TTA Standard

정보통신단체표준(국문표준)

TTAS.KO-12.xxxx 제정일: 2017년 6월 xx일

소프트웨어 환경에서의 잡음원 엔트로피 검증 알고리즘

Entropy Evaluation Algorithms for Noise Sources in Software Environments

TTA 한국정보통신기술협회
Telecommunications Technology Association

표준초안 검토 위원회 정보보호기반 프로젝트그룹(PG501)

표준안 심의 위원회 정보보호 기술위원회(TC5)

	성명	소 속	직위	위원회 및 직위	표준번호
표준(과제) 제안	한상윤	NSR	선임연구원	정보보호기반 프로젝트그룹 위원	TTAKS.KO-12.xxxx
표준 초안 작성자	한상윤	NSR	선임연구원	정보보호기반 프로젝트그룹 위원	TTAKS.KO-12.xxxx
	서석충	NSR	선임연구원	정보보호기반 프로젝트그룹 위원	TTAKS.KO-12.xxxx
	염용진	국민대학교	교수	-	TTAKS.KO-12.xxxx
	김예원	국민대학교	연구원	-	TTAKS.KO-12.xxxx
사무국 담당	박수정	TTA	선임	-	

본 표준 발간 이전에 접수된 지식재산권 확약서 정보는 본 표준의 '부록(지식재산권 확약서 정보)'에 명시하고 있으며, 이후 접수된 지식재산권 확약서는 TTA 웹사이트에서 확인할 수 있습니다.
본 표준과 관련하여 접수된 확약서 외의 지식재산권이 존재할 수 있습니다.

발행인 : 한국정보통신기술협회 회장
발행처 : 한국정보통신기술협회
13591, 경기도 성남시 분당구 분당로 47
Tel : 031-724-0114, Fax : 031-724-0109
발행일 : 2017.6.

서 문

1 표준의 목적

이 표준의 목적은 소프트웨어 환경에서 결정론적 난수발생기(DRBG)에 사용되는 잡음원에 대한 엔트로피 검증 알고리즘을 규정한다.

2 주요 내용 요약

이 표준은 소프트웨어 환경에서의 난수발생기에 사용되는 잡음원 엔트로피를 검증할 수 있는 알고리즘을 규정한다. 특히 소프트웨어 환경에서의 잡음원 데이터 수집방법과 이를 바탕으로 한 엔트로피 검증 알고리즘을 규정한다.

3 인용 표준과의 비교

3.1 인용 표준과의 관련성

각각의 인용 표준에서 소프트웨어 환경에서의 잡음원에 적합한 테스트를 다음과 같이 선정하였고, 가변적인 검정기준을 가지도록 일반화하여 이 표준을 작성하였다.
적은 데이터를 이용한 통계적 난수성 테스트를 위해 AIS.31에서 Class P1의 T1~T5를 선정하였고, 이 표준에서는 P-value를 적용하여 가변적인 검정기준을 적용할 수 있는 알고리즘으로 작성하였다. 그리고 Shannon 엔트로피 측정을 위해 AIS.31에서 Coron 검정을 기반으로 한 엔트로피 측정 방법인 T8을 선정하였다.
최소 엔트로피(Min-entropy) 측정을 위해 SP 800-90B(2nd Draft)에서 Non-IID 테스트의 최빈값 추정(The Most Common Value Estimate), 충돌 추정(The Collision Estimate), 마르코브 추정(The Markov Estimate), 압축 추정(The Compression Estimate)을 선정하였다. 또한, 잡음원 건전성 테스트를 위해 SP 800-90B에서 건전성 테스트인 반복 횟수 테스트(Repetition Count Test)와 적응성 비율 테스트(Adaptive Proportion Test)를 선정하였다. 잡음원 수집 방법, 상호정보량(Mutual Information) 기반의 엔트로피 추정 방법 등은 위의 표준들에 포함되지 않는 내용이다.

3.2 인용 표준과 본 표준의 비교표

TTAK.KO-12.xxxx	AIS.31 (2001)	NIST SP 800-90B (2nd Draft, 2016)	비고
5.1 통계적 난수성 테스트	F. Statistical tests (T1~T5)		가변적인 검정기준을 적용하여 선택적 추가
5.2 잡음원 건전성 테스트		4.4 Approved Continuous Health Tests (Repetition Count Test, Adaptive Proportion Test)	선택적 추가
6.1 최소 엔트로피 측정		6.3 Estimator (The Most Common Value Estimate, The Collision Estimate, The Markov Estimate, The Compression Estimate)	선택적 추가
6.2 Shannon 엔트로피 측정	F. Statistical tests (T8)		일부 채택

Preface

1 Purpose

The standard specifies evaluation and statistical testing algorithms for DRBG in software environments.

2 Summary

The standard specifies evaluation and statistical testing algorithms for DRBG resource in software environments. In particular, this standard specifies resource collecting methods and entropy validation algorithms in S/W environments.

3 Relationship to Reference Standards

The standard is referenced some part from AIS.31 and NIST SP 800-90B (2nd Draft). This standard has some selective adoption after generalization depending on software environments.
From AIS.31, this standard is applied several statistical test methods(T1-T5) of Class P1 to calculate small sample data and applying P-value. Also, this standard is adopted Shannon entropy estimation(T8).
In Addition, This standard is adopted some min-entropy estimators from Non-IID test method of NIST SP 800-90B. eg, The Most Common Value Estimate, The Collision Estimate, The Markov Estimate, and the Compression Estimate.
And, health tests is adopted Repetition Count Test and Adaptive Proportion Test from NIST SP 800-90B.

TTAK.KO-12.xxxx	AIS.31 (2001)	NIST SP 800-90B (2nd Draft, 2016)	Remarks
5.1 Statistical Tests	F. Statistical tests (T1-T5)		Selective adoption after generalization
5.2 Health Tests		4.4 Approved Continuous Health Tests (Repetition Count Test, Adaptive Proportion Test)	Selective adoption
6.1 Minimum Entropy Estimator		6.3. Estimator (The Most Common Value Estimate, The Collision Estimate, The Markov Estimate, The Compression Estimate)	Selective adoption
6.2 Shannon Entropy Estimator	F. Statistical tests (T8)		Some adoption

목 차

소프트웨어 환경에서의 난수발생기 잡음원 엔트로피 검증 알고리즘 (Entropy Evaluation Algorithms for Noise Sources in S/W Environments)

1 적용 범위

이 표준은 암호 제품을 개발하는 개발자가 운영체제에서 잡음원을 수집하는 경우와 수집된 잡음원을 응용하는 경우에 적용될 수 있다. 이 표준을 바탕으로 소프트웨어 환경에서 수집된 잡음원을 검증하고 통계적으로 시험할 수 있다.

따라서, 소프트웨어 환경의 잡음원을 이용하여 암호 제품을 개발하거나 해당 제품을 바탕으로 시험·검증을 준비하는 기관 그리고 시험·검증기관에서 본 표준을 적용할 수 있다.

소프트웨어에서 획득하는 잡음원의 특징을 기반으로 적용하는 알고리즘이므로 하드웨어 기반의 엔트로피 검증에 직접적으로 적용에 어려움이 있다. 하지만, 하드웨어 기반의 난수에 대한 엔트로피 및 난수성 평가방법은 NIST SP 800-80B, SP 800-22, BSI AIS.31, Dieharder 등이 방법론과 사용하는 가능한 도구까지 제공되고 있으니, 해당 내용을 바탕으로 검증이 가능하다.

2 인용 표준

BSI AIS.31 (2001), A Proposal for : Functionality Classes and evaluation methodology for true(physical) random number generators. - F. Statistical tests[1)]

(2nd Draft) NIST SP 800-90B (2016), Recommendation for the Entropy Sources Used for Random Bit Generation. - 4.4 Approved Continuous Health test[2)], 6.3 Estimator[3)]

---

1) “5.1절 통계적 난수성 테스트”에서 T1~T5에 P-value를 적용하여 일반화한 테스트 방법론을 사용했으며, T8을 선정하여 이 표준의 “6.2절 Shannon 엔트로피 측정”에서 사용하였다.

2) “4.4 Approved Continuous Health Tests”에서 반복 횟수 테스트(Repetition Count Test)와 적응성 비율 테스트(Adaptive Proportion Test)를 선정하여 이 표준의 “5.2 잡음원 건전성 테스트”에서 사용하였다.

3) Non-IID 테스트 중 최빈값 추정(The Most Common Value Estimate), 충돌 추정(The Collision Estimate), 마르코브 추정(The Markov Estimate), 압축 추정(The Compression Estimate)을 선정하여 이 표준의 “6.1절 최소 엔트로피 측정”에서 사용하였다.

3 용어 정의

3.1 충돌(Collision)
데이터세트 내에서 동일한 샘플 값이 발생한 경우

3.2 엔트로피(Entropy)
데이터가 가지는 정보량을 수치적으로 나타낸 것
무질서도(Disorder) 또는 난수성(Randomness)을 나타내며 난수에 가까울수록 엔트로피가 높음

3.3. IID(Independent and Identically Distributed)
난수열 사이에 확률분포가 동일하며, 서로 독립적인 형태

3.4. 마르코브 모델(Markov model)
난수열 I번째가 그 이전의 n개의 난수열에 의존할 경우 n차 마르코브 모델(n-th order Markov model)이라고 함

3.5. 최소 엔트로피(Min-entropy)
m-비트의 난수 X에 대한 min-entropy는 가장 많은 X의 형태가 나타날 경우이며, 이는 난수 X의 엔트로피 중 가장 낮은 임계값이며 가장 안 좋은 경우임

3.6. 유의 확률(P-value)
가설 검정에서 귀무가설(Null hypothesis(H_0): 난수는 랜덤하다)을 지지할 확률
꼬리 확률(tail probability)로 불리기도 함
유의 확률이 유의 수준보다 크다는 의미는 귀무가설을 지지할 확률이 높음을 의미함
즉, 테스트를 통과한 데이터가 설정한 유의 수준에 대해서 랜덤하다는 것을 의미함

4 약어, 기호 및 입력 데이터 구성 방식

4.1 약어

DRBG	Deterministic Random Bit Generator
NIST	National Institute of Standards and Technology
SP	NIST Special Publication

4.2 기호

$[a, b]$	a와 b를 포함한 a와 b 사이의 모든 실수들의 집합
$A = \{x_1, x_2, ..., x_k\}$	디지털화된 잡음원으로부터 출력 가능한 모든 크기가 ℓ-비트인 샘플 x_i들의 집합 ($k = 2^{\ell}$)
$b_1, ..., b_N$	크기가 ℓ-비트인 샘플 s_i로 정렬된 데이터세트 $S = (s_1, ..., s_L)$의 비트열 표현 ($N = L\ell$)
CF	$CF(1/z) = igamc(k+1, z)z^{-k-1}e^{z}$으로, CF은 NIST SP 800-90B 등을 참조하여 구현함
cor	$cor = \dfrac{\sum_{i=0}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=0}^{n}(x_i - \overline{x})^2 \times \sum_{i=0}^{n}(y_i - \overline{y})^2}}, where -1 \le cor \le 1$ 첫 번째 바이트열과 두 번째 바이트열에 대한 상관관계를 계산할 때, $\overline{x}$와 $\overline{y}$는 각각 첫 번째 바이트열과 두 번째 바이트열에 대한 평균이고, x_i와 y_i는 각각 첫 번째 바이트열과 두 번째 바이트열에서 관찰된 값임
$erfc$	여오차 함수(Complementary Error Function)로, 표준정규분포의 오류함수임 $erfc(z) = \dfrac{2}{\sqrt{\pi}}\int_{z}^{\infty} e^{-x^2}dx$ $erfc$은 NIST SP 800-22 등을 참조하여 구현함

H 통계량으로 유도된 크기가 ℓ-비트인 샘플 당 Shannon 엔트로피
수치계산의 결과로 얻어진 H가 0보다 작은 경우 $0 \le H$이 되도록 보정 과정을 거침
H가 ℓ보다 큰 경우 $H = \ell$이 되도록 보정 과정을 거침

H_∞ 통계량으로 유도된 크기가 ℓ-비트인 샘플 당 최소 엔트로피
수치계산의 결과로 얻어진 H_∞가 0보다 작은 경우 $0 \le H_\infty$이 되도록 보정 과정을 거침
H_∞가 ℓ보다 큰 경우 $H_\infty = \ell$이 되도록 보정 과정을 거침

