Journal of Korean
Data & Information Science Society
2002, Vol. 13, No.2 pp. 77 85

Use of the estimated critical values adapting a regression equation for the approximate entropy test[1)]

Kyung-joon Cha[2)] · Je-seon Ryu[3)]

Abstract

The statistical testing methods have been widely recognized to determine the plain and cipher texts. In fact, the randomness for a sequence from an encryption algorithm is necessary to guarantee security and reliance of cipher algorithm. Thus, the statistical randomness tests are used to discover cipher text. In this paper, we would provide the critical value for an approximate entropy test by estimating the nonlinear regression equation when the number of sequence and the level of significance are given. Thus, we can discern plan and cipher text for real problem with given number of sequence and the level of significance. Also, we confirm the fitness of the estimated critical values from the rate of success for plain or cipher text.

주제어 : 근사 엔트로피 검정, 평문, 암호문, 임계값, 회귀식

1. 서론

최근의 많은 과학기술의 응용 기법에 있어 난수열 혹은 모의난수열에 대한 요구가 급증되고 있다. 예를 들어 암호시스템에서의 비밀키는 보통 난수성을 따르는 형태로 생성되어야 한다. 또한 많은 암호화 단계에서의 프로토콜에서 이러한 난수(random number) 혹은 모의난수(pseudo-random number, quasi-random number)가 요구되고 있다. 이는 암호 알고리즘의 설계자와 사용자가 암호체계를 사용하기에 앞서 제3자의 암호해독 공격으로부터 안전성과 신뢰성을 보장하기 위해 난수성을 만족하는 이진 수열을 이용하기 때문이다. 또한, 암호 알고리즘 설계 시에 출력된 암호문이 난수성을 만족할 때 좋은 암호 알고리즘으로 판정되고 있으며

1. "This work was supported by the research fund of Hanyang University(HY-2002)"

2. Professor, Dept. of Mathematics, Hanyang University, 17 Haengdang-dong, Seongdong-Gu, Seoul, 133-791, Korea.
E-mail : kjcha@hanyang.ac.kr
3. Ph.D student, Dept. of Mathematics, Hanyang University, 17 Haengdang-dong, Seongdong-Gu, Seoul, 133-791, Korea.

난수성을 검정하기 위해 암호 알고리즘을 통해 출력된 결과의 통계적인 성질을 살펴봄으로서 판정할 수 있다.

주어진 이진 수열에 대한 난수성의 정의를 통계적 관점에서 볼 때 그 수열이 갖고 있는 각 항들 사이의 독립성을 보장하기 위한 상호 연관성(association)이나 빈도의 편향성(unbiasedness) 등의 통계적 성질을 찾아내는데 있다. 즉, 한 수열이 통계적 검정을 통과하였다고 해서 그 수열이 랜덤하다는 것을 의미하는 것이 아니라 검정방법의 성질을 만족하는 난수적 성질을 지니고 있다는 것을 의미한다. 따라서, 여러 가지 통계적 검정을 통과하였을 때 그 수열의 난수성에 더 많은 신뢰성을 부여할 수 있을 것이다. 기존에 사용된 통계적 검정방법으로는 Beker와 Piper (1982)의 frequency test, Mood (1940)의 runs distribution test, Good (1953, 1957)의 serial test 등이 있으며 그 외의 몇 가지 검정방법이 Sheskin (1997)과 Gibbons (1985)에 소개되어 있다. 또한 박홍구 등(2000)과 한국정보보호진흥원 (1999)은 이러한 통계적 방법들을 이용하여 이진 수열에 대한 통계적 검정에서의 최소한의 표본크기에 관하여 연구를 하였다.

엔트로피 관점에서 이진 수열에 대한 난수성을 고려해 보면, 암호학적으로 중요한 정량화된 측도인 비트 당 엔트로피를 추정하여 수열의 난수성에 대하여 검정하는 데에 있다. 엔트로피 검정방법은 기존의 통계적 관점의 검정에 비해 좀더 일반적인 통계적 모형을 기반으로 하며, 공격자가 비밀키 소스의 통계적 결점에 관한 지식을 부당하게 이용할 때 공격자가 수행하는 최적키 탐색 전략의 시간과 관련이 있는 소스의 비트 당 엔트로피를 측정한다.(Maurer, 1992 참조) 또한, 다른 길이의 인접한 부분수열의 빈도를 이용하여 주어진 수열의 길이와 부분수열의 길이에 대하여 엔트로피 값을 추정함으로서 주어진 수열의 추정된 엔트로피의 편차를 측정한다.(Pincus, 1991, FIPS, 1997 참조)

