通信情報保護學會誌
第1卷, 第3號, 1991. 12

Entropy에 의한 Randomness 검정법†

최봉대* · 신양우** · 이경현***

1. 서 론

본 논문에서는 난수발생기의 source가 BMS_p이거나 M-memory를 갖는 마르코프연쇄로 모형화되었을 경우에 비트당 entropy와 관련이 있는 새로운 통계적 검정법을 제안한다. 기존에 알려진 여러가지 검정법은 0 혹은 1의 분포의 편향성, 연속된 비트간의 상관성 중의 한 종류만을 검정할 수 있는 반면에, 이 새로운 검정법은 비트간의 독립성과 안정성을 동시에 검정할 수 있을 뿐만 아니라 암호학적으로 중요한 측도인 비트당 entropy를 측정하여 암호학적인 약점을 검정할 수 있다. 제2절에서는 entropy의 개념을 도입하고 entropy를 불확실성의 정도 또는 한 실험에 의하여 얻어진 정보량으로 해석할 수 있음을 설명하고, 앞으로 사용될 entropy의 성질을 기술한다. 제3절에서는 먼저 M-memory을 갖는 마르코프연쇄를 정의하고, M=1인 경우, 즉, 보통 의미의 마르코프연쇄의 entropy를 정의한다. 다음에 비밀키의 통계적 결점을 바탕으로 하여 키를 찾는 적의 최적 전략(optimal strategy) 문제의 분석이 제시되고, 이 최적 전략이 비트당 entropy와 밀접한 관계가 있음을 보인다. 제4절에서는 비트당 entropy에 관련이 있는 새로운 통계량을 도입하고 이 통계량이 이진수열을 생성하는 난수발생기의 source가 $i.i.d.$ symmetric인 경우, BMS_p인 경우, M-memory를 갖는 마르코프연쇄인 각각 경우의 특성을 조사하고 각 경우에 새로운 통계량의 평균과 분산을 구한다. 새로운 통계량은 중심극한 정리에 의하여 근사적으로 정규분포를 따르므로 위의 평균과 분산을 이용하여 통계적 검정을 시행할 수 있다. 끝으로 5절에서는 여러 난수 발생기에서 생성된 이진 난수열에 대하여 entropy검정을 시행한다.

2. Entropy

Entropy의 개념은 여러 종류의 정보(informations)의 전송을 위한 이론적 모형을 만드는 단계에서 C. Shannon (1984)에 의하여 도입되었다.

$\{A_1, A_2, \cdots, A_n\}$이 확률공간 (Ω, F, P)의 분할(partition)이란 $\{A_i\}$이 서로소이고

† 본 논문은 과학 기술처의 출연금에 의하여 이루어졌습니다.
* 한국과학기술원 ** 창원대학교 *** 한국전자통신연구소

$\bigcup_{i=1}^{n} A_i = \Omega$일때를 말한다. 즉 한 실험에서 $\{A_i\}$들 중에 단 한사건만 그리고 반드시 일어나는 것을 의미한다. 예를 들면 주사위를 던지는 실험에서 $A_i=\{i\}$라 하면 $\{A_i \mid 1\leq i\leq 6\}$은 분할이 된다.

$\{A_i \mid 1\leqslant i\leqslant n\}$가 분할이고 $P(A_i)=p_i$라 두면 $p_i>0$, $\sum_{i=1}^{n} p_i=1$이다. 이때

$$A=\begin{pmatrix} A_1 & A_2 & \cdots & A_n \\ p_1 & p_2 & \cdots & p_n \end{pmatrix}$$

을 유한 체계(finite scheme)이라고 한다. 예를 들면 공정한 주사위를 던지는 실험인 경우에 다음과 같은 유한체계를 갖는다.

$$\begin{pmatrix} A_1 & A_2 & A_3 & A_4 & A_5 & A_6 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{pmatrix}.$$

모든 유한 체계는 다음과 같은 의미에서 불확실성(uncertainty)의 상태를 나타낸다. 한 실험의 결과는 사건 A_1, A_2, …, A_n 중의 어느 하나로 나타나고 실험하기 이전에는 각 사건이 일어날 확률만을 알고 있기 때문이다. 불확실성의 정도는 유한 체계에 따라서 다르다.

예를 들면 다음 두개의 유한 체계에서

$$\begin{pmatrix} A_1 & A_2 \\ 0.5 & 0.5 \end{pmatrix},\ \begin{pmatrix} A_1 & A_2 \\ 0.99 & 0.01 \end{pmatrix},$$

첫번째의 체계는 두번째 보다 훨씬 더 많은 불확실성을 나타낸다 ; 두번째 체계에서는 실험의 결과는 거의 확실시("almost surely") A_1이 되지만, 첫번째 체계에서는 어떠한 예측도 하기가 힘들다. 다음 체계

$$\begin{pmatrix} A_1 & A_2 \\ 0.3 & 0.7 \end{pmatrix}$$

은 위의 두 체계의 중간 정도의 불확실성을 나타낸다.

주어진 유한 체계에 합리적인 방법으로 불확실성의 양을 측정하는 측도를 도입하는 것이 바람직할 것이며, 이것이 다음에 정의하는 entropy의 개념이다.

정의 : $A=\begin{pmatrix} A_1 & A_2 & \cdots & A_n \\ P_1 & P_2 & \cdots & P_n \end{pmatrix}$ 가 유한 체계일때

$$H(p_1, p_2, \cdots, p_n) = -\sum_{k=1}^{n} p_k \log p_k$$

을 A의 entropy라고 한다.

주의 : 지수의 기저는 임의로 택해도 좋으나 일반적으로 2를 취한다. 그리고 $p_k=0$인 경우 $p_k \log p_k=0$으로 정의한다.

불확실성의 합리적인 측도가 가져야 하는 성질들을 위에서 정의한 entropy $H(p_1, p_2, \cdots, p_n)$가 갖고 있음을 직관과 이론적으로 설명을 하고저 한다. 먼저 우리는 다음 사실을 쉽게 알 수 있다.

(a) $H(p_1, p_2, \cdots, p_n)=0$일 필요충분조건은 p_1, p_2, …, p_n 중에 어느 하나가 1이고 나머지 모두는 0이다.

위의 사실을 실험의 결과가 시행이전에 확실하게 예측될 수 있어서 그의 결과에는 불확실성이 전혀 없는 경우를 설명하는 것이다.

(b) $\phi(x)=\begin{cases} 0, & x=0 \\ x\log x, & x\neq 0 \end{cases}$

$\phi(x)$는 convex 함수이다. 즉, $x, y\in[0, \infty]$, $\alpha+\beta=1$, $\alpha, \beta>0$이면

$\phi(\alpha x+\beta y)<\alpha\phi(x)+\beta\phi(y)$이다.

수학적 귀납법에 의하여, $x_i\in[0, \infty]$, $\alpha_i\geqslant 0$, $\sum_{i=1}^{n}\alpha_i=1$이면

$$\phi\left(\sum_{i=1}^{n}\alpha_i x_i\right) \leq \sum_{i=1}^{n}\alpha_i\phi(x_i)$$

이다.

(c) $H(p_1, p_2, \cdots, p_n)\leqslant \log n$

증명 :

$\phi(x)=x\log x$라 하고 윗식에서 $x_k=p_k, \alpha_k=\frac{1}{n}$라 두면,

$$\phi\left(\frac{1}{n}\right)=\frac{1}{n}\log\frac{1}{n}\leq\frac{1}{n}\sum_{k=1}^{n}p_k\log p_k=-\frac{1}{n}H(p_1,p_2,\cdots,p_n)$$

을 얻는다. 따라서

$$H(p_1,p_2,\cdots,p_n)\leq\log n=H(\frac{1}{n},\frac{1}{n},\cdots,\frac{1}{n})$$

이다.