$igamc$ 불완전 감마 함수(Incomplete Gamma Function)로, 카이제곱 분포의 오류함수임

$$igamc(a,x) = \int_x^\infty \frac{t^{a-1} \bullet e^{-t}}{\Gamma(a)} dt$$

$igamc$은 NIST SP 800-22 등을 참조하여 구현함

k 출력 가능한 크기가 ℓ-비트인 샘플 값의 총 개수 또는 A의 위수 ($k = 2^\ell$)

$P-value$ 통계량으로 유도된 유의 확률
수치계산의 결과로 얻어진 $P-value$가 0보다 작거나 1보다 큰 경우, $0 \le P-value \le 1$이 되도록 보정 과정을 거침

$S = (s_1, ..., s_L)$ 크기가 ℓ-비트인 샘플 $s_i (\in A)$로 정렬된 데이터세트 S

x_i 디지털화된 잡음원으로부터 출력된 크기가 ℓ-비트인 샘플

Γ 감마 함수(Gamma Function)

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t}\, dt$$

Γ은 NIST SP 800-22 등을 참조하여 구현함

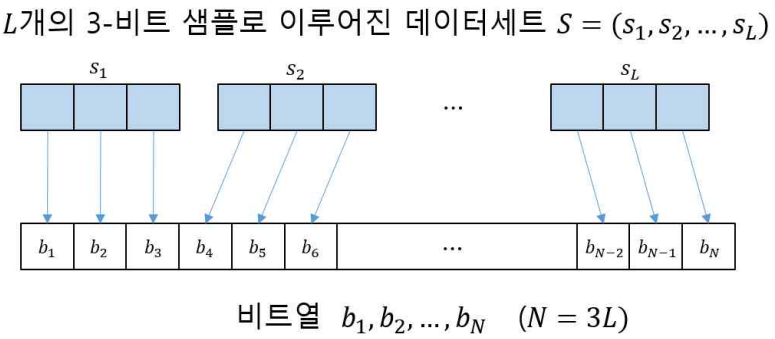
4.3 입력 데이터 구성 방식

4.3.1 비트열 구성 방식

후처리(Post-processing)와 같은 보정을 거치지 않은 디지털화된 단일 잡음원을 수집한다. 수집한 L개의 ℓ-비트 샘플로 이루어진 데이터세트 $S = (s_1, ..., s_L)$은 다음의 과정을 통해 비트열 $b_1, ..., b_N$ $(N = L\ell)$로 변환된다.

입력 : L개의 ℓ-비트 샘플로 이루어진 데이터세트 $S = (s_1, ..., s_L)$
알고리즘 : 1) 1부터 L까지의 i에 대해, 1.1) 1부터 ℓ까지의 j에 대해, $b_{j+i\ell} = (s_i \gg (\ell - j)) \& 1$ // & : 1-비트에 대한 & 연산자
출력 : 비트열 $b_1, ..., b_N$ $(N = L\ell)$

위의 알고리즘은 다음 그림과 같이 진행된다.



비트열 $b_1, ..., b_N$은 “5.1 통계적 난수성 테스트”과 “6.2 Shannon 엔트로피 측정”에서의 알고리즘의 입력으로 사용된다.

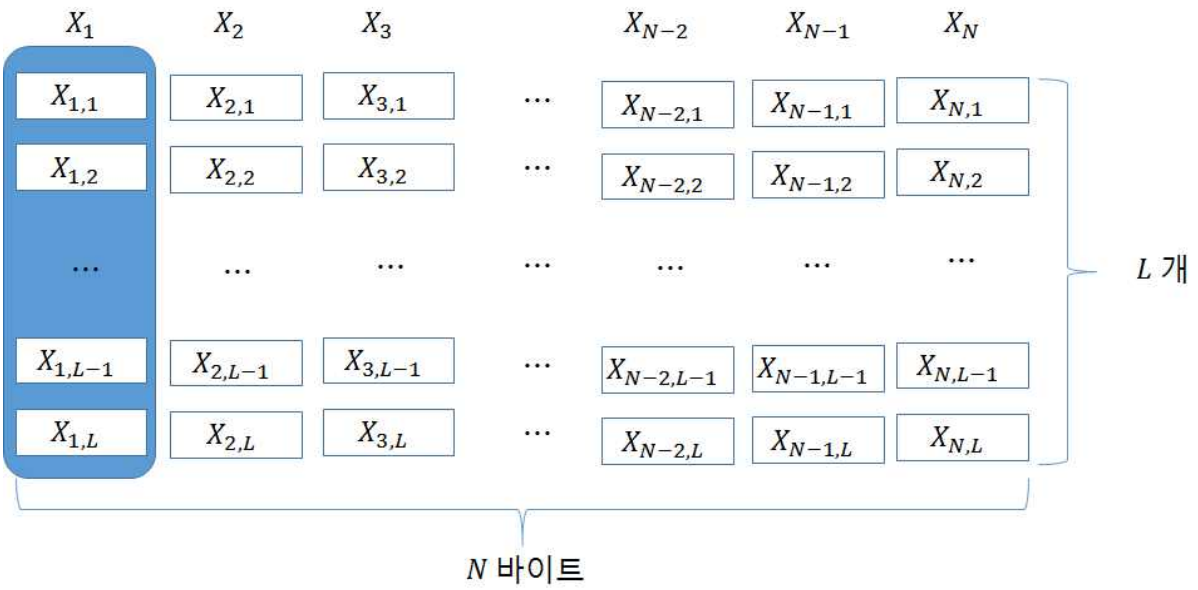
4.3.2 데이터세트 구성 방식

후처리와 같은 보정을 거치지 않은 디지털화된 단일 잡음원을 수집한다. 수집한 L개의 ℓ-비트 샘플로 이루어진 데이터세트 $S = (s_1, ..., s_L)$은 “5.2 잡음원 건전성 테스트”과 “6.1 최소 엔트로피 측정”에서의 알고리즘의 입력으로 사용된다. 이때 각 알고리즘의 계산량, 필요한 메모리 크기 등이 잡음원의 샘플 공간에 영향을 받기 때문에 최대 8-비트 샘플로 이루어진 데이터세트 $S = (s_1, ..., s_L)$만이 입력으로 사용된다.

수집한 잡음원 샘플의 크기가 8-비트보다 큰 경우($\ell > 8$), 샘플 중에서 가장 많이 변하는 비트를 선정하여 8-비트 이하의 샘플로 변환한다. 변환하는 방법은 NIST SP 800-90B(2nd Draft)의 6.4절을 참조한다.

4.3.3 바이트열 구성 방식

후처리와 같은 보정을 거치지 않은 디지털화된 단일 바이트 잡음원을 수집한다.



수집한 N-바이트 잡음원 샘플 L개로 이루어진 바이트열 $X = (X_1, ..., X_N)$은 "7. 상관관계 기반 엔트로피 측정"에서의 알고리즘의 입력으로 사용된다.

5 통계적 테스트

5.1 통계적 난수성 테스트

5.1.1 모노비트(Monobit) 검정

모노비트 검정은 비트열에서 0의 개수와 1의 개수가 균등한지 확인한다. 모노비트 검정 방법은 입력받은 비트열에서 1의 개수를 측정하고, 측정한 1의 개수가 (이상적인) 균등 분포에서의 기댓값에 근사하는지 확인하기 위해 표준정규분포의 오류함수를 이용하여 유의 확률인 $P-value$를 출력하는 것이다.

입력 : 비트열 $b_1, ..., b_N$, 비트열의 길이 N

알고리즘 :

1) 비트열 $b_1, ..., b_N$에서 1의 개수를 측정하여 변수 ctr에 저장한다.
2) 각 변수 μ와 변수 σ에 (이상적인) 균등분포에서의 기댓값과 표준편차를 저장한다.

$$\mu \leftarrow \frac{N}{2}, \quad \sigma \leftarrow \sqrt{\frac{N}{4}}$$

3) $S_{obs} \leftarrow \left| \frac{ctr-\mu}{\sigma} \right|$
4) 표준정규분포의 오류함수 $erfc$를 이용하여 $P-value$를 구한다.

$$P-value \leftarrow erfc\left(\frac{S_{obs}}{\sqrt{2}}\right)$$

출력 : 유의 확률 $P-value$

5.1.2 포커(Poker) 검정

포커 검정은 비트열을 크기가 4-비트인 샘플로 이루어진 데이터세트로 변형시키고, 데이터세트에서 같은 값을 가지는 샘플의 개수가 균등한지 확인한다. 포커 검정 방법은 입력받은 비트열을 크기가 4-비트인 샘플로 이루어진 데이터세트로 변형시키고, 데이터세트에서 같은 값을 가지는 샘플의 개수를 측정하는 것이다. 그런 다음, 측정한 각 샘플의 개수가 샘플크기가 4-비트인 균등분포에서의 기댓값에 근사하는지 확인하기 위해 카이제곱분포의 오류함수를 이용하여 유의 확률인 $P-value$를 출력한다.

입력 : 비트열 $b_1, ..., b_N$, 비트열의 길이 N

알고리즘 :

1) 비트열 $b_1, ..., b_N$을 4-비트 단위의 샘플 d_j로 이루어진 데이터세트 D로 변형시킨다.
   이때 $d_j = 8b_{4j-3} + 4b_{4j-2} + 2b_{4j-1} + b_{4j}$ $\left(j = 1, ..., \frac{N}{4}\right)$이다.
2) 크기가 16인 배열 E를 0으로 초기화한다.
   $E[1] = 0, E[2] = 0, ..., E[16] = 0$
3) 1부터 N까지인 j에 대해, $E[d_j] \leftarrow (E[d_j] + 1)$
4) 카이제곱분포 통계량 z를 다음과 같이 구한다.
$$z \leftarrow \left(\frac{16}{(N/4)}\left(\sum_{i=1}^{16}(E[i])^2\right) - (N/4)\right)$$
5) 다음과 같이 카이제곱분포의 오류함수 $igamc$를 이용하여 $P-value$를 구한다.
$$P-value \leftarrow igamc\left(\frac{15}{2}, \frac{z}{2}\right)$$

출력 : 유의 확률 $P-value$

5.1.3 런(Run) 검정

런은 비트열에서 값이 1 또는 0인 비트가 연속적으로 발생하는 경우를 의미한다. 런 검정은 비트열에서 길이가 l인 런의 발생횟수가 (이상적인) 균등분포에서의 기댓값에 가까운 지 확인한다. 런 검정 방법은 입력받은 비트열에서 각각의 길이가 1, 2, 3, 4, 5, 6 이상인 런의 발생횟수를 측정한다. 이때 길이가 6 이상인 런을 길이가 6인 런으로 여긴다. 측정한 각 길이의 런의 발생횟수가 (이상적인) 균등분포에서의 기댓값에 근사하는지 확인하기 위해 카이제곱분포의 오류함수를 이용하여 유의 확률인 $P-value$를 출력한다. 런 검정은 2개의 유의 확률을 출력하는데, 값이 1인 런을 사용하였을 때의 유의 확률인 $P-value_1$과 값이 0인 런을 사용하였을 때의 유의 확률인 $P-value_0$이다.

입력 : 비트열 $b_1, ..., b_N$, 비트열의 길이 N

알고리즘 :

1) 크기가 2×6인 행렬 R을 0으로 초기화한다.
   $R[1][1] = 0, ..., R[1][6] = 0, R[2][1] = 0, ..., R[2][6] = 0$
2) 길이가 6 이상인 런을 길이가 6인 런이라고 정한다.
3) 1부터 6까지의 i에 대해, 비트열 $b_1, ..., b_N$에서 길이가 i이고 값이 1인 런의 개수를 측정하여 $R[2][i]$에 저장하고, 길이가 i이고 값이 0인 런의 개수를 측정하여 $R[1][i]$에 저장한다.