본 논문에서는 이진 수열이 갖고 있는 난수성 판정에 관하여 근사 엔트로피 검정방법을 적용한다. 2장에서는 보편적으로 많이 사용되고 있는 엔트로피 정량화 방법인 근사 엔트로피 검정방법을 주어진 암호문과 평문 이진 수열에 적용하는 방법에 대해서 설명한다. 주어진 수열의 길이와 유의수준에서, 모의실험을 통한 임계값을 계산하고, 이를 이용하여 임의의 수열에서의 근사 엔트로피 임계값을 결정하기 위해 비선형 회귀식을 추정한다. 난수성 판정을 위한 이진수열에 대하여 6개의 임의의 문서파일과 1개씩의 실행파일 및 압축파일을 선정하고, 이를 난수성이 만족되지 않는 수열로서 평문으로 가정한다. 임의의 키와 위에서 주어진 평문들을 이용하여 DES(Data Encryption Standard)(FIPS, 1997, 현대암호학, 1991 참조)를 CBC(Cipher Block Chaining) 모드로 생성한 수열을, 난수성을 만족하는 암호문으로 가정한다. 3장에서는 각각의 수열에서 근사 엔트로피 통계량을 계산하고, 2장에서 결정된 실제의 임계값과 추정된 임계값을 각각 비교하여, 이진 수열에 대한 난수성을 검증함과 동시에, 추정된 임계값에 대한 적합성을 확인한다. 4장에서는 제안된 근사 엔트로피 방법의 임계값 결정 및 실험결과 등에 대한 전체적인 결론을 제시한다.

2. 분석 및 비교 방법

2.1 자료의 선정

본 연구에서 데이터는 임의로 선택한 파일에서 생성된 이진 수열을 이용하여 암호 시스템을 이용한 새로운 이진 수열을 추출해내는 과정으로 연결된다. 실제 자료에 대한 공정성을 고려하면서 길이가 130kb인 임의의 파일을 선택하였다. 파일은 종류별로 6개의 문서파일(cpp,

doc, hwp, pdf, ps, tex)과 1개씩의 실행파일(exe)과 압축파일(zip)을 무작위로 선정하였다. Visual C++ 프로그램을 이용하여 위의 파일들에 대한 이진 수열을 추출하고, 이를 평문으로 정의한다. 다음에 DES를 CBC모드로 적용하여 새로운 이진 수열을 생성하고, 이를 암호문으로 정의한다. 생성된 각각의 암호문과 평문을 근사 엔트로피 검정방법에 적용하여 통계량을 계산하고, 임계값과 비교하여 난수성을 만족하는지의 여부에 따라, 평문과 암호문에 대한 성공률을 계산한다.

2.2 근사 엔트로피 검정방법

Pincus (1991)와 Pincus와 Singer (1996)는 근사 엔트로피를 이용하여 불규칙성의 정도에 관하여 일련의 자료에서의 난수성의 척도의 개념을 공식화하였다. 근사 엔트로피의 개념은 다음과 같다.

양의 정수 N과 $m \le N$인 비음 정수 m이 주어졌을 때 양의 실수 r과 실수열 $u^N : (u(1), u(2), \cdots, u(N))$에 관하여

$$x(i) : (u(i), u(i+1), \cdots, u(i+m-1))$$
$$x(j) : (u(j), u(j+1), \cdots, u(j+m-1))$$

이라 하면, m 블록 $x(i)$와 $x(j)$ 사이의 거리는

$$d(x(i), x(j)) = Max_{k=1,\cdots,m} | u(i+k-1) - u(j+k-1) |$$

으로 정의된다. 또한,

$$Q_i^m(r) = \frac{d(x(i), x(j)) \le r\text{인 } j \le N-m+1\text{의 개수}}{N-m+1}$$

라고 하면, $Q_i^m(r)$은 주어진 블록과 근사적으로 같은, 블록들의 상대적 빈도라고 할 수 있다.