위의 사실 (c)는 Ω의 모든 n개의 사건으로 이루어진 체계 가운데 가장 큰 entropy가 모든 사건이 $\frac{1}{n}$의 확률을 갖는 체계임을 보여준다. 이것은 entropy의 직관적인 해석과 일치한다.

(d) 두개의 유한 체계

$$A=\begin{pmatrix}A_1, & A_2, & \cdots & A_n\\ p_1 & p_2 & \cdots & p_n\end{pmatrix},$$

$$B=\begin{pmatrix}B_1 & B_2 & \cdots & B_m\\ q_1 & q_2 & \cdots & q_m\end{pmatrix},$$

이 주어졌을때, $\{A_k\cap B_l \mid 1<k<n,\ 1<l<m\}$과 $\pi_{kl}=P(A_k\cap B_l)$은 유한 체계를 이루게 되고, 이것을 체계 A와 B의 곱(product)라고 부르고 AB로 적는다.

만약 체계 A와 B가 독립 즉, $\pi_{kl}=P(A_k\cap B_l)=P(A_k)P(B_l)=p_kq_l$이면 $H(AB)=H(A)+H(B)$이다. 단, $H(A)$ $H(B)$, $H(AB)$는 각 체계 A, B, AB의 entropy이다.

왜냐하면,

$$-H(AB)=\sum_k\sum_l\pi_{kl}\log\pi_{kl}=\sum_k\sum_l p_kq_l(\log p_k+\log q_l)$$

$$=\sum_k p_k\log p_k+\sum_l q_l\log q_l=-H(A)-H(B)$$

가 성립하기 때문이다.

(e) 체계 A와 B가 독립이 아닌 종속인 경우를 생각하여 보자.

체계 A의 한 사건 A_k가 일어났다는 가정하에 체계 B의 원소 B_l의 조건부 확률을 q_{kl}로 두면, 즉,

$$q_{kl}=P(B_l|A_k)=\frac{\pi_{kl}}{p_k},$$

A_k상에 유도된 체계 $\{B_l\cap A_k \mid 1<l<m\}$의 조건부 entropy는 $H_k(B)=-\sum_{l=1}^{m}q_{kl}\log q_{kl}$이다. 그리고 난후 k에 관하여 평균을 취하면 $H(B\mid A)=-\sum_{k=1}^{n}p_k\sum_{l=1}^{m}q_{kl}\log q_{kl}$을 얻는다. 이것을 주어진 체계 A에 관하여 체계 B의 조건부 entropy라고 한다. 이것은 다시 요약하면 다음과 같다.

정의 : 체계 $A=\{A_1, A_2, \cdots, A_n\}$,
$B=\{B_1, B_2, \cdots, B_m\}$에 대하여

$$H(B|A)=-\sum_{k=1}^{m}P(A_k)\sum_{l=1}^{m}\frac{P(A_k\cap B_l)}{P(A_k)}\log\frac{P(A_k\cap B_l)}{P(A_k)}$$

$$=-\sum_{k,l}P(A_k\cap B_l)\log\frac{P(A_k\cap B_l)}{P(A_k)}\geq 0$$

을 주어진 체계 A에 관한 체계 B의 조건부 entropy라고 한다. 조건부 entropy에 관하여 다음 사실이 성립한다.

(e. 1) $A=\{\phi, \Omega\} \Rightarrow H(B \mid A)=H(B)$

(e. 2) $H(AB)=H(A)+H(B \mid A)$

(e. 3) $H(B \mid A)<H(B)$

주의 : (e. 1)의 사실은 A는 trivial 한 실험의 결과를 나타내므로 A로 부터는 아무런 정보도 얻을 수 없다는 것을 나타낸다. (e. 3)의 사실은 A의 결과를 아는 것은 평균적으로 B의 불확실성을 감소시킨다는 것이다.

증명 : (e. 1)은 조건부 정의로부터 쉽게 얻을 수 있고 (e. 2)는 (e. 1)에 주어진 독립인 경우와 마찬가지의 방법을 따라가면 얻을 수 있다.

(e. 3)의 증명 ; $P(A_k)=p_k$, $P(B_l)=q_l$, $P(A_k \cap B_l)=\pi_{kl}$, $q_{kl}=\dfrac{\pi_{kl}}{p_k}$ 로 두면

$$H(B|A) = -\sum_k p_k \sum_l q_{kl} \log q_{kl}$$

이 된다. $f(x)=x\log x$는 convex 함수이고, 부등식 $\sum_k \lambda_k f(x_k) > f(\sum_k \lambda_k x_k)$을 만족하므로 $\lambda_k=p_k$, $x_k=q_{kl}$로 두면 임의의 l에 대하여

$$\sum_k p_k q_{kl} \log q_{kl} \leq (\sum_k p_k q_{kl}) \log(\sum_k p_k q_{kl}) = q_l \log q_l$$

을 얻는다. 양변에 l에 관하여 합하면

$$-H(B \mid A) > -H(B)$$

을 얻을 수 있다.

(f) 불확실성은 실험에 의하여 얻어지는 정보와 같다는 개념을 설명하고자 한다.

실험의 결과가 주어진 유한 체계 A에 의하여 기술되어지는 경우에, 그 실험을 시행한 후에는 실제적으로 어느 사건이 일어났는가 하는 정보(information)을 얻게 되고, 유한 체계의 불확실성이 완전히 제거되어진다. 따라서 어떠한 실험을 시행하므로써 얻어지는 정보는 실험이전에 존재했던 불확실성을 제거하는 것과 같이 볼 수 있다. 불확실성이 크면 클수록 그것을 제거하므로써 얻어지는 정보의 양이 많다는 것으로 간주한다. Entropy의 성질로부터 정보의 양이 불확실성의 측도인 entropy에 비례하는 것으로 택하는 것이 편리함을 알 수 있다. 예를 들면 유한 체계 A와 B, 그리고 곱 AB를 생각하여 보자. AB가 일어난다 함은 A와 B가 독립이면, $H(AB)=H(A)+H(B)$가 되고 따라서 AB의 정보량은 A와 B에 의하여 얻어진 정보량의 합이 된다는 것은 자연스러운 것이다. 비례상수를 1로 취하므로써, 한 유한 체계의 실현으로 얻어지는 정보의 양은 그 유한 체계의 entropy로 정의한다. 이러한 약정으로 인하여 entropy의 개념이 정보이론에 매우 유용하게 이용되고 있다. 따라서 얻어진 정보의 양의 관점에서의 (e. 2)식의 해석은 A와 B의 실현에 의하여 얻어진 정보의 양은 A의 실현에 의하여 얻어진 정보의 양에 A의 실현이 이루어진 후에 B의 실현에 의하여 부가적인 정보의 양의 수학적 기대치를 합한 것과 같다. (e. 3)의 해석은 B의 실현에 의하여 얻어진 정보의 양은 그 이전에 다른 A의 실현이 이루어졌다면 감소한다.

(g) 우리가 증명한 entropy의 기본적인 성질을 열거하면 다음과 같다.

(g. 1) $H(p_1, p_2, \cdots, p_n)$는 $p_k=\dfrac{1}{n}(1 \leqslant k \leqslant n)$인 경우에 최대치를 갖는다.