4) 크기가 6인 배열 μ에 다음과 같이 (이상적인) 균등분포에서의 기댓값을 저장한다. 1부터 6까지의 i에 대해,

$$\mu[i] \leftarrow \frac{N-i+3}{2^{i+2}}$$

5) 값이 1인 런과 값이 0인 런에 대한 카이제곱분포 통계량 z_1, z_0을 다음과 같이 구한다.

$$z_1 \leftarrow \sum_{i=1}^{6}\left(\frac{R[2][i]-\mu[i]}{\mu[i]}\right)^2, \quad z_0 \leftarrow \sum_{i=1}^{6}\left(\frac{R[1][i]-\mu[i]}{\mu[i]}\right)^2$$

6) 카이제곱분포의 오류함수 $igamc$를 이용하여 $P-value_1$와 $P-value_0$를 구한다.

$$P-value_1 \leftarrow igamc\left(\frac{5}{2}, \frac{z_1}{2}\right), \quad P-value_0 \leftarrow igamc\left(\frac{5}{2}, \frac{z_0}{2}\right)$$

출력 : 값이 1인 런에 대한 유의 확률 $P-value_1$,
값이 0인 런에 대한 유의 확률 $P-value_0$

5.1.4 롱 런(Long Run) 검정

롱 런은 비트열에서 값이 1 또는 0인 비트가 비이상적으로 연속적으로 발생하는 경우를 의미한다. 롱 런 검정은 비트열에서 가장 긴 길이의 런이 발생하는지 검정한다. 롱 런 검정 방법은 입력받은 비트열에서 가장 긴 런의 길이를 측정하고, 피보나치 l-스텝 수 (Fibonacci l-step number)를 활용하여 유의 확률인 $P-value$를 출력한다.

입력 : 비트열 b_1, ..., b_N, 비트열의 길이 N

알고리즘 :

1) 비트열 b_1, ..., b_N에서 발생하는 값이 0인 가장 긴 런의 길이는 변수 MR_0에 저장하고, 값이 1인 가장 긴 런의 길이는 변수 MR_1에 저장한다.
2) 비트열 b_1, ..., b_N에서 발생하는 가장 긴 런의 길이는 다음과 같이 계산되어 변수 M에 저장된다.

$$M \leftarrow \max\{MR_0, MR_1\}$$

3) 피보나치 l-스텝 수의 정리를 활용하여 다음과 같이 $P-value$를 구한다.

$$P-value \leftarrow \frac{N-M}{2^M}$$

출력 : 유의 확률 $P-value$

5.1.5 자기 상관관계(Autocorrelation) 검정

자기 상관관계 검정은 비트열을 절반으로 나누어 앞의 절반의 비트열과 뒤의 절반의 비트열이 상관관계를 가지는지 검정한다. 자기 상관관계 검정 방법은 입력받은 비트열을 절반으로 나누어 앞의 절반의 비트열에서 τ에 따른 비트 간의 상관관계를 구하고, 그 값과 (이상적인) 균등분포에서의 기댓값의 차이가 가장 클 때의 τ_0를 찾는 것이다. 그런 다음, 뒤의 절반의 비트열에서 찾은 τ_0에 따른 비트 간의 상관관계를 구하고, 그 값이 (이상적인) 균등분포에서의 기댓값에 근사하는지 확인하기 위해 표준정규분포의 오류함수를 이용하여 유의 확률인 $P-value$을 출력한다.

입력 : 비트열 $b_1, ..., b_N$, 비트열의 길이 N

알고리즘 :

1) 비트열 $b_1, ..., b_N$을 절반으로 나눈다.
2) 앞의 절반의 비트열 $b_1, ..., b_{\frac{N}{2}}$에서 새로운 수열 $Z_\tau \leftarrow \sum_{i=1}^{N/4}(b_i \oplus b_{i+\tau})$, $\left(\tau = 1, ..., \frac{N}{4}\right)$을 생성한다.
3) 각 변수 μ와 변수 σ에 (이상적인) 균등분포에서의 기댓값과 표준편차를 저장한다.
$$\mu \leftarrow \frac{N}{8}, \quad \sigma \leftarrow \sqrt{\frac{N}{16}}$$
4) Z_τ $(\tau = 1, ..., \frac{N}{4})$와 μ의 차이가 가장 클 때의 τ를 찾고, 변수 τ_0에 저장한다. 찾은 τ가 여러 개일 경우, 가장 작은 τ를 τ_0에 저장한다.
5) 뒤의 절반의 비트열 $b_{\frac{N}{2}+1}, ..., b_N$에 대해 다음과 같이 계산하여 변수 c에 저장한다.
$$c \leftarrow \sum_{i=N/2+1}^{3N/4}(b_i \oplus b_{i+\tau_0})$$
6) $S_{obs} \leftarrow \left|\frac{c-\mu}{\sigma}\right|$
7) 표준정규분포의 오류함수 $erfc$를 이용하여 $P-value$를 구한다.
$$P-value \leftarrow erfc\left(\frac{S_{obs}}{\sqrt{2}}\right)$$

출력 : 유의 확률 $P-value$

5.2 잡음원 건전성 테스트

잡음원 건전성 테스트는 잡음원 샘플이 출력되는 동안 이루어지는 실시간 테스트로, 잡음원 샘플을 출력되는 과정에서의 고장을 탐지하거나 동작 환경의 변화로 인해 발생하는 예상하지 않은 동작을 탐지하는 것이 목적이다.

5.2.1 반복 횟수(Repetition count) 테스트

반복 횟수 테스트는 실시간 테스트로, 잡음원의 샘플 당 엔트로피 추정치에 따른 값보다 더 많이 연속적으로 동일한 샘플이 출력되는 돌발고장을 감지한다. 반복 횟수 테스트 방법은 샘플이 컷오프(Cutoff) 값 C보다 더 많이 연속적으로 반복되어 생성되면 오류가 발생했다는 의미로 True를 출력하고, 그렇지 않으면 False를 출력하는 것이다.

반복 횟수 테스트의 입력값인 컷오프 값 C는 $\alpha \geq 2^{-\hat{H}(C-1)}$를 만족시키는 최솟값으로, 다음과 같이 계산된다.

$$C = \left\lceil 1 + \frac{-\log_2 \alpha}{\hat{H}} \right\rceil$$

여기서 α는 허용될 수 있는 유의수준이고, $\hat{H}$는 잡음원의 샘플 당 엔트로피 추정치이다. (※ 일반적으로 유의수준 α은 2^{-40}또는 2^{-50}을 사용한다.)

입력 : 데이터세트 $S = (s_1, ..., s_L)$, 데이터세트의 크기 L, 컷오프 값 C

알고리즘 :

1) 변수 $t \leftarrow s_1$
2) 카운트인 변수 $count$을 1로 초기화한다. ($count \leftarrow 1$)
3) 2부터 L까지의 i에 대해,
   - 3.1) 만약 $s_i = t$이면,
     - 3.1.1) $count \leftarrow (count + 1)$
     - 3.1.2) $count = C$이면,
       오류가 존재하므로, True를 출력한다.
   - 3.2) 만약 $s_i \neq t$이면,
     - 3.2.1) $t \leftarrow s_i$
     - 3.2.2) $count \leftarrow 1$
4) False를 출력한다.

출력 : 오류 발생 여부 T/F

5.2.2 적응성 비율(Adaptive proportion) 테스트

적응성 비율 테스트는 실시간 테스트로, 잡음원의 샘플 당 엔트로피 추정치에 따른 값보다 더 빈번하게 샘플이 출력되는 경우를 감지한다. 적응성 비율 테스트 방법은 윈도우 크기 W내에서 현재 샘플 값이 반복되는 횟수를 계산하고, 반복 횟수가 컷오프 값 C보다 크면 오류가 발생했다는 의미로 True를 출력하고, 그렇지 않으면 False를 출력하는 것이다.
적응성 비율 테스트의 입력값인 컷오프 값 C는 윈도우 크기 W에서 C개 이상의 동일한 샘플을 관찰되는 확률이 최대 유의수준 α인 것으로 계산된다. 이때 샘플이 발생할 확률은 $p = 2^{-\hat{H}}$이고, $\hat{H}$는 잡음원의 샘플 당 엔트로피 추정치이다.
(※ 일반적으로 유의수준 α은 최대 2^{-30}을 사용한다. 또한 이진 잡음원일 경우 $W = 1024$ 이며, 그렇지 않은 경우 $W = 512$를 사용한다.)

입력 : 데이터세트 $S = (s_1, ..., s_W)$, 데이터세트의 크기 W, 컷오프 값 C

알고리즘 :

1) $t \leftarrow s_1$
2) 카운터인 변수 $count$을 1로 초기화한다. ($count \leftarrow 1$)
3) 2부터 W까지의 i에 대해,
   3.1) $s_i = t$이면, $count \leftarrow (count + 1)$
4) $count > C$이면, 오류가 존재하므로 True를 출력한다.
5) $count \leq C$이면, False를 출력한다.

출력 : 오류 발생 여부 T/F

---

6 확률론 및 정보이론 기반 엔트로피 측정

6.1 최소 엔트로피 측정

6.1.1 최빈값(The Most Common Value) 추정

최빈값 추정은 데이터세트에서 가장 많이 발생하는 샘플의 상대도수로 최소 엔트로피를 구한다. 최빈값 추정 방법은 입력받은 데이터세트에서 가장 많이 발생하는 샘플의 비율 $\hat{p}$를 구하고, $\hat{p}$에 대한 99%의 신뢰구간의 상한을 이용하여 샘플 당 최소 엔트로피인 H_{∞}를 추정하는 것이다.

입력 : 데이터세트 $S = (s_1, ..., s_L)$, 데이터세트의 크기 L,
출력 가능한 샘플 값의 총 개수 k

알고리즘 :

1) 데이터세트 S에서 최빈값의 비율 $\hat{p}$를 구한다.

$$\hat{p} \leftarrow \max_{1 \le i \le k} \frac{S\text{에서 } x_i\text{의 개수}}{L}$$

2) $\hat{p}$에 대해 99%의 신뢰구간을 구성하고, 최빈값의 확률 p_u는 이 신뢰구간의 상한을 이용하여 다음과 같이 구한다.

$$p_u \leftarrow \min\left(1,\ \hat{p} + 2.576\sqrt{\frac{\hat{p}(1-\hat{p})}{L}}\right)$$

3) $H_{\infty} \leftarrow (-\log_2(p_u))$

출력 : 샘플 당 최소 엔트로피 H_{∞}

6.1.2 충돌(Collision) 추정

충돌 추정은 데이터세트에서 충돌이 발생하는 시간을 기반으로 하여 최소 엔트로피를 추정한다. 충돌 추정 방법은 입력받은 데이터세트에서 첫 번째 충돌이 일어나기까지의 샘플의 평균수를 측정하고, 측정한 값에 대한 99%의 신뢰구간의 하한과 이진 탐색(Binary search)를 이용하여 샘플 당 최소 엔트로피인 H_{∞}를 추정하는 것이다.

입력 : 데이터세트 $S = (s_1, ..., s_L)$, 데이터세트의 크기 L,
출력 가능한 샘플 값의 총 개수 k

알고리즘 :

1) $v \leftarrow 1, \quad index \leftarrow 1$
2) s_{index}에서 시작하여, $index \le i < j$인 i에 대해 $s_i = s_j$를 만족하는 가장 작은 j를 찾는다.
3) $t_v \leftarrow (j - index + 1), \quad v \leftarrow (v+1), \quad index \leftarrow (j+1)$
4) 데이터세트 S의 마지막에 도달할 때($j = L$)까지 (2 ~ 3)의 과정을 반복한다.
5) $v \leftarrow (v-1)$
6) $\overline{X} \leftarrow \frac{1}{v}\sum_{i=1}^{v} t_i, \quad \hat{\sigma} \leftarrow \sqrt{\frac{1}{v}\sum_{i=1}^{v}(t_i - \overline{X})^2}$
7) $\overline{X'} \leftarrow \left(\overline{X} - 2.576\frac{\hat{\sigma}}{\sqrt{v}}\right)$
8) 이진 탐색을 이용하여 아래의 식을 만족하는 변수 p에 대한 해를 구한다.

$$\overline{X'} = \left(pq^{-2}\left(1 + \frac{1}{k}(p^{-1} - q^{-1})\right) CF(q) - pq^{-1}\frac{1}{k}(p^{-1} - q^{-1})\right)$$

여기서 $q = \frac{1-p}{k-1}$, $p \ge q$이다.

9) 이진 탐색이 해를 찾으면, $H_\infty \leftarrow (-\log_2(p))$
10) 해를 찾지 못하면, $H_\infty \leftarrow (-\log_2(k))$

출력 : 샘플 당 최소 엔트로피 H_∞

6.1.3 마르코브(Markov) 추정

마르코브 추정은 데이터세트에서 연속되는 값들 간의 의존성을 측정하여 최소 엔트로피를 구한다. 마르코브 추정 방법은 바운딩행렬(Bounding matrix)사용하여 전이확률을 계산하고, 이를 이용하여 샘플 당 최소 엔트로피인 H_∞를 추정하는 것이다.

