된 윈도우 내 최빈값(Multi most common in window prediction estimate), 래그(The lag prediction estimate), 다중 마코브(The multiMMC prediction estimate), LZ78Y 압축(The LZ78Y prediction estimate)을 이용한 추정 방법은 본 논문에서 그 특성을 규명하고자 하는 프레딕터 추정 방법으로 다음 장에서 다룬다.

Ⅲ. 프레딕터

프레딕터(predictor)는 SP 800-90B가 업데이트 되면서 새롭게 추가된 Non-IID 트랙의 엔트로피 추정 방법이다. 프레딕터란 Shannon에 의해 도입된 개념으로 잡음원 샘플사이에 종속성이 의심되거나 샘플 간 주기적 특성이 존재할 때 활용될 수 있는 엔트로피 측정방법이다[5,9]. 개정된 SP 800-90B는 Non-IID로 판정된 잡음원의 엔트로피를 보다 엄밀히 측정하기 위해 프레딕터를 추가하였다. 이때 다수의 보조 프레딕터를 이용하여 잡음원의 엔트로피를 추정하는 다중결합 프레딕터(Ensemble predictor) 모델을 채택하여, 잡음원의 샘플 간 종속성 검출의 정확성을 높이고 있다[9]. SP 800-90B에서 사용하는 프레딕터 추정 방법으로는 윈도우 내 최빈값 추정, 래그 추정, 다중 마코브 추정, LZ78Y 압축 추정의 4 가지가 있다. 3.1절에서는 각 프레딕터 추정 방법에 대한 특징을, 3.2절에서는 프레딕터 엔트로피 추정의 이론적 근거를 살펴본다.

3.1 SP 800-90B 프레딕터

SP 800-90B가 개정되면서 Non-IID 트랙의 두드러진 변화는 잡음원의 주기적 특성을 검출하기 위한 프레딕터 추정 방법이 추가된 점이다. SP 800-90B에서

사용하고 있는 프레딕터 추정 방법은 표 2와 같다. 또한 프레딕터는 그 종류에 따라 검출하는 종속성이 다르지만, 전체적인 과정은 그림 2와 같이 초기화(Initialize), 예측(Predict), 관측(Observe), 비교(Compare), 업데이트(Update)의 5 단계로 구성된다[9,13].

이때 프레딕터 추정 방법은 샘플의 통계적 특성을 이용하기 보다는 프레딕터의 예측 공격 성공을 이용하여 추정하는 것이 특징이다. 프레딕터의 엔트로피 추정에 대한 내용은 3.2절에서 살펴본다.

표 2. SP 800-90B 프레딕터 추정 방법에 대한 설명
Table 2. Description of SP 800-90B Predictor

Predictor	Description
The multi most common in Window Prediction Estimate	Using the most common sample in window, predict a next sample and estimate entropy of samples.
The Lag prediction Estimate	Using a position, predict a next sample and estimate entropy of samples.
The MultiMMC Prediction Estimate	Using from 1st order to 16th order markov chain, predict a next sample and estimate entropy of samples.
The LZ78Y Prediction Estimate	Using a LZ78 encoding rule[14], predict a next sample and estimate entropy of samples.

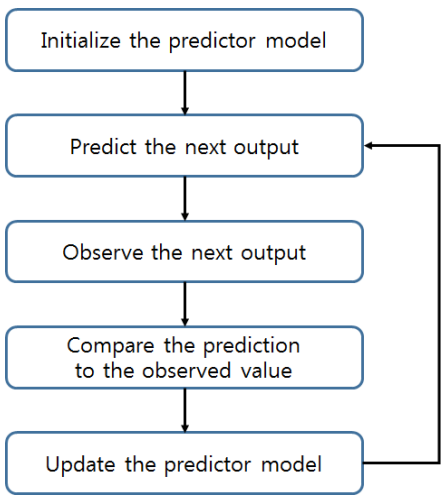

그림 2. 프레딕터의 엔트로피 추정 진행과정
Fig. 2. Process of SP 800-90B Predictor

3.2 프레딕터의 엔트로피 추정

프레딕터를 이용한 잡음원 엔트로피 추정은 기존의 통계적 성질을 이용하여 엔트로피를 추정하는 것과 다르게 프레딕터의 예측 공격 성공 여부를 이용하여 엔트로피를 추정한다[4,9]. 프레딕터의 엔트로피 추정에는 예측 성공 비율로 추정하는 전체 엔트로피(Global Entropy, H_{global})와 가장 긴 연속 예측 성공으로 추정하는 구간 엔트로피(Local Entropy, H_{local})의 두 가지 개념이 사용된다. 또한 프레딕터의 최종 엔트로피 $H_{predictor}$은 H_{global}, H_{local} 중 최솟값으로 결정된다. 즉,

$$H_{predictor} = \min\{H_{global}, H_{local}\}$$

3.2.1 전체 엔트로피(Global Entropy)[4,9,15]