(g. 2) $H(AB)=H(A)+B(B \mid A)$

(g. 3) $H(p_1, p_2, \cdots, p_n, 0)=H(p_1, p_2, \cdots, p_n)$.

정리 2. 1 [5] : $\Delta_n=\{(p_1, p_2, \cdots, p_n) \in R^n \mid p_k \geq 0, \sum_{k=1}^{n} p_k=1\}$라 두면

함수 $H : \bigcup_{n=1}^{\infty} \Delta_n \to R$가 위의 (g. 1)-(g. 3)을 만족하고 또한 $H|_{\Delta_n} : \Delta_n \to R$가 연속이면, 적당한 $\lambda>0$가 존재하여

$$H(p_1, p_2, \cdots, p_n) = -\lambda \sum_{k=1}^{n} p_k \log p_k$$

로 나타난다.

위의 정리는 entropy의 정의가 실제적 의미의

불확실성의 측도(또는 정보의 양)으로 해석하였을 때 만족하여야 할 성질을 갖는 것으로는 유일하다는 것을 보인다.

3. 마르코프연쇄와 그의 entropy

일반적으로 난수 발생기 (혹은 source)에서 생성된 난수열이 가지고 있는 약점은 빈도의 편향성(bias)과 각 항 사이의 상호관련성이다. 2진 난수 발생기 S가 확률변수열 U_1, U_2, …에 의하여 난수를 생성한다고 하자. n번째 항 U_n이 바로 앞의 M 비트에만 의존한다면, 즉, U_1, U_2, …, U_{n-1}이 주어졌을 때 U_n의 분포가 U_{n-1}, … U_{n-M}에만 의존할때 $\{U_n\}$을 M-memory를 갖는 마르코프연쇄라고 부른다. 위의 사실을 수학적으로 나타내면,

$$P_{U_n|U_{n-1},\cdots,U_1}(u_n|u_{n-1},\cdots,u_1) = P(U_n = u_n|(U_{n-1},\cdots,U_1) = (u_{n-1},\cdots u_1))$$

을 조건부 확률이라 하면, 모든 $n>M$, $(u_1, \cdots, u_n) \in B^n$에 대하여

$$\begin{aligned} &P_{U_n|U_{n-1},\cdots,U_1}(u_n|u_{n-1},\cdots,u_1) \\ &\quad = P_{U_n|U_{n-1},\cdots,U_{n-M}}(u_n|u_{n-1},\cdots,u_{n-M}) \end{aligned} \tag{3. 1}$$

여기서 $B=\{0, 1\}$이다. (3. 1)을 만족하는 최소의 정수 M을 S의 memory라고 한다. M이 0인 경우, 즉, U_n이 U_{n-1}, …, U_1과 독립일때 S는 memoryless source라고 한다. 특히 memoryless source S에서 확률 p로 1값을 갖는 비트를, 확률 $1-p$로 0값을 갖는 비트를 생성할때 source S를 BMS_p (Baised Memoryless Source)라고 한다. $M=1$인 경우 우리는 보통 의미의 마르코프연쇄를 얻는다. $\Sigma_n=[U_n-1, \cdots, U_{n-M}]$은 회수 n이전의 M개의 비트의 상태를 나타낸다. $\Sigma_1=[U_0, \cdots, U_{-M+1}]$은 초기상태를 나타내고 이때 U_0, U_{-1}, …, U_{-M+1}은 가상의 확률변수들이다. 특히 확률변수열 $\{U_n\}$이 (3. 1)을 만족하고 또한 모든 $n>M$, $u \in B$, $\alpha \in B^n$에 대하여 $P_{U_{n+1}|\Sigma_n}(u|\sigma)=P_{U_{n+1}|\Sigma_n}(u|\sigma)$을 만족할 때 source S를 stationary 하다고 한다. memory M을 갖는 stationary source S는 초기상태의 분포 P_{Σ_1}과 전이확률 $P_{U_1|\Sigma_1}(u_1|\sigma)$에 의해서 완전히 결정되어 진다. 만일 random vector Σ_n의 한 state $(u_1, u_2, \cdots, u_M)$ $(\in B^M)$이 정수 j의 2진수일 때 $(u_1, u_2, \cdots, u_M)$을 j와 동일시하면 $\{\Sigma_n\}$은 상태공간을 $[0, 2^M-1]$로 갖는 보통의미의 Markov chain이 된다. 그러므로 M-memory를 갖는 stationary source S는 상태공간 $[0, 2^M-1]$을 갖는 homogeneous 1-memory Markov chain으로 모형화 될 수 있다. 만일 $\{\Sigma_n\}$이 ergodic Markov chain이면 $\{\Sigma_n\}$의 극한분포

$$\lim_{n\to\infty} P(\sum\nolimits_n = j) = P_j,\ 0 \le j \le 2^M - 1$$

가 존재하고 다음의 관계식을 만족한다.

$$\sum_{j=0}^{2^M-1} P_j = 1,$$

$$P_j = \sum_{k=0}^{2^M-1} P_{\Sigma_2|\Sigma_1}(j|k)P_k,\ \ 0 \le j \le 2^M - 1.$$

보통 의미의 마르코프연쇄에 관한 entropy를 도입하기로 하자. 전이확률 $(p_{ik})(i, k=1, 2, \cdots, n)$와 상태공간 $\{1, 2, \cdots, n\}$와 상태공간 $\{1, 2, \cdots, n\}$을 갖는 ergodic이고 stationary인 마르코프연쇄에서 극한분포를 P_j라 하면

$$P_j = \sum_{k=1}^{n} P_k p_{kj}$$

를 만족한다. 만약 마르코프연쇄가 현재 state i에

있으면, 다음에 다른 state로의 전이확률이 (p_{i1}, p_{i2}, …, p_{in})이 되고, 이것의 entropy

$$H_i = -\sum_{k=1}^{n} p_{ik} \log p_{ik}$$

는 i에만 의존하고, H_i는 마르코프연쇄가 state i에서 출발하여 한 단계 앞으로 나아갔을 때 얻어진 정보량의 측도로 간주 되어질 수 있다. 모든 초기 state i에 관한 H_i의 평균

$$H = \sum_{i=1}^{n} P_i H_i = -\sum_{i=1}^{n}\sum_{k=1}^{n} P_i p_{ik} \log p_{ik}$$

는 마르코프연쇄가 한 단계 앞으로 움직였을 때 얻어진 평균정보량의 측도로 볼 수 있다. 이것을 마르코프연쇄의 entropy라고 부른다.

좋은 암호계는 적의 공격에 대해서 안정되어야 한다. 즉, 좋은 암호계란 권한이 없는 암호 해독자가 암호방식을 해독하기 위하여 key를 찾을 때 어떠한 통계적인 공격방식도 key space에 있는 모든 key를 하나하나 적용하는 방식(exhausitive key search) 보다 본질적으로 더 빠른 방식이 없도록 고안된 것이다. 만약 비밀 key가 truly random이 아니면, 즉, 모든 key가 같은 확률을 가지지 않으면, 적의 최적의 전략은 확률이 가장 높은 키값으로부터 시작하여 확률이 낮은 키 값으로 순서대로 적용할 것이다. Z를 비트의 길이가 n인 비밀키라 하고 $P_Z(z)$를 Z의 값이 z일 확률이라고 하자. 이때 Z_1, Z_2, …, Z_{2^n}은 $P_Z(z_i)$의 값이 큰 순서대로 key space의 원소들에 번호를 붙여 놓은것이라 하자. 즉, $P_Z(z_1) > P_Z(z_2) > \cdots > P_Z(z_{2^n})$.