연속되는 값들 간의 의존성을 측정하기 위해서는 많은 데이터가 필요하고, 필요한 데이터 크기는 잡음원에서 나올 수 있는 샘플 값의 개수 k에 따라 달라진다. 따라서 마르코브 추정에서 수용할 수 있는 k은 최대 2^6이다.

입력 : 데이터세트 $S = (s_1, ..., s_L)$, 데이터세트의 크기 L,
출력 가능한 샘플 값의 총 개수 k

알고리즘 :

1) $d \leftarrow 128$

2) $\alpha \leftarrow \min\left(0.99^{k^2},\ 0.99^{d}\right)$

3) 크기가 k인 배열 O와 배열 P를 생성한다.

4) 1부터 k까지의 i에 대해, 데이터세트 S에서 x_i가 발생한 횟수를 $O[i]$에 저장한다.

5) $\epsilon \leftarrow \sqrt{\frac{\log_2\left(\frac{1}{1-\alpha}\right)}{2L}}$

6) 1부터 k까지의 i에 대해, 다음과 같이 각 샘플 값에 대한 초기 확률을 구하여 $P[i]$에 저장한다.

$$P[i] \leftarrow \min\left\{1,\ \frac{O[i]}{L} + \epsilon\right\}$$

7) $O[s_L] \leftarrow (O[s_L] - 1)$

8) 크기가 $k \times k$인 행렬 R와 바운딩행렬 B를 생성하고, 크기가 k인 배열 E를 생성한다.

9) 1부터 k까지의 i와 j에 대해, 데이터세트 S에서 샘플 값이 x_i 값과 같고 다음 순서의 샘플 값이 x_j 값과 같은 경우의 횟수를 $R[i][j]$에 저장한다.

10) 1부터 k까지의 i에 대해,

$$E[i] \leftarrow \sqrt{\frac{\log_2\left(\frac{1}{1-\alpha}\right)}{2\,O[i]}}$$

11) 1부터 k까지의 i와 j에 대해,

$$B[i][j] \leftarrow \begin{cases} 1 & \text{if} \quad O[i] = 0 \\ \min\left\{1,\ \frac{R[i][j]}{O[i]} + E[i]\right\} & \text{otherwise} \end{cases}$$

12) 1부터 $d-1$까지의 m에 대해,

12.1) 크기가 k인 배열 M를 생성한다.

12.2) 1부터 k까지의 n에 대해,

12.2.1) 크기가 k인 배열 P'을 생성한다.

12.2.2) 1부터 k까지의 i에 대해, $P'[i] \leftarrow (P[i] \times B[i][n])$

12.2.3) $M[n] \leftarrow \max_{i=1..k}(P'[i])$

12.3) 1부터 k까지의 i에 대해, $P[i] \leftarrow M[i]$

13) $\hat{p}_{\max} \leftarrow \max_{i=1..k}(P[i])$

14) $H_\infty \leftarrow \left(-\frac{1}{d} log_2(\hat{p}_{\max})\right)$

출력 : 샘플 당 최소 엔트로피 H_∞

6.1.4 압축(Compression) 추정

압축 추정은 데이터세트가 압축될 수 있는 정도를 기반으로 하여 최소 엔트로피를 추정한다. 압축 추정 방법은 Maurer Universal Statistic를 이용한 방법이다. 입력받은 데이터세트를 두 개의 서로 다른 그룹으로 나누고 첫 번째 그룹을 이용하여 딕셔너리(Dictionary)를 생성한다. 생성한 딕셔너리를 기반으로 하고 두 번째 그룹을 테스트 데이터로 사용하여 같은 샘플을 출력하는 데 필요한 샘플의 평균수를 측정한다. 측정한 값에 대한 99%의 신뢰구간의 하한과 이진 탐색을 이용하여 샘플 당 최소 엔트로피인 H_{∞}를 추정하는 것이 압축 추정 방법이다.

(※ 알고리즘 출력 결과의 정확도를 높이기 위해 일반적으로 $L \geq 100{,}000$으로 사용한다. 따라서 소프트웨어 환경에서의 잡음원 샘플을 충분히 수집할 수 있는 경우에 압축 추정 방법을 사용한다.)

입력 : 데이터세트 $S=(s_1, ..., s_L)$, 데이터세트의 크기 L,
출력 가능한 샘플 값의 총 개수 k

알고리즘 :

1) 딕셔너리로 사용하기 위해 크기가 k인 배열 $dict$를 생성하여 0으로 초기화한다.

$$dict[1]=0,\ dict[2]=0,\ ...,\ dict[k]=0$$

2) 1부터 $d=1{,}000$까지의 i에 대해, $dict[s_i] \leftarrow i$

3) 생성한 딕셔너리에 대해 테스트 데이터를 실행한다.

3.1) 크기가 $v=L-d$인 배열 D를 생성한다.

3.2) $d+1$부터 L까지의 i에 대해,

3.2.1) 만약 $dict[s_i] \neq 0$이면,

3.2.1.1) $D[i-d] \leftarrow (i - dict[s_i])$

3.2.1.2) $dict[s_i] \leftarrow i$

3.2.2) 만약 $dict[s_i] = 0$이면,

3.2.2.1) $dict[s_i] \leftarrow i$

3.2.2.2) $D[i-d] \leftarrow i$

4) $b \leftarrow (\lfloor \log_2(k-1) \rfloor + 1)$

5) $\overline{X} \leftarrow \dfrac{\sum_{i=1}^{v} \log_2 D[i]}{v}$, $c \leftarrow (0.7 - \dfrac{0.8}{b} + \dfrac{\left(4+\frac{32}{b}\right)v^{-3/b}}{15})$,

$$\hat{\sigma} \leftarrow c\sqrt{\frac{\sum_{i=1}^{v}(\log_2 D[i])^2}{v} - \overline{X}^2}$$

6) $\overline{X'} \leftarrow (\overline{X} - \dfrac{2.576\,\hat{\sigma}}{\sqrt{v}})$

---

7) 이진 탐색을 이용하여 아래의 식을 만족하는 변수 p에 대한 해를 구한다.

$$\overline{X'} = (G(p) + (n-1)G(q))$$

여기서 $q = \frac{1-p}{k-1}$, $G(z) = \frac{1}{v}\sum_{t=d+1}^{L}\sum_{u=1}^{t}\log_2(u)F(z, t, u)$,

$F(z, t, u) = \begin{cases} z^2(1-z)^{u-1} & \text{if } u < t \\ z(1-z)^{t-1} & \text{if } u = t \end{cases}$ 이다.

8) 이진 탐색이 해를 찾으면, $H_\infty \leftarrow (-\log_2(p))$

9) 해를 찾지 못하면, $H_\infty \leftarrow (-\log_2(k))$

출력 : 샘플 당 최소 엔트로피 H_∞

6.2 Shannon 엔트로피 측정

6.2.1 엔트로피 테스트

엔트로피 테스트는 Maurer의 Universal statistical test를 향상시킨 테스트로, Coron 검정에 따라 수행된다. 비트열 $b_1, ..., b_N$ $(N=(Q+K)T)$을 길이가 T인 워드(Word)로 구성된 수열 $w_1, ..., w_{Q+K}$로 변환하여 앞의 Q개의 워드는 초기화 부분으로 사용하고 나머지 K개의 워드는 테스트 부분으로 사용한다. 엔트로피 테스트 방법은 워드 간의 충돌이 발생하는 최소 거리를 이용하여 T-비트 워드의 Shannon 엔트로피인 H를 추정한다.

(※ 알고리즘 출력 결과의 정확도를 높이기 위해 일반적으로 $Q \geq 10 \times 2^T$과 $K \geq 1{,}000 \times 2^T$로 사용한다. 따라서 소프트웨어 환경에서의 잡음원 샘플을 충분히 수집할 수 있는 경우에 압축 추정 방법을 사용한다.)

입력 : 워드 길이 T, 변수 Q, 변수 K
비트열 $b_1, ..., b_N$ $(N=(Q+K)T)$

알고리즘 :

1) 1부터 $Q+K$까지의 i에 대해, 길이가 T인 각각의 중복이 없는 워드 $w_i = (b_{(i-1)T+1} \| b_{(i-1)T+2} \| ... \| b_{iT})$를 구성한다.
2) 크기가 K인 배열 D를 생성한다.
3) $Q+1$부터 $Q+K$까지의 m에 대해,
   3.1) 1부터 $m-1$까지의 i에 대해, w_m과 같은 값을 가지는 w_{m-i}이 없으면
   $D[m-Q] \leftarrow m$

3.2) 1부터 $m-1$까지의 i에 대해, w_m과 같은 값을 가지는 w_{m-i}이 있으면

$$D[m-Q] \leftarrow \min\{i \mid 1 \le i < m,\ w_m = w_{m-i}\}$$

4) $H \leftarrow \frac{1}{K}\sum_{j=1}^{K} g(D[j])$

여기서 $g(i) = \frac{1}{\log(2)}\sum_{z=1}^{i-1}\frac{1}{z}$이다.

출력 : T-비트 워드의 Shannon 엔트로피 H

6.2.2 상호정보량(Mutual Information) 기반의 엔트로피 추정 방법

상호정보량 기반의 엔트로피 추정 방법은 길이 T의 워드가 가지는 Shannon 엔트로피를 측정하는 Maurer, Coron 방법과 빈도 측정 방식을 결합한 알고리즘으로 상호정보량의 계산효율을 높일 수 있는 엔트로피 측정법이다. 인접한 워드의 상호정보량을 계산하여 워드 열의 독립성을 계산할 때 Coron의 방법보다 효율적이며, 결합 엔트로피의 빠른 계산에도 활용할 수 있다.

(※ 알고리즘 출력 결과의 정확도를 높이기 위해 일반적으로 $Q \ge 10 \times 2^T$과 $K \ge 1{,}000 \times 2^T$로 사용한다. 따라서 소프트웨어 환경에서의 잡음원 샘플을 충분히 수집할 수 있는 경우에 압축 추정 방법을 사용한다.)

입력 : 워드 길이 T, 테스트 파라미터 Q, K,
비트열 $b_1, ..., b_N$ $(N=(Q+K)\,T)$

알고리즘 :

1) 1부터 $Q+K$까지의 i에 대해, 길이가 T인 각각의 중복이 없는 워드 $w_i = (b_{(i-1)T+1} \,\|\, b_{(i-1)T+2} \,\|\, ... \,\|\, b_{iT})$를 구성한다.

[초기화 과정]

2) 카운터로 사용할 2^T 크기의 배열 E를 0으로 초기화한다.

$E[1]=0,\ E[2]=0,\ ...,\ E[2^T]=0$

3) 인덱스 i를 1부터 Q까지 증가시키면서, w_1부터 w_Q까지 각 블록의 빈도를 저장한다.

$$E[w_i] \leftarrow (E[w_i]+1), \quad (i=1,2,...,Q)$$

[계산 과정]

4) 변수를 초기화한다.

$$H \leftarrow 0,\ count \leftarrow 0,\ i \leftarrow (Q+1)$$

5) $i \leq K+Q$ 일 동안 다음 과정을 반복한다.

5.1) $count$를 1 증가시킨다. ($count \leftarrow (count+1)$)

5.2) 카운터 배열의 $E[w_i]$의 값을 1 증가시킨다. ($E[w_i] \leftarrow (E[w_i]+1)$)

5.3) $H \leftarrow (H+((\log_2 i)/E[w_i]))$ 으로 업데이트 한다.

5.4) i을 1 증가 시킨다. ($i \leftarrow (i+1)$)

6) $H \leftarrow (H/count)$ 로 업데이트 한다.

출력 : T-비트 워드의 Shannon 엔트로피 H

7 바이트 상관관계 기반 엔트로피 측정

소프트웨어 잡음원은 일반적으로 바이트 형태의 다양한 가변길이를 갖고 있으며, 바이트 사이의 종속이 존재한다. 바이트 상관관계 기반 엔트로피 측정은 최소변화 분포 수로 필터링을 한 샘플을 대상으로 바이트별 엔트로피를 추정하고 바이트간의 상관관계를 반영하여 최종 엔트로피를 추정하는 것이다.