Pincus와 Singer (1996)는 수열 u의 근사 엔트로피를 측정하는 계산적 방법으로서

$$ApEn(m, r, N)(u) = \phi^m(r) - \phi^{m+1}(r), \quad m \ge 1$$

을 제시하였으며, 여기서

$$\phi^m(r) = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} \log_2 Q_i^m(r)$$

이다.

0과 1만을 포함하는 이진 수열에서, $r < 1$이며, 두 m 블록 사이의 거리 $d(x(i), x(j))$는 0과 1의 값만을 갖는다. 따라서, $r = 0$으로 제한할 수 있으며, 근사 엔트로피 통계량 $ApEn$은 r과 독립적으로 사용될 수 있다.

Chatterjee외 3인(2000)에 의하면, 수열 u^N이 N 랜덤하다는 것은 u^N 수열이 $m = 0, 1, \cdots, m_{crit}(N)$에 대하여 $\{m, N\}$ 랜덤하다는 것이다. 이때, $m_{crit}(N) = \max(m ; 2^{2^m} \le N)$이다. 따라서 길이가 m_{crit}인 블록으로 길이가 N인 이진 수열의 근사 엔트로피 비율을 추정하는 것이 가능하다.

이제, 관측된 이진 수열이 성공률 p인 베르누이 확률과정에 의해 생성되었을 때, 그 수열의 랜덤성을 결정하기 위해 사용되는 엔트로피 검정을 유도할 것이다. 성공률이 p인 p 랜덤 베르누이 수열에서 $\psi^m(p)$를 ϕ^m의 기대 엔트로피라고 하면

$$\phi^m(p) = \sum_{t=0}^{m} \binom{m}{t} p^{m-t}(1-p)^t \log p^{m-t}(1-p)^t \quad (1)$$

으로 정의된다. 이때, $T_1(p,m)$을 (추정된) 엔트로피 ϕ^{m+1}과 식(1)에서 계산된 기대 엔트로피 ψ^{m+1}의 평균편차라고 정의하자. 즉,

$$T_1(p,m) = \frac{\phi^{m+1} - \psi^{m+1}(p)}{m+1} \quad (2)$$

일 때, $T_1(p,m)$은 p 랜덤 베르누이 과정에 의해 생성된 유한 수열에 대하여 0에 가까운 값을 갖게 된다. 즉, $T_1(p,m)$ 값이 작으면 랜덤한 수열로, 큰 값을 가지면 랜덤하지 않은 수열로 해석할 수 있다. 식(2)에서 $p = 0.5$인 베르누이 수열의 모든 m에 대하여 $\psi^m(p) = -m\log 2$이므로

$$T_1(0.5,m) = \frac{\phi^{m+1} + m\log 2}{m+1} \quad (3)$$

이다. 따라서, 식(3)에서의 통계량 $T_1(0.5,m)$을 고려했을 때, 가설검정을 하기 위한 귀무가설과 대립가설을 각각

H_0 ; 수열이 랜덤하다

H_1 ; 수열이 랜덤하지 않다

라고 하면, 주어진 m과 유의수준에서 $T_1(0.5,m)$값이 커질 때, 귀무가설 H_0을 기각할 수 있다. 따라서, H_0이 참일 때, 관측된 m 블록의 빈도는 수열의 길이 N이 증가하면 기대 빈도로 근사한다. 즉, $N \to \infty$일 때, $T_1(p)$는 확률적으로 0에 근사한다. 결국, N이 증가할 때 관측된 엔트로피 ϕ^{m+1}은 기대 엔트로피 ψ^{m+1}에 근사하며, 식(2)에서의 분자는 0에 근사한다.

