NIST SP 800-22 파라미터 조정이 테스트 결과에 미치는 영향에 관한 연구

박형준, 이명훈, 홍종필
충북대학교
phj952000@chungbuk.ac.kr, hun990508@chungbuk.ac.kr, jphong@chungbuk.ac.kr

A Study on the Impact of Parameter Adjustment in NIST SP 800-22 on Test Results

Park Hyeong Jun, Lee Myeong Hun, Hong Jong Phil
Chungbuk National Univ.

요 약

본 논문은 NIST SP 800-22 테스트 파라미터 선택이 난수성 검증 결과에 미치는 민감도를 실험적으로 분석한다. 동일한 RNG임에도 파라미터 설정에 따라 테스트 결과가 달라질 수 있음을 확인하였다. 실제 제작된 진성 난수 발생기(TRNG)를 대상으로 NIST SP 800-22의 하위 테스트인 Frequency Test에서 하위 비트스트림 길이(n)를 변화시키며 통과율을 측정하였다. 그 결과, n 값 조절만으로 통과율이 97%에서 100%까지 변화함을 확인하였다. 이는 RNG 검증 시 최적의 테스트 파라미터 선정이 필수적임을 보여준다.

Ⅰ. 서론

난수 발생기(RNG)는 암호 키 생성, 보안 프로토콜 등 다양한 분야에서 핵심적인 역할을 수행하며, RNG의 난수성은 시스템 보안의 근간이 된다. RNG의 난수성 평가는 필수적이며, 국제 표준으로 널리 사용되는 NIST SP 800-22 테스트가 그 대표적인 방법이다. NIST 테스트는 전체 비트스트림을 여러 하위 비트스트림으로 분할하고, 하위 비트스트림 길이(n) 등 다양한 파라미터를 통해 평가를 수행한다.

NIST 지침은 신뢰성 있는 검증 결과를 위해 RNG 특성에 맞는 적절한 파라미터 설정을 요구하고 있다. 그러나 RNG가 동일한 무작위성을 가지더라도, 검증자가 설정하는 파라미터에 따라 테스트 결과가 달라질 수 있는 문제가 존재한다. 이는 RNG의 실제 난수성 품질과 NIST 테스트 통과 여부가 반드시 일치하지 않을 수 있음을 의미하며, 이러한 파라미터 민감도 문제는 난수성 검증의 신뢰도 저하로 이어질 수 있다.

따라서 본 논문에서는 NIST 테스트 파라미터 선택이 검증 결과에 미치는 영향을 실험적으로 분석하였다. NIST SP 800-22의 15개 하위 테스트 중 기본적인 Frequency Test를 선정하여, 실제 제작된 진성 난수 발생기(TRNG)를 대상으로 하위 비트스트림 길이(n)의 변화에 따른 통과율 민감도를 분석하였다. 본 연구는 RNG 설계 및 검증 시 파라미터 선택의 중요성을 강조한다.

Ⅱ. 본론

2.1. Frequency Test

$$X_i = \epsilon_i - 1 = \pm 1 \quad (1)$$

$$S_n = X_1 + X_2 + \cdots + X_n \quad (2)$$

$$erfc(z) = \frac{2}{\pi}\int_z^{\infty} e^{-u^2} du, \ (z = \frac{|s_n|}{\sqrt{2 \times n}}) \quad (3)$$

Frequency Test는 주어진 하위 비트스트림(ϵ) 내 0과 1의 개수가 통계적으로 기대되는 비율과 일치하는지 검증하여 난수열의 가장 기본적인 편향성(Bias) 유무를 확인하는 테스트이다. 이 테스트는 ϵ의 길이를 n이라 할 때, 각 비트 {0, 1}를 수식 (1)을 사용하여 {-1, +1}로 변환한다. 그리고 수식 (2)를 사용하여 누적 합(S_n)을 계산한다. 이후 수식 (3)을 이용하여 P-Value 값을 계산하고 그 값이 유의수준(α) 이상인 경우 해당 비트스트림은 통과라고 판단한다. 유의수준(α)은 암호학에서 일반적으로 사용되는 0.01로 설정한다.