전체 엔트로피 H_{global}은 프레딕터의 예측을 성공하는 기대확률을 이용하여 잡음원의 엔트로피를 추정하는 방법으로 예측 성공비율 $p_{global} = S/N$을 이용한다. 여기에서 N은 프레딕터가 예측한 횟수, S는 예측에 성공한 횟수를 의미한다. 이때 H_{global}은 99% 모비율 추정을 이용하여 아래와 같이 계산한다.

$$H_{global} = -\log_2\left(p_{global} + 2.576\sqrt{p_{global}(1-p_{global})/(N-1)}\right)$$

3.2.2 구간 엔트로피(Local Entropy)[4,9,15]

H_{local}은 프레딕터의 연속적인 예측성공을 이용하여 잡음원의 엔트로피를 추정하는 방법이다. 프레딕터에 의해 연속적으로 예측성공이 된다는 의미는 구간에서 잡음원의 주기적인 특성이 강하게 나타남을 의미하며, 이는 H_{global}에 의해 검출되기 구간의 종속성을 검출해내어 잡음원의 엔트로피를 추정하는 역할을 한다. 이때 H_{local} 계산에 필수적인 p_{local}은 유의수준(significance level) α를 이용하여 다음과 같이 계산한다.

$$\alpha = \frac{1-p_{local}\times x}{(r+1-rx)q}\times\frac{1}{x^{n+1}}$$

이때 $q = 1-p_{local}$, r은 프레딕터의 연속 성공 횟수 + 1, n은 프레딕터의 총 예측 횟수를 의미하며, x는 점화식(recurrence relation) $x_j = 1+q\times p_{local}^{r}\times x_{j-1}^{r+1}$을 10번 반복하면, 해에 근사하다는 성질을 이용하여 얻은 값이다. 이때, 위 식을 이진 탐색(binary search)

을 이용하면 p_{local} 을 계산할 수 있고, 엔트로피의 정의에 의해 구간 엔트로피를 다음과 같이 계산한다.

$$H_{local} = -\log_2(p_{local})$$

Ⅳ. 실험 및 결과

4.1 실험 목적

SP 800-90B[4]는 잡음원의 주기적인 특성을 검출해내는 데 적합하다고 알려진 프레딕터(predictor)를 Non-IID 트랙에 추가함으로써 주기적 특성을 가진 잡음원에 대해 보다 엄밀하게 추정하고 있다. 하지만, 현재까지 프레딕터가 검출해낼 수 있는 잡음원의 주기적 특성에 대해서는 알려져 있지 않다. 이에 본 논문에서는 프레딕터가 검출해낼 수 있는 잡음원의 주기적인 특성을 규명하기 위해, 잡음원에 주기적인 종속성을 부여하는 세 가지 시나리오를 구상하여 그 사실을 확인한다. 또한 이를 통해 프레딕터 추정 방법의 역할과 SP 800-90B에 추가된 의미를 실험적으로 규명한다.

4.2 실험 환경 및 방법

Quantis는 스위스 ID Quantique SA에서 출시한 양자난수발생기로, 양자 물리 잡음원(이하, raw data)을 추출 행렬(Extraction matrix)에 통과시키는 원리로 최종 난수를 출력한다. 현재 Quantis는 사용자가 추출 행렬을 선택하고 최종 난수를 출력할 수 있도록 하는 소프트웨어를 제공하고 있다[16]. 이를 활용하여 본 논문에서는 주기적 종속성을 부여한 2048×1792 크기의 추출 행렬을 생성하고, 그 추출 행렬로부터 종속성을 부여받은 최종 난수를 출력하였다. 실험의 신뢰성을 위해 변형된 추출 행렬을 통과한 최종 난수 130만 비트의 샘플 10개를 출력하여 프레딕터에 적합한 잡음원의 성질을 분석하는데 사용하였다.