주어진 source S로부터 n-bit key Z를 만들었을 때 최소한 δ의 확률로 key를 찾는데, 성공하기 위하여 암호해독자가 최적의 전략을 따라서 test 해야할 최소한의 시행회수를 $\mu_s(n, \delta)$라 하자. 즉,

$$\mu_s(n,\delta) = \min\{k : \sum_{i=1}^{k} P_Z(z_i) \geq \delta\}.$$

이때 $\log_2 \mu_s(n, \frac{1}{2})$을 effective key size라 한다. 여기서 $\delta = \frac{1}{2}$을 선택한 것은 임의적인 것이며, 일반적으로 key의 길이 n이 충분히 클때 δ가 0이나 1에 극단적으로 가깝지 않으면 $\log_2 \mu_s(n, \delta)$는 δ에 거의 의존하지 않는다. 만약 S가 truly random source라면, $P_Z(z_i) = \frac{1}{2}$, $i = 1, 2, \cdots, 2^n$이므로 effective key size는 $\log_2 \mu_s(n, \frac{1}{2}) = n - 1$이다.

Effective key size와 entropy의 관련성을 아래의 정리 3. 1 (M-memory를 갖는 source의 경우)와 정리 3. 2(BMS$_p$ source의 경우)에서 알 수 있다.

정리 3. 1 (Shannon(1948)) [5] : M-memory를 갖는 Markov chain인 stationary source S로 부터 key Z가 만들어졌을 때 다음의 식이 성립한다.

$$\lim_{n\to\infty} \frac{\log_2 \mu_S(n,\delta)}{n} = H_S,\ 0 < \delta < 1 \qquad (3.2)$$

여기서 $H_S = -\sum_{j=0} P_j \sum_{k=0} P_{\Sigma_S|\Sigma_I}(k \mid j) \log_2 P_{\Sigma_Z|\Sigma_I}(k \mid j)$는 M-memory을 갖는 Markov chain인 source S의 단위 비트당 entropy이다.

길이 n인 모든 key의 개수는 2^n인 반면 정리 3. 1로부터 $\mu_s(n, \delta)$는 약 2^{nH_S}임을 나타낸다.

다음은 key source S가 BMS$_p$인 경우를 생각하자. 우리는 일반성을 잃지 않고 $0 < p \leq \frac{1}{2}$을 가정할 수 있다. Z가 BMS$_p$로부터 만들어졌다면 Z의 확률분포는 $P_Z(z) = P^{\omega(z)}(1-p)^{n-\omega(z)}$로 주어진다. 여기서 $\omega(z)$는 z의 성분중에 있는 1의 개수이다. 약 $\frac{1}{2}$ 의 확률로 성공하기 위하여 해독자는 $\omega(z) < pn$인 모든 key값 z에 대하여 조사를 하여야 한다. 이 때 effective key size는

$$\log_2 \mu_{BMS_p}\left(n, \frac{1}{2}\right) \approx \log_2 \sum_{i=0}^{pn} \binom{n}{i}. \qquad (3.3)$$

다음의 부등식

$$\frac{1}{\sqrt{8t(n-t)/n}} 2^{nH(\frac{t}{n})} \leq \binom{n}{i} \leq 2^{nH(\frac{t}{n})},\ t \leq \frac{n}{2} \qquad (3.4)$$

에서 $t=np$를 대입한 다음 $\log_2$를 취한후 n으로 나누고 n을 ∞로 보내면 다음을 얻는다. 여기서 $H(x)=-x\log_2 x-(1-x)\log_2(1-x)$는 binary entropy이다.

정리 3. 2 :

$$\lim_{n\to\infty}\frac{\log_2 \mu_{BMS_p}(n,\delta)}{n}=H(p),\ 0<\delta<1 \qquad (3.5)$$

어떤 source (혹은 generator)를 이용하여 암호계의 key를 만들었을때 source의 단위 비트당 entropy는 그 암호계의 안정성과 밀접한 관련이 있다는 것을 위의 결과들로부터 알 수 있다. 그러므로 암호키를 만드는데 사용된 source의 단위 비트당 entropy는 암호계의 안정성을 측정하는 하나의 측도가 될 수 있다.

4. Entropy 관점에서의 Randomness 검정법

이 절에서 우리는 비트당 entropy와 밀접한 관련이 있는 통계량 T를 정의한다. $s^N=s_0s_2\cdots s_{N-1}$을 전체의 길이가 N인 이진 수열이라 하자. 먼저 s^N에서 길이가 L인 $\frac{N}{L}$ 개의 겹치지 않는 block을 만든다. 처음 Q개의 block을 초기화를 위한 block으로 사용하고 나머지 $K=\frac{N}{L}-Q$개의 block은 test를 위해 사용한다.

$b_n(s^N)=[s_{Ln}, s_{Ln+1}, \cdots, s_{Ln+L-1}]$, $0\leqslant n\leqslant Q+K-1$, 을 n번째 block이라 하고 $Q\leqslant n\leqslant Q+K-1$에 대하여

$$A_n(s^N)=\begin{cases}\min\{i|1\le i\le n,\ b_n(s^N)=b_{n-i}(s^N)\} & \text{if } \{\cdots\}\neq\phi\\ n & \text{if } \{\cdots\}=\phi\end{cases}$$

라고 두자. 즉 $A_n(s^N)$는 n번째 block $b_n(s^N)$과 처음으로 일치하게 되는 $b_{n-i}(s^N)$가 존재할 때 i로 정의하며, 존재하지 않는 경우 n으로 정의한다. 이때 통계량 T를 다음과 같이 정의한다.

$$T(s^N)=\frac{1}{K}\sum_{n=Q}^{Q+K-1}\log_2 A_n(s^N). \qquad (4.1)$$

모의실험을 위하여 통계량 T의 모수 L, Q와 K를 다음의 값을 추천한다. $8\leqslant L\leqslant 16$, $Q\geqslant 30\cdot 2^L$ 그리고 K는 클수록 좋다(예를들면 $K=10^4$ 혹은 $K=10^5$). 이러한 Q의 선택은 거의 확률 1로써 random 수열에서 모든 L-bit 형태가 최소한 한번씩 처음부터 Q blocks 내에 나타나도록 한 것이다. 통계량 $T(s^N)$을 계산하는 알고리즘은 PASCAL 형태의 기호로 표시하면 다음과 같다.

```
FOR i : =0 TO 2^L-1 DO Tab [i] : =0 ;
FOR n : =0 TO Q-1 DO Tab [b_n(s^N)] : =n ;
        sum=0.0 ;
FOR n : =Q TO Q+K-1 DO BEGIN
        sum : =sum+log2(n-Tab [b_n(s^N)]) ;
        Tab [b_n(s^N)] : =n ;
END
        T : =sum/K
```

$L\to\infty$일 때 통계량 T와 단위 비트당 entropy와의 관계는 다음의 정리 4. 1(truly ramdom source의 경우), 정리 4. 2 (BMS_p source의 경우), 정리 4. 3 (M-memory를 갖는 source의 경우)로 주어진다.