2.3 근사 엔트로피 검정에서 임계값 결정

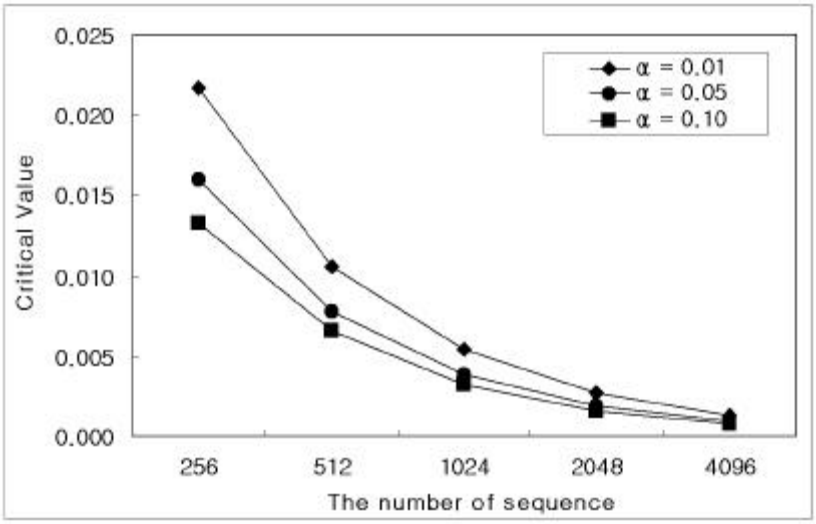
Chatterjee외 3인(2000)을 보면, 식(3)에서의 $T_1(p,m)$의 이론적인 분포는 쉽게 다룰 수 없으므로 5000번의 실험을 통하여 $p = 0.4$와 $p = 0.5$, 그리고 유의수준이 $\alpha = 0.05$에 대하여 $T_1(p,m)$의 임계값을 얻을 수 있었다. 본 연구에서는 $T_1(0.5,m)$에 대한 0.01, 0.05, 그리고 0.10의 유의수준과 임의의 이진 수열의 길이 N에 따른 회귀방정식을 이용한 추정 임계값을 제공하고자 한다. 이러한 문제를 해결하기 위해 $p = 0.5$인 베르누이 확률과정을 고려하였고, 실제 이진 수열의 랜덤성을 조사하기 위해 길이가 $N = 2^8, 2^9, 2^{10}, 2^{11}, 2^{12}$인 각각의 랜덤한 이진 수열을 가지고 10000번의 실험 후 각각의 유의수준 $\alpha = 0.01, 0.05, 0.10$에 대한 임계값을 생성하였으며, 이때의 통계량은

$$T_1(0.5,m) = \frac{\phi^{m+1} - \psi^{m+1}(0.5)}{m+1}$$

을 사용하였다. <표 1>은 임의의 이진 수열에 대하여 각각의 수열의 길이 N과 유의수준에 대한 10000번의 실험을 수행한 T_1의 임계값을 보여주고 있다.

<표 1> 수열의 개수 $N = 2^8, 2^9, 2^{10}, 2^{11}, 2^{12}$과 유의수준 $\alpha = 0.01, 0.05, 0.10$이 주어졌을 때, 10000번의 실험 후 결정된 근사 엔트로피 통계량 T_1의 임계값

N \ α	0.01	0.05	0.10
256	0.0216612220	0.0160188068	0.0132444933
512	0.0106148301	0.0077741937	0.0065182655
1024	0.0053803610	0.0038349925	0.0032030667
2048	0.0026892232	0.0019269289	0.0016101718
4096	0.0013132284	0.0009617510	0.0008091151



<그림 1> 수열의 길이 $N = 2^8, 2^9, 2^{10}, 2^{11}, 2^{12}$과 유의수준 $\alpha = 0,01, 0.05, 0.10$이 주어졌을 때, 계산된 임계값의 산점도

<그림 1>을 통하여 임의의 수열의 길이 N과 유의수준에 대한 임계값을 추정하기 위해 비선형 회귀식으로

$$T_1 = \beta_0 e^{-\beta_1 \log_2 N} + \varepsilon$$

을 사용하였으며, 밑수가 2인 로그함수는 본 논문에서의 수열의 길이 N이 2의 지수로 결정되었기 때문이다. 여기서 β_0와 β_1은 회귀계수이고 ε은 평균이 0이고 분산이 σ^2인 정규분포를 따르는 오차이다. 실제로 Chatterjee외 3인(2000)은 상수항이 있는 모형

$$T_1 = \alpha + \beta e^{-\gamma N} + \varepsilon$$

을 제안하였는데, 이는 더 많은 수열에 대하여 임계값의 상한이 결정되어 유의한 임계값을 얻을 수 없으므로, 본 연구에서는 상수항이 없는 모형을 적용하였다.