- 실험 환경 : Intel(R) Core(TM) i7-4790K CPU @ 4.00GHz 4.00GHz. 16.0GB RAM
- 실험 대상 : Quantis raw data 1,300만 비트
- 실험 도구 : IDQ Quantis-PCIe-4M

프레딕터가 검출할 수 있는 주기적 특성을 규명하기 위한 세 가지 시나리오를 구상하였다. 시나리오 1은 최종 난수가 주기적으로 1 비트의 같은 값이 출력하게 되는 경우, 시나리오 2는 데이터의 이전 비트들을 비트별 연산자인 배타적 논리합(XOR) 연산을 이용하여 추출 행렬에 1 비트의 주기적인 종속성을 부여하여 최종 난수를 출력하는 경우, 마지막 시나리오 3은 시나리오 1, 2와 같은 계수(Rank)를 주지만 최대 계수(Full-Rank)에서 줄어든 만큼의 비트에 종속성을 주는 출력 행렬을 사용하여 최종 난수를 출력한 경우로 하여 실험을 진행하였다. 실험에 사용된 주기는 16(계수 1680), 32(계수 1736), 64(계수 1764), 128(계수 1778), 256(계수 1785)이며, 엔트로피 추정에는 NIST에서 제공한 SP 800-90B EntropyAssessment-master Python 코드를 활용하였다[17]. 또한 부록의 표 3은 최종 난수와 비교를 위한 대조군으로 추출 행렬을 거치지 않은 잡음원을 Non-IID 트랙을 실행한 결과이며, Non-IID 트랙의 최소 엔트로피를 결정한 추정 방법을 쉽게 알 수 있도록 엔트로피 추정치를 다른 색과 밑줄로 표현하였다.

4.3 실험 결과

4.3.1 주기가 16인 출력 난수

첫 번째는 주기가 16인 출력 난수에 대한 실험으로 추출 행렬의 계수가 1680인 경우이다. 주기가 16인 특성이 부여된 추출 행렬을 이용하여 시나리오 1, 시나리오 2, 시나리오 3을 진행하였다. 부록의 표 4, 5, 6은 각 시나리오에 대한 엔트로피 추정 방법으로 확인한 결과이다.

실험 결과 주기가 16일 때, 시나리오 1은 10개 데이터 모두 프레딕터에 의해 최소 엔트로피가 결정됨을 확인할 수 있었다. 또한 시나리오 2, 시나리오 3은 프레딕터 이외의 다른 추정 방법으로 잡음원의 최소 엔트로피가 결정됨을 확인할 수 있다. 첫 번째 실험을 통해 프레딕터는 추출 행렬의 주기성에 영향을 받기 보다는 데이터 값의 주기성에 영향을 받음을 알 수 있다.

4.3.2 주기가 32인 출력 난수

세 번째는 주기가 32인 출력 난수에 대한 실험으로, 추출 행렬의 계수가 1736인 경우이다. 주기가 64인 특성이 부여된 행렬을 이용하여 시나리오 1, 시나리오 2, 시나리오 3을 진행하였다. 부록의 표 7은 시나리오 1에 대한 엔트로피 추정 방법으로 확인한 결과이다.

실험 결과 주기가 32일 때, 시나리오 1은 10개 데이터 중 6개 데이터가 프레딕터에 의해 데이터의 최소 엔트로피가 결정됨을 확인할 수 있었다. 또한 시나리오 2, 시나리오 3은 프레딕터 이외의 다른 방법으로 데이터의 최소 엔트로피가 결정됨을 확인할 수 있다.

4.3.3 주기가 64인 출력 난수

세 번째는 주기가 64인 출력 난수에 대한 실험으로, 추출 행렬의 계수가 1764인 경우이다. 주기가 64인 특성이 부여된 행렬을 이용하여 시나리오 1, 시나리오 2, 시나리오 3을 진행하였다. 부록의 표 8은 시나리오 1에 대한 엔트로피 추정 결과이다.

실험 결과 주기가 64일 때, 시나리오 1은 10개 데이터 중 3개 데이터가 프레딕터에 의해 데이터의 최소 엔트로피가 결정됨을 확인할 수 있었다. 또한 시나리오 2, 시나리오 3은 프레딕터 이외의 다른 방법으로 데이터의 최소 엔트로피가 결정됨을 확인할 수 있다.