엔트로피 계산 알고리즘은 다음과 같다.

입력 : N-바이트로 구성된 L개의 잡음원에 대한 바이트열 $X=(X_1, ..., X_N)$

여기서, $X_i=(X_{i,1}, X_{i,2} ..., X_{i,L})$, $1 \le i \le N$, $L=256\times\alpha$ $(\alpha \ge 1)$

최소변화 분포수: min_distub

알고리즘 :

[분포 계산을 통한 바이트별 필터링 수행]

1) 1부터 N까지의 i에 대해,

1.1) $X_{i,1}, X_{i,2}, ..., X_{i,L}$의 분포가 min_distub 보다 작으면 해당 컬럼을 제외시킨다.

1.2) min_distub 보다 큰 k 바이트열인 $X'=(X_1, X_2 ..., X_k)$ 만을 선택한다. ($k \le N$)

[필터링된 바이트열에 대하여 바이트열간 상관관계 계산]

2) 1부터 k까지의 i와 j에 대해, Pearson 상관관계를 적용한 correlation을 계산한다.

$$C_{i,j} = \begin{cases} cor(X_i, X_j) & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases}$$

[필터링된 바이트열에 대하여 엔트로피 계산]

3) 1부터 k까지의 i에 대해, 엔트로피 계산을 선택한다.

(여기서, 확률분포 p는 바이트열 X_i의 구성요소 $(X_{i,1}, X_{i,2} ..., X_{i,L})$에 대한 확률임)

3.1) 확률 분포 기반으로 섀논 엔트로피를 계산한다.

$$e_i = \sum_{t=1}^{L}(-p_t \log_2 p_t) \quad // Shannon\ Entropy \text{ 적용 시}$$

3.2) 발생빈도에 따라 최대 확률 기반인 최소 엔트로피를 계산한다.

$$e_i = -\log_2(\max(p_1, ..., p_L)) \quad // Min-entropy \text{ 적용 시}$$

[상관관계를 반영하여 엔트로피 재산정]

4) 1부터 k까지의 i와 j에 대해,

4.1) i번째 컬럼에 대한 상관관계 총합을 계산한다.

$$C_i' \mathrel{+}= |C_{i,j}|$$

4.2) i번째 컬럼에 대한 엔트로피를 재산정한다.

$$e'_i \leftarrow e_i\left[1-\frac{1}{k}C_i^{'}\right]$$

[최종 엔트로피 계산]

5) 1부터 k까지의 i에 대해, $e_f += e_i$

출력 : 잡음원에 대한 최종 엔트로피 e_f

부 록 Ⅰ-1

지식재산권 확약서 정보

Ⅰ-1.1 지식재산권 확약서(1) : 해당 사항 없음

- 발명의 명칭
- 권리자의 성명
- 등록(출원) 번호
- 등록(출원) 연월일
- 실시조건
- 확약서 접수일

Ⅰ-1.2 지식재산권 확약서(2) : 해당 사항 없음

- 발명의 명칭
- 권리자의 성명
- 등록(출원) 번호
- 등록(출원) 연월일
- 실시조건
- 확약서 접수일

부 록 Ⅰ-2

시험인증 관련 사항

Ⅰ-2.1 시험인증 대상 여부 : 해당 사항 없음

Ⅰ-2.2 시험표준 제정 현황 : 해당 사항 없음

부 록 Ⅰ-3

본 표준의 연계(family) 표준

해당 사항 없음

부 록 Ⅰ-4

참고 문헌

[1] E. Barker and J. kelsey, “Recommendation for the Entropy Sources Used for Random Bit Generation”, NIST SP 800-90B(2nd Draft), January, 2016.

[2] E. Barker and J. kelsey, “Recommendation for the Entropy Sources Used for Random Bit Generation”, NIST SP 800-90B(1st Draft), August, 2012.

[3] Wolfgang Killmann and Werner Schindler, “A Proposal for : Functionality Classes and evaluation methodology for true(physical) random number generators”, BSI AIS.31, September, 2001.

[4] Young-Sik Kim, Yongjin Yeom, and Hee Bong Choi, “Online Test Based on Mutual Information for True Random Number Generators”, J. Korean Math. Soc. 50, No. 4, 2013.

[5] 박호중, 강주성, 염용진, “진난수발생기용 난수성 검정 방법 AIS.31에 대한 확률론적 분석 및 보안성 평가 적용 방법”, 한국정보보호학회논문지, 2016.

[6] Andrew Rukhin, et al. “A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications,” NIST Special Publication 800-22 revision 1a, Apr. 2010.

부 록 Ⅰ-5

영문표준 해설서

해당 사항 없음

부 록 Ⅰ-6

표준의 이력

판수	채택일	표준번호	내용	담당 위원회
제1판	2017.06.xx	제정 TTAKS.KO-12.xxxx	소프트웨어 환경에서의 잡음원 엔트로피 검증 알고리즘	정보보호기반 (PG501)

부 록 II

참조구현 값

본 부록에서는 통계적 테스트 및 수학적 엔트로피 측정 알고리즘의 구현 적합성 확인을 위한 참조구현 값(Test vectors)을 제공한다.

참조구현 값은 각 알고리즘의 입력값, 출력값, 단계별 계산 결과를 포함한다. 개별 참조구현 값의 배열은 해당 단위의 띄어쓰기로 한다.

참조구현 값으로 사용된 각 알고리즘의 입력값은 잡음원으로 사용하지 않는다.
참조구현 값을 작성하기 위해 각 알고리즘을 구현할 때, $erfc$, $igamc$, Γ 함수는 NIST SP 800-22을 참조하였고, CF 함수는 NIST SP 800-90B를 참조하였다.

II.1 통계적 테스트

II.1.1 통계적 난수성 테스트

N=50,000인 비트열 b_1, ...,b_N을 아래의 의사 코드로 구성하여 모노비트, 포커, 런, 롱런, 자기 상관관계 검정의 입력으로 사용한다.

```
unsigned char cState[16]={1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0};
for (int i=1; i<N+1; i++)
{
    b_i = cState[16];
    unsigned char temp=(cState[16]^cState[15]^cState[13]^cState[4])&1;
    for (int j=16; j>1; j--)
      cState[j]=cState[j-1];
    cState[1]=temp;
}
```

Ⅱ.1.1.1 모노비트 검정

[단계 1] ctr:
ctr = 24960

[단계 2] μ, σ:
μ = 25000
σ = 111.803

[단계 3] S_{obs}:
S_{obs} = 0.357

[단계 4] $P-value$:
$P-value$ = 0.720

[출력] $P-value$:
$P-value$ = 0.720

Ⅱ.1.1.2 포커 검정

[단계 3] $E[i]\,(1 \le i \le 16)$:
$E[i]$

823	776	837	757	731	748	782	797	748	731
809	772	823	792	811	763				

[단계 4] z:
z = 21.643

[단계 5] $P-value$:
$P-value$ = 0.117

[출력] $P-value$:
$P-value$ = 0.117

Ⅱ.1.1.3 런 검정

[단계 3] $R[i][j]\,(1 \le i \le 2,\ 1 \le j \le 6)$:
$R[1][j]$

6248	3114	1572	770	393	401

$R[2][j]$

6267 3107 1557 789 395 383

[단계 4] μ:

μ = 195.308

[단계 5] z_1, z_0:

z_1 = 0.923

z_0 = 1.109

[단계 6] $P-value_1$, $P-value_0$:

$P-value_1$ = 0.968

$P-value_0$ = 0.953

[출력] $P-value_1$, $P-value_0$:

$P-value_1$ = 0.968

$P-value_0$ = 0.953

II.1.1.4 롱 런 검정

[단계 2] M:

M = 16

[단계 3] $P-value$:

$P-value$ = 0.762

[출력] $P-value$:

$P-value$ = 0.762

II.1.1.5 자기 상관관계 검정

[단계 3] μ, σ:

μ = 6250

σ = 55.901

[단계 4] τ_0:

τ_0 = 6171

[단계 5] c:

c = 6263

[단계 6] S_{obs}:

S_{obs} = 0.232

[단계 7] $P-value$:

$P-value$ = 0.816

[출력] $P-value$:

$P-value$ = 0.816

Ⅱ.1.2 잡음원 건전성 테스트

50,000개(L=50,000)의 4-bit(ℓ=4) 샘플로 이루어진 데이터세트 $S=(s_1, ..., s_L)$을 아래의 의사 코드로 구성하여 반복 횟수 테스트와 적응 비율 테스트의 입력으로 사용한다.

```
unsigned char cState[16]={1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0};
for (int i=1; i<L+1; i++)
{
    s_i = 0;
    for (int n=1; n<ℓ+1; n++)
    {
      s_i ^= (cState[16]<<(ℓ-n));
      unsigned char temp=(cState[16]^cState[15]^cState[13]^cState[4])&1;
      for (int j=16; j>1; j--)
        cState[j]=cState[j-1];
      cState[1]=temp;
    }
}
```

Ⅱ.1.2.1 반복 횟수 테스트

반복 횟수 테스트의 입력으로 사용되는 컷오프 값 C를 12로 설정한다. (C=12)

[단계 3.1.1] $count$:

$count$ = 2

[출력] : 오류 발생 여부 T/F

False

Ⅱ.1.2.2 적응 비율 테스트

반복 횟수 테스트의 입력으로 사용되는 윈도우 크기 W와 컷오프 값 C를 W=512, C=80으로 설정한다.

[단계 3.1] $count$: $count$ = 31
[출력] : 오류 발생 여부 T/F False

Ⅱ.2 확률론 및 정보이론 기반 엔트로피 측정

Ⅱ.2.1 최소 엔트로피 측정

50,000개(L=50,000)의 4-bit(ℓ=4) 샘플로 이루어진 데이터세트 $S=(s_1, ..., s_L)$을 아래의 의사 코드로 구성하여 최빈값, 충돌, 마르코브, 압축 추정의 입력으로 사용한다.

```
unsigned char cState[16]={1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0};
for (int i=1; i<L+1; i++)
{
    s_i = 0;
    for (int n=1; n<ℓ+1; n++)
    {
      s_i ^= (cState[16]<<(ℓ-n));
      unsigned char temp=(cState[16]^cState[15]^cState[13]^cState[4])&1;
      for (int j=16; j>1; j--)
        cState[j]=cState[j-1];
      cState[1]=temp;
    }
}
```

II.2.1.1 최빈값 추정

[단계 1] $\hat{p} \leftarrow \max_{1 \le i \le k} \frac{S\text{에서 } x_i\text{의 개수}}{L}$:

$\hat{p}$ = 0.064

[단계 2] $p_u \leftarrow \min\left(1,\ \hat{p} + 2.576\sqrt{\frac{\hat{p}(1-\hat{p})}{L}}\right)$:

p_u = 0.067

[단계 3] $H_\infty \leftarrow (-\log_2(p_u))$:

H_∞ = 3.907

[출력] H_∞:

H_∞ = 3.907

II.2.1.2 충돌 추정

[단계 5] $v \leftarrow (v-1)$:

v = 8725

[단계 6] $\overline{X} \leftarrow \frac{1}{v}\sum_{i=1}^{v} t_i, \quad \hat{\sigma} \leftarrow \sqrt{\frac{1}{v}\sum_{i=1}^{v}(t_i - \overline{X})^2}$:

$\overline{X}$ = 5.732

$\hat{\sigma}$ = 2.229

[단계 7] $\overline{X'} \leftarrow \left(\overline{X} - 2.576\frac{\hat{\sigma}}{\sqrt{v}}\right)$:

$\overline{X'}$ = 5.669

[단계 9] $H_\infty \leftarrow (-\log_2(p))$:

p = 0.096

H_∞ = 3.375

[출력] H_∞:

H_∞ = 3.375

II.2.1.3 마르코브 추정

[단계 2] $\alpha \leftarrow \min\left(0.99^{k^2},\ 0.99^{d}\right)$:

k = 16

α = 0.076

[단계 5] $\epsilon \leftarrow \sqrt{\frac{\log_2\left(\frac{1}{1-\alpha}\right)}{2L}}$:

ϵ = 0.001

[단계 6] $P[i] \leftarrow \min\left\{1,\ \frac{O[i]}{L} + \epsilon\right\},\ (1 \le i \le k)$:

$P[i]$

0.064	0.064	0.065	0.063	0.064	0.063	0.062	0.063
0.063	0.063	0.065	0.064	0.065	0.064	0.063	0.064