정리 4. 1 : $R^N=R_1R_2\cdots R_N$이 $i.i.d.$ symmetric binary sequence이라 할 때 다음 식이 성립한다.

$$\lim_{L\to\infty}(E(T(R^N))-L)=-0.832746$$

$$\lim_{L\to\infty} K \cdot \mathrm{Var}(T(R^N)) = 3.423715$$

정리 4. 2 : $U_{\mathrm{BMS}_p}{}^N$가 길이 N인 BMS_p의 output이라 할 때 다음 식이 성립한다.

$$\lim_{L\to\infty} \left(E(T(U^N_{\mathrm{BMS}_p})) - LH(p)\right) = -0.832746$$

정리 4. 3 : U_s^N가 M-memory를 갖는 stationary source S의 output이면

$\lim_{L\to\infty} \dfrac{E(T(U_s^N))}{L} = H_s$가 된다.

이진수열 U^N이 주어졌을 때 통계량 $T(U^N)$의 평균 $E(\mathrm{T}(\mathrm{U}^N))$와 분산 $\mathrm{Var}(T(\mathrm{U}^N))$를 구하자.

---

(1) Source가 BMS_p인 경우

$p=\dfrac{1}{2}$인 경우가 독립이고 일양분포를 갖는 경우에 해당된다. $U_{\mathrm{BMS}_p}{}^N$의 block들인 $b_n(U_{\mathrm{BMS}_p}{}^N)$은 서로 독립이므로

$$\begin{aligned}
&P(A_n(U^N_{\mathrm{BMS}_p}) = i)\\
&= \sum_{b\in B^N} P\Big(b_n(U^N_{\mathrm{BMS}_p}) = b,\ b_{n-1}(U^N_{\mathrm{BMS}_p}) \neq b, \cdots,\\
&\qquad b_{n-i+1}(U^N_{\mathrm{BMS}_p}) \neq b,\ b_{n-i}(U^N_{\mathrm{BMS}_p}) = b\Big)\\
&= \sum_{b\in B^N} \left[P(b_n(U^N_{\mathrm{BMS}_p}) = b)\right]^2 \left[1 - P(b_n(U^N_{\mathrm{BMS}_p}) = b)\right]^{i-1}\\
&= \sum_{k=0}^{L} \binom{L}{k} \left(p^k(1-p)^{L-k}\right)^2 \left(1 - p^k(1-p)^{L-k}\right)^{i-1}
\end{aligned}$$

이다. 그러므로

$$\begin{aligned}
E(T(U^N_{\mathrm{BMS}_p})) &= \frac{1}{K} \sum_{n=Q}^{Q+K-1} \sum_{i=1}^{n} P\left(A_n(U^N_{\mathrm{BMS}_p}) = i\right) \log_2 i\\
&= \frac{1}{K} \sum_{n=Q}^{Q+K-1} \sum_{k=0}^{L} \binom{L}{k} \left(p^k(1-p)^{L-k}\right)^2 \sum_{i=1}^{n} \left(1 - p^k(1-p)^{L-k}\right)^{i-1} \log_2 i
\end{aligned}$$

이다. $Q\to\infty$이면 $\sum_{i=Q}^{\infty}(1-p^k(1-\mathrm{p})^{L-k})^{i-1}\to 0$이므로 $A_n(U_{\mathrm{BMS}_p}{}^N)$은 n에 거의 의존하지 않게 된다. 그러므로 이 때는

$$E\left(T(U^N_{\mathrm{BMS}_p})\right) = \sum_{k=0}^{L} \binom{L}{k} \left(p^k(1-p)^{L-k}\right)^2 \sum_{i=1}^{\infty} \left(1 - p^k(1-p)^{L-k}\right)^{i-1} \log_2 i \qquad (4.2)$$

이다. 또한 $Q\to\infty$일때

$$K \cdot \mathrm{Var}\left(T(U^N_{\mathrm{BMS}_p})\right) = \mathrm{Var}\left(\log_2 A_n(U^N_{\mathrm{BMS}_p})\right)$$
$$= E\left(\log_2 A_n(U^N_{\mathrm{BMS}_p})\right)^2 - \left(E(\log_2 A_n(U^N_{\mathrm{BMS}_p}))\right)^2$$
$$= \sum_{k=0}^{L}\binom{L}{k}(p^k-(1-p)^{L-k})^2\sum_{i=1}^{\infty}(1-p^k(1-p)^{L-k})^{i-1}(\log_2 i)^2$$
$$- \left[E(\log_2 A_n(U^N_{\mathrm{BMS}_p}))\right]^2 \qquad (4.3)$$

이다. 위의 계산에서 $Q\to\infty$를 가정한 것은 실제의 검정에서 Q를 충분히 크게 선택하므로 오차가 아주 작은 것으로 볼 수 있어서 실제로 이용하는 데 문제가 없다. 여러가지 block의 크기 L에 대하여 $p=\frac{1}{2}$인 경우에 (4. 2)와 (4. 3)식에 의하여 얻어진 $T(U_{\mathrm{BMS}}{}^N)$ 의 평균 $E(T(U_{\mathrm{BMS}}{}^N))$과 분산 $K\cdot\mathrm{Var}(T(U_{\mathrm{BMS}}{}^N))$의 값은 〈표 1〉에 주어지고, 난수발생기의 source가 $\mathrm{BMS}^N\frac{1}{2}$이라는 가설검정에 이용된다. 또한 〈표 1〉은 정리 4. 1이 좋은 근사치를 제공함을 보여준다.

〈표 1〉 $T(U_{BMS}{}^N)$의 평균과 분산

L	$E(T(U^N))$	$K\cdot Var(T(U^N))$	L	$E(T(U^N))$	$K\cdot Var(T(U^N))$
1	0.73264948	0.68977	9	8.17642476	3.31120
2	1.53743829	1.33774	10	9.17232431	3.35646
3	2.40160681	1.90133	11	10.17003231	3.38409
4	3.31122472	0.35774	12	11.16876491	3.40065
5	4.25342659	2.70455	13	12.16807031	3.41043
6	5.21770525	2.95403	14	12.16769261	3.41614
7	6.19625065	3.12539	15	14.16748841	3.41943
8	7.18366555	3.23866	16	15.16726881	3.42130

(2) Source가 M-memory Markov Chain인 경우 이진 수열 U_s^N를 M-memory를 갖는 source S에서 출력된 길이 N인 2진 수열이라고 하자. $E(T(U_s^N))$와 $\mathrm{Var}(T(U_s^N))$를 구하기 위해서는 먼저 확률

$$P(A_n(U_s^N)=i)=\sum_{b\in B^n}P(b_n(U_s^N)=b, b_{n-1}(U_s^N)\neq b,\cdots,b_{n-i+1}(U_s^N)\neq b, b_{n-i}(U_s^N)=b)$$

를 구해야한다.

$$P(b_n(U_s^N)=b, b_{n-1}(U_s^N)\neq b,\cdots,b_{n-i+1}(U_s^N)\neq b, b_{n-i}(U_s^N)=b))$$
$$=P(b_n(U_s^N)=b, b_{n-i}(U_s^N)=b)$$

$$-\sum_{k=1}^{i-1} P(b_n(U_s^N)=b, b_{n-k}(U_s^N)=b, b_{n-i}(U_s^N)=b)+\cdots$$

$$+(-1)^k \sum_{\substack{i_1,\cdots,i_k=1 \\ i_1<i_2<\cdots<i_k}}^{i-1} P(b_n(U_s^N)=b, b_{n-i_1}(U_s^N),\cdots,b_{n-i_k}(U_s^N)=b, b_{n-i}(U_s^N)=b)$$

$$+\cdots+(-1)^{i-1}P(b_n(U_s^N)=b, b_{n-1}(U_s^N),\cdots,b_{n-i+1}(U_s^N)=b, b_{n-i}(U_s^N)=b)$$

이므로 $Q\leqq n_1<n_2<\cdots<n_k$일때, $P(b_{n1}(U_s^N)=b,\ b_{n2}(U_s^N)=b,\ \cdots,\ b_{nk}(U_s^N)=b)$를 구하면 된다. 먼저 block의 길이 L이 memory의 크기 M의 배수라고 가정하자. 즉 $L=Mm$. 여기서 m은 양의 정수이다.