4.3.4 주기가 128, 256인 출력 난수

네 번째는 주기가 128, 256인 출력 난수에 대한 실험으로, 추출 행렬의 계수가 1778, 1785인 경우이다. 주기가 128, 256인 특성이 부여된 행렬을 이용하여 시나리오 1, 시나리오 2, 시나리오 3을 진행하였다. 부록의 표 9, 10은 시나리오 1에 대한 엔트로피 추정 방법으로 확인한 결과이다.

실험 결과 주기가 128일 때, 시나리오 1, 시나리오 2, 시나리오 3 모두 프레딕터가 데이터의 최소 엔트로피를 결정하는 경우는 확인할 수 없었다.

실험을 총 정리해보면, 그림 3과 같이 프레딕터 추정 방법은 잡음원에 주입된 주기적인 종속성 검출보다는 잡음원의 주기적으로 같은 값이 반복되는 경우를 검출해내는데 우수하다고 할 수 있다. 실험 결과로부터 프레딕터 추정 방법이 SP 800-90B에 추가된 이유는 주기적으로 같은 값을 갖는 잡음원을 보다 엄밀하게 평가하기 위함으로 분석된다.

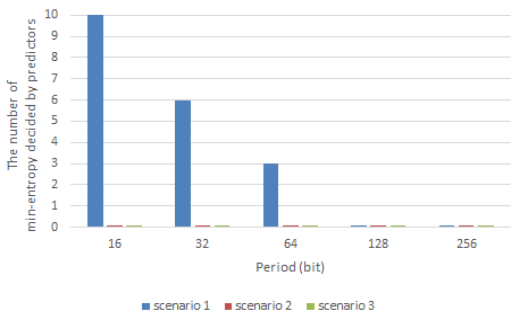

그림 3. 주기에 따른 시나리오 1, 2, 3의 결과
Fig. 3. The result of scenario 1, 2 and 3 by period

V. 결 론

본 논문에서는 NIST SP 800-90B Non-IID 트랙의 프레딕터 추정 방법이 검출해낼 수 있는 잡음원의 특성을 실험적으로 규명하였다. 실험을 통하여 프레딕터 추정 방법은 잡음원에 주기적인 종속성 검출보다는 주기적인 값이 반복되는 잡음원을 검출해내는데 유리하다고 할 수 있다. 또한 주기 구간이 증가함에 따라 프레딕터가 검출해내는 횟수가 감소함을 통해 프레딕터는 주기적 특성이 큰 잡음원 검출에 우수하다고 분석할 수 있다. 이를 통해 프레딕터가 SP 800-90B 2차 안(Second Draft)에 추가된 이유는 주기적인 값을 출력하는 잡음원의 안전성을 이전보다 엄밀하게 평가하기 위함으로 분석된다. 또한 현재 이론적으로 분석이 부족한 4가지 프레딕터 추정 방법에 대한 확률론적 원리를 추후 연구로 진행할 예정이다.

부록 - 실험 시나리오에 따른 엔트로피 추정 결과

표 3. 초기 잡음원의 Non-IID Test 결과
Table 3. Result of Non-IID test on the raw data

	MCW	Collision	Markov	Compre-s sion	t-tuple	LRS	MCW Pre	Lag Pre	MMC Pre	LZ78Y Pre
data 1	0.989175	**0.857893**	0.992547	1	0.928093	0.979621	0.98653	0.988918	0.989806	0.993047
data 2	0.989863	**0.863009**	0.991997	1	0.924966	0.981472	0.993429	0.994368	0.992367	0.991745
data 3	0.986829	**0.855342**	0.9908	1	0.94471	0.952605	0.991955	0.989605	0.988507	0.99121
data 4	0.986677	**0.850253**	0.988777	1	0.928093	0.981472	0.990463	0.992587	0.993248	0.993391
data 5	0.988107	**0.886258**	0.992696	1	0.918966	1.00022	0.994771	0.990503	0.99342	0.995672
data 6	0.988507	**0.830075**	0.985814	0.936605	0.924966	0.830473	0.989813	0.992836	0.989997	0.986533
data 7	0.990035	**0.868143**	0.993619	1	0.924966	0.999679	0.993161	0.986648	0.992004	1.00032
data 8	0.986619	**0.840129**	0.990553	1	0.910534	0.875365	0.990348	0.988727	0.988526	0.993774
data 9	0.987935	**0.896712**	0.992899	1	0.918966	1.00011	0.996536	0.989376	0.992674	0.991611
data 10	0.98702	**0.875879**	0.986418	0.943362	0.928093	0.994214	0.989947	0.989681	0.983611	0.987658