실험에서 구한 임계값에 관한 비선형 회귀모형식을 최소제곱법으로 적용해 보면, 임의의 수열의 길이 N과 각각의 유의수준에 대하여

$$\alpha = 0.01,\quad \widehat{T_1} = 5.741778468e^{-0.69791\log_2 N},\ (R^2 = 0.99)$$

$$\alpha = 0.05,\quad \widehat{T_1} = 4.348308853e^{-0.70204\log_2 N},\ (R^2 = 0.99),$$

$$\alpha = 0.10,\quad \widehat{T_1} = 3.520732273e^{-0.69891\log_2 N},\ (R^2 = 0.99)$$

으로 추정되었다. 여기서 R^2은 결정계수이며, Chatterjee 외 3인(2000)의 결과보다 더 정확한 것으로 판정되었다.

이러한 결과로서, 임의의 수열이 주어졌을 때 수열의 길이 N과 유의수준이 결정되고 근사 엔트로피 통계량 T_1을 계산하여 추정된 임계값과 비교한 후, 주어진 이진 수열에 대한 난수성을 판단할 수 있는 것이다. 실제로, 주어진 수열의 길이가 위의 실험과 같은 경우 각각의 유의수준에 대하여 추정된 임계값을 <표 2>에서와 같이 얻을 수 있다. <표 1>과 <표 2>를 보면 회귀식을 이용하여 추정된 임계값이 모의실험을 통하여 얻은 임계값과 거의 차이가 없음을 알 수 있다.

3장에서는 5개의 문서파일에 대한 이진 수열의 근사 엔트로피 통계량을 계산하고, <표 1>에서 얻어진 실제의 임계값과 <표 2>에서 얻어진 추정된 임계값을 이용하여, 수열의 난수성을 판단하고자 한다. 평문이 평문으로, 그리고 암호문이 암호문으로 판정되는 성공률에 대한 차이를 비교하여, 임계값에 따른 차이가 없을 때, 모형에서의 적합도를 바탕으로($R^2 = 0.9999$), 추정된 임계값이 실제의 임계값을 대신할 수 있으리라는 결론을 얻을 수 있을 것이다.

<표 2> 수열의 개수 $N = 2^8, 2^9, 2^{10}, 2^{11}, 2^{12}$과 유의수준 $\alpha = 0.01, 0.05, 0.10$이 주어졌을 때, 회귀방정식을 이용한 근사 엔트로피 통계량 T_1의 추정된 임계값

N \ α	0.01	0.05	0.10
256	0.021590303	0.015819167	0.013133212
512	0.010743858	0.007839557	0.006528873
1024	0.005346405	0.003885075	0.003245678
2048	0.002660501	0.001925339	0.001613514
4096	0.001323930	0.000954147	0.000802121

<표 3> 선택된 문서파일에 대한 근사 엔트로피 검정에서의 평균 T_1 값과 성공률

	평문			암호문		
파일	평균 T_1	계산된 임계값의 성공률(%)	추정된 임계값의 성공률(%)	평균 T_1	계산된 임계값의 성공률(%)	추정된 임계값의 성공률(%)
cpp	0.011290830	100	100	0.000463782	96	95
doc	0.108111400	100	100	0.000459994	97	97
exe	0.061973840	100	100	0.000492556	95	95
hwp	0.012048764	52	53	0.000443164	96	96
pdf	0.001535578	47	47	0.000466323	95	95
ps	0.053868700	100	100	0.000399024	98	98
tex	0.011387640	100	100	0.000460892	98	98
zip	0.072551400	100	100	0.000447775	97	97

3. 결과 및 분석

아래의 결과는 선택된 8개의 파일에 대하여 생성된 평문과 암호문을 이용하여 얻어진 근사 엔트로피 검정방법의 결과를 보여주고 있다.

근사 엔트로피 검정을 수행하기 위해, 길이가 $N = 2^{12} = 4096$인 전체수열을 100개로 정하였다. 각각의 길이가 $N = 4096$인 전체수열에 대하여 $m_{crit} = 3$으로 고려할 수 있으며, 이진 수열의 엔트로피 값 ϕ^{m+1}을 추정한다. 이를 이용하여 식(3)에 있는 통계량 $T_1(0.5, m)$의 값을 구하고, <표 1>과 <표 2>에서 제공된 임계값과 비교하여 암호문과 평문에 대한 판정을 한다. 이때, 근사 엔트로피 통계량 $T_1(0.5, m)$의 값이 작으면 랜덤한 수열, 즉 암호문으로, 크면 랜덤하지 않은 수열, 즉 평문으로 판정한다. 유의수준이 $\alpha = 0.05$일 때의 임계값을 이용하였으며, 실험된 이진 수열의 평문, 암호문 성공률에 기인하여 실제의 임계값에 대한 추정된 임계값의 유의성을 확인한다. 본 연구에서는 근사 엔트로피 검정을 위해 S-plus2000을 이용하였다.