U_s^N를 계산하기 위하여 아래와 같이 U_s^N의 n번째 block $b_n(U_s^N)$를 $\{\Sigma_n\}$의 block으로 나타내면 다음과 같이 된다.

$$b_n(U_s^N)=[U_{Ln}, U_{Ln+1},\cdots,U_{Ln+L-1}]$$
$$=\left[\sum_{mnM}, \sum_{(mn+1)M},\cdots,\sum_{(mn+m-1)M}\right].$$

기호의 편의를 위하여 $\Gamma_n=[\Sigma_{mnM},\ \Sigma_{(mn+1)M},\ \cdots,\ \Sigma_{(mn+m-1)M}]$이라고 두자. $b\in B^N$일때 b의 n번째 block $(b_{Ln},\ b_{Ln+1},\ \cdots,\ b_{Ln+L-1})$을 다시 크기 M인 m개의 block으로 나누자. 즉

$$(b_{Ln},\ \cdots,\ b_{Ln+L-1})=(B_{n0},\ \cdots,\ B_{nN-1})$$

여기서 $B_{nk}=(b_{Ln+kM},\ b_{Ln+kM+1},\ \cdots,\ b_{Ln+(k+1)M-1})$이다. m비트 B_{nk}에 대응하는 십진수 (양의 정수)를 i_k라 하면

$$\{b_n(U_s^N)=b\}=\{\Gamma_n=(i_0,i_1,\cdots,i_{m-1})\}$$
$$=\{\sum_{mnM}=i_0, \sum_{(mn+1)M}=i_1,\cdots,\sum_{(mn+m-1)M}=i_{m-1}\}$$

이다. 그러므로

$$P(b_n(U_s^N)=b)=P(\{\Gamma_n=(i_0,i_1,\cdots,i_{m-1})\})$$
$$=P(\sum_{mnM}=i_0,\cdots,\sum_{(mn+m-1)M}=i_{m-1})$$
$$=P(\sum_{mnM}=i_0)P^M(i_1,i_2)\cdots P^M(i_{m-2},i_{m-1})$$

을 얻는다. 여기서 $P^M(i,j)$는 Markov chain $\{\Sigma_n\}$의 M-step transition probability matrix의 (i,j)-성분이다. $\lim_{n\to\infty}P(\Sigma_n=i)=P(i)$를 $\{\Sigma_n\}$의 steady-state probability라 할때, $Q\to\infty$ 이면 $P(\Sigma_{mnM}=i_0)\to P(i_0)$, $n>Q$이다. 그러므로 $Q\to\infty$ 일때

$$P(b_n(U_s^N)=b)=P(i_0)P^M(i_0,i_1)P^M(i_1,i_2)\cdots P^M(i_{m-2},i_{m-1}). \qquad (4.4)$$

이다. 또한 $j>k\geqq Q$일때, 마르코프의 성질에 의해서

$$\begin{aligned} &P(b_j(U_s^N)=b|b_k(U_s^N)=b) \\ &\quad=P(\sum_{mjM}=i_0,\sum_{(mj+1)M}=i_1,\cdots, \\ &\qquad\sum_{(mj+m-1)M}=i_{m-1}|\sum_{(mk+m-1)M}=i_{m-1}) \\ &\quad=P^{M(m(j-k)-1)}(i_{m-1},i_0)P^M(i_0,i_1)\cdots P^M(i_{m-2},i_{m-1}) \end{aligned} \tag{4.5}$$

을 얻는다. (4. 5)식에서 $j-k$, M, m 중에서 어느 하나가 충분히 크면 $M(m(j-k)-1)$도 충분히 커지므로 $P^{M(m(j-k)-1}(j_{m-1},i_0)\approx P(i_0)$가 된다. 일반적으로 $Q\leqq n_1<n_2<\cdots<n_k$일때,

$$\begin{aligned} &P(b_{n_1}(U_s^N)=b,b_{n_2}(U_s^N)=b,\cdots,b_{n_k}(U_s^N)=b) \\ &=P(b_{n_k}(U_s^N)=b|b_{n_{k-1}}(U_s^N)=b)\cdots P(b_{n_2}(U_s^N)=b|b_{n_1}(U_s^N)=b)P(b_{n_1}(U_s^N)=b) \end{aligned} \tag{4.6}$$

이므로 (4. 6)의 오른편은 (4. 4)와 (4. 5) 결과를 이용하여 계산할 수 있다. 또한

$$\begin{aligned} &\sum_{b\in B^n}P(b_{n_1}(U_s^N)=b,b_{n_2}(U_s^N)=b,\cdots b_{n_k}(U_s^N)=b) \\ &=\sum_{i_1=0}^{2^M-1}\sum_{i_2=0}^{2^M-1}\cdots\sum_{i_m=0}^{2^M-1}P(\Gamma_{n_1}=(i_1,i_2,\cdots,i_m),\cdots\Gamma_{n_k}=(i_1,i_2,\cdots i_m)) \end{aligned}$$

이므로 확률 $P(A_n(U_s^N)=i)$를 (4. 4)－(4. 6)으로부터 계산할 수 있다. $Q\to\infty$ 이면 $E(\log_2 A_n(U_s^N))$는 n에 의존하지 않는다. 그러므로

$$E(\log_2 A_n(U_s^N))=\sum_{i=1}^{\infty}P(A_n(U_s^N)=i)\log_2 i$$

이고

$$\begin{aligned} K\cdot\mathrm{Var}(T(U_s^N))&=\mathrm{Var}(\log_2 A_n(U_s^N)) \\ &=\sum_{i=1}^{\infty}P(A_n(U_s^N)=i)(\log_2 i)^2-[E(T(U_s^N))]^2 \end{aligned}$$

이다. $\mathrm{Var}(T(U_s^N))$를 계산할 때 우리는 $A_n(U_s^N)$가 서로 독립이라고 가정하였다.

이진 수열 U_S^N의 source S가 BMS_p인가 M-memory source인가를 entropy 관점에서 검정하기 위하여는 먼저 S의 memory 크기를 추정하고 $T(U_S^N)$의 평균 $E(T(U_S^N))$와 분산 $\mathrm{Var}(T(U_S^N))$를 구한다. 이때의 유의수준 α에 대한 기각역은

$$\left|\frac{T(s^N)-E(T(U^N))}{\sqrt{\mathrm{Var}\,(T(U_S^N))}}\right| \geq z_{\frac{\alpha}{2}} \qquad (4.\ 7)$$

이다. 여기서 $1-\alpha=\int_{-\infty}^{2\alpha}\frac{1}{\sqrt{2\pi}}\ e^{-\frac{x^2}{2}}\ dx$이다.

S의 memory가 0, 즉 독립일때 2진수열 (U_S^N)이 위 기각영역 (4. 7)에 들지 않으면 entropy에 의한 검정을 통과한 것이며, 이때 U_S^N의 source는 BMS_p라는 것을 의미한다. 마찬가지로 S의 memory가 M일때 U_S^N가 위 검정을 통과하면 Source S가 M-memory를 갖는 마르코프 연쇄라는 것을 의미한다.