표 4. 시나리오 1 - 16을 주기로 동일한 값이 발생하는 난수열
Table 4. Scenario 1 - Result of Non-IID tests on data with fixed value at every 16 samples

	MCW	Collision	Markov	Compre-s sion	t-tuple	LRS	MCW Pre	Lag Pre	MMC Pre	LZ78Y Pre
data 1	0.410186	0.865574	0.582428	0.607146	0.416109	0.517732	0.410181	**0.404112**	0.404347	0.410233
data 2	0.409268	0.852795	0.580588	0.604533	0.415184	0.519834	0.409219	**0.406435**	0.407936	0.409257
data 3	0.408496	0.837609	0.579044	0.602526	0.414406	0.467891	**0.408462**	0.412902	0.409057	0.408529
data 4	0.409705	0.863009	0.581486	0.607683	0.415625	0.468849	0.409729	**0.406609**	0.412249	0.409752
data 5	0.4102	0.868143	0.582458	0.607951	0.416124	0.502984	0.410152	**0.409738**	0.40977	0.410262
data 6	0.410885	0.875879	0.583856	0.609965	0.416815	0.495494	0.41091	0.407903	**0.407689**	0.410918
data 7	0.406199	0.811411	0.574459	0.594659	0.41209	0.519102	**0.406163**	0.408704	0.407413	0.406231
data 8	0.409545	0.835093	0.581143	0.604399	0.415463	0.496569	0.409511	0.413048	**0.407834**	0.409592
data 9	0.406809	0.820091	0.575675	0.597388	0.412705	0.513799	**0.406759**	0.411851	0.411067	0.406798
data 10	0.409967	0.852795	0.582012	0.606744	0.415889	0.514967	0.409933	0.409927	**0.408358**	0.410029

표 5. 시나리오 2 - 행 XOR 연산을 이용하여 16 비트 주기의 종속성을 부여받은 난수열
Table 5. Scenario 2 - Result of Non-IID tests on data with dependency at every 16 samples

	MCW	Collision	Markov	Compre-s sion	t-tuple	LRS	MCW Pre	Lag Pre	MMC Pre	LZ78Y Pre
data 1	0.982573	**0.838868**	0.985404	1	0.930553	0.914244	0.993436	0.990962	0.98584	0.983975
data 2	0.987913	**0.82258**	0.990726	0.959025	0.939301	1.00055	0.989275	0.987596	0.990273	0.995213
data 3	0.98752	**0.863009**	0.988853	1	0.909969	1.00001	0.992713	0.987269	0.988874	0.985544
data 4	0.986451	**0.840129**	0.991457	1	0.913095	0.987183	0.988467	0.986789	0.9954	0.992012
data 5	0.988896	**0.928538**	0.986191	1	0.926407	0.951928	0.989932	0.994292	0.981855	0.987595
data 6	0.986276	1	0.985862	1	**0.903968**	0.999697	0.987178	0.986876	0.986756	0.985719
data 7	0.981138	**0.857893**	0.989431	1	0.939301	0.999716	0.988467	1.00047	0.986145	0.983779
data 8	0.98883	**0.837609**	0.99379	1	0.932681	1.00003	0.991836	0.996904	0.990885	0.993809
data 9	0.989682	**0.875879**	0.992371	1	0.913095	0.958646	0.988204	0.986221	0.988699	0.993722
data 10	0.985447	**0.847715**	0.991237	1	0.939301	0.999925	0.990785	0.993152	0.988961	0.988097

표 6. 시나리오 3 - 16 비트 주기와 같은 크기의 종속성이 부여된 난수열
Table 6. Scenario 3 - Result of Non-IID tests on data with same dependency to Scenario 1, 2