위의 <표 3>은 선택된 5개의 파일에 대하여 근사 엔트로피 검정을 수행한 결과를 보여주고 있다. 주어진 전체수열 100개에 대하여 각각의 근사 엔트로피 통계량을 구하고, 그에 대한 평균 T_1 값과 성공률을 나타내고 있다. 평균 T_1 값은 각 파일에 대하여 계산된 100개의 통계량 $T_1(0.5, m_{crit} = 3)$들의 평균값이며, 성공률은 100개의 주어진 수열에 대하여 평문이 평문으로 판정되는, 그리고 암호문이 암호문으로 판정되는 비율을 나타낸다.

<표 3>을 보면 평문에서의 평균 T_1 값은 0.001535578에서 0.108111400으로서 암호문에 비해 상당히 높고 불안정하다고 볼 수 있다. 또한, 성공률에 있어서 pdf와 hwp 파일을 제외한 모든 파일에서 100%의 성공률을 보이고 있다. pdf와 hwp 파일에 대한 이진 수열에 대하여 계산된 엔트로피가 추정된 엔트로피와의 편차가 작은 부분수열이 많이 있는 것으로 판단되며, 결국 성공률이 낮아지게 된 것으로 해석할 수 있다. 또

한, 암호문에서의 평균 T_1 값은 0.000399024에서 0.000492556으로 상당히 낮고 안정되어 있다고 할 수 있다. 성공률에 있어서도 모든 파일에서 95% 이상으로 안정적이라는 것을 볼 수 있다.

<표 1>에서의 계산된 임계값과 <표 2>에서의 추정된 임계값을 이용하여 각각의 이진 수열에 대한 성공률을 비교한 결과에서는, cpp 파일에 대한 암호문 수열에서 각각 96%와 95%로, hwp 파일에 대한 평문 수열에서 52%와 53%로 약간의 차이가 있었을 뿐, 다른 모든 이진 수열에서는 같은 성공률을 보이고 있다. 따라서, 비선형 회귀식의 적합도(R^2)가 모두 0.99로서 상당히 높게 안정되어 있으며, 실제의 이진수열에 대한 성공률 비교에서도 거의 차이가 없는 것으로 미루어, 본 연구에서 제안된 비선형 회귀식을 근사 엔트로피 검정을 위한 임계값의 추정을 위하여 유용하게 사용할 수 있을 것으로 보인다.

4. 결론

본 연구에서는 근사 엔트로피 통계량에 대한 임계값을 결정하기 위해 수열의 길이 N 이 $N = 2^8, 2^9, 2^{10}, 2^{11}, 2^{12}$ 일 때와 유의수준 α 가 $\alpha = 0,01,\ 0.05$, 그리고 0.10 일 때의 임계값을, 10000번의 모의실험을 통해 계산하였으며, 이를 이용하여 각각의 유의수준에 따른 비선형 회귀모형을 최소제곱법으로 추정하였다.

실험에 사용되는 이진 수열로서 6개의 주어진 문서파일(cpp, doc, hwp, pdf, ps, tex)과 실행파일(exe) 그리고 압축파일(zip)에서 생성된 이진 수열(평문)과 DES의 CBC 모드를 통해 생성된 새로운 이진 수열(암호문)에 대하여 각각의 길이가 $N = 4096$ 인 수열 100개를 선택하였으며, 조건에 의하여 길이 $m_{crit} = 3$ 인 블록을 이용하여 근사 엔트로피 통계량을 계산했다. 유의수준이 $\alpha = 0.05$ 일 때, <표 1>의 계산된 임계값과 <표 2>의 추정된 임계값을 이용하여 계산된 근사 엔트로피 통계량에 대한 난수성 검정을 시행하였으며, 평문과 암호문을 판정하는 성공률에 있어서 거의 차이가 없다는 결과를 얻었다. 또한, 비선형 회귀식의 적합도가 상당히 높았으므로 ($R^2 = 0.99$), 본 연구에서 제안된 근사엔트로피 임계값의 비선형 방정식은 임계값의 추정에 유용하게 사용될 수 있으리라 사료된다.