5. 모의실험 및 결과

이진 난수 발생기 (Random bit generator)에서 생성된 의사이진수열의 randomness를 검정하는 판정법은 완전한 난수 (truly random number)가 가지고 있는 어떤 특성에서의 벗어남(예를들면, 비트간의 독립성 결여, 혹은 0과 1의 빈도의 편향성 등)을 발견하면 생성자를 기각하는 것이다. 이 절에서는 entropy 관점에서의 검정을 시행하여 이진수열의 각항 사이의 독립성과 빈도의 안정성이 동시에 검정됨을 모의실험을 통하여 보이고자 한다.

(1) 난수 발생기

모의 실험을 위하여 우리가 사용한 난수 발생기는 선형합동법 (Linear congruential generator), 선형귀환 쉬프트 레지스터(Linear feedback shift register)와 비선형 알고리즘으로서 승산 시스템(Multiplication system), J-K플립플롭 시스템 (J-K Flip Flop system), Geffe 시스템 그리고 상호대칭 시스템이다.

(i) 선형 합동법(Linear congruential generator)

Linear congruential generator는 다음과 같이 정의된다.

$x_i=(ax_{i-1}+c)\ \mathrm{mod}\ M$

여기서 승수 a, 쉬프트 (shift) c, 모듈러스 (modulus) M은 정수이다. 우리는 $a=7^5=16807$, $c=0$, $M=2^{31}-1=2147483647$를 선택했다. 이 생성자의 주기는 $2^{31}-2$이다. x_i가 $\frac{M}{2}$ 보다 크면 $b_i=1$로 정의하고 x_i가 $\frac{M}{2}$보다 크지 않으면 $b_i=0$로 정의하여 이진수열 $\{b_i\}$을 얻는다.

(ii) 선형귀환 쉬프트 레지스터(Linear feedback shift register)

쉬프트 레지스터에 의해 생성되는 이진수열은

$b_i=(a_1b_{i-1}+\cdots+a_db_{i-d})\ \mathrm{mod}\ 2$ ····· (5. 1)

에 의해 얻어진다.

(5. 1)에 다항식 $f(x)=x^d+a_1x^{d-1}+\cdots+a_d$이 대응된다. 우리는

$f(x)=x^p+x^q+1$ ····· (5. 2)

과 같은 형태의 다항식을 사용한다.

특성 다항식 (5. 2)에 대응하는 쉬프트 레지스터를 이용하여 이진수열을 생성하는 알고리즘은 다음과 같다.

Algorithm 5. 1

1. $Y\leftarrow X$(X는 b_{i+p-1}, $b_{i+p-2}\cdots b_i$의 형태로 되어 있다.)
2. Y를 q bit 만큼 오른쪽으로 이동시키고 빈자리는 0으로 채운다.
3. $Y\leftarrow X\leftarrow Y$ XOR X (여기서 *XOR*은 exclusive *OR*연산을 의미한다.)
4. Y를 $p-q$ 비트만큼 왼쪽으로 이동시키고 빈자리는 0으로 채운다.
5. $X\leftarrow Y$ XOR X (X는 다시 $b_{i+2p-1}\ b_{i+2p-2}\cdots b_{i+p}$로 구성된다.)

쉬프트 레지스터에 의해 생성되는 비트열의 주기는 2^p-1이하이다. 최대주기 2^p-1이 되도록 하는 (p, q)쌍의 예를 들어 보면 다음표와 같다.

p	q	p	q
15	1, 4, 7, 8, 11, 14	31	3, 6, 7, 13, 18, 24, 25, 28
17	3, 5, 6, 11, 12, 14	33	13, 20
18	7, 11	35	2, 33
20	3, 17	36	11, 25
21	2, 19	89	38, 51
22	1, 21	98	27, 71
23	5, 9, 14, 18	521	32, 439
25	3, 7, 18, 22	607	273, 334
28	3, 9, 13, 15, 19, 25		
29	2, 27		

이진 수열 $\{b_i\}$의 주기가 최대주기인 2^p-1이 되는 선형 쉬프트 레지스터를, 최대주기를 갖는 선형 쉬프트 레지스터라 하고 m-LESR (maximum length Linear Feedback Shift Register)이라고 정의한다.

(iii) 비선형 알고리즘

앞에서 언급한 m-LFSR 한개로 구성된 생성자의 약점을 보완하여 몇 개의 m-LFSR을 비선형 논리 구조로 결합하여 난수 생성자로 이용하는 방법이 많이 제안되었다.

(a) 승산(multiplication) 시스템

m-LFSR 2개의 출력을 서로 곱하여 최종의 출력수열을 발생하는 시스템을 승산시스템이라고 한다. m-LFSR 1의 출력수열이 $\{a_i\}$이고 m-LFSR 2의 출력수열이 $\{b_i\}$인 경우 승산시스템의 출력수열 $\{c_i\}$는

$$c_i=a_i\times b_i$$

가 된다.

(b) J-K 플립-플롭 (J-K Flip-Flop)

m-LFSR 2개의 J-K 플립-플롭에 의해 조합하여 출력수열을 발생하는 시스템이다. m-LFSR 1의 출력수열이 $\{a_i\}$이고 m-LFSR 2의 출력수열이 $\{b_i\}$일 때 J-K 플립-플롭 시스템의 출력수열은 $\{c_i\}$는

$$c_i=((a_i+b_i+1)c_{i-1}+a_i) \bmod 2,\ c_0=0$$

가 된다.

(c) Geffe 시스템

Geffe 시스템은 3개의 m-LFSR로 구성된다. m-LFSR 1, m-LFSR 2, m-LFSR 3의 출력수열을 각각 $\{a_i\}$, $\{b_i\}$, $\{c_i\}$라고 하면 Geffe 시스템의 출력수열 $\{g_i\}$는 다음과 같이 생성된다.

$$g_i=a_ib_i\oplus c_ib_i\oplus c_i$$

m-LFSR 1, m-LFSR 2, m-LFSR 3의 차수가 m, n, k이고 각 쌍마다 서로소가 될 때 Geffe 시스템에서 발생되는 출력수열의 주기는 $(2^m-1)(2^n-1)(2^k-1)$이 된다.

(d) 상호 대칭 시스템

Geffe 시스템과 마찬가지로 3개의 m-LFSR로 구성되며 m-LFSR 1, m-LFSR 2, m-LFSR 3의 출력수열을 각각 $\{a_i\}$, $\{b_i\}$, $\{c_i\}$라고 하면 시스템 출력수열 $\{s_i\}$는

$$s_i=a_ib_i\oplus b_ic_i\oplus c_ia_i$$

가 된다. m-LFSR 1, m-LFSR 2, m-LFSR 3의 차수가 각각 m, n, k이고 서로소일 때 출력되는 수열의 주기는 $(2^m-1)(2^n-1)(2^k-1)$이 된다.