[단계 7] $O[s_L] \leftarrow (O[s_L] - 1)$:

$o[s_L]$ = 3192

[단계 10] $E[i] \leftarrow \sqrt{\frac{\log_2\left(\frac{1}{1-\alpha}\right)}{2\,O[i]}}$, $(1 \le i \le k)$:

$E[i]$

0.004270	0.004264	0.004243	0.004310	0.004282	0.004299
0.004324	0.004301	0.004306	0.004301	0.004248	0.004271
0.004236	0.004258	0.004308	0.004273		

[단계 14] $H_\infty \leftarrow \left(-\frac{1}{d} log_2(\hat{p}_{\max})\right)$:

$-\log_2(\hat{p}_{\max})$ = 488.876

H_∞ = 3.819

[출력] H_∞:

H_∞ = 3.819

Ⅱ.2.1.4 압축 추정

[단계 4] $b \leftarrow (\lfloor \log_2(k-1) \rfloor + 1)$:

b = 4

[단계 5] $\overline{X} \leftarrow \dfrac{\sum_{i=1}^{v} \log_2 D[i]}{v}$, $c \leftarrow \left(0.7 - \dfrac{0.8}{b} + \dfrac{\left(4 + \dfrac{32}{b}\right) v^{-3/b}}{15}\right)$,

$\hat{\sigma} \leftarrow c \sqrt{\dfrac{\sum_{i=1}^{v} (\log_2 D[i])^2}{v} - \overline{X}^2}$:

$\overline{X}$ = 3.309

c = 0.500

$\hat{\sigma}$ = 0.769

[단계 6] $\overline{X'} \leftarrow \left(\overline{X} - \dfrac{2.576\,\hat{\sigma}}{\sqrt{v}}\right)$:

$\overline{X'}$ = 3.3

[단계 8] $H_\infty \leftarrow (-\log_2(p))$:

p = 0.098

H_∞ = 3.354

[출력] H_∞:

H_∞ = 3.354

Ⅱ.2.2 Shannon 엔트로피 측정

N=216인 비트열 $b_1, ..., b_N$을 아래의 의사 코드로 구성하여 엔트로피 테스트와 상호정보량 기반의 엔트로피 추정 방법의 입력으로 사용한다.

```
unsigned char cState[16]={1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0};
for (i=1; i<N+1; i++)
{
  b_i = cState[16];
  unsigned char temp=(cState[16]^cState[15]^cState[13]^cState[4])&1;
  for (int j=16; j>1; j--)
    cState[j]=cState[j-1];
  cState[1]=temp;
}
```

엔트로피 테스트와 상호정보량 기반의 엔트로피 추정 방법의 입력의 입력으로 사용되는 워드 길이 T, 변수 Q, 변수 K를 T = 4, Q = 4, K = 50로 설정한다.

Ⅱ.2.2.1 엔트로피 테스트

[단계 4] $H \leftarrow \frac{1}{K}\sum_{j=1}^{K} g(D[j])$:

$\sum_{j=1}^{K} g(D[j])$ = 198.362

H = 3.967

[출력] H:

H = 3.967

Ⅱ.2.2.2 상호정보량 기반의 엔트로피 추정 방법

[단계 3] 수행 이후의 $E[j]$ $(j=1,...,2^T)$:

$E[j]$ $(j=1,...,2^T)$

0 0 0 0 0 4 0 0 0 0 0 0 0 0 0 0

[단계 5] 수행 이후의 $E[j]$ $(j=1,...,2^T)$:

$E[j]$ $(j=1,...,2^T)$

2 2 6 2 5 9 3 1 2 4 3 0 3 3 6 3

[단계 5] 수행 이후의 H:

H = 112.736

[단계 6] H:

H = 2.255

[출력] H:

H = 2.255

부 록 III

소프트웨어 잡음원 검증 사례

III.1 바이트 상관관계 기반 엔트로피 측정

□ 잡음원 수집

- 운영체제: Windows 7 (32비트)
- 수집함수명: QueryPerformanceCounter
- 잡음원 샘플 크기: 8바이트

	1st	2nd	3rd	4th	5th	6th	7th	8th
1	00	00	00	03	7C	9F	D7	9E
2	00	00	00	03	EC	C6	AE	A0
3	00	00	00	03	4D	4A	9B	A2
4	00	00	00	03	E9	3A	85	A4
5	00	00	00	03	7A	C2	70	A6
6	00	00	00	03	70	B1	5C	A8
7	00	00	00	03	56	CA	48	AA
8	00	00	00	03	35	8F	34	AC
9	00	00	00	03	C0	58	20	AE
10	00	00	00	03	1	27	0C	B0
11	00	00	00	03	43	6A	F8	B1
12	00	00	00	03	C2	F1	E4	B3
13	00	00	00	03	94	B2	D0	B5
14	00	00	00	03	A8	2	BD	B7
15	00	00	00	03	69	BF	A7	B9
16	00	00	00	03	1A	53	94	BB
17	00	00	00	03	44	13	7F	BD
18	00	00	00	03	6F	6A	6B	BF
19	00	00	00	03	F6	DA	57	C1
20	00	00	00	03	2E	2D	45	C3
21	00	00	00	03	6F	58	33	C5
22	00	00	00	03	1E	52	1D	C7
23	00	00	00	03	28	1A	0D	C9
24	00	00	00	03	29	B9	F3	CA
25	00	00	00	03	3A	7F	DE	CC
26	00	00	00	03	BD	A6	C8	CE
27	00	00	00	03	68	3A	B4	D0
28	00	00	00	03	F6	AD	9F	D2
29	00	00	00	03	65	57	8B	D4
30	00	00	00	03	12	E3	76	D6
31	00	00	00	03	97	7E	62	D8

	1st	2nd	3rd	4th	5th	6th	7th	8th
32	00	00	00	03	60	3B	4E	DA
33	00	00	00	03	DA	A1	3A	DC
34	00	00	00	03	33	3B	27	DE
35	00	00	00	03	A3	F4	11	E0
36	00	00	00	03	1D	22	FF	E1
37	00	00	00	03	38	97	EB	E3
38	00	00	00	03	21	B5	D7	E5
39	00	00	00	03	2D	1C	C4	E7
40	00	00	00	03	78	B2	AF	E9
41	00	00	00	03	81	4E	9B	EB
42	00	00	00	03	3E	8	87	ED
43	00	00	00	03	2F	88	72	EF
44	00	00	00	03	96	E6	5D	F1
45	00	00	00	03	F8	82	49	F3
46	00	00	00	03	D3	7	36	F5
47	00	00	00	03	CA	C0	20	F7
48	00	00	00	03	FC	5D	0C	F9
49	00	00	00	03	ED	F8	F7	FA
50	00	00	00	03	F7	62	E3	FC
51	00	00	00	03	8A	FA	CE	FE
52	00	00	00	04	56	95	BA	0
53	00	00	00	04	0A	2F	A6	2
54	00	00	00	04	E8	CE	91	4
55	00	00	00	04	D3	64	7D	6
56	00	00	00	04	25	1	69	8
57	00	00	00	04	9B	96	54	0A
58	00	00	00	04	A8	2D	40	0C
59	00	00	00	04	D0	AB	2C	0E
60	00	00	00	04	FF	2A	19	10
61	00	00	00	04	77	99	5	12
62	00	00	00	04	0F	7	F2	13
63	00	00	00	04	A8	78	DE	15
64	00	00	00	04	A3	36	C9	17
65	00	00	00	04	0E	D1	B4	19
66	00	00	00	04	8E	6B	A0	1B
67	00	00	00	04	6A	4	8C	1D
68	00	00	00	04	D5	9F	77	1F
69	00	00	00	04	EE	3A	63	21
70	00	00	00	04	5E	CE	4E	23
71	00	00	00	04	29	66	3A	25
72	00	00	00	04	7	5	26	27
73	00	00	00	04	86	9A	11	29
74	00	00	00	04	6C	38	FD	2A
75	00	00	00	04	11	D2	E8	2C
76	00	00	00	04	2F	3D	DA	2E
77	00	00	00	04	0B	67	C0	30
78	00	00	00	04	DF	D2	AB	32

	1st	2nd	3rd	4th	5th	6th	7th	8th
79	00	00	00	04	5E	6F	97	34
80	00	00	00	04	3E	7	83	36
81	00	00	00	04	FC	9F	6E	38
82	00	00	00	04	1A	BA	59	3A
83	00	00	00	04	E4	D9	45	3C
84	00	00	00	04	59	4E	32	3E
85	00	00	00	04	C7	BA	1E	40
86	00	00	00	04	B5	31	0B	42
87	00	00	00	04	E1	A7	F7	43
88	00	00	00	04	C2	21	E4	45
89	00	00	00	04	A8	DC	CE	47
90	00	00	00	04	42	4F	BB	49
91	00	00	00	04	60	E4	A7	4B
92	00	00	00	04	C7	78	93	4D
93	00	00	00	04	8A	44	7D	4F
94	00	00	00	04	39	AC	6A	51
95	00	00	00	04	5E	95	56	53
96	00	00	00	04	53	DE	41	55
97	00	00	00	04	96	77	2D	57
98	00	00	00	04	39	0D	19	59
99	00	00	00	04	19	AA	4	5B
100	00	00	00	04	75	22	F1	5C
101	00	00	00	04	BD	DF	DB	5E
102	00	00	00	04	2D	7A	C7	60
103	00	00	00	04	2A	3A	B3	62
104	00	00	00	04	44	93	9E	64
105	00	00	00	04	7C	76	89	66
106	00	00	00	04	4D	EC	74	68
107	00	00	00	04	47	88	60	6A
108	00	00	00	04	0F	22	4C	6C
109	00	00	00	04	9A	BE	37	6E
110	00	00	00	04	E5	50	23	70
111	00	00	00	04	B0	F2	0E	72
112	00	00	00	04	CB	85	FA	73
113	00	00	00	04	43	44	E6	75
114	00	00	00	04	A6	BC	D1	77
115	00	00	00	04	A7	55	BD	79
116	00	00	00	04	2B	21	A9	7B
117	00	00	00	04	ED	65	95	7D
118	00	00	00	04	E2	D8	81	7F
119	00	00	00	04	94	4F	6E	81
120	00	00	00	04	36	CD	5A	83
121	00	00	00	04	FE	5E	47	85
122	00	00	00	04	7	B0	33	87
123	00	00	00	04	42	91	1D	89
124	00	00	00	04	48	7	0A	8B
125	00	00	00	04	7	C8	F4	8C

	1st	2nd	3rd	4th	5th	6th	7th	8th
126	00	00	00	04	77	36	E1	8E
127	00	00	00	04	D4	D3	CC	90
128	00	00	00	04	0F	6E	B8	92
129	00	00	00	04	98	6	A4	94
130	00	00	00	04	8E	71	8F	96
131	00	00	00	04	1B	40	7B	98
132	00	00	00	04	60	D4	66	9A
133	00	00	00	04	29	6D	52	9C
134	00	00	00	04	B8	7	3E	9E
135	00	00	00	04	1E	CE	28	A0
136	00	00	00	04	6	3F	15	A2
137	00	00	00	04	80	D7	0	A4
138	00	00	00	04	64	77	EC	A5
139	00	00	00	04	60	31	D7	A7
140	00	00	00	04	BC	A3	C3	A9
141	00	00	00	04	CC	15	B0	AB
142	00	00	00	04	FB	5B	9E	AD
143	00	00	00	04	E4	D4	8A	AF
144	00	00	00	04	93	90	75	B1
145	00	00	00	04	FC	7	62	B3
146	00	00	00	04	94	78	4E	B5
147	00	00	00	04	DE	C9	3B	B7
148	00	00	00	04	1E	3A	28	B9
149	00	00	00	04	D8	B3	14	BB
150	00	00	00	04	79	4B	0	BD
151	00	00	00	04	57	B9	EC	BE
152	00	00	00	04	42	33	D9	C0
153	00	00	00	04	43	8F	C6	C2
154	00	00	00	04	E5	2A	B2	C4
155	00	00	00	04	2F	AD	9D	C6
156	00	00	00	04	9A	25	8A	C8
157	00	00	00	04	F4	EC	74	CA
158	00	00	00	04	67	AB	5F	CC
159	00	00	00	04	AD	41	4B	CE
160	00	00	00	04	61	B9	37	D0
161	00	00	00	04	1B	79	22	D2
162	00	00	00	04	A0	EF	0E	D4
163	00	00	00	04	8E	A6	F9	D5
164	00	00	00	04	32	1E	E6	D7
165	00	00	00	04	CC	E3	D0	D9
166	00	00	00	04	55	3E	BB	DB
167	00	00	00	04	AC	8C	AA	DD
168	00	00	00	04	B6	83	94	DF
169	00	00	00	04	0B	1D	80	E1
170	00	00	00	04	B7	B7	6B	E3
171	00	00	00	04	B2	4C	57	E5
172	00	00	00	04	1C	E9	42	E7