	MCW	Collision	Markov	Compre-s sion	t-tuple	LRS	MCW Pre	Lag Pre	MMC Pre	LZ78Y Pre
data 1	0.903667	0.635865	0.843896	0.673711	**0.447485**	0.837321	0.831328	0.865574	0.831328	0.831328
data 2	0.903399	0.634771	0.842649	0.672447	**0.444858**	0.820473	0.800328	0.831328	0.800328	0.800328
data 3	0.898529	0.638056	0.835279	0.669013	**0.442239**	0.799961	0.865574	0.901968	0.865574	0.865574
data 4	0.900212	0.642448	0.838117	0.670274	**0.442527**	0.831121	0.743791	0.771181	0.743791	0.743791
data 5	0.903193	0.645751	0.843811	0.673009	**0.441837**	0.830674	0.915942	0.915605	0.865574	0.865574
data 6	0.902247	0.632039	0.837167	0.667824	**0.441609**	0.820473	0.831328	0.865574	0.865574	0.865574
data 7	0.900068	0.644649	0.839397	0.671816	**0.451997**	0.820473	0.865574	0.916955	0.865574	0.865574
data 8	0.907769	0.630403	0.845568	0.67329	**0.447849**	0.83804	0.865574	0.911521	0.901968	0.865574
data 9	0.89976	0.638056	0.841795	0.676664	**0.445568**	0.835907	0.800328	0.831328	0.800328	0.800328
data 10	0.906861	0.646853	0.840033	0.6644	**0.440755**	0.830411	0.865574	0.901968	0.901968	0.901968

표 7. 시나리오 1 - 32 비트를 주기로 동일한 값이 발생하는 난수열
Table 7. Scenario 1 - Result of Non-IID tests on data with fixed value at every 32 samples

	MCW	Collision	Markov	Compre-s sion	t-tuple	LRS	MCW Pre	Lag Pre	MMC Pre	LZ78Y Pre
data 1	0.671173	0.817606	0.732235	0.905262	0.679112	0.786259	0.673478	0.670253	**0.666033**	0.671302
data 2	0.671454	0.832582	0.733645	0.925524	0.679394	0.78895	0.673443	**0.666672**	0.672745	0.671495
data 3	**0.667748**	0.875879	0.730029	0.902297	0.675661	0.798627	0.669888	0.672073	0.668669	0.667823
data 4	0.668865	0.832582	0.729588	0.894419	0.676787	0.761866	0.670955	**0.665574**	0.671782	0.668941
data 5	0.672276	0.870717	0.736005	0.946074	0.680223	0.780449	0.674391	**0.668296**	0.673551	0.672335
data 6	0.671874	0.955606	0.735588	0.947432	0.679817	0.813664	0.674005	**0.669344**	0.672237	0.671897
data 7	**0.66698**	0.817606	0.727984	0.89311	0.674887	0.818385	0.668979	0.67083	0.668023	0.667212
data 8	**0.671471**	0.901968	0.732871	0.910219	0.679411	0.794503	0.673549	0.671128	0.675779	0.671687
data 9	**0.66773**	0.881059	0.731292	0.920849	0.675643	0.817144	0.669731	0.668138	0.672657	0.667771
data 10	0.672942	0.860449	0.733685	0.914196	0.680893	0.817837	0.673141	0.676218	**0.672511**	0.67497

표 8. 시나리오 1 - 64 비트를 주기로 동일한 값이 발생하는 난수열
Table 8. Scenario 1 - Result of Non-IID tests on data with fixed value at every 64 samples

	MCW	Collision	Markov	Compre-s sion	t-tuple	LRS	MCW Pre	Lag Pre	MMC Pre	LZ78Y Pre
data 1	0.82278	0.845182	0.84515	1	0.831818	0.864824	0.832085	**0.822321**	0.822855	0.822963
data 2	**0.822079**	0.857893	0.846193	1	0.831113	0.939351	0.828504	0.822263	0.822291	0.822282
data 3	**0.817517**	0.878466	0.838879	1	0.826517	0.938443	0.823607	0.819562	0.817805	0.817582
data 4	0.818195	0.820091	0.838588	0.962453	0.827201	0.941134	0.818319	0.818348	**0.816169**	0.825654
data 5	**0.822897**	0.886258	0.85074	1	0.831936	0.941134	0.829443	0.824969	0.823108	0.823002
data 6	0.823325	0.891476	0.844436	1	0.832367	0.958741	0.831909	**0.820863**	0.823614	0.82343
data 7	**0.816955**	0.84139	0.83843	0.967954	0.825951	0.943476	0.825498	0.824073	0.817262	0.817117
data 8	**0.824201**	0.845182	0.843814	1	0.833249	0.923929	0.833242	0.824579	0.824413	0.824677
data 9	**0.818292**	0.835093	0.841407	1	0.827298	0.963476	0.824523	0.826958	0.818833	0.818513
data 10	**0.823247**	0.881059	0.84682	1	0.832288	0.941134	0.831419	0.8238	0.823906	0.82345