근사 엔트로피 방법은 Pincus (1991)에 의해 시스템 내에서의 복잡성을 표현하는 정량화 측도로서 제안되었다. Rukhin (2000)은 근사 엔트로피 통계량의 극한적 분포를 이용하여 χ^2 -분포 가정을 하였다. Chatterjee 외 3인(2000)과 본 논문에서 제시하고 있는 근사 엔트로피 검정방법은 비모수적 방법으로서, 모의실험을 통한 근사 엔트로피 통계량의 임계값을 결정하는 것이다. 또한, Chatterjee 외 3인(2000)에서는 임의의 이진 수열에 대한 근사 엔트로피 검정방법에 관하여 소개하고 있으나, 본 논문은 실제의 다양한 문서파일의 평문과 암호문 이진수열을 생성하여 적용하였다는 점에서 그 중요성이 있다고 할 수 있다. 특히, 보다 정확한 임계값에 대한 비선형 회귀모형을 제공하고 있으며, 이를 실제의 임계값과 비교하여 거의 차이가 없다는 사실도 실험으로 확인하였다.

평문과 암호문을 판정하는 알고리즘에서 실시간 분석을 할 때, 이진 수열에 대한 난수성을 보다 정밀하게 판단하기 위해서는 많은 길이의 수열이, 정밀성보다는 신속하게 판단하기 위해서는 적은 길이의 수열이 요구되어야 할 것이다. 또한, 유의수준에

따른 유의한 판정을 원하는 경우도 있을 것이다. 이와 같이 수열의 크기와 유의수준을 다양한 기준에 따라 정하게 될 때, 본 연구에서 계산된 근사 엔트로피 임계값에 관한 비선형 회귀식은 매우 유용하게 사용될 수 있으며, 안전성과 신뢰성을 보장하기 위한 암호 알고리즘의 개발에도 중요한 역할을 할 것으로 기대된다.

참고문헌

1. 박홍구, 차경준, 장호종, 송정환, 박성준, "실험계획법을 이용한 평문, 암호문 식별방법의 검정크기에 관한 연구", 통신정보보호학회논문지, 19, 4, pp.71-84, 2000.
2. 한국정보보호진흥원, "암호통신문 탐지기술 연구 및 S/W 개발", 1999.
3. 현대암호학, 한국전자 통신연구소, pp.58-66, 1991.
4. Beker, H. and Piper, F. (1982). *Cipher System: The Protection of Communications*, John Wiley & Sons, New York.
5. Chatterjee, S., Yilmaz, M.R., Habibullah, M. and Laudato, M. (2000) An approximate entropy test for randomness, *Commun. Statist.-Theory Meth.*, 29, 3, 655-675.
6. FIPS, *Data Encryption Standard*, Federal Information Processing Standard(FIPS) Publication, 46, National Bureau of Standards, Washington DC., 1977.
7. Gibbons, J. D. (1985). *Nonparametric Statistical Inference*, 2nd. edition, Marcel Dekker Inc, NewYork.
8. Good, I. J. (1953). The Serial Test for Sampling Numbers and Other Tests for Randomness, *Proceedings of the Cambridge Philosophical Society*, 49, 276-284.
9. Good, I. J. (1957). On the Serial Test for Random Sequences, *Annals of Mathematical Statistics*, 28, 262-264.
10. Maurer, U. M. (1992). A universal statistical test for random bit generator", *Journal of Cryptology*, 5, 2, 89-105.
11. Mood, A. M. (1940). The distribution Theory of Runs, *Annals of Mathematical Statistics*, 11, 367-392.
12. Pincus, S. (1991). Approximate entropy as a measure of system complexity, *Proceedings of the National Academy of Sciences*, 88, 2297-2301.
13. Pincus, S. and Singer, B. H. (1996). Randomness and degrees of irregularity, *Proceedings of the National Academy of Sciences*, 93, 2083-2088.
14. Rukhin, A. (2000). Approximate entropy for testing randomness, *Journal of Applied Probability*, 37, 88-100.
15. Sheskin, D. J. (1997). *Handbook of Parametric and Nonparametric Statistical Procedures*, CRC Press Inc.15.

[2002년 4월 접수, 2002년 9월 채택]