	1st	2nd	3rd	4th	5th	6th	7th	8th
173	00	00	00	04	F3	87	2E	E9
174	00	00	00	04	76	20	1A	EB
175	00	00	00	04	2F	BB	5	ED
176	00	00	00	04	18	5C	F1	EE
177	00	00	00	04	C0	ED	DC	F0
178	00	00	00	04	DD	81	C8	F2
179	00	00	00	04	C4	E7	B4	F4
180	00	00	00	04	26	74	A1	F6
181	00	00	00	04	1D	F1	8E	F8
182	00	00	00	04	91	86	7A	FA
183	00	00	00	04	BC	F3	66	FC
184	00	00	00	04	8A	8D	52	FE
185	00	00	00	05	89	AF	3F	0
186	00	00	00	05	FB	DC	2D	2
187	00	00	00	05	CF	76	18	4
188	00	00	00	05	98	42	4	6
189	00	00	00	05	FB	F9	EE	7
190	00	00	00	05	2E	BD	D9	9
191	00	00	00	05	75	2E	C6	0B
192	00	00	00	05	E9	89	B2	0D
193	00	00	00	05	D5	62	9D	0F
194	00	00	00	05	7B	D0	89	11
195	00	00	00	05	79	98	74	13
196	00	00	00	05	19	7	61	15
197	00	00	00	05	BC	8E	4C	17
198	00	00	00	05	4	1B	39	19
199	00	00	00	05	19	27	24	1B
200	00	00	00	05	CF	95	0E	1D
201	00	00	00	05	20	35	FA	1E
202	00	00	00	05	DD	A7	E6	20
203	00	00	00	05	EC	40	D2	22
204	00	00	00	05	AD	DB	BD	24
205	00	00	00	05	E2	57	AA	26
206	00	00	00	05	3F	9	95	28
207	00	00	00	05	95	82	81	2A
208	00	00	00	05	32	F7	6D	2C
209	00	00	00	05	5E	E3	57	2E
210	00	00	00	05	F8	E4	42	30
211	00	00	00	05	EE	E8	2F	32
212	00	00	00	05	B7	32	19	34
213	00	00	00	05	4	17	6	36
214	00	00	00	05	BC	DA	F1	37
215	00	00	00	05	E9	E2	DF	39
216	00	00	00	05	40	CE	C9	3B
217	00	00	00	05	6D	90	BB	3D
218	00	00	00	05	CF	21	A9	3F
219	00	00	00	05	BB	74	99	41

	1st	2nd	3rd	4th	5th	6th	7th	8th
220	00	00	00	05	9	52	80	43
221	00	00	00	05	F5	82	66	45
222	00	00	00	05	B3	25	52	47
223	00	00	00	05	15	88	3E	49
224	00	00	00	05	1D	AD	2C	4B
225	00	00	00	05	B6	AE	15	4D
226	00	00	00	05	9B	28	2	4F
227	00	00	00	05	22	ED	F2	50
228	00	00	00	05	BA	C8	DC	52
229	00	00	00	05	5E	CF	C5	54
230	00	00	00	05	26	31	AF	56
231	00	00	00	05	90	CF	9A	58
232	00	00	00	05	6	17	86	5A
233	00	00	00	05	9E	AF	71	5C
234	00	00	00	05	3E	4D	5D	5E
235	00	00	00	05	A4	E7	48	60
236	00	00	00	05	C5	22	34	62
237	00	00	00	05	21	62	20	64
238	00	00	00	05	9F	9E	0B	66
239	00	00	00	05	2C	50	F7	67
240	00	00	00	05	67	E7	E2	69
241	00	00	00	05	3	86	CE	6B
242	00	00	00	05	59	20	BA	6D
243	00	00	00	05	26	B7	A5	6F
244	00	00	00	05	AC	28	92	71
245	00	00	00	05	FF	EA	7C	73
246	00	00	00	05	66	62	69	75
247	00	00	00	05	27	AC	56	77
248	00	00	00	05	E0	46	42	79
249	00	00	00	05	F9	BD	2F	7B
250	00	00	00	05	A5	E5	20	7D
251	00	00	00	05	D6	E3	6	7F
252	00	00	00	05	F8	22	F4	80
253	00	00	00	05	46	98	E0	82
254	00	00	00	05	5B	2E	CE	84
255	00	00	00	05	75	77	B9	86
256	00	00	00	05	E3	79	A4	88
분포수	1	1	1	3	192	164	166	158
Shannon	0.00	0.00	0.00	0.00	8.00	7.75	7.71	7.66
Min	0.00	0.00	0.00	0.00	2.00	1.48	1.51	1.38

□ 바이트 필터링

(분포수: 4)

	1st	2nd	3rd	4th
1	7C	9F	D7	9E
2	EC	C6	AE	A0
3	4D	4A	9B	A2
4	E9	3A	85	A4
5	7A	C2	70	A6
6	70	B1	5C	A8
7	56	CA	48	AA
8	35	8F	34	AC
9	C0	58	20	AE
10	1	27	0C	B0
11	43	6A	F8	B1
12	C2	F1	E4	B3
13	94	B2	D0	B5
14	A8	2	BD	B7
15	69	BF	A7	B9
16	1A	53	94	BB
17	44	13	7F	BD
18	6F	6A	6B	BF
19	F6	DA	57	C1
20	2E	2D	45	C3
21	6F	58	33	C5
22	1E	52	1D	C7
23	28	1A	0D	C9
24	29	B9	F3	CA
25	3A	7F	DE	CC
26	BD	A6	C8	CE
27	68	3A	B4	D0
28	F6	AD	9F	D2
29	65	57	8B	D4
30	12	E3	76	D6
31	97	7E	62	D8
32	60	3B	4E	DA
33	DA	A1	3A	DC
34	33	3B	27	DE
35	A3	F4	11	E0
36	1D	22	FF	E1
37	38	97	EB	E3
38	21	B5	D7	E5
39	2D	1C	C4	E7
40	78	B2	AF	E9
41	81	4E	9B	EB
42	3E	8	87	ED
43	2F	88	72	EF
44	96	E6	5D	F1
45	F8	82	49	F3

	1st	2nd	3rd	4th
46	D3	7	36	F5
47	CA	C0	20	F7
48	FC	5D	0C	F9
49	ED	F8	F7	FA
50	F7	62	E3	FC
51	8A	FA	CE	FE
52	56	95	BA	0
53	0A	2F	A6	2
54	E8	CE	91	4
55	D3	64	7D	6
56	25	1	69	8
57	9B	96	54	0A
58	A8	2D	40	0C
59	D0	AB	2C	0E
60	FF	2A	19	10
61	77	99	5	12
62	0F	7	F2	13
63	A8	78	DE	15
64	A3	36	C9	17
65	0E	D1	B4	19
66	8E	6B	A0	1B
67	6A	4	8C	1D
68	D5	9F	77	1F
69	EE	3A	63	21
70	5E	CE	4E	23
71	29	66	3A	25
72	7	5	26	27
73	86	9A	11	29
74	6C	38	FD	2A
75	11	D2	E8	2C
76	2F	3D	DA	2E
77	0B	67	C0	30
78	DF	D2	AB	32
79	5E	6F	97	34
80	3E	7	83	36
81	FC	9F	6E	38
82	1A	BA	59	3A
83	E4	D9	45	3C
84	59	4E	32	3E
85	C7	BA	1E	40
86	B5	31	0B	42
87	E1	A7	F7	43
88	C2	21	E4	45
89	A8	DC	CE	47
90	42	4F	BB	49
91	60	E4	A7	4B
92	C7	78	93	4D

	1st	2nd	3rd	4th
93	8A	44	7D	4F
94	39	AC	6A	51
95	5E	95	56	53
96	53	DE	41	55
97	96	77	2D	57
98	39	0D	19	59
99	19	AA	4	5B
100	75	22	F1	5C
101	BD	DF	DB	5E
102	2D	7A	C7	60
103	2A	3A	B3	62
104	44	93	9E	64
105	7C	76	89	66
106	4D	EC	74	68
107	47	88	60	6A
108	0F	22	4C	6C
109	9A	BE	37	6E
110	E5	50	23	70
111	B0	F2	0E	72
112	CB	85	FA	73
113	43	44	E6	75
114	A6	BC	D1	77
115	A7	55	BD	79
116	2B	21	A9	7B
117	ED	65	95	7D
118	E2	D8	81	7F
119	94	4F	6E	81
120	36	CD	5A	83
121	FE	5E	47	85
122	7	B0	33	87
123	42	91	1D	89
124	48	7	0A	8B
125	7	C8	F4	8C
126	77	36	E1	8E
127	D4	D3	CC	90
128	0F	6E	B8	92
129	98	6	A4	94
130	8E	71	8F	96
131	1B	40	7B	98
132	60	D4	66	9A
133	29	6D	52	9C
134	B8	7	3E	9E
135	1E	CE	28	A0
136	6	3F	15	A2
137	80	D7	0	A4
138	64	77	EC	A5
139	60	31	D7	A7

	1st	2nd	3rd	4th
140	BC	A3	C3	A9
141	CC	15	B0	AB
142	FB	5B	9E	AD
143	E4	D4	8A	AF
144	93	90	75	B1
145	FC	7	62	B3
146	94	78	4E	B5
147	DE	C9	3B	B7
148	1E	3A	28	B9
149	D8	B3	14	BB
150	79	4B	0	BD
151	57	B9	EC	BE
152	42	33	D9	C0
153	43	8F	C6	C2
154	E5	2A	B2	C4
155	2F	AD	9D	C6
156	9A	25	8A	C8
157	F4	EC	74	CA
158	67	AB	5F	CC
159	AD	41	4B	CE
160	61	B9	37	D0
161	1B	79	22	D2
162	A0	EF	0E	D4
163	8E	A6	F9	D5
164	32	1E	E6	D7
165	CC	E3	D0	D9
166	55	3E	BB	DB
167	AC	8C	AA	DD
168	B6	83	94	DF
169	0B	1D	80	E1
170	B7	B7	6B	E3
171	B2	4C	57	E5
172	1C	E9	42	E7
173	F3	87	2E	E9
174	76	20	1A	EB
175	2F	BB	5	ED
176	18	5C	F1	EE
177	C0	ED	DC	F0
178	DD	81	C8	F2
179	C4	E7	B4	F4
180	26	74	A1	F6
181	1D	F1	8E	F8
182	91	86	7A	FA
183	BC	F3	66	FC
184	8A	8D	52	FE
185	89	AF	3F	0
186	FB	DC	2D	2

	1st	2nd	3rd	4th
187	CF	76	18	4
188	98	42	4	6
189	FB	F9	EE	7
190	2E	BD	D9	9
191	75	2E	C6	0B
192	E9	89	B2	0D
193	D5	62	9D	0F
194	7B	D0	89	11
195	79	98	74	13
196	19	7	61	15
197	BC	8E	4C	17
198	4	1B	39	19
199	19	27	24	1B
200	CF	95	0E	1D
201	20	35	FA	1E
202	DD	A7	E6	20
203	EC	40	D2	22
204	AD	DB	BD	24
205	E2	57	AA	26
206	3F	9	95	28
207	95	82	81	2A
208	32	F7	6D	2C
209	5E	E3	57	2E
210	F8	E4	42	30
211	EE	E8	2F	32
212	B7	32	19	34
213	4	17	6	36
214	BC	DA	F1	37
215	E9	E2	DF	39
216	40	CE	C9	3B
217	6D	90	BB	3D
218	CF	21	A9	3F
219	BB	74	99	41
220	9	52	80	43
221	F5	82	66	45
222	B3	25	52	47
223	15	88	3E	49
224	1D	AD	2C	4B
225	B6	AE	15	4D
226	9B	28	2	4F
227	22	ED	F2	50
228	BA	C8	DC	52
229	5E	CF	C5	54
230	26	31	AF	56
231	90	CF	9A	58
232	6	17	86	5A
233	9E	AF	71	5C