(2) 결과 및 논의

위의 6가지 이진 난수발생기를 귀무가설 H_0 : BMS$\frac{1}{2}$로 설정하여 entropy에 의한 randomness 검정법으로 검정통계량 (4. 1)을 이용하였다. 〈표 2〉에서 난수 생성자 1은 선형합동법(Linear congruential generator)를 2는 선형귀환 쉬프터 레지스터를, 3은 승산 시스템을 4는 J-K 플립플롭을 5는 Geffe 시스템을 6은 상호대칭 시스템을 나타낸다. 생성자 2에 해당하는 다항식은 (5. 2)이며, 이 때

〈표 2〉 Entropy 검정에 대한 모의실험 결과표

발생기 블럭의 수(L)	1	2	3	4	5	6
8	0.5157 Pass	0.5623 Pass	81.9629	1.0729 Pass	0.8631 Pass	0.1335 Pass
9	0.5486 Pass	0.1565 Pass	90.9500	0.9609 Pass	0.8972 Pass	0.4181 Pass
10	1.3476 Pass	0.0682 Pass	100.8102	0.0587 Pass	0.2593 Pass	0.1121 Pass
11	0.7509 Pass	0.0964 Pass	112.0477	0.3316 Pass	0.3736 Pass	0.1570 Pass
12	0.6861 Pass	1.1293 Pass	122.7078	1.1927 Pass	0.9653 Pass	1.4350 Pass
13	0.5567 Pass	0.7932 Pass	132.2253	0.2737 Pass	0.1721 Pass	0.1321 Pass
14	0.0361 Pass	0.3607 Pass	148.6450	0.0998 Pass	0.5972 Pass	0.8375 Pass
15	0.4421 Pass	1.4815 Pass	156.5084	0.0418 Pass	0.9124 Pass	1.2133 Pass
16	1.3233 Pass	1.2649 Pass	163.9852	0.6239 Pass	0.1429 Pass	0.2817 Pass

난수발생기 1. 선형합동법
2. 선형귀환 쉬프트 레지스터
3. 승산시스템
4. J-K 플립-플롭시스템
5. Geffe 시스템
6. 상호대칭시스템

사용한 비트의 길이 :
$8 \leqslant L \leqslant 10$: $(30 \times 2^L + 10000) \times L$
$11 \leqslant L \leqslant 16$: $(5 \times 2^L + 10000) \times L$
유의수준 : 5%

$p=28$, $q=9$를 사용하였다. 이 때 생성되는 이진 수열의 주기는 $2^{28}-1=268435455$이다. 생성자 3과 4에서 사용된 선형귀환 쉬프트 레지스터는 $(p, q)=(31, 13)$과 $(p, q)=(33, 13)$인 경우이며 생성자 5와 6에 사용된 선형귀환 쉬프트 레지스터는 $(p, q)=(31, 13)$ $(p, q)=(33, 13)$ 그리고 $(p, q)=(28, 9)$인 경우이다.

Block의 크기는 8부터 16까지에 대하여 적용하였으며, 초기화를 위한 block의 수 Q는 L이 8과 11 사이에 있을 때는 $Q=30 \cdot 2^L$개를 L이 12와 16 사이에 있을 때는 $Q=5 \cdot 2^L$개를 사용하였다. 검정을 위한 block의 수는 $K=10000$개를 사용하였다. 그러므로 각각의 경우 사용된 총비트의 수는 $N=(Q+K)L$이다.

귀무가설 H_0 : 난수 발생기의 source가 BMS $\frac{1}{2}$ 이다.

위의 가설검정을 위하여 사용하는 검정통계량 $T(s^N)$은 (4. 1) 식이며 기각역은

$$\left|\frac{T(s^N) - E(T(U^N_{\mathrm{BMS}_{\frac{1}{2}}}))}{\sqrt{\mathrm{Var}\,(T(U^N_{\mathrm{BMS}_{\frac{1}{2}}}))}}\right| \geq z_{\alpha}$$

이다. 여기서 $E(T(U_{\mathrm{BMS}^n\frac{1}{2}}))$과 Var $(T(U_{\mathrm{BMS}^N\frac{1}{2}}))$는 〈표 1〉에서 얻는다. 이때 이 검정법을 통과하였다함은 source가 BMS $\frac{1}{2}$ 임을 의미하므로 생성된 수열은 독립이고 일양분포로 부터 나온 것으로 볼 수 있다.

이 때 승산 시스템을 제외한 모든 난수 발생기가 5% 유의수준에서 entropy에 의한 검정을 통과하였다. 기존의 전통적인 통계적 검정방법에 비하여 entropy에 의한 검정방법은 source가 BMS $\frac{1}{2}$ 인가 하는 것을 검정하므로 비트간의 독립성과 빈도의 안정성을 동시에 검정할 수 있고, 또한 암호학적으로 중요한 측도인 비트당 entropy와 관련이 있다는 점에서 새로운 randomness의 검정법으로 타당하다.

참 고 문 헌

1. H. Beker and F. Piper, Ciper Systems The Protection of Communicatins, John Wily and Sons, 1982.

2. P. Bratky, B. L. Fox and L. F. Schrage, A Guide to Simulation, Springer-Verlag. 1983.

3. S. W. Golombo, Shift Register Sequences, Holden-Day, 1967.

4. I. J. Good, On the serial tests for random sequences, Ann. Math. statist., 28, 262-264, 1967.

5. A. I. Khinchin, Mathematical Foundation of Information Theory, Dover Pubulication, Inc., 1957.

6. D. E. Knuth, The Art of Computer Programming, Vol. 2, Semi Numerical Algorithms, Addison-Wesley Publishing Company, 1981.

7. A. N. Kolmogorov and V. A. Uspenskii, Algorithms and Randomness, Theory Prob. Appl., Vol. 32, No. 3, 389-412.

8. J. Lehoczky, Statistical Methods, Handbooks in Operations Research and Management Science, Vol. 2, Stochastic Models (ed) by D. P. Heyman & M. J. Sobel, pp. 255-294, North-Holland, 1990.

9. U. M. Maurer, A universal statistical test for random bit generators, Crypto '90, 401-413.

10. A. M. Mood, The distribution theory of runs, Ann. Math. Statist., 11, 367-392, 1940.

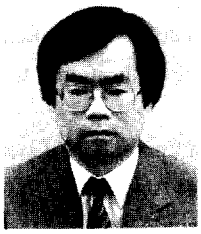
11. B. Riphey, Stochastic Simulation, John Wiley & Sons, 1987.

12. C. E. Shanon, A mathematical theroy of communication, Bell Syst. Tech. J. Vol. 27, 379-423, 623-656, 1948.

13. J. M. Wozen Craft and B. Reiffen, sequential Decoding, Cambridge, MA. Techn. Press of the MIT, 1960.

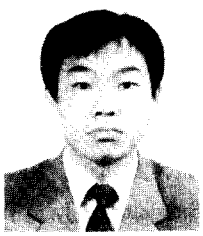
14. 현대 암호학, 한국전자통신연구소편저, 1991.

□ 著者紹介



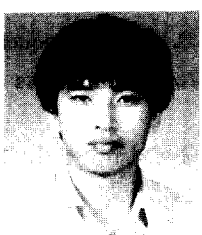
崔 鳳 大(正會員)

慶北大學校 師範大學 數學科(理學士)
慶北大學校 大學院 數學科(理學碩士)
Ohio State University 大學院 數學科(理學博士)
University of North Carolina 訪問教授
慶北大學校 數學科, 專任講師, 助教授
韓國科學技術院 數學科, 助教授, 副教授, 教授
현재 韓國科學技術院 數學科 教授
大韓數學會 編輯理事



신 양 우(正會員)

慶北大學校 自然科學大學 數學科(理學士)
韓國科學技術院 應用數學科(理學碩士)
韓國科學技術院 數學科(理學博士)
韓國科學技術院 應用數學科 助教
현재 昌原大學校 自然科學大學 統計學科 專任講師



이 경 현(正會員)

慶北大學校 師範大學 數學教育科(理學士)
韓國科學技術院 應用數學科(理學碩士)
현재 韓國科學技術院 數學科 博士課程在學中
韓國電子通信研究所 先任研究員