표 9. 시나리오 1 - 128 비트를 주기로 동일한 값이 발생하는 난수열
Table 9. Scenario 1 - Result of Non-IID tests on data with fixed value at every 128 samples

	MCW	Collision	Markov	Compre-s sion	t-tuple	LRS	MCW Pre	Lag Pre	MMC Pre	LZ78Y Pre
data 1	0.904264	**0.850253**	0.915721	1	0.892861	0.98671	0.91925	0.909266	0.904468	0.904551
data 2	0.904676	**0.852795**	0.917956	1	0.906928	0.994836	0.919604	0.905262	0.905396	0.905066
data 3	0.897811	**0.870717**	0.906199	1	0.895536	0.941134	0.910758	0.903141	0.898076	0.898076
data 4	0.900849	**0.863009**	0.909426	1	0.898276	0.98464	0.913949	0.898251	0.901032	0.901012
data 5	0.904243	**0.870717**	0.919318	1	0.895536	0.993681	0.91798	0.903985	0.90453	0.904324
data 6	0.904099	**0.837609**	0.90953	0.956289	0.898276	978847	0.918521	0.901084	0.904613	0.904263
data 7	0.899678	**0.838868**	0.907811	0.965889	0.903923	0.981676	0.914759	0.903676	0.900087	0.89982
data 8	0.906139	**0.830075**	0.91538	1	0.906928	0.993758	0.921688	0.905881	0.906447	0.906613
data 9	0.899945	**0.837609**	0.912521	1	0.89025	0.988029	0.913431	0.904974	0.90093	0.900498
data 10	0.906449	**0.863009**	0.915074	1	0.892861	0.987365	0.919937	0.903079	0.907809	0.90717

표 10. 시나리오 1 - 256 비트를 주기로 동일한 값이 발생하는 난수열
Table 10. Scenario 1 - Result of Non-IID tests on data with fixed value at every 256 samples

	MCW	Collision	Markov	Compre-s sion	t-tuple	LRS	MCW Pre	Lag Pre	MMC Pre	LZ78Y Pre
data 1	0.946909	**0.842653**	0.955046	1	0.913095	0.9819	0.974093	0.951979	0.947078	0.947274
data 2	0.947588	**0.850253**	0.956183	1	0.928462	0.998821	0.973596	0.953299	0.949266	0.948272
data 3	0.940156	**0.850253**	0.944202	1	0.928462	0.968706	0.966498	0.950574	0.941085	0.940646
data 4	0.942079	**0.855342**	0.947995	1	0.928462	0.989571	0.966541	0.947768	0.943052	0.942612
data 5	0.946336	**0.868143**	0.95697	1	0.934847	0.998799	0.971132	0.94499	0.947163	0.946744
data 6	0.9457	0.845182	0.947677	0.961767	0.913095	**0.820473**	0.974872	0.943888	0.947184	0.946871
data 7	0.942756	**0.833837**	0.945736	0.950151	0.913095	0.969654	0.966734	0.945647	0.943454	0.94331
data 8	0.950223	**0.837609**	0.957161	1	0.924386	0.951928	0.97245	0.947365	0.951393	0.950866
data 9	0.942249	**0.857893**	0.95097	1	0.913095	0.969654	0.971823	0.95068	0.943602	0.943183
data 10	0.950521	**0.888864**	0.954558	1	0.909969	0.992473	0.983231	0.947938	0.952564	0.951037

References

[1] NIST, *Recommendation for the Entropy Sources Used for Random Bit Generation*, NIST DRAFT Special Publication 800-90B, Aug. 2012.

[2] NIST, *Recommendation for Random Bit Generator(RBG) Constructions*, NIST Special Publication 800-90C, Aug. 2012.

[3] H. Kang, Y. Yeom, and J. S. Kang, "An implementation of integrated tool for statistical randomness tests and entropy estimations," in *Proc. KICS Winter Conf. 2016*, pp. 229-230, Jeongseon, Korea, Jan. 2016.