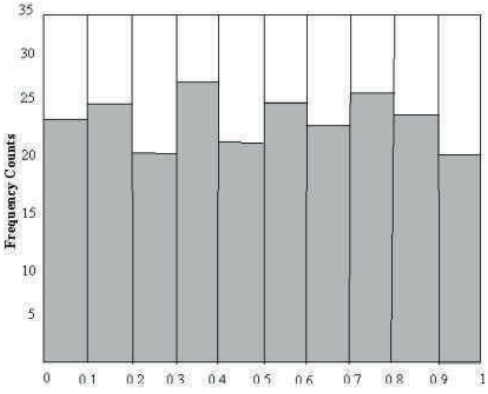

그림 1. 각 비트스트림의 P-Value 분포 균일성 분석 결과

NIST 테스트는 RNG의 무작위성을 최종적으로 평가하기 위해 Two-Level Test 방식을 필수적으로 요구한다. 이는 하위 비트스트림 길이(n)의 하위 비트스트림을 m개 추출하여 각 비트스트림에서 얻어진 m개의 P-Value($P_1, P_2, \cdots, P_m$)의 분포가 균일성을 띄는지 검증하는 과정이다.

$$\chi^2 = \sum_{i=1}^{10} \frac{(F_i - s/10)^2}{s/10} \quad (4)$$

$$igamc(\frac{\chi^2}{2}) = \frac{\int_x^{\infty} e^{-t} t^{3.5} dt}{\int_0^{\infty} e^{-t} t^{3.5} dt}, \ (x = \frac{\chi^2}{2}) \quad (5)$$

각 비트스트림의 P-Value들을 10개의 구간으로 나누어 m개의 P-Value 분포를 검증하기 위해 카이제곱(χ^2)을 계산한다. 카이제곱(χ^2) 은 수식

(4)를 이용하여 계산한다. 여기서, F_i은 i번째 구간 내에 존재하는 P-Value의 개수이며, s는 전체 P-Value 샘플의 개수를 의미한다. 이후 수식 (5)를 통해 최종 P-Value 값을 계산한다. 이 때, 0.0001 이상인 경우 최종적으로 통과했다고 판단한다. 본 실험에서는 하위 비트스트림의 길이(n) 변화에 따른 테스트 결과를 확인하기 위해 하위 비트스트림의 개수는 100으로 고정시키고 진행한다.

2.2. 실험 및 결과

본 연구는 진성 난수 발생기(True Random Number Generator, TRNG)를 대상으로 NIST SP 800-22 테스트 파라미터가 통계적 평가에 미치는 민감도를 실험적으로 분석하였다. 동일한 TRNG에서 생성한 100억 비트를 사용하여 충분한 표본 크기로 통계적 신뢰도와 결과 재현성을 확보하였다. 하위 비트스트림의 길이(n)를 주요 실험 변수로 설정하였으며, 각 실험에서 사용하는 비트 수는 n × 100으로 고정하였다. 즉, n이 100만일 경우 각 비트스트림에서 100개의 하위 비트스트림을 구성해 총 100회 테스트가 수행되고, n이 500만일 경우 20개의 하위 비트스트림을 구성하여 총 20회 테스트가 수행된다. 이처럼 하위 비트스트림 길이(n)의 변화에 따른 영향만을 독립적으로 관찰할 수 있도록 설계하였다. 이를 통해 n 값 변화에 따른 검정력 조절과 결과 영향에 대해 분석하였다.

표 (1)은 실험 결과를 보여준다. 하위 비트스트림 길이(n)가 3,000,000 이상일 경우 모든 테스트에서 통과율이 100%로 유지되었으나, 1,000,000과 2,000,000일 때는 각각 96%, 98%로 다소 낮아지는 현상이 관찰되었다. 이는 하위 비트스트림 길이(n)가 작을 경우, 단일 비트스트림에 대한 검정력이 과도하게 높아져 TRNG가 가지는 미세한 통계적 편차까지 통계적으로 유의한 것으로 판단되어 통과율이 낮아지는 경향을 보인다. 반면, 하위 비트스트림 길이(n)가 커질수록 검정력은 상대적으로 낮아지고, 동일한 수준의 편차는 유의하지 않은 것으로 간주되어 통과 처리된다. 또한, 하위 비트스트림 길이(n)가 커짐에 따라 각 비트스트림 내 통계량이 기대값에 수렴하는 큰 수의 법칙의 효과가 나타나면서, 무작위성이 보다 정확히 반영되고 전체 통과율도 안정적으로 100%에 수렴하는 경향을 나타낸다.