	1st	2nd	3rd	4th
234	3E	4D	5D	5E
235	A4	E7	48	60
236	C5	22	34	62
237	21	62	20	64
238	9F	9E	0B	66
239	2C	50	F7	67
240	67	E7	E2	69
241	3	86	CE	6B
242	59	20	BA	6D
243	26	B7	A5	6F
244	AC	28	92	71
245	FF	EA	7C	73
246	66	62	69	75
247	27	AC	56	77
248	E0	46	42	79
249	F9	BD	2F	7B
250	A5	E5	20	7D
251	D6	E3	6	7F
252	F8	22	F4	80
253	46	98	E0	82
254	5B	2E	CE	84
255	75	77	B9	86
256	E3	79	A4	88
분포수	192	164	166	158
Shannon	8.00	7.75	7.71	7.66
Min	2.00	1.48	1.51	1.38

(분포수: 160)

	1st	2nd	3rd
1	7C	9F	D7
2	EC	C6	AE
3	4D	4A	9B
4	E9	3A	85
5	7A	C2	70
6	70	B1	5C
7	56	CA	48
8	35	8F	34
9	C0	58	20
10	1	27	0C
11	43	6A	F8
12	C2	F1	E4
13	94	B2	D0
14	A8	2	BD
15	69	BF	A7
16	1A	53	94
17	44	13	7F
18	6F	6A	6B
19	F6	DA	57
20	2E	2D	45
21	6F	58	33
22	1E	52	1D
23	28	1A	0D
24	29	B9	F3
25	3A	7F	DE
26	BD	A6	C8
27	68	3A	B4
28	F6	AD	9F
29	65	57	8B
30	12	E3	76
31	97	7E	62
32	60	3B	4E
33	DA	A1	3A
34	33	3B	27
35	A3	F4	11
36	1D	22	FF
37	38	97	EB
38	21	B5	D7
39	2D	1C	C4
40	78	B2	AF
41	81	4E	9B
42	3E	8	87
43	2F	88	72
44	96	E6	5D
45	F8	82	49
46	D3	7	36

	1st	2nd	3rd
47	CA	C0	20
48	FC	5D	0C
49	ED	F8	F7
50	F7	62	E3
51	8A	FA	CE
52	56	95	BA
53	0A	2F	A6
54	E8	CE	91
55	D3	64	7D
56	25	1	69
57	9B	96	54
58	A8	2D	40
59	D0	AB	2C
60	FF	2A	19
61	77	99	5
62	0F	7	F2
63	A8	78	DE
64	A3	36	C9
65	0E	D1	B4
66	8E	6B	A0
67	6A	4	8C
68	D5	9F	77
69	EE	3A	63
70	5E	CE	4E
71	29	66	3A
72	7	5	26
73	86	9A	11
74	6C	38	FD
75	11	D2	E8
76	2F	3D	DA
77	0B	67	C0
78	DF	D2	AB
79	5E	6F	97
80	3E	7	83
81	FC	9F	6E
82	1A	BA	59
83	E4	D9	45
84	59	4E	32
85	C7	BA	1E
86	B5	31	0B
87	E1	A7	F7
88	C2	21	E4
89	A8	DC	CE
90	42	4F	BB
91	60	E4	A7
92	C7	78	93
93	8A	44	7D

	1st	2nd	3rd
94	39	AC	6A
95	5E	95	56
96	53	DE	41
97	96	77	2D
98	39	0D	19
99	19	AA	4
100	75	22	F1
101	BD	DF	DB
102	2D	7A	C7
103	2A	3A	B3
104	44	93	9E
105	7C	76	89
106	4D	EC	74
107	47	88	60
108	0F	22	4C
109	9A	BE	37
110	E5	50	23
111	B0	F2	0E
112	CB	85	FA
113	43	44	E6
114	A6	BC	D1
115	A7	55	BD
116	2B	21	A9
117	ED	65	95
118	E2	D8	81
119	94	4F	6E
120	36	CD	5A
121	FE	5E	47
122	7	B0	33
123	42	91	1D
124	48	7	0A
125	7	C8	F4
126	77	36	E1
127	D4	D3	CC
128	0F	6E	B8
129	98	6	A4
130	8E	71	8F
131	1B	40	7B
132	60	D4	66
133	29	6D	52
134	B8	7	3E
135	1E	CE	28
136	6	3F	15
137	80	D7	0
138	64	77	EC
139	60	31	D7
140	BC	A3	C3

	1st	2nd	3rd
141	CC	15	B0
142	FB	5B	9E
143	E4	D4	8A
144	93	90	75
145	FC	7	62
146	94	78	4E
147	DE	C9	3B
148	1E	3A	28
149	D8	B3	14
150	79	4B	0
151	57	B9	EC
152	42	33	D9
153	43	8F	C6
154	E5	2A	B2
155	2F	AD	9D
156	9A	25	8A
157	F4	EC	74
158	67	AB	5F
159	AD	41	4B
160	61	B9	37
161	1B	79	22
162	A0	EF	0E
163	8E	A6	F9
164	32	1E	E6
165	CC	E3	D0
166	55	3E	BB
167	AC	8C	AA
168	B6	83	94
169	0B	1D	80
170	B7	B7	6B
171	B2	4C	57
172	1C	E9	42
173	F3	87	2E
174	76	20	1A
175	2F	BB	5
176	18	5C	F1
177	C0	ED	DC
178	DD	81	C8
179	C4	E7	B4
180	26	74	A1
181	1D	F1	8E
182	91	86	7A
183	BC	F3	66
184	8A	8D	52
185	89	AF	3F
186	FB	DC	2D
187	CF	76	18

	1st	2nd	3rd
188	98	42	4
189	FB	F9	EE
190	2E	BD	D9
191	75	2E	C6
192	E9	89	B2
193	D5	62	9D
194	7B	D0	89
195	79	98	74
196	19	7	61
197	BC	8E	4C
198	4	1B	39
199	19	27	24
200	CF	95	0E
201	20	35	FA
202	DD	A7	E6
203	EC	40	D2
204	AD	DB	BD
205	E2	57	AA
206	3F	9	95
207	95	82	81
208	32	F7	6D
209	5E	E3	57
210	F8	E4	42
211	EE	E8	2F
212	B7	32	19
213	4	17	6
214	BC	DA	F1
215	E9	E2	DF
216	40	CE	C9
217	6D	90	BB
218	CF	21	A9
219	BB	74	99
220	9	52	80
221	F5	82	66
222	B3	25	52
223	15	88	3E
224	1D	AD	2C
225	B6	AE	15
226	9B	28	2
227	22	ED	F2
228	BA	C8	DC
229	5E	CF	C5
230	26	31	AF
231	90	CF	9A
232	6	17	86
233	9E	AF	71
234	3E	4D	5D

	1st	2nd	3rd
235	A4	E7	48
236	C5	22	34
237	21	62	20
238	9F	9E	0B
239	2C	50	F7
240	67	E7	E2
241	3	86	CE
242	59	20	BA
243	26	B7	A5
244	AC	28	92
245	FF	EA	7C
246	66	62	69
247	27	AC	56
248	E0	46	42
249	F9	BD	2F
250	A5	E5	20
251	D6	E3	6
252	F8	22	F4
253	46	98	E0
254	5B	2E	CE
255	75	77	B9
256	E3	79	A4
분포수	192	164	166
Shannon	8.00	7.75	7.71
Min	2.00	1.48	1.51

□ 바이트 필터링에 대한 상관관계 (분포수: 160)

	1st	2nd	3rd
1st	-	0.0192	0.0541
2nd	0.0192	-	0.0291
3rd	0.0541	0.0291	-

□ 상관관계를 반영한 엔트로피

		1st	2nd	3rd
상관관계 반영 전	Shannon	8.00	7.75	7.71
	Min	2.00	1.48	1.51
상관관계 반영 후	Shannon	7.804	7.627	7.499
	Min	1.951	1.453	1.466

III.2 엔트로피 측정 알고리즘 사용 사례

- 운영체제: Windows 7 (64비트)
- 수집함수명: GetTickCount
- 잡음원 샘플 크기: 32비트

소프트웨어 환경에서의 잡음원은 데이터의 수집 간격에 의존적이고, 통계적 테스트 및 엔트로피 검증에 필요한 충분한 크기의 데이터를 수집하기 어려울 수 있다. 이러한 특성을 고려하여 다음과 같은 수집 옵션을 적용하여 GetTickCount를 수집한다.

- 수집 시간 : 200ms(millisecond) 내에서 랜덤한 시간 간격으로 수집
- 수집 샘플 개수 : 2,000개, 20,000개

6.1절에서의 알고리즘을 사용하여 수집한 GetTickCount의 최소 엔트로피를 측정한다. 이때 수집한 GetTickCount 샘플의 크기가 8-비트보다 크므로, 가장 많이 변하는 비트를 선정하여 8-비트 이하의 샘플로 변환해야한다(4.3.2절). GetTickCount는 운영체제가 시작된 이후 경과한 시간이기 때문에 하위 비트일수록 가장 많이 변한다. 따라서 빅 엔디언(Big-endian) 방식으로 GetTickCount의 하위 8-비트를 GetTickCount의 샘플로 변환하여 6.1절에서의 각 알고리즘의 입력으로 사용한다.
6-비트 크기 이하의 샘플에 대해서만 수용할 수 있는 마르코브 추정을 제외한 6.1절에서의 최빈값 추정, 충돌 추정, 압축 추정을 사용하여 최소 엔트로피를 측정한다. 3가지의 최소 엔트로피 중에서 최소인 엔트로피를 GetTickCount의 최소 엔트로피로 추정한다.

- 수집 샘플 개수가 2,000개인 경우

	최빈값 추정	충돌 추정	압축 추정
입력	GetTickCount를 8-비트 크기로 변환한 샘플로 이루어진 데이터세트, 데이터세트의 크기 $L=2{,}000$, 출력 가능한 샘플 값의 총 개수 $k=256$		
출력 (샘플 당 최소 엔트로피 H_{∞})	6.325	4.581	4.698
GetTickCount의 최소 엔트로피	4.581 = min(6.325, 4.581, 4.698)		

- 수집 샘플 개수가 20,000개인 경우

	최빈값 추정	충돌 추정	압축 추정
입력	GetTickCount를 8-비트 크기로 변환한 샘플로 이루어진 데이터세트, 데이터세트의 크기 $L=20{,}000$, 출력 가능한 샘플 값의 총 개수 $k=256$		
출력 (샘플 당 최소 엔트로피 H_∞)	7.340	8.000	6.841
GetTickCount의 최소 엔트로피	6.841 = min(7.340, 8.000, 6.841)		

6.2절에서의 알고리즘(엔트로피 테스트, 상호정보량 기반의 엔트로피 추정 방법)을 사용하여 GetTickCount의 Shannon 엔트로피를 측정한다. 이때 위의 최소 엔트로피 알고리즘의 입력으로 사용한 데이터세트를 비트열로 변환하여 6.2절의 각 알고리즘의 입력으로 사용한다(4.3.1절). 각 알고리즘의 또 다른 입력인 워드 길이 T를 8로 설정하고 비트열 길이에 따라 변수 Q와 변수 K를 설정하여 8-비트 워드의 Shannon 엔트로피 H를 추정한다.

- 비트열의 길이가 16,000인 경우(수집한 8-비트 샘플 개수가 2,000개인 경우)

	엔트로피 테스트	상호정보량 기반의 엔트로피 추정 방법
입력	비트열 $b_1, \ldots, b_N$ ($N=16{,}000(=(Q+K)T)$), 워드 길이 $T=8$, 변수 $Q=20$, 변수 $K=1{,}980$	
출력 ($T(=8)$-비트 워드의 Shannon 엔트로피 H)	2.104	0.644

- 비트열의 길이가 160,000인 경우(수집한 8-비트 샘플 개수가 20,000개인 경우)

	엔트로피 테스트	상호정보량 기반의 엔트로피 추정 방법
입력	비트열 $b_1, \ldots, b_N$ ($N=160{,}000(=(Q+K)T)$), 워드 길이 $T=8$, 변수 $Q=200$, 변수 $K=19{,}800$	
출력 ($T(=8)$-비트 워드의 Shannon 엔트로피 H)	7.899	2.918