[4] NIST, *Recommendation for the Entropy Sources Used for Random Bit Generation*, (Second DRAFT)NIST Special Publication 800-90B, Jan. 2016.

[5] C. E. Shannon, "Prediction and entropy of printed English," *Bell Syst. Tech. J.*, vol. 30, no. 1, pp. 50-64, 1951.

[6] K. Horvath, H. Stögner, A. Uhl, and G. Weinhandel, "Lossless compression of polar iris image data," Pattern Recognition and Image Anal., vol. 6669, pp. 329-337, 2011.

[7] N. Chater and C. D. Manning, “Probabilistic models of language processing and acquisition,” *Trends in Cognitive Sci.*, vol. 10, no. 7, pp. 335-344, 2006.

[8] R. Collobert, J. Weston, L, Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa, “Natural language processing (almost) from scratch,” *J. Machine Learning Res.*, pp. 2493-2537, 2011.

[9] J. Kelsey, K. A. McKay, and M. S. Turan, “Predictive models for min-entropy estimation,” *CHES 2015*, vol. 9293, pp. 373-392, Sept. 2015.

[10] Y. Kim and K. Yi, “Safety comparision analysis against known/chosen plaintext attack of RBF (Random Block Feedback) mode to other block cipher modes of operation,” *J. KICS*, vol. 39B no. 05, pp. 317-322, 2014.

[11] H. Park, M. Bae, J. S. Kang, and Y. Yeom, “Key derivation functions using the dual key agreement based on QKD and RSA crypto-system,” *J. KICS*, vol. 41 no. 04, pp. 479-488, 2016.

[12] K. J. Ha, C. H. Seo, and D. Y. Kim, “Design of validation system for a crypto-algorithm implementation,” *J. KICS*, vol. 39B no. 04, pp. 242-250, 2014.

[13] H. Park, M. Bae, Y. Yeom, and J. S. Kang, “A study on the predictor of Non-IID track in SP 800-90B,” in *Proc. KICS Int. Conf. Commun.*, pp. 115-116, Jeju, Korea, Jun. 2016.

[14] D. Salomon, *Data Compression: The Complete Reference Fourth Edition*, Springer, pp. 189-192, 2007.

[15] W. Feller, *An Introduction to Probability Theory and Its Applications Third Edition*, John Wiley & Sons, Inc, pp. 303-341, 1950.

[16] *ID Quantique SA*, Retrieved Aug., 28 from http://www.idquantique.com/random-number-generation/quantis-random-number-generator/.

[17] NIST, *SP800-90B_EntropyAssessment*, Retrieved Aug., 28 from https://github.com/usnistgov/SP800-90B_EntropyAssessment.

박 호 중 (Hojoong Park)

2015년 2월 : 국민대학교 수학과 학사

2015년 3월~현재 : 국민대학교 금융정보보안학과 석사

<관심분야> 암호이론, 정보보호 알고리즘 및 프로토콜, 난수성 분석

배 민 영 (Minyoung Bae)

2016년 2월 : 국민대학교 수학과 학사

2016년 3월~현재 : 국민대학교 금융정보보안학과 석사

<관심분야> 암호이론, 병렬구현, 정보보호 프로토콜

염 용 진 (Yongjin Yeom)

1991년 2월 : 서울대학교 수학과 학사

1994년 2월 : 서울대학교 수학과 석사

1999년 2월 : 서울대학교 수학과 박사

2000년 4월~2012년 2월 : ETRI 부설연구소 책임연구원/팀장

2006년 12월~2007년 12월 : Columbia 대학교 방문연구원

2012년~현재 : 국민대학교 수학과 부교수

2013년~현재 : 국민대학교 BK21+ 미래 금융정보보안 인력양성사업단 교수

<관심분야> 암호구현 및 분석, 보안시스템 평가

강 주 성 (Ju-Sung Kang)

1989년 2월 : 고려대학교 수학과 학사
1991년 2월 : 고려대학교 수학과 석사
1996년 2월 : 고려대학교 수학과 박사
1997년~2004년 : 한국전자통신연구원 선임연구원/팀장
2001년~2002년, 2010년 : 벨기에 루벤대학 COSIC 방문 연구원
2004년~현재 : 국민대학교 수학과 교수
2013년~현재 : 국민대학교 BK21+ 미래 금융정보보안 인력양성사업단 교수
<관심분야> 암호이론, 정보보안 프로토콜, 안전성 분석 및 평가