n	테스트 횟수	통과율
1,000,000	100	96%
2,000,000	50	98%
3,000,000	33	100%
4,000,000	25	100%
5,000,000	20	100%
6,000,000	16	100%
7,000,000	14	100%
8,000,000	12	100%
9,000,000	11	100%
10,000,000	10	100%

표 1. n 값에 따른 Frequency Test 통과율

이러한 결과는 하위 비트스트림의 길이(n)가 통계적 테스트 결과에 실질적인 영향을 미친다는 점을 실증적으로 보여준다. 동일한 난수 데이터를 사용했음에도 불구하고, 하위 비트스트림의 길이(n) 설정에 따라 통과율이 달라지는 현상은 테스트 파라미터가 평가 결과를 결정짓는 중요한 요인임을 뒷받침한다

Ⅲ. 결론

연구에서는 동일한 진성 난수 발생기(TRNG)에서 생성된 대규모 난수 데이터를 기반으로, NIST SP 800-22 테스트 중 하위 비트스트림 길이(n)라는 핵심 파라미터가 통계적 평가 결과에 미치는 영향을 실험적으로 분석하였다. 그 결과, 하위 비트스트림 길이(n)가 작을 경우 단일 비트스트림에 대한 검정력이 과도하게 높아 미세한 통계적 편차까지 유의한 것으로 판단되어 통과율이 낮아지는 반면, n이 커질수록 검정력은 완화되고 큰 수의 법칙에 의해 통계량이 안정적으로 기대값에 수렴함에 따라 통과율이 100%에 안정적으로 수렴하는 현상이 확인되었다. 이러한 결과는 난수성 평가에서 테스트 파라미터의 설정이 평가 신뢰도와 결과 해석에 결정적인 영향을 미침을 시사한다. 특히, 동일한 난수 데이터를 대상으로 하더라도 하위 비트스트림 길이(n)의 설정에 따라 통계적 검증 결과가 크게 달라질 수 있으므로, RNG 품질 평가 시 단순히 테스트 통과 여부만을 기준으로 판단하는 것은 위험하다. 따라서, RNG 설계 및 평가 과정에서는 NIST 테스트 결과를 해석할 때 파라미터 민감도를 고려한 체계적 분석과 최적 파라미터 설정이 필수적임을 본 연구를 통해 강조한다.

향후 연구에서는 본 실험에서 다룬 Frequency Test를 포함한 NIST SP 800-22 내 다른 테스트 항목들에 대해서도 파라미터 민감도를 종합적으로 분석하여, RNG 특성에 최적화된 평가 방법론을 제시할 계획이다.

ACKNOWLEDGMENT

이 논문은 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 지역지능화혁신인재양성사업임(IITP-2025 - RS - 2020-II201462)

이 논문은 2025년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No. RS-2020-NR049604)

참 고 문 헌

[1] L. E. Bassham et al., A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications. Gaithersburg, MD, USA: National Institute of Standards and Technology (NIST), Apr. 2010.

[2] F. Pareschi, R. Rovatti, and G. Setti, “Second-level NIST randomness tests for improving test reliability,” in Proc. IEEE Int. Symp. Circuits Syst., May 2007, pp. 1437 - 1440.

[3] D. Chen, H. Chen, L. Fan and K. Luo, "Error Analysis of NIST SP 800-22 Test Suite," in IEEE Transactions on Information Forensics and Security, vol. 18, pp. 3745-3759, 2023, doi: 10.1109/TIFS.2023.3287391.

[4] H. Sackrowitz and E. Samuel-Cahn, “P values as random variables – Expected P values,” Amer. Statistician, vol. 53, no. 4, pp. 326 - 331, Nov. 1999.